**VI Congresso Internacional da**

**Red Iberoamericana de Estudios Jurídicos 1812**

**“A DEFESA DO ESTADO DE DIREITO NO SÉCULO XXI”**

**AUTORES:**

*Ana Paula Felipe* | Facilitadora Restaurativa e Mediadora de Conflitos (Privada e Judicial) e Pesquisadora Pós-Doutorado InEAC/Universidade Federal Fluminense | Universidade Federal Fluminense | anapaulaffelipe@gmail.com

*Fernanda Duarte* | Professora Associada do Departamento de Ciências Judiciárias, Faculdade de Direito/Niterói-UFF, Pesquisadora NUPEJ e INCT-InEA, Juíza Federal | Universidade Federal Fluminense | fernanda\_duarte@id.uff.br

*Filipa Pais d’Aguiar* | Professora Auxiliar e Investigadora Integrada | Centro Universitário Lusíada - Lisboa, Centro de Estudos Jurídicos, Económicos, Internacionais e Ambientais - CEJEIA, e Investigadora Convidada IGADI | 13001224@lis.ulusiada.pt

*Nuno Silva* | Professor Auxiliar e Investigador | Centro Universitário Lusíada - Lisboa, e Centro de Investigação em Organizações, Mercados e Gestão Industrial - COMEGI | nsas@ulusiada.pt

**EIXO TEMÁTICO n.º 4:**

Novas ameaças ao Estado de Direito

**TÍTULO DA COMUNICAÇÃO:**

Estado de Direito e Inteligência Artificial: Na encruzilhada entre escolhas humanas e escolhas de IA?

**RESUMO:**

*Objecto de estudo -* A disseminação global e ampla utilização dos sistemas de Inteligência Artificial, robots e tecnologias conexas (IA), particularmente, na última década, significaram, antes do mais, a necessidade de repensar o significado de *escolha.* O presente estudo tem como objecto de pesquisa a compreensão e ponderação sobre o significante e significado do signo *escolha* entendido enquanto capacidade de decidir entre uma miríade de opções possíveis. *Relevância da temática –* Escolhas humanas e escolhas algorítmicas*, i.e.,* resultantes de um sistema de IA, distinguem-se, desde logo e por ora, pela existência de um substracto orgânico, de uma mente e dos dispositivos como a consciência e sentimentos. Assim, a escolha humana terá subjacente dimensões e princípios éticos que as escolhas geradas por um algoritmo poderão, eventualmente, ignorar. Tudo dependerá, verdadeiramente, da escolha humana em fazer incluir ou excluir, *a priori*, princípios éticos no momento da construção dos sistemas de IA bem como a observância de códigos de conduta pelos *designers* deste tipo de sistemas. A questão torna-se, particularmente, relevante no caso dos sistemas de IA que operam em áreas sensíveis como a saúde, educação, eleições, *etc*…, ou seja, sistemas de IA que representam um risco elevado em áreas sensíveis e exigem o reforço da protecção dos Direitos Humanos e do Estado de Direito Democrático. Além da distinção fundamental entre escolhas humanas e algoritmos de IA, o presente estudo busca explorar uma terceira via: a interdependência entre humanos e IA no processo decisório. À medida que a IA se integra cada vez mais em áreas sensíveis, como saúde, educação e processos eleitorais, emergem questões cruciais sobre o grau de autonomia que sistemas de IA devem ter e o nível de supervisão humana necessário para garantir a proteção dos direitos fundamentais. A pesquisa propõe uma análise das implicações dessa colaboração entre IA e humanos, sugerindo que a verdadeira escolha não é apenas sobre a inclusão de princípios éticos nos algoritmos, mas também sobre o controle contínuo e responsabilização dos seres humanos que supervisionam esses sistemas. Essa perspectiva amplia o escopo da investigação ao considerar o papel contínuo do ser humano não apenas na fase de criação dos algoritmos, mas também na sua aplicação e monitoramento diário, garantindo que a IA seja uma ferramenta em prol da preservação do Estado de Direito e da proteção dos Direitos Humanos, e não um instrumento de erosão desses valores. De fato, é preocupante a capacidade da IA em trabalhar com a imensa quantidade de Leis de um Estado de Direto e interpretar um universo de casos de Jurisprudência, dadas as atuais limitações técnicas que podem gerar dissonância jurídica, ao mesmo tempo que supera muitíssimo a capacidade de escolha estatística de qualquer ser humano. *Objectivos da pesquisa* – Destarte, procuramos, primeiro, conhecer a real possibilidade de integração de princípios éticos na construção de sistemas de IA e da sua monitorização por seres humanos; segundo, aquilatar o impacto daquela integração na protecção dos Direitos Humanos e do Estado de Direito Democrático; terceiro, identificar as medidas adequadas a uma efectiva protecção dos Direitos Humanos e do Estado de Direito Democrático. O enfoque do presente estudo dirige-se às soluções identificadas, particularmente, mas não exclusivamente, nos ordenamentos jurídicos português e brasileiro. *Metodologia adoptada-* Para tal, será aplicado o método juscomparativo, o método de abordagem hipotético-dedutivo e a análise categorial de conteúdo. Considerando a proliferação de medidas legislativas, no âmbito da utilização de sistemas de IA e da protecção dos Direitos Humanos e do Estado de Direito Democrático, serão consideradas seguintes *questões:* quais os efeitos das medidas proibitivas e das medidas preventivas, na utilização de sistemas de IA? Qual o significado da adopção destas medidas na construção dogmática dos Direitos Humanos e do Estado de Direito Democrático? Além destas questões, serão também apresentados *resultados parciais* relativos à apreciação do impacto da adopção destas medidas na relação entre ordenamentos jurídicos, particularmente, no seu espaço digital.

**Dados de Divulgação – Autores:**



**Ana Paula Felipe | Universidade Federal Fluminense, RJ - Brasil**

Diretora do Observatório de Pesquisa do Instituto Ives Gandra. Professora do Instituto Ives Gandra. Coordenadora do Programa de Políticas Familiares da Associação Guardiões da Infância e da Juventude. Pesquisadora do Instituto de Estudos Comparativos em Administração de Conflitos (NEPEAC/InEAC/PROPPI).

****

**Fernanda Duarte** **| Universidade Federal Fluminense, RJ- Brasil**

fernanda\_duarte@id.uff.br

Professora Associada da Faculdade de Direito da Universidade Federal Fluminense e pesquisadora do NUPEJ/PROPPI/UFF (Núcleo de Pesquisa e Extensão sobre Ciências do Poder Judiciário) e do Instituto de Estudos Comparativos em Administração de Conflitos (NEPEAC/InEAC/PROPPI). Coordenadora do Convênio Acadêmico UFF e Universidade Lusíada (Lisboa/Porto, Portugal). Coordenadora do CRN1 "Comparative Constitutional Law and Legal Culture: Asia and the Americas" no âmbito da Law and Society Association/LSA- EUA. Bolsista do Departamento de Estado Norte-Americano (Programa International Visitors IV. Colaboradora do ALR Int da Thomson Reuters. Global Ethics Fellow Alumna of the Carnegie Council for Ethics in International Affairs (Nova York, EUA). Professora visitante na Mercer School of Law (Macon/GA, EUA). Professora visitante na Birmingham City University – BCU , GB. Professora convidada do" War and Peace – Oxford Iniciative for Global Ethics" (Oxford University, GB).



