Universidade do Minho
Escola de Direito

Victor Moreira Mulin Leal

O Regulamento Geral sobre a Proteção de Dados como um instrumento de regulação de uma inteligência artificial de confiança à luz das orientações éticas da Comissão Europeia

Junho de 2021

O Regulamento Geral sobre a Proteção de Dados como um instrumento de regulação de uma inteligência artificial de confiança à luz das orientações éticas da Comissão Europeia
Victor Moreira Mulin Leal
UMinho | 2021

Universidade do Minho
Escola de Direito

Victor Moreira Mulin Leal

O Regulamento Geral sobre a Proteção de Dados como um instrumento de regulação de uma inteligência artificial de confiança à luz das orientações éticas da Comissão Europeia

Dissertação de Mestrado

Mestrado em Direito e Informática

Trabalho realizado sob a orientação de
Profª Doutora Joana Abreu
Professor Doutor Paulo Novais

Junho de 2021

Agradecimentos:

No momento em que eu escrevo esse agradecimento, passaram-se aproximadamente 2 anos e 8 meses do dia em que eu saí do Brasil para iniciar o Mestrado em Portugal. E de certa forma, eu sinto que eu devo agradecer a cada pessoa que cruzou o meu destino durante todo esse período. É claro que a escrita da dissertação não levou tanto tempo, mas cada experiência que eu tive, foi decisiva para o Autor alcançar a maturidade que tem hoje, o que reflete diretamente na dissertação.

Um primeiro agradecimento vai ao meu pai e minha mãe, Eli Mulin Leal e Benilda Moreira Leal, que apesar da distância física nunca permitiram que eu me sentisse sozinho, mesmo existindo um oceano entre nós.

Também devo agradecer aos meus irmãos, Myller e Allan, que sempre me deram força para que eu persistisse no caminho que eu acreditava. Um agradecimento aos meus sobrinhos, Victor Hugo, Alice e Elisa (e agora o Lucas), que passavam me tranquilidade toda a vez que eu pensava neles.

Impossível esquecer da Mariana Costa, pois se não fosse pela sua força, coragem, sinceridade, paciência, suporte, carinho, atenção (e outras mil qualidades), eu talvez sequer teria saído do Brasil, e nesse momento poderia estar vivendo uma vida totalmente diferente.

Agora eis a questão: como agradecer aos amigos? Essa experiência de 2 anos e meio, 4 países e 8 residências me trouxe dezenas de inesquecíveis amigos, como: Dude, Rafael, Rael, Diego, Matheus, Marília, Tayrone, Vera, Mara, Lucas, Raphael, Bruno, João, Diogo, Ygor, Tahena, Gabriela, Lícia, Pollyne, Alfredo, Dona Hermínia, Cookie, Francisco, Tobias, Mathew, Isabela, Ana, Letícia, Nemailla, Dima, Caio, Lídia e Shrada. Todas essas pessoas, e outras mais que não pude lembrar, me ajudaram direta ou indiretamente a superar muitos obstáculos. E por isso, eternizo aqui essa menção.

Por fim, também gostaria de agradecer aos meus orientadores, que mostraram o caminho e permitiram que eu seguisse em frente na busca gradual pelos meus objetivos.

DECLARAÇÃO DE INTEGRIDADE

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri à prática de plágio nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração.

Mais, declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho.

O REGULAMENTO GERAL SOBRE A PROTEÇÃO DE DADOS COMO UM INSTRUMENTO DE REGULAÇÃO DE UMA INTELIGÊNCIA ARTIFICIAL DE CONFIANÇA À LUZ DAS ORIENTAÇÕES ÉTICAS DA COMISSÃO EUROPEIA

RESUMO:

É inegável que a inteligência artificial (IA) gera importantes oportunidades para a sociedade, como o aumento do bem-estar dos indivíduos e o avanço da economia. Contudo, se mal utilizada, a IA também pode provocar riscos aos interesses da sociedade e aos direitos fundamentais dos indivíduos. A ausência de uma regulação específica voltada para a IA torna esse cenário ainda mais complexo, uma vez que os danos à privacidade, equidade e segurança podem ser irreparáveis. A partir desse contexto, princípios éticos e de direitos se apresentam para construir uma IA ética e de confiança. No entanto, ainda pela ausência de uma legislação específica, torna-se importante perceber se o Regulamento Geral de Proteção de Dados (RGPD) possui os instrumentos adequados para transformar o desígnio europeu de uma IA ética e de confiança em realidade. A fim de verificar essa possibilidade, os dispositivos do RGPD, e especialmente a Avaliação de Impacto na Proteção de Dados (AIPD), serão analisados a luz de uma IA europeia ética e de confiança.

PALAVRAS-CHAVE: avaliação de risco; IA ética; princípios éticos; regulação baseada em risco.

THE GENERAL DATA PROTECTION REGULATION AS AN INSTRUMENT FOR REGULATING TRUSTWORTHY ARTIFICIAL INTELLIGENCE IN THE LIGHT OF THE EUROPEAN COMMISSION'S ETHICAL GUIDELINES

ABSTRACT:

It is undeniable that artificial intelligence (AI) generates significant opportunities for society, such as increasing the well-being of individuals and advancing the economy. However, if misused, AI can also pose risks to the interests of society and the fundamental rights of individuals. The lack of specific regulation for AI makes this scenario even more complex since the harm to privacy, fairness, and safety can be irreversible. From this context, ethical and rights principles are presented to build a trustworthy and ethical AI. However, by the absence of specific legislation, it becomes important to understand whether the General Data Protection Regulation (GDPR) is an appropriate instrument to turn the European design of an ethical and trustworthy AI into reality. In order to verify this possibility, the provisions of the GDPR, and mainly the Data Protection Impact Assessment (DPIA), will be analyzed in the light of an ethical and trustworthy European AI.

KEYWORDS: risk assessment; ethical AI; ethical principles; risk-based regulation.

O REGULAMENTO GERAL SOBRE A PROTEÇÃO DE DADOS COMO UM INSTRUMENTO DE REGULAÇÃO DE UMA INTELIGÊNCIA ARTIFICIAL DE CONFIANÇA À LUZ DAS ORIENTAÇÕES ÉTICAS DA COMISSÃO EUROPEIA

Sumário

INTRODUÇÃO

A atual digitalização das relações comerciais, governamentais e sociais tem como consequência a possibilidade de criação de grandes volumes de dados que dizem respeito maioritariamente aos aspetos pessoais dos indivíduos. Esse grande volume de dados, definido como *Big Data*[1], vem sendo utilizado como combustível no aprimoramento da Inteligência Artificial (IA). Este fato fica evidenciado através da ascensão dos algoritmos de aprendizagem de máquina, um subcampo da IA. Nesse sentido, por mais que a sinergia entre dados e IA possa gerar oportunidades para a sociedade, como aperfeiçoamento de diversas áreas de negócios e disponibilização de ferramentas que ajudam no enfrentamento de problemas atuais[2], não se pode acreditar que algoritmos sejam isentos de subjetividade, erro ou manipulação[3]. Nesse sentido, qualquer visão benevolente a respeito da IA não passa de uma ficção cuidadosamente construída[4].

Afirma-se isso pois, ao mesmo tempo em que sistemas baseados em IA passaram a proporcionar uma maior comodidade e eficiência nas realizações de atividades humanas, a sociedade também passou a perceber um aumento da vigilância na esfera privada das pessoas[5]. Através da vigilância e de um maior conhecimento sobre as preferências íntimas dos indivíduos, a IA passou a ser utilizada nas áreas do *marketing* e da política para, através de técnicas de perfilamento e predição, promover a manipulação do comportamento de consumidores e eleitores, mitigando suas respetivas autodeterminações pessoais em detrimento da vontade econômica ou ideológica de terceiros[6].

Além disso, tendo em consideração que os dados são um retrato da sociedade, ao serem treinados através de dados produzidos e coletados de uma sociedade estruturalmente discriminatória, os algoritmos de aprendizagem de máquina passam a replicar e potencializar o comportamento

---

[1] Ou *Dataquake.*

[2] Como pode ser visto em questão a análise de discurso de ódio e apoio na resolução alternativa de conflitos em: Martins R., Gomes M., Almeida JJ, Novais P., Henriques P., Hate Speech Classification in social media Using Emotional Analysis, 7th Brazilian Conference on Intelligent Systems (BRACIS), IEEE, 2018, p.61. Disponível: https://doi.org/10.1109/BRACIS.2018.00019; Andrade F., Novais P., Carneiro D., Zeleznikow J., Neves J., Using BATNAs and WATNAs in Online Dispute Resolution, in New Frontiers in Artificial Intelligence, Kumiyo Nakakoji, Yohei Murakami and Eric McCready (Eds), Springer - Lecture Notes in Artificial Intelligence 6284, ISBN 978-3-642-14887-3, pp 5-18, 2010. Disponível: http://dx.doi.org/10.1007/978-3-642-14888-0_2; Carneiro D., Novais P., Andrade F., Zeleznikow J., Neves J., The Legal Precedent in Online Dispute Resolution, in Legal Knowledge and Information Systems, ed. Guido Governatori (proceedings of the Jurix 2009 - the 22nd International Conference on Legal Knowledge and Information Systems, Rotterdam, The Netherlands), IOS press, ISBN 978-1-60750-082-7, pp 47–52, 2009.Disonível: https://dl.acm.org/doi/10.5555/1671082.1671089; Carneiro D., Novais P., Andrade F., Zeleznikow J., Neves J., Using Case Based Reasoning and Principled Negotiation to provide Decision Support for Dispute Resolution, Knowledge and Information Systems Journal, Springer, ISSN: 0219-1377, Vol 36, Issue 3, pp 789-826, 2013. Disponível: http://dx.doi.org/10.1007/s10115-012-0563-0.

[3] Bioni, Bruno Ricardo e Luciano, Maria. O Princípio da Precaução na Regulação de Inteligência Artificial: seriam as Leis de Proteção de Dados o seu portal de entrada? Inteligência Artificial e Direito, 2ª edição. Revista dos Tribunais, 2020, p. 206

[4] Gillespie, Tarleton. "The relevance of algorithms". In (Ed.), Media Technologies: Essayson Communication, Materiality, and Society. Cambridge: The MIT Press, 2014.

[5] Clarke, R. Profiling: a hidden challenge to the regulation of data surveillance. Journal of Law and Information Science, Australia, v. 4, n. 2, December. 1993. Available at: http://www.austlii.edu.au/au/journals/JILawInfoSci/1993/26.htmID; Magrani, Eduardo. Entre dados e robôs "Ética e privacidade na era da hiperconectividade". Publisher Arquipélago. 2019. ISBN 978-85-5450-029-0

[6] Ebers, Martin, Chapter 2: Regulating AI and Robotics: Ethical and Legal Challenges (April 17, 2019). Martin Ebers/Susana Navas Navarro (eds.), Algorithms and Law, Cambridge, Cambridge University Press, 2019 (Forthcoming), Available at SSRN: https://ssrn.com/abstract=3392379 or http://dx.doi.org/10.2139/ssrn.3392379; European Data Protection Supervisor (EDPS), Opinion 3/2018 on online manipulation and personal data, March 19, 2018.

humano, criando padrões generalizados de discriminação que afetam diretamente os indivíduos, notadamente grupos de pessoas vulneráveis[7].

Não bastassem tais complicações, esses riscos ainda são agravados pela falta de transparência e de explicabilidade decorrente de algumas técnicas de IA[8]. Isso porque, em alguns casos, técnicas como a aprendizagem profunda impedem tanto o utilizador quanto o *developer* de perceberem as razões as quais levaram o modelo algorítmico a tomar determinada decisão.[9] Essa característica, que impossibilita a auditoria do processo decisório, limita e impede o exercício de direitos de ordem fundamental.

Apesar da mencionada situação, a Comissão Europeia (Comissão) tem dado a devida atenção ao tema, sobretudo no que concerne ao desenvolvimento de diretrizes políticas e jurídicas voltadas para a regulação de uma IA responsável, ética e antropocêntrica[10]. Nesse sentido, a Comissão publicou as 'Orientações Éticas para uma IA de Confiança'[11], um documento com valor de *soft law* que apresenta o ideal de uma IA de confiança, que é composta por três aspetos: o sólido, o ético e o legal.

No que diz respeito ao aspeto sólido, uma IA de confiança deve funcionar de forma segura e fiável, sendo, portanto, baseada na segurança por padrão, onde critérios sólidos de segurança são aplicados em todos os ciclos de vida da aplicação[12]. O aspeto ético, por sua vez, indica que um sistema de IA deve respeitar os princípios e valores éticos, sendo eles: (i) o respeito pela autonomia humana; (ii) a prevenção de danos; (iii) a equidade; e (iv) a explicabilidade das decisões. Contudo, as Orientações éticas deixaram, intencionalmente, de abordar o aspeto legal da regulação da IA de confiança, pelo que apenas afirmaram que os sistemas de IA devem respeitar as fontes jurídicas de direito primário e de direito derivado da União Europeia.

A respeito das fontes jurídicas da União, destaca-se ainda que a Comissão recentemente tornou pública a sua proposta de regulamento sobre inteligência artificial[13]. Contudo, tal proposta ainda precisa enfrentar todo o processo legislativo para que possa produzir efeitos legais, o que pode levar um período de tempo considerável. Isso significa que o cenário europeu ainda se encontra sem uma regulamentação específica para a IA, exigindo assim que diplomas setoriais sejam responsáveis pela

[7] European Parliament. Artificial intelligence: From ethics to policy. EPRS | European Parliamentary Research Service. Scientific Foresight Unit (STOA). PE 641.507 – June 2020

[8] Pasquale, Frank. *The black box society:* the secret algorithms that control money and information. Cambridge: Harvard University Press, 2015; Explainable AI: the basics Policy briefing Issued: November 2019 DES6051. The Royal Society

[9] European Parliament. Artificial intelligence: From ethics to policy.

[10] Comissão Europeia. Livro Branco sobre Inteligência Artificial.

[11] Orientações éticas para uma IA de Confiança. Grupo de Especialistas de Alto Nível em Inteligência Artificial. Comissão Europeia. Junho de 2018

[12] Ibid.

[13] European Commission. Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain union legislative acts. Brussels, 21.4.2021 COM(2021) 206 final 2021/0106 (Proposta de Regulação de IA)

tutela de sistemas de IA[14], como é o caso do Regulamento Geral sobre Proteção de Dados Pessoais (RGPD).

Isso posto, e considerando que o RGPD é legitimamente aplicável na tutela de sistemas de IA que tratam dados pessoais, delimitou-se como objetivo principal de pesquisa identificar se o RGPD possui uma proteção jurídica adequada na regulação de uma IA de confiança. Além disso, baseado no princípio da responsabilidade (artigo 5.º, n.º 2 do RGPD), também será analisado se a Avaliação de Impacto da Proteção de Dados[15] (AIPD) pode demonstrar a conformidade de um sistema de IA de confiança, em que será necessário demonstrar respeito ao RGPD e aos princípios éticos referentes ao respeito a autonomia humana, prevenção de danos, equidade e, por fim, explicabilidade.

Assim sendo, para que o objetivo principal desse trabalho seja alcançado, será necessário construir algumas bases de conhecimento, que serão realizadas através dos seguintes passos: (i) entender os fundamentos técnicos da IA e os respetivos riscos que essa tecnologia causa para a sociedade; (ii) apresentar uma visão geral da estratégia europeia de IA, o que permite entender as diretrizes que a Comissão vem traçando para a regulação e desenvolvimento de uma IA de Confiança. Nesse contexto, haverá uma maior profundidade para que se entenda as Orientações éticas para uma IA de Confiança, o que inclui os seus princípios éticos; (iii) identificar se o RGPD cumpre um papel adequado na regulação de uma IA de Confiança, o que inclui o estudo sobre como a AIPD pode promover conformidade de um sistema de IA; e, por fim, (iv) indicar uma metodologia para a AIPD em sistemas de IA com fim de buscar uma IA de Confiança.

[14] Considerando que uma proposta de regulação não produz efeitos legais, que não se tem certeza de quando a proposta será aprovada, e se ela será aprovada no mesmo molde apresentado, conclui-se que a premissa inicial dessa pesquisa ainda se mantem, ou seja, não existe atualmente um regulamento especificamente aplicado a IA. Isso posto, continua válida, atual e necessária uma análise da adequação do RGPD como instrumento de regulação do ideal europeia de uma IA de Confiança.

[15] De acordo com o Considerando 75 do RGPD, a DPIA possui como objetivo a análise de risco, principalmente quando este risco estiver ligado à direitos e liberdades das pessoas singulares, "cuja probabilidade e gravidade podem ser variáveis (e o risco) poderá resultar de operações de tratamento de dados pessoais suscetíveis de causar danos físicos, materiais ou imateriais"

CAPÍTULO PRIMEIRO

UMA PERSPETIVA DOS FUNDAMENTOS DA INTELIGÊNCIA ARTIFICIAL PARA A ÁREA JURÍDICA

1.1. Inteligência Artificial: mito, ficção ou realidade.

Engana-se completamente aquele que ainda não acredita no potencial da Inteligência Artificial (IA) e, por isso, ignora que faz uso diário, direta ou indiretamente, de aplicações baseadas nessa tecnologia. A vida em sociedade tornou-se completamente dependente de algoritmos, e tarefas que, até há pouco tempo, somente poderiam ser realizadas por seres humanos, agora vêm sendo executadas por sistemas baseados em IA e alimentados por dados, como "dirigir carros, analisar dados médicos ou avaliar e executar transações financeiras complexas - sem controle humano ativo ou supervisão"[16]. No entanto, os algoritmos não só realizam tarefas que poderiam ser executadas pelos seres humanos, mas também acabam por desempenhar um papel importante no que concerne a tomada de decisões diárias, de forma a influenciar todos os aspetos da vida em sociedade.

Dentro desse aspeto de influência, algoritmos de 'autoaprendizagem', por exemplo, são responsáveis pelos resultados das pesquisas que são realizadas na web, selecionando os sites, os anúncios e os artigos que são apresentados com base em motores de busca, palavras-chave e perfis de utilizador construído através de todo um histórico de dados[17]. Além disso, 'agentes de software' são responsáveis pela otimização de carteiras de investimento e avaliações de riscos de crédito. Pode parecer irreal, contudo, nos mercados financeiros, mais de 70% do volume de negociação é realizado através de algoritmos[18].

Se durante séculos a tomada de decisão foi realizada através de conhecimento e intuição, agora esse processo decisório é realizado por computadores ou, pelo menos, preparado por eles. Nesse contexto, Citron e Pasquale afirmam que hoje se vive em uma "*scored society*"[19], e que empresas coletam, analisam, compartilham e utilizam dados a fim de avaliar e classificar indivíduos, utilizando padrões para prever o comportamento provável das pessoas através de sistemas de pontuação. Essas pontuações geram consequências existenciais para as pessoas, uma vez que esses

[16] Ebers, Martin, Chapter 2: Regulating AI and Robotics: Ethical and Legal Challenges (April 17, 2019). Martin Ebers and Susana Navas Navarro (eds.), Algorithms and Law, Cambridge, Cambridge University Press, 2019, p.4.

[17] Christl, Wolfie, Corporate Surveillance in Everyday Life. How Companies Collect, Combine, Analyze, Trade, and Use Personal Data on Billions. A Report by Cracked Labs, June 2017, https://crackedlabs.org/dl/CrackedLabs_Christl_CorporateSurveillance.pdf

[18] BI Intelligence, The Evolution of Robo-Advising: How automated investment products are disrupting and enhancing the wealth management industry, 2017.

[19] Citron, Danielle Keats e Pasquale, Frank A. Pasquale, The Scored Society: Due Process for Automated Predictions, 2014 vol 89 Washington Law Review, p. 1.

algoritmos preditivos "decidem cada vez mais se alguém é convidado para uma entrevista de emprego, aprovado para um cartão de crédito ou empréstimo, ou qualificado para fazer uma apólice de seguro"[20].

No entanto, não é somente no setor privado que a IA ganha cada vez mais espaço. As instituições governamentais, através dos setores do fisco, utilizam algoritmos para prever abusos e fraudes nas declarações fiscais, reservando casos para análise humana[21]. Além disso, também há aumento do uso de sistemas baseados em IA na área do 'policiamento preditivo', em que agências de repressão criminal utilizam algoritmos para detetar, responder e prever a ocorrência de crimes[22].

Outro setor em que a IA ganha cada vez mais destaque, é o setor de saúde, em que "sistemas médicos especialistas baseados em algoritmos de 'autoaprendizagem' avaliam a literatura médica e os dados pessoais dos pacientes, auxiliando os médicos no seu diagnóstico e tratamento, seja lendo imagens e registos médicos, detetando doenças, prevendo riscos desconhecidos do paciente, ou selecionando o medicamento certo"[23].

Tem-se também o desenvolvimento de dispositivos que tornam possível conectar o cérebro humano aos computadores[24]. Através dos chamados *brain-computer interfaces* (BCI) já é possível que pessoas, com graves problemas de paralisia, se comuniquem com um computador apenas através da atividade cerebral. Nesse contexto, Elon Musk, em seu projeto Neuralink, previu que as máquinas serão controladas no futuro apenas por pensamentos[25]. Dessa forma, a fronteira entre o homem e a máquina está se tornando cada vez mais cinzenta.

A verdade é que a IA está em toda e qualquer parte da sociedade, incorporando-se gradualmente no dia a dia da vida das pessoas, seja através de assistentes robóticos inteligentes, aspiradores de pó, drones e carros automáticos[26]. Isso implica a necessidade do início de um debate sobre questões éticas e legais que estão diretamente relacionadas aos efeitos positivos e aos impactos negativos iminentes no uso dessa tecnologia. No entanto, para que se esteja preparado para adentrar essa, é preciso, anteriormente, entender a sua parte técnica e conceitual.

1.2. Conceitos básicos da Inteligência Artificial.

[20] Ebers, Martin, Chapter 2: Regulating AI and Robotics: Ethical and Legal Challenges... p.4.
[21] DeBarr, David e Harwood, Maury, Relational Mining for Compliance Risk, Presented at the Internal Revenue Service Research Conference, 2004, p. 178-179. Disponível em: http://www.irs.gov/pub/irs-soi/04debarr.pdf
[22] Barrett, Lindsey, Reasonably Suspicious Algorithms: Predictive Policing at the United States Border, N.Y.U. Review of Law & Social Change, 2017, vol 41, p.334 ss. Disponível em: https://socialchangenyu.com/wp-content/uploads/2017/09/barrett_digital_9-6-17.pdf
[23] Ebers, Martin, Chapter 2: Regulating AI and Robotics: Ethical and Legal Challenges... p.5; Gray, Alex, 7 amazing ways artificial intelligence is used in healthcare, World Economic Forum website, 2018, disponível: https://www.weforum.org/agenda/2018/09/7-amazing-ways-artificial-intelligence-is-used-in-healthcare.
[24] Ebers, Martin, Chapter 2: Regulating AI and Robotics: Ethical and Legal Challenges...p.5.
[25] Hawkins, Andrew J., Elon Musk thinks humans need to become cyborgs or risk irrelevance. The Verge, 2017. Disponível em «https://www.theverge.com/2017/2/13/14597434/elon-musk-human-machine-symbiosis-self-driving-cars».
[26] Ebers, Martin, Chapter 2: Regulating AI and Robotics: Ethical and Legal Challenges... p.5.

1.2.1. Algoritmos, Inteligência Artificial e Robótica: aspetos gerais

Ao acompanhar as inúmeras inovações mencionadas acima, percebe-se que os termos 'algoritmos', 'inteligência artificial' e 'robótica' constantemente aparecem em conjunto. De facto, esses são termos que possuem uma estreita conexão, uma vez que todos eles acabam sendo parte de uma estrutura de conhecimento maior. Contudo, atribuir-lhes um mesmo sentido é um erro técnico que deve ser evitado, porque cada um deles representa algo diferente.

O termo algoritmo, por exemplo, não é algo novo. Na verdade, ele surgiu na Espanha, durante o século XII, quando alguns manuscritos, de um matemático árabe chamado Muḥammad ibn Mūsā al-Khwārizmī, descreviam "métodos de adição, subtração, multiplicação e divisão com o sistema numérico hindu-rábico"[27]. Nesse contexto, o termo 'algoritmo' ganhou o seu primeiro significado como sendo um "método específico, passo a passo, de executar a aritmética elementar escrita"[28].

Desde então, os algoritmos continuam exercendo a sua mesma função. No entanto, com o avanço da tecnologia, esse método foi sendo adequado a diversas tecnologias, uma delas, a da computação. Dessa forma, o termo algoritmo sofreu uma atualização em seu conceito, e hoje ele é conhecido como um "conjunto de passos definidos [e] estruturados para processar dados e instruções para produzir um resultado"[29]. Perceba que, ao adicionar dados e instruções nessa estrutura, o algoritmo tornou-se, de certa forma, parte integrante do desenvolvimento de um *software*, podendo esse *software* até mesmo ser baseado em um sistema de IA.

Por falar em IA, desde já é importante esclarecer que não há um conceito definitivo para o que seja IA, seja por problemas essenciais de nomenclatura[30], seja pela existência de diversos campos de pesquisa e atuação[31]. No entanto, para que se tenha ao menos uma noção geral do que essa tecnologia representa, pode-se afirmar que a IA seria, portanto, "um termo que se refere ao amplo ramo da ciência da computação que estuda e projeta máquinas inteligentes"[32].

Outra representação que poderia ser dada remete ao fato de a IA ser um campo da ciência que, através de seu estudo, "percebe o mundo ao seu redor, formando planos e tomando decisões para atingir determinados objetivos"[33]. Como campo da ciência, e para que ela possa realmente 'perceber o

---

[27] Miyazaki, Shintaro. Algorhythmics: Understanding micro-temporality in computational cultures, Computational Culture, 2012 p. 2, http://computationalculture.net/algorhythmics-understanding-micro-temporality-in-computational-cultures

[28] Ibid.

[29] Kitchin, Rob. Thinking critically about and researching algorithms. Information, Communication & Society, 2017 vol. 20, n. 1, 14–29, p.16. Disponível:http://futuredata.stanford.edu/classes/cs345s/handouts/kitchin.pdf

[30] Problema esse que será tratado na próxima seção.

[31] Problema que também será tratado na seção n.º 3

[32] McCarthy, John. What is Artificial Intelligence? Basic Question. Computer Science Department, 2007. Disponível: http://www-formal.stanford.edu/jmc/whatisai/ .

[33] Maine, Vishal e Sabri, Samer. Machine Learning for Humans. 2017, p.9 Disponível: «https://everythingcomputerscience.com/books/Machine%20Learning%20for%20Humans.pdf».

mundo a sua volta', a IA é fundamentada em diversas outras disciplinas, como matemática, lógica, filosofia, probabilidade, psicologia, neurociência, linguística e, ainda, pela teoria da decisão[34].

Ambos os conceitos são demasiadamente abrangentes. Por isso, é importante compreender suas diferentes características, como cada um desses elementos é usado na prática e, por sua vez, a necessária plasticidade e adaptabilidade de tais conceitos de modo a assumirem face à constante evolução tecnológica.

Essa mesma especificidade se aplica ao termo 'robô'. Uma vez que é complexa a tarefa de encontrar lhe definir um conceito tendo em vista possuis aplicações diversificadas, como membros protéticos e exoesqueletos ortopédicos, robôs industriais, de cuidados ou cirúrgicos e robôs cortadores de relva ou como aspiradores de pó[35]. No entanto, caso seja necessária a busca por uma definição, Ebers afirma que um robô pode ser definido como "uma máquina que tem uma presença física, [que] pode ser programada e tem algum nível de autonomia que depende, entre outros, dos algoritmos de IA usados em seu sistema"[36]. Em outras palavras, um robô é uma "IA em ação no mundo físico"[37].

Após essa breve explicação, o estudo a partir de agora será dedicado a seções mais específicas a respeito da IA, como as dimensões, os tipos de algoritmos e os impactos da IA. Nesse sentido, a fim de referir-se ao termo sistemas baseados em IA, poderá ser usado alguns dos seguintes termos ou palavras: IA, algoritmo, aprendizagem própria ou inteligente, sistemas inteligentes ou autónomos.

1.2.2. Inteligência Artificial: a complexidade da busca por uma definição.

Um dos grandes desafios na área da IA é o de chegar a um acordo sobre sua definição. Ocorre que esse não é só um problema somente para o campo da ciência da computação, mas também do direito. Isso porque uma definição jurídica deve ser segura para poder discernir o objeto que está a ser tutelado, mas também flexível o suficiente para se adequar a possíveis evoluções tecnológicas.

O termo "inteligência artificial" foi utilizado pela primeira vez em 1956, quando o pesquisador de Stanford, John McCarthy, apresentou uma primeira definição sobre a inteligência artificial como sendo "a ciência e engenharia de produzir máquinas inteligentes"[38]. Novais classifica essa definição como 'interessante', na medida em que ela "diz muitas coisas e nada de especial ao mesmo tempo"[39],

[34] Maine, Vishal e Sabri, Samer, "Machine Learning for Humans". (2017)
[35] Ebers, Martin, Chapter 2: Regulating AI and Robotics: Ethical and Legal Challenges... p.7.
[36] Ibid.
[37] Grupo de Peritos de Alto Nível sobre a Inteligência Artificial. Uma definição de IA: Principais capacidades e Disciplinas Científicas, Bruxelas, p.5.
[38] McCarthy, John. What is Artificial Intelligence? Basic Question...
[39] Carneiro D., Novais P., Neves J., Conflict Resolution and its Context. From the Analysis of Behavioural Patterns to Efficient Decision-Making, Chapter 4 Artificial Intelligence in Online Dispute Resolution. Springer-Verlag, 279 pages, 2014, p. 61-62. http://dx.doi.org/10.1007/978-3-319-06239-6

uma vez que é bastante genérica e acaba se baseando numa definição de 'inteligência' que sequer é consensual.

Lugar apresenta um posicionamento semelhante, uma vez que existe ainda uma incompreensão sobre o que seria a própria 'inteligência' em si. Ele faz essa afirmação pois, por mais que seja possível reconhecer um comportamento inteligente, ainda é improvável que "alguém seja capaz de definir inteligência de uma maneira que seja específica o suficiente para auxiliar na avaliação de um programa de computador supostamente inteligente, e, ao mesmo tempo, captura[r] a vitalidade e a complexidade da mente humana"[40]. Além disso, Novais afirma que nos próprios seres humanos, a inteligência "é geralmente definida em função de um conjunto de capacidades, incluindo (...) a aprendizagem, autoconsciência, comunicação, raciocínio, abstrato ou planeamento"[41] .

Assim, antes de definir o que seria IA, deve-se definir propriamente o conceito de 'inteligência'. Com essa finalidade, Luger apresenta uma série de perguntas que possuem o objetivo de estimular a reflexão do leitor, como: "[i] a inteligência é uma faculdade única, ou é apenas um nome dado a uma coleção de habilidades distintas e não correlacionadas? [ii] até que ponto a inteligência pode ser aprendida em oposição à sua existência prévia? [iii] o que exatamente acontece quando ocorre a aprendizagem? [iv] o que é criatividade? [v] o que é intuição? [vi] o que é autoconsciência e que papel ela tem para a inteligência? [vii] é mesmo possível se obter inteligência num computador, ou uma entidade inteligente requer a riqueza das sensações e experiências que só podem ser encontradas numa existência biológica?"[42].

A complexidade de se alcançar uma definição para a inteligência também pode ser percebida ao buscar o significado da própria palavra nos dicionários mais tradicionais. Nesse sentido, note-se que, de acordo com o dicionário Merriam-Webster, a inteligência é considerada como "a capacidade de aprender ou entender ou lidar com situações novas ou situações de tentativa"[43], enquanto o dicionário Cambridge a define como a "capacidade de aprender, entender e fazer julgamentos ou ter opiniões que são baseadas na razão"[44]. De facto, percebe-se que a inteligência está relacionada com a capacidade de aprender e entender coisas. No entanto, enquanto na primeira definição se evidencia a questão das tentativas, na segunda já se apresenta a componente da razão.

Perceba que a definição do que é inteligência já é um problema para o campo da semântica. Dessa forma, alcançar uma definição do que seja uma 'Inteligência Artificial' se transforma em uma tarefa ainda mais difícil. Aliás, rendendo-se a esse facto, Luger conforta-se no sentido de que "a

---

[40] Ibid.

[41] [41] Carneiro D., Novais P., Neves J., Conflict Resolution and its Context... p.62.

[42] Luger, George F. Inteligência Artificial (2013), 9. 23-24.

[43]Merriam-webster, "Intelligence", acedido 19 de Setembro de 2019, <https://www.merriam-webster.com/dictionary/intelligence?utm_campaign=sd&utm_medium=serp&utm_source=jsonld>.

[44] Cambridge Dictionary, "Inteligence", acedido 19 de Setembro de 2019, https://dictionary.cambridge.org/pt/dicionario/ingles/intelligence

dificuldade em encontrar uma definição precisa de IA é completamente apropriada"[45], uma vez que a própria disciplina relacionada com a IA ainda é muito jovem e que, por conta disso, a sua estrutura, o seu interesse e os seus métodos não são, ainda, tão claramente definidos como os outros ramos de uma ciência mais 'madura'[46].

Contudo, por mais que essa tarefa se mostre cada vez mais desafiadora, sempre há pesquisadores e cientistas dispostos a contribuir de alguma forma. Quanto a esses contributos, eles não são poucos, mas esbarram na segunda razão da complexidade de determinar uma definição para a IA, que é caracterizada pelos inúmeros campos de pesquisa e desenvolvimento da IA. Como prova disso, a partir de agora serão apresentadas ao menos quatro dimensões dessa tecnologia, o que demonstrará ainda mais a complexidade do tema.

1.2.3. As quatro dimensões da Inteligência Artificial

Apesar de as matérias e as disciplinas que fundamentam a IA remontem há milénios de anos atrás, como a matemática e a lógica, a realidade é que a IA pode ser considerada uma das áreas mais recentes das ciências e das engenharias. Isso porque os seus primeiros trabalhos ocorreram somente após a Segunda Guerra Mundial, e seu nome, como se conhece hoje, foi primeiramente anunciado apenas no ano de 1956[47].

Como mencionado anteriormente, é demasiadamente complexo resumir a IA em apenas uma definição. Isso porque, ao analisar cada um dos seus quatro grupos de estudo, será facilmente percebido que há, pelo menos, quatro diferentes definições, onde cada definição foi desenvolvida por pessoas diferentes com metodologias diferentes. Esses quatro grupos serão analisados abaixo, e eles se dividem pela seguinte perspetiva: (i) agindo de forma humana; (ii) pensando de forma humana; (iii) pensando racionalmente; e (iv) agindo racionalmente. Como se pode perceber, duas correntes se importam mais com o comportamento, enquanto as outras duas com o pensamento.

1.2.3.1. Agindo de forma humana: a abordagem do teste de Turing

Ao longo da década de 90, as definições apresentadas por cientistas e pesquisadores afirmavam que a IA era "arte de criar máquinas que executam funções que exigem inteligência quando

[45] Luger, George F. Inteligência Artificial (2013), p. 24.
[46] Ibid.
[47] Russell, Stuart e Norvig, Peter. Inteligência Artificial: uma abordagem moderna, ed. Elsevier Editora Ltda., Inteligência Artificial, 3a Edição (Rio de Janeiro, 2013), p. 25.

executadas por pessoas (Kurzweil)"[48]. Outra definição da época afirmava que era um "estudo de como os computadores podem fazer tarefas que hoje são mais bem desempenhadas pelas pessoas (Rich and Knight)"[49]. Como se pode perceber, as duas definições partem da comparação com a forma como os seres humanos agem e executam suas tarefas.

O projeto histórico que representa esse campo de estudo é conhecido como o teste de Turing[50], que foi proposto por Alan Turing, em 1950. Seu objetivo era obter uma definição operacional de inteligência, a partir do qual um sistema seria reconhecido como inteligente se, perante uma mesma situação, ele agisse da mesma forma que uma pessoa inteligente agiria.

Com esse objetivo em mente, o Teste de Turing submetia um computador a um teste em que ele seria aprovado "se um interrogador humano, depois de propor algumas perguntas por escrito, não consegui[sse] descobrir se as respostas escritas vêm de uma pessoa ou de um computador"[51]. Assim, apesar de esse teste ter sido desenvolvido em 1950, ele ainda é importante para o contexto da IA.

Além disso, mesmo naquela época, esse era um teste que demandava muitas capacidades de um sistema computacional[52], entre elas (i) processamento de linguagem natural, a fim de permitir que o sistema possa se comunicar com sucesso em um idioma natural (i.e., português, inglês ou espanhol); (ii) representação de conhecimento, visando armazenar tudo aquilo que o sistema 'sabe' e 'ouve'; (iii) raciocínio automatizado, a fim de utilizar as informações já armazenadas para a finalidade de tirar novas conclusões e responder a perguntas realizadas; e (iv) aprendizagem de máquina, para que o sistema detetasse e extrapolasse padrões e se adaptasse a novas circunstâncias. Além disso, o teste ainda submetia o sistema a testes de perceção de objetos, o que demandava do sistema as capacidades de (v) visão computacional e (vi) robótica, a fim de que fosse possível se movimentar e manipular objetos.

Ao usar seis disciplinas que compõem a maior parte de toda a ciência de IA, é muito importante, até hoje, reconhecer os méritos de Turing, uma vez que seu teste permanece relevante mesmo após 70 anos.

1.2.3.2. Pensando de forma humana: a estratégia de modelagem cognitiva

De acordo com essa área da IA, para que um sistema pensasse de forma humana, primeiramente era necessário "determinar como os seres humanos pensam [o que denota a

[48] Russell, Stuart e Norvig, Peter. Inteligência Artificial... p.26.
[49] Ibid.
[50] Ibid.
[51] Russell, Stuart e Norvig, Peter. Inteligência Artificial... p.26.
[52] Russell, Stuart e Norvig, Peter. Inteligência Artificial... p.26-27

necessidade de] penetrar nos componentes reais da mente humana"[53]. Nesse sentido, existiam três maneiras para que se pudesse entender como uma mente humana pensa, sendo elas: (i) a introspeção, ao procurar captar os próprios pensamentos à medida que eles se desenvolvem; (ii) os experimentos psicológicos, ao observar uma pessoa em ação; e, por último, (iii) através de imagens cerebrais, observando o cérebro em ação.

Dessa forma, e de acordo com essa corrente, "depois [de termos] uma teoria da mente suficientemente precisa, será possível expressar a teoria como um programa de computador. Se os comportamentos de entrada/saída e sincronização do programa coincidirem com os comportamentos humanos correspondentes, isso será a evidência de que alguns dos mecanismos do programa também podem estar operando nos seres humanos"[54].

1.2.3.3. Pensando racionalmente: a abordagem das 'leis do pensamento'

De acordo com a racionalidade, essa corrente se baseia muito no que foi desenvolvido pelo filósofo Aristóteles no que se refere aos silogismos, ou seja, um tipo de argumento lógico que aplica o raciocínio dedutivo para extrair uma conclusão de duas ou mais proposições, que se supõem verdadeiras[55]. Percebe-se, portanto, que essa corrente é muito baseada no campo da lógica[56].

Com base nessa corrente, por volta de 1965, alguns 'sistemas inteligentes' começaram a ser desenvolvidos com base em "declarações sobre todos os tipos de coisas no mundo e sobre as relações entre elas"[57], e que, em princípio, "poderiam resolver qualquer problema solucionável descrito em notação lógica"[58]. No entanto, essa abordagem acabou por enfrentar dois grandes obstáculos.

Quanto ao primeiro obstáculo, percebeu-se que não era simples transformar conhecimentos informais em conhecimentos formais exigidos pela linguagem lógica. Quanto ao segundo, percebeu-se que "há uma grande diferença entre resolver um problema na teoria e resolvê-lo na prática"[59] e, sem uma orientação sobre as prioridades de etapas de raciocínio, a tentativa indiscriminada poderia esgotar os recursos computacionais de qualquer computador.

[53] Russell, Stuart e Norvig, Peter. Inteligência Artificial... p.27.
[54] Russell, Stuart e Norvig, Peter. Inteligência Artificial... p.27.
[55] Ibid.
[56] Russell, Stuart e Norvig, Peter. Inteligência Artificial... p.28.
[57] Ibid.
[58] Ibid.
[59] Ibid.

1.2.3.4. Agindo racionalmente: a abordagem de agente racional

Para entender essa conceção, primeiro é interessante diferenciar um agente de um agente racional. Isso porque, enquanto um 'agente' é meramente algo que age; um 'agente racional' será aquele que "age para alcançar o melhor resultado ou - quando há incerteza - o melhor resultado esperado"[60].

Além disso, também é importante entender que as 'leis do pensamento', representadas pela lógica, pelo silogismo e pelas inferências corretas da corrente anterior, são somente uma parte daquilo que se espera de um agente racional. Isso porque o raciocínio lógico é exatamente aquilo que se espera das ações de um agente racional, ator central dessa corrente.

Por outro lado, a fim de complementar essa racionalidade, é entendido que, em algumas situações, "não existe nenhuma ação comprovadamente correta a ser realizada, contudo, mesmo assim algo tem de ser feito. [E que] também existem modos de agir racionalmente que não se pode dizer que envolvem inferências"[61] como, por exemplo, a execução de um ato reflexo, derivado da ação instintiva de se recolher a mão quanto esta está em contato com o fogo.

Dentro de todo esse contexto, Russell e Norvig entendem que a abordagem do agente racional possui duas vantagens sobre todas as demais teorias apresentadas anteriormente. A primeira, pois ela não se restringe somente às 'leis do pensamento', pois a lógica é apenas uma parte da racionalidade. Em segundo, que a abordagem do agente racional é "mais acessível ao desenvolvimento científico do que as estratégias baseadas no comportamento ou no pensamento humano"[62].

1.2.4. A ascensão da Aprendizagem de Máquina

A IA definitivamente é um ramo da ciência que possui diversas aplicações, como foi mencionado anteriormente ao comentar várias das técnicas utilizadas no teste de Turing. A aprendizagem de máquina[63], por outro lado, deve ser necessariamente compreendido como um subcampo da IA. No entanto, antes que se inicie efetivamente o estudo sobre ela, é importante entender o fenômeno que ocorreu através da computação e da digitalização das relações sociais, e que veio a possibilitar a criação de grandes quantidades de dados que atualmente alimentam a IA e a aprendizagem de máquina.

[60] Russell, Stuart e Norvig, Peter. Inteligência Artificial... p.28-29.
[61] Russell, Stuart e Norvig, Peter. Inteligência Artificial... p.29.
[62] Ibid.
[63] Também conhecido como Machine Learning.

Tal fenômeno foi chamado, pelo *Massachusetts Institute of Technology* (M.I.T.), de '*dataquake*'[64]. E sua origem foi parcialmente desencadeada por alguns fatores, como a introdução no mercado de computadores pessoais com interfaces gráficas mais amigáveis, a expansão dos telemóveis e da difusão acessível de internet. Todos esses fatores vieram a tornar o computador, seja ele o pessoal ou o telemóvel, um dispositivo praticamente omnipresente em nossa sociedade.

Como consequência da democratização da tecnologia, a sociedade começou a passar por um processo de digitalização. As documentações anteriormente realizadas em papel deram lugar as bases de dados em suporte digital, que se tornaram rapidamente o principal meio de armazenamento de informação. Além disso, paralelamente, a comunicação digital também acabou por substituir a analógica, e com isso o meio digital também se tornou a "principal forma de transferência de informação"[65].

Esse novo contexto de transferência e armazenamento de dados fez com que "todo o tipo de informação, não só números e texto, mas também imagem, vídeo, áudio etc., fossem armazenados, processados, e - graças à conectividade *online* - transferidos digitalmente"[66]. Nesse contexto, toda essa quantidade de dados, que agora era gerada diariamente e que se referia a todos os setores da sociedade, foi considerada como sendo o mencionado fenômeno '*dataquake*'. Como é de se esperar, esse fenômeno, que alguns também chamam de *Big Data*, se tornou diretamente "responsável por desencadear o interesse generalizado na análise de dados e na aprendizagem de máquinas"[67].

Como se pode perceber, a ascensão da aprendizagem de máquinas está diretamente ligada ao momento em que a sociedade passou, não só a gerar mais dados, mas também a armazená-los e, consequentemente, buscar e possibilitar meios de processá-los em grande escala. No entanto, para explicar a ideia principal do que seria a aprendizagem de máquina, primeiro é importante relembrar que, por muito tempo, para se resolver um problema computacional, um programador era a pessoa responsável por definir, através da codificação de um algoritmo, o que um computador deveria necessariamente fazer[68].

Contudo, após décadas de uma "revolução silenciosa no campo da ciência da computação"[69], ao invés de programadores codificarem algoritmos com instruções específicas, o fortalecimento da aprendizagem de máquina passou a permitir que os próprios computadores 'aprendessem' a partir de

[64] Alpaydin, Ethem. Machine Learning, the new AI. The MIT press essencial knowledge series. 2016 Massachusetts Institute of Technology. ISBN 9780262529518, p.10.
[65] Alpaydin, Ethem. Machine Learning, the new AI. (2016), Preface.
[66] Ibid.
[67] Ibid.
[68] Alpaydin, Ethem. Machine Learning, the new AI. (2016), p. 11
[69] Alpaydin, Ethem. Machine Learning, the new AI. (2016), Preface.

um banco de dados. Isso possibilitou que as próprias máquinas melhorassem a si próprias sem serem explicitamente programadas[70].

Nesse novo contexto, "programas de computador aprendiam [e adaptavam] seu comportamento automaticamente, [objetivando] atender melhor aos requisitos de sua tarefa"[71]. Essa mudança de paradigma foi fundamental para que fosse possível o surgimento de programas que promovessem o reconhecimento facial, entendessem a comunicação baseada em uma linguagem natural, pudessem dirigir um carro e até mesmo recomendar um filme.

Em resumo, ao falar em aprendizagem de máquina, deve-se entendê-la como "um subcampo da ciência da computação e da estatística [que possui] fortes laços com a IA, [pois fornece] métodos, teorias e domínios de aplicação para o campo"[72]. Além disso, a aprendizagem de máquina pode ser considerada "uma máquina ou software que pode aprender automaticamente"[73]. Não da maneira como os humanos aprendem, mas "com base em um processo computacional e estatístico"[74].

Nesse contexto, os dados são considerados a fonte de alimentação, ou melhor, de aprendizagem, possibilitando que "os algoritmos de aprendizagem possam detetar padrões ou regras nos dados e fazer previsões para dados futuros"[75], sem que, para isso, seja obrigatoriamente necessária a "instruções explicitamente programadas"[76].

Dito tudo isso, há três coisas que são importantes para o resto do desenvolvimento deste estudo. Em primeiro lugar, que dentro do contexto de aprendizagem de máquinas, 'aprender' tem o significado de "qualquer mudança num sistema que melhore o seu desempenho na segunda vez que ele repetir a mesma tarefa, ou uma outra tarefa da mesma população"[77]. Para que haja essa aprendizagem, será necessário criar um modelo específico para a realização de cada tarefa que se pretende concretizar, onde cada modelo, a depender da sua finalidade, será submetido a um determinado tipo de 'treino'.

Em segundo, que os termos algoritmo e modelo possuem sentidos ligeiramente diferentes. Algoritmo, como já mencionado, consiste em uma instrução estruturada em uma sequência de passos. Modelo, por outro lado, deve ser considerado como "um algoritmo baseado em uma função

[70] Ebers, Martin, Chapter 2: Regulating AI and Robotics: Ethical and Legal Challenges... p.7.
[71] Alpaydin, Ethem. Machine Learning, the new AI. (2016), Preface
[72] P.Anitha, G.Krithka, Mani Deepak Choudhry, Machine Learning Techniques for learning features of any kind of data: A Case Study. International Journal of Advanced Research in Computer Engineering & Technology (2014), ISSN: 2278-1323, p. 4324.
[73] Coeckelbergh, Mark. Ética da IA MIT Press Essential Knowledge. Edição do Kindle, p. 204.
[74] Ibid.
[75] Ibid.
[76] Anitha, P, Krithka, G. e Choudhry, Mani Deepak. Machine Learning Techniques for learning features of any kind of data. A Case Study. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3, Issue 12, 2014, p.4324.
[77] Luger, George F. Inteligência Artificial, p.332.

matemática que gera uma saída com base nos padrões aprendidos com os dados de treino no processo de treino"[78].

Em terceiro, que a aprendizagem de máquina também possui subcampos conhecidos como técnicas de aprendizagem de máquina e que, consequentemente, irão impor, a cada tipo de modelo, um 'treino' diferente. Essas técnicas podem ser denominadas por: (i) aprendizagem supervisionada; (ii) aprendizagem não supervisionada; e, por fim, (iii) aprendizagem por reforço[79].

1.2.4.1. Aprendizagem supervisionada

A aprendizagem supervisionada é uma técnica da aprendizagem de máquina que visa possibilitar que uma máquina aprenda uma tarefa específica. No entanto, como o próprio nome diz, todo o treino dessa máquina será, de alguma forma, supervisionado por um 'professor'.

