PORTUGAL E A REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL: ENTRE O RGPD, O *AI ACT* E A COOPERAÇÃO INTERNACIONAL PARA UMA ÉTICA ALGORÍTMICA GLOBAL

Tiago Matos*

Sumário: 1. Fundamentação Teórica. 1.1. Contexto Histórico. 1.2. Princípios Constitucionais Portugueses. 2. Contexto Jurídico Atual em Portugal. 2.1. Regulamento Geral sobre a Proteção de Dados (RGPD). 2.2. Plano Nacional para a Inteligência Artificial (2025-2030). 2.3. Conexão entre RGPD e Plano Nacional para a IA. 3. Desafios Éticos e Jurídicos. 3.1. Vieses Algorítmicos e Discriminação. 3.2. Responsabilidade Civil por Decisões Autónomas. 3.3. Privacidade e Consentimento. 4. Perspetivas Futuras e Recomendações. 4.1. Flexibilidade Regulatória. 4.2. Cooperação Internacional. 5. Conclusão.

Table of Contents: 1. Theoretical Basis. 1.1. Historical Context. 1.2. Portuguese Constitutional Principles. 2. Current Legal Context in Portugal. 2.1. General Data Protection Regulation (GDPR). 2.2. National Plan for Artificial Intelligence (2025-2030). 2.3. Connection between GDPR and National Plan for AI. 3. Ethical and Legal Challenges. 3.1. Algorithmic Biases and Discrimination. 3.2. Civil Liability for Autonomous Decisions. 3.3. Privacy and Consent. 4. Future Prospects and Recommendations. 4.1. Regulatory Flexibility. 4.2. International Cooperation. 5. Conclusion.

Resumo: A integração da inteligência artificial (IA) na sociedade portuguesa enfrenta desafios éticos e jurídicos críticos, desde a salvaguarda dos direitos fundamentais até à responsabilidade civil por decisões automatizadas. Este artigo analisa o contexto histórico e jurídico atual em Portugal, destacando

* Solicitador, Mestrando em Direito, na vertente das Ciências Jurídico-Empresariais na Universidade Lusófona do Porto, a frequentar a Pós-Graduação em Direito Imobiliário na CIDP/FDUL, tiagojosematos@gmail.com.

a influência do Regulamento Geral sobre a Proteção de Dados (RGPD) e do Plano Nacional para a IA (2025–2030) na promoção de transparência, explicabilidade e não discriminação. Discutem-se problemas como os vieses algorítmicos, a indefinição da responsabilidade civil em sistemas autónomos e a complexidade do consentimento informado, propondo soluções como auditorias independentes, regimes de responsabilidade objetiva e *sandboxes* regulatórias. O estudo conclui que a governança da IA exige uma abordagem multidisciplinar, alinhada com os princípios constitucionais e com a cooperação internacional, para garantir um desenvolvimento ético e inclusivo.

Abstract: The integration of artificial intelligence (AI) into Portuguese society faces critical ethical and legal challenges, from safeguarding fundamental rights to civil liability for automated decisions. This article examines the historical and legal context in Portugal, highlighting the influence of the General Data Protection Regulation (GDPR) and the National AI Plan (2025–2030) in promoting transparency, explainability, and non-discrimination. It addresses issues such as algorithmic bias, the ambiguity of liability for autonomous systems, and the complexity of informed consent, proposing solutions like independent audits, objective liability regimes, and regulatory sandboxes. The study concludes that AI governance requires a multidisciplinary approach aligned with constitutional principles and international cooperation to ensure ethical and inclusive development.

Palavras-chave: Inteligência Artificial; Ética Algorítmica; Responsabilidade Civil; RGPD; Viés Algorítmico; Plano Nacional para a IA.

Keywords: Artificial Intelligence; Algorithmic Ethics; Civil Liability; GDPR; Algorithmic Bias; National AI Plan.

1. Fundamentação Teórica

1.1. Contexto Histórico

A reflexão ética sobre a inteligência artificial (IA) não é recente. Desde o século XX, com o avanço da computação e da automação, emergiram preocupações quanto à responsabilidade e à transparência em decisões tomadas por sistemas automatizados. Autores como Lima Vaz advertiram para a tendência da sociedade contemporânea em "mitificar" a tecnologia, desconsiderando os riscos éticos e sociais associados ao seu uso indiscriminado. Esta crítica revela a necessidade de uma abordagem prudente e filosófica face às promessas da técnica, especialmente quando esta impacta diretamente a vida humana[188].

Durante as décadas de 1970 e 1980, começaram a surgir os primeiros quadros teóricos no domínio da ética aplicada à IA, centrando-se na responsabilização pelas decisões algorítmicas e na prevenção de discriminações sistémicas introduzidas por dados enviesados. Esta linha de pensamento evoluiu para incorporar preocupações mais amplas, como os riscos existenciais da IA geral e a preservação dos direitos humanos em contextos automatizados.[189]

Em Portugal, o debate sobre a regulação da IA tem vindo a ganhar relevância em especial, e mais importância após a entrada em vigor do Regulamento Geral sobre a Proteção de Dados (RGPD) em 2018, e com a preparação para a aplicação do Regulamento Europeu da Inteligência Artificial (AI Act). A Lei n.º 58/2019, que assegura a execução do RGPD na ordem jurídica nacional, introduz limites significativos à tomada de decisões exclusivamente automatizadas que produzam efeitos jurídicos relevantes sobre os indivíduos. Este enquadramento reflete uma continuidade histórica na preocupação do legislador português com a salvaguarda dos direitos fundamentais face à inovação tecnológica.

