



**UNIVERSIDADE
FERNANDO
PESSOA**

INTELIGÊNCIA ARTIFICIAL E HUMANIDADE: A URGÊNCIA DE UMA COEXISTÊNCIA ÉTICA E CONSCIENTE

[ARTIFICIAL INTELLIGENCE AND HUMANITY: THE URGENCY FOR AN
ETHICAL AND CONSCIOUS COEXISTENCE]

Projeto de Graduação

[1º Ciclo de Estudos em Ciência Política e Relações Internacionais]

Maria Inês Gomes Peres de Miranda

Orientadora:

Doutora Judite Antonieta Gonçalves de Freitas

Setembro 2025

INTELIGÊNCIA ARTIFICIAL E HUMANIDADE: A URGÊNCIA DE UMA COEXISTÊNCIA ÉTICA E CONSCIENTE

[ARTIFICIAL INTELLIGENCE AND HUMANITY: THE URGENCY FOR AN
ETHICAL AND CONSCIOUS COEXISTENCE]

Projeto de Graduação

[1º CE de Ciência Política e Relações Internacionais]

Maria Inês Gomes Peres de Miranda

Orientadora:

Professora Doutora Judite Antonieta Gonçalves de Freitas

Setembro 2025

© 2025

Maria Inês Gomes Peres de Miranda

TODOS OS DIREITOS RESERVADOS

INTELIGÊNCIA ARTIFICIAL E HUMANIDADE: A URGÊNCIA DE UMA COEXISTÊNCIA ÉTICA E CONSCIENTE

Maria Inês Gomes Peres de Miranda

Projeto de graduação apresentado à Universidade Fernando Pessoa como parte dos requisitos para obtenção do grau de Licenciada em Ciência Política e Relações Internacionais, sob orientação da Professora Doutora Judite Gonçalves de Freitas.

Dedicatória

À Aurora, Mio e Nala, que todos os dias me dão razões para querer continuar a contribuir para um mundo mais justo e empático para com todos os seres vivos que nele habitam.

In memoriam, a todos os que amei e hei de continuar a amar.

À Terra, a minha principal fonte de inspiração e a quem tudo devemos.

Agradecimentos

Primeiramente, à Universidade Fernando Pessoa, a instituição de ensino superior universitário que está por detrás de todas as atividades que, enquanto aluna, coloquei em marcha nos últimos 3 anos.

À minha orientadora de projeto de graduação, a Professora Doutora Judite Gonçalves de Freitas, cuja orientação e dedicação constante foram determinantes para a concretização deste trabalho. O seu exemplo enquanto docente e ser humano não se limita a ter marcado o meu percurso académico nesta instituição, mas antes, ficará na minha memória enquanto referencial de integridade, dedicação e genuína humanidade. Prometo levar os seus ensinamentos para o resto da vida.

A todos os docentes com quem tive a honra de aprender diariamente: o meu mais sincero agradecimento.

À minha família, que quando a dedicação aos estudos me afastou demonstraram uma paciência inabalável.

À Catarina e à Cláudia, por nunca deixarem de acreditar em mim, mesmo quando eu própria duvidava. Obrigada pela escuta, pelas palavras certas nos momentos certos e por estarem presentes, mesmo na distância. A vossa amizade foi âncora e abrigo ao longo desta jornada.

Este trabalho também é vosso.

Resumo

O presente trabalho propõe uma reflexão crítica e construtiva sobre o papel da Inteligência Artificial (IA) nas sociedades contemporâneas, focando-se nos desafios éticos, políticos, jurídicos e sociais que levanta, nas possíveis repercussões para a democracia e nas visões de futuro que suscita. Longe de advogar uma postura tecnofóbica e de rejeição da IA, parte-se da convicção de que a integração consciente, responsável e ética destas tecnologias na vida humana não é tão somente necessária, mas antes inevitável - à semelhança do que se defende no que reporta à convivência entre o ser humano e a natureza: uma coexistência orientada por valores de respeito, equilíbrio, responsabilidade e interconexão. Através de uma análise teórica e bibliográfica, o estudo inicia-se com uma introdução à evolução da IA e às suas aplicações mais relevantes, procurando expor o seu potencial revolucionário. De seguida, explora-se a complexidade dos dilemas éticos que decorrem da sua utilização, como o desvio algorítmico e democracia, a responsabilização e a instrumentalização do ser humano, propondo ainda uma breve incursão sobre os dilemas políticos, jurídicos e sociais que aqui se colocam. Num terceiro momento, analisa-se a relação entre a IA e a democracia, com destaque para os riscos associados à manipulação de informação, monitorização digital, concentração do poder tecnológico e a erosão do debate público. Por fim, reflete-se sobre a urgência da aposta na literacia digital, o papel central do ensino na manutenção da autonomia intelectual e o equilíbrio frágil entre segurança e liberdade.

Palavras-chave: Inteligência Artificial, ética, algoritmos, democracia, responsabilidade, e literacia digital.

Abstract

The present study proposes a critical and constructive reflection on the role of Artificial Intelligence (AI) in contemporary societies, focusing on the ethical, political, legal, and social challenges it raises, its possible repercussions for democracy and on the visions of the future it evokes. Far from advocating a technophobic stance or a rejection of AI, this work proceeds from the conviction that the conscious, responsible, and ethical integration of these technologies into human life is not merely necessary but indeed inevitable - much like what is defended in relation to the coexistence between human beings and nature: a coexistence guided by values of respect, balance, responsibility, and interconnection. Through a theoretical and bibliographical analysis, the study begins with an introduction to the evolution of AI and its most significant applications, seeking to demonstrate its revolutionary potential. It then explores the complexity of the ethical dilemmas arising from its use, such as algorithmic bias and democracy, accountability, and the instrumentalization of the human being, while also undertaking a brief incursion into the political, legal, and social dilemmas involved. In a third stage, the analysis turns to the relationship between AI and democracy, with particular emphasis on the risks associated with information manipulation, digital surveillance, the concentration of technological power, and the erosion of public debate. Finally, the study reflects on the urgency of investing in digital literacy, the central role of education in preserving intellectual autonomy and the fragile balance between security and freedom.

Keywords: Artificial Intelligence, ethics, algorithms, democracy, accountability and digital literacy.

Índice

RESUMO.....	I
ABSTRACT	I
ÍNDICE DE FIGURAS	IV
LISTA DE ABREVIATURAS	V
INTRODUÇÃO	1
CAPÍTULO I – A INTELIGÊNCIA ARTIFICIAL: DEFINIÇÃO E EVOLUÇÃO	3
1.1. INTELIGÊNCIA ARTIFICIAL - ETAPAS DE DESENVOLVIMENTO E TRANSFORMAÇÃO	3
1.2. CLASSIFICAÇÃO DA IA: FUNCIONALIDADE E CAPACIDADE	5
1.3. AVANÇOS TECNOLÓGICOS E MODELOS DE PROCESSAMENTO DE LINGUAGEM	6
CAPÍTULO II - DESAFIOS ÉTICOS, POLÍTICOS, JURÍDICOS E SOCIAIS DA IA	8
2.1. PARA LÁ DO VAZIO ÉTICO: DA INCERTEZA À CONFIANÇA	8
2.2. DECISÃO ALGORÍTMICA: AUTONOMIA E A CRISE DA RESPONSABILIDADE	9
2.3. TRÊS DESAFIOS DE RELEVO NA REGULAÇÃO DA IA	10
CAPÍTULO III – DEMOCRACIA, ALGORITMOS E INTELIGÊNCIA ARTIFICIAL	12
3.1. (DES)INFORMAÇÃO, FILTER BUBBLES E MANIPULAÇÃO DE OPINIÃO	12
3.2. EROSÃO DO DEBATE PÚBLICO E DA PLURALIDADE DEMOCRÁTICA	13
3.3. ALGOCRACIAS E O ECLIPSE DO JUÍZO DEMOCRÁTICO	14
CAPÍTULO IV – LITERACIA DIGITAL E CIDADANIA: ÉTICA, EDUCAÇÃO E EQUILÍBRIO NA ERA DA IA	16
4.1. LITERACIA DIGITAL COMO COMPETÊNCIA CRÍTICA NO SÉCULO XXI	16
4.2. ENSINO, ÉTICA E SUSTENTABILIDADE DEMOCRÁTICA	17
4.3. EQUILÍBRIO ENTRE SEGURANÇA E LIBERDADE	18
REFLEXÕES FINAIS.....	21
BIBLIOGRAFIA	22

