



Inteligência Artificial (IA): Inquietações Sociais, Propostas Éticas e Orientações Políticas

LIVRO BRANCO

**Conselho Nacional de Ética
para as Ciências da Vida**

Maio de 2024



Conselho
Nacional de
Ética para as
Ciências da Vida



Conselho
Nacional de
Ética para as
Ciências da Vida

© **Conselho Nacional de Ética para as Ciências da Vida, 2024**

Inteligência Artificial (IA): Inquietações Sociais, Propostas Éticas e Orientações Políticas
Livro Branco

Relatores/Membros do Grupo de Trabalho CNECV dedicado à Ética, Inteligência Artificial
e Ciências da Vida: Rui Nunes, Miguel Ricou, Inês Godinho e M. Patrão Neves.

Apoio aos trabalhos: Cíntia Águas e Joana Araújo
Imagem de capa: iStock

Conselho Nacional de Ética para as Ciências da Vida
Avenida D. Carlos I, n.º 134 - 5º
1200-651 LISBOA PORTUGAL
Tel. +351 213 910 884
email: geral@cnecv.pt
www.cnecv.pt





A Inteligência Artificial (IA) está presente nos mais diversos domínios do conhecimento e da atividade humanas, em particular no âmbito das ciências da vida e da saúde, em que tem vindo a ser introduzida a um ritmo acelerado e numa crescente diversidade de áreas de atuação, especialidades médicas e atos clínicos.

Ciente da premência desta reflexão, o Conselho Nacional de Ética para as Ciências da Vida (CNECV), no cumprimento da sua missão de acompanhamento e deliberação sobre os problemas éticos suscitados pelos progressos científicos e os avanços tecnológicos, e sobretudo na prossecução da sua missão de contribuir para a qualificação das políticas públicas e para a formação da sociedade em geral em matéria de implicações éticas dos progressos científico-tecnológicos, entendeu constituir um Grupo de Trabalho dedicado a esta problemática específica.

O presente Livro Branco é um documento aberto e dinâmico que, como tal, não foi sujeito a votação, resultado da reflexão do Grupo de Trabalho em matéria de IA e do Plenário, tendo ainda beneficiado dos contributos de especialistas em sede de audições públicas e das sugestões dos cidadãos na discussão das matérias que precederam a sua elaboração. A presente iniciativa marca a permanente abertura do CNECV à sociedade, a qual, fruto de um extenso calendário de trabalhos, enriqueceu a visão prospetiva que agora se apresenta ao Legislador.



Conselho
Nacional de
Ética para as
Ciências da Vida



INTELIGÊNCIA ARTIFICIAL: INQUIETAÇÕES SOCIAIS, PROPOSTAS ÉTICAS E ORIENTAÇÕES POLÍTICAS

Livro Branco

Índice

1. A questão terminológica.....	6
2. O percurso histórico	7
<i>Da pré-história da IA ...</i>	7
<i>... à história da IA.....</i>	8
<i>Do segundo impulso... ..</i>	9
<i>...à presente dinâmica</i>	11
<i>A IA generativa.....</i>	11
3. A diversidade de aplicações da IA.....	13
4. Regulação ético-normativa da IA	16
4.1. Regulação normativa mundial	16
4.1.1. <i>Recomendação da UNESCO sobre a Ética da Inteligência Artificial.....</i>	16
4.1.2. <i>Guia da Organização Mundial de Saúde sobre Ética e Governação da IA para a Saúde</i>	17
4.2. Regulação normativa europeia.....	19
4.2.1. <i>Guia de Orientações Éticas para uma IA de Confiança do Grupo de Peritos de Alto Nível sobre a Inteligência Artificial.....</i>	19
4.2.2. <i>O Regulamento da IA da UE (AI Act)</i>	20
4.3. Instrumentos nacionais	22
<i>Ambiente controlado (Regulatory sandbox).....</i>	23
5. A aplicação da IA no plano das ciências da vida e seus desafios éticos.....	23
5.1. Investigação biomédica e (o impacto do) reconhecimento de padrões.....	24
5.2. Assistência clínica e (o impacto da) assistência digital	31
5.3. Administração da saúde, gestão hospitalar e (o impacto da) intervenção à distância (telessaúde).....	40
5.4. Administração da saúde pública e (o impacto da) codificação de dados	47
5.5. Ensino e educação em saúde e (o impacto da) realidade virtual.....	51
6. Visão prospetiva.....	57
Audições.....	61

* Relatores/Membros do Grupo de Trabalho: Rui Nunes, Miguel Ricou, Inês Godinho e M. Patrão Neves.



Inteligência Artificial (IA): inquietações sociais, propostas éticas e orientações políticas Livro Branco

1. A questão terminológica

A expressão “inteligência artificial” data de 1956, tendo sido forjada e introduzida no domínio então nascente da computação pelos cientistas norte-americanos Marvin Minsky, John McCarthy, Nathaniel Rochester e Claude Shannon aquando da organização da *Dartmouth Summer Research Project on Artificial Intelligence*, em New Hampshire, nos Estados Unidos da América. Ela reflete, de forma objetiva e clara, o estado da arte do desenvolvimento da automação então em debate e centrado na possibilidade de criar máquinas, computadores capazes de simularem (informaticamente) as características da inteligência humana (as suas faculdades cognitivas), nomeadamente a aprendizagem e o uso de linguagem, e de mimetizarem desempenhos cognitivos humanos¹, numa aceção cognitivista do cérebro humano (na esteira da conceção mecanicista do ser humano). A expressão “inteligência artificial” é, ela própria, mimética.

Poucos anos mais tarde, em 1959, Minsky e McCarthy fundam o *Artificial Intelligence Laboratory* (atual *MIT Computer Science and Artificial Intelligence Laboratory*)² o que contribuiu fortemente para a disseminação da expressão “inteligência artificial”.

Entretanto, e à medida que a noção de “inteligência artificial” se vai popularizando, sobretudo para além do meio científico e académico, vão surgindo cada vez mais vozes críticas, acusando a inadequação do termo para designar a realidade em causa.

Os argumentos contra a utilização do conceito de “inteligência artificial” são diversos. Um primeiro, tematicamente mais distante, poder-se-á reportar à dificuldade de definição da própria inteligência humana³, verificada desde o século XIX, em que se torna objeto de estudo, assim se comprometendo a conceção mimética da expressão.

¹ No convite para o Dartmouth Summer Project lê-se: “*We propose that a 2-month, 10-man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.*”, J. McCarthy, M. L. Minsky, C.E. Shannon, N. Rochester, “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence”, 31 August, 1955, (acedido em Agosto de 2023) <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>

² O *AI Lab* foi criado, em 1959, como *AI Project*. Mais tarde, em 1963, foi criado o *Laboratory for Computer Science (LCS)* como *Project MAC (Multiple Access Computing or Machine-Aided Cognition)*, estando ambos na origem do atual *MIT Computer Science and Artificial Intelligence Laboratory (CSAIL)*. A fusão entre os dois laboratórios deu-se apenas em 2003.

³ «Inteligência», etimologicamente, significa «capacidade de estabelecer relações entre coisas». Porém, não há uma definição única e consensual, a qual está fortemente dependente do domínio científico que a aborda (filosofia, psicologia, neurociência, etc.) e, dentro deste, da escola de pensamento que a estuda.



Na esteira deste, e já beneficiando de desenvolvimentos científicos da segunda metade do século XX, ergue-se o de o ser humano manifestar uma multiplicidade de inteligências, assim se rejeitando toda a perspetiva monolítica de “inteligência”. E, todavia, a informática parece reduzi-la a apenas uma vertente, a lógico-matemática.

Ainda mais determinantes são os argumentos que denunciam o conceito como equívoco, por este sugerir que os algoritmos podem tomar decisões conscientes e racionais à semelhança dos seres humanos, o que consideram não se verificar no presente ou no futuro⁴. De facto, para além do que possa ser classificado como uma questão terminológica, a sanar no contexto científico ao nível das ciências da computação e no sentido de uma crescente objetividade e rigor, a definição de “inteligência artificial” faz emergir também uma problemática societal, na medida em que induz perceções por sua vez condicionantes de abordagens e comportamentos em relação a esta tecnologia. As definições não são axiologicamente neutras e, neste caso particular, podem influenciar o sentido de desenvolvimento.

Não existindo uma definição consensual de “inteligência artificial”, mas considerando-se útil apresentar uma proposta abrangente para a mesma, adotamos a do *High-Level Expert Group on Artificial Intelligence* (instituído pela Comissão Europeia em 2018) pelo imediato e fácil acesso que proporciona aos conteúdos a que nos reportamos: “Inteligência artificial (IA) refere-se a sistemas que apresentam comportamento inteligente, analisando o seu ambiente e decidindo ações - com algum grau de autonomia - para atingir objetivos específicos. Os sistemas de IA podem ser puramente baseados em *software*, atuando no mundo virtual (por exemplo, assistentes de voz, *software* de análise de imagem, motores de busca, sistemas de reconhecimento de fala e rosto) ou podem ser incorporados em dispositivos de *hardware* (por exemplo, robôs avançados, carros autónomos, drones ou Internet das Coisas).”⁵

2. O percurso histórico

Da pré-história da IA ...

Pode-se fazer recuar a origem da Inteligência Artificial (IA), a sua pré-história, muito remotamente, à invenção da primeira máquina de calcular, por Wilhelm Schickard, em 1623, ou, preferencialmente, numa mais direta proximidade, a 1936 e à invenção do

⁴ Luc Julia, engenheiro informático cocriador da Siri, defende esta mesma ideia na sua obra *L'Intelligence artificielle n'existe pas*, Paris, Ed. First, 2019, 200 pp. Vale a pena referir também, entre muitos outros críticos, o engenheiro e especialista em cibernética Nathanaël Leroy, que reputa a expressão “inteligência artificial” de oximoro, na medida em que a máquina é apenas simulacro, reflexo da inteligência humana que a criou.

⁵ European Commission. High-Level Expert Group on Artificial Intelligence. (2019) *Ethics Guidelines for Trustworthy AI*.



primeiro modelo de computador capaz de armazenar e processar informação, pelo matemático e cientista da computação britânico Alan Turing, muitas vezes identificado como “o pai da IA”. Com efeito, a IA surge quase simultaneamente com a informática. Desencadeia-se, assim, a designada era computacional. Esta beneficia também do contexto favorável do pós-segunda guerra mundial, com a evolução tecnológica que o caracteriza, e ainda do desenvolvimento da cibernética, popularizada pelo trabalho do matemático norte-americano Norbert Wiener que a define, em 1948, numa publicação com o mesmo nome, como “o estudo científico do controlo e comunicação no animal e na máquina”. Esta ciência dos sistemas autorregulados corresponderia, para Wiener, a uma tentativa de unificação dos domínios então emergentes do automático.

As máquinas pareciam então poder ser inteligentes, tal como Turing perguntava em *Computing Machinery and Intelligence*, de 1950: poderão as máquinas pensar? Ou, dito de outra forma, um computador comportar-se de forma inteligente, como um ser humano?⁶.

Um primeiro ensaio de resposta é dado através do “teste de Turing”, que se popularizou como o “jogo da imitação”: será possível um interrogador distinguir as respostas dadas por um computador das respostas dadas por um ser humano? Ou seja, poderão as máquinas passar por humanos, ou imitar os humanos? A resposta de Turing é, já então, resolutamente positiva num horizonte calculado de 50 anos.

... à história da IA

Estamos, pois, no âmbito da tentativa de simulação ou imitação computacional do raciocínio humano, no domínio originário do que Minsky e McCarthy vêm a designar por “inteligência artificial”, assinalando assim também o início da sua história. Esta é uma IA classificada como “simbólica” na medida em que traduz conhecimentos por símbolos e raciocínios por proposições lógicas. Curiosamente, é ainda no texto precursor de 1950 que Turing vem propor que a estratégia a seguir não deveria ser, como até então, a de tentar “escrever um programa que permitisse a uma máquina passar o jogo da imitação” (reproduzindo partes do raciocínio humano), mas antes a de escrever “um programa que permitisse a uma máquina aprender com a experiência, tal como acontece com um bebé”⁷. Denota, assim, uma posição conexionista, caracterizada pelo empenho de reprodução do funcionamento do cérebro humano pela máquina; contrariamente ao que se verifica na perspetiva cognitivista, que surgiu sensivelmente ao mesmo tempo, e

⁶ Alan Turing. (1950) *Computing Machinery and Intelligence*, *Mind* 59 (236): 433-460.

⁷ Alan Turing, *op. cit.*



se caracteriza pela tentativa de recriar mecanicamente as leis do pensamento humano, através de uma manipulação de símbolos.

É neste sentido, no da aprendizagem automática ou "*machine learning*" (capacidade de os algoritmos aprenderem através da sua análise de dados, de se modificarem sem serem programados para tal, numa lógica ou análise preditiva), que a IA evolui nas décadas que se seguem através do desenvolvimento dos computadores, em capacidade de armazenamento e velocidade de processamento, e do aperfeiçoamento dos algoritmos na sua aplicação específica a problemas particulares.

A face mais visível, para o cidadão comum, do progresso desta etapa da aprendizagem automática consiste no aperfeiçoamento de jogos de computador: o primeiro - um jogo de damas chinesas - foi programado por Arthur Samuel, em 1952, sendo capaz de melhorar o seu desempenho a cada novo jogo⁸. Entretanto, o inicial extraordinário progresso da IA⁹, com o conseqüente forte investimento que vinha suscitando, foi esmorecendo, principalmente a partir de 1974, data que assinala o início do designado "primeiro inverno da IA", o qual se prolonga até cerca de 1980. Este período é desencadeado pelo crescimento excessivo de expectativas (irrealistas) e caracterizado por frustrações, críticas, pessimismo e cortes no financiamento, tendo sido sobretudo determinado pela reduzida capacidade de memória e baixa velocidade de processamento dos computadores de então para permitirem uma maior complexidade e mais ousado poder de realização da IA.

Do segundo impulso...

Um segundo momento de mais forte desenvolvimento da IA inicia-se principalmente a partir da década de 80 do século passado, sobretudo devido a dois fatores determinantes.

⁸ O confronto direto entre o raciocínio humano e a capacidade da máquina, em diversos jogos, é característico desta fase mas vai-se prolongando até ao presente, sendo que o sucesso da máquina sobre o humano se verifica já numa fase bem mais avançada, destacando-se, no xadrez, em 1996, o Deep Blue, um software da IBM que vence o campeão mundial, Kasparov; do jogo Go, em 2017, o AlphaGo ganhou contra o melhor do mundo; de poker, em 2019, o Pluribus venceu uma maratona de 12 dias, competindo com 5 jogadores.

⁹ Destacam-se alguns progressos revolucionários para o desenvolvimento da IA como sejam: em 1957, Frank Rosenblatt constrói o primeiro computador (o Perceptron) para criar redes neuronais, capaz de reconhecimento de padrões; em 1959, J. McCarthy inventa uma linguagem de programação funcional, Lisp; em 1963, Donald Michie constrói um dos primeiros programas (Menace) capaz de aprender a jogar Tic-Tac-Toe (o "jogo do galo"); em 1967, é introduzido o algoritmo de aprendizagem supervisionada KNN (K-Nearest Neighbors), básico da *machine learning*; em 1970, Seppo Linnainmaa apresenta o modelo de diferenciação automática, usado para treinar redes neuronais artificiais o qual, aliás, será importante para quase todos os algoritmos de otimização de *deep learning*.



