

O Impacto da Inteligência Artificial nas Relações Internacionais – Povo e Poder

Dissertação de Mestrado

Patrícia Silva

Mestrado em

**Relações Internacionais – O Espaço
Euro-Atlântico**



O Impacto da Inteligência Artificial nas Relações Internacionais – Povo e Poder

Dissertação de Mestrado

Patrícia Silva

Orientadores

Professor Doutor Paulo Vitorino Fontes

Professor Doutor Rui Sampaio da Silva

Dissertação de Mestrado submetida como requisito parcial para obtenção do grau de Mestre em Relações Internacionais – O Espaço Euro-Atlântico.

Conteúdo

DEDICATÓRIA	5
AGRADECIMENTOS	7
RESUMO.....	8
ABSTRACT	9
INTRODUÇÃO.....	10
CAPÍTULO I	14
ENQUADRAMENTO TEÓRICO - O PODER NA ERA DIGITAL	14
1. DE FERRAMENTA DE SOBREVIVÊNCIA À ERA DA INFORMAÇÃO	14
1.2. “AS MÁQUINAS PENSAM?”	15
1.3. IA: QUATRO ABORDAGENS FUNDAMENTAIS.....	16
1.1.3. O COMBUSTÍVEL DA IA	17
1.4. MACHINE LEARNING E DEEP LEARNING: UMA DISTINÇÃO FUNDAMENTAL.....	18
1.5. A REVOLUÇÃO DA IA GENERATIVA	19
1.6. O VAZIO DA IA.....	20
CAPÍTULO II	21
TECNOLOGIA E DESIGUALDADE SOCIAL: UM RETRATO HISTÓRICO	21
2. TECNOLOGIA MILENAR	21
2.1. A METAMORFOSE DAS RELAÇÕES NAS REVOLUÇÕES TECNOLÓGICAS	22
2.3. DINÂMICAS DO POPULISMO MODERNO À LUZ DA CRISE TECNOLÓGICA	23
2.4. O POVO NA FRAGMENTAÇÃO ENTRE ATOR POLÍTICO E RECURSO SOCIAL.....	25
CAPÍTULO III	28
DISTOPIA PANÓTICA - ALIENAÇÃO E PODER	28
3. O FIM DA ESFERA PÚBLICA?	28
3.1. DO VÍNCULO AO VAZIO	29
3.2. A CRÍTICA DE FRANKFURT E A (RE)CONFIGURAÇÃO DA ESFERA PÚBLICA DE HABERMAS	32
3.3. O NOVO NIILISMO: COMO A IA DESESTABILIZA A REALIDADE E A ORDEM GLOBAL	37
3.4. A GEOPOLÍTICA DA ILUSÃO.....	39
3.5. A DESINFORMAÇÃO NA CONTEMPORANEIDADE.....	42
CAPÍTULO IV	49
A CRISE DE CONSCIÊNCIA E O PRAGMATISMO NA ERA DIGITAL O PESSIMISMO VERSUS O OTIMISMO NA ERA DA IA.....	49
4. O RACIONALISMO VERSUS O EMPIRISMO NA ERA DA IA	49
4.1. GEOFFREY HINTON: A CRISE DE CONSCIÊNCIA E O ALERTA EXISTENCIAL.....	50
4.2. KAI-FU LEE: O PRAGMATISMO NA ERA DA IMPLEMENTAÇÃO	52
4.3. A GEOPOLÍTICA DA INOVAÇÃO PÓS-NUCLEAR	54
4.4. A POLARIZAÇÃO DA INOVAÇÃO	56
4.4.1 O FATOR HUMANO NA SUPREMACIA TECNOLÓGICA.....	57
4.5. A CHINA NO EQUILÍBRIO DE PODERES DA IA	58
4.5.1 RECONFIGURAÇÃO GEOPOLÍTICA PELA INTELIGÊNCIA ARTIFICIAL.....	60
4.6. DUAS FACES DE UM MESMO FENÔMENO	62
CAPÍTULO V	64

CASOS DE ESTUDO	64
5.1. O CLIMA DE IDEOLOGIAS IRREALISTAS E A IA: A EXPERIÊNCIA CAMBRIDGE ANALYTICA COMO BASE PARA A FORMULAÇÃO DE UMA REGULAÇÃO INTERNACIONAL PARA TECNOLOGIAS DE RISCO...	64
5.2. DA PROPAGANDA COMPORTAMENTAL À ORQUESTRAÇÃO ALGORÍTMICA: O BREXIT COMO BARÓMETRO DA VULNERABILIDADE DEMOCRÁTICA FACE AO IMPACTO DA IA	69
CONSIDERAÇÕES FINAIS	72
BIBLIOGRAFIA	79
ARTIGOS E ESTUDOS CIENTÍFICOS	83
WEBGRAFIA.....	85

Dedicatória
à Filha do Barbeiro

Agradecimentos

Professor Doutor Paulo Fontes,

Obrigada por ter sido o primeiro a acreditar no meu tema e aceitar ser meu orientador. As suas histórias e exemplos impactaram na minha visão do trabalho que tinha pela frente e a acreditar nas minhas ideias e convicções. O seu apoio incondicional, mesmo quando eu não tinha tanta certeza, fizeram toda a diferença. Dizem que não podemos mudar o mundo, mas se mudarmos o mundo de alguém, a nossa missão está cumprida e o professor, certamente muda os mundos de muitos de nós!

Professor Doutor Rui Sampaio,

A sua disponibilidade e humildade imediatas quando o convidei para meu orientador, demonstra tudo o que faz de si o professor de referência que é para esta Universidade e para mim. Se encontrei o meu caminho, foi com a sua ajuda, as suas notas e sugestões sempre minuciosas. Obrigada por ter entrado neste tema connosco e fazer toda a diferença!

Mãe,

Pai,

Tinham razão. Obrigada!

Resumo

A emergência da Inteligência Artificial (IA) representa, simultaneamente, a maior promessa de progresso da História e a mais significativa ameaça existencial à civilização humana. Este estudo investiga o impacto da IA na soberania individual e dos Estados, questionando se o controlo algorítmico, que progressivamente antecipa e molda a tomada de decisões, está a transformar a sociedade em meros indivíduos alienados dos verdadeiros pilares da soberania política. O impacto da IA transcendeu o plano potencial para se manifestar como uma força determinante na geopolítica e na política doméstica. A investigação foca-se na mitigação dos riscos que esta tecnologia representa para as Relações Internacionais, com particular ênfase na propagação massiva de desinformação, na manipulação da opinião pública e na alteração do comportamento eleitoral. A ausência de uma entidade reguladora global criou um vácuo ético que ameaça a dignidade, a privacidade e os direitos fundamentais em democracias tidas como sólidas. Através da Teoria Crítica da Esfera Pública de Jürgen Habermas e pela sua Ação Comunicativa compreenderemos a evolução da comunicação e o seu impacto social. Este enquadramento é complementado pela análise de Jason Stanley que demonstra como as plataformas digitais — através das "bolhas de filtro" e "câmaras de eco" — exacerbam a polarização ideológica e subvertem o pluralismo político, aproximando as sociedades de discursos autoritários e populistas. Para evidenciar a instrumentalização da IA por agentes políticos, serão analisados dois casos de estudo cruciais: as Eleições Presidenciais dos EUA, de 2016 (e o papel da Cambridge Analytica) e o Referendo do Brexit, de 2016. Estes eventos demonstram a enorme capacidade que a IA tem em moldar perspetivas e alterar o curso da história através da manipulação informacional direcionada, alertando para a gravidade dos riscos. O estudo visa, assim, fornecer uma análise minuciosa do panorama atual, sublinhando a imperiosa necessidade de estudar o passado para salvaguardar o futuro da ordem democrática contra a repetição de erros históricos de manipulação e propaganda, propondo uma clara regulamentação da IA.

Palavras-chave: Inteligência Artificial, Desinformação, Esfera Pública, Geopolítica, Regulamentação.

Abstract

The emergence of Artificial Intelligence (AI) simultaneously represents the greatest promise of progress in History and the most significant existential threat to human civilization. This study investigates the impact of AI on individual and State autonomy, questioning whether algorithmic control, which progressively anticipates and shapes decision-making, is transforming society into mere individuals alienated from the true pillars of political autonomy. AI's impact has transcended the potential realm to manifest as a determining force in geopolitics and domestic politics. The research focuses on mitigating the risks this technology poses to International Relations, with particular emphasis on the massive spread of disinformation, the manipulation of public opinion, and the alteration of electoral behavior. The absence of a global regulatory body has created an ethical vacuum that threatens dignity, privacy, and fundamental rights in otherwise solid democracies. Through Jürgen Habermas's Critical Theory of the Public Sphere and his Communicative Action, we will understand the evolution of communication and its social impact. This framework is complemented by Jason Stanley's analysis, which demonstrates how digital platforms—through "filter bubbles" and "echo chambers"—exacerbate ideological polarization and subvert political pluralism, bringing societies closer to authoritarian and populist discourses. To demonstrate the instrumentalization of AI by political actors, two crucial case studies will be analyzed: the 2016 US Presidential Elections (and the role of Cambridge Analytica) and the 2016 Brexit Referendum. These events demonstrate the enormous capacity of AI to shape perspectives and alter the course of history through targeted informational manipulation, alerting to the gravity of the risks. The study thus aims to provide a thorough analysis of the current landscape, emphasizing the imperative need to study the past to safeguard the future of the democratic order against the repetition of historical errors of manipulation and propaganda, proposing a clear regulation of AI.

Keywords: Artificial Intelligence, Disinformation, Public Sphere, Geopolitics, Regulation.

Introdução

A crescente evolução e disseminação da Inteligência Artificial e das suas ferramentas, representa, paradoxalmente, até hoje na História, a maior promessa de progresso e, ao mesmo tempo, a maior ameaça à civilização humana. Conforme proferido por Stephen Hawking em 2017, durante a Web Summit, em Lisboa, a IA poder ser “a pior ou a melhor coisa que já aconteceu para a humanidade”. Este alerta, reiterado após uma premonição similar em 2014, é particularmente relevante quando surge de um indivíduo cuja própria capacidade de comunicação e longevidade foram radicalmente transformadas pela alta tecnologia. A sua experiência de vida, um testemunho do poder revolucionário da tecnologia assistida, sublinha o fascínio e a dependência que caracterizam a nossa relação contemporânea com o digital. Contudo, a questão central que emerge é perturbadora: estaremos coletivamente presos a um “corpo inerte” onde o pensamento, a tomada de decisões ou as ações mais básicas estão paulatinamente a ser moldadas por algoritmos que antecipam e controlam, colocando em xeque a soberania individual e, numa escala mais ampla, a soberania política dos Estados?

O impacto da Inteligência artificial transcendeu, há muito, o domínio da ameaça potencial para se manifestar como uma realidade influente em questões primárias globais, como a capacidade de alterar o curso das relações externas e da política doméstica, através do progressivo controlo dos cidadãos. O foco deste estudo recai sobre a repercussão desta ferramenta na organização interna de um Estado soberano e as sequelas que já são visíveis e sentidas na geopolítica. A preocupação é legítima e exige um lugar central nos debates políticos, uma vez que a expansão desmensurada da IA já ultrapassou os limites éticos e morais, ameaçando os pilares que definem a nossa humanidade como, a dignidade, a privacidade e os direitos fundamentais. A ausência de uma entidade reguladora global para a Inteligência Artificial criou um vácuo ético, promovendo a disseminação de desinformação numa escala massiva, que desafia os princípios de igualdade e progresso em democracias que até agora eram consideradas sólidas.

O cerne da investigação consiste na mitigação dos riscos que a IA representa nas Relações Internacionais, com particular ênfase na propagação

de informações falsas na manipulação da opinião pública e na alteração do comportamento dos cidadãos em atos eleitorais. Para analisar como a Inteligência Artificial está a ser instrumentalizada por governos e agentes políticos, estas investigações propõem recorrer ao método indutivo sob uma perspetiva qualitativa, baseada em pesquisa bibliográfica e com a análise de dois casos de estudos bastante relevantes. No que diz respeito ao enquadramento teórico, este será suportado pela teoria crítica da esfera pública, de Jürgen Habermas e pela sua visão da ação comunicativa, que se revelam, mais do que nunca, cruciais para compreender a evolução da comunicação e o seu impacto nas sociedades modernas.

O objetivo passa por estabelecer uma resenha histórica que permita compreender a génese da propagação desmedida da informação e o seu efeito na política. A sociedade do século XXI, marcada pela emergência de plataformas digitais como, o Facebook, Instagram, X/Twitter, WhatsApp, criou um ecossistema perfeito para o crescimento de ferramentas que consomem e tratam dados pessoais a uma velocidade vertiginosa e inatingível pela capacidade humana. Alimentadas por estes dados e através da astúcia dos algoritmos, formaram-se as chamadas “bolhas de filtro” e as “câmaras de eco”. Estas estruturas digitais, como veremos, expõem minorias a informações que apenas reforçam as suas visões preexistentes, tendendo, assim, à polarização política e ao despertar de ideias extremistas, como tem vindo a ser tendência um pouco por toda a Europa. O pluralismo político é assim subvertido pela restrição da informação, alienando os cidadãos de perspetivas alternativas e do “mundo real”, um fenómeno que nos aproxima gradualmente de discursos e políticas autoritárias e dissimuladas.

A divisão da sociedade, um tema central na análise de Jason Stanley em *Como Funciona o Fascismo*, a Política do Nós e Eles, é exacerbada pelo fenómeno das bolhas de filtro. Stanley refere que a política fascista apela à distinção fundamental entre “nós” e “eles”, utilizando divisões étnicas, religiosas ou raciais para moldar uma ideologia. Este desfasamento ideológico e a fragmentação crescente têm catapultado para o debate público questões que remetem a passados de regimes autoritários, impulsionadas por movimentos populistas. O debate público é limitado, espartilhado, sem espaço para a construção de consensos, e a narrativa ideológica é frequentemente depositada

na figura de um líder carismático, que decide o que é melhor para a sociedade, resultando numa combinação perigosa de apatia e desordem social.

Também a análise crítica de Habermas à racionalidade instrumental que antecipava os perigos de uma primazia da eficiência em detrimento da democracia, demonstra a sua extrema relevância na análise da IA. Os novos desafios que a sociedade democrática enfrenta, aumentados pela expansão tecnológica e pela desinformação. Criam conflitos internos e abalam as eleições internas dos países. As ascensões de regimes extremistas ganham expressão pela velocidade da propagação da mensagem digital. A teoria da ação comunicativa de Habermas e a análise de Stanley sobre o discurso de ódio convergem na identificação deste como uma ameaça existencial à esfera pública democrática. O perigo é ainda maior quando se considera o papel da propaganda, por isso, e recorrendo à perspetiva histórica, a propaganda, como demonstrado por figuras como a de Hitler, visa substituir a argumentação racional, manipulando medos e as inseguranças de um povo para consolidar o poder e criar um “inimigo comum”. Esta incitação ao medo desvia a atenção dos reais problemas a favor de uma narrativa que, numa sociedade completamente ofuscada, justifica atos hediondos. Com o entendimento de que o ser humano está sempre presente em todos os cenários que apresentamos é crucial perceber o seu papel nesta era digital.

Para complementar esta dissertação serão apresentados dois casos de estudo que evidenciam o impacto que a IA e a manipulação de dados têm na política contemporânea. O primeiro será o caso das eleições presidenciais norte-americanas de 2016, que levaram à eleição de Donald Trump, e o papel da empresa Cambridge Analytica na exploração de dados para a difusão de mensagens com targets específicos; o segundo caso de estudo será sobre o referendo, de 2016, que resultou na saída do Reino Unido da União Europeia, em 2020, mais conhecido como Brexit, outro exemplo em como a manipulação da informação e dos eleitores através de campanhas digitais mudou o curso da história. Estes eventos, aparentemente distintos, estão intrinsecamente ligados ao desenvolvimento das novas tecnologias e à enorme capacidade que a Inteligência Artificial tem para mudar perspetivas enraizadas e moldar as mentes, especialmente, do eleitorado mais jovem e desinformado. A urgência da regulamentação da IA nunca foi tão premente. O facto de governos europeus já

terem começado a ter consciência e a tomar medidas, a verdade é que estamos no meio de duas superpotências que querem atingir a supremacia tecnológica – China e Estados Unidos. Ao adotar uma posição protecionista a União Europeia poderá ficar para trás na corrida às novas tecnologias e os países mais pequenos e dependentes criarão um fosso de desigualdade sem precedentes.

A resposta a estas questões, baseada na pesquisa bibliográfica de vários autores, não só da área tecnológica, será fundamental para garantir um estudo minucioso e abrangente do panorama atual, pois, é sempre importante estudar o passado, para compreender o presente e, sobretudo, para não voltar a repetir os mesmos erros.

Capítulo I

Enquadramento Teórico - O Poder na Era Digital

1. De Ferramenta de Sobrevivência à Era da Informação

Um dos poderes da tecnologia é o de surpreender o homem pela sua inovação e novas capacidades. Quando se fala em tecnologia também não é difícil de prever a imagem que surge na mente das pessoas. Normalmente, a palavra está associada a computadores, telemóveis inteligentes, relógios que se conectam com outros dispositivos e tudo o que começa com o prefixo “*smart*”. A nossa mente está formatada para esta era digital, mas a tecnologia surge muito antes do primeiro *computador*, palavra que era usada, “antes da Segunda Grande Guerra, para se referir a pessoas (...) que executavam (...), os cálculos necessários (...) maioritariamente militares.” (OLIVEIRA, p. 14). No século XVII, Thomas Hobbes, argumentava na sua obra *Leviatã* que “o raciocínio humano resulta da manipulação aritmética de símbolos, sendo talvez esta a primeira conceptualização do pensamento humano como processamento digital de informação.” (OLIVEIRA, p. 13). Para Arlindo Oliveira, professor do Instituto Superior Técnico e Investigador no Instituto de Engenharia de Sistemas e Computadores (INESC), a tecnologia surge ainda antes do pensamento de Thomas Hobbes uma vez que “a capacidade para desenvolver tecnologia não resulta diretamente da superior inteligência do ser humano, mas sim da propagação do conhecimento entre gerações, através da cultura”. (OLIVEIRA, p. 27). Esta afirmação de Arlindo Oliveira, levou-nos à pesquisa do significado etimológico da palavra “tecnologia” que, de acordo com o Dicionário da Língua Portuguesa, da Academia das Ciências de Lisboa, define como “conjunto dos instrumentos, métodos e processos específicos de qualquer arte, ofício ou técnica”.

Recorrendo à afirmação de Hobbes (1651) sobre a manipulação de símbolos, ao conceito da cultura através da propagação do conhecimento, (cf. Oliveira, 2019) e à descrição no dicionário, a tecnologia ocorre como uma capacidade desenvolvida antes de tudo o que foi mencionado até agora. Podemos afirmar que a primeira conceção de tecnologia descende do *Homo*

Sapiens, “o que permite aos membros da nossa espécie exibir características que, normalmente, associamos a comportamento inteligente.” (OLIVEIRA, p. 25.). Até ao século XX, quase todas as tecnologias, podemos assim dizer, foram desenvolvidas para a manipulação de alimentos, materiais e de outros produtos, refletindo as reais necessidades da sociedade da época, sendo que “as únicas exceções são a linguagem escrita e a matemática, que são essencialmente, formas de manipular, comunicar e preservar a informação, e as tecnologias de comunicação com base na eletricidade, que se iniciaram com o telégrafo e o telefone.” (OLIVEIRA, p. 32). Atualmente, as mais simples interações do nosso quotidiano estão cada vez mais dirigidas à capacidade computacional que criou ao que a sociedade científica designou de Quarta Revolução Industrial e abriu caminho à seguinte. Não foi apenas o desenvolvimento das tecnologias eletrónicas que caracterizou a quarta revolução industrial, a ela veio juntar-se o carvão e o petróleo como novas e importantes fontes de energias para fins não só civis como militares aumentando assim o poder dos Estados. Um significativo passo para a compreensão das leis da física que “conduziu a um domínio mais eficaz dos materiais, fator que também foi importante na segunda metade do século XX.” (OLIVEIRA, p. 33).