**Filipa Pais d’Aguiar | Universidade Lusíada – Lisboa**

13001224@lis.ulusiada.pt

Professora Auxiliar da Faculdade de Direito da Universidade Lusíada – Lisboa. Investigadora do CEJEIA – Centro de Estudos Jurídicos, Económicos, Internacionais e Ambientas. Coordenadora do Projecto de Investigação “*Direitos Humanos na Era Digital: e-person e privacidade”*. Coordenadora do Protocolo Académico UFF e Universidade Lusíada (Lisboa/Porto, Portugal). Colaboradora do CRN1 "Comparative Constitutional Law and Legal Culture: Asia and the Americas", no âmbito da Law and Society Association/LSA- EUA. Investigadora convidada do IGADI.



**Nuno Silva | Universidade Lusíada – Lisboa**

nsas@ulusiada.pt

Professor Auxiliar da Faculdade de Ciências da Economia e da Empresa da Universidade Lusíada – Lisboa. Investigadora do COMEGI – Centro de Investigação em Organizações, Mercados e Gestão Industrial. Coordenador do Projecto de Investigação “YOUNGNETSHARE – Ecossistema digital para partilhar conhecimentos utilizando ferramentas de IA”. Membro da “International Society for Ethics and Information Technology” e do Comité Científico Consultivo da “ETHICOMP - International Conference on the Ethical and Social Impacts of ICT”.

**ARTIGO COMPLETO:**

**Estado de Direito e Inteligência Artificial: Na encruzilhada entre escolhas humanas e escolhas de IA?**

*1. Introdução: A Emergência de um Novo Paradigma Decisório*

A ampla disseminação dos sistemas de Inteligência Artificial (IA) e das tecnologias conexas ao longo das últimas décadas impôs desafios sem precedentes aos modelos tradicionais de tomada de decisão e ao próprio conceito de escolha. Em um contexto jurídico, a escolha é historicamente vinculada à capacidade humana de deliberar com base em valores, princípios e normas, expressando juízos de valor subjetivos e integrados a um sistema ético-cultural. A ascensão da IA como agente decisório coloca em xeque essa concepção, pois traz para o centro da discussão a autonomia algorítmica e seus impactos sobre os Direitos Humanos e o Estado de Direito.

Diante desse cenário, o presente texto propõe-se a analisar a interseção entre escolhas humanas e escolhas algorítmicas, investigando não apenas suas distinções, mas também suas interdependências no contexto de sistemas de IA aplicados a áreas sensíveis, como a saúde, a educação e os processos eleitorais, nomeadamente, a susceptibilidade de os sistemas de IA influenciarem, *a priori*, os processos de formulação do pensamento humano, ou seja, a possibilidade de manipulação do pensamento humano e, consequentemente, as escolhas humanas que dele resultem (Aguiar, 2024).

*2. O Significado da Escolha no Contexto Humano e Algorítmico*

A *escolha humana* é tradicionalmente compreendida como um processo subjetivo que envolve componentes racionais, emocionais e éticos. Como ensinam Beauchamp e Childress, (2002, p. 137), o termo "autonomia" tem origem no grego antigo — autos (próprio) e nomos (regra, governo ou lei) — e foi inicialmente utilizado para designar a capacidade de autogoverno das cidades-estado independentes da Grécia. Com o tempo, o conceito foi progressivamente transposto para o âmbito individual, adquirindo múltiplas acepções, tais como autogoverno pessoal, exercício da liberdade, direito à privacidade, autodeterminação, livre-arbítrio, capacidade de orientar o próprio comportamento e reconhecimento do indivíduo como sujeito de si mesmo.

Na tradição filosófica ocidental, tem sido enfatizado o papel da deliberação e da moralidade no processo decisório, postulando que a escolha humana não se reduz a um cálculo lógico, mas integra valores e experiências sensoriais. A afirmação dos valores liberais possibilitou, há mais de dois séculos, o florescimento de uma consideração especial, na cultura ocidental, pela liberdade de autodeterminação dos indivíduos e pelo poder de conduzirem suas vidas conforme as normas que autonomamente escolhem. Incorporada como expressão dos modelos socioeconômicos e políticos que se consolidaram na experiência ocidental, a autonomia passou a ocupar lugar de destaque entre os valores reconhecidos pela sociedade como fundamentais à sua constituição e continuidade (Barreto Neto, 2014, p. 332). Em Kant (2014), por exemplo, o conceito de autonomia constitui o fundamento central de sua ética e se traduz em liberdade. Para o filósofo, a autonomia corresponde à capacidade da vontade de se autodeterminar, ou seja, de se reger por leis que ela mesma se impõe. Trata-se de um autogoverno racional, que expressa a liberdade moral do sujeito. Nesse sentido, a verdadeira escolha autônoma não se orienta por consequências externas ou imediatas, tampouco por interesses, inclinações, prudência ou mera conformidade com regras. Ela decorre de uma adesão consciente e livre à lei moral, guiada pela razão prática. Em outras palavras, podemos dizer que, para Kant, a autonomia é a vontade própria. É governar-se por si mesmo. É a escolha que não leva em conta as consequências externas e imediatas dos atos e nem as regras, por pura prudência, inclinação, interesse ou conformidade. Isto é, a autonomia genuína se revela quando o indivíduo orienta suas ações pelo respeito à lei moral que ele próprio, de forma racional, reconhece como válida — ou seja, quando age conforme o imperativo categórico. Nessa perspectiva, a liberdade, segundo Kant, não se reduz à simples ausência de impedimentos externos, mas consiste na capacidade de autolegislação, isto é, de conduzir-se segundo princípios universais fundados na razão.

Cassirer, um dos mais destacados intérpretes do pensamento kantiano, destaca com precisão o sentido de autonomia em Kant ao afirmar: "A autonomia é aquela vinculação entre a razão teórica e a razão moral, na qual esta última tem a consciência de estar vinculando-se a si mesma" (Cassirer, 1968, p. 287). Com isso, enfatiza-se que, para Kant, a moralidade autêntica surge da capacidade da razão prática de legislar para si própria, sem imposições externas, constituindo-se como expressão da liberdade racional. Nesse sentido, “Aí reside talvez o maior contributo de Kant às teorias da autonomia: a liberdade como autonomia, que significa o desprendimento do indivíduo tanto de fatores externos [...] Para Kant, o exercício da autonomia é a liberdade mesma” (Barreto Neto, 2013).