Primeiramente, o da criação de computadores de quinta geração, com a sua simplificação e miniaturização, e também o aparecimento de computadores pessoais e a multiplicação de utilizadores e de cocriadores do progresso da IA, e ainda o aumento da capacidade de processamento e de armazenamento dos computadores, do seu melhor desempenho e maior flexibilidade. Um outro fator decisivo foi o da invenção, por Edward Feigenbaum, de *software* especializado na realização de tarefas cada vez mais complexas e específicas a determinados domínios de atividade humana, tornando-se competentes para as substituir com vantagem.

Este impulso da IA é marcado pelas técnicas de “*deep learning*” (uma variante do método de análise da “*machine learning*”, utilizando redes neuronais, uma classe específica de algoritmos) que permitem que os computadores aprendam autonomamente, a partir da sua própria experiência, do reconhecimento de padrões (fala, objetos, imagens), sem qualquer intervenção humana¹⁰.

A este segundo momento de expansão seguiu-se um período de declínio, um novo “inverno” que ocorreu, muito brevemente, na primeira metade dos anos 90, ditado então por investimentos excessivos, sobrevalorizando, por exemplo, o poder das unidades centrais de processamento.

Rapidamente, na segunda metade dos anos 90, desencadeia-se a Internet comercial e a IA foi cada vez mais intervindo em todas as áreas da atividade humana, tornando-se uma realidade de utilização quotidiana do cidadão comum o qual, todavia, nem sempre tem consciência de ser utilizador da IA.

¹⁰ Ao mesmo tempo que o método de *machine learning* se continua a desenvolver – por exemplo, em 1981, Gerald Dejong introduz o “*Explanation Based Learning*” (EBL), capaz de formar conceitos a partir da experiências, que contribuirá para a aprendizagem supervisionada e, em 1990, Kearns and Vaient introduzem um meta-algoritmo que reduz o preconceito na aprendizagem – destacam-se alguns progressos revolucionários para o desenvolvimento da IA como sejam: em 1979, Hans Motavec cria o primeiro veículo autónomo controlado por um computador e sem intervenção humana, o Stanford Cart; em 1985, Terry Sejnowski inventa uma rede neuronal capaz de pronunciar palavras à semelhança do que as crianças fazem, o NETtalk e, em 1997, Jürgen Schmidhuber e Sepp Hochreiter, através da técnica de *deep learning* designada por LSTM, criam a capacidade de reconhecimento do discurso, que veio a evoluir para as atuais aplicações comuns da Alexa (Amazon), Siri (Apple), Google Translator, entre outras. Vale a pena referir que o primeiro *chatbot*, Eliza, foi criado entre 1964-1966.



...à presente dinâmica

Desde então, e ininterruptamente, tem-se verificado um aumento da capacidade dos computadores e do volume dos dados gerados e disponíveis, aliado ao aumento da capacidade de os processar, e de um grande investimento financeiro, o que resulta em um conseqüente aumento do ritmo de desenvolvimento da IA numa lógica geométrica (determinada através da multiplicação do termo anterior, o que resultou num novo impulso a cada 18 meses) e já não aritmética (determinada através de uma soma constante, o que resultava num novo impulso a cada década).

Esta realidade, aliada ao processo paralelo de independência da intervenção e do controlo humanos, permite-nos apelidar o momento presente de “época de ouro” da IA. Simultaneamente, evidencia-se um aumento das questões éticas que a IA suscita e da sua complexidade, sendo que os desafios prioritários que coloca vão-se metamorfoseando nos seus contornos, ao ritmo das inovações tecnológicas.

A IA generativa

Talvez uma das mais impactantes inovações recentes no domínio da IA seja a da estruturação e divulgação da “inteligência artificial generativa”¹¹, assim designada por se tratar de um sistema automatizado que utiliza algoritmos para produzir, manipular ou sintetizar dados, frequentemente sob a forma de textos e imagens a partir de solicitação humana em linguagem comum.

O *ChatGPT (Generative Pre-trained Transformer)* é certamente o mais conhecido ou, na sua versão mais avançada, dinâmica e personalizada, o “agente conversacional” hoje mais popularizado, decorrente da sua disponibilização gratuita a todos os interessados, no que constitui também um marco social revolucionário na história da evolução da IA.

Trata-se de uma aplicação da IA capaz de responder a questões das mais diversas áreas, gerando textos com a aparência de terem sido escritos por humanos, isto é, sintaticamente corretos e semanticamente plausíveis, obedecendo ao critério essencial da coerência. As suas afirmações são, todavia, de ordem probabilística, dependentes dos dados disponíveis sobre a matéria em causa e podendo reproduzir preconceitos associados, além de poderem igualmente sofrer de “alucinações”, ou seja, apresentarem afirmações verosímeis, mas não verificáveis ou mesmo falsas - apesar de as versões mais recentes do *ChatGPT* reduzirem consideravelmente este risco.

¹¹ O que pode ser classificado como IA generativa terá a sua origem em 1964-1966 e na produção do *chatbot* Eliza, o primeiro software simulador de diálogos. Mas foi o desenvolvimento da capacidade de aprendizagem dos computadores que permitiu o lançamento dos mais recentes e revolucionários sistemas generativos de IA.



Algumas das limitações e erros que têm sido identificados têm também sido ultrapassados em novas versões, tendo recentemente sido lançado o *ChatGPT 4*, com novas potencialidades, especialmente no campo da imagem, já não detetáveis como sendo artificialmente produzidas.

O Hub France IA destaca quatro grandes tipos de capacidades da IA Generativa de forte impacto: a) a de redação (incluindo códigos informáticos), b) a de classificação (nomeadamente a triagem de documentos), c) a de tradução (escrita ou oral) e d) a de síntese (por exemplo, para tratamento de dados)¹². Porém, também a capacidade de criação – intelectual e artística, entre outras – é hoje a marca de água das novas versões de IA generativa, característica que torna o *ChatGPT* e programas similares tão apetecíveis nos mais diferentes domínios do conhecimento.

De facto, a designação de “IA generativa” anuncia a capacidade de o *software* criar o que não existia anteriormente, o que não é consensual, podendo-se afirmar que o sistema generativo processa dados já existentes, sintetizando-os. Poderíamos encetar um debate acerca da noção de “criação”, se esta se reporta simplesmente a uma nova combinação de elementos existentes ou se o conceito remete para um absoluto, uma designada “criação *ex nihilo*” apenas atribuída a Deus. Neste último caso, o termo “criador” não se aplicaria ao humano; no primeiro, também se aplicaria à IA generativa.

A questão não é de menor interesse, uma vez que corresponde à avaliação do progresso da IA como “quantitativo”, num desenvolvimento de si gradativamente superior, resultando na expansão da aplicação de um mesmo tipo de *software*, ou “qualitativo”, e no acesso a novas realidades de natureza distinta, o que assinala a capacidade de ultrapassar supostos limites, abrindo-se um horizonte infinito de realizações.

Utilizando uma terminologia comum ao domínio em causa, perguntar-se-ia se a IA generativa se inscreve ainda num desenvolvimento superior da “IA fraca” ou “estreita” ou se prenuncia já uma “IA forte” ou “generalista” capaz, não só de executar a tarefa para a qual foi desenhada, mas também de compreender, aprender e aplicar as suas capacidades a diferentes tarefas¹³. O questionamento ético é, ele próprio, necessariamente diferente, na medida em que, na subordinação da iniciativa tecnológica às finalidades humanas, o controlo humano mantém-se possível num progresso quantitativo da IA, mas tende a dissipar-se num progresso qualitativo.

Em qualquer caso, o surgimento do *ChatGPT* assinala uma aceleração significativa da IA, tendo já sido considerado como inaugurando uma nova era numérica.

¹² Hub France IA, «ChatGPT: Usages, Impacts et Recommendations – Note de Synthèse» (Mai, 2023) https://www.hub-franceia.fr/wp-content/uploads/2023/04/ChatGPT_Note-synthese.pdf

¹³ Patrão Neves, M., Almeida, B. (2024). Before and Beyond Artificial Intelligence: Opportunities and Challenges. In *Multidisciplinary perspectives on Artificial Intelligence and the law*. essay, Cham: Springer International Publishing AG. <https://doi.org/10.1007/978-3-031-41264-6>



3. A diversidade de aplicações da IA

O quase ininterrupto progresso da IA ao longo da história recente da humanidade reflete-se hoje numa contínua expansão aos múltiplos domínios da atividade humana em que se vai também enraizando, alterando sistemas de organização, modalidades de ação e comportamentos, à medida que se vai tornando omnipresente no quotidiano do cidadão comum. Ao introduzir-se paulatinamente nas mais diversas dimensões das sociedades atuais, a IA tem vindo a reproduzir diferentes funções tradicionalmente realizadas por humanos de forma mais rápida, mais rigorosa, mais económica, isto é, de forma mais eficiente, enformando hoje o nosso quotidiano individual e coletivo, tanto nos processos de funcionamento, como na estrutura de desenvolvimento.

Talvez o domínio em que primeira e mais decisivamente a IA se tornou preponderante tenha sido o setor industrial. Referimo-nos, por exemplo, a linhas de produção e de montagem automáticas, o que aumentou exponencialmente a produtividade e reduziu significativamente o número de trabalhadores necessários. Gerou-se, então, a percepção de que a substituição do humano pela IA o libertaria de tarefas rotinizadas e fastidiosas, possibilitando que se dedicasse preferencialmente a outras atividades, mais exigentes e criativas, ou ainda que aumentasse o seu tempo de lazer – uma perspetiva eminentemente positiva que constitui também um incentivo para o desenvolvimento da IA.

O que, entretanto, se tem vindo a verificar, é que a IA não só evidencia uma superior proficiência na execução de tarefas mecânicas, mas vem ganhando também uma capacidade superior de desempenho em atividades cognitivas exigentes e, cada vez mais, igualmente em ações criativas – até então consideradas um reduto identitário do humano e, como tal, inacessível externamente. Assim, hoje, a IA enforma diversos domínios, tais como:

- a) Justiça, (e.g. pesquisa de jurisprudência, adoção de medidas de justiça com base em casos anteriores semelhantes, existindo já projetos de instituição de um tribunal automático de justiça preditiva para decidir casos mais simples),
- b) Saúde (e.g. há intervenções em algumas especialidades médicas que podem evoluir significativamente, por exemplo na radiologia ou na oftalmologia, pelas vantagens que a IA pode apresentar em relação aos médicos na elaboração do diagnóstico e da intervenção terapêutica),
- c) Ciência (e.g. na otimização do desenho, recrutamento e duração dos ensaios clínicos, na identificação de novas moléculas, na análise e correlação de dados),
- d) Ensino (e.g. a experiência de assistentes universitários digitais tem sido muito positiva na rapidez e mesmo qualidade de resposta aos alunos),
- e) Agricultura e alimentação (e.g. robôs que removem ervas daninhas, reduzindo o uso de herbicidas, monitorização da temperatura ou da rega, aumentando a produtividade e reduzindo o impacto ambiental),
- f) Transportes (e.g. hoje decisiva na elaboração dos horários dos transportes e escolha de rotas, bem como na sincronização dos semáforos),



- g) Finança (e.g. utilização crescente nos mercados financeiros para a avaliação de risco ou para a deteção de fraudes),
- h) Marketing (e.g. na criação e, sobretudo, na personalização de conteúdos, na análise preditiva, na publicidade programática),
- i) Segurança informática (e.g. contribui fortemente para a cibersegurança),
- j) Defesa e segurança nacional (e.g. na marcação de alvos por satélites),
- k) Lazer e diversão (e.g. a aplicação de realidade aumentada em espetáculos, eventos desportivos ou culturais, bem como em turismo virtual),
- l) Artes e setor criativo (e.g. na produção de música, livros, *design* ou pinturas).

Em todas estas e outras áreas de domínio humano, a IA tem-se vindo a impor pelas suas superiores capacidades funcionais, isto é, alargando o seu espaço de atuação e reduzindo a necessidade de intervenção humana, ao abrigo do valor da eficiência, num plano de assistência e colaboração com as finalidades humanas. O mesmo se confirma nos pequenos gestos do quotidiano que a IA vem enquadrando, numa pluralidade de expressões tendencialmente infinita, sem que o cidadão comum o consciencialize, como seja nas compras *online*, mas sobretudo na publicidade, nas pesquisas na *web*, mas cada vez mais na organização das agendas e rotinas diárias e, transversalmente, na relação com a administração pública e serviços.

Neste contexto, a IA mantém-se uma designada “ferramenta”, na terminologia dos seus criadores e impulsionadores, isto é, um objeto ou meio, um “utensílio” que prolonga a ação humana, que projeta a capacidade de realização das finalidades humanas, muito para além do que lhes seria naturalmente possível. Toda a “ferramenta”, por natureza inerte, possui um valor meramente instrumental, funcional, permanecendo a agência no humano, pelo que se apresenta como axiologicamente neutra. E, não obstante, mesmo a este nível, importa apontar como o progresso da IA, em contínua aceleração, tende a eliminar as alternativas tradicionais de ação humana, reduzindo a liberdade da pessoa e alienando a sua capacidade de decisão e de relação, ou seja, passando sub-repticiamente a ter uma intervenção efetiva nos comportamentos individuais e nas relações sociais. A digitalização da sociedade em curso impõe uma mediação tecnológica às relações humanas a qual, por sua vez, acentua dependências e agrava vulnerabilidades, pela própria ausência de meios alternativos para a prossecução de finalidades diversas.

Vamos assistindo assim, também, a uma transmutação de valores – ela própria justificadora da mudança, qual círculo vicioso –, em que valores identitários humanos, como o respeito pela dignidade no seio das mais profundas vulnerabilidades ou o empenho na solidariedade, contrariando a pluralidade das injustiças globais, sucumbem perante o valor da eficiência, avaliada sob a tirania do mensurável em termos de tempo e eficácia, traduzida em termos de produtividade e de rentabilidade. Neste contexto progressivamente funcionalizador do humano, em que a proficiência do desempenho ganha preponderância, também a eficiência poderá passar de um valor instrumental que hoje possui, isto é, dependente de finalidades que a ultrapassam, para um valor em si



mesmo por que se pautariam as relações humanas, convertendo a realização de si, ou personalização, num desempenho social, ou operacionalização.

Neste horizonte, também se perfila, com uma assertividade crescente, uma alteração do estatuto da própria IA que, de ferramenta axiologicamente neutra - inerte e passiva, em que a agência reside apenas no utilizador -, se converte numa dinâmica extrínseca ao humano, com uma autonomia crescente. Transformando-se, progressivamente, de mero instrumento a protoagente.

Os sistemas de inteligência artificial são originariamente desenhados para que disponham dos elementos necessários para a sua própria evolução, numa dinâmica interna de aprendizagem a partir do seu próprio exercício, dispensando cada vez mais a intervenção humana, e sempre em nome da eficiência. A automação progride no sentido da autonomização. E este processo condiciona e induz mesmo comportamentos humanos, que se têm agora de enquadrar nas possibilidades que os sistemas inteligentes disponibilizam e que assim se tornam também agentes de mudança. A IA já não depende apenas do utilizador e vem se libertando do seu criador, progressivamente escapando ao controlo humano, num paralelo adensar da opacidade dos seus processos¹⁴.