Em primeiro lugar, haverá o chamado professor que determinará, desde tão cedo, a saída/resultado que ele deseja que o modelo alcance[80]. Paralelamente, a fim de que esse seja um treino supervisionado, será apresentado à máquina um conjunto de dados já previamente estruturado e rotulado que servirá como dados de treino. Este conjunto possibilitará que a máquina possa, através do algoritmo de aprendizagem, "aprender uma regra geral que [possa mapear] as entradas [com vista a reconhecer] as saídas"[81].

Um exemplo que pode ser apresentado ocorre quando o algoritmo precisa aprender como reconhecer um determinado tipo de animal, como um gato. Com esse objetivo em mente, o *developer* fornece ao sistema muitos exemplos de imagens de gatos, informando, a sua respetiva interpretação, ou seja, onde, em cada imagem, o gato está presente[82]. Dessa forma, após o período de aprendizagem, "o sistema será então capaz de generalizar para saber também como interpretar imagens de gatos nunca vistas"[83].

Em resumo, "um algoritmo de aprendizagem supervisionado analisa os dados de treino"[84], que já são previamente rotulados pelo professor, e dessa forma, "produz uma função inferida, que pode ser usada para mapear novos exemplos"[85], ou seja, "o sistema extrapola ou generaliza suas respostas, para que ele atue corretamente em situações não presentes no conjunto de treino"[86].

---

[78] Drexl, Hilty et al., Technical Aspects of Artificial Intelligence: An Understanding from an Intellectual Property Law Perspective, Version 1.0, 2019, p.5. Disponível: https://ssrn.com/abstract=3465577

[79] Anitha, P, Krithka, G. e Choudhry, Mani Deepak. Machine Learning Techniques for learning features of any kind of data. A Case Study... p. 4325.

[80] Anitha, P, Krithka, G. e Choudhry, Mani Deepak. Machine Learning Techniques for learning features of any kind of data. A Case Study... p. 4325-4326.

[81] Ibid.

[82] Ebers, Martin, Chapter 2: Regulating AI and Robotics: Ethical and Legal Challenges...p.8.

[83] Ibid.

[84] Anitha, P, Krithka, G. e Choudhry, Mani Deepak. Machine Learning Techniques for learning features of any kind of data. A Case Study... p. 4325-4326.

[85] Ibid.

[86] Barto, A. G. & Sutton, R. Introduction to Reinforcement Learning. Second edition. (2014-2015). The MIT Press, p. 2.

1.2.4.2. Aprendizagem não supervisionada

Diferente do treino anterior, na aprendizagem não supervisionada o conjunto de dados de treino fornecido à máquina não possui qualquer sinal de rotulagem, deixando para o algoritmo de aprendizagem encontrar, por conta própria, um padrão comum, geralmente oculto, entre os dados de entrada não rotulados[87]. Através dessa abordagem, esse modelo é muito mais propício para ser utilizado de modo a identificar semelhanças, paralelos e diferenças dentro do conjunto de dados apresentado[88].

Dessa forma, ao invés de classificar uma determinada imagem como sendo um gato, o algoritmo de aprendizagem não supervisionado será capaz de, a partir de um conjunto vasto de dados não estruturados e não rotulados, agrupar dados que respetivamente representem, por exemplo, cavalos, gatos e humanos[89]. Em outras palavras, o algoritmo não saberá que aquela imagem representa um gato, mas saberá diferenciar, através de padrões comuns, que a imagem de um gato é diferente da de um cavalo que, por consequência, são ambas diferentes da de uma pessoa humana. Nesse contexto, não havendo a rotulagem prévia do que cada padrão significa, o que se pode esperar do algoritmo não supervisionado é o agrupamento, e não a classificação.

Um bom uso que se pode dar ao resultado obtido do agrupamento de dados realizado pelos algoritmos de aprendizagem não supervisionado é a projeção dos dados em alta ou baixa dimensão, possibilitando uma visualização ou análise daquele grupo de dados que, anteriormente, não estavam estruturados. Essa finalidade pode ser demonstrada a partir de um algoritmo que analisa dados sobre clientes para criar grupos com base em seu potencial de compra a fim de personalizar as ofertas[90].

1.2.4.3. Aprendizagem por reforço

A aprendizagem por reforço, assim como as anteriores, é considerada uma técnica de treino de aprendizagem de máquina. No entanto, ao invés de classificar ou de agrupar dados, o objetivo desta é de construir um meio pelo qual a máquina irá tomar uma sequência de decisões[91].

A máquina que está a ser treinada, conhecida como agente, ao enfrentar uma situação que exige uma série de tomada de decisões, deverá "aprender a atingir um [objetivo] em um ambiente incerto e potencialmente complexo"[92]. Para que alcance essa meta, existirão dois fatores incluídos nessa equação: o primeiro relativo à exploração de decisões baseadas em tentativa e erro; e o

[87] Anitha, P, Krithka, G. e Choudhry, Mani Deepak. Machine Learning Techniques for learning features of any kind of data. A Case Study... p. 4330.
[88] Drexl, Hilty et al., Technical Aspects of Artificial Intelligence... p.13.
[89] Anitha, P, Krithka, G. e Choudhry, Mani Deepak. Machine Learning Techniques for learning features of any kind of data. A Case Study... p. 4330.
[90] Drexl, Hilty et al., Technical Aspects of Artificial Intelligence... p. 8.
[91] Data Science Academy. Deep Learning Book, 'Cap. 62 - O que é Aprendizado por Reforço?' Disponível em «http://deeplearningbook.com.br/o-que-e-aprendizagem-por-reforco/». Acedido em 25/07/2020.
[92] Ibid.

segundo, que passa por uma política de recompensas atrelada à decisão ou ao conjunto de decisões tomadas[93].

Nesse sentido, a máquina se valerá da estratégia de tentativa e erro a fim de encontrar uma solução para a situação que está submetida. No entanto, para que ela ao menos tenha uma noção do que o *developer* deseja, o algoritmo de aprendizagem receberá recompensas ou penalidades pelas ações que executa, sendo que seu objetivo é maximizar a recompensa total[94]. Essa maximização está ligada ao fato de que, quanto maior o valor da recompensa recebida, mais efetiva foi a tomada de decisão. Isso também deixa a máquina "propensa a procurar maneiras inesperadas de fazê-lo"[95], maneiras essas que possivelmente mentes humanas não seriam capazes de imaginar.

Acerca da política de recompensas, essas podem ser identificadas como sendo 'as regras do jogo'. E sem qualquer sugestão ou dica de como resolver os obstáculos inseridos naquele ambiente, caberá à máquina iniciar a sua exploração através de uma série de ações aleatórias. Ao descobrir pouco a pouco as ações que possuem maior recompensa, o algoritmo de aprendizagem por reforço irá criar uma estratégia mais sofisticada, sempre tentando maximizar sua pontuação.

Muita da aprendizagem dos seres humanos também ocorre através de recompensas e penalidades, seja por pontuações em um ambiente físico ou presencial, como um elogio cada vez que se faz algo corretamente, seja por sentimentos em um ambiente mais sensorial, como a dor ao colocar a mão ao fogo. No entanto, ao contrário da mente humana, uma máquina baseada em IA pode "reunir experiência de milhares de [ações] paralelas se um algoritmo de aprendizagem por reforço for executado em uma infraestrutura de computador suficientemente poderosa"[96]. É exatamente por conta dessa capacidade que se tem nos dias atuais, por exemplo, máquinas vencendo os melhores jogadores mundiais em jogos como xadrez e *Go*, robôs e até carros autônomos.

A respeito de jogos, essa metodologia foi usada para treinar um algoritmo para jogar *Go*[97]. A máquina "jogou o jogo contra si mesma várias vezes e obteve feedback baseado apenas na pontuação final de cada jogo"[98]. Embora o algoritmo nunca tenha recebido qualquer ensinamento sobre estratégias de *Go*, ele foi capaz de melhorar suas habilidades e superar os jogadores humanos apenas com base nas recompensas que recebia[99].

---

[93] Barto, A. G. & Sutton, R. Introduction to Reinforcement Learning... p. 2-3.
[94] Ibid.
[95] Data Science Academy. Deep Learning Book, 'Cap. 62 - O que é Aprendizado por Reforço?'
[96] Ibid.
[97] Silver, D., Huang, A., Maddison, C. et al. Mastering the game of Go with deep neural networks and tree search. Nature 529, 484–489 (2016). https://doi.org/10.1038/nature16961
[98] Drexl, Hilty et al., Technical Aspects of Artificial Intelligence... p. 8.
[99] Ibid.

1.2.4.4. Redes neurais artificiais e a aprendizagem profunda

Se no intuito de simular a habilidade de voar, a racionalidade humana baseou-se, primeiramente, em como os pássaros realizavam essa ação, é consequentemente natural que os seres humanos se inspirassem no cérebro humano de modo a replicar em uma máquina aquilo que se entende por inteligência. Com isso em mente, as redes neurais biológicas humanas foram a base da criação das redes neurais artificiais (RNA), que, com o consequente avanço tecnológico, inspiraram o modelo de aprendizagem profundo[100], também conhecido como *Deep Learning.*

Dentro de todo espectro da IA, a aprendizagem profunda é, portanto, uma subclasse da aprendizagem de máquina que pretende executar o seu modelo de aprendizagem através do processamento de dados de uma maneira que se possa simular o comportamento realizado pelo cérebro humano. É importante exemplificar que esse tipo de aprendizagem possibilita exercer tarefas como reconhecimento visual, reconhecimento de fala e processamento de linguagem natural[101].

Contudo, para que seja possível entender a dinâmica da aprendizagem profunda, é necessário ter ao menos uma noção de como o cérebro humano funciona. Dessa forma, a ciência já nos demonstrou que o "cérebro humano é composto de um grande número de unidades de processamento, [ao qual são] chamadas de neurônios, e [que] cada neurônio é conectado a um grande número de outros neurônios através de conexões chamadas sinapses"[102]. Além disso, "os neurônios operam em paralelo e transferem informações entre si através destas sinapses"[103], onde acredita-se que o processamento das informações é realizado pelos neurónios, enquanto a memória em si está diretamente ligada as sinapses[104].

Nesse sentido, a estrutura das redes neurais artificiais utilizadas na aprendizagem profunda possuem uma simetria correlacionada às redes neurais biológicas. Como forma de exemplificação, deve-se imaginar que a rede neural artificial é construída por diversas e paralelas camadas, onde cada camada é composta por inúmeros 'neurônios artificiais', também conhecidos como unidades de processamento[105], que se interligam e se comunicam, através de computação, com as demais unidades com as quais estão conectados[106].

---

[100] Alpaydin, Ethem. Machine Learning, the new AI. (2016), p. 85.
[101] Bezerra, Eduardo. (2016). Introdução à Aprendizagem Profunda. Edition: 1, Chapter: 3, Publisher: SBC, Editors: Ogasawara, p.57
[102] Alpaydin, Ethem. Machine Learning, the new AI. (2016), p. 86.
[103] Ibid.
[104] Ibid.
[105] Bezerra, Eduardo. (2016). Introdução à Aprendizagem Profunda, p. 60
[106] Ibid.

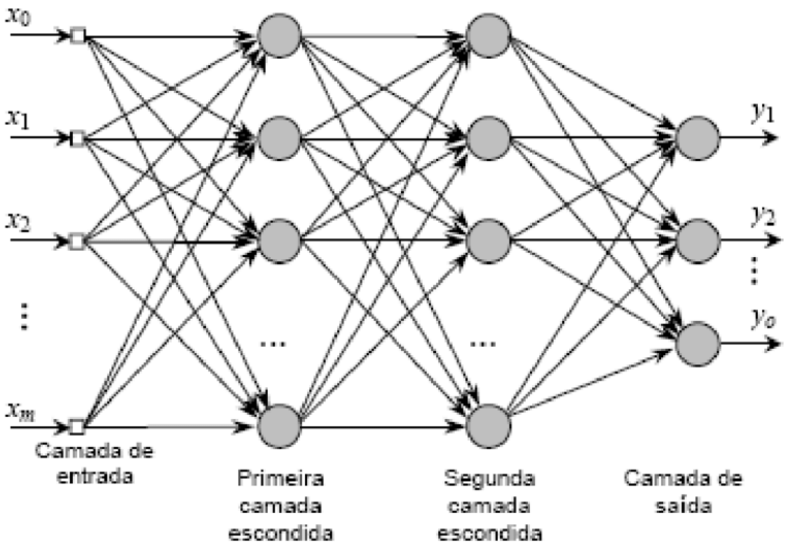

Figura1: Rede Neural Artificial Multicamadas[107]

No entanto, para fins de entendimento, há de se destacar a existência de três tipos de camadas, sendo elas a (i) camada de entrada; a (ii) camada de saída; e a (iii) camada oculta ou escondida. Em uma das extremidades é encontrada a (i) camada de entrada, que é responsável por receber os *inputs* de dados; enquanto na outra extremidade encontra-se a (ii) camada de saída, responsável por fornecer o resultado. Por exclusão, as demais camadas que se encontram no meio são denominadas como (iii) camadas ocultas, que diferentemente das demais, serão responsáveis pelo processamento interno da rede através de determinados parâmetros. Esses parâmetros também são conhecidos como pesos, que quando aplicados aos dados de entrada, fornecerão a saída prevista[108].

Nessa altura, pode ser que haja uma dúvida a respeito do que separa uma rede neural artificial, para a técnica de aprendizagem profunda. Dessa forma, destaca-se apenas que o fator de diferenciação se apresenta no fato de que, na aprendizagem profunda, haverá uma complexidade maior e um número maior no que se refere às camadas ocultas[109].

Dentro de todo o contexto, a aprendizagem profunda é uma forma mais robusta e complexa de aprendizagem, possibilitando aplicações que também se apresentam como mais complexas, como, por exemplo, os carros autônomos, os sistemas de deteção de fraudes, o reconhecimento de linguagem natural, os assistentes virtuais, o reconhecimento visual de objetos, entre outros. Além disso, podemos citar aplicações na área da saúde, como deteção de doenças com base em imagens e padrões e descoberta de novos genomas e medicamentos[110].

---

[107] Oliveira, Anderson Castro Soares. Aplicação de redes neurais artificiais na previsão da produção de álcool. Março, 2010. Pág. 4. DOI: 10.1590/S1413-70542010000200002

[108] Alpaydin, Ethem. Machine Learning, p. 90.

[109] Alpaydin, Ethem. Machine Learningp. 106-109

[110] Chatterjee, Marina. Top 20 Applications of Deep Learning in 2020 Across Industries. Great learning. 2019. Disponível: https://www.mygreatlearning.com/blog/deep-learning-applications/.

1.2.5. Singularidade e superinteligência: a Inteligência Artificial irá superar o ser humano?

Através de uma análise sobre a ascensão dos algoritmos de aprendizagem de máquina, é possível perceber o quanto a IA vem se desenvolvendo. Ações que um dia a sociedade acreditou que somente poderiam ser executadas por seres humanos, atualmente já se encontram automatizadas e realizadas por algoritmos baseados em IA. Na verdade, o conjunto entre poder computacional, dados e técnicas de aprendizagem de máquina permitiu que as máquinas executassem tarefas até mesmo inimagináveis para a própria inteligência humana.

Dessa forma, como consequência de uma IA mais preparada, complexa e melhor desenvolvida, tornou-se normal o surgimento de discussões e comparações do seu estado de evolução em relação à inteligência humana[111]. Tais discussões propuseram a classificação de sistemas baseados em IA em três possíveis categorias, sendo elas a (i) IA estreita ou fraca; a (ii) IA geral ou forte; e a (iii) singularidade ou superinteligência.

Ao que parece, essa não é uma discussão recente, uma vez que, em 2005, em seu livro *'The Singularity is Near'*[112], Kurzweil já cunhava o termo IA estreita justamente para se referir a "criação de sistemas que executam comportamentos 'inteligentes' específicos em contextos específicos"[113]. Perceba-se que, por mais que haja certo 'comportamento inteligente', esse é completamente direcionado para uma atividade extremamente própria, como jogar xadrez ou classificar imagens[114], sendo esse o ponto chave para entender essa classificação de IA.

Nesse sentido, por mais que se refira à uma IA estreita ou fraca, isso nada tem a ver com o seu potencial ao executar suas tarefas. Até mesmo porque, "geralmente, os sistemas de IA estreitos excedem muito a velocidade dos seres humanos, além de fornecerem a capacidade de gerenciar e considerar milhares de variáveis simultaneamente"[115]. Contudo, a imperfeição ou fragilidade de uma IA fraca está relacionada aos seus limites de 'inteligência', isso porque, "para um sistema restrito de IA, se alguém mudar um pouco do contexto ou da especificação de comportamento, geralmente será necessário algum nível de reprogramação ou reconfiguração humana para permitir que o sistema mantenha seu nível de inteligência"[116].

Dessa forma, por mais que tais sistemas possam exceder o desempenho da mente humana para a realização de tarefas específicas, falta aos sistemas de IA estreita o poder de adaptabilidade e generalização do seu conhecimento para ser aplicado em objetos e contextos distintos e imprevisíveis.

---

[111] Coeckelbergh, Mark. AI Ethics. MIT Press. Edição do Kindle, p. 64-65.
[112] Kurzweil, R. The singularity is near: When humans transcend biology. Penguin, 2005.
[113] Goertzel, Ben. Artificial General Intelligence: Concept, State of the Art, and Future Prospects. OpenCog Foundation. Journal of Artificial General Intelligence, 2014, DOI: 10.2478/jagi-2014-0001, p. 2-3.
[114] Coeckelbergh, Mark. Ética da IA. MIT Pressione. Edição do Kindle, p. 64-67.
[115] Kurzweil, R. The singularity is near: When humans transcend biology. (2005), p. 206.
[116] Goertzel, Ben. Artificial General Intelligence: Concept, State of the Art, and Future Prospects, p.1.

Porém, é exatamente com base nessa característica que, de forma antagônica ao conceito de IA fraca, surgiu a classificação de sistemas inteligentes de IA geral ou forte.

Para que um sistema seja classificado como de IA forte é preciso que ele apresente uma "capacidade de atingir diversos objetivos e realizar diversas tarefas, em diversos ambientes diferentes"[117], podendo, assim, "lidar com problemas e situações bastante distintas daquelas previstas por seus criadores"[118]. Diferente da IA fraca, a IA forte exige que o sistema possua ampla capacidade de generalização. No entanto, é preciso destacar que, dentro do contexto de IA forte, não há qualquer expectativa de que essa capacidade de generalidade, adaptabilidade e flexibilidade seja, de alguma maneira, infinita[119].

Em resumo, um sistema de IA forte deve executar com êxito qualquer tarefa cognitiva que um ser humano possa realizar, como: aprender, planejar, tomar decisões sob incertezas, comunicar-se em linguagem natural, contar piadas, manipular pessoas, negociar ações, ou até mesmo, reprogramar-se[120].

Contudo, por mais que exista essa classificação, há quem defenda que ainda não houve qualquer criação de sistema baseando em uma IA forte, seja porque os "modelos matemáticos tradicionais [...] não possam ser usados como ponto de partida para a criação de tais programas"[121]; ou porque os "modelos automatizados gerados pelo uso da aprendizagem automática [...] não podem ser estendidos para lidar com o diálogo humano"[122].

Diante dessa impossibilidade, algumas pessoas podem sentir-se de alguma forma desapontadas. Todavia, através de uma outra perspetiva, é certo que existem cientistas, filósofos e pesquisadores aliviados pelo fato de a humanidade ainda não ter criado nenhuma IA forte. Isso porque, se uma IA forte é capaz de desempenhar tarefas como humanos conseguem, ela poderá, em algum momento, reprogramar-se. E através de um exercício especulativo, caso um sistema consiga melhorar a si mesmo, ele poderia "destravar um ciclo de autoaperfeiçoamento recursivo que levaria a uma explosão de inteligência ao longo de um período desconhecido, variando de décadas ou a um único dia"[123].

De facto, esse tipo de especulação pode parecer algo inapropriado para se apresentar em um trabalho acadêmico como esse. No entanto, é exatamente através dessa 'explosão de inteligência' que surge a terceira e última classificação de sistemas baseados em IA, qual seja, a da singularidade ou superinteligência. Caso o desenvolvimento tecnológico alcance esse nível, é de se esperar um sistema

[117] Goertzel, Ben. Artificial General Intelligence: Concept, State of the Art, and Future Prospects, p.2.
[118] Ibid.
[119] Ibid.
[120] Maine, Vishal e Sabri, Samer. Machine Learning for Humans, p.13.
[121] Landgrebe, Jobst & Smith, Barry (2020). There is no general AI: Why Turing machines cannot pass the Turing test, p.1.
[122] Ibid.
[123] Maine, Vishal e Sabri, Samer. Machine Learning for Humans, p.11.

que possua um "intelecto que exceda muito o desempenho cognitivo dos seres humanos em praticamente todos os domínios de interesse"[124].

Dentro de um contexto em que os seres humanos podem possibilitar a criação de algo com igual ou até maior capacidade intelectual que a da própria raça humana[125], surge a imprevisibilidade de uma cadeia consequencial infinita, onde a 'criação' pode vir a criar 'algo' ainda mais inteligente que ela mesma, gerando a já mencionada 'explosão de inteligência'[126].

Por conta desses fatores, a principal questão a ser discutida refere-se aos impactos que esse desenvolvimento pode causar à sociedade, isso porque "o sucesso na criação da IA seria o maior evento da história da humanidade. Infelizmente, também pode ser o último, uma vez que há quem se preocupe até mesmo com um possível risco existencial para a humanidade[127].

Por sinal, essa é uma preocupação antiga e que remonta à 1942, onde o escritor de ficção científica, Isaac Asimov, já visualizava esse tipo de risco para a humanidade. Ele foi a primeira pessoa a abordar diretamente essa questão através das suas três leis da robótica, sendo elas: (i) um robô não pode ferir um ser humano ou, por omissão, permitir que um ser humano sofra algum mal; (ii) um robô deve obedecer às ordens que lhe sejam dadas por seres humanos, exceto nos casos em que tais ordens contrariem a primeira lei; e (iii) um robô deve proteger sua própria existência desde que tal proteção não entre em conflito com a primeira ou a segunda lei. Esse talvez tenha sido o primeiro instrumento 'ético-normativo' realizado para se construir uma IA responsável e confiável.

1.3. Os impactos da Inteligência Artificial na sociedade

Apesar de mencionar certa preocupação com os impactos que a singularidade e uma IA superinteligente poderiam causar, não há notícias do desenvolvimento bem-sucedido de uma IA forte. Portanto, não há motivos, por hora, de empreender energia para analisar os impactos inerentes a esse nível de inteligência.

Contudo, não se pode dizer o mesmo da IA fraca, que definitivamente já está em todos os lugares. O Google, por exemplo, se vale da IA tanto em seu mecanismo de pesquisa[128], quanto na indicação de novos vídeos no Youtube; o Facebook, por outro lado, utiliza a IA para publicidade direcionada e marcação de fotos[129]; Microsoft e Apple alimentam seus assistentes digitais, Cortana e

[124] Bostrom, Nick. Superintelligence: Paths, dangers, strategies. Oxford: Oxford University Press, 2014, p.26.
[125] Müller, Vincent C. and Bostrom, Nick. Future progress in artificial intelligence: A survey of expert opinion. Fundamental Issues of Artificial Intelligence. Springer, 2016, 553-571, p. 2-3
[126] Ibid.
[127] Vincent C. Müller (2014) Risks of general artificial intelligence, Journal of Experimental & Theoretical Artificial Intelligence, 26:3, 297-301, DOI: 10.1080/0952813X.2014.895110, p. 298.
[128] Coeckelbergh, Mark. Ética da IA . MIT Pressione. Edição do Kindle, p. 4-5.
[129] Ibid.

Siri, justamente com IA; carros autônomos, drones ou até mesmo armas autônomas que podem matar sem intervenção humana também são guiados por sistemas de IA[130]; no domínio da justiça, o COMPAS nos Estados Unidos, já toma decisões nos tribunais ao prever quem provavelmente poderá reincidir no crime[131]; no domínio da vigilância, a IA promove o reconhecimento facial para distinguir, não só a identidade, mas também as emoções[132].

Nesse contexto, é percetível que a presença da IA, ao mesmo tempo que radicaliza o estilo de vida das pessoas ao promover inúmeros benefícios, também pode apresentar, indiretamente, diversos riscos que, se não forem devidamente endereçados, podem causar um dano permanente e irreparável para os interesses da sociedade e os direitos dos indivíduos. A seguir serão endereçados riscos que o mal uso de sistemas de IA pode causar para áreas como: (i) privacidade e vigilância; (ii) manipulação do comportamento; (iii) opacidade dos sistemas de IA; e (iv) viés e preconceito em sistemas de decisão.

1.3.1. Privacidade e vigilância

Em seu livro, Homo Deus, o historiador e filósofo Yuval Harari levantou a seguinte questão: "O que acontecerá com a sociedade, a política e a vida cotidiana quando algoritmos não conscientes, mas altamente inteligentes, nos conhecerem melhor do que nós mesmos?[133]". Esse questionamento não é feito sem nenhum motivo, na verdade, Harari o faz com base no que se tornou o modelo de negócios da internet, que é basicamente composto por dados pessoais e serviços personalizados baseados em IA, que surpreendentemente não possuem nenhum custo financeiro ao utilizador, que realiza o 'pagamento' ao ceder esferas de sua privacidade.

Esse 'modelo de pagamento' surge pela necessidade de se obter mais dados para desenvolver melhores algoritmos. Isso porque o sucesso atual dos sistemas de IA não se baseia apenas no acesso a poder computacional barato, mas sim na disponibilidade de grandes quantidades de dados[134]. De facto, quanto maior a disponibilidade de dados, melhor será a aprendizagem de um algoritmo. Isso é algo que se extrair da fala de Peter Norvig, que na qualidade de cientista chefe do Google afirmou que "não temos algoritmos melhores que ninguém, temos apenas mais dados"[135].

É provável que o modelo de sucesso da Google também seja aplicado a outras empresas do segmento da internet, como Facebook, Amazon, Apple e Microsoft. Dessa forma, se há uma 'corrida

[130] Ibid.
[131] Ibid.
[132] Ibid.
[133] Harari, Y. N. (2016). Homo deus: A brief history of tomorrow. New York: Harper
[134] Ebers, Martin, Chapter 2: Regulating AI and Robotics: Ethical and Legal Challenges... p.22.
[135] Norvig, quoted by Scott Cleland, Google's "Infringenovation" Secrets, Forbes, October 3, 2011, https://www.forbes.com/sites/scottcleland/2011/10/03/googles-infringenovation-secrets/#78a3795430a6.

pela IA mais poderosa', ela ocorre da seguinte forma: "quanto mais utilizadores a empresa possui, mais dados pessoais podem ser coletados e tratados para treinar os algoritmos. Por sua vez, isso leva a melhores produtos e serviços, o que resulta em mais clientes e mais dados"[136].

Dentro desse contexto, não é tão difícil prever que o mercado irá impor indiretamente uma forte barreira de entrada para novos interessados. Isso torna o mercado de desenvolvimento de sistemas de IA restrito a poucas e específicas empresas, criando-se assim um verdadeiro oligopólio digital[137]. Aliás, é exatamente no sentido de confirmar essa 'previsão' que Pedro Domingues, em seu livro 'O Algoritmo Mestre', afirma que "o controle dos dados e a posse dos modelos aprendidos com eles são as metas ao redor das quais muitas das batalhas do século XXI ocorrerão – entre governos, empresas, sindicatos e pessoas"[138].

Isso posto, através de uma completa relação de interdependência entre dados e IA, que se percebe um aumento no uso de técnicas de vigilância do comportamento geral da sociedade, de modo a possibilitar e ampliar a captação de mais dados de natureza pessoal. Nesse contexto, até mesmo ambientes historicamente protegidos[139], como nossas próprias casas, passaram a ser vigiados através de sensores e assistentes virtuais.

Por sinal, esses dispositivos, que são instalados pela própria vontade do utilizador vigiado, são parte de uma estrutura muito maior chamada de Internet das Coisas (*Internet of Things* - IoT). A IoT é uma expressão cunhada para buscar a designação de "todo o conjunto de novos serviços e dispositivos que reúnem ao menos três pontos elementares: (i) conectividade, (ii) uso de sensores e (iii) capacidade computacional de processamento e de armazenamento de dados"[140]. Nesse contexto, e de acordo com Magrani, "computadores, sensores e objetos (artefactos) interagem uns com os outros e processam as informações/dados em um contexto de hiperconectividade"[141]. Onde hiperconectividade pode ser entendida como a "estreita relação entre seres humanos, objetos físicos, sensores, algoritmos, *Big Data*, Inteligência Artificial, *cloud computing*, entre outros elementos"[142].

Tudo isso, é claro, modifica de alguma forma a maneira como a sociedade vive, possibilitando inúmeros benefícios e oportunidades às pessoas, desde maior conforto, onde sensores podem adequar um ambiente profissional ou residencial de uma forma automática, tornando mais agradável e promovendo eficiência energética; até mesmo a questões de saúde, através de um monitoramento

[136] Ebers, Martin, Chapter 2: Regulating AI and Robotics: Ethical and Legal Challenges... p.22.
[137] Iyengar, Vinod. Why AI Consolidation Will Create the Worst Monopoly in U.S. History, TECHCRUNCH, 2016. Disponível: https://techcrunch.com/2016/08/24/why-ai-consolidation-will-create-the-worst-monopoly-in-us-history/ (o artigo de internet explica as principais empresas de tecnologia fizeram para adquirir a maioria das mais promissoras startups)
[138] Domingos, Pedro. O Algoritmo Mestre: como a busca pela melhor máquina de aprendizado refaz o nosso mundo. Novatec Editora, 2015, p. 46
[139] Ebers, Martin, Chapter 2: Regulating AI and Robotics: Ethical and Legal Challenges... p.23.
[140] Magrani, Eduardo. Entre dados e robôs: ética e privacidade na era da hiperconectividade. 2. edição. Porto Alegre: Arquipélago Editorial, 2019, p.24-25.
[141] Magrani, Eduardo. Entre dados e robôs... p. 25.
[142] Ibid.

constante e eficiente entre paciente e médico[143]. No entanto, deve também ser destacado que incontáveis dispositivos conectados "nos acompanharão diária e constantemente em nossas rotinas, [e estarão a] coletar, transmitir, armazenar e compartilhar uma quantidade enorme de dados, muitos deles estritamente particulares e mesmo íntimos"[144]. Dessa forma, com essa constante exposição da vida privada, sem que sequer se tenha perceção concreta sobre isso, desperta a atenção para riscos inerentes a própria privacidade das pessoas.

Nesse contexto, através de um estudo sobre a Internet das Coisas[145], a própria *Federal Trade Commission* dos Estados Unidos demonstrou certa preocupação quanto aos riscos de privacidade inerentes a IoT, uma vez que ela permite a coleta de informações pessoais sensíveis, hábitos, geolocalização precisa, números de contas financeiras, informações sobre saúde e condições físicas ao longo do tempo[146]. Através desse mesmo estudo, a *Federal Trade Commission* estima que a vigilância através da IoT gera diariamente um volume de até 150 milhões de pontos de dados[147] para cada 10.000 família[148].

No entanto, a IoT não é a única responsável pela criação desses grandes volumes de dados. Outra grande parcela do recolhimento de dados na Internet, que reforça a ideia da cultura de vigilância, ocorre automática e involuntariamente como resultado apenas das atividades *online* e *offline* do indivíduo[149]. De acordo com a *European Data Protection Supervisor* (EDPS), essas informações são denominadas como "migalhas de pão digitais"[150], em outras palavras, um rastro ou pegadas digitais deixadas pelo indivíduo. Essas informações incluem "horários e locais em que os dispositivos móveis se conectam com torres de telefonia móvel ou satélites *GPS*, endereços *IP* dos terminais, pontos de acesso *WiFi*, histórico de navegação, *'likes'* e *'shares'*, imagens coletadas por sistemas de CFTV digital, histórico de compras, engajamento em redes sociais e comportamento de navegação entre dispositivos"[151].

Existem também um conjunto de dados que são fornecidos voluntariamente pelas pessoas, tanto pela exposição de dados e informações que são disponibilizadas publicamente em redes sociais, quanto pelo preenchimento de formulários *online*[152]. Aliás, dados sensíveis são coletados através da

[143] Ibid.
[144] Ibid.
[145] Federal Trade Commission. Internet of Things: privacy & security in a connected world. FTC Staff Report, 2015
[146] Federal Trade Commission. Internet of Things: privacy & security in a connected world... p.14.
[147] De acordo com o site whatis.com: "Um 'data point' é uma unidade discreta de informações. Em um sentido geral, qualquer fato é um 'data point'. Em um contexto estatístico ou analítico, um ponto de dados geralmente é derivado de uma medida ou pesquisa e pode ser representado numericamente e/ou graficamente. O termo é equivalente a datum, a forma singular de dados." Definição disponível em: http://whatis.techtarget.com/definition/data-point. Acesso em: 03 de agosto de 2020.
[148] Federal Trade Commission. Internet of Things: privacy & security in a connected world... p.14.
[149] European Data Protection Supervisor (EDPS). Opinion 3/2018 on online manipulation and personal data. 2018, p.7.
[150] EDPS, Opinion 3/2018 on online manipulation and personal data... p.8.
[151] Ibid.
[152] Ibid.

realização *online* de testes psicológicos e de personalidade[153], que "quando combinados com os detalhes pessoais disponíveis nas redes sociais permitem uma previsão de personalidade intrincada"[154].

Por conta desse volume de dados pessoais, e com a utilização de algoritmos de aprendizagem de máquina e métodos estatísticos, tornou-se possível uma análise granular dos aspetos privados da vida de um indivíduo. O que é chamado atualmente de *Profiling*[155]. Diante de tanta informação, essa técnica permitiu a criação de perfis de utilizadores, que eram construídos a partir de uma infinidade de características físicas (como idade, altura, peso, etnia etc.) e comportamentais, tais como hábitos, preferências pessoais, sinais de personalidade, e todo e qualquer dado ou metadado que pudesse permitir a organização e categorização em um determinado perfil[156].

Além do *Profiling* possibilitar a identificação de padrões que seriam invisíveis ao olho humano, os métodos estatísticos inerentes a essa técnica também possibilitavam "traçar um quadro das tendências de futuras decisões, comportamentos e destinos de uma pessoa ou grupo"[157]. Através desses métodos de predição e análise comportamental, passou a ser possível construir uma expectativa de comportamento futuro e predefinido para cada tipo de perfil.

Magrani afirma que "o perfil formado se torna uma representação virtual da pessoa"[158], o que pode acarretar forte diminuição de suas liberdades individuais, uma vez que aqueles que tomam decisões com base nesse perfil virtual podem levar em consideração apenas análises preditivas e predefinidas[159] relacionadas a um padrão de comportamento, o que implica na completa negação da pessoalidade e individualidade de quem foi perfilado.

Nesse sentido, ao ceder a privacidade em troca de comodidade e entretenimento, cria-se um ambiente de vigilância constante. De modo que gradualmente a privacidade perde o seu valor, ao passo que o nosso comportamento, silenciosamente, torna-se um produto, uma matéria-prima, que sem percebermos é utilizado contra os nossos próprios interesses.

1.3.2. Manipulação do comportamento

O impacto que a IA tem para os aspetos de privacidade e vigilância já são, de certa forma, preocupantes. O potencial da IA em analisar grandes volumes de dados e o *profiling* de utilizadores para se obter uma projeção futura e predeterminada de comportamentos promovem, juntos, uma

[153] The Atlantic. The Dark Side of That Personality Quiz You Just Took, 2020. https://www.theatlantic.com/technology/archive/2017/07/the-internet-is-one-big-personality-test/531861/.
[154] EDPS, Opinion 3/2018 on online manipulation and personal data... p.8.
[155] Clarke, R. Profiling: A hidden challenge to the regulation of data surveillance. Journal of Law and Information Science, Australia, v. 4, n. 2, 1993. Disponível: vhttp://www.austlii.edu.au/au/journals/JILawInfoSci/1993/26.html
[156] Doneda, Danilo. Da Privacidade à Proteção de Dados Pessoais, Rio de Janeiro: Renovar. 2006, p.173
[157] Ibid.
[158] Magrani, Eduardo. Entre dados e robôs... p.127.
[159] Doneda, Danilo. Da Privacidade à Proteção de Dados Pessoais... p.173

crescente e imparável cultura da vigilância. Ocorre que, esse estado de vigilância constante de nosso comportamento pode ser apenas uma primeira parte de várias etapas.

Através da Opinião 3/2018, a EDPS afirma que a manipulação de comportamentos em ambiente digital é percebida através do culminar de um ciclo de três etapas[160], onde a primeira é a própria coleta de dados; a segunda é a criação de perfis; e a terceira a segmentação através do *microtargeting*, técnica que veio a se tornar uma promissora estratégia de *marketing* para os negócios.

Baseada na recolha e na análise psicométrica[161] de dados pessoais, o *microtargeting* tem a intenção de criar um perfil que seja passível de análise e previsão de comportamentos, preferências e interesses, tanto de indivíduos isolados, quanto de pequenos grupos aos quais ele faz parte[162]. Dessa forma, ao invés de realizar publicidade de uma forma genérica e deliberada, uma empresa pode abordar seus consumidores de uma forma individual levando em consideração exatamente a sua personalidade, seus interesses e seus comportamentos, que são previamente definidos através do seu perfil[163].

Assim como tudo que envolve tecnologia, sempre é possível perceber alguma espécie de benefício. De facto, esse tratamento com base na personalidade e em interesses permite que ações promocionais sejam completamente personalizadas, fornecendo a um indivíduo uma experiência única. Nesse sentido, através do *microtargeting*, toda a experiência em ambiente digital é personificada, e por mais que duas pessoas tenham os mesmos amigos em uma rede social, o *feed* de cada uma será único, apresentando a ela mensagens, vídeos e postagens daqueles amigos que, aparentemente, ela possui maior interesse. A ideia se repete ao aceder um programa de *streaming* de vídeo, como Youtube ou Netflix, ou até mesmo ao procurar por notícias do dia a dia, onde preferencialmente irão ser apresentadas aquelas que são compatíveis com seu perfil de utilizador.

No entanto, com a intenção de "influenciar e manipular o comportamento de cidadãos e consumidores"[164], o *microtargeting* já vem sendo utilizado por empresas e partidos políticos. Ora, nesse sentido, não é nenhuma novidade que essa técnica já foi utilizada em campanhas políticas, uma prática que resultou no conhecido escândalo da *Cambridge Analytica*[165]. A esse respeito, dados pessoais de 50 milhões de utilizadores do *Facebook* foram violados a fim de se construírem perfis de

[160] EDPS, Opinion 3/2018 on online manipulation and personal data... p.7.
[161] Ibid.
[162] De acordo com o Instituto Brasileiro de Pesquisa e Análise de Dados, psicometria pode ter os seguintes entendimentos: "No Marketing, a Psicometria pode prever a receptividade do público à um determinado produto e diagnosticar as características de programas ou campanhas em função da sua eficácia na transmissão de mensagens eficácias a seu público. A psicometria tem ganhado cada vez mais espaço em outros contextos como nas Mídias Sociais, onde a enorme quantidade de dados disponíveis garante à analistas de diversas áreas tomarem melhores decisões com base em medidas psicológicas como as de personalidade ou de orientação moral". Acessado em 05 de agosto de 2020. Disponível em https://www.ibpad.com.br/blog/comunicacao-digital/o-que-e-psicometria/
[163] EDPS, Opinion 3/2018 on online manipulation and personal data... p.7-9.
[164] Ebers, Martin, Chapter 2: Regulating AI and Robotics: Ethical and Legal Challenges... p.29.
[165] Cf. o discurso de Alexander Nix, ex-CEO da Cambridge Analytica, na Concordia Annual Summit 2016 em Nova York, https://www.youtube.com/watch?v=n8Dd5aVXLCc;

eleitores americanos[166].

Nesse contexto, Christopher Wylie, antigo funcionário da Cambridge Analytica, denunciou a estratégia de coleta e tratamento ilegal de dados que tinha como objetivo "enviar mensagens direcionadas a eleitores específicos, manipulando sua opinião política através de um algoritmo que conseguia analisar os perfis individuais e determinar traços de personalidade ligados ao comportamento *online* do eleitor, bem como seus sentimentos e medos, direcionado o conteúdo de manipulação sociopolítica com base nesses fatores"[167].

Além da manipulação em esfera política, o *microtargeting* também é aplicado no mercado de consumo, o que tem ocasionado um enfraquecimento da autonomia privada das pessoas[168]. Afirma-se isso porque, as perceções comportamentais e de personalidade, que inferem preferências futuras, permitem que as empresas "adaptem os contratos de forma a maximizar sua utilidade esperada, explorando as vulnerabilidades [...] de seus clientes"[169].

Nesse contexto, a técnica do microtargeting possibilita, por exemplo, que uma empresa ofereça determinado serviço ou produto exatamente quando o cliente está mais propenso a tomar uma decisão favorável à compra[170]. Dessa forma, "com base nos resultados de uma economia comportamental, as empresas podem explorar ou mesmo induzir comportamentos de tomada de decisão [...] nos seus clientes"[171], o que configura um abuso e, como mencionado, um potencial enfraquecimento da autonomia privada de um indivíduo.

1.3.3. Opacidade e a falta de explicabilidade das decisões

Quando alguém se vale de uma metáfora para se referir a uma 'caixa preta', somente o contexto poderá determinar o seu sentido. Isso porque uma 'caixa preta', que na verdade possui a cor laranja, pode ser uma referência a "dispositivos de gravação, como os sistemas de monitoramento de dados em aviões, trens e carros"[172]. No entanto, quando o termo 'caixa preta' é utilizado ambiente de IA, comumente ele se refere a "um sistema [de aprendizagem de máquina] cujo funcionamento é misterioso, [onde] podemos observar suas entradas e saídas [...]"[173], mas "sem entender parcial ou

[166] Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. The Guardian. Disponível: https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.
[167] Magrani, Eduardo. Entre Dados e Robôs... p. 161.
[168] Ebers, Martin, Chapter 2: Regulating AI and Robotics: Ethical and Legal Challenges... p.32; é possível entender o assunto de forma aprofundada através de: Mik, Eliza. The Erosion of Autonomy in Online Consumer Transactions. (2016). *Law, Innovation and Technology.* 8, (1), 1-38. Rikesearch Collection School Of Law. **Disponível:** https://ink.library.smu.edu.sg/sol_research/1736
[169] Ebers, Martin, Chapter 2: Regulating AI and Robotics: Ethical and Legal Challenges... p.32.
[170] Ibid.
[171] Ibid.
[172] Pasquale, Frank. The black box society: the secret algorithms that control money and information. Cambridge: Harvard University Press. 2015, p.3.
[173] Ibid.

completamente como diferentes características influenciam a previsão do modelo"[174]. Nesse sentido, sempre que for verificado esse mistério ou uma impossibilidade de explicar o porquê um algoritmo tomou determinada decisão, constata-se o que se chama de opacidade[175].

Ocorre que o termo 'opacidade' pode possuir três diferentes significados. O primeiro, quando o próprio sistema baseado em IA possui uma complexidade significativamente alta onde é impossível para um utilizador sem conhecimentos técnicos ou científicos entender o seu funcionamento[176]. O exemplo apropriado se refere a um condutor que não sabe como funcionam os sensores de um carro autônomo. O segundo significado está ligado a opacidade intencional provocada pelo pelos próprios *developers* ou proprietários de um determinado algoritmo inteligente, o que lhe é permitido por lei através da proteção intelectual e do consequente sigilo do funcionamento[177]. Como exemplo, pode-se citar os próprios algoritmos da Google e do Facebook, responsáveis pelos seus serviços de busca e de publicidade.

Por fim, o terceiro e último significado é aquele que importa para caracterizar mais um dos impactos da IA na sociedade, sendo aquilo que consiste na completa inexplicabilidade, até mesmo para os seus *developers*, de como um algoritmo de aprendizagem de máquina infere seus resultados[178]. Esse fenômeno existe e é comumente chamado de 'opacidade algorítmica', e "embora saibamos que os algoritmos de aprendizagem de máquina dependem de correlações estatísticas entre as características da entrada e o alvo, não é possível saber quais as características que o algoritmo [mais valoriza]"[179].

Nesse sentido, é importante perceber que essa opacidade algorítmica está presente em inúmeras aplicações inteligentes utilizadas diariamente. No entanto, cada uma delas pode produzir uma resposta que pode variar de algo relativamente simples, como uma mensagem de que um restaurante corresponde a uma preferência, ou uma previsão, pontuação ou um ranking[180] relativo a algo significativamente impactante, como: 'o diagnóstico de cancro do paciente é positivo, 'a solicitação de crédito foi recusada, 'esse candidato não se enquadra no perfil profissional desejado', 'réu com alta probabilidade de reincidência criminosa' ou 'alvo identificado e eliminação iniciada'[181].

Como se pode perceber, muitas das inferências e dos resultados apresentados por um algoritmo inteligente afetam significativamente a vida e os direitos de um indivíduo, e por conta disso a

[174] Anand, Avishek, Bizer, Kilian, Erlei, Alexander et al. Effects of Algorithmic Decision-Making and Interpretability on Human Behavior: Experiments using Crowdsourcing, 2018, p.1. Disponível: www.l3s.de/~gadiraju/publications/HCOMP18.pdf
[175] European Parliament. Artificial intelligence: From ethics to policy. EPRS | European Parliamentary Research Service. Scientific Foresight Unit (STOA). PE 641.507 – June 2020, p.7-8.
[176] Ibid.
[177] Ibid.
[178] Ibid.
[179] Ibid.
[180] Anand, Avishek, Bizer, Kilian, Erlei, Alexander et al. Effects of Algorithmic Decision-Making and Interpretability on Human Behavior...p.2.
[181] Inspirado nos exemplos de Müller, Vincent C. (forthcoming 2021), 'Ethics of artificial intelligence', in Anthony Elliott (ed.), The Routledge social science handbook of AI (London: Routledge), 20pp.

explicabilidade e a interpretabilidade são imprescindíveis por diversas razões[182]. Primeiramente, para que os próprios pesquisadores ou *developers* possam "entender como seu sistema ou modelo está funcionando a fim de depurar ou melhorá-lo"[183]. Por outro lado, para todo o indivíduo que, de alguma forma, vier a ser afetado por uma decisão algorítmica, possa compreender como e por que o sistema chegou a essa decisão.

Nesse sentido, uma maior opacidade em detrimento da explicabilidade resulta em três grandes problemas[184]. O primeiro, o de uma desconfiança constante sobre a tecnologia. O segundo, a impossibilidade que indivíduos afetados se valerem dos remédios jurídicos apropriados em caso de uma decisão algorítmica ilegal. E por fim, a impossibilidade de auditar o processo decisório e perceber quais características e dados pessoais estão influenciando diretamente numa decisão, o que impede, principalmente, o combate a decisões discriminatórias tomadas pelos algoritmos.

1.3.4. Injustiça e preconceito através de decisões algorítmicas baseadas em dados enviesados

É completamente redundante afirmar a importância dos dados para o treino da IA. Também seria repetitivo relembrar que é através de dados pessoais e características físicas e psicológicas que algoritmos de IA buscam padrões a fim de classificar, segmentar e perfilar utilizadores, possibilitando que sejam tomadas decisões preditivas sobre seus comportamentos. Tudo isso já foi, de facto, abordado. Contudo, a depender de quais dados são disponibilizados ou mais valorizados no momento de treinar uma IA, é possível que um algoritmo possa começar a apresentar decisões injustas ou até mesmo discriminatórias.

Há uma ficção cuidadosamente construída a respeito de uma visão benevolente da IA[185], e que por ser uma tecnologia neutra, baseada em matemática e estatística, jamais teria a vontade, a sensibilidade ou a intenção deliberada de discriminar uma pessoa com base em sua cor, sexo, idade, etnia, religião, orientação sexual, posição política ou qualquer outra condição física ou mental possível de caracterizar um indivíduo dentro de uma minoria ou de um grupo vulnerável.

Contudo, não se pode acreditar que algoritmos sejam isentos de subjetividade, erro ou manipulação[186]. Isso porque, os dados, matéria-prima que treina, norteia e fundamenta as decisões

---

[182] Anand, Avishek, Bizer, Kilian, Erlei, Alexander et al. Effects of Algorithmic Decision-Making and Interpretability on Human Behavior...p.2.
[183] Ebers, Martin, Chapter 2: Regulating AI and Robotics: Ethical and Legal Challenges... p.12.
[184] Ibid.
[185] Gillespie, Tarleton. "The relevance of algorithms". In (Ed.), Media Technologies: Essayson Communication, Materiality, and Society. Cambridge: The MIT Press, 2014.
[186] Bruno Ricardo Bioni e Maria Luciano. O Princípio da Precaução na Regulação de Inteligência Artificial: seriam as Leis de Proteção de Dados o seu portal de entrada? Inteligência Artificial e Direito, 2ª edição. Revista dos Tribunais, 2020, p. 206

algorítmicas, são um recorte, ou melhor, um reflexo da sociedade[187]. Assim sendo, por representarem e quantificarem os valores que guiam a forma de agir e de pensar de indivíduos e organizações, os dados gerados em um contexto possivelmente discriminatório irão treinar e nortear, de modo enviesado, a forma como um algoritmo de IA tomará as suas futuras decisões. Nesse sentido, e de acordo com Cortiz, é preciso que se tenha cuidado com a fonte e com a qualidade dos dados, isso porque eles influenciarão diretamente como o sistema aprenderá e se comportará, o que pode culminar na geração de impactos em larga escala[188].