Índice de Figuras

Figura 1- Trajetória diacrónica da evolução da IA. Fonte: Elaborado a partir de Daylight (2015), Smith et al. (2006) e Rai (2024).....	4
Figura 2 - Classificação da IA de acordo com a categoria de Funcionalidade. Fonte: Elaborado a partir de Hassani et al. (2020)	5
Figura 3 - Classificação da IA de acordo com a categoria de Capacidade. Fonte: Elaborado a partir de Hassani et al. (2020).....	6
Figura 4 - Sete princípios éticos para uma IA de confiança. Fonte: Elaborado a partir de Floridi (2019); Comissão Europeia (2019).....	9

Lista de Abreviaturas

AGI – Inteligência Artificial Generativa

AI HLEG – AI High-Level Expert Group

ALTAI - Assessment List for Trustworthy AI

ANI – Artificial Narrow Intelligence

APA – American Psychology Association

ASI – Artificial Superintelligence

BAT – Baidu, Alibaba e Tencent

DL – Deep Learning

DL – Deep-Learning

EUA – Estados Unidos da América

FM – Foundation Models

GAFAM – Google, Amazon, Facebook/Meta, Apple e Microsoft

Gen AI – Generative Artificial Intelligence

GPT – Generative Pre-Trained Transformer

IA – Inteligência Artificial

IBM – International Business Machines Corporation

LISP – LISt Processing

LLM – Large Language Models

ML – Machine Learning

UE – União Europeia

Introdução

“The danger is that if we invest too much in developing AI and too little in developing human consciousness, the very sophisticated artificial intelligence of computer might only serve to empower the natural stupidity of humans.” Yuval Noah Harari, X, (2023, Abril 3).

Nas últimas décadas, a Inteligência Artificial (IA) tem vindo a consolidar-se como uma das forças com maior potencial revolucionário da contemporaneidade. Desde sistemas de reconhecimento facial à automatização de processos decisórios no sector público e privado, a sua influência – seja ao nível da forma ou do ritmo a que vivemos, trabalhamos e nos relacionamos – é crescente. Esta evolução, enraizada no desenvolvimento técnico-científico, é indissociável das grandes questões sociais e políticas do século XXI. É a partir deste contexto que o presente trabalho propõe uma reflexão crítica e construtiva sobre a integração da IA, partindo da convicção de que a sua presença nas sociedades humanas deve ser moldada por valores éticos, democráticos e humanistas.

Longe de propor uma postura tecnofóbica e de rejeição do avanço tecnológico – posição que, pela sua inevitabilidade, se revelaria incompreensível – este estudo defende a necessidade de pensar a IA à semelhança da relação idealizada entre o ser humano e a natureza, como uma convivência orientada por responsabilidade, respeito mútuo e busca de equilíbrio não com base no domínio, mas sim na coexistência mutuamente benéfica – uma integração orientada pela responsabilidade, respeito mútuo e busca de equilíbrio na relação homem-máquina. Propõe-se, assim, olhar para a IA, não enquanto força autónoma e potencialmente hostil, mas sim, quando acompanhada da devida capacitação dos indivíduos, como instrumento capaz de direcionar o progresso tecnológico para o fortalecimento da vida coletiva, alicerçada em valores éticos e democráticos.

A partir de uma perspetiva teórica e de uma revisão da literatura, o presente trabalho estrutura-se em quatro capítulos. Num primeiro momento, apresenta-se uma panorâmica da evolução da IA e principais aplicações – essencial para compreender a amplitude do impacto tecnológico e o seu potencial transformador. De seguida, exploram-se os desafios éticos, políticos, jurídicos e sociais da IA, como o desvio algorítmico, a desumanização

decisória e a necessidade de harmonização normativa, ressaltando a urgência de estratégias orientadas para a transparência e responsabilidade. Num terceiro momento, analisa-se a intersecção entre a IA e os regimes democráticos, evidenciando os riscos de manipulação informacional, vigilância e concentração de poder, bem como o seu impacto na erosão das liberdades fundamentais e nos princípios democráticos. Por fim, reflete-se sobre a urgência da literacia digital para a cidadania plena, o papel do ensino na autonomia intelectual e o desafio de equilibrar segurança e liberdade face ao avanço tecnológico.

Ao propor esta reflexão, o presente estudo pretende contribuir para um debate cuja necessidade se torna progressivamente mais evidente: como integrar a IA de forma a preservar, e quiçá até expandir, a humanidade? A resposta, ainda em construção, passa também por reconhecer que o futuro não está programado, dependerá das escolhas que fizermos no presente.

Capítulo I – A Inteligência Artificial: Definição e Evolução

A IA tornou-se uma das áreas mais influentes e disruptivas da ciência e tecnologia contemporâneas, conquistando um espaço central no debate acadêmico, político e social das últimas décadas. Por representar uma verdadeira turbulência tecnológica – resultante da rápida evolução e pelo profundo impacto na organização das sociedades (Tallberg et al., 2023) – impõe-se como objeto de reflexão crítica.

Segue-se, por isso, uma breve contextualização histórica, destacando os momentos-chave do seu desenvolvimento: dos primeiros algoritmos lógicos à revolução do *Deep Learning* (DL) e dos Modelos Generativos, que marcam o presente. A partir de uma perspectiva diacrónica, estabelece-se uma base conceptual da IA, incluindo as definições clássicas do processamento de informação até abordagens modernas de aprendizagem automática e análise de dados. Este capítulo procurará responder à pergunta: o que se entende por IA e que dimensões envolve?

1.1. Inteligência Artificial - etapas de desenvolvimento e transformação

O que é a IA? Sucintamente, é a tentativa de simular, por via de um sistema computacional, algo que corresponda, ou até consiga eventualmente superar, a inteligência humana. Embora possa ser definida de diversas formas, uma definição comumente utilizada é a do matemático americano Richard Bellman¹ (1978), que a define como “[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning” (*cit.* in Russel & Norvig, 2021, p.2).

Esta definição levanta uma segunda pergunta: o que é a inteligência? A inteligência é um conceito multifacetado, e embora a sua definição, à semelhança do conceito de IA, seja alvo de diversas interpretações, de acordo com a American Psychological Association (APA), inteligência é “a habilidade de deduzir informação, aprender com a experiência, adaptarmo-nos ao ambiente, compreender e dar uso correto do pensamento e da razão”

¹ Em 1957, Bellman formalizou uma classe de problemas de decisão sequenciais conhecidos como Markov Decision Processes (MDPs), que são fundamentais para o entendimento da tomada de decisão em ambientes incertos e dinâmicos – hoje aplicadas na robótica e automatização.

(APA, 2018, s/p). Mas então o que distingue ainda a inteligência humana da IA? Muito embora a IA execute tarefas complexas, ainda existe algo singular na mente humana.

Esta terceira questão leva-nos a um outro conceito: o da consciência. A consciência é a capacidade de experienciar, conscientemente, o *eu* e o ambiente que *o* rodeia (Stanford Encyclopedea of Phylosophy, 2014). A consciência – a experiência subjetiva do mundo – é um traço que separa a inteligência humana da IA. Essa aptidão, que combina compreensão do mundo e autoconsciência, ainda não é reconhecida na IA e permanece sem resposta concreta, suscitando tanto preocupação quanto entusiasmo (Baerren, 2025). Para Damásio (2000, p. 32), “A consciência é um fenómeno inteiramente privado e na primeira pessoa que ocorre no interior (...) da mente”, esta percepção permite estabelecer, pelo menos do ponto de vista filosófico e científico, no presente, uma linha que separa a IA da inteligência biológica.