Considera-se assim que, qualquer que seja o desenvolvimento presente e futuro da IA e qualquer que seja o seu domínio de aplicação, se impõem pelo menos dois imperativos éticos ao seu progresso:

- Alicerçar-se sempre em princípios éticos identitários do humano, isto é, no mais estrito respeito pelos seus valores nucleares e direitos fundamentais;
- Centrar-se sempre no humano (uma lógica *human-centered design* em oposição à *user-centered design*), isto é, rejeitando o pragmatismo da captura por interesses sectários e/ou particulares, mantendo-a como instrumento de realização do humano, individual e socialmente considerado.

¹⁴ M. Patrão Neves (2020) "As Novas Tecnologias e os Desafios do Futuro: uma perspectiva ética", Tiago Brandão, Maria Eduarda Gonçalves [coord.], e-Book: *Ensaio sobre Ciência, Cultura e Política Científica*, Lisboa, Centro Nacional de Cultura, 2020, p. 100.



4. Regulação ético-normativa da IA

A aplicação de IA, nomeadamente à saúde, conta já com alguns instrumentos de regulação gerais e específicos no plano internacional e, especialmente, no plano europeu. Nem todos os instrumentos apresentam o mesmo valor vinculativo, ainda que todos se devam considerar relevantes para a compreensão da regulação ético-normativa da IA.

4.1. Regulação normativa mundial

No plano internacional são de mencionar, fundamentalmente, dois instrumentos de *soft law*, isto é, não vinculativos para os Estados, mas que assumem um peso normativo significativo pelo impacto dos valores éticos subjacentes.

4.1.1. Recomendação da UNESCO sobre a Ética da Inteligência Artificial¹⁵

Esta Recomendação, podendo representar o ponto incontornável da regulação da IA, assenta na ideia da ética como base dinâmica para a avaliação normativa e como barómetro das tecnologias de IA, com referência à dignidade humana, ao bem-estar e à prevenção de danos como compasso e com enraizamento na ética da ciência e tecnologia (I (1)). Com efeito, reconhece que os sistemas de IA colocam novos tipos de questões éticas, que vão desde o seu impacto na tomada de decisões, aos cuidados de saúde, à democracia e aos direitos e liberdades fundamentais (I (2) (c)). Assim, os diferentes valores e princípios indicados na Recomendação deverão ser respeitados por todos os agentes do ciclo de vida da IA e, quando necessário, promovidos pela alteração ou criação de nova legislação e diretivas. Estes instrumentos deverão, ademais, cumprir as obrigações decorrentes para os Estados-Membros em matéria de direitos humanos e estar de acordo com os Objetivos de Desenvolvimento Sustentável (ODS) definidos pela ONU¹⁶ (III (9)).

Neste contexto, a Recomendação apresenta quatro valores e dez princípios fundamentais para a IA. Os valores indicados (III.1) são:

- 1) Respeito, proteção e promoção dos direitos humanos e liberdades fundamentais e dignidade humana;
- 2) Promoção do ambiente e dos ecossistemas;
- 3) Garantia de diversidade e inclusão;
- 4) Viver em sociedades pacíficas, justas e interligadas.

¹⁵ O texto da Recomendação encontra-se disponível em:
<https://unesdoc.unesco.org/ark:/48223/pf0000381137>.

¹⁶ Cf. <https://unric.org/pt/objetivos-de-desenvolvimento-sustentavel/>.



Os princípios, traduzindo uma abordagem baseada nos direitos humanos (III.2) para a IA são:

- 1) proporcionalidade e não causar dano – o uso de sistemas de IA não deve ir além do necessário para alcançar um fim legítimo. A avaliação de risco deve ser usada para prevenir danos que possam ocorrer;
- 2) segurança e proteção – os riscos de segurança internos (dos sistemas) e externos (de ataque) devem ser evitados e abordados pelos atores da IA (programadores, fabricantes, utilizadores);
- 3) justiça (equidade) e não-discriminação – os atores de IA devem promover a justiça social, a equidade e a não-discriminação, adotando uma abordagem inclusiva que garanta que os benefícios da IA são acessíveis a todos;
- 4) sustentabilidade – as tecnologias de IA devem ser avaliadas quanto aos seus impactos em termos de sustentabilidade, entendida como um conjunto evolutivo de objetivos;
- 5) direito à privacidade e proteção de dados – a privacidade deve ser protegida e promovida durante o ciclo de vida da IA, devendo ser estabelecidos quadros de proteção de dados adequados;
- 6) supervisão humana e determinação – os Estados-membros devem garantir que os sistemas de IA não deslocam a responsabilidade humana;
- 7) transparência e explicabilidade – a colocação (no mercado) ética de sistemas de IA depende da sua transparência e explicabilidade (T&E). O patamar de T&E deve ser apropriado ao contexto, considerando eventuais tensões com outros princípios, como o da privacidade ou o da segurança;
- 8) responsabilidade e responsabilização – os sistemas de IA devem ser auditáveis e rastreáveis, devendo ser implementados mecanismos de supervisão, avaliação de impacto e auditoria, a fim de evitar conflitos com normas de direitos humanos e ameaças ao bem-estar ambiental;
- 9) consciência e literacia – o conhecimento e a compreensão pelo público de IA e de dados devem ser promovidos através de educação, empenho cívico, capacitação digital e formação ética de IA; e
- 10) governação e colaboração adaptativas e multilaterais – o direito internacional e as soberanias nacionais devem ser respeitados na utilização de dados, devendo existir a participação dos diferentes interessados para uma abordagem inclusiva à governação de IA.

4.1.2. Guia da Organização Mundial de Saúde sobre Ética e Governação da IA para a Saúde

No patamar específico dos cuidados de saúde deve ainda ter-se em consideração o Guia da Organização Mundial de Saúde sobre Ética e Governação da IA para a Saúde¹⁷.

¹⁷ Guia disponível em: <https://www.who.int/publications/i/item/9789240029200>.



Reconhecendo que a IA poderá melhorar as decisões tomadas pelos clínicos, especialmente em situações com recursos limitados em que se poderá revelar ser uma importante ferramenta, aqui se referem também os principais desafios éticos da utilização da IA para a prestação de cuidados de saúde, entre os quais a necessidade de aferir se a IA deve ser usada, o viés associado com a IA e a responsabilidade pela tomada de decisão com IA. Face aos desafios existentes e emergentes, os princípios éticos tornam-se, assim, como faróis para a orientação e o desenvolvimento destas tecnologias. O valor axial na conceção dos princípios é o da dignidade humana e do valor imanente do humano. Tratando-se de princípios éticos, configuram um dever ou responsabilidade no contexto do desenvolvimento, colocação e avaliação contínua de tecnologias e sistemas de IA para a saúde, estando enquadrados nos pilares da bioética: não-maleficência, beneficência, justiça e respeito pela autonomia.

Nesse contexto, o Guia da OMS estabelece seis princípios-chave para a utilização da IA na saúde¹⁸:

1) A proteção da autonomia – o princípio da autonomia requer que qualquer extensão de autonomia dos sistemas ou tecnologias de IA não diminua ou afete negativamente a autonomia humana, o que significa que os humanos devem permanecer no controlo total dos sistemas de saúde e das decisões médicas. De outra forma, exige-se permanente supervisão humana e adaptabilidade de decisões e a plena garantia do consentimento informado;

2) A promoção do bem-estar humano, da segurança humana e do interesse público – estando em causa o princípio da não-maleficência, exige-se o cumprimento de regras de segurança, exatidão e eficácia antes da colocação destas tecnologias no mercado, devendo ser implementadas medidas para garantir o controlo e melhoria de qualidade. Além disso, o uso das tecnologias de IA não pode resultar em nenhum dano mental ou físico;

3) A garantia de transparência, explicabilidade e inteligibilidade – a transparência exige a publicação ou documentação de informação suficiente antes do desenho e da colocação de tecnologias de IA e mesmo depois da sua aprovação para utilização. Estas tecnologias deverão ainda ser suscetíveis de explicação de acordo com a capacidade de quem é destinatário da explicação, ainda que possa existir uma troca (*trade-off*) entre a plena explicabilidade de um algoritmo (à custa da exatidão) e maior rigor (à custa da explicabilidade);

4) A promoção da responsabilidade e da responsabilização – a responsabilidade pode ser assegurada pela aplicação da “garantia humana”, *i.e.*, a avaliação, por médicos e pessoa pessoas doentes, do desenvolvimento de tecnologias de IA, através de uma supervisão com mecanismos de governação de intervenção humana (HITL). Quando existam decisões médicas por tecnologias de IA que causem dano a pessoas individuais,

¹⁸ Mais recentemente, a OMS publicou um documento contendo as considerações regulatórias para a inteligência artificial para a saúde, disponível em: <https://iris.who.int/handle/10665/373421>.



os processos de responsabilidade e de responsabilização devem identificar claramente os papéis relativos dos fabricantes e dos utilizadores clínicos na produção do dano, podendo ponderar-se modelos de responsabilidade objetiva que abranjam todos os agentes envolvidos no desenvolvimento e colocação de tecnologias de IA;

5) A garantia de inclusão e de equidade – a inclusão requer que a IA usada nos cuidados de saúde seja desenhada para permitir o mais amplo uso e acesso apropriado e equitativo, independentemente de idade, género, rendimento, capacidade ou outras características. Os dados de IA não devem ser amostras enviesadas, mas antes incluir a mais ampla representação, devendo o controlo dos dados pertencer a entidades públicas; e

6) a promoção de uma IA responsiva e sustentável – as tecnologias de IA apenas devem ser introduzidas se puderem ser plenamente integradas e sustentadas no serviço de saúde, na sequência da identificação de necessidades e atendendo aos requisitos do contexto em que são utilizadas. A sustentabilidade requer ainda um desenho ecológico das tecnologias de IA.

4.2. Regulação normativa europeia

No plano da União Europeia, são de salientar dois instrumentos, sendo que um deles constituirá o pilar da regulamentação de IA para todos os Estados-Membros.

4.2.1. Guia de Orientações Éticas para uma IA de Confiança do Grupo de Peritos de Alto Nível sobre a Inteligência Artificial

Em 8 de abril de 2019 foi publicado pela União Europeia o Guia de *Orientações Éticas para uma IA de Confiança* do Grupo de Peritos de Alto Nível sobre a Inteligência Artificial¹⁹. Partindo da premissa de que a IA de confiança deve ser lícita, ética e sólida, este Guia estabelece sete requisitos que a IA deve cumprir, nos quais estão incluídos aspetos sistémicos, individuais e sociais para que possa ser considerada de confiança.

O primeiro requisito é a ação e supervisão humanas. Com efeito, os sistemas de IA devem “capacitar” (*empower*) os seres humanos, permitindo-lhes a tomada de decisões informadas e garantindo os direitos fundamentais. Simultaneamente, devem ser assegurados sistemas de supervisão humana mediante mecanismos de governação de intervenção humana (*human-in-the-loop* – HITL), de fiscalização humana (*human-on-the-loop* – HOTL) ou de controlo humano (*human-in-command* – HIC).

O segundo requisito prende-se com a robustez técnica e a segurança. De outro modo, os sistemas de IA devem ser resilientes e seguros (garantindo a existência de um plano de recurso, caso alguma coisa corra mal), precisos, fiáveis e reprodutivos, dado ser a

¹⁹ Grupo criado pela União Europeia em junho de 2018, encontrando-se o Guia disponível em: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.



única forma de garantir que quaisquer danos acidentais possam ser minimizados e prevenidos.

O terceiro requisito gira em torno da privacidade e governação de dados. Além de ter de ser assegurado o pleno respeito pela privacidade e pela proteção de dados pessoais, devem existir mecanismos de governação que considerem a qualidade e a integridade desses dados.

Mais uma vez encontramos princípios e orientações éticas que coincidem com as de outros instrumentos, tendo como grandes eixos a supervisão humana, a segurança e a privacidade.

O quarto requisito relaciona-se com a transparência. Por outras palavras, os dados, o sistema e os modelos de negócio da IA devem ser transparentes. Além disso, os sistemas de IA e as respetivas decisões devem poder ser explicados de modo ajustado às partes interessadas em causa. Os humanos têm de estar conscientes de que estão a interagir com um sistema de IA e devem ser informados das capacidades e limitações do sistema.

O quinto requisito refere-se à diversidade, não-discriminação e equidade. Os vieses injustificados devem ser evitados, dado que podem ter múltiplas implicações negativas, desde a marginalização de grupos vulneráveis até à exacerbação de preconceitos e discriminação. Para promover a diversidade, os sistemas de IA devem ser acessíveis a todos e envolver as partes interessadas relevantes ao longo de todo o seu ciclo de vida.

O sexto requisito é atinente ao bem-estar social e ambiental. Com vista ao benefício de todas as gerações, presentes e futuras, é necessário garantir que os sistemas de IA sejam sustentáveis e respeitadores do ambiente, devendo o seu impacto social e societal ser cuidadosamente considerado.

O sétimo e último requisito é a responsabilidade. Devem ser criados mecanismos para garantir a responsabilidade e a responsabilização pelos sistemas de IA e pelos seus resultados. A suscetibilidade de auditoria, que permite a avaliação de algoritmos, dados e processos de conceção, desempenha um papel fundamental neste contexto.

4.2.2. O Regulamento da IA da UE (AI Act)

O Regulamento da IA da União Europeia (UE) (AI Act) é, indubitavelmente, o principal instrumento normativo no âmbito da União Europeia. O Regulamento pretende atingir dois objetivos simultâneos: controlar os riscos associados à IA e aumentar a confiança na IA. Ou seja, o objetivo é permitir um desenvolvimento fiável e seguro da IA na Europa, no pleno respeito pelos valores e direitos dos cidadãos, apresentando dois eixos: um ecossistema de confiança, direcionado para a proteção de direitos dos cidadãos e um ecossistema de excelência, visando a criação de valor, promovendo o reforço de investimento, inovação e utilização de IA na UE.

Deste modo, o AI Act pretende estabelecer regras harmonizadas para a colocação no mercado e em serviço de sistemas de IA, proibições de certa IA (quando os sistemas de



IA colidem com princípios éticos universais ou com direitos humanos fundamentais), regras de transparência harmonizadas para interação e, ainda, regras de monitorização do mercado e vigilância (art. 1.º).

Um dos traços originais - e ambiciosos -, à semelhança, aliás, com o Regulamento Geral sobre a Proteção de Dados (RGPD), é o âmbito de aplicação proposto (art. 2.º), pretendendo-se a sua aplicação a fornecedores de sistemas de IA (quanto à introdução na UE ou colocação em serviço na UE, mesmo que aí não estejam estabelecidos), a utilizadores de sistemas de IA (localizados na UE), a fornecedores e utilizadores de sistemas de IA, mesmo que localizados em países terceiros, mas que o *output* produzido pelo sistema seja usado na UE e, em alguns casos, também a distribuidores e importadores. Sendo este o âmbito estrito de aplicação existe um objetivo mais lato, como aconteceu aliás com o RGPD. Ou seja, servir de exemplo para a legislação de outros países sobre IA, ou mesmo para criação de regras vinculativas no plano do direito internacional.

O *AI Act* segue uma abordagem baseada na análise de risco (*risk-based approach*). Deste modo, distingue entre níveis de risco: o mínimo, o limitado, o elevado e o risco inaceitável.

No patamar do risco mínimo são considerados os programas básicos e gratuitos (p. ex., jogos com IA) e não estão previstas particulares regras ou obrigações.