1.2. Princípios Constitucionais Portugueses

A Constituição da República Portuguesa (CRP) constitui a matriz normativa fundamental para qualquer reflexão sobre ética e regulação da IA no ordenamento jurídico nacional. Os seus

[188] Sousa e Silva destaca a necessidade de transparência em sistemas algorítmicos. Ver: Sousa e Silva, Nuno (2019). *Inteligência artificial, robots e responsabilidade civil: O que é que é diferente?*. *In* Revista de Direito Civil, vol. 4, n.º 4, pp. 691-711.

[189] Como demonstram Wolff, Lehr e Yoo no seu estudo sobre as lições do RGPD para as políticas de IA, "[...] o ChatGPT não apenas capturou a imaginação pública; também está amplificando preocupações em toda a indústria, academia e decisores políticos governamentais interessados na regulação da Inteligência Artificial". Ver: Wolff, J.; Lehr, W.; Yoo, C. S. (2023). *Lessons from GDPR for AI Policymaking*. *In* Virginia Journal of Law & Technology, vol. 27, n.º 4.

princípios estruturantes impõem limites materiais ao desenvolvimento e utilização de sistemas inteligentes, devendo orientar a legislação infraconstitucional e a implementação de tecnologias públicas e privadas.

- **Artigo 1.º (República Portuguesa)** – A dignidade da pessoa humana é o valor fundacional da CRP. Sistemas de IA que incorporem ou reproduzam discriminações (ex.: preconceitos raciais, de género ou socioeconómicos) colidem com este princípio[190/191], exigindo mecanismos de auditoria, explicabilidade e correção ética.
- **Artigo 2.º (Estado de Direito Democrático)** – A exigência de segurança jurídica, um dos pilares do Estado de direito, requer que os sistemas de IA sejam compreensíveis e que as suas decisões sejam passíveis de escrutínio e contestação.[192] Esta exigência liga-se ao artigo 22.º do RGPD, que estabelece o direito do titular dos dados a não ficar sujeito a decisões exclusivamente automatizadas com efeitos jurídicos.
- **Artigo 13.º (Princípio da Igualdade)** – Todos os cidadãos têm a mesma dignidade social e são iguais perante a lei. Este preceito impõe o dever de garantir que os algoritmos usados, por exemplo, na justiça ou na saúde, não gerem desigualdades de tratamento injustificadas. A prevenção do chamado *racismo algorítmico* e outras formas de discriminação indireta é um imperativo constitucional.[193]
- **Artigo 21.º (Direito de Resistência e Defesa dos Direitos)** – Este artigo, que consagra o direito de resistir a qualquer ordem que atente contra os direitos fundamentais, reforça a obrigação de garantir que os sistemas de IA respeitam as liberdades individuais. No contexto de serviços públicos digitais, como plataformas de educação ou assistência social, torna-se essencial garantir acessibilidade e neutralidade linguística, assegurando o pleno exercício dos direitos por todos os cidadãos, incluindo os de língua portuguesa residente.

1.3. Conexão entre História e Direito Português

A evolução do debate ético sobre a IA tem convergido para a criação de marcos normativos internacionais e europeus, como o AI Act da União Europeia, que classifica os sistemas de IA com base no grau de risco (mínimo, limitado, elevado e inaceitável) e impõe obrigações reforçadas aos sistemas de alto risco.[194] Em

[190] O princípio da igualdade (CRP, Art. 13.º) exige mecanismos de auditoria algorítmica para mitigar vieses discriminatórios, tema explorado por Capellà Ricart (2024) e Grozdanovski (2021). Ver: Capellà Ricart, A. (2024). *The role of European equality bodies to address algorithmic discrimination*. *In* International Journal of Discrimination and the Law, vol. 24, n.º 3, pp. 194-216; Grozdanovski, L. (2021). *In search of effectiveness and fairness in proving algorithmic discrimination in EU law. In* Common Market Law Review, vol. 58, n.º 1, pp. 99-136.

[191] Como demonstra Chen no seu estudo sobre ética e discriminação em práticas de recrutamento baseadas em IA, a discriminação algorítmica pode manifestar-se de formas subtis, mas sistemáticas, afetando desproporcionalmente grupos vulneráveis. Ver: Chen, Z. (2023). *Ethics and discrimination in artificial intelligence-enabled recruitment practices. In* Humanities and Social Sciences Communications, vol. 10, article number 567.

[192] Como argumenta Sartor na sua análise sobre explicabilidade algorítmica e raciocínio jurídico, a explicabilidade não é apenas uma questão técnica, mas um requisito fundamental para a legitimidade democrática das decisões automatizadas. Ver: Sartor, G. (2022). *Algorithmic explainability and legal reasoning.* AI & Society. https://www.tandfonline.com/doi/abs/10.1080/20508840.2022.2033945

[193] O relatório da Agência dos Direitos Fundamentais da União Europeia sobre viés em algoritmos alerta que os algoritmos podem reproduzir e amplificar discriminações existentes na sociedade, criando novos padrões de exclusão que afetam desproporcionalmente grupos vulneráveis. Ver European Union Agency for Fundamental Rights (2022). *Bias in algorithms – Artificial intelligence and discrimination*. Vienna, p. 15 (https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf).

[194] Como analisa Almada no seu estudo sobre o *AI Act* numa perspetiva global, o *AI Act* da União Europeia representa um dos primeiros instrumentos regulatórios

Portugal, a transposição e aplicação deste regulamento exigirá um esforço legislativo e técnico para compatibilizar as exigências europeias com os princípios constitucionais nacionais, como a dignidade humana, a igualdade e a segurança jurídica.[195]

A este quadro acresce a Declaração de Bletchley (2023), adotada por diversos Estados e organismos internacionais, que sublinha a necessidade de uma cooperação global na mitigação dos riscos sistémicos associados à IA, como a manipulação de conteúdos, a desinformação e a instabilidade democrática. Estes fenómenos representam novos desafios para o Direito português, que deve evoluir para integrar mecanismos preventivos e sancionatórios eficazes face às novas formas de lesão digital.[196]

2. Contexto Jurídico Atual em Portugal

2.1. Regulamento Geral sobre a Proteção de Dados (RGPD)

A transposição do Regulamento Geral sobre a Proteção de Dados (RGPD) para o ordenamento jurídico português concretizou-se através da Lei n.º 58/2019, que estabelece limites e obrigações específicas no tratamento de dados pessoais por entidades públicas e privadas. Uma das dimensões mais relevantes desta adaptação diz respeito à regulação de decisões automatizadas com efeitos jurídicos ou impacto significativo sobre os indivíduos, conforme previsto no artigo 22.º do RGPD (e não no artigo 36.º,

dedicados à regulação da IA, baseado na experiência prévia da UE com o RGPD. Ver: Almada, M. (2025). *The EU AI Act in a Global Perspective*. SSRN Electronic Journal (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5083993).