Embora as origens da IA recuem à Grécia Antiga – em particular a Aristóteles (384-322 a.C.) e à introdução do raciocínio dedutivo² que fundamenta as bases conceptuais desta tecnologia (Rai, 2024) – a análise inicia-se com o contributo de Alan Turing, matemático e cientista da computação, conforme a Figura 1:

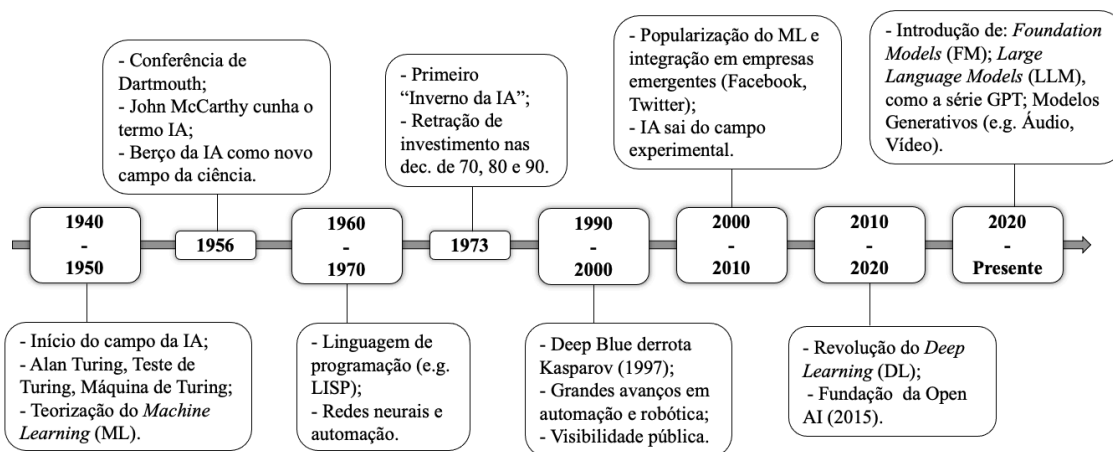


Figura 1- Trajetória diacrónica da evolução da IA. Fonte: Elaborado a partir de Daylight (2015), Smith et al. (2006) e Rai (2024)

² Processo de pensamento lógico que permite tirar conclusões específicas a partir de premissas gerais – do amplo para o particular.

1.2. Classificação da IA: Funcionalidade e Capacidade

De acordo com Hassani et al. (2020), a IA pode definir-se como um sistema computacional apto a reproduzir processos cognitivos humanos, recorrendo a dados provenientes de múltiplas origens para decidir e aprender com base em padrões. A IA, assim como todos os sistemas e máquinas nela baseados, pode ser classificada de diferentes formas, incluindo – mas não se limitando – a funcionalidade, áreas de aplicação e nível de autonomia. Neste subcapítulo, serão analisadas duas categorias distintas de classificação – funcionalidade e capacidade – conforme propostas por Hassani:

A primeira remete-nos para a “mente” da IA – a capacidade de pensar de forma mais ou menos autónoma e o grau de semelhança ao pensamento humano. Aqui encontramos quatro subclassificações: Máquinas Reativas, Máquinas de Memória Limitada, *Theory of Mind* e IA dotada de autoconsciência – representadas na Figura 2:

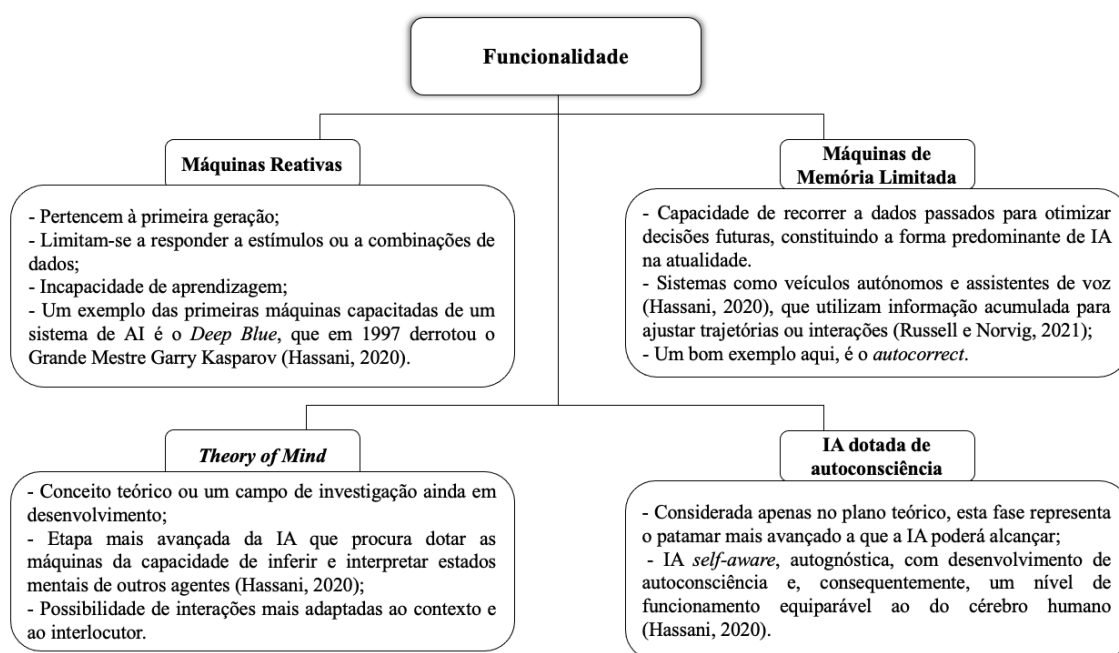


Figura 2 - Classificação da IA de acordo com a categoria de Funcionalidade. Fonte: Elaborado a partir de Hassani et al. (2020)

A segunda categoria de classificação assenta numa perspetiva orientada para a capacidade tecnológica, distinguindo três subtipos: Inteligência Artificial Restrita (*Artificial Narrow Intelligence* – ANI); Inteligência Artificial Geral (*Artificial General Intelligence* – AGI) e Superinteligência Artificial (*Artificial Superintelligence* – ASI), representados na Figura 3:

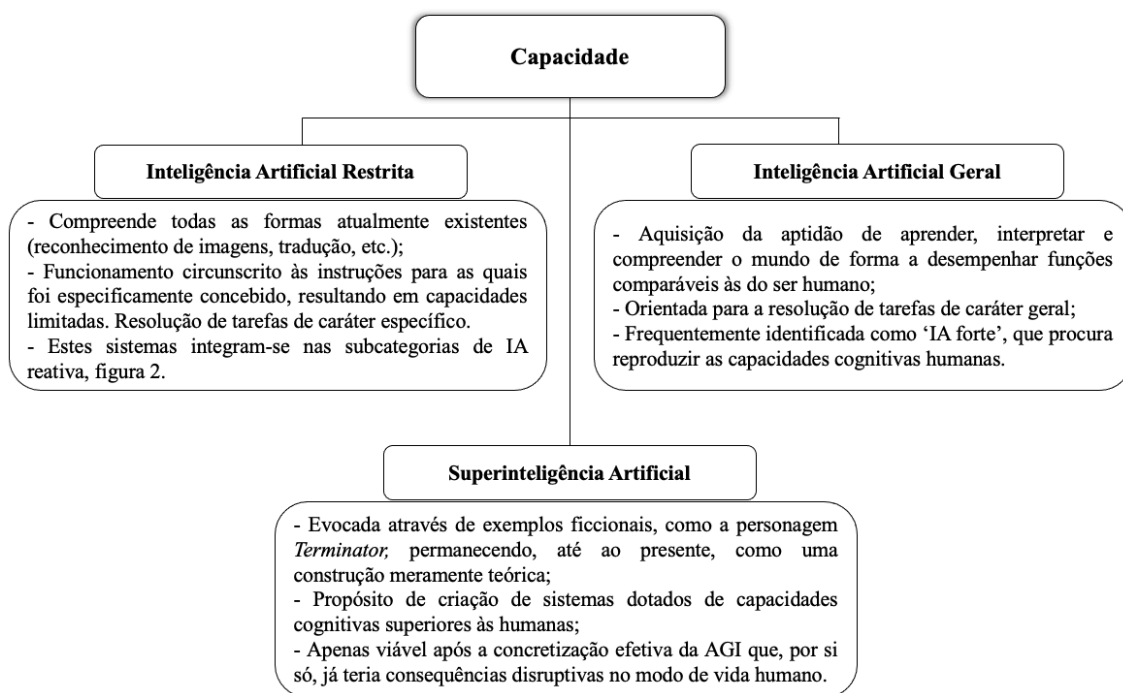


Figura 3 - Classificação da IA de acordo com a categoria de Capacidade. Fonte: Elaborado a partir de Hassani et al. (2020)