No patamar do risco limitado encontram-se os sistemas de IA em que existe uma maior interação com humanos (p. ex., *chatbot*). Aqui já existem algumas obrigações de transparência, *maxime*, a obrigação de informação dos consumidores ou pessoas humanas de que estão a interagir com um sistema de IA (art. 52.º).

O *AI Act* passou a incluir especificamente a IA generativa, impondo obrigações de transparência, como a informação de que o conteúdo foi gerado por IA, desenho de sistemas que previnam a geração de conteúdo ilícito e a publicação das fontes - beneficiárias de direito de autor - usadas para a produção de conteúdo.

O âmbito principal do *AI Act* inscreve-se no patamar do risco elevado (art. 6.º), aqueles que devam ser usados como componente de segurança de produto ou como produto e que tenham de passar um teste de conformidade para colocação no mercado (p. ex., brinquedos ou equipamento médico). O art. 6.º (2) indica ainda, por remissão para o Anexo III, que serão considerados de alto risco os sistemas de IA que se integrem nas áreas de (i) identificação biométrica e categorização de pessoas humanas, (ii) gestão e operacionalização de infraestruturas críticas, (iii) educação e formação profissional, (iv) emprego, gestão de recursos humanos e acesso ao emprego, (v) acesso e fruição de serviços e benefícios privados ou públicos, (vi) aplicação da lei, (vii) controlo de fronteiras e migração e (viii) administração da justiça e processos democráticos.

Estes sistemas de alto risco encontram-se sujeitos a requisitos específicos, plasmados nos artigos 8.º a 15.º, nomeadamente, o cumprimento dos requisitos da proposta antes



da colocação do sistema de IA²⁰, a implementação de um sistema de gestão de risco, a existência de critérios na gestão de dados (em especial, quanto ao modo de recolha de dados), a exigência de documentação técnica prévia, a manutenção de registos ("logs") durante o funcionamento dos sistemas, a transparência e a informação aos utilizadores²¹, a supervisão humana (na prevenção e minimização de risco) e, finalmente, a precisão, a robustez e a cibersegurança do sistema.

No patamar do risco inaceitável encontramos o âmbito de proibição da proposta. Com efeito, o art. 5.º *AI Act* estabelece os limites da IA permitida, ao identificar os sistemas de IA que constituem um risco inaceitável e, como tal, proibidos, a saber: (i) sistemas de IA que usem técnicas subliminares para lá da consciência humana, com vista a distorcer o comportamento de pessoa de forma a causar danos a si ou a terceiros; (ii) sistemas de IA que explorem qualquer vulnerabilidade de um grupo específico de pessoas devido à sua idade, deficiência física ou mental, com vista a distorcer o comportamento de pessoas pertencentes a esse grupo, de forma a causar dano, físico ou psicológico, a si ou a terceiros; (iii) sistemas de IA, colocados por autoridades públicas, ou a seu pedido, para avaliação ou classificação de confiança de pessoas com base no seu comportamento social; (iv) uso de identificação biométrica remota em tempo real em espaços públicos ou de acesso público. Neste último caso, a proposta exceciona – admitindo – algumas situações em que a identificação biométrica pode ser usada, ou seja, a procura de vítimas, a prevenção de ataques terroristas, situações de ameaças à vida ou à segurança (integridade) física e, finalmente, a procura de suspeitos ou arguidos²².

4.3. Instrumentos nacionais

De entre os instrumentos nacionais, um dos que assume maior relevo neste contexto é a Carta Portuguesa de Direitos Humanos na Era Digital²³, justamente por pretender constituir um instrumento base de toda a regulamentação que incida sobre direitos em ambiente digital. E, nos termos da Carta, mais concretamente, no seu artigo 9.º, a propósito do uso de IA e de robôs, nos termos do qual a utilização da inteligência artificial deve ser orientada pelo respeito dos direitos fundamentais, garantindo um justo equilíbrio entre os princípios da explicabilidade, da segurança, da transparência e da responsabilidade, que atenda às circunstâncias de cada caso concreto e estabeleça processos destinados a evitar quaisquer preconceitos e formas de discriminação (n.º 1), sendo ainda de salientar que, quanto à criação e utilização de robôs, são aplicáveis os

²⁰ Nos termos do *AI Act*, "colocação" significa a primeira colocação de um sistema de IA no mercado da UE (art. 3.º (9)).

²¹ Que se aplica também a sistemas de médio risco, como visto acima. Cf. art. 52.º *AI Act*.

²² Seguiu-se, quanto ao *AI Act*, a sistematização feita em Inês Fernandes Godinho, "Regulação da Inteligência Artificial na União Europeia e Proteção de Dados". In *Tutela Jurídica do Corpo Eletrónico. Novos Desafios ao Direito Digital*, editado por Colombo, Cristiano; Engelmann, Wilson; Faleiros Júnior, José Luiz de Moura, 691-702. Brasil: Editora Foco, 2022.

²³ Lei n.º 27/2021, de 17 de maio, alterada pela Lei n.º 15/2022, de 11 de agosto.



princípios da beneficência, da não-maleficência, do respeito pela autonomia humana e pela justiça, bem como os princípios e valores consagrados no artigo 2.º do Tratado da União Europeia, designadamente a não-discriminação e a tolerância (n.º 3).

Ambiente controlado (Regulatory sandbox)

Um dos aspetos que tem tido mais desenvolvimento na abordagem regulatória da IA, também na saúde, é o fomento de “*regulatory sandboxes*”, ou “modelo de regulação em ambiente controlado”. Com efeito, esta abordagem permite uma perceção dos possíveis riscos na fase de desenvolvimento, ainda que com exigências mínimas, concedendo uma abordagem flexível aos problemas que vão surgindo ao longo do processo. Estas “*regulatory sandboxes*” permitem, ainda, facilitar o desenvolvimento de políticas de implementação que possam mitigar riscos e consequências não desejadas, permitindo uma experimentação controlada de tecnologias de IA. A própria União Europeia tem adotado este modelo²⁴. Trata-se de um modelo avançado de regulação responsiva segundo o qual os reguladores e os regulados cooperam ativamente para a obtenção de um resultado específico e predeterminado²⁵.

5. A aplicação da IA no plano das ciências da vida e seus desafios éticos

A IA está fortemente presente no amplo domínio das ciências da vida, especificamente no âmbito humano e em particular da saúde humana em que tem vindo a ser introduzida a um ritmo acelerado numa crescente diversidade de áreas de atuação, especialidades médicas e atos clínicos. Torna-se, assim, complexo proceder a uma sistematização metodológica das suas principais áreas de intervenção, a qual depois favoreça uma reflexão ética sobre as implicações mais impactantes que poderá ter no setor da saúde.

Propomos uma estruturação em cinco planos ou grandes áreas de intervenção, distintas, mas intercetantes: investigação biomédica, assistência clínica, gestão hospitalar, administração da saúde pública, e ensino/educação em saúde. Propomos igualmente destacar, em cada um destes planos, uma das modalidades de intervenção da IA que se torne particularmente preponderante no espaço visado, sendo que qualquer aplicação possível da IA à saúde humana se exerce concomitantemente em todos os cinco planos que privilegiamos. Paralelamente, traçaremos as mais determinantes inquietudes éticas que se formulam em cada um dos planos estruturados e que exigem uma apreciação ponderada no sentido de, sem negligenciar ou subestimar os ganhos em saúde que a IA pode protagonizar, poder vir, atempadamente, a mitigar ou suprimir impactos indesejados. Tomamos como critério da nossa avaliação os princípios éticos identitários da nossa sociedade como sejam a dignidade humana e a justiça social, destacando-se também o do respeito pela autonomia e o do cuidado pela vulnerabilidade, o da

²⁴ <https://digital-strategy.ec.europa.eu/en/news/first-regulatory-sandbox-artificial-intelligence-presented>.

²⁵ Rui Nunes, *Regulação na Saúde, Vida Económica*, Porto, 2021, 4ª edição.



responsabilidade e o da solidariedade, como os princípios mais diretamente implicados no domínio da saúde.

A aplicação da IA a qualquer domínio de atividade humana e em particular ao da saúde humana, nas diferentes vertentes consideradas, promove genericamente maior rapidez, rigor e objetividade no desempenho das finalidades humanas, complementando e por vezes até substituindo a intervenção humana com vantagens ao nível da eficiência nos desempenhos, da gestão de recursos humanos, mas também económico-financeiros, diminuindo despesas e aumentando a rentabilidade. Os ganhos plurais são incontestáveis e a desenvolver; a transmutação dos valores subjacentes é evidente e a considerar. O seu carácter disruptivo implica transformações profundas e fortemente impactantes que, como tal, exigem ponderação prudente acerca dos benefícios expectáveis e dos potenciais riscos.

5.1. Investigação biomédica e (o impacto do) reconhecimento de padrões

A IA contribui para o avanço da ciência de formas muito diversas, seja através da utilização de ferramentas digitais para exponenciar os modelos clássicos de investigação em seres humanos, em animais ou na esfera ambiental, ou através da criação de formas inovadoras de realizar experimentação científica. Não existe, de facto, nenhuma área científica estanque a esta evolução, que pode mesmo revestir a forma de uma mudança de paradigma da ciência. Porquê, porque nunca como agora a interdependência entre diferentes áreas científicas foi tão profunda e transversal. No domínio da genética, por exemplo, foi a IA que permitiu, para além da genómica, desenvolver a análise e as tecnologias da proteómica. Um aspeto que denota esta evolução paradigmática é a simbiose existente entre a investigação com IA e a investigação em IA. A própria terminologia utilizada evidencia esta interdependência. As redes neuronais são disso um bom exemplo. Também conhecidas por redes neuronais artificiais (*artificial neural networks* - ANN) são um subsistema de *machine learning* e estão na base dos algoritmos de *deep learning*. Inspirados no cérebro humano, mimetizam a forma como os neurónios biológicos transmitem sinais entre si.

A IA é hoje um recurso de fácil acesso e já indispensável no domínio da investigação biomédica, numa significativa variedade de intervenções atuais, o que decorre sobretudo da sua crescente capacidade de recolha, armazenamento e processamento de quantidades gigantescas de dados, reconhecendo padrões a partir dos quais um volume crescente de informação ganha sentido e gera conhecimento, isto é, adquire valor. A IA evidencia, assim, a capacidade de identificar padrões, projetar dinâmicas, construir modelos de análise e interpretativos da realidade, permitindo conhecê-la, a qual, aliás, se intensifica com o atual aumento exponencial da produção de dados, nomeadamente no domínio da saúde. Podemos referir, a título de exemplo, modelos de IA relacionados com abordagens de avaliação de componentes (QSAR) que vão sendo usados para a descoberta e desenvolvimento de novos medicamentos (novas moléculas terapêuticas, desenvolvimento de vacinas), ainda que com algumas limitações.



Podíamos também aqui apontar a relevância da IA, sobretudo por via da realidade virtual, para a diminuição da utilização de animais em projetos científicos.

Destacamos ainda a participação direta de um algoritmo de *machine learning*, em 2020, para o desenvolvimento de um antibiótico que atua em diversos agentes patogénicos, para além do contributo continuado da IA para a descoberta de vacinas ou o desenho de novos medicamentos²⁶. Na investigação em medicina clínica, a inteligência artificial permite melhorar o diagnóstico recorrendo a imagem, ou predizer o comportamento e a resposta a estímulos. De facto, a IA tem-se revelado especialmente útil nas áreas médicas que recorrem à imagem como a radiologia²⁷ ou a dermatologia²⁸. Também é promissora no rastreio oncológico de diferentes patologias²⁹ e, nas ciências cardiovasculares e nas neurociências, tem-se revelado de enorme importância e impacto.

Não existem dúvidas de que a IA facilita a investigação e acelera o desenvolvimento científico. Porém, como é compreensível, delicadas questões éticas existem e tendem a aprofundar-se com a densificação tecnológica da IA. Designadamente, a necessidade de replicabilidade e explicabilidade, a proteção de dados e o consentimento informado, as questões éticas da inteligência artificial generativa, na interface com a criatividade intelectual e a integridade científica, e a necessidade de transparência em todo o processo (nomeadamente, através de políticas e práticas de Ciência Aberta - *Open Science*³⁰).

A **replicabilidade e a explicabilidade**, de forma isolada ou em combinação, representam um dos grandes problemas éticos da IA. A replicabilidade é um tema complexo e controverso, não específico da IA, e que envolve todos os domínios da ciência e da tecnologia, incluindo as ciências sociais. Já Karl Popper referia que “Não levamos muito a sério nem mesmo as nossas próprias observações, nem aceitamos como observações científicas, até que as tenhamos repetido e testado”³¹. A replicabilidade repousa em grande medida na avaliação da fiabilidade do método científico utilizado. Isto é, trata-se da tentativa de, aplicando o mesmo método à mesma

²⁶ Marchant J. Powerful antibiotics discovered using AI. (2020) Nature. 2020 Feb 20. doi: 10.1038/d41586-020-00018-3. Epub ahead of print. PMID: 33603175. 28

²⁷ Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJ (2018) Artificial intelligence in radiology. Nat Rev Cancer 18:500-510. <https://doi.org/10.1038/s41568-018-0016-5.8>

²⁸ Esteva A, Kuprel B, Novoa RA et al (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542:115-118. <https://doi.org/10.1038/nature21056.7>

²⁹ McKinney SM, Sieniek M, Godbole V et al (2020) International evaluation of an AI system for breast cancer screening. Nature 577:89-94. <https://doi.org/10.1038/s41586-019-1799-6>

³⁰ A Ciência Aberta (*Open Science*) representa um conjunto de princípios e práticas que visam tornar os conhecimentos gerados pela investigação científica acessíveis a todos, desde a comunidade científica à sociedade civil, em todos os domínios da ciência. A ciência aberta visa ainda garantir que a própria produção desse conhecimento seja inclusiva, equitativa e sustentável. V., para melhor contexto, a Recomendação da UNESCO sobre esta matéria - UNESCO (2021). *Recommendation on Open Science*. <https://unesdoc.unesco.org/ark:/48223/pf0000379949>.

³¹ Popper K. *The Logic of Scientific Discovery*. London, UK: Routledge; 2005



questão científica, alcançar os mesmos resultados. É assim fundamental avaliar quais os fatores que afetam a replicabilidade da investigação gerada através da IA.

A possibilidade de replicar um estudo é um dos fatores decisivos para construir uma relação de confiança entre o investigador e a comunidade científica e, por maioria de razão, entre a ciência e a sociedade em geral. É fundamental implementar medidas que visem promover a replicabilidade e assim reconhecer o mérito de uma determinada investigação, nomeadamente quando realizada em seres humanos. Se os resultados de um estudo forem consistentes com os de outros estudos, é mais seguro que se possa reclamar como novo conhecimento científico³², sendo que para alguns autores a replicação tem mais valor científico do que a própria descoberta original³³. Isto deve-se ao facto de existir uma desconfiança crescente em relação aos resultados da ciência, em parte devido a uma crise da replicabilidade³⁴. Existem diferentes motivos para este fenómeno, mas o *P-hacking*, isto é, a manipulação da análise de dados até produzir resultados estatisticamente significativos, é um fator decisivo³⁵.

