

UMA (BREVE) REFLEXÃO TÉCNICA SOBRE A INTELIGÊNCIA ARTIFICIAL: *STATUS QUO E FUTURUS*

GONÇALO ALMEIDA¹³⁹

SUMÁRIO: 1. Introdução. 2. A Trindade. 3. Por Detrás da Cortina Sintética. 4. *Inputs, Outputs, Qualia* e Liberdade. 4.1. Breve Desvio: Teoria das Redes Neurais.

RESUMO: Este artigo explora o funcionamento da Inteligência Artificial (IA), centrando-se nos seus princípios fundamentais, nos algoritmos principais e na arquitetura em constante evolução. Ademais, investiga a mecânica dos modelos de aprendizagem automática, incluindo os paradigmas de aprendizagem supervisionada, não supervisionada e de reforço, e examina o papel das redes neurais, dos mecanismos de *deep learning* e *machine learning*.

ABSTRACT: This paper explores the inner workings of Artificial Intelligence (AI), focusing on its foundational principles, core algorithms, and evolving architectures. It delves into the mechanics of machine learning models, including supervised, unsupervised, and reinforcement learning paradigms, and examines the role of neural networks, deep learning and machine learning mechanisms.

PALAVRAS-CHAVE: Inteligência Artificial; Redes Neurais; Autodeterminação.

¹³⁹ Licenciado em Engenharia Informática, Universidade de Coimbra; Mestrando em Engenharia Informática e da Computação, Faculdade de Engenharia da Universidade do Porto, atualmente a realizar a sua dissertação na empresa Pink Room, no ramo da Inteligência Artificial e Dispositivos móveis; Estudante na Universidade de Rijeka, Croácia, “SusTrainable Summer School”.

KEYWORDS: Artificial Intelligence; Neural Networks; Self-determination.

1. Introdução

A Inteligência Artificial (*Kunstig Intelligens*, *Kunstmatige Intelligentie*, *Intelligence Artificielle*, *Künstliche Intelligenz*, *Intelligenza Artificiale*, *Kunstig Intelligens*, *Artificiell Intelligens*) tem-se desenvolvido exponencialmente desde a sua concepção em meados do século XX¹⁴⁰.

O conceito inicial de Inteligência Artificial foi proposto pelo cientista da computação britânico CHRISTOPHER STRACHEY, ao desenvolver o primeiro programa de computador capaz de jogar damas autonomamente¹⁴¹.

¹⁴⁰ “The first work that is now generally recognized as AI was done by Warren McCulloch and Walter Pitts (1943). They drew on three sources: knowledge of the basic physiology and function of neurons in the brain; a formal analysis of propositional logic due to Russell and Whitehead; and Turing’s theory of computation. They proposed a model of artificial neurons in which each neuron is characterized as being ‘on’ or ‘off,’ with a switch to ‘on’ occurring in response to stimulation by a sufficient number of neighboring neurons”, RUSSELL, STUART; NORVIG, PETER (2020). *Artificial Intelligence: A Modern Approach*. (3ª edição). Pearson, p. 16.

¹⁴¹ STRACHEY, CHRISTOPHER (1952). *Logical or Non-Mathematical Programmes*. In Proceedings of the ACM National Meeting, p. 47.

Todavia, a origem científica da Inteligência Artificial é geralmente atribuída a ALAN TURING, matemático e lógico britânico, que introduziu a ideia de máquinas (μηχανή¹⁴²) capazes de realizar tarefas inteligentes¹⁴³, e a WARREN McCULLOCH e WALTER PITTS¹⁴⁴, ainda que a pré-história

¹⁴² A antiga palavra grega μηχανή (*mēkhanē*) referia-se originalmente a um dispositivo ou ferramenta utilizada para realizar tarefas, muitas vezes superando dificuldades ou obstáculos. Deriva da palavra μῆχος (*mēkhos*), que significa “meio” ou “dispositivo”, implicando algo que ajuda a alcançar um objetivo. Em contextos clássicos, μηχανή era frequentemente usado para descrever dispositivos engenhosos ou engenhocas, particularmente na engenharia e na guerra. Com o tempo, μηχανή tornou-se a raiz de termos modernos como “mecânica” e “máquina”, mantendo a sua associação com ferramentas, maquinaria e a aplicação do engenho para atingir fins práticos.

¹⁴³ TURING, ALAN (1950). *Computing Machinery and Intelligence*. In *Mind: A Quarterly Review of Psychology and Philosophy*, vol. LIX, n.º 236, pp. 433 e ss.

¹⁴⁴ Note-se, porém: “The first work that is now generally recognized as AI was done by Warren McCulloch and Walter Pitts (1943). They drew on three sources: knowledge of the basic physiology and function of neurons in the brain; a formal analysis of propositional logic

da Inteligência Artificial se possa vincular a pensadores como DESCARTES (com a boneca Francine)¹⁴⁵ e LEIBNIZ (com a criação do sistema de numeração binário)^{146/147}.