Nesse contexto, a fim de comprovar tais afirmações, demonstra-se importante apresentar ao menos dois casos reais em que dados enviesados provocaram decisões algorítmicas injustas e discriminatórias.

O primeiro caso com grande repercussão no meio acadêmico ocorreu nos EUA e foi exposto por uma investigação conduzida pela agência de jornalismo ProPublica[189]. O ponto central girou em torno do uso de um algoritmo no setor da justiça criminal, chamando de COMPAS. Esse sistema realizava uma avaliação de risco sobre cada condenado com a finalidade de prever a potencial probabilidade de reincidência criminal desse indivíduo. O resultado dessa avaliação interferia diretamente na possibilidade de liberdade condicional do preso.

O problema discriminatório desse algoritmo foi evidenciado quando a ProPublica, ao ter acesso a aproximadamente 7.000 (sete mil) avaliações de risco respetivas a um condado da Flórida, percebeu uma completa disparidade racial referente as avaliações. Ao analisar cada avaliação de risco e verificar se o réu veio ou não a reincidir criminalmente nos 2 anos seguintes após sua libertação, a ProPublica percebeu que "a fórmula tinha uma probabilidade particular de sinalizar falsamente os réus negros como futuros criminosos, rotulando-os erroneamente dessa forma quase duas vezes mais que os réus brancos"[190]. Enquanto isso, "réus brancos foram erroneamente rotulados como de baixo risco com mais frequência do que réus negros"[191].

A Northpoint, empresa com fins lucrativos que criou o algoritmo COMPAS, contesta a análise realizada pela ProPublica. No entanto, há aqui um problema talvez tão grave quanto a discriminação, a empresa não divulga publicamente os cálculos usados para chegar às pontuações de risco dos réus, portanto, não é possível realizar qualquer auditoria em seu código. Dessa forma, tanto os réus quanto o público em geral não conseguem perceber quais fatores estão a causar a grande disparidade dos resultados. Por não terem acesso à fundamentação da decisão que avaliou um indivíduo como de grave

[187] Cortiz, Diogo. O Design pode ajudar na construção de Inteligência Artificial humanística? 17° ERGODESIGN – Congresso Internacional de Ergonomia e Usabilidade de Interfaces Humano Tecnológica: Produto, Informações Ambientes Construídos e Transporte. Disponível: http://pdf.blucher.com.br.s3-sa-east-1.amazonaws.com/designproceedings/ergodesign2019/1.02.pdf

[188] Ibid.

[189] Angwin, J. et al. Machine Bias. ProPublica, 2016. Disponível: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[190] Ibid.

[191] Ibid.

risco, o próprio direito de acesso a justiça fica limitado, pois é impossível contestar uma decisão que não tem fundamento.

O segundo caso está relacionado com questões de discriminação no ambiente de trabalho, mas que foi percebido pela própria empresa e já foi descontinuado. Nessa oportunidade, os próprios especialistas em aprendizagem de máquina da *Amazon.com Inc*. descobriram que um algoritmo responsável pelo recrutamento da empresa discriminava mulheres em prol de homens[192]. A ideia do algoritmo era a de receber como *input* uma grande quantidade de currículos, analisá-los e apontar os melhores candidatos para serem contratados. No entanto, foi percebido que não havia neutralidade de gênero no novo sistema, pois ele não estava classificando candidatas mulheres para determinados cargos específicos da área de tecnologia da empresa.

Os próprios especialistas da Amazon afirmaram que esse fato ocorreu pois, ao treinarem o algoritmo para realizar essa tarefa, o alimentaram com currículos que foram enviados à empresa ao longo de um período de 10 anos. Ocorre que o algoritmo percebeu um padrão naqueles currículos, uma vez que a maioria dos candidatos eram do sexo masculino[193]. Aliás, aparentemente "o sistema da Amazon ensinou a si mesmo que candidatos do sexo masculino eram preferíveis. Ele penalizava currículos que incluíam a palavra 'feminino', como em 'capitão do clube de xadrez feminino'[194].

Como se pode perceber, em todos os casos há um problema nítido que envolve tanto a injustiça quanto a discriminação causada por algoritmos. Nesse contexto, pode-se afirmar que preconceitos culturais e estereótipos foram determinantes para que o algoritmo tomasse sua decisão, mesmo sem saber exatamente o que aquilo definitivamente significava. Como Cortiz afirmou, é preciso entender que os dados são um reflexo da sociedade[195], e nesse sentido, ter cuidado com a própria fonte e com a qualidade dos dados, uma vez que eles irão influenciar diretamente como o sistema aprenderá e se comportará.

O problema em questão remonta a como ser justo com diferentes grupos, sejam minorias, maiorias ou vulneráveis, onde 'justo' implica na necessidade de atenção e cuidado não só com a qualidade e veracidade dos dados, mas também com a representatividade precisa de todos os grupos e comunidades[196]. O impacto social da utilização incorreta de dados, não só no setor privado, mas principalmente no setor público, pode ser catastrófico, marginalizando e limitando cada vez mais o exercício de direitos por parte indivíduos pertencentes a um determinado tipo de grupo.

Nesse sentido, a fim de evitar o estímulo cada vez maior de injustiça e a discriminação

---

[192] Dastin, Jeffrey. Amazon scraps secret AI recruiting tool that showed bias against women. Reuters, 2018. Disponível: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruting-tool-that-show-bias-against-women-idUSKCN1MK08G

[193] Ibid.

[194] Ibid.

[195] Cortiz, Diogo. O Design pode ajudar na construção de Inteligência Artificial humanística?...

[196] European Parliament. Artificial intelligence: From ethics to policy... p.8-9.

algorítmica, assim como fortalecer o enfrentando à cultura de vigilância e à manipulação de comportamentos de indivíduos, torna-se cada vez mais necessário aprofundar o debate sobre como regular e minimizar os impactos inerentes ao uso da IA. Isto posto, para cumprir tal objetivo, este trabalho irá se dedicar ao estudo das diretrizes europeias voltadas para a IA. Tais diretrizes possuem uma abordagem baseada na confiança e no objetivo de construir uma orientação ética para o desenvolvimento de uma IA de confiança.

CAPÍTULO SEGUNDO

A ESTRATÉGIA EUROPEIA PARA A INTELIGÊNCIA ARTIFICIAL E OS ASPECTOS JURÍDICOS E ÉTICOS NECESSÁRIOS PARA O DESENVOLVIMENTO DE UMA IA CONFIANÇA.

2.1. A Inteligência Artificial na União Europeia

Apesar da existência de graves riscos provocados por sistemas de IA, as oportunidades trazidas por essa tecnologia aperfeiçoam gradualmente a nossa sociedade. Esse contexto faz com que seja necessário discutir o uso e desenvolvimento da IA. Tais discussões tornam-se ainda mais imperativas ao se levar em conta que não há, a nível da União Europeia, um diploma jurídico que regule especificamente a IA[197]. Isso implica que qualquer análise jurídica sobre a IA seja realizada com base nos direitos fundamentais e em regulações setoriais, como é o caso da proteção de dados[198]. Contudo, por mais que existam bases legais para análise da IA, a ausência de um diploma específico permite possíveis lacunas que poderão ser prejudiciais para a proteção de indivíduos e da sociedade.

Diante desse cenário, é importante destacar que a União Europeia está ciente de todas essas preocupações, pelo que desde 2018 vem adotando políticas e estratégias nessa área. A própria Comissão Europeia (Comissão) já apresentou planos de desenvolvimento e adoção de sistemas de IA visando o fortalecimento do Mercado Único Digital. Além disso, paralelamente, a Comissão também se mantém alerta ao avaliar os impactos potencialmente prejudiciais da IA. Aliás, apesar da referida lacuna legislativa, é importante destacar as diversas iniciativas relativas ao estudo de um possível quadro regulamentar para a IA, onde, atualmente, a Comissão dedica seus esforços para desenvolver diretrizes políticas e orientações jurídicas que norteiam a regulação e o desenvolvimento de uma IA ética, de confiança e centrada em valores humanos[199].

O ideal europeu de uma IA de confiança foi recentemente materializados através das 'Orientações Éticas para uma IA de Confiança'[200], desenvolvidas em 2019 pelo Grupo de Alto Nível em IA (GPAN IA). No entanto, antes de analisá-la profundamente, mostra-se pertinente entender a estratégia europeia para a IA que vinha sendo desenvolvida anteriormente. Assim sendo, a próxima seção irá se dedicar a analisar cronologicamente os documentos que compõem a referida estratégia, o

[197] Após finalizar toda a pesquisa, essa dissertação foi surpreendida com a proposta de regulamento que estabelece regras harmonizadas em matéria de inteligência artificial realizada pela Comissão Europeia. Contudo, é importante destacar que ainda se trata de uma proposta e que precisará ser analisada e aprovada, não possuindo força legal vinculante. Acessível: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence

[198] The impact of the General Data Protection Regulation (GDPR) on artificial intelligence. EPRS | European Parliamentary Research Service Scientific Foresight Unit (STOA) PE 641.530. 2020, p.32.

[199] Comissão Europeia. Livro Branco sobre a inteligência artificial - Uma abordagem europeia virada para a excelência e a confiança. COM (2020) 65 final. Bruxelas, 2020, p.2.

[200] Grupo de Peritos de Alto Nível em IA (GPAN IA). Orientações Éticas para uma IA de Confiança. 2019

que permitirá um maior entendimento da dimensão, do contexto e dos agentes envolvidos atualmente nos assuntos relacionados a IA no cenário europeu.

Apenas se note que a próxima seção tem, de facto, a intenção de entender a estratégia da Comissão Europeia para a IA através de uma perspetiva conceitual e política, sem adentrar especificamente em conceitos jurídicos. O estudo e a análise dos aspetos jurídicos de uma IA de confiança será realizado na seção 2.3.

2.2. A estratégia europeia para a Inteligência Artificial: uma abordagem centrada na confiança e nos valores humanos.

2.2.1. Declaração de Cooperação sobre a Inteligência Artificial

No dia 10 de abril de 2018, vinte e cinco (25) Estados-Membros assinaram a 'Declaração de Cooperação sobre a IA'[201]. Diferente de regulamentos e de diretivas, uma declaração é um documento sem força jurídica vinculativa, ou seja, por mais que se trate de uma declaração oficial, ela apenas descreve a intenção dos seus signatários, não os vinculando legalmente a tomada de nenhuma ação.

Através desta Declaração, os respetivos Estados-Membros concordaram em cooperar e trabalhar em conjunto em três diferentes vertentes. A primeira vertente diz respeito a garantir a competitividade da Europa na investigação e implantação de IA através do reforço da tecnologia e da capacidade industrial de IA da Europa. A segunda, voltada para a educação e requalificação dos cidadãos europeus, visa enfrentar os desafios socioeconómicos criados pela transformação dos mercados de trabalho. A terceira vertente, tendo em consideração os impactos que a IA pode causar em diversas áreas, pretende "garantir um quadro ético e jurídico adequado, baseado nos direitos fundamentais e nos valores da União, incluindo a privacidade e a proteção de dados pessoais, bem como princípios como a transparência e a responsabilização."[202].

Por mais que a Declaração seja enxuta e simplificada, ela cria a base de uma abordagem política tripla baseada num compromisso assumido pelos Estados-Membros de endereçar aspetos como competitividade, educação e regulação através de padrões comuns em termos de investimento. Essa abordagem influenciar diversos outros documentos europeus.

Além disso, é importante destacar que os Estados-Membros têm se comprometido também em lidar com a IA de uma forma abrangente e integrada, e quando necessário, "rever e modernizar as

[201] Declaração de Cooperação sobre a Inteligência Artificial. Disponível: https://ec.europa.eu/digital-single-market/en/news/eu-member-states-sign-cooperate-artificial-intelligence

[202] Stix. Charlotte. A survey of the European Union's artificial intelligence ecosystem. 2019, p.11. Disponível: https://ec.europa.eu/jrc/communities/sites/default/files/ff3afe_1513c6bf2d81400eac182642105d4d6f.pdf

políticas nacionais para assegurar que as oportunidades decorrentes da IA sejam aproveitadas e os desafios emergentes abordados"[203]. Ainda nesse sentido, destaca-se que a Declaração também reforçou sua preocupação quanto a uma IA antropocêntrica, uma vez que ela pretende "garantir que os seres humanos permaneçam no centro do desenvolvimento, implantação e tomada de decisões da IA"[204], além de impedir a criação e o uso prejudiciais de aplicativos de IA.

A Declaração de Cooperação sobre a IA deve ser percebida como um meio pelo qual os Estados-Membros demonstraram fortemente as suas intenções em colaborar no desenvolvimento de um ecossistema europeu de IA. Destaca-se que a harmonia e o trabalho em conjunto devem ser considerados como elementos-chave para o desenvolvimento de uma IA à nível europeu. Isso porque a cooperação entre Estados-Membros impede uma possível fragmentação política e legislativa, que levaria a erros que poderiam enfraquecer o Mercado Único Digital.

2.2.2. Comunicação da Comissão: Inteligência Artificial para a Europa

Em 25 de abril de 2018, na sequência de um convite do Conselho Europeu para apresentar uma abordagem europeia à IA[205], a Comissão apresentou sua Comunicação 'Inteligência Artificial para a Europa' [206], documento que pode ser denominado como a base da sua estratégia para a IA. Através dessa Comunicação, a Comissão afirmou que "a União Europeia deve adotar uma abordagem coordenada, a fim de tirar o máximo partido das oportunidades oferecidas pela IA e fazer face aos novos desafios que esta tecnologia provoca, [liderando] o caminho no desenvolvimento e utilização da IA para o bem comum, tendo por base os seus valores e pontos fortes"[207].

Perceba que existem dois pontos centrais na proposta da CE. O primeiro é o trabalho coordenado, que confere maior força às políticas europeias face àquelas adotadas por países terceiros e evita, como já mencionado, possíveis políticas fragmentadas. O segundo ponto refere-se ao desenvolvimento de uma tecnologia que, além de ser construída de acordo com os valores humanos, deve necessariamente estar a serviço das pessoas e do bem comum da sociedade.

Para que a União Europeia (União) alcance os objetivos estabelecidos, a referida Comunicação apresenta três passos que possibilitam o fortalecimento do ecossistema europeu de IA, sendo eles: (i) reforçar a capacidade industrial e tecnológica da União e a adoção da IA na economia; (ii) preparar as

[203] Declaração de Cooperação sobre a Inteligência Artificial...
[204] Ibid.
[205] Conselho Europeu. Conclusões do encontro de 19 de outubro de 2017. Bruxelas. Disponível: http://data.consilium.europa.eu/doc/document/ST-14-2017-INIT/en/pdf
[206] Comissão Europeia. Comunicação da Comissão ao Parlamento Europeu, ao Conselho Europeu, ao Conselho, ao Comité Económico a Social e ao Comité Das Regiões: Inteligência artificial para a Europa. COM (2018) 237 final, Bruxelas.
[207] Ibid.

mudanças socioeconómicas; e (iii) garantir um quadro ético e jurídico apropriado a fim de lidar com o desenvolvimento e implantação da IA.

Quanto ao primeiro passo, o de (i) reforçar a capacidade industrial e tecnológica da União e a adoção da IA na economia, a Comissão propõe ações como a de atrair e acelerar investimentos; reforçar a investigação e a inovação; levar a IA a todas as pequenas empresas e potenciais utilizadores; apoiar a realização de testes e a experimentação; e também disponibilizar cada vez mais dados.

Quanto ao segundo passo, o de (ii) preparar as mudanças socioeconómicas relativas à IA, a Comissão possui alguns projetos. Um deles é o de 'Estimular o talento, a diversidade e a interdisciplinaridade' através de "iniciativas que incentivem mais jovens a optarem por carreiras em áreas ligadas à IA e a domínios associados à mesma"[208]. Outro projeto consiste em 'Não deixar ninguém para trás', investindo assim em programas de formação e requalificação profissional para os profissionais que possam ser afetados pelas mudanças proporcionadas pela automatização de operações dentro do mercado de trabalho[209]. Nesse contexto, a Comissão afirma que uma "estratégia de antecipação e de investimento nas pessoas é essencial numa abordagem à IA inclusiva e centrada no ser humano"[210].

Relativamente ao terceiro passo, de (iii) garantir um quadro ético e jurídico apropriado a fim de lidar com o desenvolvimento e implantação da IA, a Comissão acredita que é importante estabelecer uma abordagem sustentável acerca da IA, e para isso é "necessário gerar um clima de confiança e responsabilidade em torno do desenvolvimento e da utilização da IA"[211].

Nesse sentido, apesar da ausência de uma legislação específica referente a IA[212], a Comunicação reforça que a União ainda assim dispõe de um quadro regulamentar sólido e equilibrado, como por exemplo o Regulamento Geral de Proteção de Dados (RGPD), que garante um elevado nível de proteção dos dados pessoais. Dentro desse contexto, a Comissão reforça que acompanhará de perto a aplicação do RGPD no contexto da IA, mas que também confia nas diversas propostas apresentadas no âmbito da estratégia para o Mercado Único Digital, "como o regulamento sobre a livre circulação de dados não pessoais, o Regulamento Privacidade e Comunicações Eletrónicas e o Regulamento Cibersegurança"[213].

[208] Ibid.
[209] Ibid.
[210] Ibid.
[211] Ibid.
[212] À época da Comunicação não existia sequer a proposta.
[213] Ibid.

2.2.3. Plano Coordenado para Promover o Desenvolvimento e a Utilização da IA na Europa

Publicado em 7 de dezembro de 2018 e desenvolvido com base na Declaração de Cooperação sobre a Inteligência Artificial e na Comunicação da Comissão 'Inteligência artificial para a Europeia', o Plano coordenado para o desenvolvimento e uso da inteligência artificial na Europa (Plano Coordenado) tem como seu principal objetivo que "os Estados-Membros e a União [alinhem-se] em esforços para um desenvolvimento responsável da AI no nível global"[214], tornando-se a "região líder mundial em desenvolvimento e implantação de IA de ponta, ética e segura [e] promovendo uma abordagem centrada no ser humano no contexto global"[215].

A ideia de desenvolvimento deste Plano Coordenado surgiu com a intenção de definir coletivamente um caminho a seguir, de modo que a União possa "falar a uma só voz a países terceiros e ao mundo em geral sobre este tema"[216]. Trata-se, portanto, de um fator muito importante, pois a harmonia entre os Estados-Membros nos assuntos que tocam a IA impede fragmentação e insegurança jurídica. Dessa forma, a União segue por um caminho semelhante àquele que originou o RGPD, promovendo união e harmonização em sua aplicação, ao invés de segmentação e insegurança.

Para que esses objetivos sejam alcançados, o Plano Coordenado da Comissão depende da sinergia entre os Estados-Membros e diversas partes interessadas, como empresas de tecnologia, indústria, academia e sociedade civil. Essa cooperação entre inúmeros agentes, dos mais diversos setores, é, talvez, a única saída para que a Comissão possa definitivamente liderar um movimento harmônico e unido que torne possível alcançar a chamada IA de confiança.

Da análise das muitas abordagens promovidas pelo Plano Coordenado, uma delas passa por aumentar os investimentos e reforçar a excelência em tecnologias e aplicativos de IA confiáveis, sendo, além disso, éticos e seguros por *design*. Interessante destacar que a Comissão entende que a 'ética por padrão' deve ser um princípio fundamental para uma IA desenvolvida na Europa, segundo o qual princípios éticos e legais devem ser, necessariamente, implementados desde o início do desenvolvimento de IA. Aliás, o Plano Coordenado afirma que estes princípios deveriam ser construídos com base no RGPD, promoverem conformidade com o direito da concorrência e, por fim, garantir que os dados utilizados para treino da IA sejam livres de qualquer viés, garantindo assim uma luta contra a discriminação e injustiças[217].

---

[214] Comissão Europeia. Comunicação da Comissão ao Parlamento Europeu, ao Conselho Europeu, ao Conselho, ao Comité Económico a Social e ao Comité Das Regiões: Plano Coordenado para a Inteligência Artificial. COM (2018) 795 final. Bruxelas.

[215] Ibid.

[216] Ibid.

[217] Ibid.

2.2.4. Grupo de Peritos de Alto Nível em Inteligência Artificial

A União pretende que a sua IA seja baseada em valores éticos e sociais, tendo como referência a Carta dos Direitos Fundamentais. Para apoiar esse projeto ambicioso, e como forma de materializar toda a cooperação e a sinergia entre todos os agentes e *stakeholders*, a Comissão criou um grupo de pesquisadores que foi estabelecido em junho de 2018 pela Comunicação 'Inteligência Artificial para a Europa', sendo chamado de Grupo de Peritos de Alto Nível em IA (GPAN IA)[218].

O GPAN IA tem a função de contribuir para a "definição de estratégias, políticas e prioridades da União Europeia em matéria de IA, além de aconselhar a Comissão Europeia em desafios de curto e longo prazo, bem como em oportunidades decorrentes da IA"[219]. A sua composição é formada por 52 membros que são escolhidos por meio de uma chamada aberta da Comissão Europeia, com o grupo final composto por 23 membros da indústria, 19 da academia e 10 da sociedade civil[220].

O GPAN IA desenvolveu dois trabalhos que norteiam essa pesquisa, sendo eles: 'Uma Definição de IA: principais capacidades e disciplinas científicas'[221] e as 'Orientações Éticas para uma IA de Confiança'[222].

2.2.5. Livro Branco sobre Inteligência Artificial - Uma abordagem europeia virada para a Excelência e a Confiança

De acordo com o glossário encontrado na EUR-Lex[223], um Livro Branco é um documento que contém propostas de ação da União Europeia em uma área específica, possuindo também o objetivo de lançar um debate entre todos os agentes e *stakeholders* a fim de chegar a um consenso político. Nesse sentido, um Livro Branco é um documento que apresenta o posicionamento da União e convida os participantes a apresentarem também as suas razões, atuando como *soft law* e operando, a posteriori, como um instrumento interpretativo para os atos normativos a adotar.

No caso em questão, a Comissão apresentou suas propostas para construir uma IA com uma abordagem europeia voltada para excelência e a confiança, visando o desenvolvimento fiável e seguro da IA na Europa. A fim de alcançar esse objetivo, o Livro Branco planeja construir dois ecossistemas diferentes, um baseado na excelência e outro baseado na confiança.

---

[218] Comissão Europeia. Grupo de Especialistas de Alto Nível em AI. Disponível: https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

[219] Stix. Charlotte. A survey of the European Union's artificial intelligence ecosystem... p. 13-14

[220] Ibid.

[221] Comissão Europeia. Grupo de Peritos de Alto Nível sobre IA. Uma Definição de IA: principais capacidades e disciplinas científicas'. 2018

[222] Comissão Europeia. Grupo de Peritos de Alto Nível sobre IA. Orientações Éticas para uma IA de Confiança. 2019. Disponível: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[223] EUR-Lex. Glossary of sumaries. *White Paper*. https://eur-lex.europa.eu/summary/glossary/white_paper.html

De acordo com o ecossistema de excelência, a Comissão planeja o desenvolvimento de um quadro político que visa alinhar os esforços a todos a nível europeu, nacional e regional para mobilizar recursos ao longo de toda a cadeia de valor, seja na investigação, na inovação ou na criação de incentivos para acelerar a adoção de soluções baseadas em IA. Visando sempre equilibrar o mercado, o foco dessa abordagem recai na promoção das pequenas e médias empresas (PME). Além disso, esse ecossistema é formado por oito propostas para diferentes áreas[224], pelo que pode-se citar: (i) obtenção de recursos para investimento e revisão do Plano Coordenado de 2018; (ii) criação de c e de teste que possam combinar investimentos europeus, nacionais e privados; (iii) desenvolvimento de competências para que se possua os melhores profissionais e os melhores mestrados a nível mundial no domínio da IA; (iv) criação de um centro de inovação digital com alto grau de especialização em IA por estado-membro; (v) promoção de parcerias com o sector privado; (vi) incentivo ao setor público a adotar sistemas de IA; (vii) criação de um verdadeiro espaço de dados europeu com garantia de acesso fácil e simplificado a dados e às infraestruturas de computação; e, por fim, (viii) cooperação com agentes internacionais a fim de influenciar debates a nível mundial. Como se pode perceber aqui, não há, necessariamente, nada de novo. Como se pode perceber, a União Europeia vem fortalecendo a abordagem política tripla para a IA, baseada em investimento, regulação e educação pública. Como mencionado, essas propostas são baseadas nos planos econômicos e políticos que a Comissão já apresentou em Declarações e Comunicações anteriores.

Por outro lado, há o ecossistema de confiança que consiste no desenvolvimento de um quadro regulamentar que pretende "garantir o respeito das regras da União, incluindo as regras de proteção dos direitos fundamentais e dos direitos dos consumidores"[225]. Dessa forma, com uma abordagem antropocêntrica e baseada nas 'Orientações Éticas para uma IA de Confiança'[226], o ecossistema de confiança direciona seus objetivos no sentido de: (a) fomentar o sentimento de confiança dos cidadãos para com as aplicações em IA; e (b) garantir segurança jurídica necessária para que as empresas e organizações públicas possam buscar maior inovação com base na IA, sem que a regulamentação represente alguma espécie de engessamento.

A fim de (a) fomentar a confiança dos cidadãos em aplicações de IA, a Comissão apresenta sete requisitos principais que sistemas baseados em IA deveriam apresentar[227], sendo eles (i) iniciativa e controlo por humanos; (ii) robustez e segurança; (iii) privacidade e governação dos dados; (iv) transparência; (v) diversidade, não discriminação e equidade; (vi) bem-estar societal e ambiental; e (vii) responsabilização. Esses requisitos possuem grande importância pois eles são desdobramentos dos

[224] Comissão Europeia. Livro Branco sobre a inteligência artificial... p.5-9
[225] Comissão Europeia. Livro Branco sobre a inteligência artificial... p.3
[226] GPAN IA. Orientações Éticas para uma IA de Confiança...
[227] Comissão Europeia. Livro Branco sobre a inteligência artificial... p.10

princípios éticos que irão compor uma IA de confiança. E nesse sentido serão melhor analisados na próxima seção.

Por outro lado, a Comissão também pretende promover (b) segurança jurídica necessária para incentivar a utilização e o desenvolvimento de sistemas de IA por empresas e organizações públicas. Para que isso seja possivelmente alcançado a Comissão planeja como alternativa um quadro regulatório específico para IA que seja baseado em uma abordagem de risco[228]. Através dessa abordagem, a Comissão deixa de lado a pretensão de tutelar IA como um todo, passando apenas a regular usos específicos que possam provocar risco elevado de dano aos interesses da sociedade e aos direitos fundamentais dos indivíduos.

Dessa forma, essa abordagem possui como elemento-chave a tutela jurídica de sistemas de IA que somente sejam considerados de alto risco e que, por isso, deverão respeitar uma legislação mais rigorosa. Observe-se que os sistemas de IA considerados de baixo risco, que não estarão sujeitos a esse regulamento, ainda assim deverão manter-se em conformidade com a demais legislação atualmente em vigor.

De acordo com o Livro Branco sobre IA, para ser considerado como um sistema de alto risco, tanto o 'setor' de utilização quanto o 'uso pretendido' do sistema de IA deverão estar inseridos em alguma situação de risco que possa ser considerada como tal. Destaca-se apenas que, enquanto a análise do 'setor' ocorre de forma mais objetiva, devendo já estar previamente apontado pelas autoridades como de alto risco, a análise do 'uso pretendido' é mais subjetiva, e depende de um risco significativo que será analisado pela perspetiva da segurança, dos direitos dos consumidores ou dos direitos fundamentais[229]. A própria Comissão também já prevê casos excecionais[230] em que apenas o uso pretendido já oferece um risco tão elevado que o requisito de cumulação dupla (risco e uso pretendido) deixa de ser necessário. Como exemplo é apontado o uso de tecnologias de vigilância intrusiva e de tecnologias de aspetos biométricos.

Dessa forma, é importante destacar que a abordagem baseada em risco, com a necessidade de presença de dois critérios cumulativos (ou identificação de uma das possíveis exceções), garante que o âmbito de aplicação do quadro regulamentar seja bem delimitado, proporcionando assim segurança jurídica para as empresas. Além disso, o estrito âmbito de aplicação incentiva o uso e desenvolvimento de IA pelas PME, uma vez que, cientes do baixo risco de aplicações que queiram utilizar, não sofrerão com demandas legislativas mais rigorosas. Por outro lado, a delimitação baseada em risco também permite que as autoridades e os agentes responsáveis efetuem um trabalho de

[228] Comissão Europeia. Livro Branco sobre a inteligência artificial... p.19
[229] Ibid.
[230] Comissão Europeia. Livro Branco sobre a inteligência artificial... p.20.

fiscalização mais específico e profundo em sistemas de alto risco, o que contribui para a proteção dos diretos fundamentais dos utilizadores.

Como se pode perceber da análise de todos os documentos que foram apresentados até esse momento, ficou clara a diretriz política que norteia o caminho da Comissão em seu desígnio de desenvolver uma IA para a Europa que seja ética, de confiança e centrada em valores humanos. A presente seção tinha como objetivo entender quais são os planos e os projetos que a Comissão possui em relação a IA. Objetivo este que deve ser considerado concluído. Dessa forma, esse trabalho acadêmico passa a analisar os aspetos jurídicos que compõem a mencionada IA de confiança.

2.3. Uma perspetiva introdutória a uma IA de Confiança

2.3.1. Uma definição de IA à luz das diretrizes da Comissão Europeia

A IA é uma tecnologia que está em constante evolução. Além disso, existem atualmente uma gama diversificada de aplicações baseadas em IA, podendo, por exemplo, servir de filtro de spam e, ao mesmo tempo, ser usada num sistema de armas autônomo. Assim sendo, a escolha por uma má definição pode ser simplesmente ineficaz ou prejudicial, já que pode não representar a realidade da IA ao longo do tempo, das suas aplicações e do seu processo de evolução[231]. Nesse sentido, a Comissão reconhece que "em qualquer novo instrumento jurídico, a definição de IA terá de ser suficientemente flexível para ter em conta os progressos técnicos, sendo suficientemente precisa para proporcionar a segurança jurídica necessária"[232].

Como se pode perceber, a Comissão busca uma definição que possa ser flexível e precisa ao mesmo tempo. Essa, de facto, é uma tarefa bastante complexa, mas que ao ser alcançada, promoverá a segurança jurídica necessária tanto para a proteção dos direitos dos utilizadores quanto dos interesses das empresas e dos desenvolvedores de IA, permitindo a adoção de uma legislação suficientemente adaptável e versátil no tempo e perante os incrementos tecnológicos e digitais expectáveis.

A primeira definição de IA apresentada foi pela Comunicação sobre IA para a Europa, onde afirmou-se que "o conceito de inteligência artificial (IA) aplica-se a sistemas que apresentam um comportamento inteligente, analisando o seu ambiente e tomando medidas — com um determinado nível de autonomia — para atingir objetivos específicos"[233].

---

[231] Ressalta-se aqui que esse trabalho não tentará resolver essa questão, até porque esse não é o objeto central de pesquisa. Isto posto, continuaremos a basear nossa pesquisa nos trabalhos que já foram publicados pela Comissão e pelo GPAN IA.

[232] Comissão Europeia. Livro Branco sobre a inteligência artificial... p.19.

[233] Comissão Europeia. Inteligência artificial para a Europa...

Contudo, essa definição é baseada no próprio conceito de inteligência, o que torna o processo de definição mais complexo e incerto. Nesse sentido, Matthew U. Scherer afirma que "a dificuldade em definir inteligência artificial não está no conceito de artificialidade, mas sim na ambiguidade conceptual da inteligência"[234]. Scherer faz essa afirmação pois, como os humanos são, teoricamente, os únicos seres vivos universalmente reconhecidas como possuidores de comportamento inteligente, as definições de inteligência quase sempre tendem a estar ligadas às características humanas[235]. Assim, torna-se particularmente complexo tentar definir algo como inteligente, principalmente uma máquina ou uma tecnologia, quando temos de compará-la com comportamentos humanos[236]. Ocorre que essa foi a primeira definição apresentada pela Comissão, e embora haja críticas a seu respeito, ela foi aperfeiçoada pelo GPAN IA.

De acordo com 'Uma definição de IA', o GPAN IA apresentou a sua própria definição, pelo que afirmava que: "os sistemas de inteligência artificial (IA) são sistemas de software (e eventualmente também de hardware) concebidos por seres humanos[237], que, tendo recebido um objetivo complexo, atuam na dimensão física ou digital percecionando o seu ambiente mediante a aquisição de dados, interpretando os dados estruturados ou não estruturados recolhidos, raciocinando sobre o conhecimento ou processando as informações resultantes desses dados e decidindo as melhores ações a adotar para atingir o objetivo estabelecido. Os sistemas de IA podem utilizar regras simbólicas ou aprender um modelo numérico, bem como adaptar o seu comportamento mediante uma análise do modo como o ambiente foi afetado pelas suas ações anteriores"[238].

Além de uma definição direta acerca dos atos que envolvem a IA, o GPAN IA fez questão de nominar as técnicas e as abordagens que possivelmente podem estar atreladas à IA. Nesse sentido, o GPAN IA menciona "a aprendizagem automática (de que a aprendizagem profunda e a aprendizagem por reforço são exemplos específicos), o raciocínio automático (que inclui o planeamento, a programação, a representação do conhecimento e o raciocínio, a pesquisa e a otimização) e a robótica (que inclui o controlo, a perceção, os sensores e atuadores, bem como a integração de todas as outras técnicas em sistemas ciberfísicos)"[239].

A definição de IA realizada pelo GPAN IA tenta abordar os atos que permeiam as ações que a IA possivelmente pode realizar. De facto, levando em consideração que a evolução da IA ainda seja baseada nesse modelo de operação, a definição do GPAN IA definitivamente alcançou o conceito de

[234] Scherer, Matthew U. Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, And Strategies. Harvard Journal of Law & Technology Volume 29, Number 2 Spring 2016. p 360.
[235] Ibid.
[236] Os problemas envolvidos na conceituação ou definição do que é inteligência foram abordados no primeiro capítulo, especificadamente na seção 1.2.2.
[237] Os seres humanos concebem os sistemas de IA diretamente, mas também podem utilizar técnicas de IA para otimizar a sua conceção (Nota realizada pelo próprio GPAN IA).
[238] GPAN IA. Uma definição de IA... p. 8.
[239] Ibid.

flexibilidade requerido pela Comissão. Afirma-se isso pois essa definição possui competência para se manter eficaz ao longo do tempo e perante a evolução da tecnologia. Além disso, essa definição aparentemente cumpre com a característica de precisão, uma vez que é capaz de identificar e enquadrar simples, complexas ou distintas aplicações baseadas em IA. Isto posto, é possível afirmar que tal definição pode promover a segurança jurídica necessária para o setor da IA.

Contudo, há ainda o sentimento de que essa não será a última definição apresentada pela Comissão sobre a IA. Pelo valor econômico que a IA tem agregado ao Mercado Único Digital, assim como pela atenção que a IA tem gerado na sociedade e na academia, é possível que a busca por uma definição sobre a IA ainda seja profundamente debatida. No entanto, tal definição promovida pelo GPAN IA é compreensivelmente útil e bem-vinda ao presente trabalho acadêmico, pelo que será tomada como a definição de IA para efeitos de investigação.

2.3.2. Por que precisamos de uma IA de confiança?

Apesar do potencial referente aos benefícios substanciais para os indivíduos e a sociedade, algumas aplicações de IA também são suscetíveis de provocar danos que podem ser difíceis de prever, identificar ou medir[240]. Nesse sentindo, a IA, como qualquer outra tecnologia, não somente traz inúmeras oportunidades, mas também suscita uma variedade de desafios éticos, legais e sociais[241].

Inúmeros projetos, mas com diferentes nomenclaturas, vêm sendo desenvolvidos para dar uma resposta ética e legal a essas provocações. Como exemplo, pode-se citar (i) a 'IA Benéfica' desenvolvida pelo *Future of Life Institute*[242]; (ii) a 'IA Responsável', utilizada pela Universidade de Montreal[243] e pela Governação Nacional Chinesa[244]; e as duas iniciativas sobre a 'IA Ética', sendo (iii) uma liderada pela UK House of Lords[245], e (iv) a outra por Luciano Floridi[246].

Há, ainda, a iniciativa que é objeto desse estudo, a 'IA de confiança' que, como mencionado, foi elaborada a partir das 'Orientações Éticas para uma IA de Confiança' publicada pelo GPAN IA. As Orientações criaram um quadro ético para a IA que tem desempenhado um papel importante no

[240] GPAN IA. Orientações Éticas para uma IA de Confiança... p. 16.
[241] Floridi, Luciano. Establishing the rules for building trustworthy AI. Nature Machine Intelligence, 2019, 1(6), 261–262. https://doi.org/10.1038/s42256-019-0055-y
[242] Future of Life Institute. Asilomar AI Princples. 2017. Disponível: https://futureoflife.org/ai-principles/
[243] Université de Montréal. Montreal Declaration for a Responsible Developmen t of AI. 2017. Disponível: https://www.montrealdeclaration-responsibleai.com/the-declaration.
[244] Chinese National Governance Committee for the New Generation Artificial Intelligence. Governance Principles for the New Generation Artificial Intelligence–Developing Responsible Artificial Intelligence. 2019. Disponível: https://www.chinadaily.com.cn/a/201906/17/WS5d07486ba3103dbf14328ab7.html.
[245] Université de Montréal. Montreal Declaration for a Responsible Developmen t of AI. 2017. Disponível: https://www.montrealdeclaration-responsibleai.com/the-declaration.
[246] Floridi, L., Cowls, J., Beltrametti, M., et al. AI4People: An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 2018, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

cenário internacional, a ponto de influenciarem os trabalhos da OECD[247] e do Gabinete de Política Científica e Tecnológica da Casa Branca[248], que adotaram a mesma terminologia.

No entanto, assim como afirma Scott Tietes, independentemente da terminologia adotada, todas essas iniciativas possuem em si o mesmo objetivo, que é o de fazer com que o desenvolvimento e a aplicação da IA sejam realizados de forma que seus benefícios sejam maximizados, enquanto seus riscos e perigos sejam mitigados ou prevenidos[249]. Nesse sentido, e de acordo com o GPAN IA, a busca por uma IA de confiança encontra respaldo na ideia de que os indivíduos, empresas e sociedades só serão capazes de alcançar todo o potencial da IA se "esta tecnologia, incluindo os processos e as pessoas que lhe estão subjacentes, for digna de confiança"[250].

Inúmeras razões levaram ao GPAN IA a escolher a 'confiança' como o sentimento a ser perseguido. O GPAN IA se baseia na premissa de que a fiabilidade é a "condição prévia essencial para as pessoas e sociedades desenvolverem, implantarem e utilizarem os sistemas de IA"[251]. Assim sendo, a incapacidade de demonstrar que os sistemas de IA – e as pessoas envolvidas – são confiáveis, prejudicará e limitará a aceitação desses sistemas. Além disso, o GPAN IA também acredita que uma abordagem baseada na confiança poderá incentivar e promover uma competitividade responsável, aliada a uma inovação responsável e sustentável no domínio da IA[252].

A partir da análise da estratégia europeia para a IA, pôde-se perceber que a Comissão está determinada em colocar os cidadãos no centro dos seus esforços, sendo esse o caminho que a Comissão acredita que colocará a Europa em uma posição de liderança em matéria de sistemas de IA. Além disso, é um desígnio da Comissão criar uma cultura de IA de confiança para a Europa, "mediante a qual os benefícios da IA possam ser usufruídos por todos de uma forma que garanta o respeito aos valores fundamentais: os direitos fundamentais, a democracia e o Estado de direito"[253]

A IA de confiança não se limita apenas a ser uma resposta para os riscos provocados pela IA, uma vez que permite representar uma característica reputacional de uma IA europeia antropocêntrica. Dessa forma, a Comissão consegue promover a proteção dos interesses da sua sociedade e dos direitos dos seus residentes, como também consegue agregar um valor reputacional à sua indústria, o que possibilita promover e ampliar o poder econômico europeu.

---

[247] OECD. Principles on AI. 2019. Disponível: https://www.oecd.org/going-digital/ai/principles/.
[248] Vought, Russel T. Guidance for Regulation of Artificial Intelligence Applications. Executive Office of the President of United States of America. 2020, p. 4-5 Disponível: https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf
[249] Thiebes, S., Lins, S. & Sunyaev, A. Trustworthy artificial intelligence. Electron Markets (2020). https://doi.org/10.1007/s12525-020-00441-4, p.2.
[250] GPAN IA. Orientações Éticas para uma IA de Confiança... p. 45.
[251] GPAN IA. Orientações Éticas para uma IA de Confiança... p.6.
[252] Ibid.
[253] GPAN IA. Orientações Éticas para uma IA de Confiança... p.45-46.

2.3.3. Uma IA de confiança e as suas componentes «Legal», «Ética» e «Sólida»

Em 8 de abril de 2019, o GPAN IA publicou suas 'Orientações éticas para IA de Confiança' (Orientações). As Orientações éticas consistem em um documento que sistematiza um quadro para alcançar uma IA de confiança baseada nos direitos fundamentais consagrados na Carta dos Direitos Fundamentais da União Europeia e nos direitos humanos[254]. Apesar desse quadro não possuir qualquer força vinculativa, as partes interessadas podem voluntariamente seguir a sua orientação. Essa voluntariedade representa uma forma de operacionalizar o compromisso de alcançar IA confiável.

Aliás, é importante destacar que as Orientações afirmam claramente que não refletem uma posição oficial da Comissão e não possuem como objetivo substituir qualquer forma de formulação de políticas ou regulamentação atual ou futura. Em vez disso, suas diretrizes têm como objetivo oferecer uma orientação para que uma aplicação de IA possa ser desenvolvida no sentido de representar uma IA de confiança. Outra característica importante remete-se ao fato de que as Orientações pretendem ser um documento dinâmico, que será revisto e atualizado com o tempo e que servirá como um ponto de partida para a investigação e debate a nível mundial sobre um quadro ético para sistemas de IA[255].

Nesse sentido, as Orientações compõem um documento estruturado propositalmente em 03 (três) capítulos. O Capítulo I apresenta as bases de uma IA de confiança através do estabelecimento de direitos fundamentais e princípios éticos. O Capítulo II apresenta um meio de materialização de uma IA de confiança através da presença de sete requisitos essenciais que todo o sistema de IA deve possuir. Tais requisitos são originados através dos princípios estabelecidos anteriormente. O Capítulo III propõe a operacionalização de uma IA de confiança através dos requisitos anteriormente mencionados. Para esse fim, é apresentada uma 'lista de autoavaliação' para ajudar a operacionalizar uma IA de confiança. Essa lista fornece perguntas práticas, mas não exaustivas, para as partes interessadas realizarem uma inquirição durante o processo de *design* de desenvolvimento do sistema de IA.

É importante destacar que, de acordo com o próprio GPAN IA, as Orientações pretendem ser mais do que uma simples lista de princípios éticos. Nesse sentido, a estrutura e o desenvolvimento desse quadro ocorre de forma a facultar aos interessados o acesso às orientações éticas para uma IA de confiança através de três níveis de abstração. O Capítulo I apresenta o maior nível de abstração, pois é baseado no estudo de direitos fundamentais e princípios éticos. O Capítulo II, por sua vez, torna-se menos abstrato na medida em que são apresentados requisitos práticos que uma IA deve possuir para ser considerada em conformidade com os direitos e princípios previamente estabelecidos. Por

[254] GPAN IA. Orientações Éticas para uma IA de Confiança... p.7.
[255] GPAN IA. Orientações Éticas para uma IA de Confiança... p.3-4.

fim, o Capítulo III possui o menor nível de abstração, na medida em que apresenta uma lista de avaliação para efetivamente operacionalizar tais princípios e requisitos em sistemas sociotécnicos[256].

Assim sendo, por já ter ciência do que é o documento em questão, da metodologia adotada em sua construção e da forma como ele deve ser utilizado, torna-se importante conhecer o que torna um sistema inteligente em uma IA de confiança.

Para que uma IA seja de confiança, ela deve ser desenvolvida a partir de três componentes[257]. Em primeiro lugar, ela deverá ser «Legal», o que implica no respeito a toda legislação e regulamentação aplicável. Ela também deverá ser «Ética», pelo que deve garantir respeito a observância de princípios e valores éticos. Por fim, a IA de confiança deverá ser «Sólida», o que significa que ela deverá garantir segurança e robustez pelos pontos de vista técnico e social.

É importante destacar que, de acordo com o GPAN IA, a fiabilidade de uma IA de confiança diz respeito tanto ao próprio sistema de IA quanto a todos os processos que fazem parte do seu ciclo de vida[258]. Note que esse pressuposto remete diretamente a abordagem promovida pelo Plano Coordenado de IA, que alertava sobre a necessidade de sistemas de IA éticos e seguros por padrão[259], ou seja, segundo o qual princípios éticos e legais seriam, necessariamente, implementados desde o início do desenvolvimento de um sistema de IA.

Apesar das componentes «Legal,», «Ética» e «Sólida» serem necessárias para uma IA de confiança, somente elas, por si, não são suficientes para alcançar esse fim[260]. Isso porque situações e aplicações diferentes suscitam novos desafios e condições específicas que deverão ser necessariamente observadas. Através desse raciocínio, pode-se perceber que um algoritmo que reconhece um perfil e indica um filme ou uma música não provoca as mesmas preocupações que um carro autônomo ou um sistema que identifica doenças e propõe certos tratamentos médicos.

A introdução da característica «Legal» como uma componente da IA de confiança é uma das principais adições que ocorreram quando da publicação da versão final das Orientações em comparação com o seu rascunho predecessor. O que, de certa forma, iremos perceber que é uma importante modificação, mas que está longe de ter a profundidade que era esperada.

Evidentemente as Orientações reconhecem que os sistemas de IA não podem operar através de um vácuo jurídico, ou seja, às margens da lei. Nesse sentido, conclui-se que os sistemas de IA, assim como toda e qualquer tecnologia, deve respeitar os diversos diplomas jurídicos que possuem força vinculativa, sejam eles a nível nacional, europeu ou internacional.

---

[256] GPAN IA. Orientações Éticas para uma IA de Confiança... p.2.
[257] GPAN IA. Orientações Éticas para uma IA de Confiança... p.6.
[258] Ibid.
[259] Comissão Europeia. Plano Coordenado para a Inteligência Artificial...
[260] GPAN IA. Orientações Éticas para uma IA de Confiança... p.6.

Através de uma perspetiva de interação entre Direito e IA, diversas são as fontes jurídicas relevante, pelo que se pode citar[261] as de direito primário da União, como os Tratados da União Europeia e a sua Carta dos Direitos Fundamentais; assim como os de direito derivado da União, como o Regulamento Geral sobre a Proteção de Dados, o Regulamento de Livre Fluxo de Dados Não Pessoais, as diretivas relativas à não discriminação e as diretivas nos domínios da defesa do consumidor. Observe-se apenas que essa é uma lista apenas exemplificativa, e os diplomas jurídicos relevantes não se limitam a esses.

No entanto, as Orientações não discutem em mais detalhes as obrigações legais que se aplicam a um sistema de IA de confiança, limitando a exposição da componente «Legal» somente essas explicações.

Nesse sentido, embora se detete tal lacuna nas Orientações, a realidade é que o GPAN IA não teve a pretensão de analisar a fundo as implicações e as obrigações legais referentes a uma IA de confiança, gerando uma oportunidade de que isso venha a ser levado a cabo no âmbito das dinâmicas legislativas próprias da União, explicáveis à luz do princípio do equilíbrio institucional.

Independentemente dessa opção feita pelo GPAN IA, para se alcançar uma IA de confiança é necessário que um sistema de IA respeito as três componentes. Nesse sentido, comprometer-se apenas com a componente «Legal» não faz um sistema inteligente ser reconhecido como de confiança. Isso porque existem ações que respeitam estritamente a ordem jurídica, mas podem ser consideradas fundamentalmente antiéticas. Nesse sentido, não se pode afirmar que, ao se alcançar uma IA Legal, esta será automaticamente uma IA Ética.

Por mais que a ética possa ser tida como uma influência jurídica, "a legislação nem sempre acompanha a rapidez da evolução tecnológica e, por vezes, pode estar desfasada das normas éticas ou não se adequar, pura e simplesmente, ao tratamento de certas questões"[262]. Assim sendo, para que os sistemas baseados em IA possam ser considerados de confiança, necessariamente deverão ir além da mera legalidade e estarem também em harmonia com as normas éticas.

As Orientações delimitaram alguns princípios éticos que devem ser seguidos no que concerne o desenvolvimento de uma IA de confiança, sendo eles: respeito pela autonomia humana; prevenção de dados; equidade e explicabilidade.