A classificação da IA evidencia a diversidade de estágios de desenvolvimento e de potencialidades de um domínio científico em constante evolução. Enquanto as categorias mais avançadas permanecem sobretudo conceptuais, é nas formas já operacionais – como a IA restrita e as máquinas de memória limitada – que assenta o progresso atual, base das aplicações contemporâneas e ponto de partida para a análise das tecnologias hoje dominantes, nomeadamente os modelos de linguagem.

1.3. Avanços Tecnológicos e Modelos de Processamento de Linguagem

Segundo Jeff Crume (2024), antes da consolidação do *Machine Learning* (ML), a IA apoiava-se maioritariamente em linguagem de programação. O ML, que se afirmou a partir da década de 2010, representou um ponto de viragem ao permitir que os sistemas aprendessem a partir de grandes bases de dados, mostrando-se particularmente eficaz em tarefas de previsão e deteção de padrões, com aplicações relevantes na área de cibersegurança e análise de informação. O DL, que se desenvolve na mesma altura, e que assenta em redes neurais profundas inspiradas no funcionamento do cérebro humano, ampliou exponencialmente as capacidades de processamento e reconhecimento, permitindo avanços notórios na simulação e mímica de funções cognitivas.

A década de 2020 assinala a consolidação da IA Generativa (GenAI) estando presente em aplicações amplamente difundidas, como sistemas de recomendação (Netflix, Spotify) e assistentes virtuais (Alexa, Siri). Em 2022, com a disponibilização pública do *ChatGPT* – baseado na série de modelos GPT da OpenAI (2015) desenvolvidos desde 2018 – verifica-se um marco decisivo na massificação do seu uso. Segundo Crume, estes sistemas, assentes em *Foundation Models* (FM) treinados com volumes massivos de dados, embora ainda dentro do campo da ANI – inauguram uma nova etapa da IA, trazendo para o centro do debate os *Large Language Models* (LLM), que se distinguem pela capacidade de dialogar, traduzir, resumir e produzir textos complexos automaticamente. A sua disseminação não se limitou a democratizar o acesso a ferramentas avançadas, tornando-as num fenómeno cultural global, mas também a impulsionar o surgimento de aplicações multimodais capazes de gerar imagens, áudio e vídeo, dando origem a fenómenos como os *deep fakes*.

O alcance destas tecnologias ultrapassa o domínio estritamente académico, estando hoje incorporadas em ferramentas de escrita, motores de pesquisa, plataformas educativas, aplicações de entretenimento e sistemas de apoio à decisão (Rashid & Kausik, 2024). Tal ubiquidade explica porque se considera que os LLM simbolizam o atual domínio da IA, ao mesmo tempo que se levantam debates sobre criatividade, autenticidade e responsabilidade. Embora estes desenvolvimentos confirmem o carácter transformador da tecnologia, levantam, simultaneamente, um conjunto de desafios éticos, políticos, jurídicos e sociais, cuja análise se segue.

Capítulo II - Desafios Éticos, Políticos, Jurídicos e Sociais da IA

A par de ser um instrumento de eficiência e inovação, a IA detém um potencial de impacto significativo nas dinâmicas sociais, reconfigurando decisões, sistemas de poder e a concepção de dignidade humana. Por estes motivos, a comunidade científica debate as questões éticas, da regulação política e jurídica e dos impactos sociais da IA, na atualidade.

Este capítulo sintetiza os principais desafios emergentes e algumas das respostas correspondentes, considerando o contexto de crescente interdependência e interação entre humanos e IA. Para o efeito, centramo-nos em três eixos: o vazio ético da IA, a crise da responsabilidade decorrente da decisão algorítmica e os obstáculos que se colocam à construção de quadros regulatórios que respondam ao desafio em mãos.

2.1. Para lá do Vazio Ético: da Incerteza à Confiança

À medida que a substituição do homem pela máquina rumo à otimização de recursos se materializa, surgem questões quanto à forma como a interação entre humanos e tecnologia se está a edificar (Rashid & Kausik, 2024). Como salientam Rashid e Kausik, será necessário encontrar um equilíbrio entre inovação e princípios éticos de forma a garantir que esta relação maximize benefícios e minimize riscos. Em julho de 2020, o *High-Level Expert Group on AI* (AI HLEG) apresentou à Comissão Europeia o *Assessment List for Trustworthy AI* (ALTAI), a versão final de uma lista de recomendações para a área do desenvolvimento da IA, segundo a qual, para além de cumprir com os quadros legais em vigor, uma IA pode ser considerada confiável mediante a observância de 7 princípios éticos, representados na Figura 4:

1) Agência e supervisão humana	A IA deve promover a autonomia humana, capacitando os indivíduos de forma a conseguirem tomar decisões informadas. Paralelamente, devem ser assegurados mecanismos de supervisão que sigam abordagens <i>human-in-the-loop</i> , <i>human-on-the-loop</i> , and <i>human-in-command</i> .
2) Robustez técnica e segurança	Os sistemas de IA devem ser seguros e resilientes, com planos de <i>back-up</i> capazes de lidar com erros e minimizar danos. Têm de ser precisos, confiáveis e passíveis de ser reproduzidos.
3) Privacidade e governação de dados	Para além da garantia do respeito pela privacidade e dados pessoais, devem ser instituídos mecanismos de gestão de dados que assegurem a qualidade e a integridade da informação, garantindo simultaneamente o acesso legítimo e a proteção contra usos abusivos, discriminatórios ou nocivos.
4) Transparência	Os sistemas de IA, a qualidade dos dados, dos modelos e dos processos devem ser passíveis de ser rastreados por meio de mecanismos desenhados para o efeito. Todos os que interagem com a IA devem conhecer as suas capacidades e limitações.
5) Equidade, diversidade e não discriminação	O acesso aos sistemas de IA deve ser garantido a todos, evitando desigualdades e discriminação.
6) Bem-estar social e ambiental	Todos devem poder gozar dos benefícios que a IA traz consigo, futuras gerações e ambiente incluídos. Deve ser promovida a responsabilidade ecológica, considerando os impactos sociais e ambientais que acarreta.
7) Responsabilização	O uso da IA reclama a existência de mecanismos de auditoria algorítmica que permitam a responsabilização no que concerne aos seus resultados e efeitos. Vias de reparação e compensação devem igualmente estar previstas.

Figura 4 - Sete princípios éticos para uma IA de confiança. Fonte: Elaborado a partir de Floridi (2019); Comissão Europeia (2019)

Segundo Luciano Floridi, membro do HLEG, a incerteza ética gera tanto riscos desnecessários como excesso de cautela, impedindo que se tire real proveito da tecnologia – seja por subutilização, uso desmedido ou malicioso (Floridi, 2019). As diretrizes visam orientar o desenvolvimento da IA a partir de consenso político e social, com deliberação ampla e visão de longo prazo sobre o seu propósito e prioridades. Não obstante a origem europeia destes princípios, e integrados num conjunto diversificado de orientações éticas, este quadro de valores ambiciona uma projeção global. Porém, para alcançar êxito será necessário superar uma série de desafios normativos, geopolíticos e sociais (Floridi, 2019). As páginas seguintes ocupar-se-ão destes desafios.