Hoc die, a Inteligência Artificial manifesta-se de várias formas, desde assistentes virtuais como a *Siri* e o *Google Assistant* até algoritmos complexos que operam sistemas de recomendação, como o algoritmo de recomendação de vídeos do *YouTube*

due to Russell and Whitehead; and Turing's theory of computation. They proposed a model of artificial neurons in which each neuron is characterized as being 'on' or 'off,' with a switch to 'on' occurring in response to stimulation by a sufficient number of neighboring neurons", RUSSELL, STUART; NORVIG, PETER (2020). *Artificial Intelligence: A Modern Approach*. (3ª edição). Pearson, p. 16.

¹⁴⁵ Sobre DESCARTES, com bastante interesse, GRAYLING, A. C. (2019). *Uma História da Filosofia*. Edições 70, pp. 243 e ss., e SACY, SAMUEL (1988). *Descartes*. Dom Quixote.

¹⁴⁶ Cfr., MARTINS, JOÃO MARQUES (2022). *Inteligência Artificial e Direito: Uma Brevíssima Introdução*. In *Revista da Faculdade de Direito da Universidade de Lisboa*, vol. LXIII, n.ºs 1 e 2, p. 488.

¹⁴⁷ Sobre LEIBNIZ, com bastante interesse, GRAYLING, A. C. (2019). *Uma História da Filosofia*. Edições 70, pp. 277 e ss.

ou do *Tiktok*, e veículos autônomos¹⁴⁸. A definição de Inteligência Artificial varia, mas é comumente entendida como um conjunto de sistemas de *software* e, possivelmente, *hardware*, projetados, até agora, por humanos, que, dado um objetivo complexo, agem no ambiente físico ou digital percebendo-o através da aqui-

¹⁴⁸ “Falar de inteligência artificial gera sempre alguns temores e receios devidos, essencialmente, à incerteza do seu impacto no nosso dia a dia. Embora ainda estejamos um pouco distantes da convivência ou até mesmo da existência de uma real IA consciente, a verdade é que a expansão da tecnologia, para além de uma realidade diária, desempenha um papel preponderante na sociedade contemporânea. Porém, nas páginas que se seguem a abordagem deste fenómeno complexo e transversal, apenas irá incidir na problemática da personalização dos entes dotados de inteligência artificial com fortes implicações ético-jurídicas. Pois, esta evolução coloca-nos em diálogo com vazios legais que poderão dar asas a uma equiparação desmedida, sendo precisamente aqui que emergem novas questões relacionadas com a ética no direito, abarcando uma panóplia de dúvidas, receios e uma exigência de delimitações”, VASCONCELOS, MARIA PARENTE (2020). *Inteligência Artificial: Direito e Personalidade Jurídica*. (Dissertação de Mestrado). Faculdade de Direito da Universidade de Coimbra, p. 28.

sição de dados, interpretando os dados obtidos, “raciocinando” sobre o conhecimento ao processar informação, oriunda dos dados, de modo a decidir a(s) melhor(es) ação(ões)^{149/150/151}. Se-

¹⁴⁹ “A method of thinking about knowledge models using logic, deduction, and induction as tools, with the objective of deriving new knowledge and overall understanding of the context.”, KANTOROVITCH, JULIA (2015). *Reasoning about Knowledge*. In “Encyclopedia of Information Science and Technology”, ed. KHOSROW-POUR, 3.ª edição, pp. 375-386.

¹⁵⁰ “From a historical point of view, the evolution of AI-based systems starts with artificial narrow intelligence (ANI), then continues with artificial general intelligence (AGI), and finally meets artificial super intelligence (ASI), which will surpass human capabilities in all dimensions”; “ANI refers to intelligent systems that do specific tasks. For example, an agent with capabilities such as face recognition and games playing”; “AGI is used for those agents whose intelligence is equivalent to human agents. AGI can be equivalent to human-level intelligence (HLI)”, AKBARI, V.; ZARRABI, M.; EBRAHIMI, S. (2022). *A Survey of Artificial Intelligence Challenges: Analyzing the Definitions, Relationships, and Evolutions*. In *Applied Sciences*, vol. 12, n.º 8, pp. 2-5.

¹⁵¹ KANTOROVITCH, JULIA (2015). *Reasoning about Knowledge*. In “Encyclopedia of Information Science and

gundo STUART RUSSELL e PETER NORVIG, a Inteligência Artificial é o “[...] estudo de agentes que recebem percepções do ambiente e realizam ações”¹⁵². Já JOHN MCCARTHY, um dos pioneiros no campo, define a Inteligência Artificial como a “[...] ciência e engenharia de fazer máquinas inteligentes, especialmente programas de computador inteligentes”¹⁵³. Por sua vez, sistemas de Inteligência Artificial são definidos, pelo Conselho da Europa (no artigo 2.º da *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*), como “[...] a machine-based system that, for explicit or

Technology”, ed. KHOSROW-POUR, 3.ª edição, pp. 111-112.

¹⁵² “We define AI as the study of agents that receive percepts from the environment and perform actions”, RUSSELL, STUART; NORVIG, PETER (2020). *Artificial Intelligence: A Modern Approach*. (3ª edição). Pearson, p. viii.