Outras questões também problemáticas reportam-se à limitação da importante variabilidade nas metodologias e, conseqüentemente, da qualidade da investigação que decorreria da atribuição ao algoritmo da melhor ou mesmo da única resposta a considerar perante as hipóteses formuladas. Com efeito, um dos pressupostos da investigação científica, que contribui para a sua validade, diz respeito à replicabilidade dos resultados em diferentes momentos e recorrendo à utilização de diferentes instrumentos e bases de apoio. A intervenção preponderante da IA na investigação biomédica, além de perturbar a determinação rigorosa da autoria do trabalho, pode dificultar a descrição da metodologia utilizada, colocar em causa a variabilidade das abordagens em investigação e, assim, comprometer a replicação dos estudos e condicionar a necessária dúvida científica.

Por seu turno, a excessiva profusão de dados, mesmo de metadados, originou a convicção generalizada que apenas com acesso a dados se encontram padrões relevantes e se atinge a evolução do conhecimento. O perigo da busca sem limites da significância estatística, assim como a conceção generalizada de que mais publicações significa mais e melhor conhecimento científico, encontra na IA o caldo ideal para a sua utilização sem limites controláveis. Circunstância que pode ser exponenciada por uma reduzida (eventualmente nula) explicabilidade, da IA. Ou seja, como é que se garante que a investigação científica recorrendo a IA é replicável se não se conhece ao certo o método utilizado, porque ele não é compressível à inteligência humana? Em teoria, a

³² National Academies of Sciences. Reproducibility and Replicability in Science. [National Academies Press \(US\)](#); Washington (DC): 2019 May 7. ISBN-13: 978-0-309-48616-3

³³ John Ioannidis. Why replication has more scientific value than original discovery. *Behavioral and Brain Sciences* 2018; 41 e137 <http://www.nature.com/articles/s41562-016-0021>

³⁴ Monya Baker. Is there a reproducibility crisis? *Nature* 533 (7604); 2016: 452-454

³⁵ Wicherts J, Veldkamp C, Augusteijn H, et al. Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid *p*-Hacking. *Frontiers in Psychology* 2016; 7 (1832); 2016: 1-12 <https://doi.org/10.3389/fpsyg.2016.01832>



solução reside em desenvolver ferramentas que garantam a explicabilidade da IA e assegurar, simultaneamente, que todos os resultados da investigação são publicados. Isso significa total transparência e que todos os passos previstos no método científico possam ser verificados e reconstruídos.

Por seu turno, **a proteção de dados e o consentimento informado** são imperativos que as sociedades avançadas devem proteger, sobretudo no domínio da IA. A investigação biomédica em seres humanos, recorrendo a IA, pode utilizar dados pessoais, ou produtos biológicos humanos (e outros) integrados em biobancos. A sua recolha, armazenamento e utilização, com dados sensíveis associados, gera progressivamente inquietude junto da sociedade, a ponto de se ter configurado recentemente um direito básico ao esquecimento, ou seja, o direito que assiste a qualquer pessoa de expurgar o sistema dos seus dados pessoais, exceto em circunstâncias muito específicas³⁶. Pelo que a doutrina e a prática do consentimento informado devem ser revistas, de modo a se ajustarem a esta profunda transição digital.

O recurso sistemático à IA aprofunda ainda mais a dicotomia entre o respeito pelo direito à privacidade individual e a necessidade de transparência que se exige à utilização da IA, para assegurar a verificabilidade dos processos utilizados e dos resultados obtidos e gerar a confiança dos utilizadores. A transparência é fundamental para um adequado escrutínio por parte da sociedade, designadamente para uma supervisão adequada pelas comissões de ética. Mas, o exercício da transparência pode constituir-se como um risco acrescido para a privacidade individual e para a proteção de dados pessoais. Sendo fundamental para combater a falta de explicabilidade e o fenómeno designado por *black-box*, uma excessiva transparência pode, no entanto, afetar decisivamente a privacidade, dado que uma adequada e minuciosa explicação do processo pode implicar a exposição de dados pessoais sensíveis e confidenciais.

De facto, a informação presente nos registos de saúde eletrónicos, ou obtida em ensaios clínicos, é altamente sensível e tem-se tentado minimizar potenciais quebras de privacidade através da anonimização, onde as variáveis potencialmente identificáveis são removidas. Também a pseudonimização dos dados é prática frequente, através da codificação de dados e da existência de um identificador artificial. Porém, diferentes tentativas de reidentificação demonstraram a falibilidade da anonimização, pelo que outras estratégias têm sido sugeridas. Por exemplo, o recurso a dados sintéticos.

Dados sintéticos são dados artificiais que podem ser usados na investigação científica, evitando o acesso a dados pessoais. Os dados sintéticos, recorrendo usualmente a *machine learning*, originam uma representação realista da fonte original, parecendo os dados desta fonte, mas sem qualquer informação de indivíduos reais. O objetivo é preservar algumas das propriedades estatísticas desta fonte original, tal como a distribuição de dados contínuos, proporções de dados categóricos, ou correlações

³⁶ Correia, M; Rego, G; Nunes, R. Gender transition: Is there a right to be forgotten? *Health Care Analysis*, 2021; 29: 283-300.



entre variáveis. Existem ainda dúvidas sobre se os dados artificiais podem ser considerados válidos para análises formais e quais os reais riscos da sua utilização. Importa também garantir a qualidade dos dados e evitar novos vieses (por exemplo, nas revisões sistemáticas). Desde logo, importa garantir a explicabilidade do modo como foram gerados³⁷. Esta tendência evidencia ainda mais a “*datificação da ciência*”, ou seja, a evolução de uma ciência baseada em hipóteses (*hypothesis-driven science*) para uma ciência baseada em dados (*data-driven science*). Uma solução plausível poderia ser a implementação de uma política de “*código aberto*”, isto é, de *software* aberto. A falta de estímulo e de interesse nesta estratégia, desde logo do ponto de vista económico por parte dos investidores, pode implicar uma nova perspetiva dos direitos de autor ou, alternativamente, o exercício da tutela através de um novo direito das patentes.

Em qualquer caso, a IA irá necessitar de um amplo debate público e de um consenso mais alargado para a sua implementação e supervisão. Na investigação científica, a “*inteligência coletiva*” pode ser uma abordagem a considerar. A experiência, conhecimento e paixão de um determinado grupo de pessoas é o que se denomina por *inteligência coletiva (crowdsourcing)*³⁸. Isto é, o recurso a redes de especialistas de diferentes setores de modo a abordar, de um modo mais eficiente e legítimo, a escala e a complexidade de problemas tal como a supervisão da IA.

Também a **inteligência artificial generativa**, e o seu impacto **na criatividade intelectual, na integridade científica e na transparência** (e.g., *open science*) representam um desafio ético sem precedentes. De facto, ao dominar a linguagem humana, e tendo um acesso ilimitado a bases de dados que constam na Internet, a inteligência artificial generativa, tal como o *ChatGPT*, é um dos domínios mais fecundos da aplicação destas novas tecnologias digitais. No plano da investigação biomédica, está frequentemente em causa o desenvolvimento de algoritmos para a pesquisa de artigos e para a recolha e sistematização da informação, sendo hoje possível, de uma forma bastante razoável, a realização de artigos científicos de revisão por algoritmos, sem intervenção humana. Contudo, a circunstância de o *ChatGPT* recorrer a informação disponível na Internet implica que a resposta aos *prompts* seja em grande medida enviesada, não sendo possível verificar a validade dessa informação³⁹. Isto porque a resposta às questões formuladas depende das bases de dados disponibilizadas, o que à partida se encontra fortemente condicionado pelas fontes, que são geralmente as entidades com mais acesso a estas plataformas. O que implica distorções científicas e mesmo culturais.

Reconhecendo-se o possível contributo do algoritmo para a análise dos dados e redação do artigo, registam-se também várias falhas, nomeadamente quando a mesma pergunta

³⁷ Kokosi T, Harron K. Synthetic data in medical research BMJ Medicine 2022;1: e000167. doi:10.1136/bmjmed-2022-000167

³⁸ Matt Ryan, Dane Gambrell, Beth Simone Noveck, Using Collective Intelligence to Solve Public Problems, Nesta/The Govlab, Brooklyn 2020.

³⁹ Lee, P; Bubeck, S; Petro, J. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. New England Journal of Medicine. 2023, Mar 30;388(13):1233-1239. doi: 10.1056/NEJMs2214184.



é feita de forma diferente, várias vezes. O mesmo algoritmo poderá, então, dar respostas diversas, o que suscita a questão adicional da responsabilidade pelo trabalho realizado. Aliás, esta mesma interrogação remete igualmente para a da atribuição da autoria do trabalho, uma vez que os documentos produzidos através destes algoritmos se apresentam como inéditos, embora tenham resultado da composição de informação existente. De facto, já foi publicado um artigo tendo o *ChatGPT* como coautor⁴⁰, não obstante esta possibilidade ter sido posterior e genericamente rejeitada.

Nestes casos de produção de trabalho científico-académico por um algoritmo, a sua eventual assinatura por uma pessoa humana, não podendo ser classificada como plágio, não deixaria de ser fraudulenta, aspeto particularmente importante quando a progressão na carreira e a avaliação de desempenho (pessoal e institucional) se fazem, sobretudo, em função das publicações realizadas e cuja autoria é escrupulosamente averiguada. Pelo que também a garantia da integridade científica deve ser necessariamente reapreciada face à IA generativa⁴¹. Universidades, institutos de investigação, e outras entidades públicas ou privadas onde se realize investigação alavancada pela IA devem implementar medidas concretas que visem preservar a integridade da ciência e dos investigadores.

Paralelamente, a utilização destes algoritmos permite que investigadores venham a assinar artigos fora das suas áreas de especialidade, pelo que sem a reflexão crítica desejável, o que afeta a qualidade da investigação. Também os investigadores que melhor dominarem a linguagem algorítmica terão melhores resultados, que mascaram o real nível de conhecimentos do domínio em causa. Por exemplo, sugeriu-se a utilização da IA para melhorar o trabalho já efetuado e publicado⁴² o que suscita naturalmente significativas questões éticas se esta evolução de textos já publicados não for claramente reconhecida e divulgada.

Valorizando-se a rapidez, eficiência e rigor que confere à investigação científica, importa igualmente ponderar que a IA coloca em causa, por exemplo, o atual sistema de registo de patentes, originando dificuldades reais na atribuição dos méritos e das mais-valias associadas à inovação produzida. Simultaneamente, a crescente capacidade de processamento da IA tenderá a reduzir o nível de segurança da anonimização ou configurar uma pseudonimização, o que poderá comprometer tanto a utilização desses mesmos dados secundários como, em sentido oposto, a privacidade das pessoas a que se referem. Também o regime da propriedade intelectual terá de ser reapreciado, dado que a excessiva proteção da propriedade intelectual pode ser um obstáculo à

⁴⁰ Kung, T. H., Cheatham, M., GPT, C., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., & Tseng, V. (2022). Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models. *MedRxiv*, 2022.12.19.22283643. <https://doi.org/10.1101/2022.12.19.22283643> 29

⁴¹ The European Code of Conduct for Research Integrity. Revised edition. ALLEA - All European Academies, 2023. <https://allea.org/code-of-conduct/>

⁴² Porsdam Mann, S., Earp, B. D., Møller, N., Vynn, S., & Savulescu, J. (2023). AUTOGEN: A Personalized Large Language Model for Academic Enhancement—Ethics and Proof of Principle. *American Journal of Bioethics*. <https://doi.org/10.1080/15265161.2023.2233356>



transparência e explicabilidade da IA generativa. Pelo que se impõe uma minuciosa avaliação, de preferência *ex ante*, dos sistemas de IA e do seu impacto ético, para além de um reforço do sistema de supervisão, regulação e acompanhamento ao longo do ciclo de vida da utilização da IA.

Por seu turno, assiste-se a um movimento imparável de subalternização dos investigadores com a progressiva substituição do trabalho humano pelos sistemas de IA⁴³. O que, para além de questões de outra natureza - tal como a dimensão autorrealizadora do trabalho ou a progressiva substituição do ser humano pela automação e o conseqüente impacto nos salários, contribuições sociais e nível de vida - coloca em causa um elemento central na evolução da ciência que é a criatividade. Isto é, se a atual deriva das políticas científicas e a excessiva padronização dos métodos científicos já torna difícil pensar “fora da caixa”, o recurso sistemático à inteligência artificial generativa torna difícil exponenciar a criatividade humana e, virtualmente, impossível usufruir do acaso na ciência. Note-se que o “feliz acaso na ciência” (*serendipity*) foi e é responsável por incríveis benefícios científicos (descoberta da penicilina, da heparina ou dos raios X).

Muitas descobertas científicas que revolucionaram o mundo tiveram importantes contributos do acaso na ciência que estará particularmente limitado com a IA generativa. Pelo que soluções inovadoras devem ser concebidas para que a criatividade humana continue a estar no epicentro do desenvolvimento científico.

Partindo do pressuposto que os sistemas de inteligência artificial generativa mais evoluídos serão dominados por alguns grupos económicos colocam-se, também, questões éticas relacionadas com o modo como se armazena e processa a informação, o diferente *ethos* do setor privado *versus* entidades públicas, ou mesmo fenómenos de sistemática exploração laboral.

Também o impacto intergeracional deve ser avaliado, dado que as novas gerações estão “naturalmente” incluídas digitalmente, o que pode originar um aprofundamento das desigualdades sociais e mesmo económicas entre as diferentes gerações. Pelo que uma solução ética e socialmente responsável seria, para além de um incremento geral na literacia digital, a promoção de uma intensa colaboração entre os diferentes parceiros envolvidos parceria multidisciplinar e multicultural, sendo esta uma forma prática de inteligência coletiva, de democracia em saúde e de *open science*.

Em suma, a aplicação da IA à investigação biomédica contribui decisivamente para acelerar o progresso científico, pela capacidade de recolher e processar mais dados, gerando mais e melhor conhecimento numa maior diversidade de áreas comunicantes. Simultaneamente, importa ter em atenção as alterações disruptivas que um progresso científico marcado pela agregação e padronização de dados provoca em vários planos de atividade humana, nomeadamente, nos da investigação, académico e jurídico.

⁴³ S. Bianchini et al. “Artificial intelligence in science: An emerging general method of invention” Research policy 51, 2022, p. 2.



No plano científico, destaca-se a relevância de manter dúvidas e hipóteses científicas, de implementar diferentes metodologias explicáveis e de poder proceder à replicação dos estudos; no académico, destaca-se a importância da atribuição da autoria do trabalho realizado e de promover a originalidade da análise e interpretação da realidade; e no plano jurídico destaca-se a questão das patentes, cuja legislação específica, visando a proteção dos direitos de propriedade intelectual e o encorajamento da investigação científica, corre o risco de se tornar obsoleta, dada a primazia crescente da IA no processo de investigação, quando foi estabelecida no pressuposto de os investigadores serem necessariamente humanos.

Aspetos Éticos Relevantes da IA na Investigação Biomédica

- Não existe nenhuma área científica estanque ao impacto da IA, podendo mesmo tratar-se de uma mudança de paradigma da ciência;
- Na investigação em ciências biomédicas, a IA suscita complexas questões éticas que tendem a aprofundar-se com a evolução das tecnologias digitais;
- A replicabilidade e a explicabilidade representam um dos grandes problemas éticos da IA. Importa implementar medidas concretas para ultrapassar a crise de replicação na ciência promovida pela IA;
- A utilização massiva de dados pessoais implica um reforço importante da sua proteção, de novas modalidades de anonimização e de inovadoras estratégias para garantir a privacidade individual, tal como o recurso aos dados sintéticos gerados por IA;
- A doutrina e a prática do consentimento informado devem ser revistas para se ajustarem à profunda transição digital promovida pela IA;
- A inteligência artificial generativa representa um desafio ético sem precedentes, apesar dos inúmeros benefícios subjacentes. Deve ser clarificada a questão da autoria científica e as políticas institucionais de integridade científica devem ser reapreciadas face à IA generativa;
- O recurso sistemático à IA não deve colocar em causa a criatividade intelectual dos investigadores.