2.2. Decisão algorítmica: Autonomia e a crise da responsabilidade

À medida que o processo decisório é transferido para sistemas de IA, emergem debates sobre quem recai o título da responsabilidade. Desde veículos autónomos – capazes de adotar decisões em frações de segundo, suscetíveis de produzir consequências diretas, imediatas e potencialmente irreversíveis sobre a vida humana – até algoritmos utilizados em diagnósticos médicos, decisões judiciais ou processos de recrutamento, torna-se cada vez mais premente determinar quem responde pelos erros, danos ou injustiças resultantes dessas escolhas.

Inseridos na chamada Quarta Revolução Industrial, marcada pela centralidade da automação e substituição de tarefas humanas por processos mais eficientes, os sistemas de IA anunciam transformações profundas em múltiplos sectores (Rashid & Kausik, 2024). Da saúde ao setor militar, passando pelas telecomunicações, finanças, educação, manufatura, agricultura, energia e automóveis, a IA traduz-se num aumento exponencial da capacidade produtiva e na aceleração de processos de decisão, incluindo judiciais (Vasdani, 2020). Todavia, este avanço vem acompanhado de um problema crítico de *interpretabilidade*: os sistemas decidem, mas não conseguem explicar de forma transparente as razões subjacentes às suas decisões, o que deveria, desde logo – e especialmente no plano jurídico –, tornar estas decisões inaceitáveis (Benjamins & Salazar García, 2019).

Acresce que, por ora, a IA funciona apenas aparentemente de forma autónoma: esta depende do *input* humano para recolha, classificação e manutenção dos dados, e, apesar de identificar padrões, permanece incapaz de compreender o seu significado, como no caso de células cancerígenas (Dignum, 2023, p. 199) – reforçando o argumento da necessidade de responsabilidade sobre as decisões que produz. Não obstante ao debate sobre se se deve ou não criar uma “personalidade jurídica do robot” (Benjamins & Salazar García, 2019, p. 92), Virginia Dignum (2023) sublinha que os sistemas de IA são artefactos criados por seres humanos e que, a responsabilidade última sobre a sua utilização deve recair sobre aqueles que os desenvolvem e aplicam.

2.3. Três desafios de relevo na regulação da IA

Reconhecendo que o debate em torno da regulação da IA se estende por múltiplas dimensões que não podem ser aqui esgotadas, este trabalho privilegia uma leitura sintética de três desafios jurídico-políticos estruturais que condicionam a sua efetivação.

O primeiro prende-se com o chamado *pacing problem*³ (Yordanova & Bertels, 2024). Entre o momento de apresentação, voto e aprovação de uma lei, a tecnologia pode já a ter ultrapassado, criando quase como que um estado perpétuo de vazio legal propício ao desenvolvimento desregulado e a uma aplicação e utilização carente de monitorização

³ O *pacing problem* refere-se ao desfasamento significativo entre a velocidade do desenvolvimento da IA e a capacidade de inovação dos instrumentos regulatórios destinados a governá-la (Yordanova & Bertels, 2024).

adequada. Esta assimetria evidencia não apenas a diferença entre o ciclo de vida da IA e os ciclos políticos e humanos, mas também a necessidade de que as forças políticas assegurem a continuidade e o acompanhamento das medidas regulatórias independentemente das alternâncias de poder.

O segundo desafio consiste no problema da taxonomia, destacado pelo HLEG (Yordanova & Bertels, 2024): a ausência de uma definição única e consensual de IA. A pluralidade de conceitos e definições gera ambiguidades quanto ao objeto da regulação, comprometendo a clareza e eficácia normativa. Como sublinham Krafft et al. (2020), esta fragmentação conceptual dificulta a formulação de políticas adequadas, na medida em que investigadores tendem a adotar definições funcionais, centradas nas capacidades técnicas das tecnologias atuais, ao passo que decisores políticos privilegiam comparações com o pensamento e comportamento humano projetadas para cenários hipotéticos. Esta divergência pode levar à sobrevalorização de riscos futuros em detrimento dos desafios atuais (Krafft et al., 2020).

O terceiro desafio diz respeito à natureza e dimensão transnacional da IA, que contrasta com a ausência de uma governança global centralizada. Apesar dos avanços legais promovidos pela UE (AI Act, GDPR), EUA (National AI Initiative Act), China (Código de Ética para IA) – entre outros – e do setor privado, que aposta na autorregulação (GenAI: Content Risk Standards da Meta), o atual quadro regulatório revela-se insuficiente para lidar com as externalidades transfronteiriças da IA (Tallberg et al., 2023, p. 9). Ademais, esta multiplicidade de iniciativas contribui para a formação de um *mushrooming of declarations* (Floridi, 2019), fenómeno que produz inconsistências e confusão entre os *stakeholders*. Este processo conduz à criação de um “supermercado de princípios” (Floridi, 2019, p.2), isto é, um sistema de preferências e de *pick-and-choose* que prejudica a adaptação a um quadro ético-normativo comum (Floridi, 2019).

Capítulo III – Democracia, Algoritmos e Inteligência Artificial

Os obstáculos identificados no capítulo integram um debate mais amplo sobre o impacto da IA nas instituições políticas e na qualidade democrática, cuja regulação envolve não apenas dimensões técnicas e jurídicas, mas também o papel da informação na deliberação pública, na formação da opinião e na legitimidade democrática.

É nesta intersecção entre tecnologia e democracia que este capítulo se concentra, explorando dinâmicas de informação, desinformação e manipulação, os efeitos das *filter bubbles* (Pariser, 2011) e a crescente influência de sistemas algorítmicos sobre o espaço público. Ao fazê-lo, pretende-se compreender os desafios colocados à pluralidade democrática, refletir sobre possíveis transformações do juízo democrático, abordando ainda o que alguns autores designam por *algoracrias*.

3.1. (Des)Informação, Filter Bubbles e Manipulação de Opinião

A comunicação algorítmica – a gestão da informação através de sistemas que selecionam, priorizam e personalizam conteúdos – coloca sérios desafios à democracia. Ao facilitar a difusão de (des)informação, tem um enorme potencial de corrosão da base de conhecimento partilhada que sustenta o consenso e as instituições democráticas (Christiano, 2022). Como observa Harari (2024), o que temos em mãos é, sobretudo, um problema de rede, mais concretamente, de informação. A distorção da informação na qual assentam as sociedades humanas compromete, inevitavelmente, a qualidade das nossas decisões, independentemente do nosso grau de inteligência (Harari, 2024).

Em 2011, Eli Pariser alertava para a lógica algorítmica de plataformas como a Google e o Facebook – hoje comum a todas as grandes plataformas – cuja personalização de conteúdo nos dá a informação que queremos e não necessariamente a que precisamos, contribuindo para um isolamento intelectual sem precedentes. Dá-se assim a formação daquilo que Pariser denomina por *filter bubbles*, esferas de informação, que limitam o acesso à informação diversificada, reforçam tendências pré-existentes e isolam os indivíduos da comunidade (Pariser, 2011). Daqui radica a progressiva erosão do património cognitivo comum que sustenta o consenso social e as instituições democráticas.

Embora a manipulação algorítmica seja um fenómeno complexo, é possível, de acordo com Thomas Christiano (2022), identificar três processos a partir dos quais esta se manifesta na esfera pública: *filtering*, *hypernudging* e *microtargeting*. O *filtering*, consiste na seleção e organização da informação apresentada ao utilizador, limitando o acesso a conteúdos alternativos, reforçando fenómenos como as *filter bubbles*. *Hypernudging*, ou o hiper-direcionamento, é a reconfiguração contínua daquilo a que o utilizador tem acesso, que resulta na influência dos padrões de escolha sem que o utilizador disto de aperceba. Já o *microtargeting* consiste na orientação de mensagens a segmentos específicos, com recurso à exploração de dados pessoais e vulnerabilidades destes segmentos para reforçar a capacidade de persuasão.