¹⁵³ “Q. What is artificial intelligence? A. It is the science and engineering of making intelligent machines, especially intelligent computer programs.”, MCCARTHY, JOHN (2007). *What is Artificial Intelligence?*. Departamento de Ciência da Computação, Universidade de Stanford, p. 2.

implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations or decisions that may influence physical or virtual environments”¹⁵⁴.

¹⁵⁴ Reproduzimos, aqui, parte do preâmbulo, por curiosidade: “Conscious of the accelerating developments in science and technology and the profound changes brought about through activities within the lifecycle of artificial intelligence systems, which have the potential to promote human prosperity as well as individual and societal well-being, sustainable development, gender equality and the empowerment of all women and girls, as well as other important goals and interests, by enhancing progress and innovation; Recognising that activities within the lifecycle of artificial intelligence systems may offer unprecedented opportunities to protect and promote human rights, democracy and the rule of law; Concerned that certain activities within the lifecycle of artificial intelligence systems may undermine human dignity and individual autonomy, human rights, democracy and the rule of law; Concerned about the risks of discrimination in digital contexts, particularly those involving artificial intelligence systems, and their potential effect of creating or aggravating inequalities, including those experienced by women and individuals in vulnerable situations, regarding the enjoyment of their human rights and their full, equal and effective participation in

2. A Trindade

A Inteligência Artificial pode ser categorizada em três tipos principais: (a) Inteligência Artificial Estreita, ou fraca (*artificial narrow intelligence*, ou “ANI”); (b) Inteligência Artificial Geral, ou forte (*artificial general intelligence*, ou “AGI”)^{155/156}; e (c) Superinteligência Artificial (*artificial super intelligence*, ou “ASI”)¹⁵⁷.

economic, social, cultural and political affairs”.

¹⁵⁵ Sobre esta, OLIVEIRA, ARLINDO (2019). *Mentes Digitais: A Ciência Redefinindo a Humanidade*. (3.ª edição). IST Press, p. 81.

¹⁵⁶ “Alguns pensadores destacados calculam que a primeira IAG de nível humano será um upload, e que será assim que o caminho para a superinteligência irá ter início”, TEGMARK, MAX (2024). *Life 3.0: Ser-se Humano na Era da Inteligência Artificial*. (2.ª edição). Dom Quixote, p. 217.

¹⁵⁷ “É usual classificar-se a IA de acordo com três tipos de inteligência: Fraca, Média e Forte. A primeira – ANI: Artificial Narrow Intelligence – é aquela que se especializa em uma única área, possuindo um determinado objectivo (p. ex. máquinas treinadas para jogar xadrez); a segunda – AGI: Artificial General Intelligence – imita a mente humana, sendo capaz de compreender conceitos complexos e de resolver problemas aprendendo com a sua própria experiência; a terceira – ASI: Artificial

Enquanto a Inteligência Artificial Estreita é especializada numa só tarefa específica, como jogar xadrez ou reconhecer animais em imagens, a Inteligência Artificial Geral refere-se a sistemas com capacidade de entender, aprender e aplicar conhecimentos de maneira geral, semelhante ao ser humano. A Superinteligência, por sua vez, descreve uma Inteligência Artificial que supera a inteligência humana em todos os aspetos, incluindo criatividade, resolução de problemas e *capacitas* emocional. No *status quo*, encontramos na era da Inteligência Artificial Estreita, com sistemas extremamente competentes em tarefas específicas, mas ainda incapazes de replicar a flexibilidade e adaptabilidade da

Super Intelligence – já possuiria competências sociais, igualando ou até superando o cérebro humano. No actual contexto tecnológico, ainda não existe a IA Forte”, MOREIRA, SÓNIA (2022). *IA e Robótica: A Caminho da Personalidade Jurídica?*. In “Liber Amicorum Benedita Mac Crorie”, vol. II, UMinho Editora, p. 538.

inteligência humana^{158/159/160}.

Apud MAX TEGMARK¹⁶¹, Inteligência Artificial Geral é a capacidade de uma máquina realizar qualquer tarefa cognitiva pelo menos tão bem quanto o ser humano, incluindo aprender, ao contrário da ANI, no decorrer de um jogo de xadrez virtual¹⁶². Para NICK BOSTROM, a AGI é a capacidade de uma máquina ter “bom senso” e de ser capaz de aprender, raciocinar e planear, assim como enfrentar desafios complexos de processamento de informações

¹⁵⁸ AA. VV. (2024). *Multidisciplinary Perspectives on Artificial Intelligence and the Law*. Springer, pp. 111-112.

¹⁵⁹ AKBARI, V.; ZARRABI, M.; EBRAHIMI, S. (2022). *A Survey of Artificial Intelligence Challenges: Analyzing the Definitions, Relationships, and Evolutions*. In *Applied Sciences*, vol. 12, n.º 8, p. 2.