5.2. Assistência clínica e (o impacto da) assistência digital

A IA tem o potencial de transformar profundamente a vida das pessoas em todos os domínios da sociedade e em todas as atividades económicas. Mas é na saúde, paradoxalmente, que este impacto se irá sentir com maior acuidade, dada a enorme variedade de domínios em que a IA (incluindo o *machine learning*, o *deep learning* e as redes neuronais) terá uma profunda aplicação prática. De facto, é no plano da assistência



clínica que o desenvolvimento da IA e a sua integração no domínio da saúde humana mais tem avançado, numa diversidade de aplicações de difícil sistematização e de uma praticamente impossível enunciação exaustiva.

Uma das áreas da medicina onde a IA poderá ter mais impacto é o domínio do **diagnóstico** e mesmo do diagnóstico diferencial. Designadamente, através da análise de dados provenientes da interpretação de imagem (radiologia convencional, tomografia computadorizada, ressonância magnética, ou ultrassonografia), da observação do fundo de olho através de fotografias da retina, análises histológicas para anatomopatologia, ou observação de lesões cutâneas na dermatologia. Ou, por outro lado, recorrendo à leitura computadorizada avançada de eletrocardiograma ou ao estudo sanguíneo através da contagem e estudo das células brancas presentes na circulação.

Porém, o **tratamento de dados** por IA merece especial atenção e uma particular consideração ética, uma vez que o progresso verificado na ciência de dados não se traduz apenas em um aumento da *performance*, da rapidez e da capacidade de armazenamento. A grande diferença é que os dados são simultaneamente gerados por pessoas e por programas de *software* através da IA e de complexos algoritmos, sendo esta intersecção de elevada complexidade. A intensa conectividade proporcionada pela IA, operando em rede, coloca uma questão técnica de difícil resolução, e que se designa por "*data set shift*". Isto é, a hipótese de que a mudança de dados para treinar um modelo de *machine learning* para dados encontrados no mundo real, por exemplo provenientes de pessoas doentes, possa alterar a natureza da intervenção sobre esses mesmos dados, não existindo, assim, correspondência entre o *dataset* sobre o qual o sistema de IA foi desenvolvido e aquele em que irá ser implementado. Isto pode dever-se a diferenças existentes entre diferentes sistemas de saúde, na população de pessoas doentes, ou mesmo nas práticas clínicas⁴⁴.

Algumas questões que se podem colocar na aplicação da IA ao diagnóstico clínico são: qual a norma concreta a seguir? como é que o viés segundo o qual a IA e os algoritmos de *machine learning* foram instruídos pode influenciar a sua aplicação no mundo real? como são os valores éticos humanos incluídos na IA e nos algoritmos de *machine learning* de modo que os resultados obtidos venham a refletir os problemas reais dos profissionais de saúde? quais os *standards* da IA na investigação clínica de intervenção? que estratégia deve ser assumida pelos reguladores de modo a evitar a publicidade enganosa da IA, nomeadamente face à problemática, aparentemente inultrapassável, da sua difícil explicabilidade (*black-box*)?

Na radiologia, por exemplo, é expectável um futuro promissor para a IA. Algumas das funções dos algoritmos utilizados na assistência a radiologistas passam pela deteção de imagens suspeitas, melhoramento de imagens, auxílio na triagem, ou mesmo na

⁴⁴ Haug CJ, Drazen J. (2023). Artificial Intelligence and Machine Learning in Clinical Medicine. New England Journal of Medicine, 388(13), 1201-1208. <https://doi.org/10.1056/NEJMra2302038>



quantificação⁴⁵, avaliando, por exemplo, a densidade do tecido mamário ou pulmonar, estudando estruturas anatómicas cerebrais ou quantificando o fluxo cardíaco. Não restam dúvidas de que os modelos radiológicos de IA podem expandir as capacidades interpretativas dos médicos especialistas em radiologia, dispendo mesmo da capacidade de predizer resultados clínicos.

Porém, algumas questões com incidência ética relevante devem ser apreciadas. Se, por um lado, a IA na radiologia poderá originar um aumento do acesso a exames radiológicos num contexto de escassez de recursos, em países com sistemas mais ou menos desenvolvidos, nomeadamente em grandes extensões territoriais onde o acesso à saúde é limitado, a sua utilização recorrente pode estar dependente de médicos não especialistas nesta área médica. O que coloca o problema, pelo menos potencial, de falta de competência reconhecida profissionalmente, de ausência de conhecimentos médicos apropriados, de diminuição da confiança nos resultados obtidos, de falhas de supervisão e verificação e mesmo de alterações profundas da identidade e autonomia profissionais.

Estas dúvidas são naturalmente extensíveis aos diferentes domínios da saúde nos quais a IA possa vir a alcançar uma posição predominante, antevendo-se que, na medicina de precisão – recorrendo à genómica, à proteómica, ou à metabolómica – todas estas questões éticas sejam ainda mais exponenciadas.

Estas questões éticas são extensíveis também à **assistência virtual** através do recurso a *chatbots* que funcionam como assistentes digitais (algoritmos para a decisão clínica), disponibilizando uma assistência personalizada a partir dos dados (e.g. história clínica, sintomas) que lhe forem transmitidos pelo utilizador. Estão acessíveis numa diversidade imensa de plataformas, como sejam *websites*, *apps*, assistentes de voz e em aplicações interativas que permitem uma comunicação rápida e a obtenção de decisões simples. Esta tecnologia poderá ser útil no atendimento remoto de utentes e na triagem de doentes, bem como na elaboração de diagnósticos e recomendação de terapêuticas, libertando os recursos humanos para tarefas mais complexas.

Contudo, não podemos deixar de apontar a existência de um risco real de se criar um distanciamento e afastamento entre profissionais de saúde e pessoas doentes e respetivas famílias, que se instala desde o contacto remoto em que pode estar envolvido um *chatbot*, que poderá não se identificar como tal, à receção na unidade de saúde e no processo administrativo-burocrático. A mediação tecnológica pode despersonalizar o encontro clínico e padronizar as relações. Pode também erguer novas barreiras junto de segmentos da população com menor acesso ou mais baixa literacia tecnológica, agravando vulnerabilidades e a injustiça social. No que se refere ao exercício profissional, pode diminuir a liberdade, imaginação e intuição nos processos e nas decisões, podendo ainda tornar-se difícil a um médico descartar o diagnóstico de um

⁴⁵ Rajpurkar P, Lungren MP. (2023). The current and future state of AI interpretation of medical images. *New England Journal of Medicine*. 388: 1981-90. <https://doi.org/10.1056/NEJMra2301725>



assistente digital, supostamente mais rigoroso, bem como à pessoa solicitar uma segunda opinião.

Em termos gerais, as pessoas, sobretudo doentes e seus familiares, tenderão a não duvidar da decisão algorítmica, o que poderá conduzir a alterações no seu próprio processo de decisão, na relação com a patologia e com os profissionais de saúde; paralelamente, as pessoas serão tendencialmente menos tolerantes com as falhas dos algoritmos do que com as dos humanos. A consideração destes aspetos suscita inevitavelmente a problematização das várias dimensões da responsabilidade implicadas e os seus principais destinatários. Por exemplo, o profissional de saúde poderá ter cada vez menos capacidade de verificação das decisões tomadas pelos algoritmos porque vai perdendo a sua prática, ou porque desconhece o seu funcionamento e o *rationale* envolvido na tomada de decisão. Neste sentido, as decisões serão tendencialmente mais opacas, pondo em causa o princípio da verificabilidade, central para a correta avaliação de riscos e para promover a necessária prestação de contas junto das pessoas e da sociedade.

Ou seja, a existência de assistentes virtuais baseados na inteligência artificial, recorrendo a *chatbots* de distinta natureza, tem o potencial de transformar substantivamente a prática dos cuidados de saúde. Isto porque, pelo menos potencialmente, não se trata de uma mera ferramenta de auxílio na relação clínica e no processo de tomada de decisão. Pelo menos seis aspetos podem ter impacto na natureza intrínseca da relação clínica⁴⁶: desigualdade no acesso a cuidados de saúde de qualidade, transparência na relação com as pessoas doentes, risco de viés social em sistemas de IA, diluição da perceção de bem-estar por parte da pessoa doente, risco de desqualificação profissional e de diluição da responsabilidade, e impacto no direito à privacidade.

Antecipa-se que os assistentes virtuais possam assumir funções e responsabilidades assistenciais progressivamente crescentes, existindo também uma crescente dificuldade em determinar se as respostas providenciadas estão fundamentadas em factos e evidências cientificamente validados. Esta função de "*scribe and coach*", para além de implicar ética e juridicamente a determinação da responsabilidade relativa de cada uma das partes, acarreta a necessidade de repensar definitivamente práticas já estabelecidas, como a obtenção do consentimento informado ou os registos de saúde eletrónicos.

A preocupação será acrescida quando estas aplicações tiverem a capacidade, não apenas de efetuar diagnóstico diferencial, mas de produzir recomendações de tratamento, o que suscita desde logo a questão de determinar se a orientação terapêutica decorrente dos *prompts* e respetivas respostas é a recomendação adequada de acordo com as *leges artis*. Importa também determinar o modo como é verificada a

⁴⁶ Mittelstadt B. (2021). The impact of artificial intelligence on the doctor-patient relationship. Council of Europe, December. <https://www.coe.int/en/web/bioethics/report-impact-of-ai-on-the-doctor-patient-relationship>



sua veracidade, e se o programa tem a capacidade, ou não, de detetar o erro e corrigir a opção terapêutica.

A **Mobile Health** (mHealth) tem sido outra área de grande desenvolvimento. Esta refere-se à utilização de diversos dispositivos, sobretudo objetos de uso pessoal como, por exemplo, telemóveis ou relógios inteligentes, que podem ser usados para automonitorização e autoavaliação, medindo e registando, entre outros, valores de pressão arterial, frequência cardíaca, ECG e atividade cardíaca, fluxo sanguíneo, temperatura, atividade física, padrões e qualidade de sono, frequência respiratória, níveis de stress, calorias consumidas. Muitos destes sistemas de monitorização trabalham automática e continuamente, transferindo, agregando e interpretando os dados recolhidos.

A **telessaúde** também recorre a estas novas tecnologias ao, por exemplo, utilizar imagens de lesões de pele e outras captadas por câmaras de telemóveis pessoais para apreciar situações clínicas. Em qualquer caso, a transmissão de dados de saúde, obtidos pelos cidadãos, aos profissionais de saúde que os assistem, torna-se cada vez mais frequente no seu desempenho como auxiliares de saúde. Importa também referir os milhares de aplicações associadas a intervenções em saúde digital: em 2021 existiam mais de 350 000, com um aumento de 90 000 nesse mesmo ano, e com um crescimento de cerca de 250 novas aplicações por dia⁴⁷. Destacam-se as aplicações direcionadas para patologias específicas - como a saúde mental, a diabetes e as doenças cardiovasculares - ou as que possibilitam o acesso rápido a certificados de vacinação, processos clínicos ou exames de diagnóstico (MySNS, SNS 24).

Esta ampla variedade, disponibilidade e utilização de aplicações digitais potencializa o seu desempenho como auxiliar efetivo, mas também, por vezes, como substituto eficaz de cuidados clínicos prestados por profissionais, sendo perspetivado como um acelerador no alargamento da cobertura universal dos serviços de saúde. As aplicações digitais podem facilitar a comunicação entre as pessoas e os profissionais de saúde, melhorar a gestão do tratamento, monitorizar a pessoa em tempo real e promover a acessibilidade aos tratamentos - vantagens evidentes que, simultaneamente, suscitam questões de fiabilidade em relação aos dados, gerando também expectativas e receios que se podem tornar irrealistas ou assustadores. As aplicações digitais permitem igualmente aos utilizadores uma maior compreensão das relações entre o seu comportamento e a sua saúde, encorajando a uma mudança de comportamentos individuais, e a uma maior autonomia dos cidadãos no âmbito da saúde.

Estas são também vantagens que importa ponderar, a par da consciência de que a autonomia, exercida frequentemente a partir de conhecimentos superficiais e visando sempre uma maior segurança, pode conduzir à resistência de formulação de dúvidas ou

⁴⁷ "Digital health apps balloon to more than 350,000 available on the market, according to IQVIA report". (2021). Journal Article. Mobile Health News. Digital Health Ed. 4 August 2021
<https://www.mobihealthnews.com/news/digital-health-apps-balloon-more-350000-available-market%20according-iqvia-report>



revisão de posições, contribuindo em última análise para tomadas de decisão prejudiciais à saúde. Além disso, urge reconhecer que parte significativa do sucesso das intervenções em saúde está associada à construção de uma relação de confiança entre o profissional de saúde e a pessoa doente, o que pode ser comprometido num contexto exacerbado de autonomização individual.

Tal como já se aludiu, na panóplia de aplicações da IA no contexto clínico destacamos os **algoritmos de deteção de objetos, dados e imagens**, com rigor e rapidez únicos, que têm o potencial de revolucionar especialidades como a oncologia, a radiologia, a patologia, a oftalmologia e a cardiologia, implicando também a diminuição do número de especialistas nestas áreas, para além do seu desempenho transversal nos cuidados de saúde. Estas aplicações da IA colocam também insanáveis questões éticas relacionadas, por exemplo, com a privacidade individual e a proteção de dados pessoais. De facto, a obtenção de estes dados de saúde, incluindo som e imagem, por aplicações e sistemas fora da esfera pública torna impossível qualquer garantia total de proteção, ou mesmo de controlo da sua utilização futura com finalidades distintas das originalmente pretendidas. Nomeadamente quando se trate de aplicações, e respetivas infraestruturas digitais que lhes servem de suporte, cuja titularidade seja de organizações exteriores à União Europeia e, portanto, sem os mesmos requisitos éticos e legais, tal como os previstos nos já referidos RGPD⁴⁸ ou no *AI Act*, a primeira legislação compreensiva sobre IA em todo o planeta⁴⁹.

Estes algoritmos têm capacidade de autoaprendizagem, evoluindo assim no seu poder de **diagnóstico**, apresentando já hoje taxas de sucesso superiores às dos próprios profissionais de saúde, numa pressão para o seu recurso massivo ou rotinizado, o que poderá acarretar uma redução da capacidade dos profissionais a curto prazo. Uma outra aplicação de menção incontornável é a da utilização terapêutica da realidade virtual (RV), por exemplo, na realização de processos de dessensibilização, promovendo a exposição das pessoas a estímulos fóbicos⁵⁰. Se esta tecnologia partiu da utilização de simuladores de voo, os atuais avanços tecnológicos permitem um maior nível de imersibilidade e uma mais fácil acessibilidade. A RV permite intervenções em contextos virtuais, sobretudo em circunstâncias em que essa exposição *in vivo* seria difícil ou arriscada, sendo os estímulos mais bem controlados pelo profissional e obtendo-se uma resposta (*feedback*) imediata sobre a forma como a pessoa reage à intervenção⁵¹. Pode também oferecer intervenções

⁴⁸ European Union. (2016). General data protection regulation. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. <http://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:32016R0679>.