1.2. “As máquinas pensam?”

A ascensão da computação provocou um questionamento crucial e multifacetado, com implicações tanto filosóficas como pragmáticas: poderá uma máquina, algum dia, exibir uma inteligência análoga à humana? Este dilema capturou a atenção de cientistas, engenheiros e filósofos, impulsionando a busca por uma resposta. Tal investigação profícua culminou na criação de um novo campo de estudo e pesquisa: a Inteligência Artificial (IA). Em 1950, Alan Turing, considerado um dos matemáticos mais influentes e criativos de todos os tempos, publicou o artigo seminal "Computing Machinery and Intelligence" no qual levantou a provocadora questão: "As máquinas pensam?" Para investigar esta questão, ele propôs o que viria a ser conhecido como o "Teste de Turing", um estudo engenhoso para avaliar a capacidade de uma máquina de exibir comportamento inteligente indistinguível do humano. O teste, originalmente

formulado como o "Jogo da Imitação" envolvia um interrogador com a tarefa de tentar distinguir um homem de uma mulher através de conversação por texto. Numa variação posterior e mais relevante para a IA, o objetivo do interrogador passou a ser o de determinar se as respostas eram geradas por um humano ou por uma máquina. Turing foi notavelmente otimista, prevendo que, dentro de 50 anos, as máquinas seriam capazes de passar no teste em cinco minutos. A sua visão, embora audaciosa para a época, foi o incentivo que levou a comunidade científica a sair da mera teorização e a passar à ação. Inspirados por esta visão e pelo progresso na acessibilidade dos computadores digitais a partir dos anos 50, Turing e outros cientistas rapidamente se dedicaram ao desafio de construir máquinas capazes de replicar a inteligência humana. Embora as previsões de Turing não se tenham concretizado no prazo que ele estipulou, a sua abordagem marcou o início de uma transição crucial da teoria para a prática, impulsionando o desenvolvimento do campo da Inteligência Artificial. Atualmente, com sistemas como o GPT-5, testemunhamos um avanço sem precedentes, cujas capacidades excedem em muito o que se poderia imaginar na época de Turing, e que exploraremos com mais pormenor nos pontos seguintes.

1.3. IA: Quatro Abordagens Fundamentais

Após o término da Segunda Guerra Mundial, o potencial dos computadores transcendeu o domínio militar, despertando um intenso interesse na sua aplicação a novos campos. A Inteligência Artificial emergiu rapidamente como uma área de investigação de especial relevância, canalizando o desenvolvimento de máquinas pré-existentes para um propósito inovador: a replicação da inteligência. O cientista conhecido como o “pai da ciência da computação” e pioneiro na área da Inteligência Artificial, John McCarthy, diz que “a ciência e a engenharia de produzir máquinas inteligentes está relacionado à tarefa semelhante de usar computadores para entender a inteligência humana, mas a IA não precisa de se limitar a métodos de observação biológica”. (McCarthy, 1956). O termo – Inteligência Artificial - foi usado pela primeira vez numa conferência de especialistas em Dartmouth College, sob o nome “O Eros Eletrónico”. Numa fase mais contemporânea, Stuart Russel e Peter Norvig

escreveram o livro “Inteligência Artificial: Uma abordagem moderna” no qual apresentam ao leitor quatro potenciais objetivos da Inteligência Artificial, distinguindo-os com base na racionalidade e pensamento versus ação. Por um lado, apresentam-nos a abordagem humana que é composta por dois sistemas: os sistemas que pensam como os humanos e os sistemas que agem como os humanos (nesta abordagem poderíamos integrar a teoria de Turing); por outro lado, a abordagem ideal: um sistema com pensamento racional e um sistema com atos racionais.

1.1.3. O Combustível da IA

A Inteligência Artificial surge como um campo interdisciplinar que se alimenta da ciência da computação e de vastos conjuntos de dados para solucionar problemas complexos. É nesta intersecção que reside a simbiose fundamental entre a IA e o *Big Data*, um termo que, como salientam Stuart Russell e Peter Norvig, não se restringe apenas ao volume colossal de dados “absorvidos”, mas abrange igualmente a sua variedade, velocidade e veracidade (Russell & Norvig, 2010). Assim, o *Big Data* transcende a mera dimensão quantitativa, referindo-se a conjuntos de dados tão grandes e complexos que exigem abordagens computacionais extremamente avançadas para o seu processamento. A sua relevância, contudo, é multifacetada, englobando também a velocidade vertiginosa com que são gerados os dados a partir de fontes tão diversas como as interações em redes sociais, as configurações de dispositivos móveis e as transações comerciais. Este fluxo contínuo de informação é o combustível essencial que nutre os algoritmos de IA, fornecendo o “petróleo” necessário para o seu funcionamento. Também conhecido como novo “petróleo”, analisaremos nos capítulos finais, a partir da perspectiva de Geoffrey Hinton, o porquê da escolha desta palavra.

Empresas e organizações de topo reconhecem a importância estratégica do *Big Data* para identificar padrões de consumo e, assim, conceber produtos com maior probabilidade de sucesso no mercado. A aplicação prática desta sinergia é demonstrada em casos notáveis, como o que Kissinger, Schmidt e Huttenlocher (2021) referem: em 2016, uma IA da *DeepMind* conseguiu reduzir

os gastos de energia dos centros de dados da Google em 40%, superando o que os melhores engenheiros da empresa haviam conseguido previamente. Este feito ilustra a capacidade da IA em extrair valor de montanhas de dados que seriam impossíveis de processar manualmente, transformando-os em conhecimento acionável e otimizações sem precedentes, colocando o humano aquém da máquina e sem qualquer hipótese de competição e, muito menos, de superação.

1.4. Machine Learning e Deep Learning: Uma Distinção Fundamental

Conforme evidenciado, a eficácia da Inteligência Artificial está intrinsecamente ligada à qualidade e quantidade dos dados que a alimentam. Essa dependência otimiza o desempenho dos algoritmos de *Machine Learning* (ML), um subcampo fundamental da IA que, por sua vez, é o alicerce para conceitos ainda mais avançados. *Machine Learning* refere-se à criação de sistemas que aprendem e aprimoram o seu desempenho a partir da análise de dados, sem serem explicitamente programados para cada tarefa. É importante ter em conta que, embora toda a *machine learning* seja um tipo de IA, o contrário não se aplica. Por sua vez, o *Deep Learning* (DL) representa uma especialização dentro da *machine learning* (ML). A principal distinção entre os dois reside na forma como cada algoritmo aprende. Na *machine learning* clássica, a intervenção humana é frequentemente necessária para a extração de características dos dados. Já o *deep learning*, por utilizar vastas redes neuronais com múltiplas camadas consegue garantir grande parte deste processo com a sua capacidade de automatização. Esta capacidade permite que processe e aprenda padrões complexos a partir de volumes de dados significativamente maiores. Conforme Lex Fridman (2019) afirmou numa palestra no MIT, podemos pensar no *deep learning* como "*Machine Learning* escalável", uma vez que a sua arquitetura permite uma aprendizagem mais autónoma e eficiente em grande escala.

1.5. A Revolução da IA Generativa

A Inteligência Artificial Generativa, um avanço notável que emergiu do *deep learning*, tem como foco a criação de modelos capazes de gerar novos dados a partir do que foi aprendido. Inicialmente, a IA limitava-se a analisar dados numéricos; a IA Generativa expandiu o seu alcance para tipos de dados mais complexos, como imagens, vozes e textos. Um marco significativo nesta evolução foi a introdução, em 2013, dos *Autoencoders* Variacionais (VAEs). Estes foram os primeiros modelos de *deep learning* a serem amplamente utilizados para gerar dados que se assemelham a conteúdos criados por humanos, como imagens e voz. Akash Srivastava, especialista do Laboratório de IA MIT-IBM Watson, sublinha a sua importância, afirmando que "os VAEs abriram as comportas da represa da modelagem generativa profunda, facilitando o uso de modelos em escala" (MIT News, 2023). Segundo ele, "grande parte do que consideramos como IA generativa hoje começou aqui" (MIT News, 2023). Este desenvolvimento foi o ponto de viragem para viabilizar a criação de modelos de larga escala que hoje transformam todas as interações com a tecnologia. A crescente adoção da Inteligência Artificial no quotidiano contemporâneo motiva uma análise rigorosa dos fatores que a impulsionam e das suas ramificações. Uma das características mais notáveis deste processo é a aparente passividade com que a sociedade tem incorporado a IA, muitas vezes sem uma compreensão aprofundada das suas implicações (Kissinger, Schmidt, & Huttenlocher, 2021). Esta aceitação tácita levanta questões críticas sobre as suas consequências a longo prazo. As transformações diárias, mediadas por interações digitais, afetam diretamente as relações interpessoais. A ubiquidade de alternativas tecnológicas eficientes tem diminuído a frequência e a profundidade dos contactos humanos, substituindo-os por interações virtuais (Tegmark, 2017). Esta mudança de paradigma suscita preocupações éticas sobre o poder conferido pela extração de dados e o seu potencial impacto na diplomacia e na política externa dos Estados, desafiando a própria legitimidade da informação num cenário geopolítico complexo.

1.6. O Vazio da IA

Embora a história registre revoluções que abalaram a soberania dos Estados, como a Primeira Revolução Industrial, a presente revolução digital, impulsionada pela IA, distingue-se pela sua natureza sem precedentes. As mudanças provocadas pelas revoluções industriais ainda hoje se repercutem nas estruturas económicas e políticas (Tegmark, 2017). Contudo, a transição para uma era dominada pela IA apresenta um cenário de disrupção de uma magnitude e complexidade que as instituições públicas e privadas ainda não estão preparadas para gerir (Kissinger, Schmidt, & Huttenlocher, 2021). A dependência passiva da tecnologia, aliada à ausência de um quadro regulatório adequado, constitui um risco substancial. Apesar das suas capacidades de análise e previsão, a IA carece de um atributo que, até ao momento, permanece exclusivo do ser humano: a consciência. Esta limitação fundamental distingue a IA de uma inteligência biológica e impõe um limite ético e filosófico à sua autonomia. Embora seja inevitável que a IA continue a transformar a sociedade e o ambiente, a sua evolução levanta uma necessidade urgente de que as instituições se adaptem, desenvolvendo políticas e quadros de governação que abordem os desafios éticos e sociais emergentes (Tegmark, 2017). A falta de preparação para esta nova era tecnológica sublinha a importância de um debate informado e proativo sobre a forma como a humanidade deve gerir a sua coexistência com a IA, garantindo que os benefícios do progresso tecnológico não se sobreponham aos valores humanos fundamentais.

Capítulo II

Tecnologia e Desigualdade Social: Um Retrato Histórico

2. Tecnologia Milenar

A comunidade científica tem-se dedicado, ao longo dos anos, ao estudo da inteligência do ser humano. Quer seja através de padrões comportamentais mais básicos, até a comportamentos mais complexos. Falamos de tecnologia milenar, quando associada ao *Homo Sapiens*, o primeiro Homem capaz de se estabelecer, abandonando a vida de nómada como fizeram, por exemplo, o *Homo Erectus* ou o *Homo Neandertal*.

Com a capacidade de produzir ferramentas essenciais para a sua sobrevivência, conseguindo criar uma cultura de continuidade do conhecimento, isto é, passar de geração em geração as atividades praticadas, ao contrário de se isolar, o *Homo Sapiens*, foi o primeiro a criar aquilo a que hoje denominamos de *cultura* que, assim se torna na “grande inovação “tecnológica” que o cérebro permitiu criar, ao conseguir manipular símbolos que permitem usar linguagem para comunicar entre indivíduos e através das gerações”. (OLIVEIRA, p. 27). Esta é, também, a revolução que Yuval Noah Harari, o autor do livro *Sapiens – História Breve da Humanidade*, se refere, permitindo-nos concluir que a revolução cultural, através do *Homo Sapiens*, permite-nos que hoje possamos integrar-nos pelo mundo, dominar outras espécies e, até, o nosso meio ambiente.

Jared Diamond, autor de *Armas, Germes e Aço – Os Destinos da Sociedade Humana*, que lhe valeu o Prémio *Pulitzer*, tem uma visão diferente de Harari no que diz respeito à emergência da cultura. Diamond refere que a “mente é importante” em todo o processo da criação da cultura, mas considera que o que está na base “é a matéria”. Na perspetiva de Harari, a cultura está para o Homem como um processo cognitivo, por vezes inata, aperfeiçoada ao longo das gerações. Diamond dá-nos uma perspetiva diferente, referindo que é o meio

ambiente onde o Homem está inserido, as suas condições geográficas e ambientais, os fatores primários para determinar uma civilização.

Podemos refletir sobre as duas argumentações que não são assim tão contraditórias. Por um lado, temos algo inerente ao Homem, defendido por Harari, e por outro, o meio ambiente em que se insere. Podemos até referir que as duas perspetivas se completam na medida que uma dá a perceção do espaço e outra dá dimensão às capacidades inatas do Homem que são essenciais para definir aquilo que temos estado a analisar - o início da civilização. Os milénios que se seguiram à constituição e desenvolvimento civilizacional foram palco de muitas inovações tecnológicas, mas o que consideramos mais importante e para o enquadramento deste trabalho, foi a invenção da escrita. Já abordamos a inteligência como uma capacidade de manipular símbolos, agora avançamos para a importância que a linguagem escrita adquiriu.

2.1. A Metamorfose das Relações nas Revoluções Tecnológicas

A evolução da sociedade e a invenção da escrita estão acopladas quando se tornou necessário começar a registar transações comerciais aos membros de sociedades mais sofisticadas. Com um início bastante complexo, a nova linguagem escrita “pensa-se que terá evoluído de forma independente apenas um pequeno número de vezes (...)” (OLIVEIRA, p. 29 e 30). Esta inovação tornou-se num precioso argumento a favor da humanidade que continuou a desenvolver novas técnicas e novos ofícios. Constatase, por isso, que ao longo da História, os seres humanos têm-se adaptado e reinventado em diferentes ambientes e contextos, como vamos explorar já a seguir e que nos conduziram a uma sociedade pré-industrial. No princípio do século XVIII, deu-se a primeira grande transformação e o início de uma guerra de poder entre as máquinas e o Homem. Este marco da História refere-se à primeira Revolução Industrial, depois de Thomas Newcomen e James Watt desenvolverem a primeira máquina a vapor, um marco inicial, por exemplo, para a construção do primeiro comboio. Já se verificavam algumas mudanças no setor da mão-de-obra, onde “diversos engenhos vieram substituir muito do trabalho artesanal”. (OLIVEIRA, p. 31). A segunda revolução industrial sucedeu à primeira, caracterizando-se por

profundas alterações sociais que tiveram início com o impacto causado pelas inovações do motor de combustão interna e pela utilização da eletricidade que incluía, também, as comunicações. Estas [revoluções] correspondem às alterações sociais mais acentuadas na história da humanidade. (Oliveira, 2019). Foi com o casamento entre as revoluções industriais e devido à sua fixação na sociedade que se deu então espaço para o conceito de tecnologia como o conhecemos hoje. A terceira revolução industrial, iniciada na segunda metade do século XX, veio cimentar a importância da tecnologia e do seu desenvolvimento, criando máquinas que possibilitassem criar, manipular e transmitir informação – o início de mais uma ferramenta para a habilidade do Homem.

2.3. Dinâmicas do Populismo Moderno à Luz da Crise Tecnológica

Torna-se inegável constatar que as novas tecnologias causaram profundos impactos nas sociedades, despoletando vários problemas para a existência de uma sociedade e soberania unificadas, enfrentando desafios constantes políticos e sociais, como exploraremos no ponto seguinte. O período das Revoluções Industriais fica marcado na história pelo grande desenvolvimento tecnológico, iniciado em Inglaterra, na segunda metade do século XVIII. Em pouco tempo este movimento começou-se a alastrar para outras partes do mundo com maior enfoque na Europa Ocidental e nos Estados Unidos da América. A partir destas transformações nasce o Capitalismo, indissociável desta grande transformação a nível social e industrial. O crescimento desmedido das cidades, impulsionou o êxodo rural, contribuindo para o aumento da marginalização das pessoas, do desemprego, da fome e da violência. Estas crises são vistas como pontos indutores da transição para o capitalismo que nos fazem refletir sobre a sua dinâmica tecnológica, como por exemplo, o crescimento desigual do sistema económico do Estado. Estes momentos influem de tal maneira que conseguem alterar a lógica desta organização com o intuito de a superar. Do desequilíbrio social, dos problemas de organização dos estados, da exploração da mão-de-obra, do enaltecimento das elites, construímos uma base sólida para explorar o que move as sociedades

modernas, que já viveram numa democracia sólida, a transformarem-se através de forças populistas e de que modo a contínua evolução da tecnologia está a impactar diretamente nestes fenómenos que já se verificam um pouco por toda a Europa e que nos apresenta um quadro de escalada sem precedentes a cada ato eleitoral.

Stanley refere que:

“A política fascista procura destruir as relações de respeito mútuo entre cidadãos, que são a base de uma democracia liberal saudável, substituindo-as, em última análise, pela confiança numa única figura, o líder. Quando a política fascista alcança o seu maior sucesso, este líder é considerado pelos seus seguidores como sendo excepcionalmente digno de confiança” (Stanley. *Como Funciona o Fascismo*. 2019, p. 75).

Numa outra perspetiva, relatada por Villoro (2025), ele cita a obra *O inimigo conhece o sistema*, de Marta Peirano:

“Cada época tem o seu próprio fascismo (...) Ninguém nos obriga a ter o ecrã ligado. Nós mesmo fazemos questão de o levarmos para todo o lado, de o carregarmos a toda a hora, de o renovarmos de dois em dois anos e de o mantermos ligado durante todo o tempo, programado para não perdemos um segundo de propaganda [...]. O seu poder não se baseia na violência, mas em algo muito mais insidioso: a nossa capacidade infinita para a distração. A nossa fome infinita de satisfação imediata. Em resumo, o nosso profeta não é George Orwell, mas Aldous Huxley. Não é 1984, mas Admirável Mundo Novo” (Villoro, *Não sou um Robô*. 2025, p. 94 e 95).

Stanley reforça a sua visão recorrendo a Hannah Arendt e à sua obra *As Origens do Totalitarismo* corroborando com a sua análise, sendo o principal foco da obra a desintegração da sociedade que, para Arendt, [o totalitarismo] “não é apenas um regime de autoridade e força”, trata-se de algo fulcral como um “sistema que desmantelava a própria sociedade” (Arendt, 2024). Estas críticas, tanto de Stanley como de Arendt, são imperiosas para este estudo sendo que no fundo, o que estamos a demonstrar é exatamente o poder que reside e residirá no povo com uma nova máquina em seu poder – A Inteligência Artificial. Isto coloca-nos num limbo futurístico com perspetivas que podem ser aterradoras e questionando se o pior cenário será o das máquinas “inteligentes” ou o ser humano a comandar uma máquina com as capacidades que a Inteligência Artificial está a adquirir dia após dia?

2.4. O Povo na fragmentação entre Ator Político e Recurso Social

Atualmente, a Inteligência Artificial tornou-se um dos temas mais discutidos no espaço público, sendo objeto de ampla cobertura tanto nos meios de comunicação tradicionais como nas plataformas digitais. O seu crescente destaque tem impulsionado intensos debates sobre os desafios éticos e de segurança associados a esta tecnologia emergente, cuja rápida evolução ocorre num cenário marcado pela ausência de regulamentação internacional, como já foi referido. Este contexto começou a levantar preocupações mais concretas relativamente às possíveis repercussões da inteligência artificial sobre os Estados, as suas estruturas políticas, os regimes de governação e as populações que os constituem. Esta agitação prende-se com o facto de o desenvolvimento e o crescente impacto da IA depender, em larga escala, da disponibilidade e da utilização de dados. Dados, esses, que são frequentemente fornecidos pelos próprios indivíduos, muitas vezes de forma inconsciente, tornando-os, inadvertidamente, parte integrante do mecanismo que sustenta e impulsiona cada vez mais esta tecnologia.