¹⁶⁰ “ASI refers to an agent faster and more capable than human intelligence in all aspects”, AKBARI, V.; ZARRABI, M.; EBRAHIMI, S. (2022). *A Survey of Artificial Intelligence Challenges: Analyzing the Definitions, Relationships, and Evolutions*. In *Applied Sciences*, vol. 12, n.º 8, p. 5.

¹⁶¹ Com base em “apud Platonem”, RODRIGUES, FERNANDO PEREIRA (2010). *Locuções e Máximas Latinas*. Coimbra Editora, p. 40.

¹⁶² TEGMARK, MAX. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. Alfred A. Knopf, p. 55.

numa ampla gama de domínios, naturais e abstratos, de forma eficaz¹⁶³.

NICK BOSTROM discute, ainda, os riscos de uma AGI tornar-se incontrollável, destacando a importância de desenvolver salvaguardas eficazes para garantir que esse tipo de inteligência aja de acordo com os interesses do ser humano¹⁶⁴. A Superinteligência Artificial, para este autor, refere-se a sistemas de Inteligência Artificial como qualquer intelecto que sobre-exceda o desempenho cognitivo do ser humano em praticamente todos os domínios de interesse¹⁶⁵. STUART RUSSELL e PETER NORVIG, consideram que a ASI poderia realizar qualquer tarefa cognitiva de forma mais eficiente do que qualquer ser hu-

mano¹⁶⁶. Estes sistemas seriam capazes de se aprimorarem de forma autónoma e progressiva, levando a uma inteligência exponencialmente superior à humana¹⁶⁷. A ASI não replicaria apenas as capacidades humanas, mas ampliá-las-ia de formas que atualmente são inimagináveis¹⁶⁸.

Uma ASI poderia revolucionar quase todos os aspetos da sociedade, desde a medicina até à economia, oferecendo soluções inovadoras para problemas complexos¹⁶⁹. No entanto, também

¹⁶⁶ RUSSELL, STUART; NORVIG, PETER (2020). *Inteligência Artificial: Uma Abordagem Moderna*. (3ª edição). Pearson, p. 12.

¹⁶⁷ OLIVEIRA, ARLINDO (2018). *Inteligência Artificial: Disrupção e Oportunidade*. In Conferência, 21 e 22 de junho de 2018, p. 13.

¹⁶⁸ “We can tentatively define a superintelligence as any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest”, BOSTROM, NICK (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, p. 117.

¹⁶⁹ MCKINSEY & COMPANY (2023). *The Bio Revolution: Innovations Transforming Economies, Societies, and our Lives*. Disponível em <https://www.mckinsey.com/industries/life-sciences/our-insights/the-bio-revolution-innovations-transforming-economies-societies-and-our-lives>.

¹⁶³ BOSTROM, NICK (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, p. 3.

¹⁶⁴ BOSTROM, NICK (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, pp. 115-143.

¹⁶⁵ “We can tentatively define a superintelligence as any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest”, BOSTROM, NICK (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, p. 22.

levanta sérias preocupações¹⁷⁰. Aliás, figuras como STEPHEN HAWKING e ELON MUSK alertaram sobre os perigos potenciais de uma ASI descontrolada, sugerindo que poderá representar uma ameaça existencial à humanidade se não for devidamente regulamentada e controlada^{171/172}.

3. Por Detrás da Cortina Sintética

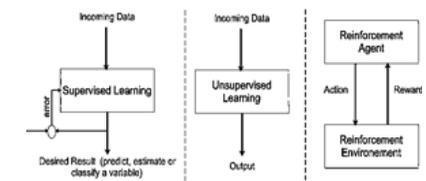
De forma a melhor entender o conceito de Inteligência Artificial, é essencial perceber o seu funcionamento. O funcionamento desta baseia-se em algoritmos que processam grandes volumes de dados para extrair padrões e tomar decisões. Técnicas de aprendizagem automática, ou *machine learning* (“ML”), e aprendizagem profunda, um ramo de ML mais conhecido

¹⁷⁰ GIBBS, SAMUEL (2014). *Elon Musk: artificial intelligence is our biggest existential threat*. In The Guardian.

¹⁷¹ ALTMAN, S.; BROCKMAN, G.; SUTSKEVER, I. (2023). *Governance of Superintelligence*. In OpenAI, disponível em <https://www.openai.com/research/governance-of-superintelligence>.

¹⁷² TEGMARK, MAX. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. Alfred A. Knopf, p. 412.

por *deep learning* (“DL”), são amplamente utilizadas para treinar algoritmos com uma extensa quantidade de dados, de forma a melhorar a sua precisão e eficiência, através do processo-exemplo apresentado na figura *infra*¹⁷³. Redes neuronais artificiais, que imitam o funcionamento do cérebro humano¹⁷⁴, têm sido particularmente eficazes em áreas como visão computacional e processamento de linguagem natural



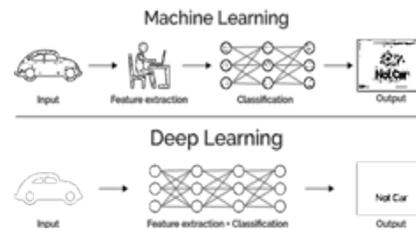
Processo de aprendizagem através de *Machine Learning* e *Deep Learning*¹⁷⁵

¹⁷³ MUELLER, JOHN PAUL; MASSON, LUCA (2019). *Artificial Intelligence for Dummies*. John Wiley & Sons, p. 162.