Traslada para o universo jurídico ocidental, a autonomia se apresenta como um princípio decorrente da dignidade humana, como expressão da liberdade de decidir, de escolher. Nesse sentido, princípio da autonomia da vontade é reflexo da liberdade que nos é conferida pelo ordenamento jurídico, com berço em norma constitucional, entendida como o estatuto jurídico do político conforme nos ensina Canotilho (1995) e se constitui em um dos pilares fundamentais do Direito, transversal a diversos ramos e manifestações jurídicas. A liberdade de decisão e de ação constitui um dos pilares fundamentais do Direito, transversal a diversos ramos e manifestações jurídicas. Conceitos como autonomia e escolha são expressões dessa liberdade, e adquirem contornos próprios em áreas como, por exemplo, o Direito Civil e a Bioética. Em cada uma dessas esferas, a autonomia é reconhecida como valor jurídico e ético, que assegura às pessoas a capacidade de autogoverno, dentro dos limites estabelecidos pela ordem jurídica. No Direito Civil, a autonomia da vontade manifesta-se especialmente no âmbito das relações contratuais. Trata-se do princípio segundo o qual as partes têm liberdade para celebrar contratos e estabelecer suas cláusulas, desde que respeitados os limites da lei, da ordem pública e da função social dos contratos. A autonomia contratual, porém, não é absoluta. Seu exercício exige conhecimento, consciência e equilíbrio entre as partes, razão pela qual o aconselhamento jurídico revela-se essencial. É por meio de uma orientação qualificada que se assegura que a vontade seja livre, informada e verdadeiramente consentida — evitando-se vícios como erro, dolo ou coação. Já na Bioética, o princípio da autonomia assume um caráter profundamente ligado à dignidade humana. Pressupõe o respeito à capacidade individual de tomar decisões sobre a própria vida, o próprio corpo e os próprios cuidados em saúde. É o direito de autodeterminação que garante ao paciente, por exemplo, a possibilidade de aceitar ou recusar tratamentos médicos. Essa liberdade de escolha é reconhecida como elemento indispensável para um atendimento humanizado, ético e de qualidade, no qual o sujeito é visto como agente ativo de sua trajetória terapêutica. A autonomia, nesse campo, é inseparável da ideia de consentimento livre e esclarecido — eixo estruturante da relação médico-paciente. Assim, seja no plano individual, seja no plano institucional, o Direito, nas sociedades ocidentais, valoriza a autonomia como fundamento da liberdade, da responsabilidade e da legitimidade das decisões. O respeito à escolha — livre, consciente e informada — é, em última instância, uma condição necessária para a realização de uma ordem jurídica justa e democrática.

Num outro giro, a questão da autonomia também pode ser apreciada à luz da neurociência que vem ensinando que no processo das escolhas humanas importa, além do substrato orgânico, também os sentimentos homeostáticos, ou seja, os sentimentos que resultam de processos de homeostasia entendidos como o “[…] processo de manutenção dos parâmetros fisiológicos de um organismo vivo (por exemplo, temperatura, pH, níveis de nutrientes, operações viscerais) dentro da amplitude mais conducente ao ótimo funcionamento e à sobrevivência.” (Damásio, 2020, p. 113). As emoções constituem, por isso, ações involuntárias externas e internas que apoiam a homeostasia e traduzem-se em determinados comportamentos (ex. alterações do ritmo cardíaco, reagir com medo, alegria, etc.) (Damásio, 2020, p. 114). Assim, as escolhas humanas, além de resultarem do pensamento racional são também resultado dos sentimentos, e sentimentos homeostáticos. De facto, e como vimos afirmando (Aguiar, 2024, p. 6-7), as escolhas humanas e o seu carácter único, individual e original resulta, precisamente do facto de o ser humano sentir e raciocinar, dependendo das exigências de cada momento. Para Damásio, a escolha humana é, por isso, expressão da “[…] *guerra entre o sentimento e a razão.”* (Damásio, 2020, p. 20)[[1]](#footnote-1), pelo que “Vivemos a vida a sentir, ou a raciocinar, ou ambas as coisas, dependendo daquilo que nos é exigido pelas circunstâncias.” (Damásio, 2020, p. 20). Também M. Sigman refere como a informação emocional recebida pelo cérebro, através dos sentidos, e a sua manifestação em variáveis corporais são suscetíveis de influenciar o nosso processo de decisão (Sigman, 2015, p. 75-76). Além da informação emocional, Sigman destaca como as “boas conversas”[[2]](#footnote-2) podem contribuir, da perspetiva da neurociência, para o processo de escolhas humanas visto que “As boas conversas temperam, permitem compreender diferentes pontos de vista e atenuam a violência.” (Sigman, 2023, p. 68), sendo que, cientificamente, “[…] a boa conversa é a melhor forma de desenvolver uma ampla liberdade de pensamento […]” (Sigman, 2023, p. 68). Mas, liberdade do pensamento humano significa contemplar o ser humano na sua qualidade de ser natural, enquanto concretização da natureza na medida em que as decisões humanas livres “[…] são livremente determinadas pelos resultados das interações fugazes e riquíssimas entre os milhares de milhões de neurónios no nosso cérebro: são livres quando é o interagir desses neurónios a determiná-las.” (Rovelli, 2022, p. 67). A individualidade traduz-se, assim, neste interagir dos neurónios ou, como bem lembra Feynman, num esquema ou numa dança de átomos no nosso cérebro (Feynman, 1955, p. 14)[[3]](#footnote-3). Ser-se livre significa, por isso, que os nossos comportamentos não são simplesmente determinados pelas leis da natureza, mas pelas leis da natureza que atuam no nosso cérebro (Rovelli, 2022, p. 67). Muito bem…, mas, aqui chegados, deveremos questionar: afinal, qual a relevância deste contexto das escolhas humanas nas escolhas algorítmicas? O que importa ao Direito e aos sistemas de IA o contexto neurocientífico das escolhas humanas? Bom, desde logo, a forma como a utilização de sistemas de realidade virtual, em sede de audiência de instrução e julgamento, encerram o potencial de influenciar as decisões do juiz (enquanto ser humano) levam-nos a questionar a observância do princípio da livre apreciação da prova, o princípio da imediação e o respeito pela Dignidade da Pessoa Humana, neste contexto (Duarte e Aguiar, 2023, p. 233 ss.). De facto, a utilização de sistemas de realidade virtual significa uma imersão global do ser humano numa realidade que não é natural, mas virtual, construída a partir de IA. De todo o modo, os sentimentos experimentados pelo ser humano no âmbito da realidade virtual, não deixam de ser sentimentos humanos, não importando a sua fonte. Como relembra Harari, “Dor é dor, medo é medo, amor é amor […]. Não importa se o medo que sentimos tem origem numa coleção e átomos no mundo exterior ou em sinais elétricos manipulados por um computador. O medo continua a ser real. […]” (Harari, 2021, p. 289). Ainda assim, e na medida em que agimos eticamente a partir do entendimento e não do instinto (Humphrey, 2023, p. 236), existe a possibilidade de o ser humano escolher, por exemplo, utilizar ou não utilizar a realidade virtual nestes contextos. De todo o modo, escolhendo utilizar, deverá procurar compreender o significado global da sua utilização bem como das suas consequências. A escolha humana que se determina pela utilização de um determinado sistema de IA deverá, por isso, ser uma escolha educada, informada e esclarecida, sobre todos os aspectos. Esta literacia digital é particularmente relevante nas dimensões sensíveis da sociedade como a justiça, saúde, defesa e educação (Mazzi *et al.*, 2023). De outra forma, o sentimento de autonomia que os utilizadores experimentam na tomada de decisões algorítmicas é apenas uma ilusão (Lu, 2024).

Por outro lado, os sistemas de IA, por mais avançados que sejam, operam sob um paradigma distinto, norteado por algoritmos que em sua definição mais fundamental, são conjuntos finitos de instruções bem definidas, organizadas em uma sequência lógica destinada a resolver um problema ou realizar uma tarefa. Do ponto de vista da ciência da computação, constituem o núcleo do funcionamento de qualquer sistema automatizado, pois orientam a máquina sobre como processar entradas (dados) para gerar saídas (resultados). Contudo, os algoritmos não são apenas estruturas técnicas abstratas: eles refletem escolhas humanas — conscientes ou não — quanto aos critérios de análise, às formas de categorização e aos objetivos finais do processamento. Nesse sentido, algoritmos carregam valores, pressupostos e priorizações que, ainda que implícitos, influenciam diretamente os seus resultados.