[195] A análise detalhada do *AI Act* realizada por Jozak revela que este regulamento visa oferecer uma compreensão atual das implicações do *AI Act* para todos os intervenientes envolvidos no ecossistema de IA. Ver: Jozak, T. (2025). *2024 EU AI Act: A Detailed Analysis*. SSRN Electronic Journal (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5196856).

[196] O estudo de Renda sobre os possíveis danos da inteligência artificial e o *AI Act* da UE argumenta que a regulação baseada em risco da IA apresenta tanto potencial quanto limitações, particularmente no que se refere à noção de dano aos direitos fundamentais. Ver: Renda, A. (2024). *Possible harms of artificial intelligence and the EU AI Act*. *In* Journal of Risk Research, https://www.tandfonline.com/doi/full/10.1080/13669877.2024.2350720

como é por vezes confundido). Este regime jurídico consagra o direito dos cidadãos a não ficarem sujeitos exclusivamente a decisões baseadas em tratamentos automatizados, incluindo a definição de perfis, quando tais decisões afetem substancialmente os seus direitos. Este princípio é particularmente importante no contexto da inteligência artificial, onde algoritmos podem ser utilizados para decisões em áreas sensíveis como o acesso a apoios sociais, crédito bancário, habitação ou emprego. Por exemplo, se uma plataforma digital como o Portugal 2030 utilizar IA para selecionar candidaturas a fundos comunitários, os candidatos têm o direito de ser informados sobre os critérios algorítmicos utilizados[197][198], solicitar revisão humana da decisão e obter explicações claras e compreensíveis sobre os fatores decisivos.[199] Além disso, entidades que processem dados com recurso a IA em contextos de risco elevado estão obrigadas a realizar uma avaliação de impacto sobre a proteção de dados (DPIA), de forma a antecipar e mitigar riscos relacionados com discriminação, falta de transparência ou violação de direitos fundamentais.

[197] A exigência de explicabilidade algorítmica para cumprir o direito à contestação do RGPD é abordada por Thalpage e Asatiani *et al*. Ver: Thalpage, N. (2023). *Unlocking the black box: Explainable artificial intelligence (XAI) for trust and transparency in AI systems*. *In* Journal of Digital Art & Humanities, vol. 4, n.º 1, pp. 31-36; Asatiani, A. *et al.* (2020). *Challenges of explaining the behavior of black-box AI systems*. *In* MIS Quarterly Executive, vol. 19, n.º 4, pp. 259-278.

[198] Conforme estabelecido no relatório da OECD sobre transparência algorítmica no setor público. Ver: OECD. AI. (2024). *Algorithmic Transparency in the Public Sector: A state-of-the-art report of algorithmic transparency instruments*. Global Partnership on AI, p. 23 (https://wp.oecd.ai/app/uploads/2024/12/14-Algorithmic-Transparency-in-the-Public-Sector-A-state-of-the-art-report-of-algorithmic-transparency-instruments.pdf).

[199] A análise de Wolff, Lehr e Yoo sobre as lições do RGPD para as políticas de IA revela que "a extensão em que o proposto AI Act depende da autorregulação e a complexidade técnica da aplicação são suscetíveis de representar desafios significativos para a aplicação, com base na implementação dos elementos mais tecnologicamente focados e de autorregulação do RGPD". Ver: Wolff, J.; Lehr, W.; Yoo, C. S. (2023). *Lessons from GDPR for AI Policymaking*. *In* Virginia Journal of Law & Technology, vol. 27, n.º 4.

2.2. Plano Nacional para a Inteligência Artificial (2025-2030)[200]

O Plano Nacional para a Inteligência Artificial (2025-2030), aprovado pelo Governo português, representa um esforço estratégico para promover o desenvolvimento e a adoção ética da IA no país. Alinhado com os valores e diretrizes da União Europeia, este plano visa garantir que a inovação tecnológica seja acompanhada de responsabilidade ética, inclusão social e salvaguarda dos direitos fundamentais. Entre os seus pilares estruturantes, destacam-se as auditorias algorítmicas, que fomentam a criação de mecanismos de auditoria e certificação de algoritmos, especialmente em setores de alto risco como saúde, justiça e educação, assegurando que as decisões sejam transparentes, explicáveis e livres de vieses. O plano também enfatiza a inclusão e acessibilidade[201], sublinhando a importância da participação de grupos vulneráveis[202], como pessoas com deficiência, populações com baixa literacia digital e minorias linguísticas, no desenho e acesso a tecnologias baseadas em IA, garantindo que plataformas educacionais e serviços públicos digitais assegurem acessibilidade e equidade no acesso à informação e oportunidades.[203] A cooperação internacional é outro ponto-chave, com Portugal a participar ativamente em iniciativas globais como a Global Partnership on AI (GPAI), promovendo a harmonização de normas éticas, jurídicas e técnicas à escala global, essencial para enfrentar desafios como a manipulação de conteúdos, a disseminação de desinformação automatizada e a vigilância digital. Por fim, o plano prevê a criação de *sandboxes* regulatórias[204], ambientes controlados para testar soluções inovadoras com supervisão regulatória, permitindo conciliar inovação tecnológica com segurança jurídica, sendo particularmente úteis para startups, universidades e empresas tecnológicas.[205]