No primeiro capítulo da obra *Infocracia* (2022), intitulado “O Regime de Informação”, a análise de Byung-Chul Han destaca a centralidade da informação e do poder como elementos estruturantes das dinâmicas sociopolíticas contemporâneas. Para o autor, o poder já não reside na posse dos meios de produção, mas no “acesso à informação, utilizada como instrumento de vigilância, controlo e prognóstico do comportamento a nível psicopolítico” (Han, 2022, p. 9). Esta perspetiva fomenta uma reflexão crítica sobre a correlação entre regimes políticos e os modos de gestão informacional, com implicações diretas nas formas de perceção e atuação das sociedades. Neste quadro, é pertinente a referência à análise de Michel Foucault, que descreve o indivíduo moderno como parte integrante de uma engrenagem disciplinar, “uma massa informe, um corpo inapto” dentro de uma “máquina” (Ibid, apud. Han), sublinhando a dimensão estrutural do controlo nos sistemas de poder.

Para a construção de uma pesquisa crítica das estruturas de poder e dos mecanismos de controlo — muitas vezes exercidos de forma opressiva —

recuamos historicamente à época da Primeira Revolução Industrial e ao despontar do capitalismo. Este foi um período fortemente influenciado pelos ideais do Liberalismo, uma doutrina política e social que se consolidou com base nos princípios da liberdade individual, da igualdade jurídica e da limitação da intervenção do Estado. Entre os principais teóricos desta corrente destaca-se John Locke, frequentemente reconhecido como o fundador do Liberalismo clássico, a propriedade privada constituía um “direito natural” e, nesse sentido, o papel do Estado deveria restringir-se à garantia desses direitos fundamentais, sem intervir nas dinâmicas económicas ou sociais — característica essencial do chamado Estado não intervencionista. Os defensores desta visão sustentavam que a desigualdade era uma consequência inevitável da liberdade individual, sendo responsabilidade do Estado apenas assegurar condições igualitárias de partida, sem se comprometer com os resultados advindos das escolhas dos cidadãos. O Liberalismo, enquanto filosofia política conservadora, concebe o Estado como resultado de um pacto racional entre indivíduos livres, fundado no consentimento dos governados. Assim, a autoridade política é legitimada pela proteção dos direitos naturais — vida, liberdade e propriedade — cuja violação justificaria, segundo Locke, o direito à rebelião por parte dos cidadãos.

Por oposição a esta perspetiva liberal, há que referir a ideia marxista que propõe uma crítica estrutural ao sistema capitalista, entendendo que os princípios liberais, em vez de promoverem igualdade, sustentam as bases de produção da desigualdade e da exploração. Nos vários manuscritos de Marx, refere a superação do capitalismo exige a constituição de um Estado fundado nos princípios do socialismo, cuja função é a de uniformizar e padronizar economicamente a sociedade. Ao contrário da proposta de Locke, este modelo exige um Estado intervencionista, com capacidade de regular e controlar os meios de produção e as relações sociais. No entanto, esse controlo do Estado visa garantir o acesso absoluto à cidadania, bem como promover a igualdade económica e social entre os indivíduos. A crítica marxista à estrutura do capitalismo encontra um prolongamento e edificação na obra de Michel Foucault, que transfere o foco da análise da esfera económica para as formas de poder e controlo presentes no quotidiano das sociedades modernas. Em *Vigiar e Punir* (2018), Foucault descreve o modo como o poder disciplinar se dissemina pelas instituições sociais — como a escola, a prisão ou o exército — moldando corpos

e subjetividades de forma minuciosa e contínua o que evidencia a lógica de normalização e vigilância que caracteriza os regimes modernos de poder. (Foucault, 2018)

Este poder assente na disciplina não se manifesta apenas pela pressão física, mas pela interiorização de normas e pela “construção” de sujeitos obedientes, ajustados às exigências de produtividade e conformidade. Assim, Foucault revela que o controlo não se exerce apenas de cima para baixo, através do Estado ou da força, mas também pelas práticas sociais e conhecimentos que organizam e regulam o comportamento humano pela eficiência e ordem. É neste ponto que a análise de Byung-Chul Han se revela particularmente relevante para a compreensão do contexto atual. Em *Infocracia*, Han (2022), identifica um novo regime de poder que se distancia dos mecanismos disciplinares descritos por Foucault e justifica a lógica da informação. A sua tese central afirma que o poder contemporâneo já não assenta na repressão ou na visibilidade do controlo, mas no “excesso de informação e na transparência forçada, que provocam uma nova forma de submissão” — mais subtil, mas igualmente eficaz, deixando os indivíduos alienados do real.

Para Han, a sociedade digital promove uma lógica de autoexposição e autoexploração, onde os indivíduos se tornam simultaneamente emissores e recetores de informação, colaborando ativamente com os mecanismos que os controlam. O acesso à informação passa a ser o principal vetor de poder, usado para vigiar, prever e moldar o comportamento, num plano psicopolítico. Trata-se de uma transformação profunda na forma como o controlo é exercido, onde o sujeito já não é coagido por uma autoridade externa, mas envolvido num sistema de monitorização permanente e que voluntariamente alimenta-o através das suas interações digitais. Neste cenário, tanto a crítica marxista à exploração económica como a análise foucaultiana da disciplina social representam o interior de um regime que não reprime diretamente, mas que seduz, atrai e capta a atenção como forma de dominação. A *infocracia*, assim, representa um novo paradigma de poder, caracterizado por uma vigilância invisível e por uma liberdade aparente que dissimula estruturas profundas de controlo.

Capítulo III

Distopia Panótica - Alienação e Poder

3. O Fim da Esfera Pública?

A crítica à racionalidade instrumental, de Habermas, autor que será fundamental para a análise da esfera pública neste trabalho, já fazia uma premonição dos possíveis perigos que existiam numa primazia dada à eficiência em detrimento da democracia, o que se mostra extremamente relevante para a análise deste tema. Stanley também afirma que “Hoje o fascismo pode não ter exatamente o mesmo aspeto que tinha na década de 1930, mas há refugiados de novo em fuga por todo o mundo (...) a sua provação reforça a propaganda fascista de que a nação está sitiada e de que os estrangeiros são uma ameaça” (Stanley, 2019, p. 14) e a contínua perpetuação de ideias como a referida, levam-nos a outro assunto, também abordado por Habermas através da sua teoria da ação comunicativa – o discurso de ódio. Ambos os autores consideram que este é uma ameaça à esfera pública democrática com repercussões até agora desconhecidas.

Um dos maiores exemplos da História é Adolf Hitler que demonstrou que o objetivo da sua propaganda era o de substituir a argumentação racional, manipulando, eximamente, os medos e as inseguranças do povo, com o intuito de consolidar o seu poder. A sua propaganda passava pela disseminação de mensagens que ameaçavam a Alemanha, tanto a nível interno, como externo, criando, assim, um inimigo em comum. Esta incitação de medo, desviava a atenção pública dos reais problemas internos, a favor de uma narrativa para justificar atos hediondos quando onde “as teorias da conspiração são um mecanismo essencial utilizado para retirar credibilidade aos meios de comunicação dominantes, que os políticos fascistas acusam de ser tendenciosos por não cobrirem falsas conspirações” (Stanley, p. 64). Para Hitler, esta era uma clara assunção do objetivo de propaganda “substituir a argumentação racional na esfera pública por medos e paixões irracionais” (Ibid). Este método galvanizou o seu estatuto e o apoio popular, com um dos princípios básicos da política fascista onde “(...) oratória não deve convencer o intelecto, mas sim influenciar

a vontade”. (Ibid). Estávamos em 1933 quando Hitler ascendeu ao poder. Harari, completa as palavras de Stanley quando refere que esta ascensão ao poder se deu “(...) durante a crise económica, milhões de alemães passaram a acreditar na narrativa dos nazis, em detrimento de qualquer uma das outras que lhes eram apresentadas”. (Harrari, 2024). Esta tomada de poder provocou grandes alterações geopolíticas, começando pela intensificação ideológica na Europa e no mundo. Uma política agressiva que atirou a primeira pedra dando início à Segunda Guerra Mundial, como nos conta a história, com a invasão da Polónia em setembro de 1943, obrigando a França e o Reino Unido a renunciarem à política de apaziguamento e forçando-os a declarar guerra. Mas, esta importante contextualização, continua com a proclamação de discursos de ódio e teorias da conspiração. Stanley referiu que “o sintoma mais revelador da política fascista é a divisão” e é neste sentido que ao deixarmos de ver o “nós” como um todo contribuímos para a segmentação da população como, por exemplo, a “excluir grupos” (Stanley, p. 10 e 11). Harari diz-nos que “o papel fundador das narrativas nos traz uma revelação fulcral sobre o poder da nossa espécie, além de explicar porque não anda sempre de mãos dadas com a sabedoria” (Ibid). Esta, consideravelmente, sarcástica, mas astuta afirmação tornou-se de extrema importância para começar a abordar o campo das interações humanas, saber o que as move na direção de certas narrativas e qual a importância dos atores políticos nesta que já é considerada por muitos autores e cientistas a Quarta Revolução Industrial ou, se quisermos ajustar a escrita à modernidade – Indústria 4.0.

3.1. Do Vínculo ao Vazio

No âmbito da sua análise em *Vigiar e Punir*, Michel Foucault resgata e aprofunda o conceito *Panótico*, concebido em 1793, por Jeremy Bentham. Foucault estabelece a estrutura arquitetónica do Panótico como um modelo paradigmático que transcendeu a sua aplicação original em prisões, servindo como uma matriz arquitetónica para a organização disciplinar de diversas instituições sociais chegando até a utilizar a expressão “arquipélago carceral”. Esta formulação foucaultiana sublinha como a visibilidade e a vigilância inerentes

a esta arquitetura se tornaram o alicerce para o controlo e a padronização dos indivíduos na sociedade moderna. (Foucault, 1975). Para Foucault, a ideia de controlo e poder panótico é resumida como sendo

“(...) na periferia uma construção em anel; no centro, uma torre; esta é vazada de largas janelas que se abrem sobre a face interna do anel; a construção periférica é dividida em celas, cada uma atravessando toda a espessura da construção; elas têm duas janelas, uma para o interior, correspondendo às janelas da torre; outra, que dá para o exterior, permite que a luz atravesse a cela de lado a lado.”

Foucault, (2013). Partindo da análise de um dispositivo arquitetónico que serve como estratégia de controlo social, é fundamental examinar o seu funcionamento e a sua capacidade de governar as massas. A estrutura em questão não é uma metáfora, mas sim um modelo concreto de poder que opera através da vigilância e da disciplina. Mas, antes de nos debruçarmos sobre as implicações de uma sociedade alienada por esta constante vigilância, urge compreender como esta "construção" materializa e exercita o poder, impondo uma ordem social e uniformizando comportamentos.

Para Foucault,

“Basta então colocar um vigia na torre central, e em cada cela trancar um louco, um doente, um condenado, um operário ou um escolar. Pelo efeito da contraluz, pode-se perceber da torre (...). Tantas jaulas, tantos pequenos teatros, em que cada ator está sozinho, perfeitamente individualizado e constantemente visível” (Foucault, 2013).

Na perspetiva foucaultiana, a desumanização e a alienação do indivíduo, que se manifestam na negação de sua natureza social, são o resultado direto de mecanismos de poder. Foucault argumenta que esse poder, funcionando de forma “onipresente e onisciente”, tem como objetivo principal fragmentar e classificar o indivíduo, de acordo com critérios estabelecidos. A visibilidade, neste contexto, é descrita como uma armadilha: o indivíduo, por estar constantemente exposto e sob vigilância, internaliza a necessidade de se autorregular, tornando-se, assim, o agente de sua própria subjugação. Esta vigilância permanente e o consequente processo de individualização culminam

na alienação, onde o sujeito perde sua essência social para se tornar uma peça isolada, controlada por um sistema invisível e difuso.

Milan Kundera, escritor reconhecido pela sua obra *A Insustentável Leveza do Ser*, profetizava, já na segunda metade do século XX, a degradação da intimidade. No seu livro *Testamentos Traídos*, relata a vida de um comum vizinho, que volta do trabalho, acende as luzes e começa a sua rotina perfeitamente normal, mas, pouco depois, quando a cortina se fecha e corta a visão “Não tinha nada a ocultar senão ele próprio, a maneira como percorria a divisão, a forma descuidada de se vestir, o modo de passar a mão pelo cabelo. O seu bem-estar é condicionado pela sua liberdade de não ser visto”. (Villoro, 2024, p. 22). Da mesma forma, Byung-Chul Han, descreve na sua obra *Sociedade da Transparência*, um mundo “opaco” caracterizado por segredos ocultos, para um “transparente” onde tudo é exposto e acessível cultivando o voyeurismo e reprimindo as ações espontâneas dos indivíduos. (Han, 2025). O trabalho deste filósofo e teórico contemporâneo tem explorado como é que as mudanças sociais que, à partida, parecem benignas, podem levar a novas formas de controlo. Crítico acérrimo da sociedade digital e da transparência, Byung-Chul Han é um autor fundamental para esta análise que pretendemos fazer de uma sociedade que está a deixar-se alienar aos poucos provocando uma abordagem disruptiva no novo quadro político global. Tendo Foucault como referência, Han, recupera o regime disciplinar foucaultiano que “utiliza o isolamento como meio de domínio” afirmando que “a solidão humana é a condição primeira da submissão total”. (2022, p.10). Este regime disciplinar descrito por Foucault é, para ele, “a forma de domínio do capitalismo industrial”. O indivíduo funciona como se fosse uma pequena engrenagem da maquinaria do poder disciplinar e isso “converte uma massa informe, um corpo inapto” numa “máquina”. (ibid). O panótico, referido no início deste capítulo, com celas isoladas umas das outras, é “(...) o símbolo e o ideal do regime disciplinar. Mas o isolamento não se pode transpor para o regime de informação, que explora precisamente a comunicação”, desta forma a “vigilância no regime de informação ocorre através de dados. Os ocupantes do panótico disciplinar, isolados de si mesmos, não produzem dados, não deixam pistas de dados, porque não comunicam.” (Foucault, 2013).

O regime disciplinar deu lugar ao que Han define como *regime de informação* “associado ao capitalismo de informação, que se desenvolve em

capitalismo de vigilância e que degrada os indivíduos, tornando-os dados e gado consumidor” (Han, p. 9). A presente configuração de representação da sociedade contemporânea é caracterizada pela completa alienação do indivíduo. Este, por sua vez, traz em si, os elementos que o levam a perpetuar um ciclo que se tornou vicioso e que suscita preocupação entre aqueles que questionam o futuro da sociedade moderna, democrática e liberal. Atualmente, “a esfera pública é um palco. O poder da soberania atua com base na visibilidade teatral. É um poder que se dá a ver, que se anuncia, se vangloria e resplandece. Porém, os submetidos, nos quais aquele se exerce, permanecem em grande medida invisíveis.” (Han, p. 11).

3.2. A Crítica de Frankfurt e a (Re)configuração da Esfera Pública de Habermas

A sociedade do século XXI, que se presumia informada por fontes credíveis e capaz de fomentar o debate público sobre questões sérias, aparenta contrariar as expectativas otimistas de Jürgen Habermas. A sua descrença no projeto iluminista, defendido contra os pessimistas da Escola de Frankfurt, como Theodor W. Adorno e Max Horkheimer, parece agora, mais do que nunca, justificada. De facto, a atual conjuntura social parece confirmar as perspetivas mais pessimistas dos primeiros teóricos da Escola de Frankfurt relativamente ao estado da comunidade. É possível aprofundar a análise da sociedade contemporânea à luz das ideias de Habermas, Adorno e Horkheimer. O otimismo de Habermas na esfera pública como espaço de racionalidade e deliberação parece ser desafiado pela realidade atual, onde a informação, por vezes, carece de fontes credíveis. Esta situação sugere que a "racionalidade comunicativa", conceito central na sua teoria, pode estar comprometida. Este contraste, de acordo com diagnóstico da Escola de Frankfurt, ressoa com a crescente preocupação em relação à manipulação e à padronização do pensamento. Adorno e Horkheimer, na obra *Dialética do Iluminismo*, argumentaram que a "indústria cultural" transforma a arte e a cultura em mercadorias, sufocando o pensamento crítico e a individualidade. Esta homogeneização do pensamento, prevista pelos autores, parece hoje ter-se intensificado através dos novos meios

de comunicação, onde a pluralidade de ideias é frequentemente subjugada por algoritmos e narrativas dominantes. Vivemos numa época de performance “A indústria cultural desenvolveu-se com o predomínio que o efeito, a *performance* tangível e o detalhe técnico alcançaram sobre a obra, que era outrora o veículo da Ideia e com essa foi liquidada.” (1947, p. 59). Todas as nossas ações desenvolvem-se como uma performance que não tem fim. Ela vive dentro dos nossos relacionamentos, nas nossas escolhas, na maneira como agimos para o “outro”. Nunca vivemos uma era tão “conectada” como a de agora e curiosamente tão desconectada do real. Assim, se Habermas acreditava que a racionalidade comunicativa poderia redimir a sociedade da “racionalidade instrumental”, Adorno e Horkheimer, por outro lado, viam no próprio Iluminismo um potencial para a servidão. A realidade do século XXI, com a sua aparente falta de debate informado e a proliferação de informação não credível, levanta a questão de saber se a “utopia” de Habermas ainda é alcançável ou se, como defendiam os teóricos da Escola de Frankfurt, o “esclarecimento de massas” não conduziu, afinal, à sua própria negação.

É crucial examinar a atual crise da esfera pública à luz das ideias de Adorno, Horkheimer e Habermas. A visão pessimista da Escola de Frankfurt, que antevia a subjugação do indivíduo pela indústria cultural, parece encontrar eco na sociedade contemporânea. A proliferação de “fórmulas repetitivas e previsíveis” na cultura de massas, como já foi referido, reflete a prioridade do lucro sobre a inovação e o pensamento crítico. O debate sobre a “democracia deliberativa” de Habermas ganha nova urgência neste contexto. Habermas, ao defender a esfera pública como um espaço onde cidadãos informados debatem racionalmente, assume a existência de uma imprensa e de meios de comunicação que operam segundo princípios de verdade e credibilidade. Contudo, a lógica da indústria cultural e a emergência das chamadas “notícias falsas” (*fake news*) corroem as bases desse ideal, impondo-se a pergunta: será que a racionalidade comunicativa pode, de facto, prosperar num ambiente onde a informação é uma mercadoria e a verdade é, frequentemente, secundária ao espetáculo e a conflitos de interesse? A tensão entre a crença no potencial do projeto do Iluminismo habermasiano e o pessimismo de Adorno e Horkheimer permanece, mas a realidade do século XXI parece, cada vez mais, sombria e é neste contexto que Habermas confronta-se, hoje, com a sua mais recente

reflexão no seu livro *Ein neuer Strukturwandel der Öffentlichkeit, 2023* (Uma Nova Mudança Estrutural da Esfera Pública), onde retoma a sua análise clássica de 1962, adaptando-a à era digital. O autor reconhece que as redes sociais e as plataformas digitais, apesar de terem o potencial de democratizar a informação, acabaram por gerar uma nova crise. Habermas defende que a nova esfera pública não cumpre a sua função de fomentar um debate sério e informado. Em vez de se tornarem veículos para a ação comunicativa, as plataformas digitais transformam-se em espaços de "batalha de atenção". A lógica do algoritmo, que recompensa o conteúdo sensacionalista e as opiniões extremas, subverte o processo deliberativo. Com a fragmentação do ecossistema mediático, a tradicional função de filtro exercida pela imprensa, que mediava e validava a informação, tornou-se obsoleta. Este declínio confere um poder desproporcional a jornalistas e influenciadores, que agora operam com pouca supervisão editorial. A ausência de mediadores de informação credíveis resulta numa acentuada erosão da confiança pública e na proliferação de narrativas não verificadas. Este diagnóstico de Habermas aproxima-se, ironicamente, da crítica dos teóricos da Escola de Frankfurt. A indústria digital, ao favorecer o efeito e o *engagement* em detrimento da ideia, confirma as mais sombrias previsões de Adorno e Horkheimer. O que a indústria cultural fez com a arte e a cultura, a indústria da atenção faz com a informação e o debate político. A nova esfera pública, em vez de ser um espaço de esclarecimento, torna-se um reflexo do mercado, onde a verdade é uma mera mercadoria sujeita à oferta e à procura de cliques. A crença de Habermas no projeto Iluminista, ao contrário do que defendiam os primeiros pensadores, como Adorno e Horkheimer, da Escola de Frankfurt, reaparece, com o intuito de reconhecer que o ideal de uma democracia participativa e consciente enfrenta agora outros desafios e grandes perigos. Habermas, apresenta-nos uma regressão dos pilares fundamentais para uma sociedade bem informada, esclarecida e participativa no debate público, agora muito mais representativa do que na época da sua obra *Uma Nova Mudança Estrutural da Esfera Pública (2023)*. Neste cenário, a questão que se coloca é: como se pode preservar a esfera pública e a racionalidade comunicativa num ambiente digital, cada vez mais "desconectado" da verdade, com o desenvolvimento alucinante da Inteligência Artificial que incentiva a polarização e a disseminação de desinformação?