Na era da inteligência artificial, os algoritmos assumem uma complexidade ainda maior. Sistemas baseados em aprendizado de máquina (*machine learning*) utilizam algoritmos capazes de modificar seu próprio comportamento a partir da experiência extraída de grandes volumes de dados. Diferentemente de algoritmos determinísticos tradicionais — em que cada entrada leva a uma saída previsível —, esses sistemas se caracterizam pela adaptabilidade e pela construção de modelos estatísticos que “aprendem” com padrões detectados nos dados. Esse processo de autoajuste, ainda que poderoso, pode tornar os algoritmos opacos e imprevisíveis, dificultando a compreensão de seus critérios decisórios mesmo por seus criadores. Além disso, algoritmos modernos frequentemente operam em redes neurais artificiais e arquiteturas de *deep learning*, que simulam em alguma medida o funcionamento do cérebro humano. Nessas estruturas, o conhecimento não é armazenado em regras explícitas, mas em ponderações distribuídas entre camadas de processamento. Isso aprofunda ainda mais o desafio da explicabilidade algorítmica — ou seja, da possibilidade de compreender e justificar as decisões tomadas por sistemas automatizados.

Ora a *escolha algorítmica* é resultado de um conjunto de regras formais, padrões estatísticos e aprendizado de máquina, que buscam maximizar a eficiência de um resultado a partir da análise preditiva de grandes volumes de dados (big data), ou seja, dependem, além de uma quantidade imensa de dados de treino, também de uma “[…] sequência de instruções que diz a um computador o que fazer.” (Domingos, 2017, p. 25), e que “[…] têm de ser suficientemente precisas e não ambíguas para serem executadas por um computador.” (Domingos, 2017, p. 27). No essencial, “[…] os dados entram no computador, o algoritmo faz o que quer com eles, e o resultado sai.” (Domingos, 2017, p. 30). Naturalmente que o resultado dependerá, não só das escolhas humanas que originaram o algoritmo e as instruções, mas também das características dos dados de treino disponibilizados pelo ser humano, sendo este derradeiro aspecto particularmente relevante no caso da GenAI. Independentemente da escolha humana inicial, certo, porém, é que (pelo menos, até certa medida) os algoritmos fazem uma escolha, uma escolha algorítmica, é certo, mas uma escolha.

Destarte, a distinção fundamental entre esses dois modelos de decisão (escolha humana e escolha algorítmica) reside no fato de que a IA, enquanto artefato tecnológico, carece de consciência e de uma estrutura cognitiva capaz de conferir intencionalidade e responsabilidade moral às suas escolhas. Contudo, sua crescente sofisticação na interpretação de dados e no aprendizado autônomo levanta a questão de até que ponto as decisões algorítmicas podem ser equiparadas às humanas e quais são os mecanismos adequados para garantir que suas escolhas permaneçam alinhadas aos valores democráticos e aos direitos fundamentais. Nesse sentido, recordando Norbert Wiener (fundador da Cibernética), o instituto Max Planck (2024) alertou para o fato de as máquinas inteligentes poderem nem sempre tomar decisões da forma que os humanos prevejam - ou desejem.

É nesse ponto que a discussão sobre autonomia se intensifica. Ainda que as máquinas não possuam consciência ou vontade no sentido humano, os sistemas de IA têm adquirido crescente capacidade de operar de forma autônoma, ou seja, sem intervenção humana direta a cada tomada de decisão. Essa “autonomia algorítmica” – que não é a autonomia humana - levanta importantes desafios éticos e jurídicos: como garantir que essas decisões estejam alinhadas com os valores e direitos fundamentais das pessoas? Como assegurar transparência, responsabilidade e possibilidade de contestação em decisões automatizadas que impactam vidas humanas?

Justifica-se, assim, a emergência da designada *machine ethics*, ou seja, a consideração de princípios éticos no processo de elaboração uma escolha algorítmica (Duarte, Aguiar, González, 2024). De facto, a criação de *robots* sem emoções poderia significar no limite, *v.g.*, que “Perante um incêndio descontrolado, o *robot* poderá salvar em primeiro lugar os ficheiros do computador, e não as pessoas, porque a sua programação pode indicar que os documentos importantes não podem ser substituídos, mas os trabalhadores sim.” (Kaku, 2014, p. 299). De todo o modo, no início da IA, estava o ser humano pelo que a escolha algorítmica deverá sempre depender, em última análise, de uma escolha humana. Por isso, será necessário que um ser humano escolha o tipo de programação do *robot* de modo a distinguir o que é ou não importante para o ser humano (Kaku, 2014, p. 229-300). Por isso, também, talvez se discuta o papel dos *artificial moral agents*, ou os agentes morais artificiais que tomam decisões éticas e se sistematizam em quatro tipos: no essencial, os *ethical impact agents* (que produzem consequências morais sem intenção); *implicit ethical agents* (que agem eticamente devido à sua programação); *explicit ethical agents* (que processam e tomam decisões morais); e os *full ethical agents* (que possuem consciência, intencionalidade e livre-arbítrio, características atualmente exclusivas dos humanos)[[4]](#footnote-4).

*3. O Papel da IA na Proteção (ou Risco) dos Direitos Humanos e do Estado de Direito*

A introdução de sistemas de IA no domínio jurídico e administrativo tem potencial para aprimorar a eficiência dos serviços públicos e a previsibilidade das decisões judiciais. No entanto, também apresenta riscos significativos, especialmente quando aplicada a domínios sensíveis como a segurança pública, a concessão de benefícios sociais e a gestão de processos eleitorais[[5]](#footnote-5).

A preocupação central reside na possibilidade de que escolhas algorítmicas, ao serem adotadas sem a devida supervisão e transparência, acabem reforçando desigualdades e preconceitos estruturais. Estudos demonstram que algoritmos de IA podem reproduzir vieses discriminatórios, seja pela inadequação dos dados de treinamento, seja pela falta de mecanismos de auditoria e controle humano (Barocas, Hardt & Narayanan, 2019). No contexto das eleições, por exemplo, sistemas automatizados de segmentação de eleitores e modulação de discursos políticos podem comprometer a integridade do processo democrático, violando direitos fundamentais à privacidade e ao acesso igualitário à informação (Kroll et al., 2017).

Segundo a definição apresentada no *The Blueprint for an AI Bill of Rights* (The White House, 2022), a discriminação algorítmica ocorre quando sistemas automatizados produzem ou contribuem para tratamentos ou impactos injustificados e desfavoráveis dirigidos a indivíduos com base em características como raça, cor, etnia, sexo, gravidez, parto e condições médicas correlatas, identidade de gênero, condição intersexo, orientação sexual, religião, idade, origem nacional, deficiência, status de veterano, informações genéticas, entre outras classificações protegidas por lei. A depender do contexto e das circunstâncias específicas, tais práticas podem configurar violação de garantias legais asseguradas por dispositivos normativos nacionais ou internacionais.

A crescente adoção de sistemas algorítmicos em diversas áreas sociais tem evidenciado a presença de vieses discriminatórios, que resultam em tratamentos desiguais de indivíduos com base em atributos como raça, classe social, gênero, orientação sexual, idade, nacionalidade, deficiência, religião e convicções políticas, como tem apontado a literatura. Tais distorções podem ter origem na interação humano-máquina, na seleção de dados enviesados ou em padrões latentes presentes nos *datasets* utilizados para treinar os sistemas de machine learning. Os efeitos discriminatórios concretos dessas práticas vão, por exemplo, desde a negativa de crédito, emprego e benefícios sociais até decisões judiciais automatizadas que penalizam injustamente determinados grupos.