Por fim, há também a necessidade de que a IA seja «Sólida». Nesse contexto, os desenvolvedores de um sistema inteligente devem garantir que as suas aplicações não venham a causar danos não intencionais à sociedade ou a seus utilizadores. Para que isso aconteça, uma IA de confiança deve funcionar de forma segura e fiável, sendo, portanto, baseada na segurança por padrão,

[261] GPAN IA. Orientações Éticas para uma IA de Confiança... p.7.
[262] GPAN IA. Orientações Éticas para uma IA de Confiança... p.8.

onde critérios sólidos de segurança são aplicados em todos os ciclos de vida da aplicação[263]. Além disso, é importante que já sejam previstos salvaguardas e planos de resposta a fim de que se evitem ou mitiguem impactos negativos não intencionais.

2.4. Um quadro para a IA de confiança: princípios, requisitos e uma lista de avaliação.

As Orientações publicadas pelo GPAN IA talvez componham um dos documentos mais importantes para o debate de regulação de IA. Isso porque, ao mesmo tempo em nos apresentam a ideia das componentes «Legal», «Ética» e «Sólida», também sintetizam um quadro técnico, ético e jurídico para alcançar uma IA de confiança.

Esse quadro foi necessariamente desenvolvido a partir dos direitos fundamentais consagrados na Carta dos Direitos Fundamentais da União Europeia. Nesse sentido, por mais que o quadro, em si, não seja vinculativo, os direitos que estão por trás dele o são. Assim sendo, não se pode obrigar nenhuma pessoa, empresa ou organização a seguir as orientações apresentadas por esse quadro. No entanto, é interessante que se perceba que tais orientações podem ser entendidas como uma forma de operacionalizar os direitos fundamentais que foram o núcleo central do seu desenvolvimento e poderão atuar como referencial interpretativo inerente à proteção de direitos fundamentais visada na União e pautada pelo princípio do nível mais elevado de proteção (nos termos dos artigos 52.º, n.ºs 3 e 4 e 53.º da CDFUE).

2.4.1. Os princípios éticos de uma IA de confiança: uma abordagem baseada no padrão jusfundamental da União Europeia

Através das Orientações éticas, a Comissão apresentou quatro princípios para uma IA de confiança, sendo eles: o respeito pela autonomia humana; a prevenção de danos; a equidade; e a explicabilidade[264]. Ocorre que esse quadro ético foi desenvolvido a partir de direitos fundamentais consagrados nos Tratados da União Europeia (TUE) e na Carta dos Direitos Fundamentais da União Europeia (CDFUE).

Destaca-se que a "CDFUE oferece um catálogo de direitos fundamentais a todos os que se submetem à jurisdição da União"[265]. E que, nesse sentido, o artigo 51.º da CDFUE determina que esses direitos "têm como destinatários as instituições, órgãos e organismos da União, na observação do

[263] Ibid.
[264] GPAN IA. Orientações Éticas para uma IA de Confiança... p. 14.
[265] Silveira, Alessandra. "Comentário ao art. 51.º", in Carta dos Direitos Fundamentais da União Europeia Comentada, Alessandra Silveira/Mariana Canotilho (coords.), Almedina, Coimbra, 2013, p.574.

princípio da subsidiariedade, bem como os Estados-Membros, apenas quando apliquem o direito da União"[266].

Uma primeira observação recai no fato de todos os agentes da União estão sujeitos às disposições da Carta[267], e não somente àqueles mencionados no artigo 13.º do TUE. Por outro lado, no que se refere aos Estado-Membros, sua interpretação deve ser feita em sentido amplo, vinculando todos os organismos ou entidades que estejam sujeitos à autoridade ou ao controlo do Estado, justamente para evitar que o Estado-Membro tire proveito da inobservância do direito da União[268].

Perceba, no entanto, que os particulares não figuram como destinatários dessa obrigação, mas sim como beneficiários, na medida em que os particulares poderão invocar o padrão jusfundamental da União sempre quando a "medida impugnada (europeia ou nacional) integre o âmbito de aplicação material do direito da União"[269], ou, em outras palavras, diga respeito às competências da União. Dessa forma, "se a União exerceu a sua competência através de um ato jurídico previsto no artigo 288.º do TFUE – seja através de um regulamento, uma diretiva ou uma decisão – esse será o *link* a partir do qual a proteção jusfundamental da União poderá ser invocada"[270]. O RGPD, fundamentado no artigo 8.º da CDFUE e no artigo 16.º, n.º 1 do TFUE, é o possível *link* quando particulares querem invocar o padrão jusfundamental da União em assuntos relacionados à proteção de dados pessoais.

Contudo, haverá casos em que não será possível identificar um *link* direto que permita invocar o padrão jusfundamental da União. Nesse sentido, e na ausência de um *link* claro e evidente com o direito da União, alguns doutrinadores, incluindo Alessandra Silveira, defendem que o particular pode invocar a própria cidadania europeia, presente no artigo 20.º do TFUE para que preencha esse requisito[271]. Essa posição também encontra respaldo na jurisprudência recente do TJUE através do caso *Zambrano*.

Dessa forma, ao partir do pressuposto que o quadro ético de uma IA de confiança fundamenta-se no padrão jusfundamental da União Europeia, pode-se concluir que essência protetiva dos princípios éticos do respeito pela autonomia humana, da prevenção de danos, da equidade e da explicabilidade, de forma indireta e consequente aos direitos fundamentais que representam, poderão ser invocadas para a proteção dos indivíduos e utilizadores e sistemas de IA que possam vir a causar algum dano, observadas as regras do artigo 51.º, 1, do CDFUE.

---

[266] Artigo 51.º, 1, da Carta

[267] Silveira, Alessandra. "Comentário ao art. 51.º", in Carta dos Direitos Fundamentais da União Europeia Comentada, p. 579.

[268] Silveira, Alessandra. "Comentário ao art. 51.º", in Carta dos Direitos Fundamentais da União Europeia Comentada, p. 580.

[269] Silveira, Alessandra, Canotilho, Mariana e Froufe, Pedro Madeira. Direito da União Europeia – Elementos de Direito e Políticas da União. Almedina, 2016, p. 51-52.

[270] Silveira, Alessandra. "Comentário ao art. 51.º", in Carta dos Direitos Fundamentais da União Europeia Comentada, p. 577.

[271] Silveira, Alessandra, Canotilho, Mariana e Froufe, Pedro Madeira. Direito da União... p. 53.

2.4.1.1. Uma releitura dos direitos fundamentais no contexto da Inteligência Artificial

Relativo aos direitos fundamentais que podem estar relacionados com os impactos da IA na sociedade, o GPAN IA apontou, principalmente, os encontrados no Artigo 2.º do TUE e os Artigos 1.º e 3.º da CDFUE[272]. Nesse sentido, pode-se afirmar que a Comissão tem o desígnio de que o desenvolvimento de uma IA de confiança seja baseado em "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"[273]. Além disso, uma IA de confiança deve assegurar a todas as pessoas o "respeito pela sua integridade física e mental"[274].

Nesse contexto, há algo interessante a se notar. Isso porque Marcelo Rebelo afirma que esses "princípios correspondem aos elementos definidores da identidade europeia"[275]. Essa é uma afirmação importante, uma vez que reforça aquilo que a Comissão afirma sobre desenvolver uma IA antropocêntrica[276] e baseada em valores europeus[277]. Esse posicionamento demonstra, cada vez mais, que a Comissão não pretende ser líder mundial em uma IA a qualquer custo. Aliás, em hipótese alguma, os direitos fundamentais e os valores europeus irão ser limitados ou mitigados em prol de uma inovação, não havendo, portanto, espaço na União para "inovar primeiro [e] consertar mais tarde"[278].

No que diz respeito aos direitos fundamentais encontrados nos Artigo 2.º da TUE, Rebelo ainda afirma que o princípio da não discriminação é inseparável da igualdade, e que a justiça é corolário do respeito pela dignidade humana. Sendo que esses princípios são "inimagináveis fora de um contexto comunitário"[279], uma vez que compõem a visão social do Estado de Direito europeu. Além do mais, Rebelo ratifica aquilo que já deve ser claro, uma vez que declara que "o âmago axiológico reside na dignidade humana"[280], pois é a partir dela que se erguem os princípios da liberdade, da democracia, da igualdade, e, por conseguinte, o próprio Estado de Direito, elementos estes que são considerados indissociáveis[281].

Assim sendo, é importante destacar que cada um desses princípios possuem um devido sentido e aplicação que lhe são próprios, pois decorrem principalmente de um desenvolvimento histórico e cultural. No entanto, assim como em significantes acontecimentos que ocorreram ao longo

[272] O Artigo 8.º da Carta referente ao direito à proteção de dados será objeto de estudo no próximo capítulo.
[273] Artigo 2.º do TUE.
[274] Artigo 3.º da Carta.
[275] Anastácio, Manuel Lopes Porto Gonçalo. Tratado de Lisboa - Anotado e Comentado. Almedina. Edição do Kindle, p. 335.
[276] Joana Covelo de Abreu afirma que devemos partir de um fundamento antropológico para buscar respostas para os possíveis riscos das tecnologias de informação na justiça eletrónica europeia. Essa afirmação pode ser comparada com o ideal da Comissão Europeia em criar uma IA centrada no ser humano. Joana Covelo de Abreu et al. O Contencioso da União Europeia e a cobrança transfronteiriça de créditos: compreendendo as soluções digitais à luz do paradigma da Justiça eletrónica europeia (e-Justice). UNIO, 2020, p.4. DOI: 10.21814/1822.65807
[277] Comissão Europeia. Livro Branco sobre a inteligência artificial... p.25.
[278] Floridi, Luciano. Establishing the rules for building trustworthy AI... p.1-2.
[279] Anastácio, Manuel Lopes Porto Gonçalo. Tratado de Lisboa... Edição do Kindle, p. 335.
[280] Ibid.
[281] Anastácio, Manuel Lopes Porto Gonçalo. Tratado de Lisboa... Edição do Kindle, p. 335.

da história, a IA também vem demonstrando capacidade de modificar e adicionar novas camadas às relações sociais já existentes. Nesse sentido, torna-se importante a realização de uma releitura desses princípios de acordo com os impactos da IA, a fim de construir uma contextualização, mas também de entender os motivos pelo qual levaram o GPAN IA a basearem a sua IA de confiança nesses direitos fundamentais.

Assim sendo, pela importância que possui em possibilitar o fortalecimento dos demais direitos, não há melhor caminho a se começar do pelo próprio respeito pela dignidade humana. Nesse sentido, à luz da filosofia de Kant, "no reino dos fins tudo tem ou um preço ou uma dignidade". Quando uma coisa tem um preço, pode-se pôr em vez dela qualquer outra como equivalente; mas quando uma coisa está acima de todo o preço, e, portanto, não possui equivalente, então ela tem dignidade"[282].

O ser humano, por ser dotado de racionalidade, possui dignidade. Nesse sentido, por mais que usualmente a palavra dignidade possa se referir aos sentimentos morais relacionados ao respeito e a honestidade, essa não é a atribuição dada por Kant. Através de uma perspetiva kantiana, a dignidade atribuída ao ser humano deve ser entendida como um valor essencial que este carrega em si, e que por isso ele possui o direito a não ser tratado como algo que possuí um preço ou que possa ser usado como meio[283]. Por isso, o ser humano, possuidor da dignidade humana, deve sempre ser tratado como o fim em si mesmo.

Assim sendo, ao trazer esse debate para o contexto atual, percebe-se que nenhum ser humano pode ter o seu valor e a sua dignidade subjugada, limitada ou relativizada, seja pelo Estado, por outras pessoas ou pela inovação relativa a uma nova tecnologia, o que inclui a IA[284]. Nesse sentido, o artigo 2.º do TUE e o artigo 1.º da Carta determinam que o respeito à dignidade humana implica que todos sejam tratados como os sujeitos de direito que o são, e jamais como meros objetos suscetíveis de serem vigiados, examinados, triados, perfilados, classificados, segmentados, condicionados ou manipulados[285].

Por isso, através de uma contextualização entre dignidade humana e IA, percebe-se a importância do desenvolvimento de um quadro ético e jurídico para uma IA de confiança. Afirma-se isso pois os sistemas de IA não podem, indiscriminadamente, serem desenvolvidos e testados diretamente com indivíduos ou no meio ambiente[286]. Ciente dos impactos e da imprevisibilidade dos danos, e devendo agir com respeito à dignidade humana, torna-se necessário que o desenvolvimento da IA seja deliberado através de um consenso sociopolítico, em vista de uma estratégia de longo prazo sobre que tipo de IA deve ser desenvolvida, qual a finalidade e quais são os padrões éticos e jurídicos

[282] Kant, Immanuel. A fundamentação da Metafísica dos Costumes. Lisboa, Portugal: Edições 70, 2011.
[283] Pagno, Luana. A Dignidade Humana em Kant, p.225.
[284] C. McCrudden, «Human Dignity and Judicial Interpretation of Human Rights», EJIL, 19(4), 2008
[285] GPAN IA. Orientações Éticas para uma IA de Confiança... p.13.
[286] Floridi. Establishing the rules for building trustworthy AI. p.1.

esperados[287]. Esse é o caminho que, por sinal, a Comissão Europeia vem traçando através das suas diretrizes, orientações e comunicações.

Quanto aos demais direitos fundamentais, colaciona-se novamente aquilo que foi afirmado por Rebelo, em que "o âmago axiológico reside na dignidade humana"[288], sendo somente a partir dela que se é possível discutir a liberdade do indivíduo; o respeito pela democracia, pela justiça e pelo Estado de direito; e a igualdade e a não discriminação.

No que é relativo à liberdade do indivíduo, todo e qualquer ser humano, independentemente de cor, gênero ou credo, deve possuir a autonomia individual necessária a fim de ser livre para tomar as decisões relativas à sua vida privada. Nesse sentido, e em um contexto em que a IA participa diretamente da nossa vida privada através de motores de busca, redes sociais[289] e *marketplace* em linha, "a liberdade do indivíduo exige a atenuação da coerção ilegítima (in)direta, da correção à autonomia mental e à saúde mental, da vigilância injustificada, do engano e da manipulação indevida"[290]. Assim sendo, o direito fundamental à liberdade do indivíduo desse ser um desígnio nos sistemas de IA para que sejam possíveis a proteção e a promoção de diversos outros direitos, como a proteção do direito de empresa, a liberdade das artes e das ciências, a liberdade de expressão, o direito ao respeito pela vida privada e familiar e a liberdade de reunião e de associação[291].

Relacionado aos direitos fundamentais, também se é preciso o respeito pela democracia, pela justiça e pelo Estado de direito, ou União de Direito[292], caso se refira à União Europeia. Nesse contexto, os sistemas baseados em IA necessitam manter e favorecer os processos democráticos, além de respeitar a pluralidade de valores e escolhas de vida dos indivíduos[293]. Como se pode perceber, trata-se de uma responsabilidade que coaduna com aquela desejada pelo direito à liberdade do indivíduo. Isso porque, ao longo da última década, temos observado diversas tentativas (bem-sucedidas) de condicionamento de opinião no intuito de obstruir, mitigar e até manipular processos democráticos[294]. Tal manipulação afronta não somente a democracia, o Estado de Direito, e a liberdade do indivíduo, como também desrespeita a dignidade humana. Afirma-se isso pois grupos e organizações se valem dos dados pessoais e técnicas de perfilamento da IA para construírem um perfil individual passível de demonstrar fraquezas e vulnerabilidades sociais e psicológicas. Pelo que usam esse conhecimento contra o próprio indivíduo, desrespeitando a sua dignidade e mitigando o seu respetivo valor essencial,

[287] Ibid.
[288] Anastácio, Manuel Lopes Porto Gonçalo. Tratado de Lisboa... Edição do Kindle, p. 335.
[289] The Social Dilemma. Documentário. Netflix, 2020
[290] GPAN IA. Orientações Éticas para uma IA de Confiança... p.13.
[291] GPAN IA. Orientações Éticas para uma IA de Confiança... p.13.
[292] Alessandra Silveira. União de direito e ordem jurídica da União Europeia. Revista Eletrônica Direito e Política, Programa de Pós-graduação Stricto Sensu em Ciência Jurídica da UNIVALI, Itajaí, v.3, n.3 3º quadrimestre de 2008. ISSN 1980-7791, p. 5.
[293] GPAN IA. Orientações Éticas para uma IA de Confiança... p.13.
[294] The Great Hack. Documentário. Netflix. 2019

a fim de alcançar um objetivo obscuro, que por vezes é traduzido em eleger ou enfraquecer uma determinada corrente política.

Por fim, também se faz prudente uma rápida análise do direito à igualdade e a não discriminação. Nesse sentido, como visto no capítulo primeiro, os dados são considerados um retrato da sociedade[295]. Assim sendo, uma sociedade[296] ou uma organização[297] que possua discriminações estruturais, irá consequentemente gerar dados particularmente enviesados. Ocorre que o uso e o desenvolvimento de sistemas de IA estão diretamente condicionados ao tratamento de dados, muitos dos quais de natureza pessoal. Isto posto, dentro de um contexto em que sistemas de IA podem reproduzir ou até intensificar tratamentos discriminatórios, os direitos a igualdade e a não discriminação ganham uma importância ainda maior, pois tornam-se os verdadeiros escudos protetores de minorias e grupos vulneráveis. Aliás, acerca desses direitos, Rebelo firma que o direito fundamental "a não discriminação é inseparável da igualdade, e esta recobre a relativa a mulheres e a homens"[298]. Nesse sentido, não só entre questões de gênero, mas também de etnia, orientação sexual, religião, idade, condição social e capacidade física. Isto posto, os direitos a igualdade e a não discriminação, quando relativos à IA, devem, primeiramente, assegurar "o respeito igualitário do valor moral e da dignidade de todos os seres humanos [...], onde a igualdade implica que as operações do sistema de IA não podem gerar resultados injustamente tendenciosos"[299]. Para que isso seja possível, o primeiro passo é certificar-se da qualidade dos dados que treinarão o sistema de IA, garantindo assim que tais dados sejam inclusivos e representem os diferentes grupos em questão.

Como se pode perceber, a presença cada vez maior da IA na sociedade tem modificado e adicionado novas e complexas camadas às relações sociais que já eram conhecidas. Por mais que cada uma dessas camadas possa trazer novos desafios, é harmônico afirmar que, por ora, os direitos fundamentais continuam a projetar suas proteções em toda a sociedade. De facto, por mais que exista um avanço tecnológico antes inimaginável, uma releitura contextualizada dos direitos fundamentais é plenamente capaz de mostrar o caminho a ser seguido. Como se era de se esperar, tanto a Comissão quanto o GPAN IA baseiam seus passos nesses direitos, sendo, portanto, nesse contexto que o quatro para uma IA de confiança nos apresenta os seus quatro princípios éticos.

[295] Cortiz, Diogo. O Design pode ajudar na construção de Inteligência Artificial humanística?...
[296] COMPAS
[297] IA Amazon
[298] Anastácio, Manuel Lopes Porto Gonçalo. Tratado de Lisboa, Kindle, p.349.
[299] GPAN IA. Orientações Éticas para uma IA de Confiança... p.13.

2.4.1.2. Os princípios éticos para uma IA de confiança

Os quatro princípios éticos de uma IA de confiança, também conhecidos como 'imperativos éticos'[300], baseiam-se nos direitos fundamentais encontrados no Artigo 2.º do TUE e são: Respeito pela autonomia humana; Prevenção de danos; Equidade; e Explicabilidade.

Destaca-se que não existe qualquer grau de hierarquia entre eles. Aliás, de acordo com situações diversas e casos específicos, é possível que possa ocorrer algum tipo de conflito acerca de suas aplicações práticas, pelo que não há uma solução rígida a seguir[301].

Por mais que sejam baseados em direitos fundamentais, tais princípios não possuem força vinculante. No entanto, esses princípios podem vir a inspirar novos instrumentos regulamentares. Além disso, eles também podem adicionar novas camadas interpretativas aos direitos fundamentais que possam vir a ser violados pela IA e pelas novas tecnologias.

2.4.1.2.1. O princípio do respeito pela autonomia humana

A primeira observação que se pode retirar do princípio do respeito pela autonomia humana é que ele é uma consequência direta e natural dos direitos fundamentais do respeito pela dignidade humana e pela liberdade do indivíduo, que possuem como núcleo central a defesa e a promoção da autonomia dos seres humanos. Além disso, ele também pode ser entendido como uma implicação indireta da observância do princípio democrático e do princípio do Estado de Direito, pois suas respetivas legitimações estão atreladas à autonomia e à liberdade dos cidadãos.

Assim sendo, uma IA de confiança, à luz do princípio do respeito pela autonomia humana, deve assegurar o respeito à autodeterminação de todo o indivíduo que venha interagir com seus sistemas. O respeito a essa autodeterminação deve ser realizado de forma plena e efetiva, principalmente no que concerne à participação das pessoas no processo democrático[302]. Nesse sentido, ao invés de "subordinar, coagir, enganar, manipular, condicionar ou arregimentar injustificadamente"[303], um sistema de IA de confiança deve ser idealizado para "aumentar, complementar e capacitar as competências cognitivas, sociais e culturais dos seres humanos"[304].

Como se pode perceber, o respeito pela autonomia humana parte da premissa de que "os indivíduos têm o direito de decidir por si mesmos sobre o tratamento que recebem ou não"[305]. Além

[300] GPAN IA. Orientações Éticas para uma IA de Confiança... p.14.
[301] GPAN IA. Orientações Éticas para uma IA de Confiança... p.16.
[302] GPAN IA. Orientações Éticas para uma IA de Confiança... p.14-15.
[303] GPAN IA. Orientações Éticas para uma IA de Confiança... p.15.
[304] Ibid.
[305] Floridi et al. AI4Peaple... p. 697.

disso, tendo em vista que ao adotar sistemas de IA nós estamos, em parte, delegando voluntariamente uma parcela do nosso poder de decisão às máquinas, o princípio do respeito pela autonomia humana se apresenta como uma busca pelo equilíbrio[306]. Nesse sentido, através de um estudo comparado entre os projetos de regulação de IA, que também colocavam a autonomia como um dos princípios éticos para a IA, Floridi conclui que "não apenas a autonomia dos humanos deve ser promovida, mas também a autonomia das máquinas deve ser restringida e tornada intrinsecamente reversível, caso a autonomia humana precise ser restabelecida"[307].

Tudo isto posto, pode-se concluir que o ponto central da discussão sobre o respeito pela autonomia humana é a proteção do valor essencial da escolha humana no que se refere às decisões significativas ou que possam implicar relevantes consequências a ele mesmo ou a terceiros. Nesse sentido o GPAN IA afirma a importância que existe na garantia do controlo e da supervisão humana sobre os sistemas de IA[308]. A supervisão humana, parte importante do respeito pela autonomia humana, foi elevada à característica essencial de uma IA de confiança, que irá ser analisada futuramente. Floridi, por sua vez, no mesmo sentido de controlo e supervisão, apresenta o conceito de 'meta-autonomia'[309], onde "os seres humanos devem sempre conservar o poder de decidir que decisões tomar, exercendo a liberdade de escolha quando necessário, e cedendo-a nos casos em que razões imperiosas, como a eficácia"[310]. Destaca-se que, através da autonomia humana nos sistemas de IA, além da necessária informação e poder de decisão que o ser humano tem de escolher o que e quando delegar, também deve ser garantida a capacidade de revogar delegações realizadas.

2.4.1.2.2. O princípio da prevenção de danos

Quanto ao princípio da prevenção de danos, por mais que seu nome já traduza a sua intenção, existem observações importantes a serem realizadas. Nesse sentido, é importante tanto perceber a influência do respeito pela dignidade humana[311] quanto do respeito pela sua integridade física e mental[312]. Pelo que, de uma forma geral, os sistemas de IA não podem afetar negativamente os seres humanos, seja ao causar ou agravar danos físicos ou mentais[313]. O GPAN IA ainda faz uma importante observação, uma vez que inclui no conceito de dano a sua vertente individual ou coletiva, assim como a possibilidade de os danos serem intangíveis, afetando o meio ambiente, os ambientes sociais,

[306] Floridi et al. AI4People... p. 698.
[307] Ibid.
[308] GPAN IA. Orientações Éticas para uma IA de Confiança... p.15.
[309] Floridi apresenta duas nomenclaturas, sendo elas o modelo de "meta-autonomy" ou de "decide-to-delegate".
[310] Floridi et al. AI4People... p. 698
[311] Artigo 2.º do TUE.
[312] Artigo 3.º da Carta.
[313] GPAN IA. Orientações Éticas para uma IA de Confiança... p.15.

culturais e políticos. Sobre essa perspetiva, Floridi apresenta duas espécies de dano, os danos acidentais, que decorrem de um 'uso excessivo' da IA, e os deliberados, que são consequências de um 'uso indevido' da IA.

Floridi ainda afirma que, em termos gerais, o objetivo deve estar centrado simplesmente em evitar danos, independentemente se são ocasionados pela intenção humana ou pelo comportamento imprevisto das máquinas[314]. Para que isso seja possível, o GPAN IA afirmam que os sistemas de IA devem ser seguros e protegidos contra utilizações más intencionadas. Esse tema será mais bem abordado no requisito relativo à solidez técnica e de segurança que toda a IA de confiança deve possuir.

2.4.1.2.3. O princípio da equidade

O princípio da equidade, imperativo ético que deve guiar todo e qualquer sistema de IA, é um reflexo direto do respeito pela dignidade humana e do direito fundamental à igualdade e a não discriminação[315]. Contudo, a equidade também pode ser considerada como uma implicação indireta do direito de liberdade do indivíduo. Afirma-se isso uma vez que tratamentos injustamente discriminatórios ou inequitativos afetam diretamente a autonomia individual das pessoas, restringindo e limitando oportunidades de desenvolvimento cultural, social ou profissional a partir de um filtro relativo características físicas ou psicológicas.

Baseado nessas premissas, o GPAN IA considerou duas dimensões para a equidade, uma delas sendo a substantiva e a outra processual. Através da dimensão substantiva[316], sistemas de IA devem garantir um serviço equitativo e justo. Além disso, para ser equitativa, uma IA de confiança deve se afastar de enviesamentos injustos, discriminação e estigmatização contra pessoas e grupos. Uma outra perspetiva substantiva da equidade é a proporcionalidade necessária que os profissionais de IA devem possuir no momento de equilibrar os seus interesses com os objetivos em causa. Nesse sentido, o GPAN IA afirma que "quando diversas medidas concorrem entre si para a consecução de um fim, deve dar-se preferência à que for menos contrária aos direitos fundamentais e às normas éticas"[317]

A segunda dimensão da equidade é a processual, e se mostra como um meio de diminuir a disparidade de forças entre empresas e desenvolvedores de IA e seus utilizadores. Diz-se isso pois, aparentemente, essa dimensão se fortalece com os preceitos jurídicos encontrados no RGPD no que

[314] Floridi et al. AI4People... p. 697
[315] Artigo 2.º do TUE.
[316] GPAN IA. Orientações Éticas para uma IA de Confiança... p.15.
[317] Ibid.

concerne decisões individuais automatizadas[318]. Isso porque a dimensão processual[319] da equidade implica na possibilidade de se contestar uma decisão tomada por um sistema de IA. Além disso, tal equidade também determina a identificação da entidade responsável pela decisão e da explicabilidade dos processos decisórios. Que por sinal, compõe o quarto e último princípio ético de uma IA de confiança.

2.4.1.2.4. O princípio da explicabilidade

Como se mencionou, o princípio da explicabilidade está diretamente relacionado com a equidade, pois permite que seja aferido e comprovado que um sistema de IA é justo. Diante desse cenário, é possível afirmar que a explicabilidade também mantém ligações com o respeito pela dignidade humana, com a liberdade do indivíduo e com o respeito à justiça, pois a explicabilidade, nesse sentido, representa a transparência. Essa, que por sinal, é uma qualidade, ou uma necessidade, em diversas relações sociais e jurídicas, pois possibilita e fortalece o direito informacional do indivíduo, componente que permite a sua autonomia.

Nesse sentido, Floridi afirma que a explicabilidade é um princípio que complementa todos os demais[320]. Enquanto isso, o GPAN IA complementa ao afirmar que "a explicabilidade é crucial para criar e manter a confiança dos utilizadores nos sistemas de IA"[321]. A explicabilidade, por sua vez, deve compreender informações relativas às partes e finalidades dos sistemas, processos transparentes e decisões explicáveis. Quanto às decisões, a explicabilidade deve ser garantida às pessoas as quais ela afeta direta ou indiretamente. Além disso, o grau de necessidade de explicabilidade também deve ser valorado quanto à gravidade das consequências de uma decisão. Afirma-se isso pois, uma indicação de música ou filme não necessita da explicabilidade que um sistema médico de IA que identifica uma doença e indica um tratamento.

Ao final da análise dos princípios éticos, percebe-se que além de estarem diretamente ligados aos diretos fundamentais, eles também são diretos ou indiretamente refletidos em requisitos juridicamente vinculativos já existentes no ordenamento europeu. Uma análise mais profunda sobre essa perfectiva, notadamente relacionada ao RGPD, será realizada no capítulo terceiro. Contudo, ainda é importante afirmar que tais princípios originaram a criação de sete requisitos voltados para uma IA de confiança.

[318] Artigo 22.°, n.° 3 do RGPD
[319] GPAN IA. Orientações Éticas para uma IA de Confiança... p.15.
[320] Floridi et al. AI4People... p.700.
[321] GPAN IA. Orientações Éticas para uma IA de Confiança... p.16.

2.4.2. A concretização de uma IA de confiança através do cumprimento de sete requisitos

Como mencionado, a quadro para uma IA de confiança foi desenvolvido em três níveis de abstração. O primeiro, e mais abstrato, promovia uma discussão sobre os princípios éticos que sustentam uma IA de confiança. Contudo, com base no desígnio de concretizar as instruções principiológica desse quadro, foram criados sete requisitos essenciais para uma IA de confiança, sendo eles, sem qualquer grau hierárquico: ação e supervisão humanas[322], solidez técnica e segurança[323], privacidade e governação dos dados, transparência, diversidade, não discriminação e equidade; bem-estar societal e ambiental; e responsabilização.

2.4.2.1. Ação e supervisão humanas ou iniciativa e controlo por humanos

Baseado no respeito à dignidade humana, no direito à liberdade individual e no princípio ético do respeito pela autonomia humana, o requisito de 'ação e supervisão humanas'[324] surge em um contexto em que sistemas de IA possuem o potencial de "moldar e influenciar o comportamento humano mediante mecanismos talvez difíceis de detetar por utilizarem processos subconscientes, incluindo várias formas desleais de manipulação, engano, arregimentação e condicionamento, todas elas suscetíveis de pôr em risco a autonomia individual"[325].

Inserido nesse contexto, o requisito em questão exige que os sistemas de IA sejam facilitadores de uma sociedade democrática, próspera e equitativa, de maneira a apoiar a ação dos indivíduos e promover os seus respetivos direitos fundamentais[326]. Nesse sentido, "os sistemas de IA devem ajudar os indivíduos a fazerem escolhas melhores e mais informadas, em conformidade com os seus objetivos"[327].

Isto posto, assim como foi tratado da análise dos direitos fundamentais e dos princípios éticos, é importante que os sistemas de IA jamais limitem ou mitiguem a autonomia das pessoas. Pelo contrário, através de uma perspetiva legal e ética, os sistemas de IA devem facilitar, apoiar e incentivar a autodeterminação individual de cada ser humano. Dessa forma, perante um sistema de IA que possa vir a afetar negativamente o exercício dos direitos fundamentais, causando risco para indivíduos ou

[322] Ao adotar esse requisito em outros documentos, como o *White Paper* para IA e a COM (2019) 168 final, A Comissão Europeia modificou seu nome, passando a adotar a nomenclatura «Iniciativa e controlo por humanos». Para fins didáticos, e por manter o mesmo sentido, esse trabalho adotará a nomenclatura original das Orientações éticas.

[323] Ao adotar esse requisito em outros documentos, como o *White Paper* para IA e a COM (2019) 168 final, A Comissão Europeia modificou seu nome, passando a adotar a nomenclatura «Robustez e segurança». Para fins didáticos, e por manter o mesmo sentido, esse trabalho adotará a nomenclatura original das Orientações éticas.

[324] Denominado posteriormente pela Comissão Europeia como 'iniciativa e controlo por humanos'.

[325] GPAN IA. Orientações Éticas para uma IA de Confiança... p.19.

[326] GPAN IA. Orientações Éticas para uma IA de Confiança... p.18-19.

[327] COM (2019) 168 final. Aumentar a confiança numa inteligência artificial centrada no ser humano.

para a sociedade, é imperativa a necessidade de realização de uma avaliação do impacto nos direitos fundamentais[328][329]. Através desse ato, deve-se incluir uma avaliação no sentido de reduzir ou justificar tais riscos perante uma sociedade democrática.

Tendo em vista a perspetiva da supervisão e do controlo por humanos, esse requisito também determina que sejam assegurados graus adequados de medidas de controlo, capacidade de adaptação e a devida explicabilidade dos sistemas baseados em IA. Essa supervisão pode ser realizada através de mecanismos de governação, pelo que as Orientações éticas apresentam algumas possibilidades que não serão tratadas nessa pesquisa[330]. A preocupação pela presença dessas características se dá para proteger os utilizadores, como também permitir que as autoridades públicas exerçam o seu poder de supervisão.

2.4.2.2. Solidez técnica e segurança ou robustez e segurança

O requisito de solidez técnica e segurança[331] é diretamente ligado ao direito fundamental do respeito pela sua integridade física e mental[332] e ao princípio ético da prevenção de danos. Esse requisito pressupõe que sistemas de IA possuam uma abordagem de prevenção de riscos, e que se comportem e executem suas ações "fiavelmente conforme o previsto, minimizando os danos não intencionais e inesperados, e prevenindo os danos inaceitáveis"[333]. Para que os sistemas de IA alcancem esse requisito, eles deverão observar quatro características[334], sendo elas a resiliência perante ataques e segurança; um plano de recurso e segurança geral; a exatidão; e a fiabilidade e reprodutibilidade.

Nesse sentido, sistemas de IA devem ser, primordialmente, protegidos contra vulnerabilidades que possam ser exploradas por pessoas mau intencionadas. Isso porque um ataque ao sistema de IA, seja à sua estrutura, aos seus dados ou ao seu modelo, podem comprometer o comportamento da IA. Esse comprometimento pode causar instabilidade no sistema, assim como promover decisões imprecisas. A segunda característica se insere paralelamente a esse contexto de possíveis vulnerabilidade e complicações, uma vez que é importante que existam salvaguardas que possibilitem um plano de recurso ou contingência em caso de problemas.

No que se refere a exatidão, deve-se entender como sendo a capacidade de realização de apreciações corretas. Tal característica é importante pois, se um sistema de IA interpreta incorretamente os dados e as informações referentes à um indivíduo, consequentemente serão

[328] GPAN IA. Orientações Éticas para uma IA de Confiança... p.19.
[329] Posteriormente será apresentada uma proposta de Avaliação de Impacto dos Direitos Fundamentais.
[330] GPAN IA. Orientações Éticas para uma IA de Confiança... p.19.
[331] Denominado posteriormente pela Comissão Europeia como 'robustez e segurança' na COM(2019) 168 final
[332] Artigo 3.º da Carta.
[333] GPAN IA. Orientações Éticas para uma IA de Confiança... p.20.
[334] Ibid.

tomadas decisões incorretas, e, portanto, passíveis de serem injustas. Nesse sentido, o nível de exatidão deve ser diretamente proporcional à intensidade que um sistema de IA afeta vidas humanas.

Por fim, e de acordo com o GPAN IA, um sistema de IA é fiável quando, a partir de diversos tipos de dados de entrada, funciona adequadamente em diferentes situações. Por outro lado, um sistema de IA possui reprodutibilidade quando, repetida as mesmas condições, é capaz de apresentar o mesmo comportamento. Além disso, o sistema de IA deve seguir, em todas as suas fases, os princípios de segurança e proteção por *design*, como também processos de minimização ou reversibilidade de consequências ou de erros não intencionais[335].

2.4.2.3. Privacidade e governação dos dados

Esse requisito é uma consequência direta do imperativo ético de prevenção de danos, além de também estar relacionado com o direito fundamental da proteção de dados. Afirma-se isso pois, como visto, os danos podem ser de natureza imaterial. Nesse sentido, a má utilização de sistemas de IA pode causar séria ameaça à privacidade dos indivíduos, principalmente quando se vive em um contexto de uma economia baseada na vigilância.

Atualmente, ao analisar os registos e as atividades digitais de uma pessoa, os sistemas de IA conseguem inferir uma série de informações, como preferências, gênero, etnia, idade, orientação sexual e até convicções políticas e religiosas[336]. Isto posto, para que sistemas de IA sejam confiáveis, esses terão de garantir o respeito aos princípios relativos ao tratamento de dados pessoais[337], assegurando um tratamento lícito, leal e transparente; com finalidade determinada; restrito a dados adequados, pertinentes e limitados. O respeito e a adequação ao RGPD devem também ser demonstradas através da adoção dos conceitos de proteção de dados por conceção e por defeito em todas as fases do ciclo de vida do sistema de IA[338].

Além disso, uma IA de confiança deve estar atenta à qualidade e a integridade dos dados, uma vez que dados inexatos, incorretos ou enviesados socialmente podem causar decisões injustas e discriminatórias. Por fim, toda empresa ou organização que realize o tratamento de dados pessoais deve envidar esforços no sentido de proteger o acesso dos dados, não permitindo acessos ilegais ou ilegítimos, caracterizadores de vazamento de dados e violações de privacidade.

[335] Aumentar a confiança numa inteligência artificial centrada no ser humano
[336] GPAN IA. Orientações Éticas para uma IA de Confiança... p.20.
[337] Artigo 5.° n.° 2 do Regulamento Geral de Proteção de Dados
[338] GPAN IA. Orientações Éticas para uma IA de Confiança... p.21.

2.4.2.4. Transparência

O requisito da transparência está diretamente relacionado com o princípio ético da explicabilidade, e comporta três características, sendo elas a rastreabilidade, a comunicação e a própria explicabilidade.

A rastreabilidade é conferida a um sistema de IA quando se é possível registar e documentar o processo de tomada de decisão e as próprias decisões em si[339]. A rastreabilidade, portanto, deve ser percebida como a possibilidade de auditar os sistemas de IA. Essa é uma característica que tem o potencial de aumentar a transparência, uma vez que se torna possível identificar os motivos pelos quais ocorreu erro em uma decisão de IA.

No que diz respeito à comunicação, o GPAN IA acredita que, através do princípio ético da explicabilidade e do requisito de transparência, uma IA de confiança jamais pode se apresentar aos seus utilizadores como se fosse um ser humano[340]. De facto, caso isso ocorra, deve ser considerada como uma ação que visa enganar ou confundir o utilizador.

Por fim o, a explicabilidade, por sua vez, diz respeito a capacidade de explicar tanto o processo técnico de um sistema de IA quanto as decisões humanas relacionadas a esse processo[341]. Nesse sentido, a explicabilidade pode ser compreendida como a exigência de que decisões tomadas por sistemas de IA possam ser compreendidas pelas partes interessadas. Assim sendo, e de acordo com orientações do GPAN IA, "sempre que um sistema de IA tenha um impacto significativo na vida das pessoas, deverá ser possível solicitar uma explicação adequada do respetivo processo de tomada de decisões"[342]. Importante destacar que a possibilidade de explicação do processo decisório impede a opacidade dos sistemas de IA. A opacidade, abordada no capítulo primeiro, consiste em um impacto negativo na vida dos utilizadores, pois os impede de exercício de direitos, como o acesso a justiça.

Além disso, outra vertente da explicabilidade está relacionada com a devida transparência no que se refere ao grau de influência e orientação que um sistema de IA tem sobre o processo de tomada de decisão dentro da organização[343]. Tal informação é crucial pois permite ter ciência de até que ponto há intervenção humana nas decisões da empresa prestadora de serviços.

[339] Aumentar a confiança numa inteligência artificial centrada no ser humano
[340] GPAN IA. Orientações Éticas para uma IA de Confiança... p.22.
[341] Ibid.
[342] Ibid.
[343] Ibid.

2.4.2.5. Diversidade, Não Discriminação e Equidade

O requisito de diversidade, não discriminação e equidade, além de ser um direito fundamental[344], também está estritamente ligado ao respeito pela dignidade humana. No caso em questão, esse requisito é um desdobramento do princípio da equidade. Sua observação é de tamanha importância, uma vez que uma IA de confiança somente se tornará realidade se esse requisito estiver presente em todo o ciclo de vida dos sistemas de IA[345].

Como visto anteriormente, os dados são considerados um retrato da sociedade[346]. Pelo que é provável que conjuntos de dados possam ser afetados por desvios históricos e por condutas discriminatórias estruturais. Nesse contexto, a utilização de dados enviesados pode originar (in)diretamente decisões injustamente discriminatórias e preconceituosas contra minorias e grupos vulneráveis. De acordo com o GPAN IA, o "enviesamento identificável e discriminatório deve ser eliminado na fase de recolha de dados"[347]. Para que isso seja realizado, é necessária a adoção de processos de supervisão e de governança de dados e dos sistemas de IA. Aliás, para o fortalecimento desse requisito, é importante que as equipes sejam formadas por pessoas de diferentes origens, culturas e disciplinas[348]. Isso porque, uma equipe baseada na diversidade, seja cultural, disciplinar ou de opinião, tem maiores possibilidades de assegurar a não discriminação e a equidade.

2.4.2.6. Bem-estar societal e ambiental.

Também de acordo com o princípio da equidade e da prevenção de danos, o requisito de bem-estar societal e ambiental representa uma preocupação com a sociedade em geral, com os seres sensíveis e com o ambiente, onde todos devem ser considerados como partes interessadas ao longo do ciclo de vida da IA[349]. para que uma IA seja de confiança "há que ter em conta o seu impacto sobre o ambiente e outros seres sencientes"[350]. Nesse sentido, os sistemas de IA devem ser desenvolvidos a partir de preceitos de sustentabilidade e responsabilidade ecológica.

Ainda de acordo com o requisito de bem-estar societal e ambiental, é preciso que os sistemas de IA sejam sustentáveis e respeitadores do meio ambiente. Para que isso seja possível, os processos de desenvolvimento, implantação e utilização do sistema devem estar atentos às escolhas menos

[344] Artigo 21.ª da Carta.
[345] GPAN IA. Orientações Éticas para uma IA de Confiança... p.23.
[346] Cortiz, Diogo. O Design pode ajudar na construção de Inteligência Artificial humanística?
[347] GPAN IA. Orientações Éticas para uma IA de Confiança... p.22.
[348] Ibid.
[349] GPAN IA. Orientações Éticas para uma IA de Confiança... p.23.
[350] Aumentar a confiança numa inteligência artificial centrada no ser humano

prejudiciais de consumo de energia[351]. Por outro lado, os sistemas de IA também devem se preocupar com o impacto social[352] que a sua utilização causa na sociedade. Nesse sentido, é importante que se realizem avaliações sobre tais impactos sobre as competências sociais, físicas e mentais de seus utilizadores.

Por fim, tal requisito também está, de certa forma, relacionado com o respeito à democracia, à justiça e ao Estado de Direito. Nesse sentido, os sistemas de IA também devem levar em consideração o impacto de seus desenvolvimentos numa perspetiva societal[353]. Assim sendo, sistemas de IA deverão avaliar os seus efeitos nas instituições, na democracia e na sociedade em geral, principalmente no que concerne ao processo democrático.

2.4.2.7. Responsabilização

Por fim, o sétimo e último requisito é voltado para a responsabilização, completando o grupo de sete requisitos essenciais para uma IA de confiança. De acordo com o GPAN IA, a responsabilização está relacionada com o princípio da equidade, pelo que exige que sejam criados mecanismos para garantir tanto a responsabilidade quanto a responsabilização pelos resultados de sistemas de IA[354]. Para que se alcance tal requisito, uma IA de confiança deve apresentar quatro características[355], sendo elas a auditabilidade; a minimização e comunicação dos impactos negativos; soluções de compromissos; e vias de recurso.

Aliás, a auditabilidade, característica que vem sendo presente em diversos requisitos, é fundamental em um contexto de responsabilização. Isso porque ela permite a avaliação de algoritmos, dados e processos decisórios. Além disso, a disponibilidade dos relatórios das auditorias contribui para o estabelecimento da transparência, e com isso, aumento da confiança. No entanto, deve ser priorizada auditorias independentes quando se perceber que uma aplicação implica em riscos críticos para a segurança ou para os direitos fundamentais[356].

Ainda em um contexto que envolvem elevados riscos, os desenvolvedores devem estar atentos para a realização de avaliação de impactos inerentes às aplicações de IA[357]. Tais avaliações devem ser realizadas antes e durante a fase de desenvolvimento. Para as aplicações alimentadas por dados pessoais, deve ser observada a necessidade e realização da avaliação de impacto da proteção de

[351] GPAN IA. Orientações Éticas para uma IA de Confiança... p.23.
[352] Ibid.
[353] Ibid.
[354] Aumentar a confiança numa inteligência artificial centrada no ser humano
[355] GPAN IA. Orientações Éticas para uma IA de Confiança... p. 24-25.
[356] GPAN IA. Orientações Éticas para uma IA de Confiança... p.24.
[357] Ibid.

dados[358]. O uso desses instrumentos irá auxiliar tanto na documentação que envolve o desenvolvimento da IA quanto no possível caminho a ser seguido para minimizar os riscos.

Quanto as soluções de compromisso, essas deverão ser adotadas sempre que, durante o desenvolvimento do sistema de IA, forem percebidos possíveis conflitos na aplicação dos requisitos de uma IA de confiança. As soluções de compromisso devem ser elaboradas de forma racional e metodológica de acordo com o conhecimento esperado[359]. O GPAN IA alerta que, encontrados conflitos que não apresentam uma solução eticamente aceitável, não se deve dar seguimento ao desenvolvimento, implantação ou uso do sistema de IA[360]. Essa é uma importante observação, pois se encontrados problemas relacionados ao risco elevado e a falta de uma compatibilidade aceitável, não há motivos para continuar o desenvolvimento de um sistema que certamente irá gerar impactos negativos a seus utilizadores.

Além dessas características, também se faz necessário prever – e dar publicidade a – mecanismos acessíveis para que os utilizadores acedam as vias de recurso adequadas. Essas vias de recurso devem incluir a possibilidade de pedido de esclarecimentos e explicações, contestação de decisões e reclamações por possíveis danos. Por último, e de acordo com a CE, deve sempre ser garantido uma reparação adequada, quando da ocorrência de impactos adversos e injustos[361].

Esses foram os requisitos essenciais de uma IA de confiança. E apesar das Orientações não possuírem força vinculativa, muitos dos requisitos encontram reflexo em diplomas jurídicos em vigor. Pelo que os requisitos podem ser considerados um caminho a ser seguido pelos desenvolvedores de sistemas de IA. A próxima seção irá apresentar o terceiro nível relativo ao quadro para uma IA de confiança, onde será possível perceber uma lista de avaliação a ser empregue para operacionalizar um sistema de IA aos padrões éticos europeus.

2.4.3. A transformação de uma ideia em ação: a Lista de Avaliação para uma IA de confiança

A 'Lista de Avaliação para uma IA de confiança'[362] (Lista de Avaliação) foi desenvolvida com base da proteção dos direitos fundamentais possui a intenção de operacionalizar uma IA de confiança através de questionamentos que aferem se os sete requisitos abordados anteriormente estão devidamente implementados. Dentre tantas informações apresentadas pelo quadro ético para uma IA de confiança, Fanny Hidvégi, gerente de políticas do grupo *Access Now*, afirma que a Lista de Avaliação

[358] Artigo 35.º do RGPD.
[359] GPAN IA. Orientações Éticas para uma IA de Confiança... p.24.
[360] Ibid.
[361] Aumentar a confiança numa inteligência artificial centrada no ser humano
[362] Grupo de Peritos de Alto Nível em IA. Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment. «https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment»

é a parte mais importante das Orientações éticas, pois a lista fornece uma perspetiva prática a respeito de como eliminar ou mitigar os danos potenciais da IA[363].

Hidvégi está correta em afirmar isso, não somente por ter sido ela uma das responsáveis pelo desenvolvimento da Lista de Avaliação final[364], mas porque, de facto, o atual documento do GPAN IA apresenta concretamente um caminho a ser seguido no que concerne ao desenvolvimento de IA de confiança. Para se ter uma ideia, a presente lista conta com uma abordagem voltada para os sete domínios e seus respetivos subdomínios de uma IA de confiança, que são a ação e supervisão humanas; a solidez técnica e segurança; a privacidade e governação dos dados; a transparência; a diversidade, não discriminação e equidade; o bem-estar societal e ambiental; e a responsabilização. Como também conta, em sua versão final, com aproximadamente 124 questionamentos.