2.3. Conexão entre RGPD e Plano Nacional para a IA

Tanto o RGPD como o Plano Nacional para a IA refletem a vontade política e jurídica de colocar os direitos fundamentais no centro da transformação digital. Enquanto o RGPD estabelece a estrutura normativa sobre proteção de dados, decisões automatizadas e consentimento, o plano de IA fornece uma visão estratégica orientada por princípios éticos e de governança responsável. Ambos convergem em pontos fundamentais, como a transparência e explicabilidade dos algoritmos, essenciais para garantir accountability; a participação cívica no desenvolvimento de políticas tecnológicas inclusivas; e a cooperação internacional, alinhada com instrumentos como a Declaração de Bletchley (2023), que reforça a importância de princípios democráticos, segurança e direitos humanos na regulação da IA. Esta convergência de instrumentos jurídicos e estratégicos posiciona Portugal como

[200] O Plano Nacional para a Inteligência Artificial (2025-2030) é frequentemente associado a designações como Estratégia Nacional de Inteligência Artificial (lançada em 2019 sob o nome INCoDe.2030) ou Agenda Nacional de Inteligência Artificial, apresentada no início de 2025. Embora estas expressões se refiram a iniciativas distintas, muitas vezes são usadas de forma intercambiável para descrever o esforço estratégico de Portugal na governança ética e tecnológica da IA. O termo "AI Portugal 2030" também surge como um nome popular para o conjunto de ações alinhadas ao horizonte temporal do plano atual (2025-2030).

[201] A proposta de *sandboxes* regulatórias para testar IA em ambientes controlados é inspirada em modelos europeus, como o *European Data Protection Board* (EDPB).

[202] A escala de transparência proposta por Dores da Silva sugere níveis graduais de exigência para sistemas de IA, garantindo que tecnologias críticas sejam auditáveis e explicáveis. Ver: Dores da Silva, J. P. (2022). *A Inteligência Artificial: Perspetiva Informática e Jurídica*.

[203] Como observa Barredo Arrieta *et al.* no seu estudo abrangente sobre IA explicável, a explicabilidade da IA tornou-se uma necessidade crítica para garantir a confiança e a aceitação dos sistemas de IA em aplicações sensíveis. Ver: Barredo Arrieta, A., *et al.* (2023). *Explainable Artificial Intelligence (XAI): What we know and what is left to be explained*. *In* Information Fusion, 99, https://www.sciencedirect.com/science/article/pii/S1566253523001148.

[204] A flexibilidade regulatória pode ser inspirada em experiências como as dos EUA, onde *sandboxes* permitem testes controlados de IA. Ver: Davtyan, T. (2024). *The US approach to AI regulation: Federal laws, policies, and strategies explained*.

[205] O relatório da OECD sobre *sandboxes* regulatórias em inteligência artificial destaca que "as sandboxes regulatórias em IA permitem às autoridades envolver empresas para testar produtos ou serviços inovadores que desafiam os quadros legais existentes".

um país comprometido com uma abordagem ética, jurídica e tecnicamente robusta à inteligência artificial, preparando o terreno para a aplicação do AI Act da União Europeia e para futuras regulamentações globais.[206/207]

3. Desafios Éticos e Jurídicos

3.1. Vieses Algorítmicos e Discriminação

Os sistemas de inteligência artificial (IA) podem perpetuar ou até intensificar desigualdades sociais, especialmente quando treinados com dados históricos tendenciosos. Este problema manifesta-se em áreas críticas como o recrutamento, a concessão de crédito ou a justiça, onde decisões automatizadas podem discriminar com base na origem étnica, no género ou na condição socioeconómica.[208] Em Portugal, tal discriminação viola o princípio da igualdade consagrado no artigo 13.º da Constituição da República Portuguesa (CRP), bem como o artigo 26.º, que proíbe qualquer forma de discriminação baseada em características pessoais ou sociais.[209] Um exemplo prático ocorre quando um algoritmo utilizado por uma plataforma digital de apoio social está treinado com dados que historicamente excluem minorias étnicas[210], podendo rejeitar automaticamente candidaturas legítimas e comprometendo o acesso equitativo a direitos sociais. Este tipo de viés compromete não só a equidade do sistema[211], mas também a dignidade da pessoa humana, protegida pelo artigo 1.º da CRP. Para enfrentar este desafio, é fundamental criar mecanismos externos de verificação que realizem auditorias independentes capazes de detetar e corrigir discriminações algorítmicas, conforme previsto no Plano Nacional para a IA (2025-2030). Além disso, é essencial garantir que os conjuntos de dados utilizados para treinar sistemas de IA reflitam a diversidade social e não reproduzam preconceitos históricos, em alinhamento com os princípios do RGPD e as diretrizes do AI Act.[212]

3.2. Responsabilidade Civil por Decisões Autónomas

O regime jurídico português consagrado no artigo 483.º do Código Civil estabelece a responsabilidade objetiva por danos causados por produtos defeituosos. Todavia, esta moldura normativa mostra-se insuficiente quando confrontada com os de-

[206] Como analisa Veale e Borgesius na sua análise do projeto do *AI Act*, o *AI Act* representa uma tentativa ambiciosa de regular a IA de forma abrangente, mas a sua eficácia dependerá da coordenação com instrumentos existentes como o RGPD. Ver: Veale, M.; Borgesius, F. J. Z. (2021). *Demystifying the Draft EU Artificial Intelligence Act*. *In* Computer Law Review International, vol. 22, n.º 4.