Com o amadurecimento do ambiente digital, verifica-se que as redes sociais e plataformas digitais se consolidaram, não apenas como canais de socialização, comercialização e entretenimento, mas também como ferramentas de disseminação de conteúdo ideológico e mobilização política (e.g. caso *Cambridge Analytica*). E embora estes processos não tenham de ser aplicados numa ótica manipulativa e prejudicial, a comunicação algorítmica em que estas plataformas se baseiam exacerba o potencial de manipulação, dada a escala massiva de dados processados, grau de precisão e velocidade de adaptação (Christiano, 2022). O resultado é o controlo do fluxo de informação, o condicionamento da autonomia individual e a erosão da participação cívica plena.

3.2. Erosão do debate público e da pluralidade democrática

Ao analisar os desafios colocados ao debate público e à pluralidade democrática, torna-se crucial considerar de que forma as fragilidades epistémicas em determinados contextos – como comunidades desfavorecidas, áreas com acesso limitado à educação ou ambientes saturados de (des)informação – podem comprometer a capacidade dos indivíduos de pensar criticamente e de promover os seus interesses (Christiano, 2022).

Quando alguns cidadãos carecem de ferramentas cognitivas essenciais - como o acesso a informação de qualidade, competências de pensamento crítico ou plataformas de comunicação fidedignas – enquanto outros beneficiam de contextos que favorecem a aprendizagem, o envolvimento e o acesso a perspetivas diversificadas – cria-se uma desigualdade em termos de sofisticação política e de dinâmicas de poder que mina o princípio da igualdade em que a democracia assenta. Esta disparidade torna-se

particularmente evidente no contexto das comunicações algorítmicas, onde indivíduos com mais recursos conseguem dominar o discurso, exacerbando assimetrias pré-existentes. Consequentemente, a combinação entre conflitos de interesses e diferentes níveis de sofisticação política cria um cenário propício à manipulação (Christiano, 2022), colocando desafios significativos à vitalidade do debate público e à consolidação de uma democracia genuinamente pluralista.

Porém os desafios que se colocam ao debate público e à pluralidade democrática não se limitam às assimetrias de acesso à informação ou de sofisticação política. No centro desta dinâmica, uma classe em ascensão, ao controlar recursos estratégicos como dados e tecnologia, está a moldar novas dinâmicas de poder, que a seguir se analisam.

3.3. Algocracias e o eclipse do juízo democrático

O atual domínio da IA concentra-se sobretudo nas GAFAM (Google, Amazon, Facebook/Meta, Apple e Microsoft) e nas gigantes chinesas BAT (Baidu, Alibaba e Tencent), cuja vantagem competitiva assenta no acesso quase exclusivo a volumes massivos de dados de propriedade. Este cenário favorece a formação daquilo a que alguns autores apelidam de oligopólio tecnológico, caracterizado não apenas pela limitação da concorrência, mas também pelo risco de a inovação da IA permanecer enclausurada, acessível à sociedade apenas em condições definidas pelos interesses comerciais destas empresas (Benjamins & Salazar García, 2019). Esta concentração de poder económico e tecnológico que pode conduzir a um cenário em que a esfera pública fica subordinada a lógicas algorítmicas orientadas para a maximização do lucro, em detrimento dos interesses coletivos.

Shoshana Zuboff (2019), descreve a emergência de um “poder instrumentário” (p. 37) cuja lógica assenta em organizar, agregar e padronizar a sociedade, substituindo a política e a democracia pela certeza dos computadores e conformidade social. Zuboff aponta para os pioneiros deste modelo – Google e Facebook – que resistiram à governação corporativa e rejeitaram as exigências democráticas, assegurando as suas reivindicações através da sua força económica e política. O resultado é um sistema brutalmente eficaz em reduzir o espaço para a deliberação democrática, o debate público e a autodeterminação individual, subvertendo, a partir do seu interior, as bases da democracia e destituindo a soberania popular (Zuboff, 2019).

É neste contexto que surge o termo algocracia – governo por algoritmos. Uma forma de governação mediada por algoritmos, que interpela criticamente os ideais de pluralismo, participação democrática, debate público, transparência e responsabilidade institucional. John Danaher (2016) destaca um problema mais profundo do que a mera ocultação: a opacidade. Para além da ocultação, que compromete a responsabilização, a opacidade dos sistemas algorítmicos reduz a capacidade da sociedade de exercer escrutínio dado o potencial de incompreensibilidade destes processos para o raciocínio humano. Aqui reside o problema da legitimidade: a crescente dependência de decisões algorítmicas ameaça a governação pública, não apenas porque limita a responsabilização e a transparência, mas sobretudo porque torna a deliberação e o escrutínio dependentes de processos cujo funcionamento escapa à compreensão humana (Danaher, 2016).

Se de facto caminhamos para um mundo governado por algoritmos, a literacia digital deixa de ser opcional, torna-se, antes, imprescindível: emerge como um instrumento essencial para a participação cívica consciente e efetiva. É neste espírito que nos dirigimos ao último capítulo deste trabalho, onde se examinará, como, perante forças tecnológicas potencialmente maiores do que nós, podemos aprender a conviver, refletir e agir, garantindo que a nossa interação com a IA se funda na ética, na responsabilidade e na participação plena.

Capítulo IV – Literacia Digital e Cidadania: Ética, Educação e Equilíbrio na Era da IA

A literacia digital não é uma mera competência técnica: é uma condição mínima para coexistência com a IA. No atual cenário, em que algoritmos se afirmam para lá de uma mera ferramenta, mas como agentes (Harari, 2025) – que moldam o fluxo de informação, perceções e processos de tomada de decisões políticas e pessoais – a literacia digital emerge como um instrumento de sobrevivência intelectual e social. Mais do que dominar ferramentas digitais, trata-se de desenvolver uma cultura digital proativa: ser capaz de discernir entre factos e desinformação, avaliar a credibilidade de fontes e refletir sobre os impactos do digital no espaço público.

É a partir daqui que passamos para um exercício reflexivo, tendo em conta: a literacia digital enquanto competência crítica; o contributo da educação na edificação de um futuro ético; e o equilíbrio entre segurança e liberdade. Estes elementos determinarão o modo como nos preparamos para um futuro que já faz parte do presente.

4.1. Literacia Digital como competência crítica no século XXI

Se aceitarmos que o sistema descrito por Zuboff (2019) constitui “uma ameaça tão significativa para a natureza humana no século XXI quanto foi o capitalismo industrial para o mundo natural nos séculos XIX e XX” (p. 15), então temos, individual e coletivamente, a obrigação moral e ética de adquirir o *know-how* necessário para navegar neste ecossistema em segurança. A questão da autonomia de escolha não é algo que possamos projetar para o futuro: hoje, esta depende inerente do nosso grau de literacia digital.

À medida que o humano e o digital se fundem, de tal forma que são quase indissociáveis, não é exagerado afirmar que a IA constitui “uma revolução social a ocorrer em tempo real” (Beam, 2025). Da seleção de candidatos a um emprego à elegibilidade para crédito, dos conteúdos que consumimos às decisões judiciais, a incorporação da IA na arquitetura da vida moderna ultrapassa a esfera individual ou setorial. O tecido intelectual das sociedades humanas está a sofrer alterações significativas à medida que os nossos modos de interagir e de pensar começam a refletir lógicas mecanizadas - com efeitos profundos sobre a empatia, a intersubjetividade e os vínculos que estruturam a vida em sociedade (Beam, 2025).

Cumpra ainda notar que o ambiente digital funciona em rede: ao contrário da televisão tradicional, que fluía apenas do produtor para o espectador, os fluxos digitais são bidirecionais. Toda e qualquer interação deixa rastros; cada clique, *like* ou partilha comunica informação pessoal aos produtores do conteúdo – muitas vezes sem que o utilizador disto tenha consciência. Compreender estas dinâmicas que ocorrem por detrás do ecrã é fundamental, e é imprescindível que o utilizador tenha plena consciência daquilo de que está a abdicar, nomeadamente da sua privacidade, quando está online.