Isto posto, e tendo em vista que um dos objetivos do trabalho é entender o que é uma IA de confiança, para que posteriormente possamos analisar se o RGPD possui o alcance necessário para auxiliar a Comissão Europeia no desenvolvimento desse desígnio, o presente trabalho irá apenas apresentar a essência da respetiva Lista de Avaliação para uma IA de confiança. Essa decisão é tomada pois a Lista de Avaliação é um documento extremamente extenso, e que a sua análise a fundo resultaria em uma fuga do objeto central de pesquisa.

Assim sendo, será iniciada a apresentação com base em domínios e respetivos subdomínios da Lista de Avaliação para uma IA de confiança. Criado esse contexto, serão enumeradas, a título exemplificativo, 25 (vinte e cinco) questionamentos a fim de que se demonstre a essência primordial desse documento. A respetiva escolha ocorre pelo grau de importância identificado, como também por já terem composto uma lista similar da professora Alexandra Aragão[365].

O primeiro dos mencionados domínios é o da de ação e supervisão humanas. Como já abordado na seção destinada aos requisitos de uma IA de confiança, esse domínio subdividisse na exigência de se promover a autonomia da ação humana e na necessidade de se comprovar o controlo sobre os sistemas através da supervisão humana. Dentre os questionamentos que podem ser realizados, abaixo há a seleção daqueles considerados como principais[366].

i. O sistema de IA foi projetado para interagir, guiar ou tomar decisões por utilizadores finais humanos que afetam os humanos ou a sociedade? Caso positivo, os utilizadores finais ou outros sujeitos são devidamente informados de que uma decisão, conteúdo, conselho ou resultado é o fruto de uma decisão algorítmica?

[363] Vincent, J. AI systems should be accountable, explainable, and unbiased, says EU. The Verge https://www.theverge.com/2019/4/8/18300149/eu-artificial-intelligence-ai-ethical-guidelines-recommendations (2019).
[364] GPAN IA. Assessment List for Trustworthy Artificial Intelligence... p.31
[365] Alexandra Aragão. Questões ético-jurídicas relativas ao uso de apps geradoras de dados de mobilidade para vigilância epidemiológica da Covid-19. Uma perspetiva Europeia. Instituto Jurídico da Faculdade de Direito da Universidade de Coimbra, p.10-12.
[366] GPAN IA. Assessment List for Trustworthy Artificial Intelligence... p.7-8

ii. O sistema de IA poderia afetar a autonomia humana interferindo no processo de tomada de decisão do utilizador final de qualquer outra forma não intencional e indesejável? Você implementou algum procedimento para evitar que o sistema de IA afete inadvertidamente a autonomia humana?

iii. O sistema de IA corre o risco de criar apego humano, estimular o comportamento viciante ou manipular o comportamento do utilizador? Se sim, você tomou medidas para lidar com as possíveis consequências negativas relativas a esses impactos?

iv. Por favor, determine se o sistema de IA (escolha quantos forem apropriados): É um sistema de autoaprendizagem ou autônomo; é supervisionado por um *Human-in-the-Loop*[367]; é supervisionado por um *Human-on-the-Loop*[368]; é supervisionado por um *Human-in-Command*[369].

v. Você estabeleceu algum mecanismo de deteção e resposta para efeitos adversos indesejáveis do sistema de IA para o utilizador final ou sujeito?

O segundo domínio é o da solidez técnica e segurança. Como já abordado na seção destinada aos requisitos de uma IA de confiança, esse domínio subdividisse no fortalecimento da segurança e na presença de resiliência perante ataques mal-intencionados. Além disso, é importante a existência de salvaguardas, o que inclui um plano de recurso e de segurança geral. Também deve se ter atenção quanto a necessária exatidão interpretativa do sistema de IA, e nas características de fiabilidade e reprodutibilidade. Dentre os questionamentos que podem ser realizados, abaixo há a seleção daqueles considerados como principais[370].

vi. O sistema de IA pode ter efeitos adversos, críticos ou prejudiciais (por exemplo, para a segurança humana ou social) em caso de riscos ou ameaças, como falhas de projeto ou técnicas, defeitos, interrupções, ataques, uso indevido, uso impróprio ou malicioso?

vii. Você implementou medidas para garantir a integridade, robustez e segurança geral do sistema de IA contra-ataques potenciais ao longo de seu ciclo de vida? Incluindo ataques cibernéticos?

viii. Você desenvolveu um mecanismo para avaliar quando o sistema de IA foi alterado para merecer uma nova revisão de sua robustez técnica e segurança?

ix. Você implementou medidas para garantir que os dados (incluindo dados de treino) usados para desenvolver o sistema de IA estejam atualizados, de alta qualidade, completos e representativos do ambiente no qual o sistema será implantado?

---

[367] Human-in-the-loop refere-se à capacidade de intervenção humana em cada ciclo de decisão do sistema, que em muitos casos não é possível nem desejável.

[368] Human-on-the-loop refere-se à capacidade de intervenção humana durante o ciclo de *design* do sistema e monitoramento da operação do sistema.

[369] Human-in-Command refere-se à capacidade de supervisionar a atividade geral do sistema de IA (incluindo seu impacto econômico, social, legal e ético mais amplo) e a capacidade de decidir quando e como usar o sistema em qualquer situação particular. Isso pode incluir a decisão de não usar um sistema de IA em uma situação específica, para estabelecer níveis de discrição humana durante o uso do sistema ou para garantir a capacidade de anular uma decisão tomada por um sistema.

[370] GPAN IA. Assessment List for Trustworthy Artificial Intelligence... p.9-11

x. Você implementou um procedimento adequado para lidar com os casos em que o sistema de IA produz resultados com uma pontuação de confiança baixa?

O terceiro domínio é o da privacidade e governação dos dados. Inicialmente, esse domínio exige o respeito à proteção de dados, o que implica a necessidade de conformidade com o RGPD. Além disso, há também a necessidade de atenção quanto a qualidade e integridade dos dados, uma vez que a ausência dessas características pode causar potencial discriminação negativa aos utilizadores. Por fim, o controlo sobre o acesso aos dados também deve ser uma prioridade, pois o contrário implica em vazamento de dados e violação da privacidade. Dentre os questionamentos que podem ser realizados, abaixo há a seleção daqueles considerados como principais[371].

xi. Você implementou medidas para garantir a conformidade com o RGPD (por exemplo, avaliação do impacto da proteção de dados, nomeação de um oficial de proteção de dados, minimização de dados etc.)?

xii. Você implementou o direito de retirar o consentimento, o direito de contestar e o direito de ser esquecido no desenvolvimento do sistema de IA?

xiii. Você alinhou o sistema de IA com os padrões relevantes (por exemplo, ISO[372], IEEE[373]) ou protocolos amplamente adotados para gerenciamento e governança de dados?

O quarto domínio é o da transparência. Como já abordado na seção destinada aos requisitos de uma IA de confiança, esse domínio subdividisse na possibilidade de rastreamento, e consequente auditoria, das decisões e dos processos decisórios; da transparência necessária relativa as comunicações realizadas entre sistemas de IA e utilizadores; e no desenvolvimento de meios que possibilitem a explicabilidade das decisões e dos processos decisórios. Dentre os questionamentos que podem ser realizados, abaixo há a seleção daqueles considerados como principais[374].

xiv. Você implementou medidas que abordam a rastreabilidade do sistema de IA durante todo o seu ciclo de vida? Você pode rastrear quais dados foram usados pelo sistema de IA para tomar determinada decisão ou recomendação? Você pode rastrear quais modelos ou regras de IA levaram à decisão ou recomendação do sistema de IA?

xv. Você explicou as decisões do sistema de IA aos utilizadores?

xvi. Você estabeleceu mecanismos para informar os utilizadores sobre a finalidade, critérios e limitações das decisões geradas pelo sistema de IA?

O quinto domínio é o da diversidade, não discriminação e equidade. Esse domínio se baseia principalmente na prevenção de enviesamentos injustos, evitando qualquer tipo de discriminação

[371] GPAN IA. Assessment List for Trustworthy Artificial Intelligence... p.12-13.
[372] https://www.iso.org/committee/6794475.html.
[373] https://standards.ieee.org/industry-connections/ec/autonomous-systems.html
[374] GPAN IA. Assessment List for Trustworthy Artificial Intelligence... p.14-15.

negativa a minorias ou grupos vulneráveis. Para que isso possa ocorrer, é importante que haja participação diversificada de pessoas, culturas e opiniões ao longo do desenvolvimento do sistema de IA. Dentre os questionamentos que podem ser realizados, abaixo há a seleção daqueles considerados como principais[375].

xvii. Você estabeleceu uma estratégia ou um conjunto de procedimentos para evitar a criação ou o reforço de um viés injusto no sistema de IA, tanto no que se refere ao uso de dados de entrada quanto ao projeto do algoritmo? Você garantiu um mecanismo que permite a sinalização de problemas relacionados a preconceito, discriminação ou baixo desempenho do sistema de IA?

xviii. Você avaliou se a interface do utilizador do sistema de IA pode ser usada por pessoas com necessidades especiais ou deficiências ou por pessoas em risco de exclusão?

xix. Você considerou um mecanismo para incluir a participação do maior número possível de partes interessadas no *design* e desenvolvimento do sistema de IA?

O sexto domínio é o do bem-estar societal e ambiental. Como já abordado na seção destinada aos requisitos de uma IA de confiança, esse domínio subdividisse na exigência de se promover uma IA sustentável e que respeite o ambiente. Além disso, é necessária uma preocupação com o impacto social que os sistemas de IA possam causar nas competências físicas e mentais dos utilizadores. Por fim, também há a necessidade de análise sobre o impacto que o sistema de IA pode causar na sociedade e na democracia como um todo, principalmente relativo aos processos democráticos. Dentre os questionamentos que podem ser realizados, abaixo há a seleção daqueles considerados como principais[376].

xx. Existem potenciais impactos negativos do sistema de IA no meio ambiente? Que impactos potenciais você identifica?

xxi. O sistema de IA pode criar o risco de desqualificação da força de trabalho? Você tomou medidas para neutralizar os riscos de desqualificação?

xxii. O sistema de IA poderia ter um impacto negativo na sociedade em geral ou na democracia? Você avaliou o impacto social do uso do sistema de IA além do utilizador final e do titular de dados, como partes interessadas potencialmente afetadas indiretamente ou a sociedade em geral? Você tomou medidas para minimizar o potencial dano social ao sistema de IA? Você tomou medidas para garantir que o sistema de IA não tenha um impacto negativo sobre a democracia?

E o sétimo e último domínio é o da responsabilização. Esse domínio, assim como já foi abordado anteriormente, prega a possibilidade de se promover auditorias nos algoritmos, nos dados e

[375] GPAN IA. Assessment List for Trustworthy Artificial Intelligence... p.16-18.
[376] GPAN IA. Assessment List for Trustworthy Artificial Intelligence... p.19-20

nos processos decisórios. Há também a necessidade de se realizar avaliações de impacto para identificar e minimizar possíveis impactos negativos. Por fim, é imprescindível que se disponibilize e se de publicidades a canais e vias de recurso a serem usadas pelos utilizadores, no sentido de estabelecer uma comunicação entre a empresa e as pessoas interessadas. Dentre os questionamentos que podem ser realizados, abaixo há a seleção daqueles considerados como principais[377].

xxiii. Você estabeleceu mecanismos que facilitam a auditabilidade do sistema de IA (por exemplo, rastreabilidade do processo de desenvolvimento, o fornecimento de dados de treino e o registo dos processos, resultados, impacto positivo e negativo do sistema de IA)? Você garantiu que o sistema de IA pode ser auditado por terceiros independentes?

xxiv. Você organizou um treino de risco e, em caso afirmativo, isso também informa sobre a possível estrutura legal aplicável ao sistema de IA?

xxv. Para aplicativos que podem afetar adversamente os indivíduos, os mecanismos de reparação por *design* foram colocados em prática?

2.5. A Proposta de Regulamento sobre Inteligência Artificial da Comissão Europeia

Em 21 de abril de 2021, a Comissão tornou pública a sua proposta de regulamento que estabelece regras harmonizadas sobre Inteligência Artificial na União Europeia[378]. Essa proposta é consequência do trabalho que a Comissão tem realizado desde o ano de 2018, e se fundamenta em diversos documentos que foram analisados nesse capítulo, entre eles "As Orientações Éticas para uma Inteligência Artificial de Confiança" e o "Livro Branco sobre Inteligência Artificial".

A proposta de regulamento tem como objetivo elevar a proteção dos interesses da sociedade e os direitos fundamentais dos indivíduos perante os riscos de dano que podem ser causados pela IA. Para que alcance esse objetivo, a proposta se baseia no conceito europeu de uma IA de confiança[379], buscando desenvolver uma IA responsável, antropocêntrica e guiada por princípios éticos, como o respeito pela autonomia humana, a equidade, a prevenção de danos e a explicabilidade.

Destaca-se que o quadro jurídico proposto pela Comissão busca o equilíbrio necessário para assegurar os direitos fundamentais e, ao mesmo tempo, proporcionar segurança jurídica para fomentar a inovação e o desenvolvimento de novos sistemas de IA[380], possibilitando assim que a União se torne um líder mundial no desenvolvimento de uma IA ética, segura e de confiança[381].

---

[377] GPAN IA. Assessment List for Trustworthy Artificial Intelligence... p.21-22.

[378] Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain union legislative acts. Brussels, 21.4.2021 COM(2021) 206 final 2021/0106 (Proposta de Regulação de IA)

[379]da Proposta de Regulação de IA, *Explanatory Memorandum*, p.1...

[380] Considerando 11 do Regulamento relativo a uma abordagem europeia para a inteligência artificial.

[381] Considerando 05 da Proposta de Regulação de IA...

A proposta europeia não optou por regular de forma ampla a tecnologia "inteligência artificial". Pelo contrário, ao adotar uma abordagem regulatória baseada em risco, a proposta optou por criar um conjunto de regras específicas que serão aplicadas a determinados usos que representem um alto risco de dano, seja aos interesses da sociedade ou aos direitos fundamentais dos indivíduos. Essa abordagem havia sido anunciada pela Comissão no Livro Branco de Inteligência Artificial.

Nesse sentido, a proposta de regulamento apresenta um grupo de sistemas e aplicações de IA que possuem um risco inaceitável, e por isso tiveram o seu uso explicitamente proibido[382]. Essa proibição se justifica, pois, alguns usos de IA violam os direitos fundamentais e valores da União, uma vez que: (i) promovem vigilância biométrica em massa e em tempo real; (ii) exploram vulnerabilidades e características de uma pessoa ou de um grupo de pessoa com a finalidade de distorcer e manipular o seu comportamento; ou (iii) pretendem avaliar ou classificar a confiabilidade de pessoas naturais com base no seu comportamento social ou características pessoais. Tais proibições não são absolutas, e alguns desses usos de IA poderão ser permitidos em situações excecionais descritas na proposta.

Além disso, a proposta apresentou um conjunto de requisitos específicos que devem ser observados e respeitados por sistemas de IA que possam representar alto risco para a saúde e segurança ou para os direitos fundamentais dos indivíduos[383], como: implementação de gestão do risco; governação dos dados; exatidão, robustez, e cibersegurança, documentação técnica; supervisão humana; e transparência[384]. Além de cumprir esses requisitos, esses sistemas de IA devem apresentar uma avaliação de conformidade[385] antes de serem colocados no mercado da União[386].

De facto, existem ainda diversos dispositivos de importância significativa para o tema, como definições, obrigações, responsabilidades, entre outros. Contudo, essa proposta não será objeto de uma análise mais profunda. Isso porque, como a sua própria nomenclatura afirma, trata-se apenas de uma proposta que foi recentemente tornada pública e ainda irá enfrentar um árduo processo legislativo, o que pode levar um período de tempo considerável.

Apesar de sua existência, o cenário europeu ainda se encontra sem uma regulamentação específica e aplicável para a IA. O que exige que diplomas setoriais sejam responsáveis pela tutela de sistemas de IA[387]. Isso posto, torna-se válido, atual e necessário analisar se o RGPD é um instrumento adequado para a regulação e desenvolvimento de sistemas de IA de confiança.

---

[382] Artigo 5.º da Proposta de Regulação de IA...

[383] Artigo 8.º e ss da Proposta de Regulação de IA...

[384] Esses requisitos são um espelho dos requisitos que consagram uma IA de confiança, e que forma analisados na seção 2.4.2.

[385] Uma avaliação de conformidade é o processo de demonstrar se os requisitos especificados relacionados com um sistema de IA foram cumpridos. Artigo 3.º (20) da Proposta de Regulação de IA...

[386] Considerando 62 e Artigo 16 (e) da Proposta de Regulação de IA...

[387] Considerando que uma proposta de regulação não produz efeitos legais, que não se tem certeza de quando a proposta será aprovada, e se ela será aprovada no mesmo molde apresentado, conclui-se que a premissa inicial dessa pesquisa ainda se mantem, ou seja, não existe atualmente um regulamento especificamente aplicado a IA. Isso posto, continua válida, atual e necessária uma análise da adequação do RGPD como instrumento de regulação do ideal europeia de uma IA de Confiança.

CAPÍTULO TERCEIRO

A BUSCA PELA COMPREENSÃO SOBRE O PAPEL E OS LIMITES DO RGPD NA REGULAÇÃO DE UMA IA DE CONFIANÇA

3.1. A Proteção de dados pessoais como direito fundamental

O direito fundamental à proteção de dados está positivado no Artigo 8.º da Carta, que afirma que "todas as pessoas têm direito à proteção dos dados de caráter pessoal que lhes digam respeito". Nesse sentido, Alessandra Silveira afirma que a Carta havia dado um passo adiante em relação a várias Constituições dos Estados-Membros da UE, uma vez que passou a "consagra[r] um direito fundamental que protege dados que não têm de ser privados e muito menos íntimos – basta que sejam pessoais"[388].

Assim, além de reconhecer a natureza fundamental do direito à proteção de dados, o preceito encontrado no artigo 8.º da Carta, composto por três números, ainda apresenta o que viriam a ser princípios, fundamentos e direitos relacionados ao tratamento de dados. Entre eles, (i) a necessidade de um tratamento leal, (ii) restrito à finalidade (iii) e com observância do consentimento do titular de dados ou de outro fundamento legítimo, além dos direitos de (iv) aceder aos dados que lhe digam respeito e (v) obter a respetiva retificação. Por fim, em seu n.º 3 ainda é mencionada a necessidade de fiscalização por parte de uma (vi) autoridade independente.

Isto posto, no sentido de fortalecer o posicionamento da Carta, o artigo 16.º, n.º 1, do TFUE também passou a reconhecer a proteção de dados de caráter pessoal como um direito fundamental de todas as pessoas. No entanto, há algo interessante a se destacar, uma vez que, de acordo com as observações de Luiz Neto Galvão, "a proteção de dados transcende muito a mera dimensão econômica do mercado interno"[389]. Para sustentar essa afirmação, Galvão destaca a posição do direito fundamental da proteção de dados dentro do TFUE, que pode ser encontrado integrado na Parte I, intitulada 'Os Princípios', sob o Título II, que consagra as 'Disposições de Aplicação Geral', do TFUE. Além disso, Galvão afirmou que o artigo 16.º do TFUE passou então a estabelecer um "verdadeiro regime geral de proteção de dados aplicável à totalidade da ação da UE"[390].

Essa última afirmação não é realizada sem qualquer fundamento. Isso porque, assim como afirma Alessandra Silveira, o artigo 16.º, n.º 2, do TFUE é quem confere ao Parlamento Europeu e o Conselho a competência de estabelecer normas relativas à proteção das pessoas singulares no que diz

[388] Silveira, Alessandra e Marques, João. Do direito a estar só ao direito ao esquecimento. Considerações sobre a proteção de dados pessoais informatizados no direito da união europeia: sentido, evolução e reforma legislativa, p.94.
[389] Anastácio, Manuel Lopes Porto Gonçalo. Tratado de Lisboa - Anotado e Comentado. Almedina, p.5114, Edição do Kindle.
[390] Ibid.

respeito "(i) ao tratamento de dados pessoais [...][391] e (ii) à livre circulação de dados entre os Estados-Membros e garantia de um nível de proteção adequado para a transferência de dados pessoais para países terceiros"[392].

Nesse sentido, possuindo as competências legislativas necessárias e percebendo a fragmentação à nível europeu acerca da proteção de dados, a União Europeia finalmente avançou com o procedimento legislativo e publicou o Regulamento (UE) 2016/679, sendo este o atual responsável pela proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados, dando observância ao padrão principiológico e jusfundamental existente nesta ordem jurídica, atuando como o diploma que tem concretizado[393][394] as dimensões essenciais do direito fundamental autonomizado no artigo 8.º da Carta, a proteção de dados pessoais.

3.2. O contexto de aplicação do RGPD e os seus objetivos

O RGPD, também conhecido como Regulamento (UE) 2016/679, talvez seja o diploma jurídico europeu mais conhecido em todo o mundo. Sua notoriedade certamente ocorreu por causa da sua extraterritorialidade[395], uma vez que, de maneira geral, estipulou regras sobre proteção de dados que deveriam ser observadas por empresas de todo o mundo, independentemente de estarem ou não sediadas na UE. Para que a extraterritorialidade[396] ocorresse, era preciso que uma empresa não sediada em nenhum dos Estados-Membros (i) ofertasse bens ou serviços para pessoas localizada na União Europeia ou (ii) efetuasse algum tipo de controle de comportamento desses mesmos titulares de dados.

Contudo, independentemente do âmbito de aplicação territorial, ao abordar esse regulamento, é importante que se entenda as suas principais características e objetivos. Assim sendo, torna-se necessário perceber o motivo pelo qual as normas de proteção de dados foram adotadas através de um regulamento.

Nesse sentido, no que diz respeito às competências legislativas, a União Europeia pode adotar diretivas ou regulamentos[397]. De acordo com o que afirma Alessandra Silveira, a diferença existente entre esses dois atos jurídicos reside no fato de que "as diretivas apenas harmonizam as normas

[391] Cujos tratamento são realizados pelas instituições, órgãos e organismos da União, bem como pelos Estados-Membros no exercício de atividades relativas à aplicação do direito da União

[392] Silveira, Alessandra e Marques, João. Do direito a estar só ao direito ao esquecimento... p.92.

[393] Silveira, Alessandra e Froufe, Pedro. Do mercado interno à cidadania de direitos: a proteção de dados pessoais como a questão jusfundamental identitária dos nossos tempos. UNIO - EU LAW JOURNAL Vol. 4, No. 2, julho 2018, p.18.

[394] Silveira, Alessandra e Froufe, Pedro. Do mercado interno à cidadania de direitos... p.17.

[395] Cordeiro, A. Barreto Menezes. Direito da Proteção de Dados. Almedina, 2020... p. 95.

[396] Artigo 3.º, n.º 2, do RGPD

[397] Artigo 288.º do TFUE: "Para exercerem as competências da União, as instituições adotam regulamentos, diretivas, decisões, recomendações e pareceres".

aplicáveis nos distintos Estados-Membros da UE, enquanto os regulamentos uniformizam o direito aplicável num dado domínio, sem necessidade de intermediação legislativa das autoridades nacionais"[398].

Como se sabe, o Regulamento relativo à proteção de dados substituiu a antiga Diretiva 95/46, que possuía o mesmo objeto de proteção. Contudo, apesar da Diretiva 95/46 ter como finalidade promover a harmonização do domínio da proteção de dados entre os Estados-Membros, o que se percebeu foi uma verdadeira fragmentação decorrida da liberdade concedida a cada Estado-Membro ao realizar a sua respetiva transposição interna. A complexidade do cenário aumenta ao perceber que tal fragmentação poderia aumentar a insegurança jurídica relativa à proteção de dados pessoais, uma vez que cidadãos europeus teriam diferentes níveis de proteção de acordo com o Estado-Membro em que viesse a se encontrar. Com essa perspetiva em mente, a União Europeia optou pela substituição da antiga Diretiva pelo, então, Regulamento 2016/679.

Os motivos dessa substituição ficam mais claros ao se analisar os objetivos do RGPD, que, encontrados em seu artigo 1.º, preveem, respetivamente, (n.º 2) "a defesa dos direitos e das liberdades fundamentais das pessoas singulares, nomeadamente ao direito à proteção de dados pessoais"; como também (n.º 3) "a promoção e o desenvolvimento da livre circulação de dados pessoais no interior da União, desde que se respeite o tratamento de dados pessoais". Nesse sentido, Alessandra Silveira afirma que "não é propriamente árduo perceber que só uma proteção equivalente em todos os Estados-Membros, garantida por uma legislação inicialmente harmonizada (e agora tendencialmente uniformizada) [...], poderia assegurar a livre circulação de dados no mercado interno"[399].

Como se pode perceber, a opção por um regulamento fortalece tanto a economia de dados quanto a proteção dos próprios titulares. Isso porque assegura que não venha a existir qualquer "obstáculo ao exercício das atividades económicas a nível da União, "uma vez que não há como distorcer a concorrência e impedir as autoridades de cumprirem as obrigações que lhes incumbem por força do direito da União"[400].

Nesse contexto, percebe-se que o RGPD passa a ser considerado como um elemento central na promoção e no fortalecimento da proteção de dados, uma vez que ele "reflete as preocupações (agora, acrescidas) de conciliação da necessária competitividade e flexibilidade das empresas/agentes económicos europeus com o reforço da proteção efetiva dos direitos fundamentais"[401]. Além disso, ressalta-se que o equilíbrio promovido pelo RGPD segue o mesmo caminho pretendido pelo quadro ético para uma IA de Confiança. Isso porque, por mais que ambos reconheçam a contribuição da

[398] Silveira, Alessandra e Marques, João. Do direito a estar só ao direito ao esquecimento... p.93.
[399] Ibid.
[400] Considerando (9) do RGPD.
[401] Silveira, Alessandra e Froufe, Pedro. Do mercado interno à cidadania de direitos... p.9.

tecnologia para o progresso econômico e social[402], tanto o RGPD quanto o quadro para uma IA de Confiança entendem ser necessário que as novas tecnologias sejam desenvolvidas de forma responsável, a fim de que se concilie inovação tecnológica e efetiva proteção dos direitos fundamentais.

3.3. A justificativa sobre a escolha do RGPD para a regulação de uma IA de confiança

Como já vem sendo abordado, sistemas de IA podem provocar danos à indivíduos e à sociedade. Dentre tantos impactos que podem ser causados por sistemas de IA, os mais críticos estão relacionados a ilegítima intrusão da privacidade; o tratamento desleal e irregular de dados pessoais; o uso de dados enviesados que podem causar injustificada discriminação; e a opacidade dos sistemas de IA que não concedem a devida explicação sobre as decisões e os processos decisórios que impactam significativamente as pessoas.

Esse contexto torna-se ainda mais complexo uma vez que, como mencionado, não há a nível da União Europeia um diploma jurídico que regule especificamente a IA. Fato este que implica que qualquer análise jurídica sobre a IA seja realizada com base nos direitos fundamentais e em regulações setoriais, como dos diplomas voltados para a concorrência, defesa do consumidor ou proteção de dados[403].

Contudo, ao analisar a essência e os alicerces do desenvolvimento de um sistema inteligente, percebe-se que a IA, principalmente quando empregue através dos algoritmos de aprendizagem de máquina[404], é necessariamente dependente do tratamento de dados[405]. Ocorre que, muitas das vezes, esses dados tratados por um sistema de IA são de natureza pessoal[406], fato este que acaba legitimando originariamente a tutela do RGPD[407]. No entanto, observe que essa é a justificativa óbvia e estritamente legal para que o RGPD possa regular sistemas de IA.

Ocorre que o intuito dessa seção é o de demonstrar por qual motivo o RGPD pode ser aplicado como uma forma de regulação para que se alcance uma IA de Confiança. Nesse sentido, é preciso perceber as razões que levam o RGPD, e não outros diplomas, a se apresentar como o mais adequado

---

[402] Considerando (2) e (6) do RGPD.

[403] Sartor, Giovani. The impact of the General Data Protection Regulation (GDPR) on artificial intelligence. EPRS | European Parliamentary Research Service Scientific Foresight Unit (STOA) PE 641.530. 2020, p.7-8

[404] Regressar ao capítulo primeiro, item 1.2.4. para maior compreensão.

[405] Artigo 4.°, n.° 2, do RGPD: «Tratamento», uma operação ou um conjunto de operações efetuadas sobre dados pessoais ou sobre conjuntos de dados pessoais, por meios automatizados ou não automatizados, tais como a recolha, o registo, a organização, a estruturação, a conservação, a adaptação ou alteração, a recuperação, a consulta, a utilização, a divulgação por transmissão, difusão ou qualquer outra forma de disponibilização, a comparação ou interconexão, a limitação, o apagamento ou a destruição;

[406] Artigo 4.°, n.° 1, do RGPD: «Dados pessoais», informação relativa a uma pessoa singular identificada ou identificável («titular dos dados»); é considerada identificável uma pessoa singular que possa ser identificada, direta ou indiretamente, em especial por referência a um identificador, como por exemplo um nome, um número de identificação, dados de localização, identificadores por via eletrónica ou a um ou mais elementos específicos da identidade física, fisiológica, genética, mental, económica, cultural ou social dessa pessoa singular;

[407] Artigo 1.°, n.° 1, do RGPD: O presente regulamento estabelece as regras relativas à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados.

para lidar com os aspetos éticos da IA. Dentro dessa perspetiva, há um intrigante debate na literatura acadêmica.

Por um lado, há acadêmicos que afirmam que o RGPD "deve ser lido estritamente como um instrumento legal que regula dados pessoais"[408]. Caso esse posicionamento seja adotado, mesmo assim haveria coerência na tutela da IA através do RGPD. Isto porque o âmbito material de aplicação do RGPD (Art. 3.º n. º1) é devidamente identificado no funcionamento de um sistema de IA, como exemplificado acima.

Por outro lado, há acadêmicos que sugerem que o RGPD seja analisado através de uma leitura mais ampla e profunda. Esses acadêmicos acreditam que, além das normas de direito estritamente legais, o RGPD também incorpora valores éticos e princípios fundamentais. Nesse sentido, Hielke Hijmans e Charles D. Raab afirmam que, pela natureza do direito da proteção de dados, há uma relação estreita entre o RGPD e a ética[409]. Para fundamentar essa afirmação, eles argumentam que o direito fundamental à proteção de dados dá ao indivíduo a alegação de que seus dados estão sendo processados de maneira justa[410]. Além disso, eles ainda indicam que existem "outras noções de valor moral por trás da proteção de dados [como] a dignidade humana e a autonomia pessoal, que são noções com uma dimensão ética óbvia"[411].

De facto, os argumentos apresentados por essa segunda corrente possuem bastante respaldo caso se analise o RGPD a partir de um filtro ético. A primeira constatação que deve ser realizada é que o RGPD fundamenta a sua regulação através de princípios, onde o tratamento deve se guiar por valores de «licitude, lealdade e transparência»[412]. Além disso, o RGPD determina que "o tratamento dos dados pessoais deverá ser concebido para servir as pessoas"[413], sendo essa uma afirmação que possui estrita ligação com a dignidade humana. Por fim, o Considerando (75) demonstra uma preocupação com a gravidade dos riscos para os direitos e liberdades das pessoas singulares, o que é exemplificado pelo Considerando (71) e (85), onde claramente se identifica que o RGPD se guia por valores referentes à autonomia humana, liberdade do indivíduo, equidade, transparência e preocupações quanto a discriminação e desvantagem social.

Como se pode perceber, através de uma perspetiva estritamente legal, o RGPD tem total legitimidade para tutelar os sistemas de IA que realizam processamento de dados pessoais. Contudo, ao levar em consideração todo o seu alinhamento à valores éticos, percebe-se que o RGPD, além de legitimidade, também possui adequação para lidar com a busca pela ética na IA. Essa característica se

[408] Evas, Tatjana. European framework on ethical aspects of artificial intelligence, robotics and related technologies. European Parliament. 2019, p.6.
[409] Hijmans, H. e Raab, CD. Ethical Dimensions of the GDPR, in M. Cole and F. Boehm (eds.), Commentary on the General Data Protection Regulation, Edward Elgar, 2018, p.6.
[410] Ibid.
[411] Ibid.
[412] Artigo 5.º, n.º 1, a), do RGPD.
[413] Considerando (4) do RGPD.

torna muito importante, uma vez que, ao tratarmos da IA de Confiança, ficou evidenciada uma lacuna referente a sua respetiva competente «Legal». Assim sendo, e após as justificativas aqui apresentadas, percebe-se que há total pertinência na escolha pelo RGPD para que este seja analisado a fim de perceber em que medida ele pode contribuir para o desenvolvimento de uma IA de Confiança.

3.4. O RGPD como regulação de uma IA de confiança

A proposta dessa seção é a de analisar aspetos específicos do RGPD a fim de perceber o papel e os limites desse regulamento no desenvolvimento e regulação de uma IA de Confiança. Nesse sentido, alguns autores[414] indicam algumas disposições do RGPD como sendo as mais relevantes para o debate sobre ética e IA[415], sendo elas: a supervisão humana através dos princípios de tratamento[416]; o dever de fornecer informações[417] e acesso a dados[418]; o direito de explicação; a proibição de tomada de decisão individual automatizada[419]; e a avaliação do impacto da proteção de dados[420].

Isso posto, a partir desse momento esse trabalho irá focar-se sobre os aspetos acima mencionados, uma vez que seria irrazoável analisar a aplicação de cada um dos 99 dispositivos do RGPD a sistemas de IA. Com isso mente, é importante delimitar desde já o objetivo dessa pesquisa, pelo que não será analisado todos os direitos encontrados no RGPD, como também não se irá abordar a proteção de dados desde a conceção e por defeito. Pelo que a inclusão dessa componente pode ser interessante para um novo e futuro estudo.

3.4.1. Os princípios relativos ao tratamento de dados como vetores de uma IA de confiança

Um dos imperativos éticos que guiam uma IA de Confiança é o respeito pela autonomia humana. Como foi visto no capítulo anterior, esse princípio é uma consequência direta e natural dos direitos fundamentais do respeito da dignidade humana e da liberdade do indivíduo, direitos que possuem como núcleo central a defesa e a promoção da autonomia dos seres humanos. Em um contexto de IA, a proteção e a promoção da autonomia humana pressupõem o respeito à autodeterminação de todo o indivíduo que venha interagir com seus sistemas.

---

[414] Brkan, M. AI-supported decision-making under the general data protection regulation. Proceedings of the 16th edition of the International Conference on Artificial Intelligence and Law. 2017; M. Brkan, 'Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond', International Journal of Law and Information Technology, Vol. 27, 2019, pp. 91–121

[415] Evas, Tatjana. European framework on ethical aspects of artificial intelligence... p.6

[416] Artigo 5.° e o Considerando 71 do RGPD.

[417] Artigo 13.° e 14.° do RGPD

[418] Artigo 15.° do RGPD

[419] Artigo 22.° do RGPD

[420] Artigo 35.° do RGPD

Além disso, uma IA de Confiança também se baseia no princípio da equidade, que é um reflexo direto do respeito da dignidade humana e do direito fundamental à igualdade e a não discriminação[421]. Nesse sentido, esse imperativo ético reafirma o necessário compromisso contra os dados enviesados e as injustificadas decisões discriminatórias realizadas por sistemas de IA.

Assim sendo, com a perspetiva desses imperativos éticos em mente, e cientes dos objetivos que eles ambicionam, torna-se importante analisar os próprios princípios relativos ao tratamento de dados encontrados no RGPD. Essa análise permitirá entender em qual medida os princípios de tratamento de dados já estão de acordo (ou não) com uma IA de Confiança.

3.4.1.1. *«Licitude, Lealdade e Transparência»*

O artigo 5.º, n.º 1, a, do RGPD determina que o tratamento de dados pessoais deve seguir os princípios da licitude, lealdade e transparência em relação ao titular dos dados. Cada um desses princípios possui um significado e uma exigência diferente, como é possível conferir a seguir.

O significado da palavra licitude remete diretamente ao contexto de legalidade e de respeito às leis. Isto posto, abre-se a possibilidade de discussão de qual é o limite estabelecido pelo princípio da licitude apresentado pelo RGPD. Afirma-se isso pois tal princípio pode possuir dois sentidos[422].

Através de seu sentido estrito, o princípio da licitude pressupõe que cada tratamento em concreto de dados pessoais respeite e esteja fundamentado a uma das causas de licitude elencadas no artigo 6.º da RGPD. De acordo com esse artigo, que se refere diretamente as bases legais de tratamento de dados, existem seis possibilidades de fundamentação[423], sendo elas o consentimento[424], a defesa de interesses vitais, execução de contratos e de diligências pré-processuais, obrigações legais, interesse público e interesses legítimos. Por outro lado, a interpretação do princípio da licitude através de seu sentido amplo, implicaria na necessidade do responsável pelo tratamento respeitar e cumprir, não só o RGPD, mas também todas as demais legislações aplicáveis.

Apesar do segundo sentido ser mais favorável a uma perspetiva de desenvolvimento de uma IA de Confiança, pois o próprio RGPD já iria impor aos sistemas de IA[425] uma obrigação vinculativa de respeito a todas as leis, esse não é o sentido adotado em questão. Isso porque o RGPD adota o sentido estrito do princípio da licitude[426]. Tal interpretação também pode ser retirada do artigo 8.º (n. º2) da

---

[421] Artigo 2.° do TUE.
[422] Cordeiro, A. Barreto Menezes. Direito da Proteção de Dados... p. 152.
[423] Observe apenas que esses fundamentos legais não serão aprofundados nesse trabalho
[424] O consentimento possui diversas camadas e diferentes peculiaridades, devendo ser analisado a norma geral e as suas especificações quando referentes a crianças ou a dados sensíveis.
[425] Que tratam dados pessoais
[426] Conforme os artigos 6.°, 7.°/3, 13.°/2, c) ou 14.°/2, d); Kuner, Christopher, Bygrave, Lee A., et al. The EU General Data Regulation (GDPR) A Commentary. Oxford University Press 2020, p. 314-315.

Carta, que menciona como fundamento do tratamento o consentimento ou outro fundamento legítimo previsto por lei.

Quanto ao princípio da transparência, pode-se afirmar que esse atravessa, horizontalmente, todo o processo de tratamento de dados, o que inclui a coleta e o processamento de dados em si, mas não se exauri após o término da relação. Nesse contexto, a Information Commissioner's Office[427] (ICO) afirma que o princípio da transparência exige que o responsável pelo tratamento seja claro, aberto e honesto sobre quem ele é, como ele trata os dados pessoais e por quais motivos e finalidades ele faz isso[428]. Dessa forma, o responsável pelo tratamento deve assegurar aos titulares o acesso às informações relativas ao tratamento de dados, permitindo-lhes a oportunidade de tomarem decisões informadas. Nesse sentido, é importante destacar que a tomada de decisão informada é um pressuposto para a autonomia humana, uma vez que possibilita que o titular de dados mantenha o poder e a consciência sobre delegar ou não as suas decisões.

Contudo, no que se refere ao princípio da transparência no RGPD e o seu reflexo na IA de Confiança, é possível afirmar que o RGPD cria um nível introdutório de transparência a ser respeitado e cumprido pelos sistemas de IA que tratam dados pessoais. Afirma-se isso pois o princípio da transparência relativo ao tratamento de dados está mais voltado para a criação de uma relação baseada em informações honestas entre as partes, onde o titular poderá, de antemão, saber quais dados serão colhidos, por quem e para que finalidades[429]. Contudo, a ideia de transparência da IA de Confiança, além de absorver a conceção apresentada pelo RGPD, também pretende criar mecanismos em que as decisões e os processos decisórios possam ser auditados e explicados. Sendo essa, talvez, uma lacuna ainda presente no RGPD, mas que será analisada em uma seção posterior.

Por fim, no que se refere ao princípio da lealdade, o primeiro comentário que desse ser realizado refere-se a perspetiva linguística adotada pela tradução do RGPD em sua versão em português. Tal observação é importante tanto para um melhor entendimento do regulamento quanto, também, para uma melhor adequação aos conceitos de uma IA de Confiança. Isto posto, na versão portuguesa, espanhola e francesa, a expressão escolhida foi lealdade (*lealtad* e *loyauté*). Enquanto na versão inglesa o termo utilizado foi *fairness*, que de acordo com o dicionário Cambridge[430] significa justiça ou equidade. Além disso, é importante destacar que ao longo do RGPD[431] e de seus considerandos[432] a expressão 'princípios do tratamento equitativo' é a utilizada para se referir ao princípio da lealdade. Assim sendo, apesar da nomenclatura dada na versão portuguesa do RGPD, este

[427] Information Commissioner's Office (ICO). Guide to the General Data Regulation (GDPR). May 2019, p.22-23.
[428] Observe que todos esses aspectos informacionais relativos à transparência no tratamento de dados foram concretizados através da criação do direito de informação dos titulares de dados, que pode ser encontrado nos artigos 13.°, 14.° e 15.° do RGPD, que serão objeto de uma seção própria.
[429] Kuner, Christopher, Bygrave, Lee A., et al. The EU General Data Regulation (GDPR) A Commentary... p. 315.
[430] Significado de *fairness* através do dicionário Cambridge. Disponível em: https://dictionary.cambridge.org/pt/dicionario/ingles-portugues/fairness
[431] Artigo 13.°/2 e 40.°/2, a)
[432] Considerando (39), p.1; (60), p.1 e 2; e (71) p.5.

trabalho opta, a partir de agora, por atrelar o sentido do termo equidade, e as suas consequências jurídicas, ao princípio da lealdade.

Ultrapassada essa questão preliminar, o princípio da lealdade (ou equidade) impõe aos responsáveis pelo tratamento de dados "a obrigação de atenderem, a todo o tempo, aos interesses e as expectativas legítimas dos titulares de dados"[433]. Também deve ser referendado que o tratamento leal (ou justo) implica que os dados não foram obtidos nem processados de outra forma por meios injustos, por engano ou sem o conhecimento da pessoa em causa[434]. Além disso, pode-se incluir também que a lealdade e a equidade pressupõem que os dados pessoais jamais possam ser usados de forma a provocar efeitos adversos injustificados em seus titulares.

Nesse sentido, é importante adicionar ao debate a posição de Menezes que afirma que o princípio da lealdade e equidade "permite contestar determinados comportamentos que dificilmente poderiam ser descritos como violadores do artigo 6.º, [pois] trata-se de um conceito aberto, passível de ser invocado em situações que contradigam o espírito do RGPD"[435]. Isto posto, é importante destacar que esse princípio pode ser compreendido como um dos elementos mais fortes que o RGPD possuiu para lidar com casos de tratamento de dados que possam causar decisões injustas e discriminatórias, uma vez que violam o princípio da lealdade e equidade.

Em complemento à essas perspetivas, e através de uma análise contextualizada a sistemas de IA, Giovanni Sartor[436] apresenta duas novas dimensões do princípio da lealdade (equidade), sendo elas denominadas de (i) equidade informacional e (ii) equidade substancial. No que diz respeito a primeira, pode-se afirmar que a equidade informacional está estritamente ligada a ideia de transparência, uma vez que exige que "os titulares dos dados não sejam enganados ou induzidos em erro no que diz respeito ao tratamento dos seus dados"[437]. Esse, por sinal, é o mandamento encontrado no Considerando (60), que afirma que "os princípios do tratamento equitativo e transparente exigem que o titular dos dados seja informado da operação de tratamento de dados e das suas finalidades", onde "o titular dos dados deverá também ser informado da definição de perfis e das consequências que daí advêm".

Por outro lado, a segunda dimensão, referente a equidade substantiva, aborda a necessidade de respeito dos direitos e dos interesses do titular de dados quanto ao conteúdo de decisões automatizadas. Nesse sentido, o considerando (71), p.2, reafirma, de maneira geral, que o responsável pelo tratamento deverá utilizar as medidas técnicas e organizativas necessárias, assim como os procedimentos adequados à definição de perfis, para que seja garantida a correção de imprecisões nos

[433] Cordeiro, A. Barreto Menezes. Direito da Proteção de Dados... p. 154.
[434] Kuner, Christopher, Bygrave, Lee A., et al. The EU General Data Regulation (GDPR) A Commentary... p. 314.
[435] Cordeiro, A. Barreto Menezes. Direito da Proteção de Dados, p. 153.
[436] Sartor, Giovani. The impact of the General Data Protection Regulation (GDPR) on artificial intelligence... p.44-45.
[437] Sartor, Giovani. The impact of the General Data Protection Regulation (GDPR) on artificial intelligence... p.44.

dados pessoais. Além disso, o responsável pelo tratamento também deve proteger os dados pessoais, levando em consideração os potenciais riscos para os interesses e direitos do titular dos dados, a fim de prevenir, por exemplo, efeitos discriminatórios.

Como se pôde perceber, ao adicionarmos o valor da equidade ao princípio da lealdade, o RGPD apresenta maiores ferramentas para lidar com situações complexas que se encontram em uma zona cinzenta[438]. Uma dessas zonas refere-se a autonomia humana, uma vez que a equidade informacional deve presar por não condicionar ou enganar os titulares de dados. Por outro lado, também é possível afirmar que a equidade substancial pressupõe o combate a dados incorretos, e possivelmente enviesados, que possam promover decisões injustamente discriminatórias. Pelo que se pode entender que o RGPD, assim como o quadro para uma IA de Confiança respeitam e promovem os direitos fundamentais da liberdade do indivíduo e da igualdade e da não discriminação.

3.4.1.2. *«Limitação das Finalidades»*

O artigo 5.º, n.º 1, b), apresenta o princípio da limitação das finalidades, que tem sido considerado há muito tempo como "uma pedra angular da proteção de dados e um pré-requisito para a maioria dos outros requisitos fundamentais"[439]. O princípio da limitação da finalidade pode ser dividido em duas partes.

A primeira, impõe uma limitação na recolha e no tratamento de dados a finalidades (i) determinadas, (ii) explícitas e (iii) legítimas. Para que essa obrigação seja respeitada, as finalidades que fundamentam o tratamento de dados pessoais devem ser determinadas antes do processo de recolha de dados iniciar. Além disso, tais finalidades devem ser comunicadas de forma explícita aos titulares de dados pessoais. Por fim, ao referir-se a finalidades legítimas, esse princípio determina a necessária observação da licitude em sentido amplo, ou seja, devendo observar respeitar os fundamentos de licitude do tratamento (artigo 6.º) e todas as demais leis aplicáveis.

A segunda parte desse princípio veda o tratamento posterior de dados de forma incompatível com as finalidades originalmente indicadas[440]. Em outras palavras, o princípio da limitação das finalidades permite que os responsáveis pelo tratamento utilizem os mesmos dados anteriormente recolhidos para realizarem um novo tratamento com base em uma outra finalidade, que até o momento não havia sido comunicada ao titular. No entanto, essa permissão é restrita somente para tratamentos

[438] Uma zona que aparentemente está de acordo com o ordenamento jurídico, mas que pode apresentar indícios e consequências não éticas ou que possam causar algum risco.

[439] Kuner, Christopher, Bygrave, Lee A., et al. The EU General Data Regulation (GDPR) A Commentary... p. 315.

[440] O restante do dispositivo informa que "o tratamento posterior para fins de arquivo de interesse público, ou para fins de investigação científica ou histórica ou para fins estatísticos, não é considerado incompatível com as finalidades iniciais, em conformidade com o artigo 89.o, n.o 1"

que não sejam incompatíveis com o originário. Caso haja incompatibilidade entre o tratamento originário e o posterior, esse último será considerado ilegal.

Essa exigência possui duas perspetivas. Isso porque, numa perspetiva geral, acautela os interesses dos titulares, uma vez que lhe asseguram que os seus dados pessoais não sejam tratados com finalidades completamente distintas daquelas previamente expectáveis. Além disso, há uma "perspetiva específica, [que visa] garantir o controlo dos respetivos direitos à autodeterminação informacional"[441]. Essa é, porventura, uma importante afirmação, uma vez que o espírito do RGPD reside na devolução do poder sobre os dados pessoais aos seus titulares. E dentro desse contexto torna-se importante notar que a autodeterminação informacional, assim como o princípio da limitação das finalidades, compreende e fortalece a autonomia humana, característica essencial para uma IA de Confiança, uma vez que esses elementos contribuem para a ideia de transparência necessária em relações baseadas em dados, e consequentemente em sistemas de IA.

3.4.1.3. *«Minimização dos Dados»*

O artigo 5.º, n. º1, c), apresenta o princípio da minimização dos dados e que é intrinsecamente ligado ao princípio da limitação das finalidades. Afirma-se isso pois esse princípio determina que os dados coletados devem ser (i) adequados, (ii) pertinentes e limitados ao que é (iii) necessário relativamente às finalidades para as quais são tratados. Nesse sentido, a adequação determina que os dados coletados sejam relativos à própria finalidade do tratamento. A pertinência é medida pela capacidade que esses dados têm para contribuir com a finalidade. E a necessidade dos dados é demonstrada pela ausência de qualquer outro meio alternativo menos invasivo que poderia ser utilizado para alcançar o mesmo fim.

Como se percebe, esse princípio, assim como o da limitação de finalidade, transmite uma ideia ao titular de que os seus dados pessoais estão sendo tratados dentro de uma expectativa plausível. Essa situação possibilita gerar um sentimento de confiança junto ao responsável pelo tratamento, e por isso o respeito desses princípios é tão importante. Além disso, a minimização de dados também contribui para a autodeterminação informacional do titular, que consegue ter ciência de quais tipos de dados inferem determinado processo decisório. Esse é um sentimento esperado dentro de uma IA de confiança.