[207] O estudo de Nizza sobre o impacto do *AI Act* europeu na dinâmica da inovação argumenta que o *AI Act* da União Europeia pode influenciar significativamente a dinâmica da inovação, particularmente no desenvolvimento de IA, exigindo um equilíbrio cuidadoso entre regulação e promoção da inovação. Ver Nizza, F. (2024). *Assessing the Impact of the European AI Act on Innovation Dynamics*. Northwestern University Law School (https://wwws.law.northwestern.edu/research-faculty/clbe/events/standardization/documents/nizza_assessing_impact_ai_act_innovation.pdf).

[208] Como demonstra o relatório da Agência dos Direitos Fundamentais da União Europeia, os algoritmos podem reproduzir e amplificar discriminações existentes na sociedade, criando novos padrões de exclusão que afetam desproporcionalmente grupos vulneráveis. Ver: Wendehorst, C. (2022). Bias in Algorithms: Artificial Intelligence and Discrimination.

[209] O estudo de Chen sobre ética e discriminação em práticas de recrutamento baseadas em IA revela que "a discriminação estatística refere-se ao preconceito resultante de critérios de avaliação que generalizam características de grupo para indivíduos". Ver Chen, Z. (2023), p.8

[210] O uso de auditorias independentes para mitigar discriminações algorítmicas é proposto por Weerts et al., que documentam casos práticos em sistemas de justiça e saúde. Ver: Weerts, H., Xenidis, R., Tarissan, F., Olsen, H. P., & Pechenizkiy, M. (2023). *Algorithmic unfairness through the lens of EU non-discrimination law*.

[211] Os vieses algorítmicos e sua relação com a discriminação são documentados por Weerts et al. e Xenidis, que propõem frameworks legais para responsabilizar desenvolvedores. Ver: Weerts, H. et al. (2023). *Algorithmic unfairness through the lens of EU non-discrimination law*. FAccT Conference 2023, ACM, 805–816; Xenidis, R. (2024). *When computers say no: Towards a legal response to algorithmic discrimination in Europe*. In B. Brożek et al. (Eds.), *Research handbook on law and technology* (pp. 222–234). Edward Elgar Publishing.

[212] Como argumenta Doshi-Velez e Kim no seu trabalho fundamental sobre interpretabilidade em machine learning, a interpretabilidade deve ser considerada desde o início do processo de desenvolvimento, não como uma reflexão posterior. Ver: Doshi-Velez, F., & Kim, B. (2017). Towards A Rigorous Science of Interpretable Machine Learning. arXiv preprint arXiv:1702.08608.

safios colocados pelos sistemas de inteligência artificial (IA) autónomos, cuja tomada de decisão ocorre, frequentemente, sem supervisão humana direta e com graus de autonomia variáveis.[213] A imputação da responsabilidade por danos decorrentes da atuação desses sistemas continua envolta em incerteza jurídica[214], não sendo clara a distribuição de obrigações entre os diferentes intervenientes: programadores, fornecedores, utilizadores ou terceiros afetados.[215]

Neste contexto, foi apresentado o a proposta de diretiva 2022/0303/COD, por parte da UE, que visa a criação de um regime jurídico específico aplicável à responsabilidade civil por danos causados por sistemas de IA.[216] O objetivo primordial é clarificar as obrigações legais de cada agente envolvido, assegurando previsibilidade normativa e justiça na indemnização das vítimas.

[213] Como observa Maia, A. R. no seu estudo pioneiro sobre responsabilidade civil na era da inteligência artificial, "a complexidade dos sistemas de IA e a sua capacidade de aprendizagem autónoma criam desafios únicos para a determinação da causalidade e da culpa, exigindo uma reavaliação dos paradigmas tradicionais da responsabilidade civil. Ver: Maia, A. R. (2021). A Responsabilidade Civil na Era da Inteligência Artificial – Qual o caminho? Julgar Online, p. 25. https://julgar.pt/wp-content/uploads/2021/05/20210511-JULGAR-A-responsabilidade-civil-na-era-da-inteligência-artificial-Ana-Rita-Maia.pdf

[214] A proposta de regime autónomo de responsabilidade civil para sistemas de IA, inspirado no AI Act, é discutida por Madiega e Soyer & Tettenborn, que destacam a necessidade de seguros obrigatórios e fundos públicos de compensação. Ver: Madiega, T. (2023). Artificial intelligence liability directive; Soyer, B., & Tettenborn, A. (2022). Artificial intelligence and civil liability—do we need a new regime?

[215] A insuficiência do Código Civil português para abordar sistemas autónomos é debatida por Maia e Pires & Silva, sugerindo regimes de responsabilidade objetiva. Ver: Maia, A. R. (2021). A responsabilidade civil na era da inteligência artificial - qual o caminho? Julgar Online; Pires, T. C. F., & da Silva, R. P. (2017). A responsabilidade civil pelos atos autônomos da inteligência artificial: Notas iniciais sobre a resolução do Parlamento Europeu. Revista Brasileira de Políticas Públicas, 7(3), 238–254.

[216] Esta proposta reconhece que os sistemas de IA apresentam características únicas que exigem adaptações específicas aos regimes tradicionais de responsabilidade civil. Ver: Proposta de Diretiva 2022/0303/COD sobre responsabilidade civil por danos causados por sistemas de IA.

Entre os principais desafios identificados encontra-se a dificuldade na atribuição da culpa.[217] A natureza complexa e opaca dos algoritmos pode tornar impossível a reconstrução causal dos danos, comprometendo os critérios tradicionais de imputação subjetiva. Acresce ainda a fragmentação regulatória internacional, que compromete a eficácia de qualquer regime isolado, sobretudo em sistemas digitais operados à escala transnacional, como os acessíveis nos países da CPLP.