No particular, destaca-se o conceito de “caixa-preta algorítmica” que se refere à dificuldade — e, por vezes, à impossibilidade — de auditar os critérios internos de decisão de sistemas de IA, sobretudo aqueles baseados em aprendizado profundo. O termo “caixa-preta”, tal como empregado por Pasquale (2015) opera como uma metáfora dual: refere-se tanto a um dispositivo de gravação — como um sistema de monitoramento de dados — quanto a um sistema cujos mecanismos internos permanecem secretos ou desconhecidos. De acordo com Pasquale, em uma sociedade de caixas-pretas, os indivíduos enfrentam diariamente ambas as dimensões desse conceito: à medida que são sistematicamente monitorados por empresas privadas e por órgãos governamentais, perdem a capacidade de compreender, de forma significativa, para onde seus dados são direcionados, como são utilizados e quais podem ser as consequências da manipulação e divulgação dessas informações por tais entidades. Essa opacidade compromete o princípio da motivação das decisões administrativas e judiciais, inviabiliza o contraditório e dificulta a reparação de eventuais erros ou injustiças, configurando, assim, um risco real ao devido processo legal.

De fato, a governança algorítmica, quando desprovida de mecanismos de explicabilidade, gera incertezas jurídicas e fragiliza a confiança social nas instituições públicas. No campo do reconhecimento biométrico, por exemplo, diversos estudos demonstram que a acurácia dos sistemas varia significativamente de acordo com o grupo populacional analisado. Buolamwini e Gebru (2018) investigaram se sistemas comerciais de reconhecimento facial tratam diferentes grupos de forma justa — especialmente quando se leva em conta a cor da pele e o gênero das pessoas[[6]](#footnote-6).

Dessa forma, a questão que se coloca não é apenas a inclusão de princípios éticos na programação de algoritmos, mas a criação de mecanismos institucionais que assegurem a responsabilização contínua dos agentes humanos responsáveis pelo desenvolvimento e monitoramento desses sistemas. É precisamente para esta necessidade que alerta o estudo desenvolvido para a UNESCO, por Innerarity (2024), sobre a IA e a Democracia bem como para a necessidade de uma governança democrática da IA, empenhada na construção de um mundo mais pacífico, inclusivo, equitativo e democrático através da proposta de diversas medidas como a educação e literacia digital, um desenvolvimento tecnológico a par de uma governança democrática; um desenvolvimento legislativo e regulatório da IA que afaste prática abusivas e promova o bem-comum, a explicabilidade, a transparência, o direito de queixa, o pluralismo e a inclusão (Innerarity, 2024). Também nesta senda, mas com um particular enfoque na ética da IA na realidade portuguesa, o estudo de Saraiva (2024) vem demonstrar a necessidade de uma análise multifacetada, de um enquadramento regulatório robusto, de lideranças eficazes e de esforços conjugados entre os vários intervenientes de modo a que os princípios éticos que orientam a IA não representem um estrangulamento da inovação mas, ao invés, um desenvolvimento no sentido do respeito dos Direitos humanos, do bem-estar e da Justiça Social.

*4. A Interdependência entre Humanos e IA na Tomada de Decisão*

Além da dicotomia entre escolhas humanas e algorítmicas, este estudo propõe explorar um terceiro eixo: a interdependência entre ambos no processo decisório. Sistemas de IA estão cada vez mais integrados à rotina dos operadores do Direito, sendo utilizados para análise de precedentes, predição de litígios e automatização de decisões administrativas. Esse cenário exige um novo modelo de governança, em que a IA não substitua, mas complemente a tomada de decisão humana[[7]](#footnote-7).

No contexto judiciário, por exemplo, a IA já é utilizada em tribunais para auxiliar na triagem de processos, na sugestão de jurisprudência aplicável e na análise probabilística de resultados (Aletras et al., 2016; Martinho, 2023[[8]](#footnote-8)). No entanto, a falta de transparência dos modelos de *machine learning* pode gerar desafios significativos para a previsibilidade e a coerência do Direito, especialmente quando decisões judiciais forem influenciadas por padrões estatísticos sem um devido escrutínio crítico. Assim, considerando o potencial dos sistemas de IA na formação do pensamento humano foi possível identificar duas problemáticas distintas (Aguiar, 2024, p. 18 ss.): por um lado, a utilização da IA enquanto mero instrumento auxiliar da atividade judicial; por outro lado, a utilização de sistemas de realidade virtual, onde a sala de audiência e os intervenientes processuais interagem no metaverso através dos seus avatares (*vide, v.g.,* Quiñones Triana, 2023). Sobre a primeira problemática, será, ainda, possível determinar que a “última palavra” caberá a um ser humano, ou seja, que a monitorização dependerá, em última análise, sempre de um ser humana. Quanto à segunda problemática, outras questões se levantam, particularmente, a questão de saber de que forma o pensamento do juiz. Enquanto ser humano é susceptível de ser vulnerável aos conteúdos apresentado no metaverso, condicionando a sua mente, *ab initio*, e, assim, as suas escolhas humanas. Ou seja, falamos aqui da possibilidade de os sistemas de IA influenciarem diretamente o pensamento humano na sua formação, explorando as suas vulnerabilidades e condicionando-o de forma indelével e afetando, por exemplo, o princípio da livre apreciação de prova cuja dimensão de “liberdade” será, talvez assim, algo condicionada. Além desta questão, também outras de índole mais imediata se colocam, como *v.g.*, a possibilidade furto de avatar ou de utilização indevida de dados pessoais. Certo é que a vulnerabilidade do pensamento humano, na sua formação, perante os desafios colocados pelos sistemas de AI, justifica o advento do âmbito de proteção dos neurodireitos, em diversos ordenamentos jurídicos (como veremos, seguidamente), na medida em que o objeto de proteção se dirige diretamente à formação do pensamento humano, ou seja, os direitos da mente.

*5. Questões Fundamentais para a Regulação da IA*

Frente a esse cenário, não basta inserir princípios éticos de forma abstrata nos códigos computacionais. A governança da IA demanda a construção de mecanismos institucionais robustos, que assegurem transparência, auditabilidade e responsabilização contínua dos agentes humanos envolvidos no desenvolvimento, na contratação e na supervisão desses sistemas. A lógica do Estado de Direito exige que todo exercício de poder — inclusive o mediado por tecnologias — esteja sujeito a controles legais, institucionais e democráticos. A responsabilização não pode se perder em uma cadeia difusa de atribuições entre empresas, programadores e órgãos públicos. A ausência de regulação específica favorece a impunidade e a perpetuação de práticas discriminatórias automatizadas. Torna-se essencial estabelecer marcos jurídicos que reconheçam a especificidade da “prova algorítmica”, disciplinem sua validade e forneçam critérios técnicos para aferir sua idoneidade.

Considerando a proliferação do debate sobre os regimes normativos que buscam regular a IA em diversos países, bem como a própria legislação já em vigor (vide por exemplo o Regulamento Europeu sobre Inteligência Artificial de 2024), duas tendências (ou necessidades) podem ser vislumbradas: a adoção de medidas proibitivas e a adoção de medidas preventivas.