[441] Cordeiro, A. Barreto Menezes. Direito da Proteção de Dados, p. 155.

3.4.1.4. *«Exatidão»*

O artigo 5.º, n. º1, d), apresenta o princípio da exatidão no tratamento de dados pessoais. Nesse sentido, esse princípio determina que os dados pessoais sejam exatos e atualizados sempre que necessário, além de serem adotadas todas as medidas adequadas para que os dados inexatos, tendo em conta as finalidades para que são tratados, sejam apagados ou retificados sem demora.

Por mais que tenha relação com os princípios anteriores, a exatidão deve ter sua importância destaca quando aplicado em sistemas de IA. Afirma-se isso pois a inexatidão acerca dos dados pode propiciar decisões equivocadas, causando um impacto enorme na vida dos seus titulares. Aliás, os riscos causados por dados inexatos podem ser ainda piores quando usados como entrada em algoritmos de criação de perfil, uma vez que tais impactos podem ser replicados de forma prejudicial ao titular.

Tanto o RGPD quando o quadro para uma IA de Confiança usam o termo exatidão. Contudo, o RGPD o apresenta como um princípio relativo ao tratamento de dados, tendo seu sentido relacionado com a exatidão dos dados em si. Enquanto isso, o quadro para uma IA de Confiança apresenta a exatidão a partir do requisito de Solidez técnica e Segurança, onde se exige exatidão quanto a capacidade de realização de decisões corretas. Nesse diapasão, é de se presumir que, a existência de dados corretos é um requisito para interpretações e decisões também corretas. Dessa forma, por mais que o princípio da exatidão não aborde a decisão em si, pode-se entender que ele, ao menos, faz parte do processo decisório. E por conta disso, também se apresenta como um caminho necessário para uma IA de Confiança.

3.4.1.5. *«Limitação da Conservação»*

O artigo 5.º, n. º1, e), apresenta o princípio da limitação da conservação, que se encontra diretamente ligado à limitação da finalidade e indiretamente à exatidão. Nesse sentido, esse princípio proíbe a conservação de dados pessoais quando já não são necessários para efeitos de tratamento. No entanto, o armazenamento mais longo é permitido para fins de arquivamento, pesquisa ou estatísticas, respeitadas as medidas de segurança.

Trata-se, portanto, de mais um princípio que impõe fortes limites aos responsáveis pelo tratamento e dados pessoais. Como já mencionado anteriormente, essas restrições concretizam um direito fundamental da proteção dos dados e possibilitam uma criação de um sentimento de confiança relativo ao tratamento. E nesse sentido, apesar de não existir qualquer menção referente a essa limitação no quadro para uma IA de confiança, é importante destacar que tais limitações contribuem

para o exercício da autonomia humana, uma vez que limitam o poder e as possibilidades de decisão dos sistemas de IA.

3.4.1.6. *«Integridade e Confidencialidade»*

O artigo 5.º, n. º1, f), apresenta os princípios da integridade e da confidencialidade. Tais princípios, em conjunto, impõem a obrigação de garantia de que os dados sejam tratados de uma forma segura, o que deve incluir a confidencialidade dos dados no sentido de os proteger contra o tratamento ilícito ou não autorizado, ou ainda contra a sua perda, destruição ou danificação acidental.

Por mais que seja mencionada a necessidade de promover segurança, essa está relacionada à área da segurança da informação ou de sistemas informáticos, sendo, portanto, voltada para a área de governança. Isto posto, não se pode confundir esse princípio com as intenções de proteção física e mental ambicionadas pela IA de Confiança. Sendo, portanto, essa uma abordagem que se mostra carente no RGPD, e que talvez possa ser aperfeiçoada futuramente através de alterações legislativas.

3.4.1.7.*«Responsabilidade»*

Por fim, o artigo 5.º, n. º2, consagra o princípio da responsabilidade, que determina que o responsável pelo tratamento é responsável pelo cumprimento dos princípios apresentados no disposto no n.º 1, devendo, além disso, comprovar que o fez. Como se pode perceber, esse preceito apresenta duas obrigações distintas. A primeira refere-se à necessidade de o responsável pelo tratamento atuar sempre no estrito cumprimento dos princípios elencados no artigo 5.º, n. º1. A segunda obrigação pressupõe a primeira e determina que o responsável pelo tratamento demonstre o cumprimento desses princípios.

Nesse sentido, é importante ressaltar que o dever de comprovar o cumprimento dos princípios relativos ao tratamento de dados é um dever autônomo, e que, caso não cumprido, o responsável pelo tratamento será responsabilizado por isso[442][443]. Nesse sentido, o próprio Grupo de Trabalho do Artigo 29 (GT29) afirmou que "o objetivo desse dispositivo [é] reafirmar e fortalecer a responsabilidade dos controladores no processamento de dados pessoais. Isso sem prejuízo de medidas concretas de responsabilização que possam complementar esse princípio"[444].

[442] Cordeiro, A. Barreto Menezes. Direito da Proteção de Dados... p. 162.
[443] Kuner, Christopher, Bygrave, Lee A., et al. The EU General Data Regulation (GDPR) A Commentary... p. 319.
[444] Grupo de Trabalho do Artigo 29 (GT29), Opinion 3/2010 on the principle of accountability, p.8.

Destaca-se, ainda, que esse não é um princípio isolado no RGPD. Isso porque inúmeros são os preceitos[445] que surgem como forma de operacionalizar o princípio da responsabilidade ou apenas em decorrência de sua influência. Nesse sentido, pode-se citar explicitamente o artigo 24.º, n.º 1, que afirma que "tendo em conta a natureza, o âmbito, o contexto e as finalidades do tratamento dos dados, bem como os riscos para os direitos e liberdades das pessoas singulares, cuja probabilidade e gravidade podem ser variáveis, o responsável pelo tratamento aplica as medidas técnicas e organizativas que forem adequadas para assegurar e poder comprovar que o tratamento é realizado em conformidade com o presente regulamento. Essas medidas são revistas e atualizadas consoante as necessidades".

Como se pode perceber, por mais uma vez o RGPD reafirma a obrigação do responsável pelo tratamento em assegurar e comprovar o cumprimento do RGPD, notadamente em relação aos riscos para os direitos e liberdades das pessoas. Contudo, apesar da obrigação imposta, o RGPD não esclarece de que forma se pode demonstrar tal cumprimento, pelo que essa fica sendo uma decisão dos responsáveis pelo tratamento.

De uma maneira geral, no que se refere ao diálogo sobre proteção de dados e IA, é importante mencionar que o princípio da responsabilidade não encontra nenhum reflexo no quadro para uma IA de confiança. No entanto, tendo em vista que ele reforça a obrigatoriedade dos princípios relativos ao tratamento e, ainda, cria um dever relacionado à demonstração dessa comprovação, pode-se entender que o princípio da responsabilidade se apresenta como uma das mais importantes ferramentas do RGPD no que se refere ao desenvolvimento de uma IA de confiança. Afirma-se isso pois o princípio da responsabilidade se materializa como sendo um meio pelo qual sistemas de IA devem comprovar que respeitam os princípios relativos ao tratamento de dados. E tendo em consideração que ao longo dessa seção foi possível perceber que diversos princípios do RGPD dialogam direta ou indiretamente com alguns princípios éticos da IA de Confiança, significa dizer que um sistema de IA que respeita o artigo 5.º do RGPD, também respeita, parcial e consequentemente, os preceitos de uma IA de confiança.

3.4.2. O direito à informação no RGPD: um passo introdutório rumo a explicabilidade exigida pela IA de Confiança.

Como vem sendo demonstrado, a transparência é um fator essencial para tanto para o RGPD quanto para a IA de Confiança. Isso porque, em um primeiro momento, a transparência assegura o acesso à informação que permite que o titular de dados tome decisões informadas, o que contribui

[445] Artigos 24.º, n.º 1; 30.º e 33.º.

para o fortalecimento da autonomia humana. Essa, por sinal, é a proteção assegurada pelo princípio da transparência relativo ao tratamento de dados (artigo 5.º, n.º 1, a, do RGPD). Ocorre que a perspetiva de transparência para a IA de Confiança, além de pressupor uma relação clara e honesta entre titular e responsável, também requer que sejam desenvolvidas ferramentas que possibilitem a auditoria e, principalmente, a explicabilidade das decisões e processos decisórios realizados por sistemas de IA.

Por sinal, a transparência em uma IA de Confiança é uma consequência do princípio ético relativo à explicabilidade. Como abordado no capítulo anterior, a explicabilidade é um imperativo ético que tem estrita relação com os direitos fundamentais relativos à liberdade do indivíduo, à equidade, à justiça e, consequentemente, à dignidade da pessoa humana. De forma que a explicabilidade se apresenta como um imperativo que fortalece e complementa todos os demais.

Nesse sentido, essa seção tem o objetivo de entender quais são os limites objetivos do princípio da transparência dentro do RGPD. Para que isso seja realizado, será analisado o direito de informação dos titulares de dados, que pode ser encontrado nos artigos 13.º e 14.º do RGPD. Além disso, esses preceitos complementam-se quando conjugados com o artigo 15.º também do RGPD, que se refere ao direito de acesso aos dados. No que concerne a esses artigos, enquanto os dois primeiros artigos constituem o que pode se chamar de dever à informação ativo, o último é denominado como dever de informação passivo[446].

3.4.2.1. Dever de informação ativo: Direito à Informação (artigo 13.º e 14.º)

O dever de informação ativo é aquele encontrado nos artigos 13.º e 14.º do RGPD, sendo denominado dessa forma pois o responsável pelo tratamento é obrigado a prestar informações independentemente de o titular realizar uma requisição ou não. Esse dever também é conhecido como o direito à informação do titular de dados, que nas palavras de Joana Covelo de Abreu, deve ser considerado como o direito mais importante do RGPD, uma vez que "sem apropriada informação, o titular dos dados muito dificilmente poderia fazer uso de todos os outros direitos facultados pelo RGPD, diminuindo sobremaneira a eficácia do diploma"[447].

De maneira geral, é importante notar que ambos os dispositivos se referem às informações que os responsáveis pelo tratamento devem facultar aos titulares de dados, pelo que o conteúdo de ambos os artigos é parcialmente idêntico. Contudo, a diferença significativa fica a cargo de como o direito à informação deve ser apresentado a depender de como os dados pessoais foram recolhidos, ou

[446] Cordeiro, A. Barreto Menezes. Direito da Proteção de Dados... p. 340.

[447] Silveira, Alessandra, Abreu, Joana Covelo de e Cabral, Tiago Sérgio. Breves apontamentos quanto aos direitos dos titulares de dados no RGPD. CEJ - Centro de Estudos Judiciários. 2020, p. 2. "no prelo" que cita Tiago Sérgio Cabral, AI Regulation in the European Union: Democratic Trends, Current Instruments and Future Initiatives (Dissertação de Mestrado: Universidade do Minho, 2019), 131 e ss.

seja, se foram (13.º) ou não (14.º) recolhidos diretamente com o titular de dados. Essa diferença faz com que os artigos apresentem ligeiras exceções, mas que não modificam significativamente a essência do dever de informação ativo dos responsáveis pelo tratamento. Isto posto, tendo em vista uma perspetiva expositiva, a análise de ambos os artigos será baseada no núcleo central que esses dispositivos apresentam para determinar os deveres de informação do responsável

Assim sendo, destaca-se que os números 1 e 2 dos artigos 13.º e 14.º possuem razões e fundamentos idênticos[448], sendo eles (i) o reconhecimento e proteção do direito de autodeterminação informacional, fator essencial para uma autonomia humana; e (ii) a concretização do princípio da transparência.

Nesse sentido, independentemente dos dados serem ou não recolhidos diretamente com o seu titular, de acordo com os artigos 13.º/1 e 14.º/1, o responsável pelo tratamento tem o dever, de maneira geral, de informar: (i) a identidade e os contatos do responsável pelo tratamento e, se for caso disso, do seu representante; (ii) os contatos do encarregado da proteção de dados; (iii) as finalidades do tratamento a que os dados pessoais se destinam, bem como o fundamento jurídico para o tratamento; (iv) os destinatários ou categorias de destinatários dos dados pessoais, se os houver; e (v) a eventual intenção de transferir internacionalmente os dados pessoais. Todas essas informações contribuem e fortalecem a ideia de autodeterminação informacional, pois demonstram transparência sobre quem está a tratar os dados, por quais motivos e quais são as intenções gerais acerca desse tratamento.

Além das informações acima mencionadas, como forma de respeitar o princípio da transparência e de garantir um tratamento equitativo e transparente, de acordo com os artigo 13.º/2 e 14.º/2, o responsável pelo tratamento também tem o dever de informar ao titular de dados: (i) prazo de conservação dos dados pessoais ou, se não for possível, os critérios usados para definir esse prazo; (ii) os vários direitos que o RGPD reconhece aos titulares; (iii) o direito de retirar o consentimento; (iv) o direito de apresentar reclamações; e, por fim, (v) a "existência de decisões automatizadas, incluindo a definição de perfis referida no artigo 22.º, n.º 1 e 4, e, pelo menos nesses casos, informações úteis relativas à lógica subjacente, bem como a importância e as consequências previstas de tal tratamento para o titular dos dados"[449].

Todas essas informações são imprescindíveis para que o princípio da transparência seja alcançado. No entanto a última delas, referente às decisões automatizadas, é a que desperta maior interesse no momento, isso porque ela se refere diretamente aos sistemas de IA e sobre a possibilidade de o RGPD ter garantido aos titulares o direito a explicação de decisões automatizadas. Afirma-se isso

[448] Cordeiro, A. Barreto Menezes. Direito da Proteção de Dados... p. 340.
[449] 13.º, nº. 2, f) e 14.º, n.º 2, g)

pois, de facto, os artigos 13.º, nº. 2, f) e 14.º, n.º 2, g), mencionam, como um dever ativo do responsável pelo tratamento, a necessidade de apresentar aos titulares de dados informações úteis sobre a lógica das decisões e sobre as consequências do referido tratamento. Nesse sentido, caso essa interpretação seja passível de consubstanciar um direito à explicabilidade de decisões tomadas por sistemas de IA, o RGPD apresentaria mais uma ferramenta que o deixaria adequado a regular o desenvolvimento de sistemas de IA de confiança.

Contudo, antes de se alcançar qualquer conclusão, é importante que seja realizado maiores aprofundamentos sobre o direito de explicação referente a sistemas de IA. Nesse sentido, Sandra Wachter e Luciano Floridi explicam, em primeiro lugar, que existem dois significados para explicabilidade em sistemas de IA e que eles precisam ser diferenciados[450]. Isso porque, as explicações podem referir-se às (i) funcionalidades do sistema, onde a explicação recairá sobre a própria lógica e as funcionalidades gerais de um sistema automatizado; ou às (ii) decisões especificadamente, onde, nesse caso, as explicações teriam como finalidade apresentar a lógica, as razões e as circunstâncias individuais de cada decisão automatizada específica.

Além disso, as explicações podem se distinguir pelo tempo em relação ao processo de tomada de decisão. Nesse sentido, Wachter e Floridi apresentam também dois tipos de decisão[451], as (i) *ex ante* e as (ii) *ex post*. Uma explicação *ex ante* é aquela que ocorre antes de uma tomada de decisão automatizada. Isso posto, uma decisão *ex ante* somente pode fazer referência às funcionalidades de um sistema de IA. Por outro lado, uma explicação *ex post* é aquela que é apresentada após a realização de uma decisão automatizada. Assim sendo, uma explicação ex post pode referir-se tanto às funcionalidades de um sistema quanto à lógica e as razões individuais de uma decisão em si.

Tendo em consideração os entendimentos e as dimensões temporais referentes ao direito à explicação, Wachter e Floridi afirmam que os artigos 13.º, nº. 2, f) e 14.º, n.º 2, g) apenas exigem uma explicação *ex ante* dos sistemas de IA, que por sinal, devem somente referir-se à lógica, às funcionalidades e às consequências esperadas de um sistema de IA[452]. A constatação dos limites dessa exigência é, por sinal, coerente com o próprio dever de informar ativo do responsável pelo tratamento de dados, que deve sempre ser exercido no momento da recolha dos dados pessoais do titular. Nesse momento, é impossível para o responsável apresentar qualquer explicação sobre as razões das decisões automatizadas em si, uma vez que ainda não há qualquer tomada de decisão baseada nos dados do titular.

---

[450] Wachter, S., Mittelstadt, B., e Floridi, L. Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. International Data Privacy Law 7, 76–99, 2016 p. 6.

[451] Ibid.

Wachter, S., Mittelstadt, B., e Floridi, L. Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation... p.15.

Isso posto, os artigos 13.º e 14.º não podem, em hipótese alguma, serem utilizados como dispositivos jurídicos para a fundamentação da exigência de uma explicação *ex post* das razões de uma decisão específica, devendo, portanto, restringirem-se tão somente as informações sobre a funcionalidade do sistema[453]. Essa constatação, por mais lógica e acertada que se apresente, não é suficiente para alcançar o princípio da explicabilidade almejado pela IA de Confiança, que além de ansiar por processos transparentes, também busca decisões explicáveis, que como percebeu-se seriam caracterizadas como *ex post*.

3.4.2.2. Dever de informação passivo: Direito de acesso aos dados (artigo 15.º)

Diferente do anterior, o direito de acesso (artigo 15.º) é reconhecido por ser um dever de informação passivo, uma vez que o responsável pelo tratamento está obrigado a prestar informações na exata medida em que o respetivo direito for exercido pelo titular[454]. Além disso, esse direito fortalece a autodeterminação informacional na medida em que, mais uma vez, permite que o titular tenha ferramentas para controlar e gerir os seus dados pessoais. Contudo, não se pode deixar de comentar que o direito de acesso é um reflexo do direito fundamental da proteção de dados conferido pelo artigo 8.º da Carta, que afirma que "todas as pessoas têm o direito de aceder aos dados coligidos que lhes digam respeito".

Nesse sentido, pode-se afirmar que o direito de acesso se subdivide em duas dimensões. A primeira refere-se ao (i) direito de obter do responsável pelo tratamento a confirmação de que os dados pessoais que lhe digam respeito são ou não objeto de tratamento. Nesse sentido, o responsável deve responder objetivamente a essa requisição. Além disso, de acordo com Tribunal de Justiça da União Europeia (TJUE), o direito de confirmação deve abranger tanto o presente quando o passado[455]. Essa, por sinal, é uma conclusão importante, uma vez que se assim não o fosse, o titular de dados jamais estaria em condições de exercer a plenitude de seus direitos relativos à proteção de dados, o que mitigaria a autodeterminação informacional.

No que concerne à segunda dimensão do direito de acesso, parte-se do pressuposto de que já há um tratamento e, por isso, concede-se ao titular (ii) o direito de aceder aos seus dados pessoais, assim como demais informações relacionadas a ele e ao tratamento. Destaca-se que, no que se refere ao alcance desse direito, o artigo 15.º possibilita que o titular requeira as mesmas informações ora abordados pelos artigos 13.º e 14.º. E por conta dessa exata simetria, o direito de acesso do artigo 15.º

[453] Cabrral, Tiago Sérgio. AI Regulation in the European Union: Democratic Trends, Current Instruments and Future Initiatives (Dissertação de Mestrado: Universidade do Minho, 2019), p. 185.
[454] Cordeiro, A. Barreto Menezes. Direito da Proteção de Dados... p. 263.
[455] TJUE 7-mai.-2009, proc. C-553/07 (*Rijkeboer*), 54;

se apresenta como uma nova possibilidade de o titular de dados requerer informações a respeito de decisões automatizadas, sua lógica subjacente e suas consequências previstas[456].

Porém, em contrapartida com os artigos 13.º e 14.º, que definitivamente diziam respeito à um momento anterior ao processamento de dados, o artigo 15.º parece implicar que os dados já foram, de alguma forma, coletados e tratados. Afirma-se isso pois o direito de acesso tem como objeto "dados pessoais que [...] são ou não objeto de tratamento"[457]. Essa inferência enseja a possibilidade de que o artigo 15.º possa justificar a requisição de uma explicação *ex post*, que como mencionado, pode referir-se tanto às funcionalidades de um sistema de IA quanto a lógica e às razões de uma decisão específica.

No entanto, apesar da esperança sobre explicabilidade *ex post*, Wachter, Floridi e Giovanni Sartor, acreditam que seja razoável duvidar que o direito de acesso concede o direito de obter explicações acerca da lógica e das razões de uma decisão automatizada específica. A fim de sustentar essa posição eles utilizam a semântica do artigo 15.º, nº 1, h). Isso porque a frase 'consequências previstas' é orientada para o futuro, o que sugere que o responsável pelo tratamento "deva informar ao titular de dados sobre possíveis consequências da tomada de decisão automatizada antes que esse processamento ocorra"[458].

Além disso, ambiguidades[459] encontradas no artigo 15.º são usadas como justificativa para negar o dever dos responsáveis pelo tratamento em oferecer uma explicação *ex post* acerca das razoes de uma decisão automatizada. Isso porque, enquanto a frase 'lógica envolvida' remete para uma explicação *ex post*, as frases 'bem como a importância e as consequências previstas' remetem para uma explicação *ex ante*. Isso posto, Wachter e Floridi afirma que o artigo 15.º, n.º 1, h) deve ser lido e compreendido de forma coerente como um todo. Assim sendo, da mesma forma como acontece com os deveres de informação ativos dos artigos 13.º e 14.º, o direito de acesso do RGPD concede apenas uma explicação *ex ante*, ou seja, referente as funcionalidades do sistema, e não sobre a lógica e as razões das decisões específicas. Isso posto, conclui-se que o princípio da transparência relativo ao tratamento de dados encontra como limite explicações apenas referentes a funcionalidades do sistema de IA.

Como se pode perceber, a fim de respeitar o princípio da transparência e construir uma relação clara e honesta entre titular e responsável pelo tratamento, o RGPD determina que o titular de dados seja informado de diversos aspetos relativos ao tratamento de dados. Entretendo, no que concerne às

---

[456] Artigo 15.º, n.º 1, h), do RGPD: A existência de decisões automatizadas, incluindo a definição de perfis, referida no artigo 22. o, n.º 1 e 4, e, pelo menos nesses casos, informações úteis relativas à lógica subjacente, bem como a importância e as consequências previstas de tal tratamento para o titular dos dados.

[457] Artigo 15.º, n.º 1, do RGPD

[458] Wachter, S., Mittelstadt, B., e Floridi, L. Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation... p.17.

[459] Sartor, Giovani. The impact of the General Data Protection Regulation (GDPR) on artificial intelligence... p. 54-55

tomadas de decisão automatizadas, o RGPD e o seu princípio da transparência limitam-se a apenas oferecer informações sobre as funcionalidades e a lógica técnica e esperada de um sistema de IA. De facto, não se pode negar que esse é um avanço se comparado a diversas outras legislações. Dessa forma, é muito importante destacar a adequação e o papel que o RGPD pode apresentar ao regular e desenvolver uma IA de confiança, e muito disso ocorre similaridades e congruências entre os dois diplomas. No entanto, deve-se ao mesmo tempo reconhecer que o RGPD possui limitações cruciais que o impede de executar o ideal europeu de um quadro ético para uma IA de confiança.

A partir desse entendimento, deve-se concluir que as proteções asseguradas pelo RGPD a partir do princípio da transparência podem ser consideradas como um 'nível introdutório' à uma IA de confiança, pelo que indica uma necessária adaptação ou evolução. A respeito disso, há então três caminhos que podem ser tomados. O primeiro caminho parte da academia e da mudança de posição doutrinária a respeito de um direito de explicação *ex post* das razões de uma decisão automatizada. Uma segunda opção necessitaria da provocação do TJUE sobre esse tema e do consequente reconhecimento do direito à explicação a partir do RGPD. A terceira e última opção implicaria uma alteração legislativa a fim de promover uma redação mais clara e objetiva acerca do presente tema.

3.4.3. O RGPD e a regulação de decisões automatizadas

Tendo em vista que o objetivo dessa pesquisa tem sido analisar a aplicação do RGPD aos sistemas de IA, e, consequentemente, a sua adequação a uma IA de confiança, torna-se imprescindível que se aborde o artigo 22.º do presente regulamento, uma vez que esse trata especificadamente da tomada de decisões automatizadas, uma referência direta à IA.

3.4.3.1. Uma proibição geral acerca das decisões exclusivamente automatizadas

O artigo 22.º, n. º1, afirma que "o titular dos dados tem o direito de não ficar sujeito a nenhuma decisão tomada exclusivamente com base no tratamento automatizado, incluindo a definição de perfis, que produza efeitos na sua esfera jurídica ou que o afete significativamente de forma similar". A partir desse dispositivo, há uma observação preliminar que deve ser realizada. Isso porque, de acordo com Mendoza e Bygrave[460], apesar do RGPD utilizar o termo 'direito', a sua real interpretação

[460] Mendoza, Isak e Bygrave, Lee A. 'The Right not to be Subject to Automated Decisions based on Profiling' University of Oslo Faculty of Law Legal Studies Research Paper Series No 20/2017(n 31), p.9.

deve ser realizada no sentido de que não é necessário que o titular de dados o exerça ativamente para que receba a respetiva proteção. Essa posição também é adotada pelo antigo GT29[461].

Assim sendo, o artigo 22.º, n.º 1, deve ser interpretado como uma verdadeira proibição geral relativa à tomada de decisões exclusivamente automatizadas. O que, de acordo com o GT 29, significa dizer que as "as pessoas estão automaticamente protegidas dos possíveis efeitos deste tipo de tratamento"[462]. Contudo, por mais que esse dispositivo aparente ser demasiadamente contundente, perceba que para que haja a proibição geral é necessária a presença de quatro condições, sendo elas: (i) uma decisão a ser tomada, (ii) baseada exclusivamente no processamento automatizado, (iii) através da utilização de perfis, (iv) e que possua algum efeito legal ou de afete significativamente o titular de dados.

Nesse sentido, qualquer decisão que, de alguma forma, sofra uma influência real ou significativa de um ser humano não será considerada como exclusivamente automatizada[463]. Dessa forma, a influência significativa humana afasta a aplicação do artigo 22.º dos sistemas de IA. Contudo, se por um lado pode-se imaginar que essa pode ser uma saída a ser adotada pelas empresas para que não respondam pelo artigo 22.º do RGPD, por outro lado pode-se perceber o fortalecimento do controle e da supervisão humana no processo de tomada de decisão automatizada, uma vez que o RGPD apresenta uma premissa de que há uma preferência pela participação humana nas tomadas de decisão. Caso isso de fato aconteça é importante reconhecer o fortalecimento da autonomia humana no que diz respeito aos sistemas de IA, e, consequentemente, de uma aproximação aos desígnios de uma IA de confiança.

Contudo, além de ausência de intervenção humana em um processo de tomada de decisão automatizado, para que a proibição geral seja aplicada também será necessária a existência de perfis. Essa exigência é mencionada pelo Considerando (71), que ao tratar sobre o direito em questão afirma que "esse tratamento inclui a definição de perfis mediante qualquer forma de tratamento automatizado de dados pessoais para avaliar aspetos pessoais relativos a uma pessoa singular".

Por fim, não se pode esquecer que a proibição geral também só é aplicada caso a decisão tomada produza efeitos legais ou afete de forma significativa o titular de dados. Nesse sentido, o GT29 afirma que "para que um tratamento de dados afete significativamente alguém, os seus efeitos devem ser suficientemente grandes ou importantes para merecerem atenção"[464], como afetar

[461] Grupo de Trabalho do Artigo 29 (GT29). Orientações sobre as decisões individuais automatizadas e a definição de perfis para efeitos do Regulamento (UE) 2016/679, fevereiro de 2018, p.21.
[462] GT29. Orientações sobre as decisões individuais automatizadas... p.22.
[463] De acordo com o GT 29: "Para que se considere haver uma intervenção humana, o responsável pelo tratamento tem de garantir que qualquer supervisão da decisão seja relevante, e não um mero gesto simbólico. Essa supervisão deve ser levada a cabo por alguém com autoridade e competência para alterar a decisão e que, no âmbito da análise, deverá tomar em consideração todos os dados pertinentes".
[464] GT29. Orientações sobre as decisões individuais automatizadas... p.24.

significativamente o comportamento ou as escolhas de alguém; provocar um impacto prolongado ou permanente; ou ocasionar alguma espécie de discriminação.

Até esse momento, pode-se perceber que o artigo 22.º possui diversas preocupações com fatores diretamente ligados à equidade, autonomia e liberdade dos indivíduos, uma vez que cria mecanismos que visam proteger a autonomia humana, seja pelo 'incentivo' da participação de pessoas dentro do processo decisório, seja pela posição contrária à sistemas de IA que constroem perfis psicológicos intrusivos para influenciar ou condicionar o comportamento de consumidores ou eleitores.

3.4.3.2. Exceções que permitem a tomada de dados exclusivamente automatizada

Apesar da rigidez da proibição determinada pelo RGPD, o artigo 22.º, n.º 2, apresenta três específicas exceções que permitirão ao responsável proceder na tomada de decisões exclusivamente automatizadas.

A primeiras das exceções ocorre quando a tomada de decisão exclusivamente automatizada "for necessária para a celebração ou a execução de um contrato entre o titular dos dados e um responsável pelo tratamento"[465]. Nesse sentido, é importante destacar que não basta, apenas, que tais tomadas de decisão sejam realizadas dentro do escopo de execução de um contrato, mas sim que a decisão exclusivamente automatizada se prove necessária para que o contrato seja celebrado ou executado[466]. Nesse sentido, "o responsável pelo tratamento deve ser capaz de demonstrar que esse tipo de tratamento é necessário, avaliando se seria possível adotar um método menos intrusivo para a privacidade"[467]. Como exemplo, o GT29 cita um sistema de recrutamento que deve lidar com a análise de milhares de candidaturas em um curto espaço de tempo. Contudo, apesar da opção exemplificativa adotada pelo GT29, deve-se destacar que o RGPD por diversas vezes se posiciona contrário ao tratamento automatizado em área laborais, uma vez que os riscos de discriminação e de impacto negativo são extremamente altos.

A segunda exceção depende de autorização pelo direito da União ou do Estado-Membro a que o responsável pelo tratamento estiver sujeito[468]. Aqui não há muito o que se discutir, pois é necessária expressa previsão legal para que o responsável possa se eximir da proibição geral. Contudo, é importante destacar que a lei ou o ato normativo apresentado deve prever medidas adequadas para salvaguardar os direitos e liberdades e os legítimos interesses do titular dos dados. Sobre as referidas salvaguardas, estas serão analisadas como um todo na seção seguinte.

---

[465] Artigo 22.º, n.º 2, a), do RGPD
[466] GT29. Orientações sobre as decisões individuais automatizadas... p.25-26.
[467] GT29. Orientações sobre as decisões individuais automatizadas... p.26.
[468] Artigo 22.º, n.º 2, b), do RGPD

A última exceção permite a tomada de decisão baseada em tratamento exclusivamente automatizado quando houver consentimento expresso do titular de dados[469]. Perceba que não se trata apenas do consentimento do titular de dados, devendo esse ato ser qualificadamente explicito. O RGPD faz essa distinção quando percebe ser necessário dar maior proteção ao titular, uma vez que determinados tipos de tratamento, como o em questão, podem provocar sérios riscos de proteção de dados[470]. Além disso, essa proteção, como se pode perceber, se reverte em um maior poder de controle individual sobre os próprios dados pessoais, fato este que confere maior autonomia para o titular de dados.

Contudo, ainda no que se refere às exceções, é importante destacar que, de acordo com o n.º 4 do artigo 22.º, as decisões automatizadas que envolvam categorias especiais de dados pessoais (artigo 9.º, n. º1), somente serão permitidas se houver o consentimento explícito do titular de dados[471] ou se o tratamento for necessário por motivos de interesse público, devendo, então, respeitar a essência do direito à proteção dos dados pessoais e ser devidamente proporcional ao objetivo visado[472]. Como se pode perceber, ciente de que as decisões automatizadas, por si, já podem provocar elevado risco, o RGPD passa a ser ainda mais rigoroso ao passo que o tratamento inclui dados de categoria especial.

E essa é uma conclusão geral que se pode concluir do artigo 22.º, isso porque, por entender os riscos envolvidos no processo de tomada de decisões exclusivamente automatizadas, para cada exceção à proibição geral de tratamento, o RGPD passa a exigir níveis proporcionais de requisitos a serem observados pelos responsáveis pelo tratamento. Essa posição demonstra uma preocupação gradativa com a proteção dos titulares de dados, e, ao mesmo, tempo, se apresenta como uma forma de não obstruir de forma irrazoável o desenvolvimento tecnológico.

3.4.3.3. As salvaguardas e os direitos relativos às tomadas de decisão exclusivamente automatizadas

Contudo, as proteções do RGPD não se limitam a somente exigências de maiores requisitos de licitude no tratamento, pelo que também há casos em que o regulamento passa a impor a implementação de salvaguardas de proteção aos direitos e interesses do titular. Nesse sentido, e de acordo com o artigo 22.º, n.º 3, sempre que a tomada de decisão se fundamente juridicamente nas alíneas a) e c) do artigo 22.º, n.º 2, ou seja, na execução de um contrato ou no consentimento

[469] Artigo 22.º, n.º 2, c), do RGPD
[470] GT29. Orientações sobre as decisões individuais automatizadas... p.26.
[471] Artigo 9.º, n.º 1, a), do RGPD
[472] Artigo 9.º, n.º 1, g), do RGPD: "...com base no direito da União ou de um Estado-Membro, deve ser proporcional ao objetivo visado, respeitar a essência do direito à proteção dos dados pessoais e prever medidas adequadas e específicas que salvaguardem os direitos fundamentais e os interesses do titular dos dados".

explícito, o responsável pelo tratamento deve "aplica[r] medidas adequadas para salvaguardar os [a] direitos e liberdades e [b] legítimos interesses do titular dos dados, designadamente o direito de, pelo menos, [i] obter intervenção humana por parte do responsável, [ii] manifestar o seu ponto de vista e [iii] contestar a decisão".

Da leitura desse dispositivo, a primeira conclusão que se pode tirar é que, de forma indireta, o princípio da responsabilidade exerce sua função no sentido de determinar que os (a) direitos e liberdades do titular de dados devem ser respeitados e protegidos. Isto posto, apesar do princípio da licitude no tratamento (alínea a do artigo 5.º, n.º 1) ser interpretado de forma estrita[473], ao fazer menção à direitos e liberdades, o artigo 22.º, n.º 3, passa a impressão de que exige uma proteção ampla, respeitando o RGPD, mas também as demais legislações e, principalmente, os direitos fundamentais. Esse entendimento pode fazer com que o RGPD seja utilizado como fundamento jurídico para que se exija maior respeito e proteção no tocante aos riscos voltados à autonomia, liberdade, equidade e não discriminação, apresentando assim ferramentas jurídicas para materializar uma IA de confiança.

Além dessa interpretação, é também necessário destacar que o artigo 22.º, nº 3, também exige medidas para salvaguardar os (b) legítimos interesses do titular de dados. Isso mais uma vez pressupõe que as finalidades do tratamento não podem, de forma alguma, serem prejudiciais aos titulares. Assim sendo, os dados pessoais não podem ser tratados de forma a constituir produtos e serviços que contrariem a vontade legítima do titular. A partir dessa constatação, o RGPD também passa a fortalecer a proteção da autonomia humana, pelo que impõe barreiras, ao menos em teoria, para o desenvolvimento e utilização de sistemas de IA que possam discriminar injustamente as pessoas ou condicionar a opinião e o comportamento dos indivíduos.

Dentro desse contexto, e de forma a fortalecer as interpretações apresentadas, Maja Brkan afirma que as referidas salvaguardas têm o objetivo de "evitar uma decisão errada ou discriminatória ou uma decisão que não respeite os direitos e interesses do sujeito dos dados"[474]. Nesse sentido, o GT29 apresenta inúmeras medidas técnicas e organizativas, como sugestões de boas práticas, que podem ajudar a cumprir esse desígnio, como: controlos periódicos de garantia da qualidade dos respetivos sistemas a fim de assegurar um tratamento equitativo e não discriminatório das pessoas; controlo de algoritmos realizado internamente ou por terceiros; medidas específicas de minimização dos dados; recurso a técnicas de anonimização ou pseudonimização; além de também explorarem mecanismos de certificação, códigos de conduta ou painéis de exame ético[475].

[473] Uma vez que determina que o respeito à tão somente o RGPD e os requisitos de tratamento lícito (artigo 6.º)

[474] Brkan, Maja. Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond. International Journal of Law and Information Technology, 2019, p.17.

[475] GT29. Orientações sobre as decisões individuais automatizadas... p.36-37.

Contudo, o artigo 22.º, n.º 3, não se limita a somente indicar a necessidade de salvaguardas, isso porque ele também faz questão de designar direitos que o titular de dados pode exercer no que concerne à tomada de decisões exclusivamente automatizadas, como o direito de (i) obter intervenção humana por parte do responsável; (ii) manifestar o seu ponto de vista; e (iii) contestar a decisão.

No que concerne ao direito de obter intervenção humana, de acordo com Maja Brkan, o titular de dados poderá solicitar que "a decisão totalmente automatizada se torne não automatizada por meio de intervenção humana"[476]. Isso não quer dizer que a decisão passará a ser totalmente manual e que todo o processo decisório será conduzido por um ser humano, mas sim que haverá, de alguma forma, a presença de um responsável que participará significativamente na tomada de decisão. Ocorre que, infelizmente, o RGPD não apresenta nenhuma orientação de como isso deverá proceder. Porém, o GT29 afirma que o processo de revisão da decisão deve ser realizado por uma pessoa que tenha autoridade e competência para alterar a decisão, devendo, ainda, avaliar exaustivamente todos os aspetos que influenciaram de forma pertinente a decisão[477].

Isto posto, apesar dos alertas sobre a complexibilidade, e até mesmo a impossibilidade, de uma pessoa revisar uma decisão automatizada baseada em dados pessoais e *big data*, é importante destacar que o direito de obter uma intervenção humana no processo decisório é, de facto, algo legalmente apropriado e socialmente desejável[478], sendo está uma abordagem estritamente compatível como que se espera da promoção do controle e da supervisão humana em uma IA de confiança.

Além da referida intervenção humana no processo decisório, o artigo 22.º, n.º 3, também menciona o direito do titular de dados em expressar o seu ponto de vista. Nesse sentido, por mais que o RGPD não apresente de forma clara qual é a devida consequência jurídica desse direito, a doutrina entende que a manifestação de opinião do titular deve ser levada em consideração em conjunto com o direito de uma intervenção humana, devendo o responsável dar uma resposta sobre o ponto de vista do titular[479].

Contudo, além desses dois direitos, o RGPD também faculta ao titular de dados o direito de contestar a decisão. Como se pode perceber, essa é uma tríade de direitos que se complementam em busca de efetividade e proteção do titular de dados, pelo que torna o procedimento de tomada de decisão em uma espécie de contraditório, onde a decisão automatizada tomada exclusivamente por um sistema de IA é contestada pelo próprio titular de dados. Mais uma vez o RGPD não apresenta qualquer informação que possa clarificar o que deve acontecer, e nesse sentido a doutrina entende que o responsável pela resposta deverá ser o responsável pelo tratamento, e não um outro funcionário da

[476] Brkan, Maja. Do algorithms rule the world?... p.38.
[477] GT29. Orientações sobre as decisões individuais automatizadas... p.30.
[478] Brkan, Maja. Do algorithms rule the world?... p.38.
[479] Ibid.

empresa[480]. Isto posto, por mais que não haja qualquer menção no artigo 22.º, nº 3, esgotados os direitos que o titular pode exercer junto ao responsável pelo tratamento, ele sempre poderá direcionar-se à autoridade de proteção de dados e realizar uma reclamação.

Como se pode perceber, tais direitos apontados pelo legislador europeu mais uma vez estabelecem o RGPD como um instrumento jurídico que visa o fortalecimento da autonomia humana frente os avanços do poder computacional. Isso se prova não só pelos princípios de tratamento que enfatizam esses ideais, mas também por direitos e deveres que o responsável pelo tratamento deve respeitar, como os acima mencionados e as salvaguardas adequadas para proteger os direitos, as liberdades e os legítimos interesses dos titulares. Percebe-se então que o RGPD apresenta uma complexa organização de dispositivos jurídicos que fortalecem o poder de controle e participação do titular no processo decisório, o que confere cada vez mais importância para a autonomia e autodeterminação informacional, características importantes para o ponto de vista de uma IA de confiança.

3.4.4. A responsabilidade em assegurar e comprovar a conformidade do tratamento de dados com o RGPD como uma forma de iniciar a operacionalização de uma IA de confiança

Assim como o princípio da transparência que deu origem aos direitos de informação e de acesso, o RGPD também manifestou outros princípios em dispositivos jurídicos. Esse é o caso do princípio da responsabilidade (Artigo 5.º, n.º 2), que a partir do artigo 24.º, n.º 1, reforça a obrigatoriedade do responsável pelo tratamento em aplicar as medidas técnicas e organizativas que forem adequadas para assegurar e poder comprovar que o tratamento é realizado em conformidade com o RGPD.

No entanto, o RGPD não prevê qualquer definição para as mencionadas 'medidas técnicas e organizativas', como tampouco fornece informações suficientes para entender internamente a diferença entre cada uma dessas medidas[481]. Nesse sentido, de acordo com a doutrina, esses termos, em conjunto, devem primeiramente ser interpretados de forma ampla, pelo que "medidas técnicas e organizativas abrangem todos os aspetos que possam influenciar o cumprimento do RGPD ou prejudicar os direitos dos titulares de dados"[482]. No entanto, através de uma análise individualizada, enquanto as medidas técnicas devem respeitar os procedimentos aplicados na realização do tratamento, as medidas organizativas fazem alusão à forma como o responsável e os próprios tratamentos são estruturados.

---

[480] Ibid.

[481] Kuner, Christopher, Bygrave, Lee A., et al. The EU General Data Regulation (GDPR) A Commentary... p. 562.

[482] Cordeiro, A. Barreto Menezes. Direito da Proteção de Dados, p. 321.

Isto posto, é importante destacar que, para que essa obrigação seja cumprida da melhor forma possível, o artigo 24.º, n. º1, determina que o responsável pelo tratamento leve em consideração "a natureza, o âmbito, o contexto e as finalidades do tratamento dos dados, bem como os riscos para os direitos e liberdades das pessoas singulares, cuja probabilidade e gravidade podem ser variáveis". Nesse sentido, é importante notar que a utilização de sistemas de IA para tratar dados de forma automatizada deve ser considerada como um contexto de tratamento, e que, a depender da finalidade, sistemas de IA podem provocar riscos elevados aos titulares de dados.

Nesse contexto, antes mesmo de se dar início ao projeto ou iniciativa, é importante que o responsável pelo tratamento avalie os riscos, os impactos e as consequências possíveis que o tratamento de dados possa provocar aos titulares. Apenas essa análise permitirá que o responsável esteja apto a aplicar as medidas técnicas e organizativas adequadas ao caso concreto. Como se pode perceber, o RGPD exige do responsável uma abordagem preventiva aos possíveis riscos. Isso ocorre porque o próprio direito fundamental da proteção de dados pessoais requer esse tipo de abordagem jurídica, tendo ela muita influência no desenvolvimento do próprio RGPD. A doutrina nomeia essa influência como uma abordagem baseada no risco (ou *risk based approach*), e que se difere da tradicional abordagem baseada em direitos.

Isso posto, antes que esse trabalho chegue ao seu objetivo principal, essa seção irá se iniciar através do entendimento do que é uma abordagem baseada em risco encontrada no RGPD. Isso é importante na medida em que é esse o tipo de abordagem que determina como que o regulamento deve ser interpretado e utilizado pelos responsáveis pelo tratamento e pelos operadores do direito.

Ultrapassado esse momento, irá se alcançar o objetivo principal dessa seção, que será, primeiramente, analisar a utilização da avaliação de impacto de proteção de dados (AIPD) para demonstrar conformidade de sistemas de IA com o RGPD. Isso é importante pois, tal análise, possibilitará perceber em que medida a utilização da AIPD, ao demonstrar o respeito aos princípios e aos direitos do RGPD, pode auxiliar na regulação e comprovação de uma IA de confiança.

3.4.4.1. O RGPD e a abordagem baseada em risco em um contexto de IA

A fim de assegurar o respeito ao direito fundamental da proteção de dados, o RGPD foi desenvolvido a partir de duas perspetivas jurídicas diferentes, mas que são complementares entre si. Afirma-se isso pois, embora o RGPD tenha sido desenvolvido a partir de uma abordagem baseada em

direitos, há nele também uma enorme influência do que a doutrina denomina de uma abordagem baseada no risco[483].

De acordo com Giovanni Sartor, a abordagem baseada em direitos, principalmente voltada para a proteção de dados, tem como foco os direitos individuais, dividindo-os em duas camadas[484]. A primeira camada, chamada de superior, refere-se ao direito fundamental à proteção de dados, direitos esses que estão sinergicamente conectados com os demais direitos fundamentais, como a dignidade da pessoa humana, a liberdade do indivíduo, a equidade e a não discriminação. No que se refere à camada inferior, essa é constituída pelos próprios direitos relativos à proteção de dados concedidos pelo RGPD[485]. Dessa forma, o que se pode concluir é que uma abordagem baseada em direitos é aquela que se preocupa com os danos causados aos indivíduos e, consequentemente, nas medidas judiciais relacionadas à proteção dos direitos.

Por outro lado, em vez que conceder direitos individuais, uma abordagem baseada em risco, "centra-se na criação de uma ecologia sustentável da informação, onde os danos são evitados através de medidas organizacionais e tecnológicas adequadas"[486]. Em outras palavras, trata-se de uma abordagem baseada na regulação de riscos através de medidas preventivas, assim como acontece nas áreas de proteção ambiental, segurança médica e alimentar e de mercados financeiros.

Ainda no que concerne à uma abordagem baseada em riscos, é importante apresentar suas múltiplas dimensões, sendo elas a clássica, a econômica e a cientifico-ideológica. A aceção clássica é aquela em que "o responsável pelo tratamento deve, a todo o tempo e em relação a todas as decisões que tome e aos atos que pratique, atender aos riscos que daí possam decorrer para os titulares de dados"[487], devendo o responsável pelo tratamento, dessa forma, adequar a sua atuação às especificidades de cada caso concreto. No que diz respeito à aceção econômica, de acordo com Raphael Gellert, uma abordagem baseada no risco se apresenta como um modelo a ser empregue pelo responsável pelo tratamento a fim de verificar em que medida um determinado tratamento em concreto viola o RGPD, quantificando também o risco dessa violação[488]. Por fim, no que se refere à aceção científico-ideológica, o termo 'abordagem baseada em risco' transmite uma ideia de que se deixa de lado os direitos dos titulares de dados para dar maior enfoque nos impactos e nas consequências efetivas que o tratamento de dados possa provocar[489]. Tendo em vista os objetivos desse trabalho

[483] Information commissioner's Office (ICO). Guidance on AI and Data Protection. Version 0.0.22. July 2020, p.66.
[484] Ibid.
[485] Ibid.
[486] Ibid.
[487] Cordeiro, A. Barreto Menezes. Direito da Proteção de Dados... p.317.
[488] Gellert, Raphaël. We Have Always Managed Risks in Data Protection Law: Understanding the Similarities and Differences Between the Right-Based and the Risk-Based Approaches to Data Protection, 2 EDPL, 2016, 481-492.
[489] Cordeiro, A. Barreto Menezes. Direito da Proteção de Dados... p.317.

acadêmico, a aceção clássica será a adotada para a análise da aplicação do RGPD, notadamente a respeito das responsabilidades do responsável pelo tratamento.

Nesse sentido, ao analisar o RGPD e a abordagem baseada em risco em um contexto relacionado à IA, torna-se necessário que o responsável pelo tratamento leve em consideração os aspetos potencializadores do tratamento automatizado no momento de avaliar os riscos para os direitos fundamentais dos indivíduos[490]. Afirma-se isso pois os sistemas de IA podem acentuar os riscos existentes, introduzir novos, ou ainda tornar os riscos mais difíceis de avaliar ou gerenciar.

Assim sendo, ao avaliar os riscos na proteção de dados que podem ser causados por sistemas de IA, é importante que, assim como determina o Considerando (76), leve-se em consideração que a probabilidade e a gravidade dos riscos para os direitos e liberdades dos titulares sofrem influência direta e proporcional da natureza, do âmbito, do contexto e das finalidades do tratamento de dados. Devendo o responsável pelo tratamento, portanto, realizar uma avaliação objetiva que determine se as operações de tratamento de dados implicam risco ou risco elevado ao titular de dados[491].