Como resposta, propõe-se a criação de um regime autónomo de responsabilidade civil para sistemas de IA, inspirado na abordagem da União Europeia com o AI Act, que estabelece categorias de risco e obrigações proporcionais à criticidade da tecnologia envolvida. Este regime permitiria, por exemplo, a responsabilidade objetiva por defeitos em sistemas de alto risco, dispensando a prova de culpa.

Para garantir a efetividade da reparação, recomenda-se ainda a instituição de um seguro obrigatório para sistemas de IA, cuja contribuição seria incorporada no preço de aquisição do sistema. Este modelo garantiria cobertura financeira adequada em caso de danos, distribuindo de forma preventiva o custo do risco entre os diversos operadores económicos.

Complementarmente, justifica-se a criação de um fundo público de compensação, inspirado no regime de indemnização por acidentes de viação, que permita assegurar justiça reparatória em situações-limite, nomeadamente quando o responsável é desconhecido, insolvente ou se recusa a indenizar. Tal fundo reforçaria o princípio da proteção das vítimas, mitigando os efeitos da ausência de imputabilidade direta nos sistemas mais opacos.

A adoção destas medidas permitiria não apenas preencher lacunas jurídicas, mas também alinhar o ordenamento português com as tendências europeias e internacionais, promovendo um ecossistema de IA mais seguro, transparente e justo.

[217] Como argumenta Molnar no seu guia sobre interpretabilidade em machine learning, a interpretabilidade não é apenas uma questão técnica, mas um requisito fundamental para a responsabilização legal. Ver: Molnar, C. (2024). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable (2ª edição). https://christophm.github.io/interpretable-ml-book/

3.3. Privacidade e Consentimento

A utilização intensiva de dados pessoais para treinar sistemas de inteligência artificial (IA) levanta preocupações relevantes quanto ao cumprimento do princípio do consentimento informado, consagrado no Regulamento Geral sobre a Proteção de Dados (RGPD) e concretizado, em Portugal, pela Lei n.º 58/2019. Estas preocupações intensificam-se quando estão em causa sistemas baseados em técnicas de aprendizagem profunda (deep learning), cuja lógica interna é, muitas vezes, opaca até mesmo para os próprios programadores. A complexidade técnica destes modelos compromete, frequentemente, a transparência do tratamento de dados, dificultando o exercício de direitos fundamentais dos titulares, nomeadamente o direito ao consentimento livre, específico, informado e inequívoco.[218]

Um exemplo paradigmático encontra-se nas plataformas de telemedicina que utilizam IA para apoio ao diagnóstico clínico. Nestes contextos, os pacientes devem ser devidamente informados sobre a forma como os seus dados são recolhidos, tratados e utilizados. No entanto, a dificuldade em tornar compreensíveis os processos algorítmicos envolvidos pode comprometer a validade do consentimento prestado, uma vez que impede o titular dos dados de compreender verdadeiramente as implicações da sua decisão.

Neste enquadramento, importa considerar soluções normativas e tecnológicas que reforcem os direitos dos titulares de dados. Uma dessas soluções é a implementação de mecanismos de consentimento dinâmico, que permitam aos utilizadores ajustar ou revogar, de forma contínua e plenamente informada, as autorizações anteriormente concedidas. Esta abordagem reconhece que o consentimento não deve ser um ato estático, mas antes um processo contínuo de autodeterminação informacional.

Simultaneamente, deve ser promovido o investimento em técnicas de explicabilidade algorítmica[219] (*explainable AI* ou XAI), que tornem as decisões automatizadas compreensíveis, auditáveis e, se necessário, contestáveis. Tal exigência encontra suporte no artigo 35.º e 36.º do RGPD, que impõe obrigações de avaliação de impacto e mitigação de riscos em tratamentos que envolvam tecnologias inovadoras. Neste contexto, o direito à explicação pós-decisão algorítmica[220] assume particular relevância, permitindo ao titular dos dados compreender, de forma retrospectiva, como os seus dados foram utilizados e quais os critérios que determinaram o resultado gerado pelo sistema.

A informação posterior à atuação do sistema é, aliás, essencial para esclarecer os resultados tanto ao utilizador como ao fornecedor da tecnologia, sobretudo em cenários de controvérsia judicial ou controlo administrativo. Este direito à explicação, mais do que uma garantia formal, constitui um instrumento fundamental para a responsabilização dos agentes envolvidos e para a promoção de uma cultura de transparência e confiança na utilização da IA.

4. Perspetivas Futuras e Recomendações

A governança da inteligência artificial (IA) exige uma abordagem colaborativa e multidisciplinar que reúna governo, academia, setor privado e sociedade civil na construção de políticas éticas e eficazes.[221] Esta abordagem é es-

[218] Como observa Babic et al. na sua crítica influente sobre explicabilidade algorítmica, para que a explicabilidade seja um requisito moral - e ainda mais para que seja um requisito legal - deve satisfazer certos critérios que frequentemente não consegue cumprir. Ver: Babic, B., Chen, D. L., Evgeniou, T., & Fayard, A. L. (2023). The Algorithmic Explainability "Bait and Switch". Minnesota Law Review. https://minnesotalawreview.org/article/the-algorithmic-explainability-bait-and-switch/

[219] A convergência entre RGPD e princípios éticos do Plano Nacional para a IA é destacada por Montagnani et al., que argumentam que a supervisão regulatória deve ser adaptada a setores de alto risco, como serviços financeiros. Ver: Montagnani, M. L., Najjar, M. C., & Davola, A. (2024). The EU regulatory approach(es) to AI liability, and its application to the financial services market.

[220] O RGPD (artigo 22.º) e o AI Act da UE reforçam o direito à explicação, exigindo que sistemas de IA sejam auditáveis e compreensíveis. Ver: RGPD, artigo 22.º; AI Act, artigo 13.º.