As medidas proibitivas tratam de normas que vedam determinadas aplicações de IA devido aos riscos elevados à sociedade, como sistemas de reconhecimento facial em espaços públicos e decisões automatizadas sem possibilidade de revisão humana. Já as medidas preventivas, por sua vez, constituem-se de estratégias de governança que buscam minimizar riscos através da implementação de auditorias, transparência algorítmica e exigência de supervisão humana em processos sensíveis.

A União Europeia propôs um marco regulatório inovador com o Regulamento Europeu de Inteligência Artificial (AI Act), aprovado em 2024. A norma estabelece uma abordagem baseada no risco, classificando os sistemas de IA em quatro categorias: risco mínimo, risco limitado, alto risco e risco inaceitável. Essa última categoria abrange um conjunto restrito de aplicações de IA consideradas particularmente prejudiciais por violarem os direitos fundamentais reconhecidos pelo direito europeu. Tais utilizações — como sistemas de pontuação social pelo Estado, manipulação subliminar de comportamento e reconhecimento facial indiscriminado em espaços públicos — são classificadas como "inaceitáveis" e, portanto, proibidas. Essa classificação visa compatibilizar a inovação tecnológica com a proteção da dignidade humana, impondo limites objetivos ao uso de IA em contextos que envolvem poder estatal, assimetrias de informação e vulnerabilidade social. Trata-se de um esforço normativo relevante para preservar o espaço de deliberação e escolha livre dos indivíduos, mesmo em ambientes fortemente automatizados. O Regulamento Europeu de Inteligência Artificial (UE, 2024) classifica determinadas aplicações como de “risco inaceitável”, vedando seu uso, salvo em exceções estritas relacionadas à segurança pública. Este nível abrange um conjunto restrito de aplicações de inteligência artificial consideradas especialmente prejudiciais, por representarem violações aos direitos fundamentais reconhecidos pela União Europeia. Essa abordagem baseada em proibições e medidas preventivas oferece um modelo relevante, pois incorpora o princípio da precaução e exige supervisão humana em aplicações sensíveis.

Por sua vez, o Brasil (2023) ainda carece de um arcabouço legal que enfrente essas questões com a densidade necessária. O Projeto de Lei 2.338/2023 (Marco Legal da IA) e o PL 3.069/2022 são tentativas embrionárias (inclusive com a proposta de criação de novos direitos — como os neurodireitos e os direitos biométricos) que merecem aperfeiçoamento, embora seus trâmites legislativos ainda não tenham se encerrado.

A adoção dessas medidas levanta questionamentos sobre sua eficácia na construção dogmática dos Direitos Humanos e do Estado de Direito. Qual o impacto dessas regulações na relação entre ordenamentos jurídicos? Como garantir um equilíbrio entre inovação tecnológica e proteção dos direitos fundamentais?

Além disso, coloca-se a delicada questão da compatibilidade entre diferentes modelos regulatórios: como harmonizar legislações nacionais com os padrões internacionais de proteção de dados, direitos humanos e segurança cibernética de modo a evitar-se um “dumping regulatório”, em que empresas migram para jurisdições menos exigentes para evitar responsabilidades?

Outro ponto de tensão refere-se à mediação entre inovação e controle. A imposição de regulação excessiva pode desencorajar o desenvolvimento de tecnologias socialmente benéficas, enquanto a ausência de normatização adequada pode gerar efeitos colaterais irreversíveis, como discriminação algorítmica, vigilância desproporcional e erosão da autonomia individual. É necessário, portanto, construir um modelo regulatório equilibrado, dinâmico e adaptável, que consiga responder aos riscos presentes sem sufocar a criatividade e a experimentação legítima.

É, por exemplo, na senda do âmbito de proteção dos *supra* referidos neurodireitos ou direitos da mente que encontramos instrumentos regulatórios destinados à proteção da sua especial vulnerabilidade, *v.g.*, na Constituição do Chile, logo em 2021 (Chile, 2021) e nas iniciativas legislativas do Brasil (*v.g.,* projeto de lei 522/22. In Brasil, 2022) e, mais recentemente o Marco da IA (projeto de lei 2338/2023. In Brasil, 2023). Por outro lado, também o Regulamento da IA na UE afasta os sistemas de IA susceptíveis de manipular o comportamento humano e explorar vulnerabilidades do ser humano (Parlamento Europeu, 2024). Como vimos testemunhando, assistimos, verdadeiramente, ao advento de fronteiras digitais (Duarte, Aguiar, González, 2025). De todo o modo, as fronteiras digitais parecem reconduzir-se, de novo, de regresso ao essencial, ou seja, ao princípio da proporcionalidade e os seus subprincípios concretizadores. Ou, de outra forma, as exceções[[9]](#footnote-9), não raras vezes, previstas nos instrumentos jurídicos regulatórios dos sistemas de IA redimensionam-se de acordo com dimensão de proporcionalidade. Esta observação leva-nos, agora, de regresso a uma novíssima velha questão, ou seja: não será já bem conhecida a última fronteira na proteção dos Direitos Humanos na era Digital? Ou seja, a necessidade de contemplar o princípio da proporcionalidade, agora repensado e teoricamente reenquadrado à luz dos novíssimos desafios colocados pelos velozes sistemas de IA. Como dizíamos, uma novíssima velha questão.

*6. Conclusão: Rumo a um Modelo de Governança Algorítmica*

Por fim, entendemos que a incorporação crescente da IA no campo jurídico não deve ser percebida como uma ameaça ao estado de direito, mas como um desafio de natureza regulatória e filosófica que demanda respostas institucionais consistentes, coerentes, robustas e bem fundamentadas. Torna-se essencial promover investigações que explorem não apenas a inclusão de princípios éticos na arquitetura dos algoritmos, mas, sobretudo, a criação de mecanismos efetivos de supervisão contínua e de responsabilização dos agentes humanos envolvidos em sua concepção, implementação e monitoramento.

A governança algorítmica deve ser pautada na transparência, na equidade e no respeito aos direitos fundamentais. Dessa forma, é possível garantir que a IA não seja utilizada como instrumento de erosão democrática, mas como uma ferramenta capaz de aprimorar a prestação jurisdicional e fortalecer os valores do estado democrático de direito.

Assim, perante a advento da tecno-democracia e o potencial influenciador que os sistemas de IA encerram na construção e impacto no pensamento humano e nas suas emoções[[10]](#footnote-10) (Rodrigues *et al.*, 2023), o caminho para o respeito pelos Direitos Humanos e a construção da Justiça Social deverá, necessariamente, integrar instrumentos de literacia digital (Mazzi *et al.*, 2023), desde os primeiros anos de vida dos seres humanos. Novamente, uma novíssima velha questão, ou seja, a questão de saber como utilizar a tecnologia direcionando-a para a construção do bem comum.

*Referências Bibliográficas*

ALETRAS, N., TSARAPATSANIS, D., PREOŢIUC-PIETRO, D., & LAMPOS, V. (2016). Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective. *PeerJ Computer Science*.

AGUIAR, Filipa Pais D’ (2024). Direitos Humanos e Inteligência Artificial: principais dimensões jurídicas, éticas, sociais e culturais, no contexto europeu e transnacional. In NETO, Carlos *et al.* (coord.) (2024) – *Mentes digitais : do zero ao infinito : Homenagem à Professora Fernanda Duarte*. Pref. Guilhermo Gama. Editora GZ: RJ. 1-36.

BRASIL (2022). Câmara dos Deputados. *Projeto de Lei 522/22*. Portal da Câmara dos Deputados. 2022. Disponível em: https://www.camara.leg.br/propostas-legislativas/2317524. Acesso em 13 Nov. 2023.