⁴⁹ European Union. (2023). Artificial intelligence act. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>

⁵⁰ Maples-Keller, J. L., Bunnell, B. E., Kim, S. & Rothbaum, B., O. (2017). The use of virtual reality technology in the treatment of anxiety and other psychiatric disorders. *Harvard Review of Psychiatry*, 25(3), 103-113. <https://doi.org/10.1097/HRP.000000000000138>

⁵¹ Freeman D, Reeve S, Robinson A, Ehlers A, Clark D, Spanlang B, Slater M. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychol Med*. Oct; 47(14): 2393-2400. <https://doi.org/10.1017/S003329171700040X> Epub 2017 Mar 22. PMID: 28325167; PMCID: PMC5964457



personalizadas e sistemáticas ao nível da reabilitação neuropsicológica, promovendo a autonomia da pessoa doente e permitindo a realização das intervenções em casa.

Em todo o caso, dado a sua aplicação em contextos clínicos ser relativamente recente, convém desenvolver estudos de acompanhamento para aferir os seus resultados a médio e longo prazo, nomeadamente a possibilidade de gerar alguma confusão entre o mundo real e virtual, com a possível criação de falsas memórias, sobretudo em crianças.

Ainda no âmbito da referência a aplicações inovadoras da IA no domínio da saúde, apontamos a associação de algoritmos de IA com a **robótica**, na chamada robótica inteligente. Os robôs têm vindo cada vez mais a ser chamados a um desempenho na assistência clínica, como coadjuvantes, por exemplo, em procedimentos cirúrgicos, mas também na prestação de cuidados vários a pessoas idosas ou a pessoas com deficiência, ou ainda a pessoas que sofreram acidentes vários, como os acidentes vasculares cerebrais (AVC), nomeadamente na manutenção, desenvolvimento ou recuperação de capacidades, neste último caso na medicina de reabilitação, para referir apenas alguns desempenhos.

Na *interface* entre a robótica e a telecirurgia, por exemplo, as questões éticas estão associadas com a essência da relação entre o médico e a pessoa doente, nomeadamente no que respeita ao exercício da autonomia, à obtenção de consentimento apropriado, à regulação de conflitos de interesse, à má-prática médica, ao direito à privacidade ou mesmo à justiça e equidade no acesso à inovação tecnológica⁵². Na medicina, a interação humana é fundamental, sendo necessário garantir que a pessoa doente se sente confortável em partilhar informação médica sensível e relevante com os robôs, ainda que dados recentes demonstrem que alguma conexão existe entre as pessoas doentes e os robôs sociais. De momento, não parece haver evidência de que os robôs consigam aplicar o conhecimento médico e ajustá-lo às idiossincrasias de cada pessoa, podendo a sua generalização, por isso, originar um aumento da desumanização da saúde.

Pode tratar-se, por outro lado, de **robôs sociais** que interagem diretamente com humanos e que se destinam sobretudo a pessoas doentes que carecem de cuidados a longo prazo, propiciando-lhes um acompanhamento permanente. Não obstante, importará, também em relação a esta ação inovadora, proceder ao acompanhamento dos impactos da relação máquina-pessoa e com as famílias. Entretanto, ao receio de desumanização dos cuidados, que neste contexto primeiramente foi formulado, vem-se acrescentando o das relações afetivas desenvolvidas pelas pessoas em relação aos algoritmos, o que se tornará mais frequente com a utilização de robôs humanoides e que deverá provocar outro tipo de consequências a identificar e ponderar, sobretudo por uma nova área de estudo denominada como a IA social, centrada nesta interação da

⁵² Ahuja AS. (2019). The impact of artificial intelligence in medicine on the future role of the physician. PeerJ, 10 <https://doi.org/10.7717/peerj.7702>



IA com os humanos. Espera-se, no entanto, que a robótica automatizada possa interagir de um modo humanizado com as pessoas doentes, ou seja, de acordo com os valores éticos da nossa sociedade, onde a empatia desempenha um papel central. Esta “IA robótica virtuosa” poderá não ser genuinamente virtuosa, mas apenas “atuar” de um modo virtuoso e compassivo⁵³. E, assim, tentar reproduzir ao máximo a essência da relação clínica.

A aplicação de tecnologia baseada em IA na saúde suscita questões complexas que se podem agravar no plano da participação das pessoas nos cuidados pela sua saúde em parceria com os profissionais. Por exemplo, se considerarmos que os profissionais de saúde, em geral, não dispõem de elevada proficiência em tecnologias deste tipo, a IA vai permitir um nível de autonomização da técnica cada vez maior, reduzindo o nível de participação do profissional nos processos de tomada de decisão. Este aspeto, só por si, coloca questões de responsabilidade sobre os procedimentos, associadas à potencial delegação de decisões clínicas ao algoritmo, mas também a possibilidade de informar com segurança sobre as reais probabilidades de falhas. Com efeito, terá de haver uma permanente e atualizada avaliação de qualidade da tecnologia disponível e da evidência científica associada. Neste âmbito, dever-se-á investir no estabelecimento dos critérios a implementar para essa avaliação e no perfil do responsável por essa mesma avaliação, garantindo ainda a disseminação dos resultados obtidos.

Acresce a conveniência de, por questões de responsabilidade individual e de potenciais conflitos de interesse, se definir que tecnologias poderão ser aconselhadas pelos profissionais de saúde, apesar da liberdade de acesso que existe no espaço virtual. O sistema regulador deverá ser reforçado de modo a avaliar, certificar e supervisionar a utilização de estas aplicações.

Ainda no plano da participação de cada um na gestão da sua saúde, poder-se-ia argumentar que competiria ao próprio aceitar ou não as condições de uma qualquer intervenção, incluindo a assunção das dúvidas associadas ao processo. Mas tal implicaria, simultaneamente, deixar as pessoas sozinhas com as suas decisões, abandoná-las. O consentimento informado, no respeito pela autonomia da pessoa, não se esgota no ato de uma qualquer decisão, mas compreende um processo de envolvimento ativo da pessoa nas intervenções clínicas que lhe dizem respeito e uma relação de confiança que promove, por sua vez, a adesão terapêutica, condição central para o sucesso da maioria das intervenções. Por outro lado, as intervenções em saúde baseadas na IA, tendencialmente mais económicas e com mais elevadas taxas de sucesso em alguns procedimentos, contribuirão para a supressão de alternativas, afetando significativamente o processo de consentimento informado: a ausência de procedimentos alternativos na ponderação de um consentimento informado pode configurar uma forma de coação.

⁵³ Constantinescu, M., & Crisp, R. (2022). Can robotic AI systems be virtuous and why does this matter? *International Journal of Social Robotics*, 14(6), 1547-1557. <https://doi.org/10.1007/s12369-022-00887-w>



Também é relevante sublinhar que as intervenções padronizadas da IA deverão apresentar um desfasamento em relação à individualidade de cada pessoa, particularmente evidente no que se refere a dimensões psicológicas e de bem-estar. Em suma, a aplicação da IA à assistência clínica é muito diversificada e ampla, encontrando-se em processo acelerado de expansão. A digitalização da assistência clínica, se contribui significativamente para o alargamento do acesso a cuidados de saúde e para a autonomia dos cidadãos na gestão da sua saúde, protagoniza por outro lado alterações disruptivas, em particular na relação profissional de saúde/cidadãos, impondo um novo paradigma marcado pela inversão da assimetria originária: o do exercício hegemónico de poder do profissional de saúde sobre a pessoa doente, num designado paternalismo médico, para o da decisão autónoma da pessoa a sugerir ao profissional de saúde, numa potencial funcionalização deste profissional.

A proatividade da pessoa não deve implicar a passividade do profissional de saúde. O objetivo é o de uma maior simetria da relação, fundada na confiança das pessoas e na fidelidade dos profissionais de saúde, e na respetiva assunção das responsabilidades de cada um. O que implica um ajustamento de todo o processo de obtenção de consentimento informado. Deverá haver, por um lado, um reforço das competências digitais dos profissionais de saúde e dos cidadãos em geral, sobretudo nas áreas da saúde e da medicina onde esta evolução seja mais evidente⁵⁴. Por outro, a implementação de um sistema regulatório suficientemente robusto para prevenir práticas que manifestamente conduzam a uma degradação da qualidade e da humanização dos cuidados de saúde⁵⁵.

Aspetos Éticos Relevantes da IA na Prática Clínica

- Desvirtuamento da relação clínica;
- Afastamento entre profissionais de saúde e as pessoas doentes;
- Erosão da relação de confiança na prática clínica;
- Diminuição da competência dos profissionais de saúde com a perda de prática clínica associada à generalização da IA;
- Falta de verificabilidade dos sistemas digitais no diagnóstico e terapêutica;
- Diluição da responsabilidade;
- Alteração do consentimento informado em contexto de IA;
- Menor proteção do direito à privacidade individual;
- Surgimento de novas barreiras pela iliteracia digital;
- Excessiva dependência da robótica inteligente.

⁵⁴ European Parliamentary Research Service. (2022). Artificial intelligence in healthcare. <https://doi.org/10.2861/568473>

⁵⁵ Office of Science and Technology Policy. (2022). Blueprint for an AI Bill of Rights: Making automated systems for the American people. (October), 1-73. <https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf>



5.3. Administração da saúde, gestão hospitalar e (o impacto da) intervenção à distância (telessaúde)

A IA tem o potencial de mudar profundamente o modo como se administra a saúde e se implementam as políticas públicas, nomeadamente no que se refere à gestão das organizações da saúde tal como os hospitais e centros de saúde. A nível macro, a IA pode ser um importante instrumento no planeamento da saúde, quer a nível dos cuidados de saúde, quer da saúde pública, incluindo a projeção das políticas de alocação de recursos em um setor tão complexo e diversificado. Ao nível das unidades de saúde, pode ser especialmente relevante na marcação integrada de consultas, na otimização dos registos de saúde eletrónicos, na promoção da segurança das pessoas doentes, na gestão de reclamações, no pagamento aos profissionais de saúde e fornecedores, na faturação, na contabilidade analítica, entre outros.

De facto, a IA pode ter um impacto significativo na: a) integração e coordenação da prestação de cuidados e otimização do acesso à saúde, b) gestão operacional das unidades de saúde, c) prestação de cuidados de saúde, designadamente através da telessaúde e outras formas de e-saúde, d) gestão integrada do registo de saúde eletrónico e da prescrição médica.

No plano da **integração e coordenação da prestação de cuidados e otimização do acesso à saúde**, aliás umas das principais falhas do sistema de saúde português, a IA pode contribuir para uma eficaz integração dos diferentes níveis de cuidados. Por exemplo, otimizando a rede de referenciação do Serviço Nacional de Saúde (SNS), quer a nível regional ou no plano nacional, permitindo que esta rede funcione como um sistema de vasos comunicantes, verdadeiramente articulado, e com capacidade de resposta em tempo real. Esta questão é central para o desenvolvimento futuro do sistema de saúde, seja no SNS, seja na sua articulação funcional com os setores privado e social. Isso suscita a criação e gestão de uma rede digital da saúde (*intranet*) verdadeiramente eficaz, com o potencial de expansão aos sistemas de saúde dos outros Estados-Membro da União Europeia, o que implicará que esteja assegurada a conectividade e interoperabilidade dos diferentes sistemas digitais ⁵⁶.

Sendo a integração dos diferentes níveis de cuidados um elemento central para uma gestão eficaz deste setor, e para garantir o acesso de todos à saúde, torna-se por demais evidente a sua premência em emergências de saúde pública, tal como a pandemia pela COVID-19. A recente transformação do SNS em uma rede de Unidades Locais de Saúde (ULS) enfatiza ainda mais a necessidade de integração sistémica e de articulação eficaz.

Por outro lado, o acesso equitativo à saúde é um dos temas mais complexos nos cuidados de saúde modernos, devido à inevitável escassez de recursos disponíveis e às necessidades crescentes dos mesmos ⁵⁷. Diferentes medidas são necessárias para

⁵⁶ World Health Organization (WHO). (2021). *Ethics and governance of artificial intelligence for health*. Retrieved from <http://apps.who.int/b>

⁵⁷ Nunes R. (2022). *Healthcare as a universal human right: Sustainability in global health*. Routledge, New York.



otimizar o acesso em condições de equidade de todos os cidadãos, como a implementação de novas modalidades de gestão das unidades de saúde de base empresarial, ou a priorização ética dos bens e serviços que podem ser efetivamente prestados a todos os cidadãos. A IA promete fazer decrescer os custos associados: estes diminuem à medida que a tecnologia se desenvolve e banaliza, não obstante, numa primeira fase, a realidade ser contrária, registando-se um aumento de custos, também derivado dos investimentos estruturais iniciais. Paralelamente, importa ter em conta o risco de o crescimento contínuo do poder da tecnologia vir a aprofundar assimetrias já existentes e criar novas, como as que separam quem tem acesso ao digital e quem não tem, aumentando assim também a vulnerabilidade das pessoas. É fundamental que a IA esteja disponível para todas as pessoas, bem como o desenvolvimento da literacia em tecnologias de saúde na perspetiva do utilizador, para que todos/as tenham as competências necessárias para a compreenderem e para promover uma boa relação com a mesma.

Ainda no âmbito do acesso à saúde, destaca-se a ameaça à sustentabilidade do próprio sistema de saúde, especialmente nos países com economias mais desenvolvidas, com uma população envelhecida, com aumento de doenças crónicas, a par da escassez de profissionais de saúde para as reais necessidades existentes. Neste sentido, a IA tem vindo a ser apresentada como uma promessa de contributo para uma melhor gestão dos serviços de saúde, nomeadamente também dos recursos humanos, reduzindo mesmo a sua intervenção em algumas áreas. Um exemplo relevante é o dos elevadíssimos investimentos em aplicações de IA, incluindo com robótica, o que, por sua vez, aumenta a pressão na sua utilização⁵⁸. Será assim necessário um reforço do sistema de regulação, de modo a evitar fenómenos de indução da procura pela oferta.

No que respeita à **gestão operacional das unidades de saúde**, por exemplo no atinente à gestão hospitalar, a IA é altamente promissora, nomeadamente em áreas como a gestão de recursos humanos, a elaboração de escalas, a aquisição de medicamentos e dispositivos médicos, etc. A IA pode também otimizar o *software* de gestão clínica com programas de contabilidade financeira adequados, ajudando os administradores hospitalares a melhor equilibrar os custos e a encontrar novas oportunidades de gerar ganhos de eficiência. A IA e a *machine learning* podem contribuir para que os sistemas de gestão clínica sejam mais eficientes, autónomos e funcionais.

Neste contexto, a codificação é um processo fundamental para os fluxos de informação que permitem a gestão hospitalar, tanto na perspetiva da eficiência das operações como, também, em termos de desempenho otimizado. A codificação influencia os processos de gestão hospitalar, que vão desde o agendamento de atividades clínicas à faturação (incluindo questões de deteção de fraude), à gestão de *stocks* ou mesmo à gestão de fluxos de transporte. A codificação, sendo um processo complexo e

⁵⁸ Stahl B. (2021). *Artificial intelligence for a better future. An ecosystem perspective on the ethics of AI and emerging digital technologies*, Springer, Cham, Switzerland.



demorado, tem progressivamente sido feita com recurso a aplicações de IA, contribuindo para a libertação de recursos humanos para outras tarefas.