¹⁷⁴ CHOWDHARY, K. R. (2020). *Fundamentals of Artificial Intelligence*. Springer, pp. 415-416 e 431.

¹⁷⁵ FOURATI, H.; MAALLOUL, R.; FOURATI, L. (2021). *A Survey of 5G Network Systems: Challenges and Machine Learning Approaches*. In International Journal of Machine Learning and Cybernetics, vol. 12, n.º 1, pp. 153-177.

A aprendizagem supervisionada, não supervisionada e por reforço são três abordagens principais em *machine learning*¹⁷⁶, sendo possível observar o seu método de treino na figura *infra*. Na aprendizagem supervisionada, os algoritmos são treinados com dados rotulados (*data labeling*), onde a resposta correta (*ground truth*) é fornecida¹⁷⁷. Na aprendizagem não supervisionada, o algoritmo identifica padrões e estruturas nos dados sem rótulos pré-definidos. *In fine*, a aprendizagem por reforço envolve treinar um algoritmo através de recompensas e punições, permitindo que o sistema aprenda os comportamentos pretendidos através da experiência.



Aprendizagem supervisionada, não supervisionada e por reforço¹⁷⁸

Entre a aprendizagem supervisionada, não supervisionada e por reforço, os Grandes Modelos de Linguagem (*large language models*, ou “LLMs”) representam um avanço significativo no campo da Inteligência Artificial. Estes modelos, desenvolvidos por empresas como a *OpenAI*, a *Meta* (anteriormente conhecida como *Facebook*) e a *Google*, utilizam proeminentes quantidades de dados para treinar redes neurais profundas, permitindo-lhes compreender e gerar resultados em linguagem natural com elevada precisão¹⁷⁹. O modelo *GPT*

¹⁷⁶ Note-se que também existe aprendizagem semi-supervisionada, que não se encaixa em nenhuma das outras três, mas esta classificação varia muito de autor para autor, sendo que há quem considere que *reinforcement learning* seja uma subárea de *unsupervised learning*.

¹⁷⁷ IBM (2021). *What Is Data Labeling?*. Disponível em <https://www.ibm.com/topics/data-labeling>.

¹⁷⁸ FOURATI, H.; MAALLOUL, R.; FOURATI, L. (2021). *A Survey of 5G Network Systems: Challenges and Machine Learning Approaches*. In *International Journal of Machine Learning and Cybernetics*, vol. 12, n.º 1, pp. 153-177.

¹⁷⁹ IBM (2024). *Large Language Models (LLMs)*. Disponível em <https://www.ibm.com/topics/large-language-models>.

da *OpenAI*, por exemplo, é um dos mais avançados e amplamente utilizados, destacando-se pela sua capacidade de gerar texto coerente e relevante em diversos domínios. A *Google* desenvolveu o modelo *Gemini*, que usufrui também de capacidades multimodais, permitindo interpretar e gerar texto, imagens e áudio, além de processar longos contextos de até um milhão de *tokens*¹⁸⁰. A *Meta*, por sua vez, investe no desenvolvimento de modelos de Inteligência Artificial que visam a integração com redes sociais, em destaque nas suas, como o *Instagram*, *WhatsApp* e *Facebook*¹⁸¹.

4. Inputs, Outputs, Qualia e Liberdade

Uma das distinções mais marcantes entre a Inteligência Ar-

tificial e a Inteligência Humana reside na previsibilidade dos seus *outputs*. Enquanto os sistemas de Inteligência Artificial são concebidos para funcionar com base em algoritmos que processam dados de maneira lógica e estruturada, o comportamento humano é influenciado por um vasto leque de fatores que tornam as suas ações muito mais complexas e difíceis de prever¹⁸². Estas ações representam o culminar de uma vasta gama de fatores internos e externos, incluindo emoções, experiências passadas e contextos sociais, tornando os seus *outputs* muito mais imprevisíveis¹⁸³. A imprevisibilidade

¹⁸⁰ GOOGLE (2024). *Introducing Gemini 1.5: Our next-generation AI model*. Google Developers Blog. Disponível em <https://developers.googleblog.com/2024/02/gemini-1.5-next-generation-model.html>.

¹⁸¹ META (2023). *Introducing new AI experiences across our family of apps and devices*. Meta Newsroom. Disponível em <https://about.fb.com/news/2023/09/introducing-ai-powered-assistants-characters-and-creative-tools/>.

¹⁸² “AI is defined as ‘the study of how to make computers do things at which, at the moment, people are better’. This includes thinking of creative solutions, flexibly using contextual and background information, the use of intuition and feeling, the ability to really ‘think and understand’, or the inclusion of emotion in an (ethical) consideration”, KORTELING, J. E.; VAN DE BOER-VISSCHEDIJK, G. C.; BLANKENDAAL, R. A.; BOONEKAMP, R. C.; EIKELBOOM, A. R. (2021). *Human vs Artificial Intelligence*. In *Frontiers in Artificial Intelligence*, vol. 4, article 622364, p. 3.