BRASIL (2023). Senado. *Projeto lei 2.338/2023*. Acesso em 14 Abril 2025. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233

BAROCAS, S., HARDT, M., & NARAYANAN, A. (2019). Fairness and Machine Learning: Limitations and Opportunities. *MIT Press*.

BARRETO NETO, H. M. (2013). O princípio constitucional da autonomia e sua implicação no Direito Penal. *Direito Penal e Criminologia:* *XXII Congresso Nacional do Conpedi*. São Paulo: Fundação José Arthur Boiteux. Disponível emhttp://www.publicadireito.com.br/artigos/?cod=e58be547528b4bf8

BARRETO NETO, H. M. (2014). O princípio constitucional da autonomia individual. *Boletim Científico Escola Superior Do Ministério Público Da União*, (42/43), 331–366. Disponível em https://escola.mpu.mp.br/publicacoescientificas/index.php/boletim/article/view/428

BEAUCHAMP, T. L., CHILDRESS, J. F. (2002). *Princípios de ética biomédica*. São Paulo: Loyola.

BUOLAMWINI, J., GEBRU, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, 81, 1–15. Disponível em: https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

CANOTILHO, J. J. G. (1995). *Direito Constitucional* (6. ed.). Coimbra: Almedina.

CASSIRER, ERNST (1968). *Kant, vida y doctrina*. México: Fondo de Cultura Económica.

CHILE (2021). Ministerio de Ciencia, Tecnología, Conocimiento e Innovación (2021) – *Ley 21383.* Biblioteca del Congreso Nacional de Chile : Chile. Disponível em: https://www.bcn.cl/leychile/navegar/imprimir?idNorma=1166983&idVersion=2021-10-25. Acesso em: 12 Out. 2023.

DAMÁSIO, António (2020). *Sentir & Saber: a caminho da consciência*. 1.ª ed. Temas e Debates, Círculo de Leitores: Lisboa.

DOMINGOS, P. (2017). *A revolução do algoritmo mestre: como a aprendizagem automática está a mudar o mundo.* 5.ª ed. Editorial Presença: Barcarena.

DUARTE, Fernanda; PAIS D’AGUIAR, Filipa (2023). Direitos Humanos e Inteligência Artificial: identificação e análise dos principais desafios globais (sob a ótica das sociedades ocidentais). Artigo de coordenadoras de Simpósio online n.º 56. *VIII CIDH: 2023*. P. 233-243. Coimbra, 2023. Disponível em: https://www.cidhcoimbra.com/\_files/ugd/8f3de9\_3d52f26c597d4d27b0d8e41ef3e73464.pdf. Acesso em: 11 Fev. 2024.

DUARTE, Fernanda; PAIS D’AGUIAR, Filipa (2025). 1…0J1U0S1T0I1Ç0A1…0 Justiça entre zeros e uns: Considerações contemplativas. In *Revista Polis* (no prelo).

FEYNMAN, Richard P. (1955). Value of Science. *Engineering and Science*. XIX (Dec. 1955) 13-15.

FEYNMAN, Richard P. (2025). *Nem sempre a brincar, Sr. Feynman*. Trad. Maria Georgina Segurado. 4.ª ed. Gradiva: Lisboa.

HARARI, Yuval Noah (2021). *21 Lições para o Século XXI*. Trad. Rita Canas Mendes. 7.ª Ed. Amadora: Elsinore.

HUMPHREY, Nicholas (2023). *Senciência : a invenção da consciência*. Trad. Pedro Vidal. 1.ª Ed. Lisboa: Bertrand.

INNERARITY, Daniel (2024). *Artificial Intelligence and Democracy.* [Em linha]. UNESCO Office Montevideo : Montevideo. [Consult. 11 Abr. 2024]. Disponível em https://unesdoc.unesco.org/ark:/48223/pf0000389736

KAKU, Michio (2014). *O futuro da mente*. Trad. Fernanda Barão e Isabel Fernandes. 1.ª Ed. Lisboa: Editorial Bizâncio.

KANT, I. (1994). Fundamentação da metafísica dos costumes. In *Textos selecionados*. São Paulo: Abril.

KROLL, J. A., HUEY, J., BAROCAS, S., FELTEN, E. W., REIDENBERG, J. R., ROBINSON, D. G., YU, H. (2017). Accountable Algorithms. *University of Pennsylvania Law Review*, 165(3), 633–705.

LU, W. (2024). Inevitable challenges of autonomy: ethical concerns in personalized algorithmic decision-making. *Humanit Soc Sci Commun* 11, 1321. Disponível em https://doi.org/10.1057/s41599-024-03864-y

MARTINHO, Andreia (2023). Digitalização, automação e Inteligência Artificial nos Tribunais Portugueses. In *a REVISTA* (03; Cultural) 167-178. Lisboa: STJ. Disponível em: https://arevista.stj.pt/edicoes/numero-3

MAX PLANCK, Institute for Biological Cybernetics (2024*). From Cybernetics to AI: the pioneering work of Norbert Wiener.* Disponível em: https://maxplanckneuroscience.org/from-cybernetics-to-ai-the-pioneering-work-of-norbert-wiener

MAZZI, Francesca; TADDEO, Mariarosaria; FLORIDI, Luciano (2023). AI in Support of the SDGs: Six Recurring Challenges and Related Opportunities Identified Through Use Cases. In Mazzi, Francesca and Floridi, Luciano, The Ethics of Artificial Intelligence for the Sustainable Development Goals (May 15, 2023). *Philosophical Studies Series* (PSSP, volume 152), 2023. Available at SSRN: https://ssrn.com/abstract=4448859. Acesso em: 13 Mar. 2025.

OBSERVATÓRIO Eleitoral Internacional in: Universidade Lusíada. CEJEIA. Observatório Eleitoral Internacional. Monteiro, Manuel (2022) - A abstenção eleitoral em tempos de pandemia covid-19: análise de 36 actos eleitorais em 26 Estados espalhados pelo mundo. Lisboa: Universidade Lusíada. *Separata de Polis*. s. 2, n. 5 (janeiro-junho 2022). ISSN 0872-8208. Disponível em: https://doi.org/10.34628/qdef-1657. Consult. 11 Abr. 2025.

PAIS D’AGUIAR, F., DUARTE, F., & GONZÁLEZ, J.A.R.L. (2025). Pode a máquina dotada de IA ser considerada um agente moral? Considerações jus-filosóficas sobre o âmbito de proteção dos neuro-direitos. *Revista Polis,* 2(9), 79–87. https://doi.org/10.34628/G2Y4-CE33. Consult. 11 Abril 2025.

QUIÑONES TRIANA, Maria Victoria (2023). Expert testemony on a court session from Metaverse to hear a case against the Colombian Ministry of Defense and the National Police. Virtual Reality Worlds and the Law : How Judges Can Keep Up with the Rapidly Evolving Technology. *UNESCO Judges Initiative Webinar Series Artificial Intelligence & the Rule of Law virtual conference.* UNESCO, 6 July 2023. Disponível em: https://youtu.be/U-SAsj8LEaA. Acesso em: 6 Jul. 2023.

PASQUALE, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.