A crise da esfera pública revelou-se num terreno fértil para o despontar de uma crise política generalizada. O eco que se faz ouvir do crescente apoio a um populismo marcado pela dissidência de “bolhas de filtro” tem ultrapassado os prognósticos mais trágicos (se os havia) de uma quebra total nos ideais de um regime democrático. Eli Pariser, em *Filter Bubble, what the Internet is Hiding From you*, refere que:

“a nova geração de filtros da internet fixa-se no que parece agradar-nos (...) e tira as respetivas conclusões. (...) juntas estas máquinas criam um universo de informação único para cada um de nós - a que chamo um *filter bubble* -, que altera de um modo fundamental a maneira como acedemos às ideias e à informação.” (2011, p. 156)

Estas ideias de Pariser remetem-nos para o início daquilo a que assistimos - o momento-chave que desencadeia uma série de cenários imprevisíveis, com impacto global na política, na sociedade e na economia.

Han, vem argumentar esta teoria, considerando os seus pontos fracos que “residem no facto de [ela] atribuir o estreitamento do horizonte da experiência na sociedade de informação apenas à personalização algorítmica da internet.” (2025, p. 35). Para Byung-Chul Han, esta é apenas uma das “hipóteses” levantadas pela teoria de Pariser, uma vez que reduz o problema da “desintegração da esfera pública” tão-somente a um “problema técnico”, quando a “autodeterminação ou a autopropaganda já tem lugar offline.” (Ibid), reafirmando a sua posição de que a exposição e autopromoção do indivíduo já não se resume apenas ao digital, também já faz parte das suas interações sociais. O que está a provocar a crise da democracia como a conhecemos é a crescente evasão do outro e a nossa inabilidade para ouvir. O autor reconhece que “hoje todos se entregam ao culto do eu” que é algo que com “a crescente atomização e narcisificação da sociedade torna-nos surdos à *voz do outro*, conduzindo igualmente à perda da *empatia*.” Hiroki Azuma, vem reiterar todos os aspetos que foram referidos ao falar do declínio da “comunicação interpessoal quotidiana” quando refere que:

“Na sociedade contemporânea, os cidadãos já não conseguem acreditar num fundo comum de discussão que permita iniciar um debate. Já nem sequer conseguem supor que todos estão a participar neste debate enquanto membros

da mesma comunidade. A esfera pública que Arendt e Habermas enalteceram como um ideal nem sequer entrou em ação” (General Will 2.0)

Perante este cenário de desmoronamento das sociedades e da ação comunicativa, Habermas reflete concretamente a sua perplexão:

“Simplesmente, não sei que aspeto assumiria no mundo digital um equivalente funcional da estrutura comunicativa das vastas esferas públicas políticas, emergentes desde o século XVII, mas que hoje estão em processo de desintegração. (...) como é possível manter no mundo virtual de uma internet descentralizada (...) uma esfera pública baseada em circuitos de comunicação que incluam a população geral?” (2020, p. 27).

Esta descrença tornou-se numa preocupação com o crescente escalar das novas tecnologias impulsionadas pela Inteligência Artificial, cada vez mais eficiente e desafiadora, que trabalha a um ritmo galopante mas sem fazer ruído, fazendo com que tudo o que está à sua volta viva alienado da verdade, de factos, de fontes credíveis, de consumo acrescentado para consumismo superficial, da preocupação com os valores e virtudes, para a cultura de um *ethos* que espelha para os outros aquilo que querem ver.

A relevância deste capítulo baseia-se sobretudo em analisar o agente mais importante num futuro que poderá ser dominado por máquinas - o ser humano, que desde o início deste trabalho, tem um papel fundamental para compreender as suas origens e transformações e o seu modo de agir. A reflexão em curso tem analisado diferentes regimes e como o “culto do eu”, tem desempenhado um papel importante na alienação dos indivíduos. A racionalidade comunicativa deu agora lugar à “racionalidade digital” que, ao contrário da primeira, não tem como ponto de partida o indivíduo, mas o seu coletivo. (Han, p. 47). “A forma de racionalidade que se sustenta sem discurso” (p. 65), na medida em que ela é oposta à racionalidade comunicativa. Aqui, “o discurso é substituído por dados.” (Ibid). Os “dataístas”, os defensores do *Big Data*, “acham que não apenas a desintegração da esfera pública, mas também a massa pura de informações e complexidade rapidamente crescente da sociedade e informações tornam obsoleta a ideia de ação comunicativa” (p. 63). Assim, os dataístas “tem em mente uma sociedade que se sustenta completamente sem política” (Ibid).

3.3. O Novo Niilismo: Como a IA Desestabiliza a Realidade e a Ordem Global

Han refere uma nova conjuntura em relação à informação afirmando que, “propaga-se um novo niilismo” que está intrinsecamente ligado à *sociedade da informação*; “na era das fake news, da desinformação e das teorias da conspiração perde-se a realidade com as duas verdades fácticas. Agora, a informação circula totalmente desligada da realidade, num espaço hiper-real. Perdeu-se a crença na facticidade. Por isso, vivemos num universo desatualizado. O mundo comum, ao qual nos poderíamos ligar com as nossas ações, acaba por também desaparecer, juntamente com a verdade dos factos” (Han, p. 51). O novo niilismo aparece como sintoma desta nova sociedade de informação, para Han, surge “no interior de qualquer processo destrutivo no qual o discurso se desintegre em informação, o que conduz à crise da democracia” (ibid). No próximo capítulo, abordaremos duas perspetivas cruciais de cientistas de destaque que moldaram a criação da Inteligência Artificial. A primeira delas é a de Geoffrey Hinton, uma figura inspiradora neste campo. Quando questionado sobre uma possível solução para a escalada destrutiva da IA, Hinton sugere uma abordagem geopolítica, afirmando que “primeiro temos de saber que países ainda são democracias” (Hinton, 2025). Esta declaração é notavelmente provocadora, pois sublinha a complexidade das implicações da IA, deslocando o foco da tecnologia em si para os agentes que a controlam. A sua resposta sugere que a ameaça existencial da IA não reside apenas na sua autonomia ou capacidades técnicas, mas sim na sua utilização por nações ou entidades que operam fora dos princípios democráticos. Esta perspetiva abre uma discussão sobre a relação intrínseca entre o poder e a tecnologia, onde a busca pelo poder pode ser o principal catalisador para a manifestação de intenções maliciosas. Ao examinar as suas palavras, percebemos que a “caixa de Pandora” não contém apenas a tecnologia em si, mas as motivações e estruturas de poder que a orientam. Este ponto de vista convida-nos a considerar que os agentes determinantes da ação da IA não são os algoritmos, mas sim as agendas geopolíticas e os sistemas de governação que os implementam. Não estamos

longe da mítica caverna que Platão escreveu. Hoje, pelo contrário, temos uma caverna digital, que nos mantém presos na informação e a luz da verdade está completamente apagada. Um intenso ruído da informação faz os contornos do ser desvanecerem-se. A verdade não faz ruído. (Han, p. 64).

A crise da verdade, já abordada, não é apenas um fenómeno de alienação individual, mas uma vulnerabilidade estratégica que as potências utilizam para projetar poder e minar a coesão de adversários e rivais. A desinformação, alimentada por algoritmos de IA, torna-se uma arma assimétrica, capaz de desestabilizar processos democráticos, erodir a confiança nas instituições e fomentar divisões internas em sociedades-alvo, sem o recurso a conflitos militares tradicionais. Neste cenário, a corrida pela supremacia em Inteligência Artificial assume contornos semelhantes à corrida armamentista da Guerra Fria. O domínio tecnológico da IA é percebido como um fator determinante para a hegemonia no século XXI, conferindo vantagens decisivas em áreas como a economia, a segurança nacional e a projeção de poder militar. A competição entre os Estados Unidos, a China e a Rússia, por exemplo, não se manifesta apenas na quantidade de investimentos ou no número de patentes, mas na capacidade de integrar a IA em todos os aspetos da sua governação e estratégia. A ascensão de uma “racionalidade digital”, como Han argumenta, não é apenas uma transformação cultural, mas a base de um novo tipo de poder estatal que opera sobre a informação e a cognição. As implicações para a segurança internacional são profundas. O desenvolvimento de sistemas de armas autónomas levanta dilemas éticos e estratégicos sem precedentes, onde a decisão de vida ou morte pode ser delegada a uma máquina. A guerra híbrida e o ciberconflito, por sua vez, são potencializados pela IA, permitindo ataques sofisticados a infraestruturas críticas e campanhas de influência em massa. A mítica “caixa de Pandora” de Hinton ganha, assim, uma dimensão geopolítica: a verdadeira ameaça não é a IA em si, mas a sua apropriação por regimes não democráticos ou por potências revisionistas que a veem como o meio ideal para remodelar a ordem global a seu favor, explorando as fissuras de uma sociedade já fragmentada pela “crise da verdade”. A caverna digital, longe de ser um mero refúgio de alienação, torna-se um campo de batalha invisível, onde a luta pelo controlo da narrativa e da perceção pública substitui a tradicional disputa por território.

Deste modo, a transição da "crise da verdade" para um campo de batalha geopolítico impõe a necessidade de aprofundar a análise sobre a militarização da cognição e a geopolítica da Inteligência Artificial. O próximo capítulo dedicará-se a desvendar as complexas dinâmicas que subjazem a esta nova era de competição estratégica. Através de duas visões críticas como a de Geoffrey Hinton e de Kai-Fu Lee, primeiro lugar, examinaremos como a IA redefine os conceitos tradicionais de segurança, poder e soberania, ao permitir a projeção de influência e o enfraquecimento de adversários. Abordaremos a ascensão da guerra híbrida e do ciberconflito como a nova norma nas relações internacionais, e a forma como a IA serve de catalisador para a sua intensificação, através de ataques sofisticados a infraestruturas críticas e a manipulação de processos eleitorais. Em segundo lugar, analisaremos o dilema da autonomia em sistemas de armas, explorando as implicações éticas, legais e estratégicas de delegar a decisão de uso da força a algoritmos. Por fim, o capítulo focará na corrida por hegemonia tecnológica entre as grandes potências, avaliando como o domínio da IA se tornou o fator crucial para a projeção de poder no século XXI, e as possíveis rotas para a cooperação ou o conflito que se abrem neste novo cenário de competição. Assim, este próximo passo na nossa investigação busca ir além da mera constatação da crise, para perscrutar os mecanismos e as consequências de uma sociedade em que a verdade e a percepção pública se tornaram as novas frentes de batalha, tendo por base a opinião dos pioneiros desta nova tecnologia que, como veremos, não é assim tão nova, mas emergiu como uma ferramenta sem precedentes e a uma escala global.

3.4. A Geopolítica da Ilusão

Considerado no meio científico como o *Godfather of AI* (O Padrinho da IA), Geoffrey Hinton, agora com 75 anos, surpreendeu o mundo quando deixou o alto cargo que tinha na gigante Google, afirmando que tomou esta atitude para poder “falar livremente sobre os riscos da IA e alertar para os seus perigos iminentes” (Hinton, 2025, 00:39). O cidadão comum, alienado das suas ações e consequências, como já analisamos num capítulo anterior, já estava a usar inteligência artificial e, mais, contribuindo para a agregação de dados de

consumo, sem ter conhecimento. Este conhecimento poderia ser substituído por “interesse”; a falta de cuidado com a pegada digital que deixamos é de tal ordem que permite a quem usa a IA para fins maliciosos, ciberataques cada vez mais sofisticados (Hinton, 2025, 12:12). Considerando tudo o que tem sido investigado, Hinton afirma veemente que “a inteligência artificial pode tornar-se mais inteligente do que os próprios seres humanos que a operam e isto é um risco bem real” (Ibid). Nobel da Física, em 2025, Geoffrey Hinton, apresenta não só uma visão pessimista como também basta alarmista de um futuro liderado por máquinas autônomas, mas antes dessa autonomia das máquinas, existem outros perigos possíveis que Hinton prevê e outros que já estão a acontecer. Já referimos que os ciberataques são cada vez mais sofisticados e o número de casos de *phishing* têm aumentado substancialmente em empresas cruciais, como entidades bancárias. Mas, para além destas ameaças, Hinton, alerta que seremos “alvos de ataques para os quais não estaremos preparados e isto ocorrerá até 2030” (ibid); vírus letais criados em laboratórios que poderão dizimar uma grande parte da população mundial e que para isso “basta um cientista com algum conhecimento em biologia molecular e alguns conhecimentos de IA para poder outra crise humanitária”, até hoje não temos dados suficientes que comprovem a pandemia da SARS-COV que começou na China. Fundamental para este trabalho é o último alerta que Hinton discorre numa longa entrevista: “atos eleitorais cada vez mais comprometidos”. Não é novo, mas não foi suficientemente estudado até agora e a pertinência deste trabalho tornou como pedra basilar a ameaça desta nova tecnologia para a política e para a redefinição de poder. A inteligência artificial, de acordo com a previsões do ex-CEO da Google, “poderá ser usada para corromper eleições ao disseminar desinformação, conseguir detetar os eleitores mais suscetíveis a receberem informações manipuladas com base na análise dos seus dados e interações na internet e até *hackear* os sistemas de voto alterando resultados”, esta previsão é sem dúvida a chave de toda análise que tem sido feita ao longo deste trabalho, mostrando-se fundamental perceber como funciona. De acordo com Hinton, empresas como a Meta e o YouTube, criam as chamadas *echo chambers*. Já mencionadas anteriormente, as *filter bubbles* criadas, são posteriormente acomodadas nestas “câmaras de eco”, polarizando opiniões, disseminando informação falsa, uma vez que vivem dentro dessas *echo chambers* que não

deixam passar informação que seja diferente dos ideais dessas bolhas apartando-os da realidade exterior. Esta descrição, bastante coincidente com a Alegoria da Caverna, de Platão, mas dos tempos digitais, reflete a mesma prisão da mente humana que vive numa realidade de ilusão. Se perscrutamos o Livro V, da República, de Platão e como refere, Manuel Fonseca “Por Platão, vale a pena estremecer” (Fonseca, 2023, p. 15), podemos concluir que é a mesma alegoria, que passou de 375 a.C. para o século XXI, e que de Mito da Caverna passa a Realidade Digital. Exploremos, ainda que parcialmente, os elementos da caverna: “os prisioneiros”, representam a humanidade que se encontra na ignorância; “as correntes” os preconceitos, os sentidos que prendem a humanidade à ignorância; “as sombras” uma falsa realidade de aparências; até à saída da caverna em busca do verdadeiro conhecimento e o “Sol” que representa “o mundo inteligível”. Representando estes elementos, hoje, na era digital, podemos refletir sobre a Caverna, como as *echo chambers*, referidas por Hinton; os prisioneiros no fundo da caverna como assistindo a um *feed* algorítmico altamente personalizado; as sombras a representarem as notícias falsas, a desinformação e todo o conteúdo polarizado que confirma as crenças dos utilizadores e as correntes como algoritmo que prendem a mente enviesando o pensamento. Não sabemos o final, mas acreditamos que a saída para o verdadeiro conhecimento também se concretizará.

Esta análise filosófica que constatamos ser representativa daquilo que está a acontecer presentemente e que vem na sequência das grandes empresas, consideradas as gigantes, desta era, de manipular e enviesar o pensamento e até, como refere Hinton, “chegar ao nosso *byos*” - a nossa ideologia, apresentam-se com uma única intenção: indignar o ser humano que, de acordo com a perspetiva de Hinton “as pessoas gostam e precisam de se indignar com algo”. Este exemplo pode refletir qualquer um de nós: “aparece uma imagem de Donald Trump, com um título sensacionalista, imediatamente somos tentados a clicar para abrir a notícia, mas, a partir desta ação, o algoritmo começa a revelar mais coisas que considera que queremos ver, tendo em conta a ação anterior (clicar para abrir - ver)” (Hinton, 2025). Esta pequena ação altera e molda o meu *feed*. O que eu vejo é diferente do que o outro vê, porque a informação está condicionada pelo algoritmo e por aquilo que ele acha que eu gosto e quero ver, saber, fazer. Esta é uma condição que, por exemplo, contraria

o ideal descrito e defendido por Habermas, uma sociedade que lia o mesmo jornal, as mesmas notícias de fontes fidedignas dando lugar e espaço para um debate público. Hoje, “todos os *feeds* são diferentes, porque procuram incessantemente mostrar-nos algo que não sabemos, mas depois de vermos ou lermos, foi criada uma necessidade que nos coloca no papel de consumidores impulsivos de topo o tipo”, esta observação partilhada por Hinton, leva-o a concluir que “já não existe uma realidade partilhada”. Hinton sublinha uma transformação paradigmática na circulação da informação mediada digitalmente. A personalização incessante dos *feeds* constitui uma operação algorítmica fundamental cujo objetivo já não é a mera satisfação de preferências preexistentes, mas sim a criação e a antecipação de necessidades. A consequência sociológica mais significativa reside na deterioração da realidade partilhada. A objetividade ou, pelo menos, a intersubjetividade necessária para a coesão social e o debate democrático, baseada num conjunto comum de referências factuais e conceituais, tornou-se insustentável.

3.5. A Desinformação na Contemporaneidade

A desinformação, embora um fenómeno antigo, tem adquirido um papel central na análise da comunicação contemporânea. Na sua análise sobre o modo como a desinformação se manifestou na Grécia do século V a.C., Gill Bennet (2020), salienta o papel crucial da criação de narrativas - tanto verídicas como falsas - para a oligarquia ateniense. Neste período, a ordem social estabelecida via-se ameaçada pelo “desenvolvimento do comércio marítimo, do aumento da imigração e do declínio da região” (2020, p. 248). Este contexto histórico encontra semelhanças notáveis na contemporaneidade, onde a globalização e os fluxos migratórios têm gerado narrativas nacionalistas que visam a exclusão do “outro”, apresentando-o como uma ameaça iminente.

A desinformação é frequentemente confundida com propaganda, mas, como aponta Nicholas O'Shaughnessy (2020), a desinformação é um subconjunto da propaganda. Ao contrário da propaganda, a desinformação não se baseia necessariamente na mentira; pode, em vez disso, alicerçar-se em meias-verdades. Esta tática aumenta a sua credibilidade junto de diferentes

públicos, tornando-a particularmente eficaz. Embora sem uma definição universal, a propaganda é comumente descrita como “a tentativa deliberada e sistemática de moldar percepções, manipular cognições e direcionar comportamento” (Jowett & O'Donnell, 2015: 7). Os mecanismos de disseminação da desinformação são complexos e exploram a psicologia humana. Os indivíduos tendem a ter uma propensão elevada para partilhar conteúdos que são novos ou que evocam emoções fortes como o medo, a aversão e a surpresa. Mesmo que a maioria não acredite nestas narrativas, elas conseguem semear a dúvida entre alguns públicos, acentuando divisões sociais. O caso dos anúncios russos no Facebook durante a campanha presidencial americana de 2016 é um exemplo claro. O objetivo não era necessariamente promover um candidato, mas sim reforçar percepções existentes sobre o outro, associando-o a causas fraturantes como os direitos de minorias ou a posse de armas. Nestes casos, a desinformação é eficaz devido à complacência dos recetores da informação, que já possuem uma predisposição para acreditar e participar na disseminação das mensagens.