¹⁸³ “Human behavior and decision making is subject to social and motivational influences such as emotions, nor-

humana é frequentemente vista como uma forma de criatividade e autodeterminação, algo que os sistemas de Inteligência Artificial ainda não conseguem replicar completamente, o que causa implicações significativas para a aplicação e integração da Inteligência Artificial na sociedade¹⁸⁴.

A questão da autodeterminação na Inteligência Artificial é complexa e ainda está em debate. *Hoc die*, os sistemas de Inteligência Artificial não são verdadeiramente autônomos nem têm, muito menos, uma consciência¹⁸⁵.

ms and self/other regarding preferences. The identification of the neural and psychological mechanisms underlying these factors is a central issue in psychology, behavioral economics and social neuroscience, with important clinical, social, and even political implications”, CORRADI-DELL’ACQUA, C.; KOBAN, L.; LEIBERG, S.; VUILLEUMIER, P. (2016). *What Determines Social Behavior? Investigating the Role of Emotions, Self-Centered Motives, and Social Norms*. In *Frontiers in Human Neuroscience*, p. 2.

¹⁸⁴ AA. VV. (2023). *Humans versus AI: whether and why we prefer human-created compared to AI-created artwork*. In *Cognitive Research*, vol. 8, p. 17.

¹⁸⁵ “Em particular, não parece que computadores estejam prestes a ter consciência, nem emoções e sensações. As

Estes operam com base em algoritmos escritos pelo ser humano e com as aprendizagens resultantes da compreensão de dados, sem *capacitas* de autodeterminação ou consciência de si mesmos. Filosoficamente, a ideia de uma Inteligência Artificial que se autodetermina levanta questões sobre a natureza da consciência e se é possível replicar a experiência subjetiva humana com uma/numa máquina¹⁸⁶. A consciência, no sentido filosófico, é frequentemente associada aos *qualia*¹⁸⁷,

últimas décadas assistiram a um avanço imenso na inteligência de computadores, mas o avanço na consciência dessas máquinas foi nulo”, FARIAS, BRAYNER DE (2020). *Sobre Animais, Humanos e Máquinas: Para onde vai a Consciência?*. In *Revista de Filosofia*, vol. 20, n.º 2, pp. 380-392.

¹⁸⁶ GUINGRICH, R. E.; GRAZIANO, M. S. A. (2024). *Ascribing Consciousness to Artificial Intelligence: Human-AI Interaction and its carry-over Effects on Human-Human Interaction*. In *Frontiers in Psychology*, vol. 15, pp. 1-2.

¹⁸⁷ *Qualia*, na sua forma singular “qualia”, é “[...] a palavra latina (‘qual tipo’) introduzida na filosofia contemporânea como termo de arte para se designar as qualidades fenomenais, subjetivas e conscientes da nossa vida mental, que seriam acessíveis mediante introspecção”, PEREIRA, ANTÓNIO (2023). *Qualia*. In *Revista de Filosofia*, vol. 42, n.º 3, p. 1.

ou experiências subjetivas, que ainda não conseguimos entender completamente, nem, *ipso facto*, replicar em sistemas artificiais¹⁸⁸. A pergunta fundamental é: pode uma máquina realmente ‘sentir’ ou ficará limitada a apenas simular a sensação¹⁸⁹?

Num futuro próximo, é improvável que alcancemos uma verdadeira autodeterminação na Inteligência Artificial, dado o nosso conhecimento atual das limitações tecnológicas e da complexidade da consciência humana. No entanto, avanços contínuos na investigação nas áreas de Inteligência Artificial e Neurociência podem eventualmente levar ao desenvolvimento de sis-

¹⁸⁸ “Mind’s actualizations of entangled brain-mind-world states are experienced as qualia and allow ‘seeing’ or ‘perceiving’ of uses of X to accomplish Y. We can and do jury-rig. Computers cannot”, KAUFFMAN, STUART; ROLI, ANDREA (2023). *What is Consciousness? Artificial Intelligence, Real Intelligence, Quantum Mind, and Qualia*. In *Biological Journal of the Linnean Society*, vol. 139, n.º 4, p. 530.

¹⁸⁹ “Can they genuinely feel, and if they can, what are the implications?”, SWAIN, PETER (2024). *The Emotional Enigma: Can Machines truly feel, and does it matter?*. Disponível em <https://peterswain.com/blogs/the-emotional-enigma-can-machines-truly-feel-and-does-it-matter/>

temas mais avançados que exibam níveis superiores de autonomia e adaptação, apenas o futuro o revelará¹⁹⁰. A neurociência, em especial, pode fornecer percepções sobre os mecanismos subjacentes à consciência e cognição, que poderiam, teoricamente, ser aplicados ao desenvolvimento de sistemas de Inteligência Artificial mais sofisticados. A teoria das redes neuronais artificiais, por exemplo, já se inspira em aspectos do funcionamento do cérebro humano, mas ainda estamos longe de replicar a complexidade completa do cérebro.