PARLAMENTO EUROPEU. *Regulamento Inteligência Artificial, P9\_TA(2024)0138*, Resolução legislativa do Parlamento Europeu, de 13 de março de 2024, sobre a proposta de regulamento do Parlamento Europeu e do Conselho que estabelece regras harmonizadas em matéria de inteligência artificial (Regulamento Inteligência Artificial) e altera determinados atos legislativos da união (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)). 2024. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2024-03-13\_PT.html#sdocta2. Acesso em: 05 Jun. 2024.

RODRIGUES, Paula; SOUSA, Ana; BORGES, Ana Pinto; BROCHADO, Ana; BARBOSA, Isabel (2023). Human(oid): virtual social media influencers. In R. Bansal, S. Qalati, & A. Chakir (Eds.). *Influencer Marketing Applications Within the Metaverse*. ISBN13: 9781668488980. Acesso em: 14 Abril 2025. Disponível em: DOI10.4018/978-1-6684-8898-0.ch001

ROVELLI, Carlo (2022). *Sete breves lições de física*. Trad. Vasco Gato. 1.ª ed., 3.ª Republicação. Penguin Random House: Lisboa.

SARAIVA, Ana Cristina (2024). *Ethics and Responsibility in Artificial Intelligence: A Global Perspective and Portugal's AI Governance*. Gabinete de Estratégia e Estudos do Ministério da Economia. Disponível em: https://www.gee.gov.pt/RePEc/ WorkingPapers/GEE\_PAPERS\_186.pdf Consult. 11 Abr. 2024. eISSN 1647-6212.

SIGMAN, Mariano (2015). *La vida secreta de la mente : nuestro cérebro cuando decidimos, sentimos y pensamos.* (Debates). 1.ª ed. Penguin Random House : Buenos Aires, Barcelona.

SIGMAN, Mariano (2023). *O poder das palavras: a arte de conversar.* Trad. Ana Mendes. Temas e Debates. 1.ª ed. Bertrand Editora: Lisboa.

THE WHITE HOUSE. (2022). *The Blueprint for an AI Bill of Rights*. Disponível em https://bidenwhitehouse.archives.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf

UE (2016). Regulamento (UE) 2016/679 (Regulamento Geral sobre a Proteção de Dados). *OJ L* 119/1, 4 de Maio de 2016. Acesso em 14 Abr. 2025. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679

UE (2024). Regulation (EU) 2024/1689 (Regulamento IA). *OJ L*, 2024/1689, 12.7.2024. Acesso em: 16Abr.2025. Disponível em: https://eur-lex.europa.eu/eli/reg/2024/1689/oj

1. Itálicos do autor. [↑](#footnote-ref-1)
2. Também Richard P. Feynman destaca o valor intemporal das “boas conversas” na pedagogia e educação das crianças e no interesse pela ciência, por saber questionar a realidade que os rodeia: “Foi assim que o meu pai me educou, com aquele tipo de exemplos e conversas, sem me pressionar – apenas belas e interessantes conversas. Motivou-me para o resto da via, e fez com que me interessasse por *todas* as ciências. (Só que sou melhor em física.). Fui, assim, apanhado – como alguém que recebeu algo de maravilhoso quando era criança e não cessa de o tentar *reencontrar*. Tal como uma criança, estou sempre à espera das maravilhas que sei que vou encontrar – talvez nem sempre, mas razoavelmente de vez em quando.” (Feynman, 2025, p. 17). Itálicos do autor. [↑](#footnote-ref-2)
3. “That is what it means when one discovers how long it takes for the atoms of the brain to be replaced by other atoms, to note that the thing which I call my individuality is only a pattern or dance. The atoms come into my brain, dance a dance, the go out; always new atoms but always doing the same dance, remembering what the dance was yesterday.” (Feynman, 1955, p. 4; também, Feynman, 2025, p. 265). [↑](#footnote-ref-3)
4. Para um maior desenvolvimento desta questão, *vide,* Pais D’Aguiar, F., Duarte, F., & A.R.L. Gonzalez, J. (2025). [↑](#footnote-ref-4)
5. Em concreto, sobre os desafios colocados pela utilização de meios electrónicos nas eleições realizadas durante a pandemia de covid-19, v*ide, v.g.,* o estudo desenvolvido pelo Observatório Eleitoral Internacional in: Universidade Lusíada. CEJEIA (2022). [↑](#footnote-ref-5)
6. As pesquisadoras analisaram dois conjuntos de dados muito usados para treinar algoritmos de análise facial: o IJB-A e o Adience. Descobriram que esses bancos são compostos, em sua maioria, por imagens de pessoas de pele clara — 79,6% no IJB-A e 86,2% no Adience — o que já indica um problema de representatividade. Para avaliar o impacto dessa desigualdade, elas criaram um novo conjunto de dados equilibrado, com diversidade de tons de pele (usando a classificação dermatológica Fitzpatrick) e gêneros. Em seguida, testaram três sistemas comerciais de classificação de gênero amplamente utilizados. Os resultados foram alarmantes: os sistemas acertaram quase sempre ao identificar homens de pele clara (erro de apenas 0,8%), mas erraram muito mais com mulheres de pele escura — o grupo com pior desempenho, com erros de até 34,7%. Isso significa que esses sistemas podem falhar mais de um terço das vezes ao tentar identificar corretamente esse grupo. O estudo mostra que, sem diversidade nos dados de treinamento e sem cuidado no desenvolvimento dos algoritmos, as tecnologias de inteligência artificial podem reforçar desigualdades sociais já existentes. Buolamwini e Gebru destacam a necessidade urgente de mais transparência, auditoria e responsabilidade ética no uso da IA, especialmente em aplicações sensíveis como o reconhecimento facial. [↑](#footnote-ref-6)
7. Esta é, aliás, uma das diversas questões propostas para reflexão jusfilosófica em Duarte & Aguiar (2025) - *1…0J1U0S1T0I1Ç0A1…0 Justiça entre zeros e uns: Considerações contemplativas*. In Revista Polis (no prelo). [↑](#footnote-ref-7)
8. Martinho sistematiza a utilização da IA no sistema judiciário português ao nível de suporte, substituição e disruptivo, In MARTINHO (2023). [↑](#footnote-ref-8)
9. *Vide, v.g*., art.º 5.º/1/h, in fine EU AI Act, Art.º 9.º RGPD e as expressões que referem subprincípios do princípio da proporcionalidade (v.g., “tratamento necessário”, “medidas adequadas e específicas”). In UE (2016). [↑](#footnote-ref-9)
10. A este propósito, *vide*, também, o trabalho de investigação desenvolvido pelo laboratório Neurodec, do COMEGI, na Universidade Lusíada: “COMEGI has an advanced and unique LABORATORY in Portugal, NEURODEC - which includes expensive equipment devoted to identifying and measuring the value functions and the relative importance of attributes given by decision-makers. These developments are essential to support sustainable decisions making the required trade-offs between alternative criteria in different areas, namely in sustainable public procurement. The NeuroDec Laboratory, a cornerstone of the project, focuses on enhancing Decision Support Systems through Experimental Evidence (DSSEE). This involved refining the use of advanced tools, including eye-tracking devices, to demonstrate the lab's capabilities. Applying to YOUNGNETSHARE project - Digital Ecosystem to Share Knowledge Using AI Tools - these experiments aim to assess cognitive and decision-making processes influenced by AI enhanced social networking tools. Objectives include identifying motivation triggers and selecting effective digital tools to facilitate learning and knowledge sharing.” (NEURODEC, 2025). Disponível em: https://comegi.ulusiada.pt/en-us/COMEGI/Research-Groups/A-Public-Systems [↑](#footnote-ref-10)