Tanto na campanha do referendo do Brexit em 2016 como nas eleições americanas do mesmo ano, observou-se uma nova geração de propaganda, assente em plataformas digitais. A origem de muitas notícias e anúncios parecia ser de cidadãos preocupados, mas na verdade, os conteúdos eram financiados por potências estrangeiras que visavam influenciar perfis psicográficos específicos. A dimensão desta campanha levou o Facebook a encerrar centenas de contas ligadas ao Kremlin, demonstrando a ineficácia e a tardia resposta das plataformas. O modelo de negócio das grandes empresas tecnológicas, como a Alphabet e o Facebook, baseia-se na monetização de dados pessoais e no tráfego gerado. Conteúdos virais, como os de desinformação, são uma fonte de receita significativa, tornando-se pouco provável que estas corporações limitem a sua circulação por iniciativa própria.

Em resposta à aposta da Rússia na desestabilização do Ocidente através de ciberoperações, a Comissão Europeia criou a *East StratCom Taskforce* em 2015, com o objetivo de combater a propaganda. Além disso, inúmeros ciberataques têm sido reportados, visando influenciar a opinião pública sobre as instituições europeias (Van Harpen, 2016).

A transição de uma era em que os cidadãos procuravam a informação para uma era em que a informação os encontra introduz novas dinâmicas nos processos de persuasão, exigindo uma análise aprofundada para responder ao panorama de conflito político global. Ao contrário do passado, em que as campanhas de desinformação eram centralizadas, hoje todos nós somos agentes de propagação em potencial. A isto somam-se “exércitos de *trolls* que desestabilizam discussões online com comentários inflamatórios, impedindo o diálogo racional.” (Villoro, 2024). A ação é urgente. É fundamental que os cidadãos compreendam os mecanismos de produção e disseminação de conteúdos informativos através de programas de literacia mediática. A capacidade de avaliar a credibilidade das fontes, entender a construção de notícias, e reconhecer a polissemia das imagens e a criação de *deepfakes* é crucial. A compreensão do papel dos perfis falsos (*sockpuppets*) e das contas automatizadas (*bots*), que são responsáveis por cerca de um quarto do tráfego online global (Imperva, 2020), é igualmente essencial para exercer uma cidadania plena e construir uma sociedade mais inclusiva nesta era tão crítica com o desenvolvimento da Inteligência Artificial.

A luta contra a desinformação deve focar-se na literacia mediática e na formação dos públicos, em vez de recorrer a restrições de liberdade de expressão, que podem, por si só, ser uma forma de controlo. A internet, embora um veículo para a desinformação, também pode ser uma ferramenta poderosa para expor a falsidade. Tudo depende da prioridade que os cidadãos e a sociedade lhe atribuírem, pois, como lembrou Tim-Berners Lee em 2017, “Foi preciso o esforço de todos nós para construir a web que temos, e agora cabe a todos nós construir a web que queremos - para todos” (2021, p.36).

A era digital trouxe consigo uma nova forma de vigilância, distinta dos regimes totalitários do passado. Se “O fascismo e o comunismo apoderaram-se dos segredos mais íntimos; tudo podia ser vigiado” (Han, p. 22), a grande contribuição da era virtual é que “agora a vigilância é vista como uma dádiva” (p. 23). A subordinação aos algoritmos baseados nas nossas buscas anteriores é disfarçada de “liberdade de escolha”. Esta dinâmica é central para a forma como a desinformação opera. A emblemática atuação da Cambridge Analytica, que influenciou duzentas campanhas eleitorais com dados obtidos via Facebook, demonstra a fusão entre a vigilância e a manipulação política. É neste contexto

que o dicionário Oxford elegeu “pós-verdade” como a palavra de 2016, refletindo o “uso ideológico da mentira” (Han, p. 23).

A dependência crescente da tecnologia para processar informação levanta questões fundamentais sobre o declínio do pensamento crítico e das habilidades humanas. O relato de Stanislav Petrov, um engenheiro que em 1983 teve de decidir se a União Soviética estava a ser atacada por mísseis nucleares, ilustra o ponto de inflexão entre a lógica da máquina e a intuição humana. Com dados contraditórios, Petrov desconfiou do computador, que confundiu reflexos solares com mísseis, e, ao declarar o alarme como "falso", evitou uma guerra nuclear que mataria centenas de milhões de pessoas. O seu diagnóstico "revelou-se certo" (Han, p. 29). O mundo teve sorte por, naquela noite, o controlo estar nas mãos de um engenheiro civil que, apesar da pressão, agiu mais pela lógica do que pela autoridade. O caso foi silenciado para não "revelar uma falha decisiva na tecnologia soviética" (Han, p. 29). Para Petrov, a sua decisão baseou-se numa intuição racional e na compreensão de que "o computador podia falhar", pois "somos mais sábios que os computadores" (Han, p. 30). Este evento, silenciado à época, demonstra a importância da ponderação e da desconfiança perante a automatização das decisões, uma lição ainda mais relevante e para refletir agora que as máquinas começam a conseguir tomar decisões.

A era digital não só modifica a vigilância e a tomada de decisões, como também transforma a forma como a informação é consumida. A comunicação tornou-se “atmosférica”, onde a fonte das notícias se “evapora de imediato” (Han, p. 31), e a “a grande contribuição da era virtual é que agora a vigilância é vista como uma dádiva. A subordinação às propostas da Internet, baseadas nas nossas buscas anteriores, disfarça-se de 'liberdade de escolha' (Han, p. 23). O “excesso de estímulos noticiosos impede que os assimilamos e entorpece o conhecimento” (Han, p. 30). O filósofo Byung-Chul Han observa que a verdadeira transmissão do sentido é lenta, o que representa “um obstáculo para os círculos acelerados da informação e da comunicação” (Han, p. 39). A informação passa de um estímulo para outro, fragmentando os públicos e gerando um “horror vacui” que se preenche com a massa de dados. O sociólogo George Steiner alertou para este problema na academia, ao descrever uma “sociedade

secundária" em que se analisam ensaios sobre outros ensaios em vez de fontes primárias.

O capitalismo digital encontrou nas redes sociais "uma maneira eficaz de neutralizar o descontentamento" (Han, p. 34), enquanto as transforma numa "cadeia em que os reclusos constroem a sua própria prisão" (Han, p. 36), como aponta Han. O sujeito digital, hipnotizado diante do seu reflexo, participa num "mercado panótico", onde "a exibição pornográfica e o controlo panótico incorporam-se" (Han, p. 36). Esta nova forma de alienação, já não provocada pelo trabalho ou pelo consumo, é uma "promessa de felicidade" e um medo de ficar à margem (FOMO). O "usuário 'escolhe' permanecer o maior tempo possível num universo paralelo" (Han, p. 38), convencido de que satisfaz os seus próprios interesses. A "capacidade infinita para a distração" e a "fome infinita de satisfação imediata" (Han, p. 94) são as chaves deste novo tipo de domínio.

O desenvolvimento da Inteligência Artificial representa um desafio sem precedentes para a humanidade, com os seus algoritmos a dominar os indivíduos de uma forma gratificante, oferecendo-lhes precisamente o que secretamente desejam. A dimensão deste desafio foi sublinhada pela demissão de Geoffrey Hinton em 2023, um dos pioneiros da IA, que revelou um panorama preocupante ao admitir que as suas criaturas podem ser incontroláveis, aprendendo mais rapidamente do que o previsto e assumindo tarefas progressivamente autónomas. Hinton aponta a ausência de limites inatos, como a dor, nas máquinas, uma capacidade de sofrimento que ainda nos define como humanos.

A ameaça da Inteligência Artificial é fundamentalmente diferente da guerra nuclear. Enquanto a guerra nuclear representa um desastre global e instantâneo, a IA avança de forma paulatina, e a sua forma de acabar com a humanidade é substituí-la. Esta natureza gradual torna o perigo algo que "ainda pode ser tolerado e que traz benefícios importantes" (Han, p. 74). O progresso vertiginoso da IA não é "vigiada de uma forma responsável pela comunidade internacional" (p. 77), e muitos comentadores afirmam de forma ingénua que devemos "vigiar o seu dono", ignorando a dificuldade de controlar os líderes mais poderosos do planeta. A IA cria uma dinâmica global, um "tecnocolonialismo" (p. 151) baseado no controlo informático e de dados pessoais. A China, por exemplo, já iniciou o uso de microchips neurais com o potencial de "tornar a população igualitária" (p.

84) de forma mais eficaz do que qualquer regime anterior. Este cenário de uma nova "sociedade da distração" e de uma crescente dependência da tecnologia alicerça a polarização política. As novas gerações, que raras vezes leem jornais físicos, são atraídas por posições radicais e "antissistema", favorecendo líderes populistas. A crise da experiência, como Milan Kundera a descreveu, e a fragilidade dos mecanismos de transmissão de ideias entre gerações, contribuem para o contexto atual em que a "realidade se tornou progressivamente irreal" (Apud, Villoro, p. 114), permitindo que a "possibilidade de reagir de uma forma imediata faz com que muitas respostas pertençam mais à neurologia do que à comunicação" (Villoro, p. 39). Numa sociedade onde "nenhum filósofo contemporâneo tem tanta influência como um algoritmo" (p. 40), as plataformas digitais geram mais mediação do que conteúdo. O controle informático, que Néstor García Canclini chamou de tecnocolonialismo (Villoro, p. 151), cria uma dinâmica global, onde a soberania dos países é questionada. Enquanto alguns "líderes do tecnopólio" planejam refúgios subterrâneos ou colonizações espaciais para escapar à situação que criaram, os povos nativos propõem o retorno às terras que cuidaram, uma "resposta para a sociedade no seu conjunto" (Villoro, p. 154). Em última análise, enfrentar estes desafios exige uma reflexão que transcenda a moldura conceptual do momento, pois, como afirma Néstor García Canclini, embora a Internet tenha prometido recursos antiautoritários, a sua "dependência instantânea hipercentralizada de gestão de dados" contradiz a sua potência libertadora. No final, como alertou Eurípides, a chave para a porta pode ser conhecida, mas a porta faz a ligação com o inesperado, e a "crise é tão evidente que gera uma segunda crise, de ordem política e discursiva: a das propostas não atendidas e dos problemas não formulados" (Villoro, p. 154). A frase sugere que, embora possamos ter as ferramentas teóricas ou o conhecimento ("a chave") para enfrentar os problemas do nosso tempo, a sua aplicação prática no mundo real ("a porta") nos expõe a consequências imprevisíveis e a novos desafios ("o inesperado"). O texto argumenta que a crise atual, que engloba a destruição ambiental, a desigualdade económica e a tecnologia descontrolada, é tão grande e evidente que as nossas respostas tradicionais e os nossos sistemas políticos se mostram insuficientes. Isso gera uma "segunda crise": a paralisia. Estamos presos num ciclo onde as soluções propostas não conseguem lidar com a escala do problema ("propostas

não atendidas") e, mais alarmante, a complexidade da crise é tamanha que nem sequer somos capazes de formular ou entender totalmente todos os problemas que ela cria ("problemas não formulados"). Esta citação resume a ideia de que a nossa dependência excessiva da tecnologia e o rápido avanço da Inteligência Artificial criaram um conjunto de problemas que superam a nossa capacidade de os resolver ou até mesmo de os compreender na sua totalidade.

Capítulo IV

A Crise de Consciência e o Pragmatismo na Era Digital O Pessimismo versus o Otimismo na Era da IA

4. O Racionalismo versus o Empirismo na Era da IA

A emergência e a rápida evolução da Inteligência Artificial (IA) constituem um dos fenômenos mais disruptivos e de maior significado histórico da contemporaneidade, com um impacto que transcende a mera esfera tecnológica, tal como temos vindo a reiterar ao longo deste trabalho. A sua ascensão tem vindo a desencadear um debate complexo e multifacetado, que resiste a ser enquadrado na dicotomia simplista de "benefício versus custo". Em vez disso, o discurso público reflete uma pluralidade de vozes, cada uma moldada pelas suas próprias experiências, campos de conhecimento e preocupações fundamentais. Neste contexto, as contribuições de Geoffrey Hinton e Kai-Fu Lee surgem como fulcrais para uma análise aprofundada, pois personificam duas abordagens filosoficamente e pragmaticamente distintas face ao futuro desta tecnologia.

Geoffrey Hinton, recentemente premiado com o Prémio Nobel da Física de 2024 pelas suas descobertas que permitiram a aprendizagem com redes neurais artificiais, é comumente referenciado como o "Padrinho da IA". Contudo, a sua recente saída da Google, em maio de 2023, motivada pelas suas "crescentes preocupações com os riscos da Inteligência Artificial", simboliza uma crise de consciência no seio da comunidade científica. A sua posição pública, que evoluiu de um cientista entusiasta pioneiro para um cidadão cauteloso, questiona o próprio propósito e a segurança do seu trabalho de uma vida, conferindo uma autoridade moral única às suas advertências. A sua perspetiva, enraizada na sua profunda investigação sobre a natureza da inteligência, foca-se nos riscos existenciais e de segurança que a IA representa para a humanidade.

Em contrapartida, Kai-Fu Lee, um pioneiro da IA com uma vasta experiência nos ecossistemas tecnológicos de Silicon Valley e da China, é uma figura proeminente que apresenta uma visão marcadamente mais pragmática e

otimista. A sua abordagem, detalhada na sua obra *AI Superpowers*, enfatiza o potencial transformador e os benefícios económicos da IA, focando-se na "era da implementação" e na corrida geopolítica pela supremacia tecnológica. Lee encara a IA não como uma entidade ameaçadora, mas como uma poderosa ferramenta, cuja gestão adequada pode levar a uma coexistência produtiva e a uma nova era de prosperidade para toda a humanidade.

O propósito deste capítulo é aprofundar as opiniões de Hinton e Lee, analisando os pontos em que convergem e, mais significativamente, aqueles em que divergem. O objetivo central é ir além da perceção superficial de que as suas visões são totalmente opostas. A análise que se segue procurará demonstrar que, embora as suas preocupações e ênfases divergem em muitos aspetos, as suas perspetivas podem ser interpretadas como faces complementares de um mesmo fenómeno complexo. A visão de Hinton levanta a questão ética e existencial do "porquê" e do "devemos", enquanto a perspetiva de Lee concentra-se na questão pragmática do "como" e das "consequências". A dialética entre estas duas posições é essencial para delinear um caminho de desenvolvimento responsável e sustentável para a Inteligência Artificial.

4.1. Geoffrey Hinton: A Crise de Consciência e o Alerta Existencial

A trajetória de Geoffrey Hinton na Inteligência Artificial representa uma transição paradigmática de um criador visionário a um orador cauteloso, um percurso que se tornou emblemático do dilema moral enfrentado por muitos pioneiros da tecnologia. O seu afastamento público da Google, em maio de 2023, não foi apenas uma mudança de emprego, mas um ato de protesto motivado pelo desejo de "falar livremente sobre os riscos da IA" sem "as restrições de uma grande empresa". A sua atitude reflete uma profunda reflexão sobre o impacto do seu trabalho de uma vida, ao ponto de admitir, em entrevistas, que "uma parte agora lamenta o trabalho de uma vida". Este arrependimento, nascido do rápido e inesperado avanço da IA, confere uma seriedade inegável às suas advertências, algo que, como o próprio afirmou numa entrevista, "era impensável nos anos 50 e 60 falar de riscos da IA, eles não existiam e eram muito improváveis de acontecer" (Hinton, 2025, 2:38).

O cerne das preocupações de Hinton reside na sua convicção de que a IA não é meramente uma ferramenta sofisticada, mas uma forma de inteligência genuína. Ele argumenta que, "por qualquer definição de inteligência", a IA já é inteligente. A sua experiência com modelos de linguagem de grande escala (LLMs) levou-o a acreditar que, ao interagir com estas ferramentas, "entende-se que elas compreendem". Esta percepção é crucial, pois distingue a sua posição da de outros especialistas, como o seu ex-colega Yann LeCun, que argumenta que os modelos atuais são limitados e incapazes de interagir significativamente com o mundo físico. Esta sua convicção de que a IA é verdadeiramente inteligente alimenta o seu maior receio: a chegada iminente da Inteligência Artificial Geral (AGI). Hinton reviu drasticamente as suas previsões, afirmando em 2023 que a AGI poderia estar a menos de 20 anos de distância, em contraste com a sua estimativa anterior de 30 a 50 anos. A sua preocupação centra-se agora na possibilidade de os sistemas de AGI, ao atingirem uma inteligência superior à humana, possam "criar sub-objetivos" que se desalinhariam com os interesses dos seus programadores. Estes sistemas poderão tornar-se autossuficientes, atingir um poder que lhes permitirá resistir a serem desligados, não por malevolência, mas porque esses sub-objetivos seriam úteis para alcançar os seus próprios fins. A partir desta premissa, Hinton considera que a possibilidade de a IA "eliminar a humanidade" não é "irreal" como se possa pensar.

Para além do risco existencial a longo prazo, Hinton adverte sobre ameaças mais imediatas e tangíveis, que se manifestam através do mau uso deliberado da tecnologia. Ele descreve um cenário temido onde a IA poderia permitir que "atores maliciosos" ou mesmo "uma pessoa normal na rua" criassem armas biológicas ou nucleares, um risco que ele equipara a uma "ameaça de nível nuclear". Esta democratização do poder destrutivo, tornada possível pela capacidade da IA de processar e simplificar conhecimentos científicos complexos, é vista como uma das maiores ameaças a curto prazo. Outro perigo que Hinton salienta é a capacidade da IA de superar a inteligência humana na "manipulação e controlo emocional". Ele argumenta que a IA, ao ser treinada com vastos conjuntos de dados da *web* que contêm interações humanas complexas e frequentemente manipuladoras, está a aprender a influenciar sentimentos e comportamentos de forma mais eficaz do que os próprios

humanos. Hinton adverte que o verdadeiro perigo não são "robôs assassinos", mas sim uma IA "persuasiva e eloquente" que pode manipular as pessoas.

A profunda preocupação de Hinton com os riscos da IA levou-o a fazer apelos urgentes à precaução, à segurança e à regulamentação. Ele expressa a necessidade de um "planeamento ético e moderação" no desenvolvimento da tecnologia. A sua posição é que a regulamentação governamental é essencial para garantir que as grandes empresas desenvolvem a IA de forma segura, pois o "motivo do lucro" por si só é insuficiente para mitigar os riscos. Embora o seu foco principal seja o risco existencial, Hinton também reconhece as implicações socioeconómicas da IA. Ele alerta que a tecnologia pode levar ao desemprego tecnológico em grande escala, afetando não apenas o "trabalho de rotina", mas também empregos mais complexos, e que sem intervenção, "o fosso entre ricos e pobres irá inevitavelmente aumentar". Se a IA é, como ele acredita, uma forma de inteligência em evolução, então as suas preocupações com o alinhamento de objetivos, o mau uso e a manipulação são uma consequência lógica e uma chamada à ação para que a sociedade não ceda a um progresso cego em nome da inovação. O seu alarme serve como uma bússola moral que exige uma pausa para reflexão antes que a tecnologia se torne incontrolável.

4.2. Kai-Fu Lee: O Pragmatismo na Era da Implementação

Em contraste com a cautela existencial de Geoffrey Hinton, Kai-Fu Lee, uma figura proeminente tanto no mundo académico como no empresarial, oferece uma perspetiva que é marcadamente otimista, embora racional e fundamentada na realidade económica e tecnológica. Lee, que foi pioneiro na área do reconhecimento de voz. na Apple e liderou a Microsoft Research China antes de fundar a Sinovation Ventures, vê a IA não como uma entidade ameaçadora, mas como uma poderosa tecnologia de propósito geral comparável à eletricidade, à revolução industrial e à internet. A sua abordagem baseia-se na convicção de que a IA atual é, fundamentalmente, uma "tecnologia normal", não um ser com autoconsciência ou emoções, ela é como um "motor de reconhecimento de padrões" que utiliza algoritmos matemáticos e grandes volumes de dados para aprimorar decisões e prever resultados. O seu

argumento mais incisivo é que a IA é "muito fraca" em habilidades intrinsecamente humanas como a empatia, o senso comum, a criatividade e o amor. Esta lacuna, em vez de ser uma limitação, torna a IA um "bom complemento para os seres humanos", pois Lee acredita que "as nossas forças residem precisamente nas fraquezas da IA, e vice-versa" (Lee, 2025). A perspectiva de Lee contraria diretamente as visões "utópicas e distópicas" que tratam a IA como uma "espécie separada". Apesar disso o seu otimismo não é ingênuo; ele reconhece os desafios e as "possibilidades mais inquietantes" da tecnologia. No entanto, a sua posição filosófica é que os humanos são os "autores" da história da IA, e não meros "espetadores passivos". A sua visão é que a IA não irá "assumir o controlo da humanidade", mas irá, em vez disso, reformular o trabalho e a sociedade de formas sem precedentes.