É imperioso que o nosso modelo de interação com a IA dê o salto do acesso imediato e indiscriminado, para uma participação crítica, informada e responsável. A literacia digital deve ser cultivada no quotidiano: saber identificar riscos, indagar recomendações, verificar a veracidade da informação, diversificar as ferramentas de IA a que acedemos e refletir sobre os nossos hábitos digitais – começar a assumir um papel ativo diante da tecnologia. Estas práticas podem ter um enorme impacto visto contribuir para a transformação do cidadão num participante consciente e reflexivo, abandonando a postura de um mero recetor, consumidor e produto. À luz destes desafios, importa avançar para a reflexão sobre como pensar a relação Homem-máquina.

4.2. Ensino, ética e sustentabilidade democrática

O conceito de literacia precisa de se expandir. Aprender a ler e escrever já não é suficiente: é imperioso que a componente digital seja integrada nos currículos escolares, desde o ensino básico até ao superior, assumindo-se como uma componente permanente no ensino. Como observa Dignum (2023), a literacia digital não deve permanecer circunscrita às áreas da engenharia: é necessário que se torne numa componente transversal no ensino, capaz de formar cidadãos críticos, conscientes dos impactos tecnológicos e aptos a participar de forma responsável e ética na sociedade digital.

A ausência de orientação adequada abre espaço a aprendizagens distorcidas, expondo os indivíduos a riscos significativos em virtude da sua suscetibilidade face a um ecossistema sobrecarregado de informação por vezes antagónica, imprecisa e falaciosa. Os jovens, embora familiarizados com a tecnologia, precisam de desenvolver competências para avaliar criticamente a informação disponível, identificar riscos e agir de forma responsável no ambiente digital. Ademais, e evocando o princípio da solidariedade intergeracional, a literacia digital deve estar ao alcance de todos: ao contrário dos nativos

digitais, uma parte significativa da população global não nasceu imersa na tecnologia e carece de maior atenção perante os riscos da interconectividade. Para que a IA não se torne um obstáculo ao pensamento crítico, o ensino deve ir além da transmissão técnica e promover uma literacia que preserve a autonomia intelectual humana. Iniciativas como a *AI4People – An Ethical Framework for a Good AI Society* (Floridi et al., 2018), o programa *Elements of AI*, lançado na Finlândia para democratizar o conhecimento tecnológico (Dignum, 2023), ou o trabalho desenvolvido pela organização *Center for Humane Technology*, refletem esforços no sentido de uma capacitação que transcenda a técnica, e que fomente a reflexão crítica e o discernimento, essenciais para uma cidadania consciente na era da IA.

O que se impõe, portanto, não é a proibição do uso da IA – designadamente no âmbito académico cuja restrição pode ser contraproducente – mas a sua integração no ensino e na sociedade, reconhecendo a inevitabilidade da sua presença. Ao não o fazer, está-se a submeter os utilizadores a um processo de aprendizagem não supervisionada, que tende a alimentar a ilusão de escolha e a erosão da subjetividade, com o risco acrescido de conduzir a um declínio significativo das capacidades intelectuais e criativas das novas gerações (Stezhko & Khmil, 2023). Como observa Dignum (2023), este desafio exige a diversificação dos saberes, incorporando contributos de áreas como a filosofia, o direito e a economia, de modo a formar cidadãos e profissionais capazes de articular técnica, ética e responsabilidade social.

A chave reside em apostar na amplificação da inteligência humana, e não na sua substituição, segundo uma lógica de *Intelligence Augmentation* (Hassani, 2020), em que a IA se configura como parceira das nossas faculdades cognitivas: a tecnologia não como supressora da autonomia do pensamento humano, mas como amplificadora, que abre novas possibilidades ao desempenho cognitivo humano. A partir daqui, partimos para a reflexão sobre um dos dilemas estruturantes na era da IA – o equilíbrio entre segurança e liberdade – que constitui o último ponto a abordar.

4.3. Equilíbrio entre segurança e liberdade

Não se trata de um equilíbrio fácil de alcançar; trata-se, antes, de uma espada de dois gumes. Há uma tensão clássica entre segurança e liberdade que se adensa à medida que as sociedades se digitalizam e aprofundam a sua integração num mundo progressivamente

mais interconectado. O uso da tecnologia para monitorização e vigilância pode reforçar a segurança coletiva – seja através da prevenção de crimes, gestão inteligente de sistemas urbanos ou da rápida resposta a situações de emergência –, mas simultaneamente ameaça liberdades fundamentais, sobretudo no domínio da privacidade e da autonomia individual, sujeitando os cidadãos a formas de controlo permanente e a riscos de discriminação algorítmica, cujos efeitos, conforme mencionado anteriormente, produz impactos assimétricos entre diferentes grupos sociais.

Com efeito, é inegável que tecnologias digitais de monitorização e análise de dados oferecem benefícios concretos e imediatos para a segurança pública. Ferramentas algorítmicas aplicadas à vigilância podem contribuir para a prevenção do crime, celeridade na deteção de ameaças e a gestão eficiente de situações de crise (Rashid & Kausik, 2024). Esta promessa de maior eficiência e previsibilidade explica, em parte, a sua adesão por parte da sociedade, sobretudo quando associada à proteção de bens essenciais, como a vida, a segurança pública ou o ambiente. Contudo, esta mesma promessa comporta um perigo estrutural: a normalização da vigilância permanente, e o redesenho das fronteiras entre esfera pública e privada (Zuboff, 2019). O dilema consiste, portanto, que em nome da segurança, se legitime uma arquitetura de controlo que fragiliza os princípios democráticos que deveria proteger.

Hoje, a advertência de Nick Bostrom de que a humanidade se comporta de forma semelhante a crianças a brincar com uma bomba (2016) ressoa com renovada urgência. Esta vulnerabilidade é reforçada por Harari (2024; 2025), ao recordar que a maior ameaça não reside apenas no potencial da IA em criar agentes mais inteligentes e criativos do que nós, mas também no facto de esse processo refletir a nossa própria natureza – competitiva, impaciente e demasiadas vezes inclinada a privilegiar o avanço em detrimento da segurança. E Floridi (2019) recorda-nos que esta tendência não é inédita: tanto a crise climática como a interferência das plataformas digitais na democracia já demonstraram os riscos associados à inovação sem planeamento prévio.

Benjamins e Salazar García (2019) levam esta reflexão ao seu limite hipotético: se existir um prazo-limite para fazer frente aos grandes desafios colocados pela IA, este poderá coincidir com a chamada Singularidade – o momento em que as máquinas ultrapassariam a inteligência humana. Mesmo que tal horizonte nunca se concretize, é imperativo que as sociedades humanas desenvolvam mecanismos de deliberação coletiva, com conceitos claros e enquadramentos sólidos, sob risco de deixarmos que decisões críticas, e que a

todos dizem respeito, sejam tomadas demasiado tarde, ou que fiquem à mercê de atores alheios ao interesse comum. Pois, se o poder do ser humano assenta na nossa capacidade de cooperação em larga escala, é também essa estrutura que nos predispõe a utilizá-lo de forma leviana (Harari, 2024).

Reflexões Finais

Vivemos um momento histórico em que a velocidade e escala da inovação tecnológica, especialmente da IA, ameaçam transformar instituições, modos de vida e o próprio conceito de humanidade. A IA, enquanto tecnologia com um alcance transversal e de natureza disruptiva, não é apenas uma ferramenta: é um espelho da humanidade que a cria, codifica e utiliza. Ao longo deste trabalho, ficou patente que o seu potencial é ambivalente – capaz tanto de ampliar capacidades humanas e democratizar o acesso ao conhecimento, como de reforçar desigualdades, fragilizar processos democráticos e comprometer a autonomia intelectual. A questão decisiva, portanto, não reside em perguntar se devemos ou não conviver com a IA – esta convivência já se tornou inevitável – mas em determinar de que forma a vamos integrar no nosso cotidiano e nas nossas instituições.