[221] A governança multidisciplinar pode seguir modelos como o grupo de peritos de alto nível da UE, responsável por avaliar o impacto tecnológico e social da IA. Ver: Dores da Silva, J. P. (2022).

sencial para equilibrar a inovação tecnológica com a proteção dos direitos fundamentais, em conformidade com os princípios da Constituição da República Portuguesa (CRP) e do Regulamento Geral sobre a Proteção de Dados (RGPD). Um exemplo prático seria a criação de um órgão regulador independente para a IA em Portugal, inspirado em modelos europeus como o European Data Protection Board (EDPB), que permitiria centralizar a supervisão dos sistemas de alto risco. Este organismo poderia assegurar auditorias algorítmicas, garantir transparência e incorporar especialistas em direito, ética, tecnologia e representantes de grupos vulneráveis. Dessa forma, garantir-se-ia o respeito pela dignidade humana, nos termos do artigo 1.º da CRP, e pelo princípio da não discriminação, conforme o artigo 26.º da CRP. Para além disso, seria importante promover a participação cívica através do envolvimento de organizações não governamentais e movimentos sociais na definição de políticas públicas, como previsto no Plano Nacional para a IA (2025–2030), de modo a assegurar maior justiça e representatividade. Por fim, a formação interdisciplinar deve ser fomentada, incentivando cursos superiores que integrem ética tecnológica, direito e ciência de dados, de forma a preparar profissionais capacitados para enfrentar os dilemas éticos inerentes ao uso da IA.[222]

4.1. Flexibilidade Regulatória

A velocidade com que a inteligência artificial evolui impõe exigências significativas a qualquer enquadramento normativo. Para que a regulação acompanhe esta transformação sem inibir a inovação, torna-se essencial adotar abordagens flexíveis e adaptativas. Uma solução prática reside na criação de *sandboxes* regulatórias – ambientes de teste

[222] A análise de governança da IA numa perspetiva sistémica revela que a governança da IA num ambiente regulatório complexo e em rápida mudança requer uma abordagem adaptativa que considere múltiplas dimensões de risco e oportunidade. Ver: Zaidan, E., & Ibrahim, I. A. (2024). AI governance in a complex and rapidly changing regulatory landscape: A global perspective. Humanities and Social Sciences Communications, 11(1).

supervisionados, nos quais novas tecnologias podem ser experimentadas com um controlo legal e ético adequado.[223] Este modelo permite avaliar, em tempo real, os riscos e benefícios das aplicações de IA, mitigando impactos adversos antes da sua implementação em larga escala.[224/225]

No contexto europeu, o *Artificial Intelligence Act* propõe uma abordagem regulatória baseada no risco, classificando os sistemas de IA em categorias (inaceitável, alto, limitado e mínimo) e impondo requisitos proporcionais à criticidade do sistema. Portugal poderá inspirar-se neste modelo europeu, adaptando-o à sua realidade institucional e tecnológica, através da criação de *sandboxes* específicas para setores sensíveis como a saúde digital, a educação ou a mobilidade autónoma.

Neste cenário, seria recomendado (e sensato) que Portugal avance com projetos-piloto em parceria com empresas tecnológicas e entidades públicas, sob a supervisão de autoridades independentes, com o objetivo de testar soluções de IA em contextos reais de prestação de serviços públicos. Além disso, é funda-

[223] Como demonstra o estudo sobre governança responsável da IA, desenvolvemos um quadro conceptual para a governança responsável da IA (definida através de práticas estruturais, relacionais e processuais), os seus antecedentes e as suas consequências. Ver: Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. The Journal of Strategic Information Systems, 34(2), 101885.

[224] O relatório da OECD sobre sandboxes regulatórias em inteligência artificial destaca que as sandboxes regulatórias em IA permitem às autoridades envolver empresas para testar produtos ou serviços inovadores que desafiam os quadros legais existentes, com os participantes a obter uma isenção de disposições legais específicas ou processos de conformidade para inovar. Ver OECD. (2023). Regulatory sandboxes in artificial intelligence. OECD Digital Economy Papers. https://www.oecd.org/en/publications/regulatory-sandboxes-in-artificial-intelligence_8f80a0e6-en.html

[225] A análise académica sobre sandboxes regulatórias no contexto do AI Act revela que o regime de sandbox sob o projeto do AI Act da UE investiga quão útil é o regime de sandbox para reconciliar inovação e segurança. Ver: Buocz, T., Pfotenhauer, S., & Eisenberger, I. (2023). Regulatory sandboxes in the AI Act: reconciling innovation and safety?. Law, Innovation and Technology, 15(2), 357-389.

mental que a legislação preveja mecanismos de revisão periódica, permitindo a atualização das normas em função do progresso tecnológico e das experiências recolhidas nos ambientes de teste. Esta flexibilidade regulatória será determinante para garantir que os sistemas de IA sejam seguros, eficazes e compatíveis com os valores constitucionais e que complementarmente consigam se ajustar a velocidade exponencial com que a inteligência artificial "evolui".

4.2. Cooperação Internacional

A ausência de um quadro regulatório uniforme à escala global constitui um dos principais obstáculos à governação eficaz da inteligência artificial. A fragmentação normativa dificulta o controlo coordenado de riscos como a desinformação, a manipulação de conteúdos ou a discriminação algorítmica, fenómenos que transcendem fronteiras e exigem respostas articuladas. Neste contexto, Portugal, enquanto Estado-membro da União Europeia e da Comunidade dos Países de Língua Portuguesa (CPLP), encontra-se numa posição estratégica para promover a convergência regulatória em torno de princípios comuns de ética digital e proteção de direitos fundamentais.