Aliás, para facilitar o entendimento do que poderiam ser considerados riscos elevados, o Considerando (75) apresenta uma extensa lista do que ele considera, em abstrato, como os maiores riscos para os titulares de dados, sendo eles os tratamentos (i) que possam causar danos físicos, materiais ou mentais aos titulares de dados, como discriminação; (ii) que se baseiem em categorias especiais de dados pessoais; (iii) que realizem avaliações através da criação de perfis psicológicos e comportamentais; (iv) que digam respeito à pessoas vulneráveis; ou (v) que incidam sobre uma grande quantidade de dados pessoais e afetem um grande número de titulares de dados.

Como se pode perceber, todas as cinco grandes exemplificações de riscos possivelmente elevados dizem respeito direta ou indiretamente ao tratamento de dados automatizado por sistemas de IA. Isso, por si só, já demonstra a necessária atenção que os responsáveis pelo tratamento devem ter quanto aos riscos e impactos que os sistemas de IA podem provocar à indivíduos e à sociedade.

Isso posto, em primeiro lugar, é importante que se reconheça que a abordagem de risco presente no RGPD o elege como um dos regulamentos mais adequados à auxiliar o desenvolvimento de uma IA de confiança, tendo em vista, principalmente, o desígnio preventivo de lidar com os riscos potencializados pela IA. Dessa forma, todo e qualquer tratamento de dados que possa ser considerado como de «risco elevado», principalmente no que se refere aos direitos e liberdades dos titulares, obriga o responsável a proceder a uma avaliação de impacto da proteção de dados, prevista nos artigos 35.º e 36.º do RGPD, objeto de análise da próxima seção desse capítulo.

[490] ICO. Guidance on AI and Data Protection... p.8.
[491] Considerando (76) do RGPD

3.4.4.2. Aspetos gerais da avaliação de impacto da proteção de dados (artigo 35.º)

Como foi percebido na seção anterior, o RGPD apresenta duas abordagens jurídicas que se complementam entre si, sendo uma baseada em direitos e outra baseada no risco. De certa forma, a fim de concretizar essas abordagens, tanto o princípio da responsabilidade (artigo 5.º, n.º 2) quanto as consequentes responsabilidades do responsável pelo tratamento (artigo 24.º, n.º 1), determinam que aquele que realiza o tratamento de dados pessoais deve assegurar e comprovar a sua respetiva conformidade com o RGPD. Nesse sentido, a fim de operacionalizar esse dever, e o RGPD apresenta, entre outros instrumentos[492], a Avaliação de Impacto sobre a Proteção de Dados (AIPD, artigo 35.º).

De acordo com o GT29, uma AIPD é um instrumento documental pelo qual o responsável pelo tratamento deverá (i) descrever como ocorre o tratamento de dados; (ii) avaliar os aspetos relativos à necessidade e proporcionalidade desse tratamento; (iii) gerir os riscos relacionados aos direitos e liberdades das titulares de dados afetados; e (iv) definir as medidas necessárias para fazer face a esses riscos[493]. Como se pode perceber, através desses objetivos, uma AIPD poderá ser considerada como um instrumento que auxiliará o responsável pelo tratamento a cumprir os requisitos e demonstrar a conformidade com o RGPD. Esse é um entendimento compartilhado entre o GP29 e a ICO.

Ocorre que, consoante a abordagem baseada em risco, não é obrigatório que se realize uma AIPD para toda e qualquer operação de tratamento de dados. Isso porque, de acordo com o artigo 35.º, nº 1, uma AIPD deve ser realizada somente quando um tratamento de dados, tendo em conta a sua a natureza, o âmbito, o contexto e as respetivas finalidades, "for suscetível de implicar um elevado risco para os direitos e liberdades das pessoas singulares"[494]. Nesse cenário, além dos referidos riscos aos direitos e liberdades dos titulares, o Considerando (75) também alerta para a possibilidade de que tratamentos de dados possam causar danos físicos, materiais ou imateriais. Isto posto, pode-se partir do pressuposto que qualquer dano ou elevado risco relacionado ao tratamento de dados enseja a realização de uma AIPD. Contudo, é importante que essa afirmação seja melhor desenvolvida.

Aliás, no que se refere à dimensão desses 'direitos e liberdades', é importante destacar que o RGPD não permite que se faça uma interpretação estritamente limitada ao direito da proteção de dados. Na verdade, o próprio GT29 afirma que a proteção de RGPD também envolve outros direitos

[492] A proteção de dados desde a conceção e por defeito (artigo 25.º), a designação do encarregado da proteção de dados (artigo 37.º), os códigos de conduta (artigo 40.º) e as certificações (artigo 42.º) também são formas pelas quais o responsável pelo tratamento assegura a conformidade do tratamento com o RGPD. No entanto, esses instrumentos não serão analisados nesse trabalho acadêmico.

[493] Grupo de Trabalho do Artigo 29 (GT29). Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD) e que determinam se o tratamento é «suscetível de resultar num elevado risco» para efeitos do Regulamento (UE) 2016/679, abril de 2017, p.4.

[494] Os números 5 e 10 do artigo 35.º do RGPD apresentam exceções à obrigatoriedade de realização de um AIPD. A primeira exceção se refere à possibilidade de ser indicada pela Autoridade de Controlo uma lista de tratamentos que não necessitam da AIPD. A segunda exceção afirma que "Se o tratamento efetuado por força do artigo 6.o, n.o 1, alínea c) ou e), tiver por fundamento jurídico o direito da União ou do Estado-Membro a que o responsável pelo tratamento está sujeito, e esse direito regular a operação ou as operações de tratamento específicas em questão, e se já tiver sido realizada uma avaliação de impacto sobre a proteção de dados no âmbito de uma avaliação de impacto geral no contexto da adoção desse fundamento jurídico, não são aplicáveis os nº. 1 a 7, salvo se os Estados-Membros considerarem necessário proceder a essa avaliação antes das atividades de tratamento."

fundamentais, como a liberdade de expressão, a liberdade de pensamento, a liberdade de circulação, a proibição de discriminação, o direito à liberdade, consciência e religião[495]. Essa posição é fundamentada no Considerando (4), que além de designar que o tratamento de dados deve servir as pessoas, também reforça que o RGPD respeita todos os direitos fundamentais e observa as liberdades e os princípios reconhecidos na Carta e consagrados nos Tratados. Nesse sentido, além de mencionar os direitos e liberdades defendidos pelo GT29, o Considerando (4) ainda menciona o respeito pela vida privada e familiar, pelo domicílio e pelas comunicações, a proteção dos dados pessoais, liberdade de informação, a liberdade de empresa, o direito à ação e a um tribunal imparcial, e a diversidade cultural, religiosa e linguística.

Isto posto, da análise conjunta dos direitos e liberdades mencionados acima, pode-se perceber que entre todos eles há uma possível vinculação relacionada ao tratamento de dados. Afirma-se isso pois, de facto, o RGPD foi concebido para proteger os dados pessoais contra tratamento injustos ou irregulares, que se processados de forma mal-intencionada, podem causar uma violação direta ou indireta aos mencionados direitos e liberdades. Como forma de ilustrar essa constatação, é possível recorrer à possibilidade dos algoritmos das redes sociais, através da definição de perfis, causarem elevados riscos aos direitos e liberdades de informação e expressão. Por outro lado, um sistema de IA utilizado para recrutamento profissional pode provocar elevados riscos referentes à discriminação. Além disso, um sistema de IA de vigilância policial também pode violar os direitos de circulação e respeito pela vida privada, pelo domicílio e pelas comunicações. Como se pode perceber, o tratamento de dados, principalmente de caráter pessoal, pode afetar e provocar elevados riscos para direitos e liberdades que estão indiretamente ligadas e dependentes da proteção de dados pessoais.

Essa é uma constatação demasiadamente importante, uma vez que promove grandes consequências para esse trabalho acadêmico. Isso porque, se o RGPD protege e respeita todos esses direitos e liberdades, o responsável pelo tratamento não pode limitar-se a realizar uma AIPD somente em casos em que haja uma probabilidade de elevado risco à proteção de dados em sentido estrito. Dessa forma, entende-se que a AIPD deve sempre ser realizada quando o tratamento de dados for suscetível de provocar um elevado risco a todo e qualquer direito e liberdade que esteja, direta ou indiretamente, vinculado à proteção de dados pessoais, devendo essa proteção ser interpretada em sentido lato. Assim sendo, é importante reconhecer a compatibilidade e amplitude do RGPD para regular e desenvolver uma IA de confiança, uma vez que o RGPD possui a capacidade de ser um regulamento efetivo na proteção, em teoria, contra os inúmeros riscos que desencadearam o debate para a criação de uma IA de confiança. Além disso, também é notável a forma como o RGPD

[495] GT29. Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD)... p.7.

demonstra a sua sustentação através de um pilar ético que encontra reflexo nos desígnios e nos direitos defendidos pela IA de confiança.

Contudo, ainda no que se refere a AIPD, resta uma dúvida a respeito do que poderia ser entendido como «suscetível de implicar» elevados riscos, uma vez que o RGPD não apresenta nenhum conceito ou explicação sobre isso. Contudo, o artigo 35.º, n.º 3, apresenta uma lista não exaustiva de casos em que se é obrigatória a realização de uma AIPD, como (i) a tomada de decisões que produza efeitos jurídicos ou afetem significativamente um usuário através do tratamento automatizado e de definições de perfis que avaliam de forma sistemática e completa os aspetos pessoais dos titulares de dados; (ii) tratamentos em grande escala de categorias especiais de dados (artigo 9.º, n.º1) ou tratamento de dados pessoais relacionados com condenações penais e infrações; e (iii) operações que visam o controlo sistemático de zonas acessíveis ao público em grande escala. Como mencionado, não se trata de uma lista exaustiva, tanto é que o artigo 35.º, n.º 4, legitima a autoridade de controlo a elaborar e tornar pública uma lista dos tipos de operações de tratamento de elevado risco que estão sujeitos à realização de uma AIPD.

A partir desse contexto, e com vistas a fornecer um conjunto mais concreto de informações a respeito de tratamento de dados que possam exigir uma AIPD, o próprio GT29 desenvolveu uma lista com critérios indicativos de operações que possuem grande possibilidade de promoverem elevados riscos aos titulares. A realização dessa lista baseou-se em diversas referências no RGPD, como o artigo 22.º, os Considerandos (71), (75) e (91) e o próprio artigo 35.º, n.º 3, resultando assim em nove critérios que devem ser observados atentamente pelos responsáveis pelo tratamento, sendo eles: (i) operações de avaliação ou classificação através de previsão e definição de perfis; (ii) decisões automatizadas que produzam efeitos jurídicos ou afetem significativamente de modo similar; (iii) tratamento utilizado para observar, monitorizar ou controlar os titulares dos dados ou um controlo sistemático de zonas acessíveis ao público; (iv) tratamento de dados sensíveis ou dados de natureza altamente pessoal; (v) tratamento de dados em grande escala[496]; (vi) operações que visem estabelecer correspondências ou combinar conjuntos de dados; (vii) dados relativos a titulares de dados vulneráveis, como crianças e incapazes, ou onde se percebe um desequilíbrio de poder em relação ao responsável pelo tratamento e o titular dos dados, como relações de trabalho e de consumo; (viii) utilização de soluções inovadoras ou aplicação de novas soluções tecnológicas ou organizacionais; ou (ix) quando o

[496] O RGPD não define o que é «grande escala», no entanto o Considerando (91) e o GT29 fornecem algumas orientações, como (a) o número de titulares de dados envolvidos, quer através de um número específico quer
através de uma percentagem da população pertinente; (b) o volume de dados e/ou a diversidade de dados diferentes a tratar; (c) a duração da atividade de tratamento de dados ou a sua pertinência; (d) a dimensão geográfica da atividade de tratamento. (GT 29. Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD), p.11).

próprio tratamento impede os titulares dos dados de exercer um direito ou de utilizar um serviço ou um contrato[497].

A partir dessa lista de nove critérios, o GT29 afirma que o tratamento de dados que apresente simultaneamente ao menos dois desses critérios, consequentemente representa a possibilidade de elevados riscos aos direitos e liberdades dos titulares, pelo que passa a ser exigível a realização de uma AIPD. Contudo, o GT29 menciona que, a depender da operação, é possível que o responsável pelo tratamento entenda pela exigência de uma AIPD mesmo que esteja presente apenas um único critério[498]. Aliás, é importante destacar que até mesmo uma avaliação preliminar com o fim de determinar se um tratamento pode ou não provocar elevados riscos deve ser documentada pelo responsável pelo tratamento.

Dessa forma, e como pode ser observado do Considerando (90), para que seja avaliada a probabilidade e a gravidade de elevados risco em um tratamento de dados, é preciso que a AIPD seja realizada antes de se iniciar o tratamento de dados. Isto posto, após a realização da análise das circunstâncias particulares do tratamento, a AIPD deve passar a ser encarada como um instrumento de apoio à tomada de decisão em relação ao tratamento, servindo como um verdadeiro roteiro para identificar e controlar os riscos aos direitos e liberdades dos titulares[499], principalmente quando o tratamento pode provocar um risco elevado e significativo, como é em casos de sistemas de IA.

Além disso, a AIPD deve ser considerada como um documento vivo e contínuo, que sofrerá constantes atualizações ao longo de todo o tratamento de dados, o que deve incluir a fase de início do projeto, a fase de desenvolvimento e, ainda, a fase de execução. A atualização da AIPD é imprescindível quando referente a tratamento de dados dinâmico e sujeito a mudanças permanentes[500]. Em complemento, o artigo 35.º, n.º 11, indica a possibilidade de o responsável pelo tratamento realizar um controle para avaliar se o tratamento é realizado em conformidade com a AIPD.

Quanto a realização do AIPD, essa deve ser garantida pelo responsável pelo tratamento (artigo 35.º, n.º 2), que poderá realizar por contra própria ou solicitando a um terceiro externo à organização. Contudo, é possível que, após a realização da AIPD, solicite-se um parecer do Encarregado da Proteção de Dados, quando designado. Nesse contexto, é interessante destacar que o artigo 35, n.º 9, permite, sempre quando for adequado, que o responsável pelo tratamento solicite a opinião dos titulares de dados ou dos seus representantes sobre o tratamento previsto. Essa é, por sinal, uma importante oportunidade para os titulares de dados manifestarem sua opinião sobre a forma e sobre as consequências do tratamento de seus próprios dados, pelo que fortalece a autodeterminação

[497] GT29. Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD)... p.11-12.
[498] GT29. Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD)... p.13.
[499] ICO. Guidance on AI and Data Protection... p.15.
[500] GT29. Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD)... p.17.

informacional, bem como, a participação do titular de dados no processo de desenvolvimento. Nesse sentido, de acordo com o GT29, caso o responsável opte por não consultar os titulares, ele deverá justificar na AIPD o motivo pelo qual tal consulta não se mostrou adequada[501].

Um dos fatores mais importantes da AIPD é, naturalmente, a sua metodologia, ou, em outras palavras, o que exatamente ela deverá abordar. Quanto a esse tema, existem diversas metodologias diferentes, no entanto, a parte essencial que deve ser respeitada pelo responsável pelo tratamento na realização de qualquer AIPD é aquela presente no artigo 35.º, n.º 7. Nesse sentido, uma AIPD deve, pelo menos, apresentar (i) uma descrição sistemática das operações de tratamento previstas, contendo a finalidade do tratamento e, se for o caso, os interesses legítimos do responsável pelo tratamento; (ii) uma avaliação da necessidade e proporcionalidade das operações de tratamento em relação aos objetivos; (iii) uma avaliação dos riscos para os direitos e liberdades dos titulares dos direitos, levando em consideração a sua natureza, o âmbito, o contexto e as finalidades do tratamento; e, por fim, (iv) as medidas previstas para fazer face aos riscos, o que deve incluir as garantias, as medidas de segurança e os procedimentos destinados a assegurar a proteção dos dados pessoais e a demonstrar a conformidade com o RGPD. Nesse sentido, é importante destacar que o responsável não tem o dever de erradicar totalmente os riscos. Contudo, a condução da AIPD deve ser orientada de maneira a minimizar e determinar se o nível de risco é ou não aceitável nas circunstâncias concretas de tratamento de dados[502].

Como mencionado, essas quatro características se referem ao núcleo central e essencial de uma AIPD, pelo que, a depender do caso concreto, o responsável pelo tratamento tem liberdade para aprofundar a análise e a avaliação do tratamento de dados para que reflita os seus níveis de preocupação com os riscos aos titulares. Nesse contexto, é importante destacar que, à luz do RGPD, a avaliação dos riscos aos direitos e liberdades dos titulares e a respetiva condução das salvaguardas devem ser realizadas a partir da perspetiva própria do titular de dados[503], e não pelo ponto de vista da organização. Isto posto, a próxima seção será destinada a apresentar uma metodologia melhor desenvolvida para a realização de uma AIPD relacionada ao tratamento de dados automatizados por sistemas de IA. Tal metodologia será baseada nos guias do GP29 e da ICO.

Dessa forma, após a realização da AIPD através da (i) avaliação dos riscos para os direitos e as liberdades dos titulares dos dados, da (ii) identificação das medidas previstas para reduzir esses riscos para um nível aceitável e da (iii) demonstração da conformidade com o RGPD, será permitido que o responsável pelo tratamento proceda ao tratamento de dados, sem, portanto, realizar qualquer consulta à autoridade de controlo. Por outro lado, caso o responsável pelo tratamento realize a AIPD e perceba

[501] GT29. Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD)... p.17-18.
[502] ICO. Data Protection Impact Assessment. Version 1.0.124. May 2018, p. 14.
[503] GT29. Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD)... p.20.

que as medidas para mitigar os riscos são inadequadas ou insuficientes, estará caracterizada a figura dos riscos residuais. Esses riscos são percebidos quando não há, não se conhece ou não foram aplicadas as medidas adequadas face os elevados riscos identificados na AIPD[504]. Caso isso aconteça, o artigo 36.º, n.º 1, determina que o responsável pelo tratamento consulte a autoridade de controlo antes de proceder ao tratamento de dados. Esse procedimento é denominado pela RGPD como uma consulta prévia[505].

Dentro desse contexto, e como resposta à consulta realizada, caso a autoridade de controlo considere que o tratamento de dados proposto viola os direitos e as garantias do RGPD, a autoridade fornecerá orientações, por escrito, ao responsável pelo tratamento (artigo 36.º, n.º 2). Ocorre que, ainda dentro desse escopo, a autoridade tem a faculdade de recorrer a todos os seus poderes garantidos pelo artigo 58.º, o que a permite, entre tantas ações, ordenar ao responsável pelo tratamento que tome medidas no sentido de cumprir as disposições do regulamento (artigo 58.º, n.º 2, d) ou, em casos extraordinários, de impor uma limitação temporária ou definitiva ao tratamento, ou mesmo a sua proibição (artigo 58.º, n.º 2, f).

Nesse sentido, a AIDP, quando conduzida de forma a respeitar o RGPD, impõe ao responsável a obrigação de identificar e minimizar os riscos à um padrão aceitável, e, caso não o faça, submeter o tratamento à uma avaliação da autoridade de controle. Essa, caso entenda que os riscos são demasiado elevados, poderá até mesmo proibir o tratamento de dados. Dessa forma, independentemente da solução adotada, adequação ou interrupção do tratamento, percebe-se que os direitos e os interesses do titular serão sempre protegidos. Assim sendo, por mais que no quadro para uma IA de confiança não haja nenhuma previsão semelhante, a não ser a lista de autoavaliação que não é uma obrigação legal, pode-se identificar na AIPD um forte instrumento de proteção dos direitos fundamentais dos titulares de dados. Como também, consequentemente, de uma IA de confiança.

Por fim, uma última característica que deve ser observada é que, apesar da orientação do GT29 incentivar a publicação de um resumo ou das conclusões da AIPD, o RGPD não promove nenhuma obrigação legal de publicação desse documento por parte do responsável pelo tratamento. Observe que, de acordo com o GT29, não há qualquer necessidade de que a AIPD publique a totalidade da avaliação, muito menos que revele segredos comerciais ou informações comerciais sensíveis. O que se espera de uma publicação parcial é, na verdade, um resumo das principais conclusões. Dessa forma, caso o responsável deseje publicar, ele o fará por voluntariedade. Nesse sentido, além de demonstrar que a AIPD foi realizada, uma publicação, mesmo que parcial, pode

[504] GT29. Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD)... p.22.

[505] De acordo com o artigo 36.º, n.º 5, esse procedimento também deve ser adotado quando o direito dos Estados-Membros exigir que os responsáveis pelo tratamento consultem a autoridade de controlo e dela obtenham uma autorização prévia em relação ao tratamento por um responsável no exercício de uma missão de interesse público, incluindo o tratamento por motivos de proteção social e de saúde pública.

ajudar a estimular um sentimento de confiança em relação as operações de tratamento, além de demonstrar responsabilidade e transparência[506].

Aliás, por mencionar o termo 'transparência', é importante destacar que os deveres de informação ativos encontrados nos artigos 13.º e 14.º do RGPD podem ser, de facto, executados através da publicação de uma AIPD. Afirma-se isso pois, de acordo com o artigo 35.º, n.º 7, a), uma AIPD deve apresentar uma descrição sistemática das operações de tratamento previstas, incluindo a respetiva finalidade. Sendo essa uma obrigação que encontra semelhanças no que concerne aos deveres do responsável quando esse promove o tratamento automatizado através de sistemas de IA, uma vez que esse tipo de tratamento impõe a necessidade de apresentar informações úteis relativas à lógica de funcionamento do tratamento (artigo 13.º, n.º 2, f e artigo 14.º, n.º 2). Além disso, de acordo com o artigo 35.º, n.º 7, c), uma AIPD deve realizar uma avaliação dos 'riscos aos direitos e liberdades' dos indivíduos. Nesse sentido, os indivíduos têm o direito, no contexto da tomada de decisão automatizada, de serem informados da 'importância e consequências previstas' relativas a esse referido tratamento.

Como se pode perceber, a avaliação proposta pela AIPD pode responder a inúmeros direitos do titular de dados, sendo, portanto, mais um fator de efetividade do princípio da responsabilidade e da transparência. Além disso, é possível perceber que esse trabalho encara a AIPD como um poderoso instrumento de gestão e avaliação de riscos capaz, paralelamente, de proteger os direitos tutelados pelo RGPD. Pelo que, através dessa perspetiva, entende-se que a AIPD responde à dupla abordagem jurídica do RGPD, ou seja, as abordagens de risco e de direito. Dessa forma, mais do que demonstrar conformidade com o atual regulamento, a utilização da AIPD em sistemas de IA pode ser um passo de operacionalização de uma IA de confiança.

Contudo, a falta de um mecanismo obrigatório de publicação da AIPD impede que esses desígnios possam ser realizados. Aliás, a ausência de obrigatoriedade legal de publicação de ao menos um resumo das conclusões da AIPD pode ser considerada a maior falha desse instrumento[507]. Isso porque, a divulgação obrigatória ao público, respeitados os segredos e as informações comerciais essenciais, permitiria uma auditoria pública realizada por outras empresas e por organizações da sociedade civil. Esse escrutínio não só ajudaria a coibir tratamentos de dados suscetíveis de causar elevados riscos aos direitos e liberdades dos titulares de dados, mas também iria influenciar e fortalecer a realização da AIPD e o seu respetivo papel de responder, concomitantemente, à abordagem

[506] GT29. Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD)... p.21.

[507] Malgieri, Gianclaudio e Kaminski, Margot E. Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations, p.19; Kloza et al, 'Data Protection Impact Assessments in the European Union: Complementing the New Legal Framework towards a More Robust Protection of Individuals', p.3; Michael Veale, Reuben Binns, and Jef Ausloos, 'When Data Protection by Design and Data Subject Rights Clash', International Data Privacy Law 8, no. 2 (1 May 2018), p.118, https://doi.org/10.1093/idpl/ipy002

de risco promovida pela RGPD, como também à uma proteção substantiva dos direitos e liberdades dos indivíduos.

Isto posto, por mais que esse trabalho acadêmico entenda que a AIPD possa ser um caminho inicial para a operacionalização de uma IA de confiança, nada disso será possível se não houver uma maneira de legalmente obrigar a publicação de um resumo das conclusões da AIPD. E para que isso seja possível, é necessário que os órgãos legislativos competentes iniciem um debate sobre esse tema.

3.5. Uma possível metodologia para a AIPD em sistemas de IA

Como analisado na seção anterior, a AIPD é um instrumento documental pelo qual o responsável pelo tratamento poderá, em teoria, assegurar e comprovar que o tratamento de dados está em conformidade com o RGPD. De acordo com o próprio artigo 35.º, n. º1, a AIPD deve ser realizada sempre que o tratamento de dados for suscetível de provocar elevados riscos aos direitos e liberdades dos titulares de dados. No que concerne aos sistemas de IA, da leitura do artigo 35.º, n.º 7 e dos critérios desenvolvidos pelo GT29, conclui-se que o tratamento automatizado é uma das tantas operações de dados que são passíveis de serem objeto de uma AIPD. Afirma-se isso pois, dentre os critérios indicados de suscetibilidade de provocar elevados riscos estão, nomeadamente, (a) a tomada de decisões através do tratamento automatizado que produzem efeitos jurídicos ou decisões que afetem significativamente um usuário; como também (b) a prática de definições de perfis que avaliam de forma completa e sistemática os aspetos pessoais dos titulares de dados.

Contudo, por mais que seja facilmente percebida a exigência relativa à realização de uma AIPD em sistemas de IA, o texto do RGPD fornece uma visão muito geral e abstrata sobre a metodologia a ser utilizada, restringindo-se a apresentar apenas quatro requisitos que devem ser enfrentados pela AIPD. Estes requisitos estão relacionados com a necessidade de apresentar (i) uma descrição sistemática das operações de tratamento previstas, contendo a finalidade do tratamento e, se for o caso, os interesses legítimos do responsável pelo tratamento; (ii) uma avaliação da necessidade e proporcionalidade das operações de tratamento em relação aos objetivos; (iii) uma avaliação dos riscos para os direitos e liberdades dos titulares dos direitos, levando em consideração a sua natureza, o âmbito, o contexto e as finalidades do tratamento; e, por fim, (iv) as medidas previstas para fazer face aos riscos, o que deve incluir as garantias, as medidas de segurança e os procedimentos destinados a assegurar a proteção dos dados pessoais e a demonstrar a conformidade com o RGPD.

Isto posto, e ao perceber que tais orientações eram abstratas e insuficientes para guiar a realização de uma AIPD em sistemas de IA, essa seção tem como objetivo apresentar uma possível

metodologia para ser utilizada em avaliações de sistemas de IA que possam causar elevados riscos aos direitos e liberdades dos titulares. Contudo, desde já é importante que se faça algumas observações.

A primeira delas é que nem o RGPD nem qualquer autoridade de controlo determina uma metodologia correta ou obrigatória, pelo que o responsável pelo tratamento pode desenvolver a sua própria forma de realizar a AIPD. Nesse sentido, a metodologia aqui desenvolvida foi construída a partir das orientações do GP29[508] e da ICO[509][510], e, portanto, não é algo criado particularmente do zero.

A segunda observação é que, a fim de organizar essa metodologia, levou-se em consideração a interpretação que esse próprio trabalho acadêmico realizou sobre o RGPD. Essa abordagem particular acaba por repercutir em uma maior dimensão protetiva em relação aos direitos e liberdades, isso porque, a partir dessa interpretação conclui-se que o RGPD protege o direito da proteção de dados, mas também respeita os diversos outros direitos fundamentais encontrados na Carta e que estão, direta ou indiretamente, relacionados com o tratamento de dados pessoais. Além disso, tenta-se aqui construir uma visão pela qual a AIPD tem a capacidade de dar uma resposta, simultaneamente, aos desígnios da abordagem baseada em risco e da abordagem baseada em direito, sendo, portanto, um meio pelo qual, através das análises dos riscos, tenta providenciar meios que possibilitem uma proteção substantiva aos direitos e liberdades dos titulares.

Isto posto, a presente metodologia é dividida em sete etapas, sendo elas: (i) identificação da necessidade de se realizar uma AIPD; (ii) descrição sistemática das operações de tratamento previstas; (iii) consideração de consulta a terceiros; (iv) avaliação da necessidade e proporcionalidade; (v) identificação e avaliação dos riscos para os direitos dos titulares; (vi) definição das medidas para mitigar os riscos (vii) documentação dos resultados e avaliação da necessidade de consulta prévia. Como se pode perceber, a sua estrutura é àquela apresentada pela ICO[511], e compreende os 4 requisitos exigidos pelo RGPD. Contudo, o seu conteúdo será uma junção entre o que a ICO[512] e o que o GT29[513] propõem, com algumas alterações particulares que são fruto da interpretação do RGPD desse trabalho.

3.5.1. Identificação da necessidade de se realizar uma AIPD

De acordo com o artigo 35.º, n.º 1, do RGPD, a realização de uma AIPD é necessária quando o tratamento de dados pode resultar em um elevado risco aos direitos e liberdades dos titulares de

[508] GT29. Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD)... p.26-27.
[509] Information commissioner's Office (ICO). Guidelines on Data protection impact assessments. Version 1.0.124. May 2018, p.29-38
[510] ICO. Guidance on AI and Data Protection...
[511] ICO. Guidelines on Data protection impact assessments... p.29-38
[512] ICO. Guidance on AI and Data Protection. Version 0.0.22. July 2020.
[513] GT29. Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD)... p.26-27.

dados. Como visto, essa é a regra geral que guia a AIPD, no entanto, definir o que é elevado risco ainda pode ser uma tarefa abstrata. Isto posto, existem duas etapas para identificar se o tratamento de dados pode ou não ser suscetível de provocar um elevado risco.

A primeira das etapas é bem concreta e objetiva, sendo baseada em uma lista de nove critérios que o GT29 desenvolveu como fatores que indicam a possibilidade de um tratamento de dados representar elevado risco aos direitos e liberdades dos titulares. Para desenvolver esses critérios, o GT29 se inspirou em diversas referências no RGPD, como o artigo 22.º, os Considerandos (71), (75) e (91) e o próprio artigo 35.º, n.º 3, dispositivo esse que exemplifica casos em que são necessários uma AIPD.

Dessa forma, a fim de identificar se um sistema de IA necessita de uma AIPD, deve-se perceber se o tratamento de dados que ele realiza indica ao menos duas das nove características seguintes, sendo elas: (i) operações de avaliação ou classificação através de previsão e definição de perfis; (ii) decisões automatizadas que produzam efeitos jurídicos ou afetem significativamente de modo similar; (iii) tratamento utilizado para observar, monitorizar ou controlar os titulares dos dados ou um controlo sistemático de zonas acessíveis ao público; (iv) tratamento de dados sensíveis ou dados de natureza altamente pessoal; (v) tratamento de dados em grande escala; (vi) operações que visem estabelecer correspondências ou combinar conjuntos de dados; (vii) dados relativos a titulares de dados vulneráveis, como crianças e incapazes, ou onde se percebe um desequilíbrio de poder em relação ao responsável pelo tratamento e o titular dos dados, como relações de trabalho e de consumo; (viii) utilização de soluções inovadoras ou aplicação de novas soluções tecnológicas ou organizacionais; ou (ix) quando o próprio tratamento impede os titulares dos dados de exercer um direito ou de utilizar um serviço ou um contrato[514]. Se o tratamento possuir ao menos duas dessas características, é imprescindível que se realize uma AIPD.

A segunda etapa para perceber se será necessária a realização de uma AIPD é baseada no artigo 35.º, n.º 4, onde o RGPD legitima a autoridade de controlo à "elabora e torna pública uma lista dos tipos de operações de tratamento sujeitos ao requisito de avaliação de impacto sobre a proteção de dados". Essa segunda etapa deve ser realizada consultando a respetiva autoridade de controlo de cada Estado-Membro em que o sistema de IA será utilizado. A realização dessa consulta é imprescindível, uma vez que determinados tipos de tratamento de dados podem não compreender os critérios apresentados anteriormente, mas ao mesmo tempo eles podem provocar um risco inadmissível pela perspetiva da autoridade de controlo competente. Isso posto, caso haja um tratamento de dados que esteja inserido nessa lista, torna-se também imprescindível a realização da AIPD.

[514] GT29. Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD)... p.11-12.

Contudo, se após a análise prévia for identificado que não é necessária a realização de uma AIPD, é necessária a respetiva documentação dessa decisão acompanhada das razões que fundamentaram essa posição[515]. Nesse sentido, e de acordo com o artigo 35.º, n.º 2, a análise prévia também deve conter um parecer do encarregado da proteção de dados, caso designado.

3.5.2. Descrição sistemática das operações de tratamento previstas (artigo 35.º, n.º 7, alínea a)

A descrição sistemática das operações de tratamento serve tanto para demonstrar a amplitude do que deve ser analisado pela AIPD, como também servem, posteriormente, para cumprir com os deveres do responsável pelo tratamento em fornecer informações úteis a respeito do tratamento automatizado. Nesse sentido, é preciso deixar claro «como» e «por que» o responsável pelo tratamento irá usar sistemas de IA para tratar os dados. Além disso, a fim de fornecer uma descrição sistemática do tratamento, o responsável deverá, em primeiro lugar, apresentar informações a respeito da natureza, do âmbito, do contexto e das finalidades do tratamento.

A natureza do processamento representa o que o responsável pelo tratamento planeja fazer com os dados pessoais, isso posto, ele deverá especificar as seguintes informações: (a) como os dados são coletados, armazenados e utilizados; (b) quem tem acesso a esses dados e com que terceiros eles são partilhados, incluindo processadores; (c) qual é o período de retenção; (d) quais são as medidas de segurança aplicadas; e (e) se há uso de alguma nova tecnologia ou algum tipo novo de processamento[516].

O âmbito do processamento refere-se propriamente ao tratamento dos dados, e o responsável deverá indicar: (a) a natureza dos dados pessoais, ou seja, sobre quais aspetos pessoais ele diz respeito; (b) se há dados de categorias especiais envolvidos no tratamento; (c) o volume e a variedade dos dados pessoais; (d) a extensão e a frequência do processamento; (e) a duração do processamento; (f) o número de sujeitos de dados envolvidos; e (g) a área geográfica coberta pelo tratamento[517]. Essas informações serão importantes para a avaliação de necessidade e proporcionalidade.

O contexto do processamento diz respeito à fatores internos e externos que podem afetar as expectativas dos titulares ou gerar impacto negativo em seus direitos. Isso posto, o responsável deve especificar: (a) a fonte dos dados; (b) até que ponto os indivíduos têm controle sobre seus dados; (c) qual a expectativa dos titulares a respeito desse tratamento; (d) se entre os titulares há crianças ou

[515] ICO. Guidelines on Data protection impact assessments... p.32
[516] ICO. Guidelines on Data protection impact assessments..., p.32
[517] Ibid.

outras pessoas vulneráveis; (d) se há conformidade como códigos de conduta aprovados (artigo 35.º, n.º 8) ou qualquer outra certificação que o responsável julgue eficiente[518].

Por fim, a finalidade do tratamento é a razão pela qual aquele dado pessoal é tratado, isso posto, o responsável deverá (a) indicar qual é o resultado pretendido através desse tratamento; (b) qual é o benefício esperado pelo responsável ou pela sociedade; e (c) apresentar as razões do legítimo interesse no tratamento, caso existam[519].

Cumprida a primeira parte, o responsável deverá registar na AIPD as fases em que os processos de IA e as decisões automatizadas podem produzir efeitos na vida dos titulares. Também será preciso que seja registada uma explicação de qualquer variação relevante ou margem de erro no desempenho do sistema que possa afetar a equidade do processamento de dados pessoais[520]. Além disso, de acordo com a ICO, é importante que o AIPD também registe "o grau de qualquer envolvimento humano no processo de tomada de decisão e em que fase isso ocorre".

3.5.3. Consideração de consulta a terceiros (artigo 35.º, n.º 2 e 9)

O artigo 35.º, n.º 9, determina que, se for adequado, o responsável pelo tratamento pode solicitar a opinião dos titulares de dados ou dos seus representantes sobre os aspetos envolvidos no tratamento. Caso seja solicitado, é importante que se documente as opiniões apresentadas. No entanto, caso entenda que essa consulta não é adequada, o responsável deverá fundamentar a sua decisão e documentar isso junto à AIPD[521]. Se os dados são compartilhados com processadores, também será importante a participação ou o aconselhamento destes, para que a AIPD possa possuir todos os pontos de vista necessários.

Por fim, a ICO recomenda que ao longo da realização da AIPD haja apoio e aconselhamento jurídico especializado, além da participação de especialistas independentes de áreas como TI, sociologia ou ética[522]. Essa é uma participação demasiadamente importante, uma vez que esses especialistas indicarão possibilidades de riscos aos titulares através de uma perspetiva multidisciplinar. Nesse sentido, tendo em conta que sistemas de IA são complexos e podem causar riscos até mesmo irreparáveis, é importante a presença do encarregado da proteção de dados desde o início do ciclo de vida do sistema de IA.

---

[518] v p.32-33
[519] ICO. Guidelines on Data protection impact assessments... p.33
[520] ICO. Guidance on AI and Data Protection... p. 17.
[521] GT29. Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD)... p.17-18.
[522] ICO. Guidelines on Data protection impact assessments... p.34

3.5.4. Avaliação da necessidade e proporcionalidade (artigo 35.º, n.º 7, alínea b)

De acordo com a ICO, para que sistemas de IA sejam usados para realização de tratamento automatizado de dados pessoais é necessário que seja comprovado que há um problema a ser resolvido e que a IA pode ser uma solução sensata para esse problema[523]. Nesse sentido, tendo em vista os riscos atrelados à IA, ela jamais poderá ser utilizada pela mera disponibilidade da tecnologia.

Nesse sentido, a fim de avaliar a necessidade e a proporcionalidade do tratamento, o responsável deve, naturalmente, responder à duas perguntas. A primeira refere-se à uma autocritica no sentido de se indagar se o projeto definitivamente auxilia a alcançar o propósito de tratamento pretendido. Caso a resposta seja positiva, o responsável deve agora se perguntar se há ou não um outro meio alternativo, razoável e menos intrusivo de se conseguir os mesmos resultados.

Ultrapassada essa primeira parte, e inspirado na metodologia para uma AIPD desenvolvida pelo GT29, o responsável pelo tratamento poderá avaliar a necessidade e a proporcionalidade através de duas perspetivas, sendo elas: (a) as medidas que contribuem para a necessidade e proporcionalidade do tratamento, e (b) as medidas que contribuem para os direitos dos titulares[524].

3.5.4.1. Necessidade e proporcionalidade referente ao tratamento de dados

No que se refere à análise da necessidade e da proporcionalidade referente ao tratamento de dados, o responsável deverá analisar (a) se as finalidades são determinadas, explicitas e legítimas; (b) se os dados são adequados, pertinentes e limitados ao que é necessário[525]; (c) se tratamento respeita as bases legais corretas; (d) se o processamento extrapola o seu propósito; e (e) se a exatidão e a qualidade dos dados estão a ser garantidas[526].

Além disso, ao utilizar sistemas de IA para realizar o tratamento de dados pessoais, o AIPD deverá levar em consideração e apresentar razões que fundamentem a escolha pelo tratamento automatizado em detrimento a um meio alternativo menos arriscados e intrusivo (se existir)[527]. Essas razões devem excecionalmente serem bem fundamentadas sempre que o responsável pelo tratamento estiver tratando dados com fundamentos no interesse pública ou interesses legítimos.

---

[523] ICO. Guidance on AI and Data Protection... p. 18.
[524] GT29. Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD)... p.26.
[525] Ibid.
[526] ICO. Guidelines on Data protection impact assessments... p.34-35.
[527] ICO. Guidance on AI and Data Protection... p. 16.

3.5.4.2. Medidas para promoção dos direitos dos titulares

No que se refere aos direitos dos titulares de dados, uma das primeiras observações que deve ser realizada refere-se à uma avaliação sobre a expectativa razoável dos titulares em terem seus dados tratados através de um sistema de IA[528]. Essa análise deve ser feita com base no princípio da transparência (artigo 5.º n.º 1, alínea a) e nos deveres de informação.

Além disso, no que se refere à necessidade e a proporcionalidade também deve ser observado como os titulares poderão exercer os seus respetivos direitos, isso posto, o responsável deverá decidir: (a) como as informações serão fornecidas aos indivíduos (artigos 12.º, 13.º e 14.º); (b) como o titular solicitará os direitos de acesso e de portabilidade (artigos 15.º e 20.º); (c) como será exercido os direitos de retificação e apagamento de dados (artigos 16.º, 17.º e 19.º); (d) como operacionalizar o direito de oposição e limitação do tratamento (artigos 18.º 19.º e 21.º)[529]; (e) sobre as maneiras de garantir que os processadores também cumpram os direitos dos titulares; (f) sobre as salvaguardas para transferências internacionais[530]; e, por fim, (g) como implementar processos para garantir que decisões automatizadas estejam sujeitas a intervenção humana, contestação ou revisão (artigo 22.º, n.º 3).

3.5.5. Identificação e avaliação dos riscos aos direitos dos titulares (artigo 35.º, n.º 7, alínea c)

Essa é talvez a etapa que mais sofra influência das interpretações que esse trabalho retirou do RGPD, isso porque aqui os referidos riscos para os direitos e liberdades serão interpretados através de uma perspetiva ampla, onde, em primeiro lugar, será levado em consideração todo e qualquer tipo de dano que o tratamento de dados puder causar, seja ele físico, emocional ou material. Além disso, os riscos causados não serão restritos somente àqueles que tiverem uma ligação direta com o direito da proteção de dados. Afirma-se isso pois, como já analisado, o Considerando (04) e o GT29 afirmam que o RGPD respeita os direitos fundamentais da Carta. Isso posto, esse trabalho passa a entender que os riscos devem dizer respeito à diretos que, direta ou indiretamente, são impactados pelo tratamento de dados.

Dessa forma, o responsável deverá identificar se o tratamento de dados levado a cabo pelo sistema de IA é suscetível a provocar riscos referentes à: (a) incapacidade de exercer direitos em geral; (b) incapacidade de ter acesso a serviços ou novas oportunidades; (c) discriminação; (d) perda do controle sobre o uso dos dados pessoais dos quais é titular; (e) alguma espécie de condicionamento ou

[528] ICO. Guidance on AI and Data Protection... p. 18.
[529] GT29. Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD)... p.26.
[530] ICO. Guidelines on Data protection impact assessments... p.34-35.

manipulação; (f) possibilidade de roubo de identidade ou de fraude; (g) prejuízos financeiros; (h) danos à reputação; (i) danos físicos; (j) possibilidade de re-identificação de dados pseudonimizados; e (l) qualquer outra desvantagem econômica ou social significativa[531].

Além da necessidade de identificar eventual risco, também é necessário quantificar esse risco, para que seja possível categorizá-lo pela sua gravidade. Note que, de acordo com a ICO, "o dano não tem de ser inevitável para se qualificar como um risco ou um risco elevado"[532]. Além disso, cada responsável é livre para definir um padrão próprio de quantificação e qualificação do risco.

3.5.6. Definição de medidas para mitigar os riscos (artigo 35.º, n.º 7, alínea d)

Essa etapa é uma continuação direta do resultado encontrado na etapa anterior. Isso posto, a fim de definir medidas adequadas a mitigar ou eliminar os riscos, o responsável pelo tratamento deverá, em primeiro lugar, registar a fonte de cada risco identificado na etapa anterior. A partir do momento que o responsável sabe qual é o risco e o que o está a originá-lo, ele passará a ser capaz de buscar medidas adequadas para sua mitigação ou eliminação.

Nesse sentido, é importante destacar que a causa de cada risco pode remeter a uma medida técnica ou organizativa diferente. Dessa forma, é importante que o responsável pelo tratamento tenha à sua disponibilização uma equipe de especialistas prontos para propor as devidas medidas. Isso posto, a ICO apresenta, em caráter exemplificativo e não exaustivo, algumas medidas que podem ser utilizadas pelo responsável pelo tratamento, como (a) redução de recolha de certos tipos de dados; (b) redução do âmbito de processamento; (c) redução dos períodos de retenção; (d) promoção de medidas adicionais de segurança; (e) capacitação e conscientização de todos os envolvidos para mitigar os riscos associados a erros humanos; (f) utilização de técnicas de anonimização ou pseudonimização de dados sempre que possível; (g) definição de orientações e políticas internas; (h) implementação de novos sistemas para ajudar os titulares a exercerem seus direitos[533]; como também (j) criação de meios que oportunizem os titulares a não optarem por não estar sujeitos à decisões automatizadas[534].

3.5.7. Documentação de resultados e avaliação de consulta prévia (artigo 36.º, n.º 1)

Encerrada as etapas anteriores da AIPD, o responsável pelo tratamento deverá (a) registrar as medidas adicionais que deverá realizar para mitigar ou eliminar os riscos identificados; (b) classificar

[531] ICO. Guidelines on Data protection impact assessments... p.35.
[532] Ibid.
[533] ICO. Guidelines on Data protection impact assessments... p.36-37.
[534] ICO. Guidance on AI and Data Protection... p. 19.

cada risco identificado como «eliminado, «reduzido» ou «aceito»; (c) identificar a presença de algum risco residual que possa ter ele sido incapaz de definir medidas adequadas face algum dos riscos; e, por fim, (d) decidir sobre a necessidade ou não de consultar previamente a autoridade de controlo[535].

Existem observações importantes sobre essa fase. Uma delas é que o responsável não necessita eliminar todos os riscos, pelo que alguns riscos, elevados ou não, podem ser aceitáveis mediante uma análise de proporcionalidade entre riscos, benefícios do processamento e dificuldades de mitigação[536]. Contudo, a existência de um elevado risco exige que seja realizada a consulta prévia (artigo 36.º, n. º1) junto com a autoridade de controle.

Além disso, é preciso que se tenha em mente que, de acordo com o artigo 35.º, n.º 2, o encarregado da proteção de dados, quando designado, deve emitir um parecer sobre a AIPD. Caso o encarregado, em seu parecer, se posicione contra o tratamento ou que apresente razões que demonstrem que o tratamento não está em conformidade com o RGPD, o responsável poderá aceitar ou ignorar essa posição. Caso opte por ignorar e dar continuidade ao tratamento, o responsável deverá fundamentar sua decisão. Razões também deverão ser apresentadas para confrontar a opinião de titulares de dados que foram consultados e se manifestaram contrariamente a algum especto do tratamento.

[535] ICO. Guidelines on Data protection impact assessments... p.37.
[536] Ibid.

CONCLUSÃO

Ao longo de todo o trabalho, foi possível realizar uma apresentação sobre os aspetos técnicos e os impactos negativos provocados pela má utilização de sistemas de IA (Capítulo 1); um estudo sobre o desígnio europeu de desenvolver um quadro ético para uma IA de confiança (Capítulo 2); e, por fim, uma análise do papel e dos limites do RGPD e da AIPD na regulação e no desenvolvimento de uma IA de confiança, apresentando uma possível metodologia para esse fim.

Além dos aspetos técnicos que compreendem a IA, foi possível perceber que, apesar dos diversos benefícios, sistemas de IA provocam impactos e riscos que podem ser difíceis de prever, identificar ou quantificar[537]. Nesse cenário, os riscos mais críticos estão relacionados a limitação da privacidade através de sistemas de IA e de uma cultura baseada em uma economia de vigilância; na possibilidade de tratamento de dados pessoais para compor perfis psicológicos a fim de condicionar ou manipular os indivíduos; na presença de dados enviesados que fazem com que sistemas de IA possam reproduzir ou intensificar ações injustificadamente preconceituosas ou discriminatórias; e, por fim, na opacidade de sistemas de IA que tomam decisões que impactam significativamente a vida de pessoas sem que, ao menos, lhe ofereçam uma explicação sobre a decisão ou sobre o processo decisório.

Ciente desse cenário, a partir de 2018 a União iniciou um debate sobre os desafios éticos, legais e sociais da IA, que teve como fruto o desígnio europeu de desenvolver o que atualmente se chama de uma IA de confiança, ou seja, um IA responsável, segura, antropocêntrica, centrada nos valores europeus e guiada por princípios éticos, como o respeito pela autonomia humana, a prevenção de danos, a equidade e a explicabilidade.

Para desenvolver esse conjunto de princípios éticos, o GPAN IA se baseou em direitos fundamentais consagrados no Artigo 2.º do TUE e no Artigo e 3.º da CDFUE, facto este que indica que a Comissão idealiza a IA europeia baseada em "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"[538], assegurando, ainda, o "respeito pela sua integridade física e mental"[539].

Em abril de 2021, baseado no que foi apresentado sobre a IA de confiança, a Comissão tornou pública a sua proposta de regulamento sobre Inteligência Artificial para a União Europeia. A proposta adotou uma abordagem regulatória baseada em risco, proibindo usos da IA que provocam um risco inadmissível e regulando usos de AI de alto risco. Essa abordagem foi escolhida uma vez que a Comissão deseja que a regulação seja equilibrada a ponto de assegurar a proteção dos interesses da

[537] GPAN IA. Orientações Éticas para uma IA de Confiança... p. 16.
[538] Artigo 2.º do TUE.
[539] Artigo 3.º da Carta.

sociedade e dos direitos fundamentais dos indivíduos sem que crie restrições desnecessárias que impeçam a inovação e o desenvolvimento de sistemas de IA.