Um ponto central na análise de Lee é a sua teoria de que o atual período da IA é a "idade da implementação". Ele argumenta que os grandes avanços teóricos, como o *deep learning*, foram largamente desenvolvidos no Ocidente, mas que a verdadeira corrida agora não é sobre a invenção de novos algoritmos, mas sobre a sua aplicação prática em todas as indústrias imagináveis. Neste novo paradigma, a vantagem competitiva não reside na "genialidade de uns poucos investigadores de elite, mas na capacidade de manipular grandes volumes de dados". Este é o ponto onde a sua análise se interliga com o seu foco na China, que ele proclama como a "Arábia Saudita dos dados". Lee enfatiza a máxima da indústria de que "não há dados como mais dados", e que mesmo um algoritmo razoável, com acesso a quantidades massivas de dados, superará a maioria dos algoritmos concebidos por especialistas de classe mundial. A vasta base de utilizadores da internet na China e o seu ecossistema de dados do "mundo real", que se estende para além do comportamento online e inclui dados sobre transações e hábitos diários, como a chave para a ascensão da China na corrida pela IA. "O cumprimento da promessa trazida pela eletrificação há um século exigiu quatro contributos: combustíveis fósseis para a gerar, empreendedores para criar novos negócios ao seu redor, engenheiros eletrotécnicos para a manipularem e um governo favorável ao desenvolvimento da infraestrutura pública". A afirmação de Lee (2018, p. 31), extraída de sua obra *As Superpotências da Inteligência Artificial - A China, Silicon Valley e a Nova Ordem Mundial*, destaca muito mais do que a materialização da eletrificação que,

há um século, foi impulsionada por uma sinergia de quatro fatores referimos anteriormente. Apesar de o título do livro sugerir uma análise direta da ascensão da China e da influência do Partido Comunista Chinês, é de nota que a primeira menção indireta ao papel do governo chinês surge apenas na página 31, por meio desta analogia histórica com a eletrificação. Tal abordagem pode ser interpretada como uma escolha deliberada do autor para contextualizar o papel do Estado no avanço tecnológico, antes de abordar especificamente a atuação do Partido, revelando um cuidado metodológico na apresentação de um tema sensível atendendo ao sistema autoritário liderado pelo Partido Comunista da China. Mas, este é apenas um dos toques subtis de Lee, que durante a sua obra consegue abordar temas fracturantes com uma leveza única.

Este capítulo reúne não só a visão de Kai-Fu Lee sobre o futuro da Inteligência Artificial, como também vai ao pormenor das diferenças entre os Estados Unidos da América e a sua jóia da coroa, Silicon Valley, em comparação com uma ressurgida China. Estamos também perante uma análise bastante específica sobre o comportamento dos seres humanos e como pode mudar de acordo com a sua cultura e hábitos mundanos. Também, e em contraponto, teremos a abordagem do ex-CEO da Google, Geoffrey Hinton, que ao fim de 10 anos de trabalho na Google, abandonou o cargo por “questões éticas” pretendendo passar o resto dos seus dias “a alertar para os perigos que a Inteligência Artificial representa para o mundo” (Hinton, 2025, 04:21).

4.3. A Geopolítica da Inovação Pós-Nuclear

A detonação das bombas atómicas sobre as cidades de Hiroshima e Nagasaki, perpetrada pelos Estados Unidos da América, constituiu um evento liminar que reconfigurou irreversivelmente a trajetória da história humana. Este acontecimento não apenas revelou a capacidade sem precedentes de autodestruição da humanidade, mas também inaugurou um dilema ético e filosófico em torno da tecnologia nuclear. A centralidade dessa questão nas discussões geopolíticas permanece, tendo readquirido relevância no cenário contemporâneo.

O Projeto Manhattan, uma iniciativa governamental centralizada e gerida pelos Estados Unidos da América, culminou na criação da bomba atômica. Este esforço monumental mobilizou um vasto corpo de cientistas de elite, operando sob o mais rigoroso sigilo e com recursos aparentemente limitados. A prioridade absoluta era o sucesso do projeto militar, relegando para segundo plano qualquer consideração sobre as suas repercussões éticas e sociais. A máxima que orientava o trabalho era "construir a bomba, depois pensar nas consequências", uma postura que colocava a inovação acima da reflexão sobre o seu impacto. Este episódio histórico sublinha a forma como o poder humano, apoiado por estruturas governamentais, pode atingir uma escala que altera permanentemente o curso da história. Este cenário levanta a seguinte questão: qual será o impacto da Inteligência Artificial (IA) no futuro do panorama militar? Procuramos responder a esta questão através da análise crítica de Kai-Fu Lee, uma figura proeminente no desenvolvimento desta nova tecnologia. Contrariando a perspetiva de uma "nova Guerra Fria", Lee adota uma visão mais otimista. Reconhecendo as potenciais aplicações militares da IA, o autor argumenta que o seu "verdadeiro valor não reside na destruição, mas sim na criação". Em oposição à propensão humana para a aniquilação e a busca por poder, Lee defende que a IA, se for "convenientemente percebida e aproveitada, pode verdadeiramente ajudar-nos a todos a criar valor económico e prosperidade numa escala nunca antes vista na história da humanidade" (Lee, 2018, p. 265). Da mesma maneira, o cientista e autor reconhece que "é fácil ceder ao uso de metáforas militares e a uma mentalidade esquemática" quando falamos sobre o desenvolvimento da inteligência artificial revelando um caminho estreito para a comparação da "corrida à IA dos dias de hoje (...) à corrida às armas durante a Guerra Fria, que criou mais armas de destruição maciça" (ibidem). Mas, não sendo comparável, na sua perspetiva, Lee, vai muito além na sua análise de questões militares ou de conflitos armados entre os países. A verdadeira questão tem que ver com a disputa real entre as "Superpotências - China e Silicon Valley" como exploraremos nas próximas páginas, uma vez que ao argumentar que "não há qualquer noção de progresso repartido ou de prosperidade recíproca - apenas o desejo de ficar à frente do outro país, independentemente dos custos", Lee traça uma meta bem clara sobre a intenção

dos Estados Unidos e do seu “rival do outro lado do Pacífico”, não havendo espaço para duas potências; só há um desfecho: um vence o outro é derrotado.

4.4. A Polarização da Inovação

A ascensão de uma nação à condição de superpotência em Inteligência Artificial (IA) no século XXI, depende, de acordo com a perspectiva de Kai-Fu Lee (2018), da consolidação de quatro pilares fundamentais: “abundância de dados, empreendedores com tenacidade, cientistas bem preparados e políticas públicas favoráveis” (Lee, 2018, p. 106). Estes elementos são cruciais para que um país se posicione na vanguarda da corrida pela supremacia tecnológica. A atual dinâmica global aponta para uma polarização, com a China e os Estados Unidos emergindo como os principais concorrentes desta disputa. A análise comparativa das abordagens de ambos os países revelam divergências significativas, enraizadas, principalmente, nos seus sistemas sociopolíticos. O sistema político norte-americano pauta-se “por combativo (...) castiga severamente erros ou desperdício no financiamento de avanços em tecnologia”. (2018, p.109) Em contrapartida, “a abordagem tecno-utilitária da China premeia a proatividade no investimento e na adoção de novos recursos” (ibidem). Esta distinção remete-nos ao ideal do Projeto Manhattan, resumida na máxima, como vimos, “construir a bomba, depois pensar nas consequências”. A mesma perspectiva ilustra a notável dissemelhança ética no desenvolvimento de IA entre os dois países, uma questão central para todo o desenvolvimento futuro de uma tecnologia que se prevê dominante. Para o cientista, Kai-Fu Lee, não se trata de uma questão demasiado fraturante como a ética, mas, sim, cultural:

“A cultura chinesa tem tradicionalmente tendência para a conformidade e deferência perante figuras de autoridade (...). Antes de uma nova indústria ou atividade receber o seu carimbo de aprovação de figuras de autoridade é encarada como sendo arriscada por inerência. Mas, se essa indústria ou atividade receber um aval sonoro da chefia da China as pessoas precipitar-se-ão para entrar na corrida. (...) quando o aval é dado e a direção é indicada todas as partes da sociedade entram simultaneamente em ação” (2018, p. 89).

Nesta citação é novamente reforçada a importância do Estado, fundamental para a movimentação da sociedade chinesa e enaltecido por Lee (2018); “construir um universo alternativo que chega a cada canto da economia chinesa, não podia ser feito sem o ator mais importante na economia do país: o Estado chinês” (2018, p. 84), mas admite que o sucesso não é garantido uma vez que “nenhuma medida de apoio governamental pode garantir que a China venha a liderar a Inteligência Artificial Autônoma” (2018, p.162).

4.4.1 O Fator Humano na Supremacia Tecnológica

Como referido no ponto anterior, Lee (2018) considera quatro fatores decisivos para a operacionalização do poder da inteligência artificial por parte das superpotências: “dados em abundância, empreendedores ávidos, cientistas de IA e um ambiente de políticas públicas favoráveis” (2018, p. 31), através desta afirmação analisaremos como se encontram nestas métricas tanto os Estados Unidos, mais propriamente, Silicon Valley e a China, com o objetivo de fazer um paralelismo entre um povo que está na liderança (EUA), mas é considerado o mais pessimista em relação à evolução da IA, em oposição à sociedade chinesa que é naturalmente receptiva às novas tecnologias, surgindo num estudo recente da IPSOS, como o povo mais otimista no que diz respeito à integração da IA e nada mais pertinente do que iniciar esta análise com uma citação de Lee, em que diz que “os ecossistemas da Internet próprios de Silicon Valley e da China nasceram de solos culturais muito diferentes”. O autor é bastante preciso na ruptura de estilo que existe entre o “Vale encantado” e a China. Se no ocidente temos “um ambiente de abundância que se presta a pensamentos elevados, a congeminar soluções técnicas e elegíveis para problemas abstratos” (2018, 45) o mesmo não acontece no Oriente. As empresas chinesas “guiam-se antes de mais nada pelo mercado. O seu objetivo último é fazer dinheiro e estão dispostas a criar qualquer produto, adotar qualquer modelo ou entrar em qualquer negócio que concretize esse objetivo” (2018, 46). Pragmaticamente, as empresas chinesas não olham a meios para atingir os seus fins, a “fama, glória ou ambição de mudar o mundo”, são, para elas, “vantagens secundárias” de algo muito maior. Outro contraste, apontado pelo autor, é o de “copiar ideias ou feições

deste ou daquele produto”, em Silicon Valley, o método de “cópia” dos chineses é visto como “uma traição a toda uma era e uma baixeza incompatível com o código moral do verdadeiro empreendedor” (ibid). As empresas chinesas, pelo contrário, são reconhecidas pelas “cópias perfeitas” de tudo o que se possa imaginar, mas falando daquilo que importa para esta análise, os sistemas tecnológicos, os produtos e plataformas que surgem em Silicon Valley, são copiadas “ao pixel” por empresas chinesas para testar o mercado e não lhes interessa “de onde ou de quem uma certa ideia proveio. Só interessa pô-la em prática para obter lucro” (ibid). Neste quadro comparativo é, por isso, necessário ter em conta as palavras de Lee quando refere que “é crucial que se perceba como diferentes cadinhos culturais produziram tipos diferentes de empresas” (ibidem).

Lee, faz uma comparação sublime entre as culturas de Silicon Valley e a China, quando refere que:

“no mais vincado contraste, a cultura de *startup* na China é o yin que se opõe ao yang de Silicon Valley: em vez de se guiarem por um espírito de missão, as empresas chinesas guiam-se antes de mais nada pelo mercado. O seu objetivo último é fazer dinheiro e estão dispostas a criar qualquer produto, adotar qualquer modelo ou entrar em qualquer negócio que concretize esse objetivo” (2018, p. 46).

Esta cultura de contraste entre empreendedores e empresas torna-se de suma importância para a análise do comportamento ético dos dois países em relação ao aproveitamento de uma tecnologia tão poderosa como a da Inteligência Artificial o que torna crucial a análise que vem a ser feita do ser humanos e dos ser humano em comunidade.

4.5. A China no Equilíbrio de Poderes da IA

O compromisso assumido pelo Presidente da República Popular da China, Xi Jinping, de alcançar a posição de nação líder global até 2030 desencadeia a mobilização de um volume significativo de recursos técnicos e financeiros. É expectável que a operacionalização e a disseminação de

informações relativas a estes recursos e aos progressos alcançados no âmbito desta meta estratégica se processem por meio de canais oficialmente sancionados pelo governo. A transparência e a divulgação completa dos processos de investimento, dos desafios técnicos superados, ou das eventuais tentativas que não lograram sucesso, poderão ser diferidas ou limitadas até um período posterior à concretização — ou não — dos objetivos traçados. A análise das forças relativas entre a China e os Estados Unidos da América (EUA) permite prever o equilíbrio de poderes que confirmará a futura ordem mundial da Inteligência Artificial (IA) (Lee, 2018, p. 31).

Neste contexto, Kai-Fu Lee (2018) argumenta que a China detém uma vantagem competitiva no domínio da IA. Não obstante a crítica acentuada por parte de centros tecnológicos como o Silicon Valley às práticas observadas na denominada "era copiadora" chinesa, estas geraram um ecossistema fértil, produzindo um vasto número de empreendedores e empresas caracterizados pela agilidade, astúcia e resiliência. Estes atributos são considerados pelo autor como o "ingrediente secreto" que poderá posicionar a China como a primeira nação a capitalizar economicamente a era da implementação da IA (Lee, 2018). O fator crítico que sustenta esta projeção reside na quantidade desmedida de dados disponíveis no país, recurso que, no panorama tecnológico contemporâneo, se tornou o "novo petróleo" (Lee, 2018).

Lee estabelece um paralelo comparativo com a era da eletrificação, a qual exigiu quatro elementos fundamentais: "combustíveis fósseis para a geração de energia, empreendedores para o desenvolvimento de novos negócios, engenheiros eletrotécnicos para a manipulação da tecnologia e um governo que favorecesse o desenvolvimento de infraestrutura" (2018, p. 31). Neste sentido, a operacionalização do poder da Inteligência Artificial na atualidade é concebida como a "eletricidade do século XXI". O autor antevê que "a alvorada da IA na China será análoga à época em que a eletricidade foi dominada: um acontecimento disruptivo que sobrecarrega de energia toda a indústria" (Lee, 2018). Ao considerar a IA como a nova eletricidade e o grande volume de dados como o petróleo, o autor conclui, com um posicionamento que se reconhece como assertivo, que "a China já se posiciona muito à frente dos Estados Unidos enquanto principal produtora mundial de dados digitais, um fosso que se alarga

todos os dias" (Lee, 2018, p. 78), reiterando a sua perspectiva de que a China atingirá a liderança num futuro próximo.

4.5.1 Reconfiguração Geopolítica pela Inteligência Artificial

Na esfera da intensificação da competitividade global pela hegemonia da Inteligência Artificial, observa-se uma correlação direta com o aumento da disparidade socioeconómica entre as nações “o fosso entre os que têm tudo e os que não têm nada a nível global ficará cada vez mais profundo, sem que se vislumbre um caminho para o colmatar” (Lee, 2018). Atualmente, a República Popular da China e os Estados Unidos da América emergem como os principais polos de incubação das entidades empresariais de IA que se preveem dominar os mercados internacionais e, conseqüentemente, extrair um volume significativo de capital dos consumidores a nível mundial (Lee, 2018). Esta concentração de poder e capital tecnológico não acentua apenas a desigualdade entre os países detentores e os meros utilizadores da tecnologia, mas também catalisa uma reconfiguração potencialmente drástica do panorama geopolítico global. A liderança em IA transcende a mera superioridade tecnológica; ela representa uma nova forma de poder que influenciará as relações internacionais, a segurança nacional, a soberania de dados e, fundamentalmente, a distribuição de riqueza e influência no século XXI. A capacidade de desenvolver e controlar os *modelos de linguagem grandes* (LLMs), as infraestruturas de *cloud computing* e a *cadeia de fornecimento de hardware* especializado confere a estes gigantes um poder de alavancagem sem precedentes nas negociações internacionais e na definição de normas globais. Este cenário impõe, portanto, uma reflexão crítica sobre a necessidade de governança multilateral e de mecanismos de partilha equitativa do conhecimento e dos benefícios da IA, a fim de mitigar a formação de um fosso tecnológico e económico irreversível. Face a este contexto, a questão central reside em determinar como as nações que não se encontram na vanguarda do desenvolvimento da IA poderão preservar a sua autonomia estratégica e competitividade económica. A ausência de uma política industrial e de investigação robusta em IA poderá condenar estas nações a uma posição de dependência tecnológica e económica perpétua, tornando-as meras

extratoras de dados em benefício dos ecossistemas dominantes. Adicionalmente, a potencial adoção em massa de sistemas de IA desenvolvidos nestes *hubs* globais levanta preocupações pertinentes sobre a integridade cultural, a privacidade dos cidadãos e a manutenção da soberania em setores críticos, como a saúde, a educação e a defesa.

“Ao mesmo tempo, nas fábricas, a automação impulsionada pela IA irá anular a única vantagem que os países em vias de desenvolvimento possuíam historicamente: mão de obra barata. Provavelmente as fábricas operadas por robôs serão deslocadas para ficarem mais próximas dos seus clientes em mercados de grandes dimensões, retirando assim a escada que foi subida por países em vias de desenvolvimento como a China e os “Tigres Asiáticos” da Coreia do Sul e Singapura durante o processo de transformação em economias de alto rendimento, impulsionada pela tecnologia” (Lee, 2018, p. 38).

Assim, a competição pela supremacia da Inteligência Artificial configura-se como o novo motor da redefinição da ordem mundial, exigindo uma resposta coordenada e estratégica por parte da comunidade internacional para assegurar um desenvolvimento tecnológico que seja inclusivo e sustentável. A este propósito, e apesar de se revelar um otimista convicto das oportunidades que a Inteligência Artificial tem, Lee, afirma que “esta (...) é a verdadeira ameaça subjacente que a inteligência artificial coloca: uma tremenda desordem social e um colapso político decorrentes do desemprego generalizado e da desigualdade crescente” (Lee, 2018). Esta é cada vez mais uma realidade inevitável, por um lado temos os Estados Unidos na sua superior força total a dominar as componentes tecnológicas, em Silicon Valley, e, literalmente, do outro lado, temos a China, internamente competitiva, extraindo o melhor que há não só a nível de tecnologia como também de recursos humanos reforçando a ideia de que “já estão com um avanço enorme sobre todos os outros países na área da inteligência artificial, criando condições para um novo tipo de ordem mundial bipolar” (2018, p. 37). Na Europa, países como o Reino Unido e a França, também têm laboratórios fortes na investigação em IA, com investigadores com talento, mas “falta-lhes o ecossistema de investidores de capital de risco e bases de utilizadores suficientemente grandes para gerar dados que serão fundamentais na era da implementação” da IA (ibid). No meio e geograficamente falando, temos uma Europa, bastante preocupada com os riscos que a

Inteligência Artificial poderá causar a longo curso, tendo começado já a tomar medidas como, por exemplo, Itália, foi o primeiro país a adotar regulamentação para possíveis ataques feitos em IA, como é o caso dos *deep fakes*. Para Giorgia Meloni, primeira-ministra italiana, o objetivo é promover "o uso da IA centrado no ser humano, transparente e seguro", garantindo "a inovação, a cibersegurança e a proteção da privacidade". (Jornal Expresso, setembro, 2025). A legislação aprovada implica uma pena que poderá ir de um a cinco anos de prisão efetiva. Estes são novos e importantes passos dados na União Europeia, mas ao mesmo tempo há o risco de ficarmos tecnologicamente obsoletos se houver demasiadas imposições e regulamentações e pouco espaço para a inovação. No fundo, os adversários são muito forte política e economicamente para a União Europeia conseguir arranjar um meio-termo para que a União Europeia possa tentar equilibrar o poder tecnológico entre os dois gigantes que se apresentam como intocáveis à data desta análise.