Uma das iniciativas internacionais mais relevantes neste domínio é a Declaração de Bletchley (2023), subscrita por múltiplos países como sinal de compromisso com um desenvolvimento seguro e ético da IA, em consonância com os Objetivos de Desenvolvimento Sustentável das Nações Unidas. A participação ativa de Portugal neste tipo de plataformas multilaterais, nomeadamente através da sua presença em eventos tecnológicos de relevo, como a Web Summit, poderá reforçar a sua capacidade de influência e diplomacia digital.[226]

Adicionalmente, no quadro da Global Partnership on Artificial Intelligence (GPAI), Portugal tem a oportunidade de estreitar relações com os países da CPLP, promovendo o intercâmbio de conhecimentos e a adaptação de padrões éticos às especificidades linguísticas e culturais que caracterizam este espaço. Esta cooperação pode assumir particular relevância na definição de critérios comuns para o uso responsável da IA em contextos administrativos, empresariais e sociais, potenciando a criação de um ecossistema digital inclusivo.[227]

Neste sentido, impõe-se destacar a necessidade de harmonização de normas regulatórias com países da CPLP, como o Brasil, Angola ou Moçambique. A proliferação de enquadramentos legais divergentes nestes territórios poderá gerar assimetrias preocupantes em matéria de proteção de dados, responsabilidade algorítmica e acesso equitativo às tecnologias digitais. Portugal deve, por isso, liderar esforços para alinhar políticas públicas de IA no seio da CPLP,[228] contribuindo para a construção de um mercado digital coeso, ético e resiliente.

Finalmente, é recomendável o fomento de mecanismos de partilha de boas práticas, através de fóruns multilaterais inspirados em modelos de cooperação já implementados entre países como a Alemanha e a França. Tais fóruns devem privilegiar a discussão de temas como auditorias algorítmicas, mitigação de vieses discriminatórios[229] e governança

[226] Como observa Zuiderveen Borgesius no seu estudo sobre discriminação algorítmica, a cooperação internacional é essencial para enfrentar os desafios transnacionais da discriminação algorítmica, uma vez que os algoritmos operam frequentemente através de fronteiras e jurisdições. Ver: Zuiderveen Borgesius, F. J. (2020). Strengthening legal protection against discrimination by algorithms and artificial intelligence. The International Journal of Human Rights, 24(10), 1585. https://doi.org/10.1080/13642987.2020.1743976

[227] O estudo sobre sandboxes éticas para IA argumenta que as sandboxes regulatórias podem servir como ferramentas para fomentar a inovação ética e responsável em sistemas de inteligência artificial, desde que sejam adequadamente concebidas e implementadas. Ver:
Moraes, T. (2024, September). Ethical AI Regulatory Sandboxes: Insights from cyberspace regulation and Internet governance. In Proceedings of the Second International Symposium on Trustworthy Autonomous Systems (pp. 1-10).

[228] A harmonização de normas éticas e técnicas à escala global é defendida por Davtyan e Zuiderveen Borgesius, essencial para evitar assimetrias regulatórias. Ver: Davtyan, T. (2024). *The US approach to AI regulation: Federal laws, policies, and strategies explained*; Zuiderveen Borgesius, F. J. (2020). *Strengthening legal protection against discrimination by algorithms and artificial intelligence*. The International Journal of Human Rights, 24(10), 1572–1593.

[229] Estudos como os de de Sequeira e Duffourc & Gerke destacam a importância de auditorias técnicas para identificar vieses e garantir reparação em casos de discriminação algorítmica.

multidisciplinar, promovendo a criação de standards internacionais que conjuguem inovação tecnológica com respeito pelos direitos humanos.

5. Conclusão

A regulação da inteligência artificial (IA) em Portugal encontra-se num ponto crítico, onde a promoção da inovação tecnológica deve ser cuidadosamente equilibrada com a salvaguarda dos direitos fundamentais. A dignidade da pessoa humana (artigo 1.º da CRP) e o princípio da não discriminação (artigo 26.º da CRP) devem estar no centro de qualquer política pública que vise integrar a IA na sociedade de forma ética, segura e inclusiva.

Embora o quadro normativo existente, sustentado pelo Regulamento Geral sobre a Proteção de Dados (RGPD) e pelo Plano Nacional para a IA (2025-2030), ofereça bases importantes para a transparência, a proteção de dados e a inclusão de grupos vulneráveis, persistem desafios significativos. Entre estes, destacam-se a dificuldade em identificar e mitigar vieses discriminatórios, a indefinição quanto à responsabilidade civil em decisões automatizadas e a complexidade do consentimento informado em sistemas cada vez mais opacos e autónomos.

Face a estes obstáculos, impõem-se medidas concretas e estruturantes, como a criação de um órgão regulador independente e multidisciplinar, com poderes de supervisão e auditoria algorítmica. Também se recomenda a implementação de sandboxes regulatórias para testar soluções tecnológicas em ambientes controlados e seguros, bem como o reforço da cooperação internacional, tanto no espaço europeu como no contexto da Comunidade dos Países de Língua Portuguesa (CPLP). Estas ações devem articular-se com os princípios delineados em iniciativas globais, como a Declaração de Bletchley (2023), que apela à responsabilidade partilhada na construção de uma IA ao serviço do bem comum.

Importa, acima de tudo, reconhecer que a ética algorítmica não é apenas um desafio técnico ou jurídico, mas um imperativo constitucional e social. A governança da IA requer um modelo inclusivo, colaborativo e dinâmico, envolvendo o Estado, a academia, o setor privado e a sociedade civil, numa lógica de corresponsabilidade e adaptação contínua.

Somente através de uma abordagem centrada nos direitos humanos, alicerçada em políticas públicas proativas e numa regulação orientada por valores, será possível moldar um futuro onde a inteligência artificial se afirme como aliada da justiça social e não como fator de reprodução de desigualdades. A construção desse futuro começa agora, com visão, responsabilidade e ação coordenada.

Ver: de Sequeira, E. V. (2024). *Estudos sobre responsabilidade civil;* Duffourc & Gerke (2023). *The proposed EU directives for AI liability leave worrying gaps likely to impact medical AI.*