Contudo, apesar de sua existência, a proposta de regulamento de IA ainda precisa enfrentar todo o processo legislativo para que possa produzir efeitos legais, o que pode levar um período de tempo considerável. Isso significa que o cenário europeu ainda se encontra sem uma regulamentação específica e aplicável para a IA, exigindo assim que diplomas setoriais sejam responsáveis pela tutela de sistemas de IA[540], como é o caso do Regulamento Geral sobre Proteção de Dados Pessoais. Essa conjuntura levou esse trabalho acadêmico a analisar se o RGPD poderia ser um instrumento adequado para regular e desenvolver uma IA de confiança.

O RGPD, por sinal, apresenta uma semelhança com o quadro ético para uma IA de Confiança, uma vez que ficou estabelecido nesse trabalho que o nível de proteção desse regulamento não se limita ao direito da proteção de dados. Na verdade, através do Considerando (4) e da posição do GT29[541], constatou-se que o RGPD respeita os direitos fundamentais e observa as liberdades e os princípios reconhecidos na Carta e consagrados nos Tratados, nomeadamente, o respeito pela vida privada e familiar, pelo domicílio e pelas comunicações; a proteção dos dados pessoais; as liberdades de informação, de expressão, de pensamento e de circulação; o direito à ação e a um tribunal imparcial; a proibição de discriminação e o direito a diversidade cultural, religiosa e linguística.

A partir dessa constatação, o RGPD nos permitiu concluir que sua proteção à direitos e liberdades deve ser interpretada de forma ampla. Nesse sentido, o RGPD, que foi concebido para proteger os dados pessoais contra tratamento injustos ou irregulares, deverá ser um fator de regulação frente a todo e qualquer tratamento de dados pessoais que possa, direta ou indiretamente, provocar algum dano físico, emocional ou material ao titular de dados. Observe que, através dessa perspetiva, o RGPD não só se equipara, como também demonstra promover uma proteção tão ampla aos indivíduos do que aquela reconhecida pelo quadro ético para uma IA de Confiança, desde que, tais riscos sejam ocasionados pelo tratamento de dados pessoais.

Isso posto, chega-se a talvez uma das conclusões basilares dessa pesquisa científica. Isso porque, se o RGPD protege e respeita todos esses direitos e liberdades, o responsável pelo tratamento não pode limitar-se a realizar uma AIPD somente em casos em que haja uma probabilidade de elevado risco à proteção de dados em sentido estrito. Dessa forma, entende-se que a AIPD deve sempre ser realizada quando o tratamento de dados for suscetível de provocar um elevado risco a todos os direitos e liberdades que estejam, direta ou indiretamente, vinculados ao tratamento de dados pessoais. Assim

[540] Considerando que uma proposta de regulação não produz efeitos legais, que não se tem certeza de quando a proposta será aprovada, e se ela será aprovada no mesmo molde apresentado, conclui-se que a premissa inicial dessa pesquisa ainda se mantem, ou seja, não existe atualmente um regulamento especificamente aplicado a IA. Isso posto, continua válida, atual e necessária uma análise da adequação do RGPD como instrumento de regulação do ideal europeia de uma IA de Confiança.

[541] GT29. Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD)... p.7.

sendo, pode-se concluir que ao avaliar um sistema de IA, a identificação de riscos promovida pela AIPD poderá também alcançar os diversos direitos fundamentais que a IA de Confiança pretende defender.

Como se pode perceber, no que diz respeito a amplitude de proteção de diretos fundamentais, o RGPD tem se mostrado adequado à regulação de uma IA de Confiança, porém, existem diversos pontos que devem ser analisados particularmente, uma vez que existem direitos que a IA de Confiança pretende promover e que o RGPD não apresenta dispositivos jurídicos suficientemente necessários para fundamentar tais desígnios. Nesse sentido, será agora apresentada as conclusões a respeito da adequação e da limitação do RGPD frente aos princípios éticos da IA de Confiança e, posteriormente, uma visão conclusiva a respeito do papel e da limitação da AIPD para a regulação de uma IA de Confiança.

Isso posto, a partir dos direitos fundamentais consagrados no artigo 2.º do TUE e do artigo 3º da Carta, o GPAN IA desenvolveu quatro princípios éticos que guiam o quadro para uma IA de Confiança, sendo eles o respeito da autonomia humana; a prevenção de danos; a equidade; e a explicabilidade[542].

O princípio ético referente ao respeito à autonomia humana pretende assegurar a autodeterminação de todo o indivíduo que venha interagir com sistemas de IA. De acordo com esse princípio ético, ao invés de "subordinar, coagir, enganar, manipular, condicionar ou arregimentar injustificadamente"[543], uma IA de Confiança deve ser idealizada para "aumentar, complementar e capacitar as competências cognitivas, sociais e culturais dos seres humanos"[544], principalmente relativas à participação das pessoas no processo democrático[545]. Além disso, ao partir do pressuposto de que, ao utilizar sistemas de IA, nós estamos, em parte, delegando voluntariamente uma parcela do nosso poder de decisão às máquinas, o princípio do respeito da autonomia humana se apresenta como uma busca pela promoção de equilíbrio nessa relação[546], onde "não apenas a autonomia dos humanos deve ser promovida, mas também a autonomia das máquinas deve ser restringida e tornada intrinsecamente reversível, caso a autonomia humana precise ser restabelecida"[547].

Por outro lado, o princípio ético da prevenção de danos baseia-se no respeito à dignidade humana[548] e ao respeito pela integridade física e mental[549] e propõe, de uma maneira geral, que os sistemas de IA não podem afetar negativamente os seres humanos, seja ao causar ou agravar danos físicos ou mentais[550]. Além disso, é importante destacar que o conceito de dano utilizado por esse

[542] GPAN IA. Orientações Éticas para uma IA de Confiança... p. 14.
[543] GPAN IA. Orientações Éticas para uma IA de Confiança... p.15.
[544] Ibid.
[545] GPAN IA. Orientações Éticas para uma IA de Confiança... p.14-15.
[546] Floridi et al. AI4People... p. 698.
[547] Ibid.
[548] Artigo 2.º do TUE.
[549] Artigo 3.º da Carta.
[550] GPAN IA. Orientações Éticas para uma IA de Confiança... p.15.

princípio ético abrange a perspetiva individual e coletiva, assim como a possibilidade de os danos serem intangíveis, afetando o meio ambiente, os ambientes sociais, culturais e políticos.

O princípio ético da equidade, por sua vez, visa garantir que sistemas de IA ofereçam um serviço justo e equitativo. Nesse sentido, uma IA de Confiança deve se afastar de enviesamentos injustos, discriminação e estigmatização contra pessoas e grupos. Além disso, outra característica importante relacionada a equidade é a proporcionalidade necessária para equilibrar os interesses com os objetivos em causa, pelo que "quando diversas medidas concorrem entre si para a consecução de um fim, deve dar-se preferência à que for menos contrária aos direitos fundamentais e às normas éticas"[551]. Por fim, uma última característica da equidade está relacionada com o equilíbrio de forças entre responsável pelo tratamento e titular de dados. Isto posto, uma característica importante a se levar em consideração é a possibilidade de implementar meios de explicabilidade dos processos decisórios e, consequentemente, meio de contestar tais decisões.

Por fim, o último princípio ético é o da explicabilidade, que além de estar ligado estritamente à equidade, complementa e fortalece todos os demais princípios éticos. Afirma-se isso pois a explicabilidade representa a transparência necessária em todas as etapas do tratamento de dados. Isto posto, para que um sistema de IA seja baseado no princípio da explicabilidade, ele deverá promover processos transparentes, seja referente à identidade do responsável pelo tratamento, seja por disponibilização clara e precisa das finalidades dos sistemas. Por fim, um dos desígnios mais importantes desse princípio ético é a possibilidade de que as decisões tomadas por sistemas de IA sejam explicáveis, seja referente às funcionalidades lógicas da decisão, seja pela explicação pela explicação das razões particulares da própria decisão em si.

Como se pode perceber, os princípios éticos acima não só demonstram as características principais que guiam o desenvolvimento um sistema de IA, como também indicam direitos, ações e respostas que são importantes para que se alcance uma IA de Confiança. Isso posto, da análise dos princípios éticos, pode-se perceber ao menos oito importantes características que uma de IA de Confiança desse possuir, sendo eles, em resumo, (i) proteção da autonomia humana contra condicionamento ou manipulação; (ii) buscar equilíbrio entre a autonomia humana e a autonomia das máquinas através da participação humana nos processos decisórios; (iii) prevenção contra danos provenientes dos sistemas de IA, sejam esses danos físicos, mentais ou imateriais; (iv) possibilidade de contestar decisões; (v) a proteção contra os danos tanto de forma individual como coletiva, incluindo os danos em sua dimensão intangível, ampliando assim a proteção aos meio ambiente e aos ambientes sociais, culturais e políticos; (vi) os tratamentos devem ser equitativos e justos, evitando qualquer tipo

[551] Ibid.

de enviesamento injusto, discriminação e estigmatização; (vii) proporcionalidade necessária entre os interesses perseguidos com os objetivos em causa; (viii) possibilidade de que as decisões tomadas por sistemas de IA sejam explicáveis, seja referente às funcionalidades lógicas da decisão, seja pela explicação das razões particulares da própria decisão em si.

Destaca-se que essa lista de características não é exaustiva, e as Orientações éticas apresentam outras que são retiradas dos sete requisitos técnicos necessários para se desenvolver uma IA de Confiança. Contudo, essas oito características podem ser entendidas como um núcleo essencial que o quadro ético espera de uma IA de Confiança. Nesse sentido, o trabalho conseguiu entender até que ponto os princípios e os direitos do RGPD respondem a esses anseios. A resposta, pelo que se pode antecipar, é de que o RGPD apresenta fundamento legal e vinculativo para que os sistemas de IA que tratam dados pessoais já estejam em conformidade, mesmo que parcialmente, com o que se espera de uma IA de Confiança. Contudo, ficaram evidenciadas lacunas no RGPD que, ou promovem esses desígnios parcialmente, ou simplesmente impedem qualquer tipo de proteção.

Nesse sentido, no que se refere às características relacionadas à proteção da autonomia humana, a primeira e melhor resposta que o RGPD possui está relacionada aos princípios relativos ao tratamento de dados (artigo 5.º, n.º 1). Afirma-se isso, pois, em termos gerais, os princípios da transparência, da lealdade (equidade) e da limitação da finalidade, permitem que o titular de dados tome decisões informadas, fator essencial para a autodeterminação informacional e, consequentemente, para a autonomia humana. Esses princípios exigem que o responsável informe ao titular de dados sobre todos os aspetos inerentes ao tratamento de dados, construindo processos decisórios transparentes. Além disso, o responsável é legalmente impedido de realizar um tratamento incompatível com as finalidades apresentadas anteriormente ao titular, restrição essa que confere maior segurança e controle ao titular sobre os seus próprios dados.

Ainda dentro desse contexto, o princípio da lealdade (equidade) obriga os responsáveis pelo tratamento "a atenderem, a todo o tempo, aos interesses e as expectativas legítimas dos titulares de dados"[552]. Esse princípio, por si só, já poderia ser invocado para proteger a autonomia dos titulares, no entanto, Giovanni Sartor ainda apresenta uma dimensão desse princípio chamada de equidade informacional, que determina que "os titulares dos dados não [podem ser] enganados ou induzidos em erro no que diz respeito ao tratamento dos seus dados"[553]. Como se pode perceber, os princípios acima mencionados, impedem que o titular de dados seja, de alguma forma, enganado pelo responsável pelo tratamento, seja antes ou depois do início do tratamento, devendo o responsável ser transparente no que se refere à forma como os dados são tratados e as respetivas finalidades desse tratamento. Tais

[552] Cordeiro, A. Barreto Menezes. Direito da Proteção de Dados, p. 154.
[553] ICO. Guidance on AI and Data Protection... p.44.

informações permitem que o titular de dados tome decisões conscientes, o que respeita e promove a sua autonomia.

Contudo, ainda dentro da proteção da autonomia humana, no que se refere ao equilíbrio entre a autonomia humana e a das máquinas, uma IA de Confiança requer uma maior participação humana nos processos decisórios e, ainda, a possibilidade de contestar as decisões tomadas por sistemas de IA. Sobre esse especto, o artigo 22.º e a proibição de tratamento exclusivamente automatizado se apresentou como uma barreira contra a completa delegação de decisões à sistemas de IA, na medida em que impõe requisitos para que esse tipo de tratamento seja realizado. Além disso, mesmo que o responsável pelo tratamento realize legalmente a tomada de decisões baseadas em tratamento exclusivamente automatizados, o titular ainda tem o direito de solicitar intervenção humana e contestar a decisão. Dessa forma, o artigo 22.º corresponde às expectativas de uma IA de Confiança no que se refere a busca pelo equilíbrio entre a autonomia dos seres humanos e das máquinas.

Isso posto, no que diz respeito a AIPD e a proteção da autonomia humana, a grande resposta que o RGPD possui nessa situação, e que também servirá para os demais princípios, refere-se ao princípio da responsabilidade e a consequente obrigação do responsável pelo tratamento em cumprir e comprovar que o tratamento de dados está sendo realizado em conformidade com o RGPD (artigo 5.º, n.º 2). Nesse sentido, a AIPD será um dos instrumentos disponíveis ao responsável pelo tratamento para operacionalizar essa obrigação, uma vez que pela AIPD será analisada a necessidade e a proporcionalidade do tratamento no sentido de perceber se as finalidades são determinadas, explicitas e legítimas; se os dados são adequados, pertinentes e limitados ao que é necessário; e se o processamento extrapola o seu propósito. Por fim, é também importante que o responsável pelo tratamento apresente razões que fundamentem a escolha pelo tratamento automatizado em detrimento a um meio alternativo menos arriscados e intrusivo (se existir).

Como se pode perceber, toda essa análise restringe a atuação do tratamento de dados à tão somente aquilo que é esperado pelo titular, o que consequentemente protege sua autonomia humana. Isso posto, é possível compreender que o RGPD e a AIPD, no que concerne ao respeito pela autonomia humana, fornece uma resposta jurídica firme, objetiva e profunda para o que se espera de uma IA de Confiança. Pelo que não há evidências diretas de uma necessária adequação legislativa.

Contudo, no que se refere ao princípio ético da prevenção de dano e das características que ele exige, o RGPD apresenta uma resposta insuficiente e parcialmente ineficaz. Para que se entenda essa lacuna existente no RGPD é preciso relembrar que a prevenção de dano de uma IA de Confiança abrange tanto a proteção contra danos físicos, materiais e imateriais, bem como a tutela de danos individuais e coletivos, pelo que se permite a proteção do meio ambiente e de ambientes sociais, culturais e políticos.

Nesse sentido, através das interpretações baseadas nos direitos fundamentais, bem como na posição do GT29, foi constatado que o RGPD protege os titulares de todo e qualquer dano que seja ocasionado pelo tratamento de dados pessoais, independente do dano ser físico, material ou emocional. Contudo, trata-se aqui de uma proteção estritamente vinculada a danos relacionados ao tratamento de dados pessoais. Isso posto, um sistema de IA que não trate dados pessoais não sofrerá qualquer tutela do RGPD se este vier a causar danos ao titular, a respetiva tutela deverá ser analisada e realizada através de outro regulamento setorial.

Além disso, é importante destacar que, tendo em consideração o objeto de proteção do RGPD, infelizmente não foi encontrado nesse regulamento uma interpretação que possibilitasse tutelar danos coletivos. Na verdade, ao longo do RGPD não há sequer menção a esse tipo de proteção, sendo esse regulamento também silente no que se refere a proteção do meio ambiente. Quanto à proteção dos processos democrático, o que se pode afirmar é que o RGPD pode apenas promover uma proteção indireta, isso porque a proteção e o respeito necessário a autonomia humana e a equidade acabam, mesmo que parcial e indiretamente, protegendo os processos democráticos.

Dessa forma, conclui-se que a busca por uma proteção integral contra danos proveniente de sistemas de IA deve ser buscada e aplicada através de outra legislação pertinente. Por outro lado, seria interessante pesquisar mais a fundo sobre como uma alteração legislativa poderia permitir que o RGPD promovesse uma proteção coletiva, que pudesse tanto tutelar danos ao meio ambiente quanto de forma direta os processos democráticos.

Por fim, respeitada as limitações apresentadas pelo RGPD, no que diz respeito o papel da AIPD frente a prevenção de danos, isso deve ser feito nas etapas cinco e seis. Isso porque, na etapa cinco, o responsável pelo tratamento deverá identificar qualquer risco em que se torna possível que os sistemas de IA provoquem algum dano aos titulares. Enquanto na etapa seis, o responsável deverá apresente medidas técnicas e organizativas adequadas para que esses danos sejam mitigados ou eliminados.

Isso posto, no que diz respeito ao próximo princípio ético, temos que a equidade pressupõe o direito de não discriminação. Por falar em discriminação, como foi visto, esse é um dos riscos mais significantes quando se analisa sistemas de IA. Como abordado no primeiro capítulo, os dados são um retrato da sociedade, o que significa que os dados coletados são, naturalmente, enviesados. Nesse sentido, se os dados não forem devidamente analisados antes de se treinar um sistema de IA, decisões injustamente discriminatórias podem afetar significativamente minorias e grupos vulneráveis. Dessa forma, o princípio da equidade exige que sistemas de IA de Confiança decidam de forma justa e equitativa, evitando qualquer tipo de enviesamento injusto, discriminação ou estigmatização. Além disso, a equidade também pressupõe uma análise de proporcionalidade entre os interesses do responsável e os próprios aspetos do tratamento em si.

Nesse sentido, além de respeitar os direitos fundamentais da equidade e da não discriminação, o RGPD também apresenta os princípios relativos ao tratamento de dados que, sob uma interpretação voltada à IA, apresentam uma proteção importante contra decisões discriminatórias. Nesse contexto, pode-se citar, novamente, o princípio da lealdade (equidade), que impõe aos responsáveis pelo tratamento de dados "a obrigação de atenderem, a todo o tempo, aos interesses e as expectativas legítimas dos titulares de dados"[554]. Isso permite concluir que os dados pessoais jamais podem ser usados de forma a provocar efeitos adversos injustificados em seus titulares, o que inclui decisões discriminatórias e injustas baseadas na análise de seus próprios dados. Nesse sentido, o princípio da lealdade se apresenta como o melhor instrumento contra decisões possivelmente discriminatórias e injustas, uma vez que o conceito de lealdade "permite contestar determinados comportamentos que dificilmente poderiam ser descritos como violadores do artigo 6.º, [pois] trata-se de um conceito aberto, passível de ser invocado em situações que contradigam o espírito do RGPD. Essa posição é fortalecida pelo Considerando (71), que afirma que, a fim de assegurar um tratamento equitativo e transparente em um contexto de tratamento automatizado e definição de perfil, o responsável pelo tratamento deverá prevenir efeitos discriminatórios contra pessoas singulares em razão de suas características pessoais ou de posicionamento político, ideológico ou religioso.

Contudo, o RGPD não apresenta somente princípios que podem auxiliar nessa questão. Isso porque, no que se refere a tomada de decisões baseadas em tratamento exclusivamente automatizado, o artigo 22.º, n.º 3, determina que sempre que a tomada de decisão se fundamente juridicamente na execução de um contrato ou no consentimento explícito, o responsável pelo tratamento deve "aplica[r] medidas adequadas para salvaguardar os direitos e liberdades e legítimos interesses do titular dos dados". Afirma-se que esse dispositivo tem a capacidade de combater a discriminação de sistemas de IA por dois motivos. O primeiro, quanto à proteção aos «interesses do titular», uma vez que transcende a doutrina e se consolida na letra da própria lei. O segundo, quanto à proteção dos «direitos e liberdades», que aqui devem ser interpretados em sentido amplo, o que inclui o direito de equidade e não discriminação. Esse entendimento permite que o RGPD seja utilizado como fundamento jurídico para que se exija maior respeito e proteção no tocante aos riscos voltados à autonomia, liberdade, equidade e não discriminação referentes a processos de tomada de decisão exclusivamente automatizada.

No que se refere a análise de proporcionalidade exigida por uma IA de Confiança, essa deve ser realizada ao longo do AIPD. Na verdade, ao longo de toda a quarta etapa da AIPD o responsável deve realizar uma avaliação da necessidade e da proporcionalidade relativa ao tratamento de dados,

[554] Cordeiro, A. Barreto Menezes. Direito da Proteção de Dados, p. 154.

como também relativa às medidas de promoção dos direitos dos titulares. Nesse sentido, a fim de buscar a devida proporcionalidade, o responsável deverá analisar se as finalidades são determinadas, explicitas e legítimas; se os dados são adequados, pertinentes e limitados ao que é necessário; se tratamento respeita as bases legais corretas; se o processamento extrapola o seu propósito; e se a exatidão e a qualidade dos dados estão a ser garantidas. Nesse sentido, é importante perceber que ao mesmo tempo que a AIPD realiza a análise de proporcionalidade, ela também faz uma aferição relativa aos dados que serão processados no tratamento, dando assim executividade aos princípios de tratamento.

Isto posto, a conclusão que se toma é no sentido de que o RGPD e a AIPD possuem um papel importante no combate a decisões injustamente discricionárias, apresentando dispositivos jurídicos que podem atuar eficientemente em todo o ciclo de vida de desenvolvimento de um sistema de IA e que, consequentemente, responde satisfatoriamente os desígnios de uma IA de Confiança.

Por fim, e de maneira geral, o princípio ético da explicabilidade é baseado na transparência e almeja que uma IA de Confiança possibilite que as decisões tomadas por sistemas de IA sejam explicáveis. Porém, é necessário destacar que a explicabilidade esperada se refere tanto às funcionalidades lógicas da decisão, quanto às próprias razões particulares e específicas de cada decisão tomada.

Nesse sentido, esse foi um dos princípios que mais despertou a necessidade de estudo e aprofundamento relativo ao RGPD. Afirma-se isso pois, a partir de uma primeira análise, havia certo entusiasmo em perceber que o RGPD poderia promover as duas dimensões da explicabilidade através do princípio da transparência (artigo 5.º, n.º 1, a). Esse entusiasmo cresceu ainda mais quando se analisou os deveres de informação ativo e passivo do responsável pelo tratamento de dados (artigo 13.º, 14.º e 15.º), que determinavam, de forma explícita, que quando da existência de decisões automatizadas, incluindo a definição de perfis, o responsável pelo tratamento deveria proporcionar informações úteis relativas à lógica da decisão, bem como a importância e as consequências previstas de tal tratamento para o titular dos dados.

Diante desse cenário, a interpretação inicial desses dispositivos levava a crer que 'informações úteis sobre a lógica da decisão' diziam respeito a uma explicação das razões que levaram o sistema de IA a decidir de uma maneira e não de outra. No entanto, o entusiasmo não durou muito tempo, uma vez que a doutrina se posiciona no sentido de que a interpretação temporal e conjunta do RGPD apenas determina que a explicabilidade das decisões tomadas de forma automatizada, incluindo a definição de perfis, se refere apenas a lógica das funcionalidades do sistema de IA, ou seja, uma explicação *ex ante*. Isto posto, percebe-se que o RGPD contempla parcialmente o desígnio de explicabilidade da IA de confiança.

Contudo, é importante que se ressalte que a impossibilidade de fundamentar um pedido de explicabilidade das razões da decisão ocorre pelo posicionamento da doutrina. Dessa forma, ao levar em consideração que esse é um tema muito abordado pela academia, é possível que futuramente a posição majoritária seja alterada, possibilitando explicações mais profundas sobre as decisões emitidas por sistemas de IA. Contudo, ainda existem outras possíveis soluções para essa limitação. Uma delas dependeria diretamente de uma decisão judicial do Tribunal de Justiça da União Europeia reconhecendo o direito de explicação das razões da decisão automatizada. Essa alternativa possui certa força pois o TJUE já reconheceu outros direitos relativos à proteção de dados, como é o caso do direito ao esquecimento. Por outro lado, outra solução dependeria de uma alteração legislativa do RGPD, a fim de promover uma redação mais clara e objetiva acerca do presente tema.

Ultrapassada a comparação entre os princípios éticos e os limites do RGPD e da AIPD no desenvolvimento e regulação de uma IA de Confiança, chegou-se ao momento de apresentar as conclusões gerais a respeito da AIPD, que como mencionado, é um instrumento ligado ao princípio da responsabilidade (artigo 5.º, n.º 2) e que permitirá o responsável pelo tratamento demonstrar sua conformidade com o RGPD[555].

E essa conclusão deve-se iniciar por uma observação refere-se ao posicionamento adotado por esse trabalho no sentido de que a AIPD tem a capacidade de servir, simultaneamente, às perspetivas jurídicas do RGPD, ou seja, a abordagem baseada em riscos e a abordagem baseada em direitos. Nesse sentido, a partir da interpretação dada aos dispositivos desse regulamento, somada a uma metodologia profunda e protetiva relacionada a IA, a AIPD se torna um poderoso instrumento de gestão e avaliação dos riscos capaz, simultaneamente, de proteger e prevenir os direitos tutelados pelo RGPD. Dessa forma, mais do que demonstrar conformidade com o RGPD, a utilização da AIPD em sistemas de IA pode ser um passo de operacionalização de uma IA de Confiança.

Essa afirmação tem sido comprovada não só pela forma como a AIPD possibilita a análise e a mitigação de riscos aos direitos e liberdades defendidas tanto pela IA de Confiança quanto pelo RGPD, mas também pelo fato de que a AIPD possibilita a operacionalização de alguns desses mesmos direitos. Afirma-se isso pois, após a realização da AIPD, o seu conteúdo tem a capacidade de demonstrar responsabilidade e transparência acerca de todo o tratamento de dados realizado por um sistema de IA. No sentido de que essa transparência, ao mesmo tempo que responde de forma eficiente aos deveres de informação encontrados nos artigos 13.º, 14.º e 15.º do RGPD, também se apresenta como uma resposta, mesmo que parcial, para os anseios de uma IA de Confiança.

[555] ICO. Guidance on AI and Data Protection... p.13.

Contudo, por mais que a AIPD tenha se mostrado um elemento importante do RGPD e que pode ser usado para se alcançar uma IA de Confiança, ela não está livre de críticas. Aliás, a falha encontrada a respeito desse instrumento é tão significante que tem a capacidade de colocar todos os esforços dessa pesquisa abaixo. Afirma-se isso pois não adianta realizar toda uma interpretação protetiva do RGPD à luz dos desígnios éticos de uma IA de Confiança, tampouco realizar uma análise profunda dos riscos relativos aos direitos e liberdades dos titulares, se após a execução de uma AIPD não houver qualquer tipo de obrigatoriedade na publicação, mesmo que parcial, de seus resultados. Como já abordado, não se quer que o responsável pelo tratamento revele segredos e informações comerciais, mas que disponibilize tão somente um resumo das conclusões. Esse resumo não só possibilitaria que a sociedade em geral pudesse ter acesso a maiores informações sobre o tratamento de dados em questão, mas também permitiria a execução de um escrutínio sobre o respetivo sistema de IA, de forma a fortalecer as perspetivas de transparência, equidade e autonomia informacional e humana.

Isto posto, por mais que esse trabalho acadêmico entenda que o RGPD apresente uma proteção jurídica suficiente para que a União Europeia possa dar os primeiros passos no sentido de construir uma IA de Confiança, e que a AIPD possa ser o instrumento técnico de operacionalizar esse desígnio, nada disso será possível se não houver uma maneira de legalmente obrigar a publicação de um resumo das conclusões da AIPD. E para que isso seja possível, finaliza-se esse trabalho indicando que é importante que se inicie um estudo aprofundando sobre os benefícios e as consequências em tornar obrigatória a publicação dos resultados de uma AIPD, a fim, assim, de convencer os órgãos legislativos competentes a debater sobre esse tema.

BIBLIOGRAFIA

Abreu, Joana Covelo de et al. O Contencioso da União Europeia e a cobrança transfronteiriça de créditos: compreendendo as soluções digitais à luz do paradigma da Justiça eletrónica europeia (e-Justice). UNIO, 2020, p.4. DOI: 10.21814/1822.65807

Alpaydin, Ethem. Machine Learning, the new AI. The MIT press essencial knowledge series. 2016 Massachusetts Institute of Technology.

Anand, Avishek, Bizer, Kilian, Erlei, Alexander et al. Effects of Algorithmic Decision-Making and Interpretability on Human Behavior: Experiments using Crowdsourcing, 2018,

Anastácio, Manuel Lopes Porto Gonçalo. Tratado de Lisboa - Anotado e Comentado. Almedina. Edição do Kindle

Andrade F., Novais P., Carneiro D., Zeleznikow J., Neves J., Using BATNAs and WATNAs in Online Dispute Resolution, in New Frontiers in Artificial Intelligence, Kumiyo Nakakoji, Yohei Murakami and Eric McCready (Eds), Springer - Lecture Notes in Artificial Intelligence 6284, ISBN 978-3-642-14887-3, pp 5-18, 2010. Disponível: http://dx.doi.org/10.1007/978-3-642-14888-0_2.

Angwin, J. et al. Machine Bias. ProPublica, 2016. Disponível: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Anitha, P, Krithka, G. e Choudhry, Mani Deepak. Machine Learning Techniques for learning features of any kind of data. A Case Study. International Jour-nal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3, Issue 12, 2014,

Aragão, Alexandra. Questões ético-jurídicas relativas ao uso de apps geradoras de dados de mobilidade para vigilância epidemiológica da Covid-19. Uma perspetiva Europeia. Instituto Jurídico da Faculdade de Direito da Universidade de Coimbra.

Barrett, Lindsey, Reasonably Suspicious Algorithms: Predictive Policing at the United States Border, N.Y.U. Review of Law & Social Change, 2017, vol 41. Disponível em: https://socialchangenyu.com/wp-content/uploads/2017/09/barrett_digital_9-6-17.pdf

Barto, A. G. & Sutton, R. Introduction to Reinforcement Learning. Second edition. (2014-2015). The MIT Press.

Bezerra, Eduardo. (2016). Introdução à Aprendizagem Profunda. Edition: 1, Chapter: 3, Publisher: SBC, Editors: Ogasawara.

Bioni, Bruno Ricardo e Luciano, Maria. O Princípio da Precaução na Regulação de Inteligência Artificial: seriam as Leis de Proteção de Dados o seu portal de entrada? Inteligência Artificial e Direito, 2ª edição. Revista dos Tribunais, 2020.

Bostrom, Nick. Superintelligence: Paths, dangers, strategies. Oxford: Oxford University Press, 2014.

Brkan, M. AI-supported decision-making under the general data protection regulation. Proceedings of the 16th edition of the International Conference on Artificial Intelligence and Law. 2017.

Brkan, Maja. Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond. International Journal of Law and Information Technology, 2019.

Cabral, Tiago Sérgio, AI Regulation in the European Union: Democratic Trends, Current Instruments and Future Initiatives (Dissertação de Mestrado: Universidade do Minho, 2019).

Cabral, Tiago Sérgio. AI Regulation in the European Union: Democratic Trends, Current Instruments and Future Initiatives (Dissertação de Mestrado: Universidade do Minho, 2019).

Carneiro D., Novais P., Andrade F., Zeleznikow J., Neves J., The Legal Precedent in Online Dispute Resolution, in Legal Knowledge and Information Systems, ed. Guido Governatori (proceedings of the Jurix 2009 - the 22nd Interna-tional Conference on Legal Knowledge and Information Systems, Rotterdam, The Netherlands), IOS press, ISBN 978-1-60750-082-7, pp 47–52, 2009. Disonível: https://dl.acm.org/doi/10.5555/1671082.1671089.

Carneiro D., Novais P., Andrade F., Zeleznikow J., Neves J., Using Case Based Reasoning and Principled Negotiation to provide Decision Support for Dispute Resolution, Knowledge and Information Systems Journal, Springer, ISSN: 0219-1377, Vol 36, Issue 3, pp 789-826, 2013. Disponível: http://dx.doi.org/10.1007/s10115-012-0563-0.

Carneiro D., Novais P., Neves J., Conflict Resolution and its Context. From the Analysis of Behavioural Patterns to Efficient Decision-Making, Springer-Verlag, 279 pages, ISBN: 978-3-319-06238-9, 2014. Disponível: http://dx.doi.org/10.1007/978-3-319-06239-6

Chinese National Governance Committee for the New Generation Artificial Intelligence. Governance Principles for the New Generation Artificial Intelligence–Developing Responsible Artificial Intelligence. 2019. Disponível: https://www.chinadaily.com.cn/a/201906/17/WS5d07486ba3103dbf14328ab7.html.

Christl, Wolfie, Corporate Surveillance in Everyday Life. How Companies Collect, Combine, Analyze, Trade, and Use Personal Data on Billions. A Report by Cracked Labs, June 2017, https://crackedlabs.org/dl/CrackedLabs_Christl_CorporateSurveillance.pdf

Citron, Danielle Keats e Pasquale, Frank A. Pasquale, The Scored Society: Due Process for Automated Predictions, 2014 vol 89 Washington Law Review.

Clarke, R. Profiling: a hidden challenge to the regulation of data surveillance. Journal of Law and Information Science, Australia, v. 4, n. 2, December. 1993. Available at: http://www.austlii.edu.au/au/journals/JlLawInfoSci/1993/26.htmlD; Magrani, Eduardo. Entre dados e robôs "Ética e privacidade na era da hiperconectividade". Publisher Arquipélago. 2019. ISBN 978-85-5450-029-0

Coeckelbergh, Mark. Ética da IA MIT Press Essential Knowledge. Edição do Kindle.

Comissão Europeia. Comunicação da Comissão ao Parlamento Europeu, ao Conselho Europeu, ao Conselho, ao Comité Económico a Social e ao Comité Das Regiões: Aumentar a confiança numa inteligência artificial centrada no ser humano. COM (2019) 168 final.

Comissão Europeia. Comunicação da Comissão ao Parlamento Europeu, ao Conselho Europeu, ao Conselho, ao Comité Económico a Social e ao Comité Das Regiões: Inteligência artificial para a Europa. COM (2018) 237 final, Bruxelas.

Comissão Europeia. Comunicação da Comissão ao Parlamento Europeu, ao Conselho Europeu, ao Conselho, ao Comité Económico a Social e ao Comité Das Regiões: Plano Coordenado para a Inteligência Artificial. COM (2018) 795 final. Bruxelas.

Comissão Europeia. Livro Branco sobre a inteligência artificial - Uma abordagem europeia virada para a excelência e a confiança. COM (2020) 65 final. Bruxelas, 2020, p.2.

Comissão Europeia. Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain union legislative acts. Brussels, 21.4.2021 COM (2021) 206 final 2021/

Conselho Europeu. Conclusões do encontro de 19 de outubro de 2017. Bruxelas. Disponível: http://data.consilium.europa.eu/doc/document/ST-14-2017-INIT/en/pdf

Cordeiro, A. Barreto Menezes. Direito da Proteção de Dados. Almedina, 2020

Cortiz, Diogo. O Design pode ajudar na construção de Inteligência Artificial humanística? 17° ERGODESIGN – Congresso Internacional de Ergonomia e Usabilidade de Interfaces Humano Tecnológica: Produto, Informações Ambientes Construídos e Transporte. Disponível: http://pdf.blucher.com.br.s3-sa-east-1.amazonaws.com/designproceedings/ergodesign2019/1.02.pdf

Dastin, Jeffrey. Amazon scraps secret AI recruiting tool that showed bias against women. Reuters, 2018. Disponível: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruting-tool-that-show-bias-against-women-idUSKCN1MK08G

Data Science Academy. Deep Learning Book, 'Cap. 62 - O que é Aprendizado por Reforço?' Disponível em «http://deeplearningbook.com.br/o-que-e-aprendizagem-por-reforco/». Acedido em 25/07/2020.

DeBarr, David e Harwood, Maury, Relational Mining for Compliance Risk, Presented at the Internal Revenue Service Research Conference, 2004, p. 178-179. Disponível em: http://www.irs.gov/pub/irs-soi/04debarr.pdf

Declaração de Cooperação sobre a Inteligência Artificial. Disponível: https://ec.europa.eu/digital-single-market/en/news/eu-member-states-sign-cooperate-artificial-intelligence

Domingos, Pedro. O Algoritmo Mestre: como a busca pela melhor máquina de aprendizado refaz o nosso mundo. Novatec Editora, 2015, p. 46

Doneda, Danilo. Da Privacidade à Proteção de Dados Pessoais, Rio de Janeiro: Renovar. 2006, p.173

Drexl, Hilty et al., Technical Aspects of Artificial Intelligence: An Understanding from an Intellectual Property Law Perspective, Version 1.0, 2019. Disponível: https://ssrn.com/abstract=3465577

Ebers, Martin, Chapter 2: Regulating AI and Robotics: Ethical and Legal Challenges (April 17, 2019). Martin Ebers/Susana Navas Navarro (eds.), Algorithms and Law, Cambridge, Cambridge University Press, 2019, Available at SSRN: https://ssrn.com/abstract=3392379 or http://dx.doi.org/10.2139/ssrn.3392379; European Data Protection Supervisor (EDPS), Opinion 3/2018 on online manipulation and personal data, March 19, 2018.

European Data Protection Supervisor (EDPS). Opinion 3/2018 on online manipulation and personal data. 2018, p.7.

European Parliament. Artificial intelligence: From ethics to policy. EPRS | European Parliamentary Research Service. Scientific Foresight Unit (STOA). PE 641.507 – June 2020

Evas, Tatjana. European framework on ethical aspects of artificial intelligence, robotics and related technologies. European Parliament. 2019, p.6.

Federal Trade Commission. Internet of Things: privacy & security in a connected world. FTC Staff Report, 2015

Floridi, L., Cowls, J., Beltrametti, M., et al. AI4People: An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 2018, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Floridi, Luciano. Establishing the rules for building trustworthy AI. Nature Machine Intelligence, 2019, 1(6), 261–262. https://doi.org/10.1038/s42256-019-0055-y

Future of Life Institute. Asilomar AI Princples. 2017. Disponível: https://futureoflife.org/ai-principles/

Gellert, Raphaël. We Have Always Managed Risks in Data Protection Law: Understanding the Similarities and Differences Between the Right-Based and the Risk-Based Approaches to Data Protection, 2 EDPL, 2016, 481-492.

Gillespie, Tarleton. "The relevance of algorithms". In (Ed.), Media Technologies: Essayson Communication, Materiality, and Society. Cambridge: The MIT Press, 2014.

Goertzel, Ben. Artificial General Intelligence: Concept, State of the Art, and Future Prospects. OpenCog Foundation. Journal of Artificial General Intelli-gence, 2014, DOI: 10.2478/jagi-2014-0001.

Grupo de Peritos de Alto Nível em IA (GPAN IA). Orientações Éticas para uma IA de Confiança. 2019

Grupo de Peritos de Alto Nível em IA. Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment. Disponível: https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

Grupo de Peritos de Alto Nível sobre a Inteligência Artificial. Uma definição de IA: Principais capacidades e Disciplinas Científicas, Bruxelas.

Grupo de Trabalho do Artigo 29 (GT29), Opinion 3/2010 on the principle of accountability.

Grupo de Trabalho do Artigo 29 (GT29). Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD) e que determinam se o tratamen-to é «suscetível de resultar num elevado risco» para efeitos do Regulamento (UE) 2016/679, abril de 2017.

Grupo de Trabalho do Artigo 29 (GT29). Orientações sobre as decisões individuais automatizadas e a definição de perfis para efeitos do Regulamento (UE) 2016/679, fevereiro de 2018.

Harari, Y. N. (2016). Homo deus: A brief history of tomorrow. New York: Harper

Hijmans, H. e Raab, CD. Ethical Dimensions of the GDPR, in M. Cole and F. Boehm (eds.), Commentary on the General Data Protection Regulation, Edward Elgar, 2018.

Information Commissioner's Office (ICO). Guide to the General Data Regulation (GDPR). May 2019

Information commissioner's Office (ICO). Guidance on AI and Data Protection. Version 0.0.22. July 2020

Information commissioner's Office (ICO). Guidelines on Data protection impact assessments. Version 1.0.124. May 2018

Kant, Immanuel. A fundamentação da Metafísica dos Costumes. Lisboa, Portugal: Edições 70, 2011.

Kitchin, Rob. Thinking critically about and researching algorithms. Information, Communication & Society, 2017 vol. 20, n. 1, 14–29. Disponível:http://futuredata.stanford.edu/classes/cs345s/handouts/kitchin.pdf

Kuner, Christopher, Bygrave, Lee A., et al. The EU General Data Regulation (GDPR) A Commentary. Oxford University Press 2020.

Kurzweil, R. The singularity is near: When humans transcend biology. Penguin, 2005.

Landgrebe, Jobst & Smith, Barry (2020). There is no general AI: Why Turing machines cannot pass the Turing test.

Luger, George F. Inteligência Artificial. 6ª Edição. 2013. Ed. Pearson Universidades. ISBN: 9788581435503

Magrani, Eduardo. Entre dados e robôs: ética e privacidade na era da hiperconectividade. 2. edição. Porto Alegre: Arquipélago Editorial, 2019, p.24-25.

Maine, Vishal e Sabri, Samer. Machine Learning for Humans. 2017. Disponível: «https://everythingcomputerscience.com/books/Machine%20Learning%20for%20Humans.pdf».

Malgieri, Gianclaudio e Kaminski, Margot E. Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations;

Martins R., Gomes M., Almeida JJ, Novais P., Henriques P., Hate Speech Classification in social media Using Emotional Analysis, 7th Brazilian Conference on Intelligent Systems (BRACIS), IEEE, 2018, p.61. Disponível: https://doi.org/10.1109/BRACIS.2018.00019.

McCarthy, John. What is Artificial Intelligence? Basic Question. Computer Science Department, 2007. Disponível: http://www-formal.stanford.edu/jmc/whatisai/ .

McCrudden, C. Human Dignity and Judicial Interpretation of Human Rights, EJIL, 19(4), 2008

Mendoza, Isak e Bygrave, Lee A. 'The Right not to be Subject to Automated Decisions based on Profiling' University of Oslo Faculty of Law Legal Studies Research Paper Series No 20/2017(n 31)

Mik, Eliza. The Erosion of Autonomy in Online Consumer Transactions. (2016). Law, Innovation and Technology. 8, (1), 1-38. Rikesearch Collection School Of Law.

Miyazaki, Shintaro. Algorhythmics: Understanding micro-temporality in computational cultures, Computational Culture, 2012 Disponível: http://computationalculture.net/algorhythmics-understanding-micro-temporality-in-computational-cultures

Müller, Vincent C. (2014) Risks of general artificial intelligence, Journal of Experimental & Theoretical Artificial Intelligence, 26:3, 297-301, DOI: 10.1080/0952813X.2014.895110

Müller, Vincent C. 'Ethics of artificial intelligence', in Anthony Elliott (ed.), The Routledge social science handbook of AI (London: Routledge), 2021.

Müller, Vincent C. and Bostrom, Nick. Future progress in artificial intelligence: A survey of expert opinion. Fundamental Issues of Artificial Intelligence. Springer, 2016, 553-571

OECD. Principles on AI. 2019. Disponível: https://www.oecd.org/going-digital/ai/principles/.

Oliveira, Anderson Castro Soares. Aplicação de redes neurais artificiais na previsão da produção de álcool. Março, 2010. Pág. 4. DOI: 10.1590/S1413-70542010000200002

Pagno, Luana. A Dignidade Humana em Kant, p.225.

Pasquale, Frank. The black box society: the secret algorithms that control money and information. Cambridge: Harvard University Press, 2015; Explainable AI: the basics Policy briefing Issued: November 2019 DES6051. The Royal Society

Russell, Stuart e Norvig, Peter. Inteligência Artificial: uma abordagem moderna, ed. Elsevier Editora Ltda., Inteligência Artificial, 3a Edição (Rio de Janeiro, 2013).

Sartor, Giovani. The impact of the General Data Protection Regulation (GDPR) on artificial intelligence. EPRS | European Parliamentary Research Service Scientific Foresight Unit (STOA) PE 641.530. 2020.

Scherer, Matthew U. Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, And Strategies. Harvard Journal of Law & Technology Volume 29, Number 2 Spring 2016.

Silveira, Alessandra e Froufe, Pedro. Do mercado interno à cidadania de direitos: a proteção de dados pessoais como a questão jusfundamental identitária dos nossos tempos. UNIO - EU LAW JOURNAL Vol. 4, No. 2, julho 2018.

Silveira, Alessandra e Marques, João. Do direito a estar só ao direito ao esquecimento. Considerações sobre a proteção de dados pessoais informatizados no direito da união europeia: sentido, evolução e reforma legislativa.

Silveira, Alessandra, Abreu, Joana Covelo de e Cabral, Tiago Sérgio. Breves apontamentos quanto aos direitos dos titulares de dados no RGPD. CEJ - Centro de Estudos Judiciários. 2020, "no prelo"

Silveira, Alessandra, Canotilho, Mariana e Froufe, Pedro Madeira. Direito da União Europeia – Elementos de Direito e Políticas da União. Almedina, 2016.

Silveira, Alessandra. "Comentário ao art. 51.º", in Carta dos Direitos Fundamentais da União Europeia Comentada, Alessandra Silveira/Mariana Canotilho (coords.), Almedina, Coimbra, 2013.

Silveira, Alessandra. União de direito e ordem jurídica da União Europeia. Revista Eletrônica Direito e Política, Programa de Pós-graduação Stricto Sensu em Ciência Jurídica da UNIVALI, Itajaí, v.3, n.3 3º quadrimestre de 2008.

Silver, D., Huang, A., Maddison, C. et al. Mastering the game of Go with deep neural networks and tree search. Nature 529, 484–489 (2016). https://doi.org/10.1038/nature16961

Stix. Charlotte. A survey of the European Union's artificial intelligence ecosystem. 2019. Disponível: https://ec.europa.eu/jrc/communities/sites/default/files/ff3afe_1513c6bf2d81400eac182642105d4d6f.pdf

The impact of the General Data Protection Regulation (GDPR) on artificial intelligence. EPRS | European Parliamentary Research Service Scientific Fore-sight Unit (STOA) PE 641.530 – June 2020

Thiebes, S., Lins, S. & Sunyaev, A. Trustworthy artificial intelligence. Electron Markets (2020). https://doi.org/10.1007/s12525-020-00441-4.

Université de Montréal. Montreal Declaration for a Responsible Developmen t of AI. 2017. Disponível: https://www.montrealdeclaration-responsibleai.com/the-declaration.

Vought, Russel T. Guidance for Regulation of Artificial Intelligence Applications. Executive Office of the President of United States of America. 2020. Disponível: https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf

Wachter, S., Mittelstadt, B., e Floridi, L. Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. International Data Privacy Law 7, 76–99, 2016.

Recursos em linha:

Cambridge Dictionary, "Inteligence", acedido 19 de Setembro de 2019, https://dictionary.cambridge.org/pt/dicionario/ingles/intelligence

Cf. o discurso de Alexander Nix, ex-CEO da Cambridge Analytica, na Concordia Annual Summit 2016 em Nova York, https://www.youtube.com/watch?v=n8Dd5aVXLCc;

Chatterjee, Marina. Top 20 Applications of Deep Learning in 2020 Across Industries. Great learning. 2019. Disponível: https://www.mygreatlearning.com/blog/deep-learning-applications/.

EUR-Lex. Glossary of sumaries. White Paper. https://eur-lex.europa.eu/summary/glossary/white_paper.html

Hawkins, Andrew J., Elon Musk thinks humans need to become cyborgs or risk irrelevance. The Verge, 2017. Disponível em «https://www.theverge.com/2017/2/13/14597434/elon-musk-human-machine-symbiosis-self-driving-cars».

https://standards.ieee.org/industry-connections/ec/autonomous-systems.html

https://www.iso.org/committee/6794475.html.

https://www.weforum.org/agenda/2018/09/7-amazing-ways-artificial-intelligence-is-used-in-healthcare.

Iyengar, Vinod. Why AI Consolidation Will Create the Worst Monopoly in U.S. History, TECHCRUNCH, 2016. Disponível: https://techcrunch.com/2016/08/24/why-ai-consolidation-will-create-the-worst-monopoly-in-us-history/

J., Vincent. AI systems should be accountable, explainable, and unbiased, says EU. The Verge https://www.theverge.com/2019/4/8/18300149/eu-artificial-intelligence-ai-ethical-guidelines-recommendations (2019).

Merriam-webster, "Intelligence", acedido 19 de setembro de 2019, <https://www.merriam-webster.com/dictionary/intelligence?utm_campaign=sd&utm_medium=serp&utm_source=jsonld>.

Norvig, quoted by Scott Cleland, Google's "Infringenovation" Secrets, Forbes, October 3, 2011, https://www.forbes.com/sites/scottcleland/2011/10/03/googles-infringenovation-secrets/#78a3795430a6.

Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. The Guardian. Disponível: https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.

Significado de fairness através do dicionário Cambridge. Disponível em: https://dictionary.cambridge.org/pt/dicionario/ingles-portugues/fairness

The Atlantic. The Dark Side of That Personality Quiz You Just Took, 2020. https://www.theatlantic.com/technology/archive/2017/07/the-internet-is-one-big-personality-test/531861/.

The Great Hack. Documentário. Netflix. 2019

The Social Dilemma. Documentário. Netflix, 2020