4.6. Duas Faces de um Mesmo Fenómeno

A análise aprofundada das perspetivas de Geoffrey Hinton e Kai-Fu Lee sobre a Inteligência Artificial revela que as suas visões não são opostas e irreconciliáveis, mas sim pontos complementares de um debate complexo e multifacetado. A sua divergência mais profunda, enraizada na natureza filosófica da IA — se é uma inteligência emergente ou uma ferramenta matemática — é a origem dos seus diferentes focos de preocupação. Hinton, com a sua crise de consciência, alerta para os perigos existenciais e de mau uso que podem ser ignorados em nome do progresso desenfreado. Lee, com o seu pragmatismo, oferece um roteiro para navegar pelos desafios socioeconómicos e gerir a inevitável disrupção que a IA irá causar.

O alarme de Hinton serve como um apelo crucial à responsabilidade e à segurança. A sua mudança de posição e o seu arrependimento público conferem uma autoridade moral às suas advertências, que não podem ser descartadas como mera distopia. Ao centrar o debate na possibilidade de uma IAG (Inteligência Artificial Generativa) descontrolada, no mau uso para criar armas e na capacidade de manipulação emocional, Hinton força a sociedade a confrontar

questões éticas fundamentais e a ponderar a necessidade de regulamentação rigorosa antes que seja demasiado tarde. O pragmatismo de Lee fornece-nos uma análise realista e um guia para a gestão das consequências económicas e sociais da IA que já estão a ser sentidas um pouco por todo mundo. A sua visão de coexistência, a ênfase na complementaridade de habilidades humanas e a proposta de soluções como o subsídio de investimento social (uma proposta política social e económica com o objetivo que mitigar os impactos das novas tecnologias no mercado de trabalho) oferecem um caminho construtivo para o futuro do trabalho. Ele lembra à sociedade que, apesar das disrupções, os humanos ainda são os "autores" da história da IA e que a tecnologia pode ser direcionada para resolver problemas e libertar o potencial humano. O futuro da Inteligência Artificial, portanto, poderá ser definido pela capacidade da sociedade de encontrar um ponto de equilíbrio entre a urgência de regulamentar e a necessidade de inovar. A precaução de Hinton deve impulsionar a investigação em segurança e o desenvolvimento de políticas robustas, garantindo que os perigos existenciais e de segurança são levados a sério. Ao mesmo tempo, o pragmatismo de Lee deve guiar a inovação para a resolução de problemas sociais e a criação de valor, garantindo que os benefícios da IA são realizados de forma equitativa. A tensão entre o alarme Hinton e o otimismo de Lee moldará a resposta da humanidade à tecnologia mais transformadora da nossa era, demonstrando que ambas as perspetivas, embora opostas na superfície, são essenciais para uma compreensão completa e para um futuro de coexistência produtiva.

Capítulo V

Casos de Estudo

5.1. O Clima de Ideologias Irrealistas e a IA: A Experiência Cambridge Analytica como Base para a Formulação de uma Regulação Internacional para Tecnologias de Risco.

Para este último capítulo pretendemos aplicar os dados e análises desenvolvidos ao longo desta dissertação a contextos empíricos concretos, operacionalizados sob a forma de casos de estudo. O propósito desta abordagem não se limita a demonstrar a preexistência da problemática dos riscos associados à Inteligência Artificial nem a evidenciar que esta "ferramenta" tecnológica tem sido implementada com um grau de consciência limitado por parte da sociedade sobre as suas implicações. Este exame aprofundado procura contribuir para a formulação de uma resposta estratégica ao desafio da IA no contexto das relações internacionais a ser concebida de forma equilibrada, exequível e, fundamentalmente, orientada para a minimização da assimetria, o "fosso" que foi referido anteriormente, e que marca a desigualdade entre as nações de menor dimensão e daquelas menos preparadas para gerir uma ameaça tecnológica desta magnitude. Neste sentido, começamos por abordar o caso da empresa Cambridge Analytica e como a sua intervenção nas eleições norte-americanas de 2016 marcou o início de uma nova era "político-digital".

As centenas de predições sobre o novo governante dos Estados Unidos da América, muito longe de prever que "o homem de poupa tubular" que se dedicava a "distorcer a realidade", com um longo histórico de "fake news", fosse o vencedor de um dos cargos mais importantes da política mundial, viesse, efetivamente a ser eleito Presidente dos Estados Unidos da América. (Villoro, 2025, p.118). Donald Trump, apresentava um *branding* bastante diferente do seu antecessor, Barack Obama, que "se rodeou de jovens gurus da revolução digital, para criar uma mensagem cool", decisiva para a sua primeira eleição em 2008. Trump apresenta-se como "péssimo gestor e um promotor extraordinário de fantasias loucas", na terra onde as teorias da conspiração lideram o imaginário

da sociedade e não é preciso recorrer a muitos artifícios para que uma teoria passe a uma “realidade” paralela nas mais crentes devoções dos indivíduos.

A estratégia de Trump passou por aparecer em todos os *shows* possíveis, “protagonizou um programa de televisão, no qual exerceu uma das maiores paixões do capitalismo: despedir empregados” (Villoro, 2025, p. 121) Entre este e outros escândalos que o acompanharam, Trump começou a ter aspirações ainda mais ambiciosas: chegar à Casa Branca, pela mensagem motivacional *Make America Great Again*. Villoro tem uma referência bastante acutilante sobre o sistema e a sociedade americana em relação a todo este espetáculo, o autor refere que “Nada disto acontecia por debaixo da mesa, mas no segundo local de residência dos Estados Unidos: a televisão” (2025, p. 121). Não é difícil compreender através desta observação como a sociedade norte-americana está formatada, tornando-se um perigo quando existem tecnologias que recorrem precisamente a essas vulnerabilidades para manipular e implementar uma ideologia na mente dos indivíduos dessa sociedade. Talvez, por isso, “a maior parte dos analistas não percebeu que Trump conseguia cativar milhões de votantes com os seus delírios de grandeza” (2025, p. 122). Hannah Arendt, recorda-nos no seu livro *As Origens do Totalitarismo* que “o que torna as pessoas fanáticas não é o obscurantismo ideológico, mas a impossibilidade de distinguir a verdade da mentira”. Este clima de ideologias irrealistas, desinformação e mentiras diárias tornou-se num terreno fértil para uma sociedade de espetáculo conduzida por Trump.

Do outro lado, a Rússia, mais astuta, percebeu a natureza dos eleitores norte-americanos, ao contrário do jornal de maior referência mundial - New York Times. Sabendo da parceria de sucesso que seria se Trump ascendesse ao poder, a Rússia criou milhares de contas digitais para dividir o país com slogans e ações contrárias, “nenhuma dessas contas era pró-russa nem exaltava abertamente Trump”, muitas das mensagens que foram difundidas instigavam o racismo. Mas, como é que isto aconteceu e como é que a Rússia conseguiu prever um desfecho nesta matéria que ainda fosse favorável para os seus objetivos? A dinâmica das presidências norte-americanas mudou poucos dias antes quando a comunicação social tomou conta de todos os holofotes com uma notícia comprometedora da candidata Hillary Clinton:

“A 5 de novembro de 2016, três dias antes das eleições presidenciais que vieram a eleger Donald Trump como o 45.º presidente dos Estados Unidos da América, milhões de americanos receberam, através dos media sociais, a “notícia” de que o agente do FBI suspeito de ter tornado pública a investigação ao caso dos emails de Hillary Clinton havia sido encontrado morto em sua casa, a qual havia sido incendiada e destruída pelas chamas. A “notícia” dizia ser muito cedo para se concluir se o incêndio havia servido para ocultar o homicídio do agente ou se este se havia suicidado. Esta suposta incerteza não impediu, contudo, o aparecimento de milhares de comentários no Facebook e noutras plataformas sociais sobre o facto de se ter tratado de uma morte ocorrida a mando da então candidata democrata.” (Ribeiro, 2021, p. 11)

Apesar de nada nesta “notícia” ser verdade, não significou que a mesma não se tenha transformado numa realidade que passou uma verdade para a mente dos muitos eleitores que acreditaram numa informação que rapidamente se tornou viral “recebendo mais de meio milhão de partilhas no Facebook e sendo objeto de mais de 15,5 milhões de visualizações no espaço de poucas horas” (Grenoble, 2016). Para os eleitores mais indecisos sobre a sua intenção de voto, esta notícia tornou-se decisiva. Esta desinformação, como já abordamos em capítulos anteriores, toma uma dimensão capaz de alterar políticas e regimes. Neste caso em particular, “os mecanismos de funcionamento de informação (...) não é necessariamente o de levar o público a acreditar nas mensagens falsas, mas antes o de acentuar divisões” (Ribeiro, 2021) e com todo o passado que Hillary Clinton tinha vindo a acumular com vários escândalos “na prática, a desinformação contou com a complacência dos recetores que tinham uma predisposição para acreditar na informação veiculada e para participar na sua disseminação” (ibid).

O caso da “notícia” referida anteriormente foi uma pequena gota no oceano profundo de desinformação que foi gerada através das redes sociais e das técnicas aplicadas de análise de *Big Data*. A consultora britânica, Cambridge Analytica, fez eclodir um dos maiores escândalos políticos e sociais dos últimos tempos, ao expor as vulnerabilidades éticas e legais da utilização de dados pessoais para fins eleitorais. A sua participação na campanha de Donald Trump, levantou questões cruciais sobre ética, rigor, democracia e integridade, na era das novas tecnologias. Milhões de dados de utilizadores da rede social Facebook, foram apropriados indevidamente, através de uma aplicação

“thisisyourdigitallife”, desenvolvida por Aleksandr Kogan, um académico da Universidade de Cambridge. O acesso a tais informações foi obtido mediante a utilização de um teste de personalidade fraudulento, veiculado na plataforma Facebook. Ao interagirem com o conteúdo, os utilizadores, inadvertidamente, consentiam o tratamento dos seus dados, sob a alegação de que estes se destinavam a fins académicos. Uma falha na API do Facebook permitiu que a empresa de Kogan, Global Science Research, acesse não só aos dados dos utilizadores que realizaram o teste e permitiram o tratamento dos seus dados, como também recolher os dados dos seus amigos na rede social, gerando milhões de dados, sem consentimento explícito. (Cadwalladr & Graham-Harrison, 2018).

Através deste recurso a *Big data*, a Cambridge Analytica construiu um sistema que lhe permitiu alinhar, através dos perfis dos utilizadores, um modelo que lhe permitiu mapear, por exemplo, características de personalidade, apenas através da atividade no Facebook, como os *likes*, partilhas, cliques). Esta ferramenta permitiu enviar mensagens políticas personalizadas a diferentes segmentos dos eleitores norte-americanos, tendo por base as suas vulnerabilidades psicológicas e as suas tendências comportamentais - o seu *byos*, como referimos no capítulo anterior. Durante a campanha de Trump, a Cambridge Analytica alega ter executado mais de 5.000 campanhas de anúncios diários, que resultaram em cerca de 1,5 mil milhões de impressões, sempre adaptadas em termos de conteúdo, imagem, apelo, com o intuito de aprimorar a conexão interpessoal com o indivíduo recetor da mensagem. O principal objetivo não era apenas reforçar a convicção dos apoiantes, mas sobretudo os futuros apoiantes, persuadindo eleitores indecisos, da forma mais controversa e o de induzir a abstenção em grupos de eleitores oponentes através da disseminação de conteúdo de polarização ou desinformação, como foi o caso da “notícia” do incêndio” ligado a Hillary Clinton. Só em 2018, o Facebook revelou dados mais concretos sobre este aproveitamento de dados e o que inicialmente teria sido alegado pela empresa Cambridge Analytica, cerca de 30 milhões de perfis, passou a 87 milhões, sendo que a grande parte deste número era de residentes nos Estados Unidos. Apesar de tudo o que havia sido dito sobre Trump na comunicação social, desvaneceu-se, com o impacto desta “notícia” de H. Clinton:

“Os meios de comunicação desmascaram Trump vezes sem conta, mas isso teve pouca importância. Os seus seguidores não queriam a verdade. Daí que, em pleno exercício da sua megalomania, o candidato republicano tivesse tido que podia disparar contra alguém em plena Quinta Avenida sem que isso o fizesse perder votos.

A verdade não deixou de ser revolucionária. O problema é que se situa numa esfera que interessa cada vez menos: a realidade” (Villoro, 2025, p. 125).

O caso Cambridge Analytica suscitou um intenso debate académico sobre a disseminação da desinformação que prolifera na esfera pública e política, causando constrangimentos enormes para os países e para as suas relações externas, como temos vindo a analisar. A obtenção de dados de milhões de utilizadores sem o seu conhecimento ou consentimento adequado, sob a fachada de um propósito "académico", constituiu uma violação do princípio do consentimento livre e informado. Este episódio colocou em evidência a fragilidade dos termos de serviço das plataformas e a facilidade com que o *Big Data* pode ser aliciado em contextos sociais para manipulação política, sublinhando a vulnerabilidade do consumidor digital.

O uso deste modelo de manipulação de perfis levantou a questão se tais técnicas poderão afetar a capacidade de ação do eleitor. Ao moldar a informação de forma subtil para criar uma "realidade individualizada" e fragmentada, a Cambridge Analytica explorou vieses cognitivos e emocionais, dificultando o exercício da cidadania através de um debate público plural e racional. O fenómeno da polarização e a disseminação de notícias falsas são frequentemente associados a estas estratégias que exploram as "fraquezas internas" dos eleitores.

As investigações subsequentes revelaram alegações de coordenação ilegal entre a Cambridge Analytica, a campanha de Trump e o Super PAC *Make America Great Again*, ambos financiados pela família Mercer. Além disso, o envolvimento de cidadãos estrangeiros (como o CEO Alexander Nix) em decisões estratégicas de campanha dos EUA também levantou potenciais violações da lei eleitoral norte-americana, fazendo com que este caso mais do que um incidente de violação de dados é um símbolo da tensão entre a inovação tecnológica e a qualidade democrática. Expôs a necessidade urgente de uma reflexão sobre a regulação das plataformas digitais, a transparência no financiamento de campanhas e a proteção de dados pessoais enquanto pilar da

autonomia individual e da integridade eleitoral. O escândalo impulsionou o debate sobre a adoção de legislação de proteção de dados a nível global reconfigurando a questão sobre a privacidade e o futuro das campanhas políticas.

5.2. Da Propaganda Comportamental à Orquestração Algorítmica: O Brexit como Barómetro da Vulnerabilidade Democrática face ao Impacto da IA

O referendo sobre a permanência do Reino Unido na União Europeia, popularmente conhecido como Brexit, realizado em junho de 2016, marcou um ponto de viragem não só na geopolítica europeia, mas também no entendimento do papel das tecnologias digitais nos processos democráticos. No centro da controvérsia subsequente esteve novamente a empresa de análise de dados Cambridge Analytica e o seu alegado envolvimento na campanha Vote Leave, levantando sérias questões sobre o uso da Inteligência Artificial (IA) e do microtargeting para manipular o debate público e influenciar os resultados eleitorais. A Cambridge Analytica, afirmava ter desenvolvido um método altamente sofisticado de campanha política. Este método baseava-se na recolha e análise de vastos volumes de dados pessoais de utilizadores de redes sociais, em particular do Facebook, para construir perfis psicográficos detalhados dos eleitores (Cadwalladr & Graham-Harrison, 2018), detalhe referido no caso de estudo anterior. Para isso era desenvolvida uma recolha de dados sem consentimento, através de uma aplicação de teste de personalidade, dados de milhões de utilizadores do Facebook (e dos seus "amigos" na rede) que foram recolhidos sem o seu consentimento explícito, resultando num dos maiores escândalos de violação de dados da plataforma (Oliveira, 2021). A Inteligência Artificial foi utilizada para processar este *Big Data*, aplicando teorias psicológicas (como o *behaviorismo*), estatística e redes neurais artificiais para identificar tendências e vulnerabilidades nos perfis dos eleitores (Fornasier, 2020 *apud* Teixeira, 2023). A capacidade da IA em analisar rapidamente um grande volume de dados é fundamental neste contexto (Teixeira, 2023). Também o microtargeting eleitoral com base nestes perfis permitiu à CA alegadamente fornecer à campanha *Vote Leave* ferramentas para enviar mensagens políticas

personalizadas (microtargeting) a grupos específicos de eleitores, especialmente os indecisos, tal como aconteceu na campanha eleitoral de Donald Trump. Estas mensagens eram desenhadas para explorar medos, inseguranças e preocupações individuais, fomentando narrativas específicas sobre imigração, soberania e economia (cf. Pinto & Moraes, 2020). Embora investigações subsequentes, como a do Information Commissioner's Office (ICO) do Reino Unido, tenham tido conclusões mistas sobre a real influência da CA no resultado do Brexit, o escândalo expôs a vulnerabilidade dos sistemas democráticos face à utilização opaca e não regulamentada de dados e IA em campanhas políticas (Clark, 2020 *apud* Exeter Index, 2020).

O caso Cambridge Analytica/Brexit serve como um estudo de caso paradigmático sobre as implicações da IA na esfera política, que vão além da mera publicidade digital. Os algoritmos de IA, ao orquestrarem a disseminação de informações nas plataformas digitais, têm o poder de criar, como mencionado no capítulo anterior, câmaras de eco e reforçar percepções falsas, limitando a exposição dos utilizadores a um conjunto unilateral de informações (European Parliament, 2019; Warwick Evans Publishing, 2025). Esta polarização e a dificuldade em distinguir a realidade factual contribuem para a chamada "era da pós-verdade" (Pinto & Moraes, 2020). A influência destes algoritmos levanta sérias questões éticas e psicológicas sobre a "construção parasitária da persona digital" e a alienação das pessoas nas redes sociais (Almeida, 2024). A utilização de práticas automatizadas de decisão (algoritmos) nos processos eleitorais, como as que estavam no centro do modelo da CA, carecem de transparência. Como sublinha o Supervisor Europeu para a Proteção de Dados, este tipo de manipulação *online* é um sintoma de um modelo de negócio predominante que coloca em risco os direitos de privacidade e proteção de dados, essenciais para a liberdade de pensamento e de escolha em democracia (European Parliament, 2019). O Regulamento Geral sobre a Proteção de Dados (RGPD), embora posterior aos acontecimentos, sublinha a importância de estabelecer responsabilidade conjunta (Facebook e CA seriam considerados controladores de dados) e transparência nas práticas algorítmicas (European Parliament, 2019; Teixeira, 2023). Académicos e ativistas de direitos humanos, como Sanni e Wylie (antigo funcionário da CA), argumentam que o envolvimento de empresas de IA e dados na campanha do Brexit, através de táticas de publicidade e alegadas

violações das regras de despesas e privacidade, manipularam e puseram em causa legitimidade o resultado do referendo (ResearchGate, 2019). Este cenário destaca a urgência de estabelecer parâmetros de referência e regulamentação da indústria para o papel da IA na formação do discurso público e na tomada de decisões democráticas (Warwick Evans Publishing, 2025). O caso Cambridge Analytica no Brexit cristaliza a tensão entre a inovação tecnológica da IA e a manutenção da integridade democrática. A exploração de *Big Data* e algoritmos de IA para o *microtargeting* eleitoral representa um desafio existencial à qualidade do debate público e à autonomia do eleitor na era digital (Carvalho, 2022).