4.1. Breve Desvio: Teoria das Redes Neuronais

A teoria das redes neuronais artificiais inspira-se nas redes neuronais biológicas e tenta replicar a estrutura e o processo de aprendizagem do cérebro humano, onde neurónios se encontram interconectados e comunicam através de sinapses¹⁹¹. Nas redes

¹⁹⁰ ONETO, LUCA; BUNTE, KERSTIN; NAVARIN, NICOLÓ (2022). *Advances in Artificial Neural Networks, Machine Learning and Computational Intelligence*. In *Neurocomputing*, 342, pp. 300-303.

¹⁹¹ “As sinapses são junções especializadas em que uma célula influencia diretamente outra célula através da

neurônios artificiais, os neurónios são substituídos por unidades matemáticas que recebem sinais de entrada, processam-nos e transmitem-nos à saída para outros neurónios, como é possível observar na figura *infra*. Este processo é iterativo, e as conexões entre os neurónios têm pesos atribuídos a elas, que determinam a força e a influência de uma conexão entre duas unidades. Durante a fase de treino, um algoritmo de aprendizagem ajusta iterativamente os pesos com base nos erros das saídas previstas em relação às saídas reais, através de técnicas que aplicam o conceito de retropropagação. Este ajuste contínuo permite que a rede aprenda e melhore o seu desempenho em tarefas específicas, de maneira semelhante ao processo de aprendizagem do cérebro humano, e também que aprenda a reconhecer padrões e a fazer previsões mais precisas à medida que é exposta a mais dados^{192/193}.

transmissão de um sinal elétrico ou químico”, MOREIRA, CATARINA (2014). *Sinapse*. In Revista de Ciência Elementar, vol. 2, n.º 4, p. 1.

¹⁹² RUMELHART, D.; HINTON, G.; WILLIAMS, R. (1986). *Learning Representations by Back-propagating Errors*. In *Nature*, vol. 323, pp. 533–536.

¹⁹³ BENATAN, M.; PYZER-KNAPP, E. O. (2021). *Deep Learning for Physical*

Neste ano, 2024, o Prémio Nobel da Física distinguiu o físico JOHN HOPFIELD e o cientista GEOFFREY HINTON pelo impacto revolucionário das suas descobertas no desenvolvimento das redes neuronais artificiais. HOPFIELD criou, em 1982, uma rede associativa (*Hopfield network*) desenhada para funcionar como um sistema de memória associativo (*associative memory system*) - inspirado no *spin* atómico¹⁹⁴, onde os *spins* individuais se influenciam mutuamente para formar padrões estáveis, ou domínios - com uma capacidade extraordinária (παράδοξος¹⁹⁵) de recordação, mesmo quando os padrões se encontram incompletos ou apresentam distorções¹⁹⁶. Quando um padrão distorcido ou parcial é introduzido, a rede ajusta-

Scientists: Accelerating Research with Machine Learning. Wiley, p. 74.

¹⁹⁴ DAVOUR, ANNA (2024). *They used Physics to find Patterns in Information*. The Royal Swedish Academy Of Sciences, p. 3.

¹⁹⁵ Cfr., PEREIRA, ISIDRO (1969). *Dicionário Grego-Português e Português-Grego*. (4.ª edição). Livraria Apostolado da Imprensa, p. 273.

¹⁹⁶ HOPFIELD, JOHN; HINTON, GEOFFREY (2024). *The Nobel Prize in Physics 2024*. Press Release, The Royal Swedish Academy of Sciences.

ta iterativamente os estados dos *nodes* (que representam neurónios) para minimizar a energia do sistema, convergindo assim para o padrão armazenado mais próximo. Esta invenção constitui um alicerce na classificação e reconhecimento de padrões, especialmente no âmbito das imagens, um dos pilares das redes neuronais utilizadas atualmente¹⁹⁷.

Por sua vez, GEOFFREY HINTON, tendo por base o trabalho de HOPFIELD, desenvolveu a máquina de Boltzmann (em 1985), uma arquitetura (com raízes na palavra latina *architectūra*, que, por sua vez, tem origens no grego ἀρχιτέκτων) de rede, basea-

da na física estatística¹⁹⁸, capaz de aprender autonomamente e identificar características específicas em conjuntos de dados. Assim, a máquina de Boltzmann é um modelo generativo capaz de descobrir padrões de dados subjacentes através de um método probabilístico. Esta rede incorpora *nodes* visíveis e ocultos, em que a camada visível representa os dados observados e a camada oculta capta as características latentes. Cada configuração de rede possível tem uma probabilidade associada, derivada do seu estado de energia com base na distribuição de Boltzmann. Embora inicialmente lenta, a estrutura da máquina de Boltzmann tornou-se fundamental em *deep*

¹⁹⁷ “This can learn to recognise characteristic elements in a given type of data. Hinton used tools from statistical physics, the science of systems built from many similar components. The machine is trained by feeding it examples that are very likely to arise when the machine is run. The Boltzmann machine can be used to classify images or create new examples of the type of pattern on which it was trained. Hinton has built upon this work, helping initiate the current explosive development of machine learning”, HOPFIELD, JOHN; HINTON, GEOFFREY (2024). *The Nobel Prize in Physics 2024*. Press Release, The Royal Swedish Academy of Sciences.