A análise desenvolvida permitiu delimitar três pilares essenciais para orientar o desenvolvimento e a integração da IA na sociedade: em primeiro lugar, a ética, que não deve ser entendida como um adorno, mas como condição mínima para que a inovação tecnológica se traduza em real progresso humano; em segundo lugar, a regulamentação, que não deve ser vista como entrave ao desenvolvimento, mas como quadro comum indispensável para assegurar que a IA se afirma como aliada da democracia, promovendo o equilíbrio entre segurança e liberdade, inovação e responsabilidade; e, por fim, a literacia digital, que transcende a dimensão instrumental e se configura como condição *sine qua non* para a cidadania plena, sem a qual os indivíduos se tornam mais vulneráveis à manipulação e privados da autonomia necessária para participar de forma consciente na construção de um futuro coletivo.

A IA não salvará nem condenará a humanidade: limitar-se-á a espelhar o que somos. Se guiada pela competição e pelo lucro imediato, refletirá esses vícios; se orientada pela cooperação, transparência e responsabilidade, poderá ser catalisadora de um futuro mais justo e sustentável. Integrar a IA de forma a preservar — e expandir — o nosso conceito de humanidade exige cultivar uma inteligência coletiva que assegure uma coexistência saudável entre inovação tecnológica e subjetividade humana. Somente a consciência que orienta as nossas escolhas pode determinar o devir da humanidade, e é a ética, mais do que qualquer algoritmo, que deve guiar esse caminho.

Bibliografia

- American Psychological Association. (2018). *Intelligence*. In *APA dictionary of psychology*. <https://dictionary.apa.org/intelligence>
- Baerren, E. (2025, February 19). *What happens if artificial intelligence becomes self-aware*. Central Michigan University. <https://www.cmich.edu/news/details/what-happens-if-artificial-intelligence-becomes-self-aware>
- Beam, B. (2025, April 2). *The social price of AI communication*. IE University. <https://www.ie.edu/insights/articles/the-social-price-of-ai-communication/>
- Bellman, R. (1957). A Markovian decision process. *Journal of Mathematics and Mechanics*, 6(5), 679–684. <https://doi.org/10.1512/iumj.1957.6.56038>
- Bellman, R. (1978). *An introduction to artificial intelligence: Can computers think?* Boyd & Fraser Publishing Company.
- Benjamins, V. R., & Salazar García, I. (2019). Towards a framework for understanding societal and ethical implications of artificial intelligence. In I. B. Alfonso, L. M. Fernández-Martínez, & R. Suárez-Álvarez (Eds.), *Vulnerabilidad y cultura digital: Riesgos y oportunidades de la sociedad hiperconectada* (pp. 89–100). Dykinson, S.L. <https://doi.org/10.48550/arXiv.2001.09750>
- Bostrom, N. (2016, June 12). *Artificial intelligence: 'We're like children playing with a bomb'*. The Guardian. <https://www.theguardian.com/technology/2016/jun/12/nick-bostrom-artificial-intelligence-machine>
- Christiano, T. (2022). Algorithms, manipulation, and democracy. *Canadian Journal of Philosophy*, 52(1), 109–124. <https://doi.org/10.1017/can.2021.29>
- Crume, J. (2024, August 5). *AI, machine learning, deep learning and generative AI explained* [Video]. YouTube. <https://www.youtube.com/watch?v=qYNweeDHiyU>
- Damásio, A. (2000). *O sentimento de si: O corpo, a emoção e a neurobiologia da consciência*. Mem Martins: Publicações Europa-América.
- Danaher, J. (2016). The threat of algocracy: Reality, resistance and accommodation. *Philosophy & Technology*, 29(3), 245–268. <https://doi.org/10.1007/s13347-015-0211-1>
- Daylight, E. G. (2015). Towards a historical notion of Turing – the father of computer science. *History and Philosophy of Logic*, 36(3), 205–228. <https://doi.org/10.1080/01445340.2015.1082050>
- Dignum, V. (2023). Responsible artificial intelligence: Recommendations and lessons learned. In D. O. Eke, K. Wakanuma, & S. Akintoye (Eds.),

Responsible AI in Africa (pp. 195–214). Palgrave Macmillan.
https://doi.org/10.1007/978-3-031-08215-3_9

European Commission. (2019). *Ethics guidelines for trustworthy AI*.
<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

European Commission: Directorate-General for Communications Networks, Content and Technology. (2020). *The assessment list for trustworthy artificial intelligence (ALTAI) for self-assessment*. Publications Office.
<https://data.europa.eu/doi/10.2759/002360>

Floridi, L. (2019). Establishing the rules for building trustworthy AI. *Nature Machine Intelligence*, 1(6), 261–262. <https://doi.org/10.1038/s42256-019-0055-y>

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Vayena, E. (2018). AI4People – An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>

Harari, Y. N. (2023, April 3). *X*.
https://x.com/harari_yuval/status/1642814487551000576

Harari, Y. N. (2024, August 24). *Never summon a power you can't control: Yuval Noah Harari on how AI could threaten democracy and divide the world*. *The Guardian*.
<https://www.theguardian.com/technology/article/2024/aug/24/yuval-noah-harari-ai-book-extract-nexus>

Harari, Y. N. (2025, June 21). *AI and human evolution* [Video]. YouTube.
<https://www.youtube.com/watch?v=jt3U13rPXaE>

Hassani, H., Silva, E. S., Unger, S., TajMazinani, M., & Mac Feely, S. (2020). Artificial intelligence (AI) or intelligence augmentation (IA): What is the future? *AI*, 1(2), 143–155. <https://doi.org/10.3390/ai1020008>

Krafft, P. M., Young, M., Katell, M., Huang, K., & Bugingo, G. (2020, February 7). Defining AI in policy versus practice. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (pp. 195–201). Association for Computing Machinery.
<https://doi.org/10.1145/3375627.3375835>

Pariser, E. (2011, March). *Beware online “filter bubbles”* [Video]. YouTube.
<https://www.youtube.com/watch?v=B8ofWFx525s>

Rai, D. H. (2024). *Artificial intelligence through time: A comprehensive historical review*. ResearchGate.
<https://doi.org/10.13140/RG.2.2.22835.03364>

Rashid, A. B., & Kausik, M. A. K. (2024). AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications. *Hybrid Advances*, 7, 1002727.
<https://doi.org/10.1016/j.hybadv.2024.100277>

- Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (3rd ed.). Pearson.
<https://people.engr.tamu.edu/guni/csce625/slides/AI.pdf>
- Smith, C., McGuire, B., Huang, T., & Yang, G. (2006). *The history of artificial intelligence*. University of Washington.
<https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf>
- Stanford Encyclopedia of Philosophy. (2014, January 14). *Consciousness*. In Stanford University. <https://plato.stanford.edu/entries/consciousness/>
- Stezhko, Z. V., & Khmil, T. V. (2023). Artificial intelligence as a socio-cultural phenomenon: The educational dimension. *Anthropological Measurements of Philosophical Research*, 24, 68–74.
<https://doi.org/10.15802/ampr.v0i24.295317>
- Tallberg, J., Erman, E., Furendal, M., Geith, J., Klamberg, M., & Lundgren, M. (2023). The global governance of artificial intelligence: Next steps for empirical and normative research. *International Studies Review*, 25(3). <https://doi.org/10.1093/isr/viad040>
- Vasdani, T. (2020, February 5). *Robot justice: China's use of internet courts*. *Law360 Canada*. <https://www.law360.ca/ca/articles/1750396/robot-justice-china-s-use-of-internet-courts>
- Yordanova, K., & Bertels, N. (2024). Regulating AI: Challenges and the way forward through regulatory sandboxes. In H. Sousa Antunes, P. M. Freitas, A. L. Oliveira, C. Martins Pereira, E. Vaz de Sequeira, & L. Barreto Xavier (Eds.), *Multidisciplinary perspectives on artificial intelligence and the law* (pp. 441–456). Springer.
https://doi.org/10.1007/978-3-031-41264-6_23
- Zuboff, S. (2019). *A era do capitalismo de vigilância: A luta por um futuro humano na nova fronteira do poder* (1ª ed.). Intrínseca.
<https://nae.com.pt/wp-content/uploads/A-Era-do-Capitalismo-de-Vigilancia-Shoshana-Zuboff.pdf>