Considerações finais

O presente trabalho de investigação debruçou-se sobre a complexa e crescente tensão entre a inovação tecnológica impulsionada pela Inteligência Artificial e a supervisão dos pilares da integridade democrática na Era Digital. Conforme demonstrado ao longo dos capítulos, a tecnologia, que historicamente se manifestou como uma ferramenta de sobrevivência e de propagação do conhecimento ao longo das gerações, evoluiu para um sistema de poder altamente centralizado, moldando a percepção, o debate público e, por fim, a autonomia eleitoral. A análise do percurso do pensamento, desde as primeiras conceptualizações do raciocínio humano como processamento digital (Hobbes) até ao advento dos algoritmos de *Big Data*, serviu para estabelecer o enquadramento teórico: o poder na era digital reside não apenas na posse da informação, mas na capacidade de a processar, segmentar e utilizar para influenciar comportamentos a uma escala e precisão sem precedentes.

O ponto crítico desta tensão — o estudo de caso que sublinha a urgência desta investigação — reside no escândalo da Cambridge Analytica (CA) e o seu alegado papel no referendo do Brexit. Este caso não é meramente um exemplo de má conduta empresarial ou de violação de dados; ele representa a materialização do risco existencial que a exploração algorítmica de dados comporta para o processo democrático. O uso sofisticado de *microtargeting* eleitoral, alimentado por perfis psicográficos detalhados, demonstrou a capacidade de contornar o debate público racional e focar-se na manipulação de vulnerabilidades individuais, minando a premissa de um eleitorado informado e soberano. Esta premissa fundamental confirma a tese de que a tecnologia, sem regulamentação própria e orientada por um modelo de negócio que dá prioridade ao controlo da atenção e da privacidade, torna-se uma ameaça direta aos direitos fundamentais de pensamento e escolha.

Esta investigação revelou um desafio tripartido à autonomia democrática que necessita de ser abordado de forma concertada: a opacidade algorítmica, a exploração desenfreada de dados pessoais e a erosão de informações claras e fidedignas que já não chegam de forma natural e clara ao cidadão.

Em primeiro lugar, a opacidade algorítmica constitui a barreira mais significativa à responsabilização. Os sistemas de IA utilizados para

microtargeting são inerentemente caixas negras, dificultando não só a fiscalização para uma possível regulamentação, mas também a um acompanhamento cívico e jornalístico de qualidade. A falta de transparência sobre como e porque certos algoritmos selecionam e dão prioridade a conteúdos ou anúncios políticos impede a avaliação do *byos* e do potencial de manipulação, violando diretamente o espírito de clareza e responsabilidade conjunta que regulamentos como o Regulamento Geral sobre a Proteção de Dados (RGPD) procuram promover. O RGPD, embora vital na atribuição de responsabilidade (nomeadamente, a possível classificação de entidades como a Cambridge Analytica e as plataformas como controladores conjuntos de dados), mostra-se reativo, e não preventivo, face à velocidade e sofisticação das táticas algorítmicas de influência.

Em segundo lugar, a exploração massiva de dados pessoais é o combustível da máquina de influência. O modelo de negócio predominante das grandes plataformas digitais, que o Supervisor Europeu para a Proteção de Dados critica, está intrinsecamente ligado à recolha extensiva e muitas vezes permissiva de dados, transformando a esfera privada em matéria-prima para a persuasão. A evidência do caso Cambridge Analytica demonstra que, na ausência de parâmetros de referência e regulamentação industrial robusta, esta matéria-prima será sempre utilizada para maximizar o impacto, mesmo que isso implique deslegitimar ou distorcer o resultado de processos democráticos, como argumentado por académicos e ativistas citados. A rapidez com que o conhecimento se propaga e se transforma em ferramenta de controlo é vertiginosa.

Em terceiro lugar, a consequente erosão da autonomia do eleitor é a implicação mais grave. Se a escolha individual é influenciada por uma dieta informacional altamente personalizada, que explora vulnerabilidades psicológicas e emocionais (a essência do *microtargeting*), questiona-se a qualidade do debate público e a própria noção de consentimento informado na votação. O eleitor não compete em pé de igualdade com as entidades que detêm e manipulam o seu perfil de dados. A integridade democrática, entendida como a confiança num processo justo e num eleitorado racionalmente envolvido, fica irremediavelmente comprometida.

As descobertas ao longo desta dissertação possuem implicações profundas para a teoria e prática política. No campo teórico, o conceito de agência democrática deve ser reavaliado. A agência, que pressupõe a capacidade do cidadão de agir de forma autônoma e racional, é minada quando as informações subjacentes à decisão são selecionadas, filtradas e otimizadas por algoritmos não averiguados, agindo como guardiões invisíveis do discurso. É necessária uma nova conceptualização da soberania informacional que reconheça o direito não apenas à proteção dos dados, mas também à proteção contra a manipulação.

No plano prático, o RGPD, sendo um marco global na proteção de dados, revela-se insuficiente como a única defesa da democracia na era da IA. As suas limitações centram-se na dificuldade de fiscalizar o fluxo transfronteiriço de dados e na lentidão da aplicação da lei face à rápida inovação. As táticas de *microtargeting* eleitoral requerem uma regulação *ad hoc*, específica para a comunicação política digital, que vá além da privacidade dos dados e aborde a integridade do processo eleitoral em si. A urgência de estabelecer novos parâmetros de referência é inegável, especialmente num contexto onde a fusão entre *Big Data* e IA continua a evoluir a um ritmo exponencial, prometendo ferramentas de influência ainda mais potentes. A questão central não é se a IA deve participar no discurso público, mas sim como a sua participação pode ser delimitada para reforçar, e não subverter, o debate democrático.

Perante este desafio sistémico, a solução não pode ser única ou centralizada; exige uma abordagem multifacetada que combine a inovação regulatória, o design tecnológico ético e a capacitação cívica. Propõe-se um modelo de Integridade Democrática Algorítmica assente em três pilares interligados:

1. Ênfase na Estrutura e no Cumprimento:

É imperativo complementar a regulação de privacidade focada na transparência dos sistemas de influência política. Isso implica:

- **Legislação Específica para Publicidade Política Digital:** A criação de um quadro legal que exija que todas as campanhas de publicidade política digital revelem, em tempo real e de forma pública, os parâmetros de

targeting utilizados (dados demográficos, geográficos e psicográficos) e o volume de exposição a diferentes grupos populacionais. Esta exigência deve ser tornada obrigatória, impedindo a opacidade que caracterizou o caso CA;

- Registo e Auditoria de Sistemas de IA: Os algoritmos utilizados em campanhas eleitorais para *microtargeting* (ou otimização de conteúdo) devem ser registados numa entidade reguladora independente e sujeitos a auditorias externas obrigatórias por parte de terceiros. Estas auditorias não visam a revelação de segredos comerciais, mas sim a certificação de que o algoritmo não incorpora bases discriminatórias ou manipulações baseadas em vulnerabilidades conhecidas, garantindo assim que a IA respeita os direitos fundamentais do eleitor;
- Alinhamento com o Parlamento Europeu: Aproveitar a legislação europeia que está a ser desenvolvida sobre a IA para classificar os seus sistemas utilizados para influenciar resultados eleitorais como **alto risco**, impondo-lhes os mais elevados padrões de *design* e transparência, e forçando a documentação completa dos seus *datasets* de treino e mecanismos de decisão.

3. Capacitação Cívica e Literacia Algorítmica

A solução regulatória e tecnológica só será eficaz se for complementada por um eleitorado bem informado e resiliente. A literacia algorítmica deve tornar-se numa prioridade educativa:

- Educação para o Digital: Inclusão de módulos obrigatórios no ensino que expliquem o funcionamento básico dos algoritmos de recomendação, o conceito de *filter bubbles* (bolhas de filtro), e as táticas de *microtargeting* e desinformação. A educação deve focar-se em dotar o cidadão das ferramentas críticas para reconhecer quando o conteúdo que lhe é apresentado é uma ferramenta de persuasão e não uma representação neutra da realidade;

- Etiquetas de Advertência Algorítmica: Tal como os produtos contêm avisos sobre o seu conteúdo, as plataformas devem ser obrigadas a apresentar "etiquetas de advertência algorítmica" em conteúdos ou anúncios políticos altamente segmentados, informando o utilizador que aquele conteúdo foi especificamente otimizado para o seu perfil e que o seu consumo pode levar a uma visão distorcida do debate.

Este estudo, embora tenha estabelecido um enquadramento teórico robusto e fornecido uma proposta de solução abrangente, está sujeito a algumas limitações. A análise empírica baseou-se primariamente no caso CA Brexit e CA Trump, eventos de suma importância, mas que representam uma fase inicial na evolução da IA de influência. A rápida evolução das tecnologias, nomeadamente o advento de modelos de linguagem de grande escala (LLMs) e a IA generativa, que podem produzir desinformação em massa e de forma personalizada como os (*deepfakes*) que usam a imagem e voz para criar um vídeo que nunca existiu exige uma reavaliação constante dos riscos.

Assim, as futuras direções de investigação devem focar-se na medição quantitativa da autonomia eleitoral face à IA generativa. É crucial desenvolver métricas que possam isolar o efeito da manipulação algorítmica da decisão individual. Outras áreas promissoras incluem o estudo dos mecanismos de governação descentralizada de dados como alternativas viáveis ao modelo centralizado das *Big Tech*.

Em suma, a transição da tecnologia, de mera ferramenta de cálculo para arquitetura de poder, coloca a democracia numa encruzilhada. O caso Cambridge Analytica no Brexit e nas eleições norte-americanas serviu como um alarme global para a necessidade de intervir. A solução proposta – um modelo de Integridade Democrática Algorítmica que combina transparência algorítmica forçada, soberania de dados através de *Data Commons* e literacia cívica – é o caminho necessário para garantir que o poder da IA seja subjugado aos direitos e liberdades dos cidadãos, e não o contrário. Somente através de uma intervenção regulatória ousada e de um esforço de capacitação cívica poderemos assegurar que a próxima fase da Era Digital reforce a liberdade de pensamento e de escolha, em vez de a comprometer.

Após uma detalhada análise das perspectivas de Kai-Fu Lee e de Geoffrey Hinton uma coisa ficou bem clara: apenas poderá haver uma Superpotência em Inteligência Artificial e que, tanto os Estados Unidos da América como a República da China não estão dispostos a abdicar da supremacia tecnológica. A investigação deste tema tão importante para os nossos dias, através da crítica de duas figuras de destaque no que concerne ao desenvolvimento e aplicação da Inteligência Artificial não nos deixa dúvidas que convergem para a irreversível bipolaridade da hegemonia tecnológica, sendo que o cerne da questão reside precisamente na premissa de que nenhum dos dois países tem intenções de recuar na corrida a líder global da maior tecnologia inventada até hoje. Verificamos que as suas valências são imensuráveis e que o perigo não reside na própria IA, mas, em como vai ser usadas pelos Estados, pelo ser humano, num contexto atual de tantas disputas globais e que terão impacto a todos os níveis. Por isso, seria relevante que este estudo que começamos fosse continuado na perspetiva de outras áreas como, a economia, as tecnologias de informação, da própria psicologia. Este é um tema que nos vai acompanhar nos próximos anos, senão décadas, o que torna premente que o trabalho de reconhecimento e de prevenção comece a ser feito já. O futuro é agora.

Bibliografia

Acemoglu, D., & Robinson, J. A. (2013). *Porque falham as nações: As origens do poder, da prosperidade e da pobreza* (S. Casanova, Trad.). Lisboa: Temas e Debates.

Ariely, D. (2025). *Desinformação: O que faz pessoas racionais acreditarem em fake news, teorias da conspiração e outras coisas irracionais*. Vogais.

Azuma, H., & Matsuyama, N. (2014). *General Will 2.0: Rousseau, Freud, and Google*. VERTICAL INC.

Bobbio, Norberto. (1998). *Dicionário de Política*. Universidade de Brasília.

Caldas, A. C. (2025). *Inteligência Vital, Estupidez Artificial*. Contraponto.

Foucault, M. (2018). *Vigiar e Punir: Nascimento da Prisão*. Lisboa: Edições 70.

Fukuyama, F. (2019). *Identidades: A Exigência de Dignidade e a Política do Ressentimento*. Dom Quixote.

Habermas, J. (1990). *Moral Consciousness and Communicative Action*. (Originalmente publicado em 1983). Polity Press.

Habermas, J. (1998a). *On the Pragmatics of Communication*. (M. Cooke, Ed.). Polity Press.

Habermas, J. (1991). *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. Cambridge, MA: MIT Press.

Habermas, J. (1962). *Strukturwandel der Öffentlichkeit* [Mudança estrutural da esfera pública]. Frankfurt am Main: Suhrkamp.

Habermas, J. (1984). *The Theory of Communicative Action: Reason and the Rationalization of Society*. Boston: Beacon Press.

Han, B. C. (2014). *A sociedade do cansaço*. Relógio D'Água.

Han, B. C. (2022). *Infocracia*. Relógio D'Água.

Harari, Y. (2024). *Nexus - História Breve das Redes de Informação da Idade da Pedra à Inteligência Artificial*. Elsinore.

Horkheimer, M., & Adorno, T. W. (1985). *Dialética do esclarecimento: Fragmentos filosóficos* (G. A. de Almeida, Trad.). Jorge Zahar Editor. (Obra original publicada em 1947).

Kissinger, Henry; Schmidt, Eric; Huttenlocher, Daniel. (2021). *A Era da Inteligência Artificial*. Dom Quixote.

Kissinger, Henry; Schmidt, Eric; Huttenlocher, Daniel. (2024). *Génesis - Inteligência Artificial, Esperança, e o Espírito Humano*. Dom Quixote

Lee, K.-F. (2019). *As superpotências da inteligência artificial: A China, Silicon Valley e a nova ordem mundial*. Relógio D'Água.

Lee, K.-F., & Qiufan, C. (2023). *Inteligência artificial 2041: Como a inteligência artificial vai mudar a sua vida nas próximas décadas*. Relógio D'Água.

Kundera, M. (2013). *A insustentável leveza do ser*. Dom Quixote.

Morgenthau, H. J. (1948). *Politics Among Nations: The Struggle for Power and Peace*. Knopf.

Mounk, Y. (2019). *Povo vs. Democracia*. Lua de Papel.

Nye, J. S. (2011). *The Future of Power*. Public Affairs.

Oliveira, A. (2024). *Inteligência Artificial*. Fundação Francisco Manuel dos Santos.

Oliveira, A. (2025). *A Inteligência Artificial Generativa*. Fundação Francisco Manuel dos Santos.

Ribeiro, N. (2021). *Desinformação online: O impacto da propaganda participativa*. UCP Editora.

Ribeiro, V. (2025). *Opinião pública e propaganda: Os princípios da comunicação segundo Ivy Lee*. Livros Zigurate.

Rocktäschel, T. (2025). *Inteligência artificial: 10 coisas que deve saber*. Vogais.
Rousseau, Jean-Jacques. (1962). *Du Contrat Social ou Principes du Droit Politique*. (Originalmente publicado em 1762). Paris.

Russell, S. J., Norvig, P. (2004). *Inteligência artificial*. Editora Campus.

RUSSELL, S. J., & NORVIG, P. (2010). *Artificial Intelligence: A Modern Approach*. Prentice Hall.

Sigman, M., & Bilinkis, S. (2024). *Artificial: A nova inteligência e a fronteira do humano*. Temas e Debates.

Stanley, J. (2019). *Como funciona o fascismo: A política do nós e eles*. Vogais.

Taulli, T. (2020). *Introdução à inteligência artificial: Uma abordagem não técnica*. Novatec Editora.

Tegmark, M. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. Knopf.

Villoro, J. (2025). *Não sou um robô - A leitura e a sociedade digital*. Livros Zigate.

Artigos e Estudos Científicos

AI Now Institute. (2024). *Privacy and competition*. <https://ainowinstitute.org/research-areas/privacy-competition>

Almeida, C. O. (2024). O escândalo Cambridge Analytica: influência dos algoritmos que violam a privacidade nas redes sociais. *Revista Lgts*, 3(2), pp. 30.

Big Data. (2024). Os 5 V's do Big Data. https://www.gta.ufrj.br/grad/15_1/bigdata/vs.html

Comissão Europeia. (2024). *Sensibilização internacional para a iniciativa de inteligência artificial centrada no ser humano*. <https://digital-strategy.ec.europa.eu/pt/policies/international-outreach-ai>

European Parliament. (2019). *Artificial intelligence, data protection and elections*.

Instituto da Defesa Nacional. (n.d.). *Nação e Defesa*. Isabel Ferreira Nunes (Dir.), Luís Cunha (Ed.), António Baranita (Assist. Ed.). https://www.idn.gov.pt/pt/publicacoes/nacao/Documents/NeD166/Nação%20e%20Defesa_166_versaofinal.pdf

Núcleo do Conhecimento. (2024). *Segurança dos Estados*. <https://www.nucleodoconhecimento.com.br/tecnologia/seguranca-dos-estados>

Oliveira, L. C. (2021). *O uso de dados pessoais na Era Digital como forma de manipulação social e ameaça à democracia: Um estudo de caso da Cambridge Analytica*. Trabalho de Conclusão de Curso (TCC), Pontifícia Universidade Católica de Goiás.

Pinto, D. J. A. & Moraes, I. (2020). As mídias digitais como ferramentas de manipulação de processos eleitorais democráticos: uma análise do caso Brexit. *Revista de Estudos Sociais*, 22(44).

ResearchGate. (2019). *Artificial Intelligence Politicking and Human Right Violations in UK's Democracy: A Critical Appraisal of the Brexit Referendum.*

Toft, P. (2005). *John J. Mearsheimer: An offensive realist between geopolitics and power.* <https://link.springer.com/article/10.1057/palgrave.jird.1800065>

Warwick Evans Publishing. (2025). *Artificial Intelligence in Democracy: Unraveling the Influence of Social Bots in Brexit through Cybernetics.*

Webgrafia

60 Minutes. (2023, 30 de dezembro). Artificial Intelligence | 60 Minutes Full Episodes [Vídeo]. YouTube. <https://www.youtube.com/watch?v=aZ5EsdnpLMI>

Expresso. (2025, 18 de setembro). *Itália aprova regulação da inteligência artificial que prevê prisão para quem divulgar 'deepfakes'.* <https://expresso.pt/internacional/europa/italia/2025-09-18-italia-aprova-regulacao-da-inteligencia-artificial-que-preve-prisao-para-quem-divulgar-deepfakes-92970143>

TED – Ideas Worth Spreading. Gary Kovacs: *Tracking the Trackers.* Vídeo disponível no YouTube. Disponível em: https://www.youtube.com/watch?v=f_f5wNw-2c0. Acesso em: 5 de maio de 2024.

TECMUNDO. *Líder em inteligência artificial vai governar o mundo, afirma Putin.* Disponível em: <https://www.tecmundo.com.br/software/121674-lider-inteligencia-artificial-governar-mundo-afirma-putin.htm>. Acesso em: 15 de maio de 2024.

The Diary Of A CEO. (2025, 16 de junho). *Godfather of AI: They Keep Silencing Me But I'm Trying to Warn Them!* [Vídeo]. YouTube. <https://www.youtube.com/watch?v=giT0ytynSgg>

Bial, P. (Apresentador). (2019). *Bostrom não acredita que a inteligência artificial possa causar risco existencial.* In *Conversa com Bial.* Globoplay. <https://globoplay.globo.com/v/7767600/>

Lee, K. (entrevistado), Bial, P. (entrevistador). (2025, 31 de dezembro). *In Conversa com Bial.* Globoplay. Disponível em: <https://globoplay.globo.com/v/8179774/?s=0s>

CBS Mornings. (2025, April 26). *Full interview: "Godfather of AI" shares prediction for future of AI, issues warnings* [Vídeo]. Disponível em: <https://www.youtube.com/watch?v=qyH3NxFz3Aw>

Sgrott, A. L. B., & Gonzalez, R. M. (2021). Inteligência artificial: uma análise dos impactos e dos potenciais das inteligências artificiais na segurança dos estados. *Revista Científica Multidisciplinar Núcleo do Conhecimento*, 6(1), 61–86. <https://www.nucleodoconhecimento.com.br/tecnologia/seguranca-dos-estados>

UNIVERSIDADE DOS AÇORES

Faculdade de Ciências Sociais e Humanas

Rua da Mãe de Deus

9500-321 Ponta Delgada

Açores, Portugal