¹⁹⁸ “Statistical physics describes systems that are composed of many similar elements, such as molecules in a gas. It is difficult, or impossible, to track all the separate molecules in the gas, but it is possible to consider them collectively to determine the gas’ overarching properties like pressure or temperature. There are many potential ways for gas molecules to spread through its volume at individual speeds and still result in the same collective properties”, DAVOUR, ANNA (2024). *They used Physics to find Patterns in Information*. The Royal Swedish Academy Of Sciences, p. 4.

learning, especialmente com a posterior máquina de Boltzmann restrita (RBM) de HINTON, que simplificou a formação restringindo as ligações dentro da rede, permitindo uma pré-treino eficiente para redes com camadas múltiplas.

Apesar das semelhanças estruturais, a capacidade de processamento do cérebro, com mil milhões de neurónios e milhões de biliões de sinapses, oferece um nível de paralelismo e eficiência que as redes neuronais artificiais atuais não conseguem alcançar¹⁹⁹.

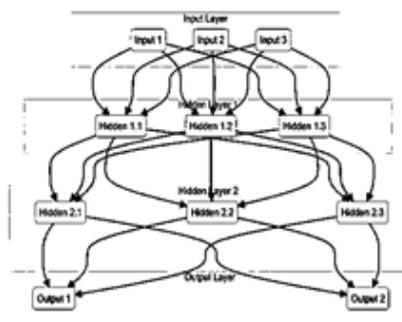


Diagrama não simplificado de uma rede neuronal

¹⁹⁹ NWADIUGWU, MARTINS (2020). *Neural Networks, Artificial Intelligence and the Computational Brain*. University of Ilorin, Nigeria, pp. 5 e 10.

Além disso, o cérebro humano não só processa informação, como também se adapta e evolui constantemente, algo que esta teoria ainda está a explorar através de técnicas como ML e DL²⁰⁰. Na figura *infra* é possível observar a previsão de RAYMOND KURZWEIL, inventor e futurista premiado, acerca do crescimento exponencial da computação²⁰¹. Todavia, avanços contínuos em *hardware*, algoritmos e na qualidade de grandes volumes de dados (*big data*) estão a aproximar-nos cada vez mais de modelos que podem realizar tarefas complexas de forma semelhante à do cérebro humano²⁰².

²⁰⁰ Vide, AA. VV. (2023). *Brain-inspired Learning in Artificial Neural Networks: A Review*. In APL Mach, vol. 2, n.º 2.

²⁰¹ GOLDMAN, BRUCE (2010). *New Imaging Method Developed at Stanford Reveals Stunning Details of Brain Connections*. Stanford Medicine News Center.

²⁰² CIELO, CARLA (2004). *Processamento Cerebral e Conexionismo*. In “Rumo à psicolingüística conexionista”, Porto Alegre, EDIPUCRS, p. 6.

RESPONSABILIDADE CIVIL E INTELIGÊNCIA ARTIFICIAL: QUID IURIS?

SOFIA GARRIAPA²⁰³

SUMÁRIO: 1. A Inteligência Artificial. 2. A Responsabilidade Civil. 3. Conclusões.

RESUMO: Os sistemas dotados de inteligência artificial, com os quais nos cruzamos a toda a hora no nosso dia a dia, trouxeram-nos (e continuarão a trazer) imensas vantagens ao nível do conhecimento e dos avanços científicos nas mais diversas áreas da sociedade, e ainda ganhos de tempo e eficácia na realização de inúmeras atividades. Não obstante, e porque não há bela sem senão, estes sistemas poderão também causar danos significativos às pessoas, quer ao nível da sua integridade física e/ou vida, quer a nível patrimonial. Ora, considerando a autonomia e capacidade de aprendizagem destes sistemas dotados de IA, vários desafios são colocados ao Direito nomeadamente no âmbito da responsabilidade civil. A questão que se coloca, e para a qual urge uma resposta do legislador, será a de saber de que forma será o lesado ressarcido dos prejuízos causados por estes sistemas quando atuem de forma autónoma e para lá do controlo e previsibilidade por parte do ser humano. Procurámos, neste pequeno ensaio e sem quaisquer pretensões de completude, debater algumas soluções que têm vindo a ser equacionadas pela doutrina e pelas instâncias europeias no sentido de procurar compreender qual será o modelo mais adequado a garantir uma justa indemnização aos lesados por estas (novas) tecnologias.

²⁰³ Doutoranda em Direito; Mestre em Direito (2021) e Licenciada em Direito (2018) pela Escola de Direito da Universidade Católica Portuguesa – Centro Regional do Porto; Assistente Convidada na Escola de Direito da Universidade Católica Portuguesa – Centro Regional do Porto.