Como em diferentes aplicações de IA, a qualidade dos dados para o rigor da codificação é o principal desafio, especialmente tendo em consideração aqueles que resultam de processos de telemonitorização. Por outro lado, atendendo ao complexo processo de gestão de fluxos informacionais no patamar da gestão hospitalar, para lá da codificação, são ainda implementadas aplicações de IA no sentido de assistir na alocação de recursos, cadeias de fornecimento e gestão de processos de pessoas doentes. Diversamente do que sucede com as aplicações de IA usadas no contexto da codificação, as aplicações de alocação de recursos integram o processo decisional de gestão. Tratando-se de recursos em saúde, deverá existir particular atenção à manutenção do humano *in the loop* (no circuito do fluir da informação), ou seja, como elemento fundamental do processo decisional e na governação dos elementos essenciais do processo de gestão, de modo a não permitir que sejam fornecidos resultados (*outputs*) que possam questionar a equidade da respetiva alocação, em especial no contexto de recursos escassos.

Em termos de gestão de fluxos organizacionais, são de referir os *autonomous mobile robots* (AMR) ⁵⁹, que permitem manter o controlo (e conseqüente reposição) de *stocks* de equipamentos e de medicamentos, capitalizando os dados recolhidos, também através de telessaúde. Os AMR podem ainda ser utilizados na higienização dos espaços, a velocidades superiores à realizada por funcionários hospitalares ou em áreas fortemente contagiosas, como se verificou em unidades de internamento de pessoas com COVID-19; são igualmente utilizados para a entrega de material (p. ex., roupa) em enfermarias de doenças infecciosas ou, nestes casos, mesmo para distribuição de medicamentos, aumentando a segurança e otimizando os recursos humanos para a pessoa doente. Outros exemplos poderiam incluir o acolhimento e prestação de informações em unidades hospitalares pelos designados e já referidos robôs sociais, funcionando ainda como rececionistas móveis. A sua permanente disponibilidade torna-os bastante vantajosos, particularmente em funções em que os recursos humanos escasseiam.

A aplicação da IA à gestão hospitalar pode, então, contribuir para uma maior eficiência dos serviços, resultando numa mais elevada produtividade a mais baixos custos, promovendo a sua sustentabilidade. A modernização da gestão hospitalar, por via da digitalização, também reduz a burocracia e a intermediação nos processos, agilizando os serviços. Simultaneamente, importa ter em atenção o impacto de alterações disruptivas na coesão das equipas, na sua comunicação interna e também na comunicação personalizada com os utentes dos serviços.

⁵⁹ Constantinescu M, Crisp R. (2022). Can Robotic AI Systems Be Virtuous and Why Does This Matter? *International Journal of Social Robotics* 14: 1547-1557 <https://doi.org/10.1007/s12369-022-00887-w>



Também no domínio dos seguros e planos de saúde, a IA tem um importante papel a desempenhar no que respeita à possibilidade de os prestadores efetuarem pedidos de cobertura de um modo eficaz e diligente, ajudando as seguradoras a adjudicar esses pedidos. Isso mesmo poderá tornar a dinâmica seguradora mais rápida, mas também mais justa e equitativa.

Na prestação de cuidados de saúde, designadamente através da telessaúde e outras formas de e-saúde, a aplicação da IA destaca-se pela possibilidade de uma intervenção clínica à distância. A adoção de mecanismos de telessaúde é vista como um fator de potenciação de equidade no acesso à saúde, na medida em que permite uma gestão eficaz e uma prestação contínua – sem limitações geográficas e inclusiva de cidadãos com menor mobilidade – de cuidados de saúde⁶⁰.

A telessaúde, na articulação da IA com as tecnologias da informação e comunicação (TIC), será talvez a vertente mais visível para o cidadão da digitalização dos serviços de saúde, permitindo a criação e disponibilização de novos serviços (teleconsulta, e-prescrição, aplicações), novos métodos de comunicação entre profissionais (remotos) e a facilidade de automonitorização, como já foi referido anteriormente, e mesmo de diagnóstico individual. A telessaúde, alargando os modos de relacionamento entre os profissionais de saúde e a pessoa doente, é uma das vertentes em que se perspetiva um maior investimento nos próximos anos ⁶¹.

Realidades como o telerrastreio e a telemonitorização remota são particularmente relevantes quanto a doentes crónicos, permitindo um acompanhamento de proximidade e uma melhor gestão de atividades clínicas, libertando, por exemplo, o recurso a serviços de urgência. Por outro lado, a telemonitorização permite ainda uma recolha contínua de dados dos doentes (crónicos), possibilitando um acompanhamento não apenas da pessoa doente, mas também da doença. As equipas multidisciplinares de monitorização acompanham os dados enviados pela pessoa, encaminhando para o médico especialista em casos justificados, o que concede às pessoas doentes uma maior qualidade de vida pela diminuição de idas à consulta, mas garantindo a proximidade no acompanhamento médico. Os dados recolhidos através destes instrumentos, mas também dos registos clínicos, relevam para o processo de codificação de dados, enquanto extração e sistematização dos mesmos através de sistemas de classificação. A codificação permite agilizar a informação disponível e a sua utilização⁶². Em todo o caso, a segurança do sistema, e a transparência, são também fundamentais para garantir a confiança dos cidadãos.

⁶⁰ OMS. (2016). Global diffusion of eHealth: making universal health coverage achievable. *Report of the third global survey on eHealth*. [Internet]. Geneva: World Health Organization.

⁶¹ Cf., p. exemplo, a Estratégia Nacional para o Ecossistema de Informação de Saúde 2020: <https://files.dre.pt/1s/2016/10/19900/0373503738.pdf> 34

⁶² Wübbeler M, Geis S, Stojanovic J, Elliott L, Gutierrez-Ibarluzea I and Lenoir-Wijnkoop I. (2021). Coding public health interventions for health technology assessments: A pilot experience with WHO's International Classification of Health Interventions (ICHI). *Front. Public Health* 9:620637. <https://doi.org/10.3389/fpubh.2021.620637>



Na prestação de cuidados de saúde, como já se referiu anteriormente, o potencial de expansão da IA é enorme devido a uma multiplicidade de áreas em que poderá utilizada - desde o diagnóstico, à cirurgia robótica, ou mesmo às múltiplas aplicações digitais existentes já hoje no mercado - tornando-se fundamental que, no plano ético, o princípio do respeito pela pessoa e pela sua autonomia não seja subtilmente condicionado pela IA. Isto é, que o paternalismo médico tradicional não seja substituído por novas formas de paternalismo digital que condicionem a liberdade de decisão e subvertam a natureza transversal que se pretende para a relação entre o profissional de saúde e a pessoa doente. Mais uma vez, a aquisição de competências e de literacia sobre IA pela generalidade da população é fundamental para que também na saúde esta evolução seja coroada de sucesso.

Na gestão integrada do registo de saúde eletrónico (RSE) e da prescrição médica a IA tem um enorme potencial de alteração dos registos clínicos e em diferentes dimensões. Por um lado, a IA generativa, por exemplo o GPT, permite já uma maior eficácia e rapidez na elaboração das notas médicas que serão registadas no RSE ⁶³. Aplicações dirigidas por IA conseguem captar as notas ditadas pelos médicos, sintetizando as entrevistas clínicas com exames radiológicos e exames laboratoriais. Mais ainda, sem intervenção direta do médico, as mais modernas aplicações analisam a interação entre o médico e a pessoa doente e “produzem” uma nota médica original para o RSE. Pressupõe-se que exista consentimento da pessoa para a sua elaboração e um pedido expresso do médico para este efeito. Antevê-se que a IA possa ainda sugerir proativamente determinada prescrição médica ou exames analíticos, incluindo o *rationale* da prescrição. Não se trata, nesta fase, de uma tecnologia isenta de problemas, pelo que é fundamental que exista a possibilidade técnica de lidar adequadamente com “alucinações”, omissões e erros destas aplicações. Ainda assim, antevê-se uma profunda alteração no modo como se elabora o registo de saúde eletrónico, com profundas implicações no plano ético e mesmo médico-legal.

Após um adequado registo das notas clínicas no RSE, a evolução futura do tratamento dos dados de saúde das pessoas doentes, nos planos individual e coletivo, antecipa-se como fortemente inovadora, dada a possibilidade de se constituírem plataformas digitais que consolidem toda a informação relacionada com as pessoas inseridas no sistema. Este conceito de *health wallet* - carteira de saúde - pretende redefinir o modo como se processa o armazenamento digital de dados de saúde, podendo revolucionar a prestação de cuidados de saúde ⁶⁴.

Esta consolidação dos dados de saúde inclui, não apenas o tradicional registo eletrónico efetuado pelos médicos e outros profissionais de saúde, mas todos os dados que estejam conectados *online*, como exames de imagiologia, análises clínicas, dados

⁶³ Lee P, Bubeck S, Petro J. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. (2023). *New England Journal of Medicine* 388: 1233-1239. DOI: 10.1056/NEJMra2301725

⁶⁴ Zanetti, JK, Nunes R. (2023). To wallet or not to wallet. The debate over digital health information storage. *Computers* 12, 114. <https://doi.org/10.3390/computers12060114>



colhidos das mais diversas aplicações digitais, dados de seguradoras, etc. As vantagens desta carteira de saúde são óbvias, nomeadamente pela centralização de dados ou pela facilidade de acesso, importando assegurar, também, a sua fiabilidade e o respeito pela autonomia e privacidade das pessoas doentes.

Entre outras vantagens já conhecidas encontra-se a possibilidade, sobretudo no espaço da União Europeia, de fácil acessibilidade aos dados, tendo em conta a existência de regimes jurídicos diferentes para algumas práticas clínicas. Casos ilustrativos são o da transplantação de órgãos ou o das diretivas antecipadas de vontade, e a consequente compatibilização internacional, respetivamente, do Registo Nacional de Não Dadores (RENDA) ou do Registo Nacional do Testamento Vital (RENTEV). A criação do Espaço Europeu de Dados de Saúde visa facilitar esta evolução⁶⁵.

Esta carteira de saúde consolidada contribuirá também para a implementação de uma medicina personalizada de um modo mais eficaz, dado ser possível uma colheita compreensiva dos dados das pessoas doentes, incluindo os dados genéticos. A interseção de dados genéticos, pessoais e do grupo em análise, permitirá conceber novas estratégias terapêuticas personalizadas, incluindo o recurso à fármaco-genética. A IA será então decisiva para uma medicina de precisão ajustada a cada pessoa doente, com evidentes benefícios no plano da prevenção e terapêutica. Mais uma vez, será importante acautelar os direitos fundamentais das pessoas e assegurar o acesso universal à medicina de precisão, para que não se aprofundem iniquidades e agravem vulnerabilidades.

A carteira consolidada de dados de saúde possibilitará, também, reduzir o erro médico, sobretudo se associada à IA generativa na análise destes dados e na interpretação das notas clínicas e exames subsidiários⁶⁶. Mais ainda, no decurso de processos médico-legais permitirá também uma maior facilidade de acesso aos dados constantes no RSE, bem como a sua disponibilização aos diferentes intervenientes no processo. De salientar a enorme relevância da informação constante nos registos de saúde em todo o processo pericial médico-legal, nomeadamente para efetuar prova e determinar a culpa.

Por seu turno, estas ferramentas podem também traduzir-se na capacitação dos cidadãos, que passam a ter maior controlo sobre a sua informação pessoal, permitindo tomar decisões mais informadas e esclarecidas. A IA é também um importante recurso para melhorar a experiência global da pessoa doente no sistema de saúde. Por exemplo, o Alder Hey Children's Hospital em Liverpool tem colaborado com o supercomputador IBM Watson para criar um "hospital cognitivo"⁶⁷, incluindo o recurso a uma aplicação ("app") que facilita a interação com as pessoas doentes. Esta aplicação pretende identificar a ansiedade da pessoa antes da consulta e fornecer informação em tempo

⁶⁵ Comissão Europeia. European Health Data Space. https://health.ec.europa.eu/ehealth-digitalhealth-and-care/european-health-data-space_en

⁶⁶ Mittelstadt, B. (2021). *The impact of artificial intelligence on the doctor-patient relationship*. Retrieved from <https://www.coe.int/en/web/bioethics/report-impact-of-ai-on-the-doctor-patient-relation>

⁶⁷ Nuffield Council on Bioethics. (2018). *Artificial intelligence (AI) in healthcare and research*. London.



real aos médicos ajudando, assim, a proporcionar um tratamento mais apropriado. Campanhas de educação para a saúde e de literacia digital neste domínio permitirão também aumentar a consciência cívica dos cidadãos, nomeadamente sobre a importância de proteger a sua privacidade pessoal.

Em síntese, antevê-se uma assinalável evolução no modo como os dados de saúde são gerados e organizados, e a forma como serão disponibilizados em tempo real. Porém, de acordo com o Parlamento Europeu, existem riscos associados à IA na medicina e na saúde que devem ser ponderados: a) danos causados à pessoa doente devido a falhas da IA, b) uso deficiente das ferramentas de IA, c) viés na IA e perpetuação das iniquidades existentes, d) falta de transparência, e) problemas de privacidade e segurança, f) falhas de *accountability*, e g) obstáculos na sua implementação⁶⁸.

Existem, assim, importantes desafios de governação ética e de regulação jurídica no plano da administração e da gestão da saúde, designadamente no que se refere à proteção de dados, à investigação em seres humanos e à prestação de cuidados. Por isso mesmo, é previsível o surgimento de uma nova área de regulação especificamente relacionada com a IA, de modo que esta seja considerada e percebida como transparente, responsável e compatível com o interesse público, deixando sempre espaço para a uma verdadeira interação humana.

⁶⁸ European Parliamentary Research Service. (2022). *Artificial intelligence in healthcare. Applications, risks, and ethical and societal impacts* <https://doi.org/10.2861/568473>



Aspetos Éticos Relevantes da IA na Administração e Gestão da Saúde

- A IA deve contribuir para uma plena integração e coordenação da prestação de cuidados e otimizar o acesso à saúde;
- Deve evitar-se o aprofundamento das assimetrias existentes e as pessoas mais vulneráveis devem ser protegidas;
- Na gestão operacional das unidades de saúde, a IA deve garantir o respeito pela autonomia e privacidade das pessoas, mantendo sempre uma relação ética e humana entre os profissionais de saúde e as pessoas doentes;
- A telessaúde e outras formas de e-saúde devem pautar-se pelo estrito cumprimento das regras da medicina baseada na evidência, e pelo contacto humano e personalizado com a pessoa doente;
- A gestão integrada do registo de saúde eletrónico tem o enorme potencial de reforçar a autonomia da pessoa doente e de disponibilizar em tempo real toda a sua informação e dados de saúde, através de uma carteira consolidada de dados de saúde;
- Deve garantir-se que a prescrição médica resultante de uma gestão integrada na interface com a IA generativa é fiável, verificável, e que respeita a autonomia das pessoas doentes.

5.4. Administração da saúde pública e (o impacto da) codificação de dados

A IA tem vindo a desenvolver uma intervenção cada vez mais preponderante também ao nível da administração pública, sendo um dos objetivos da Agenda 2030 que Portugal se torne um laboratório vivo de experimentação de novos desenvolvimentos na área da saúde⁶⁹.

O uso de IA na administração da saúde pública, sobretudo em fenómenos de saúde global, pode transformar o modo como os prestadores de serviços de saúde identificam e respondem a riscos de saúde e tomam decisões neste contexto, na medida em que a IA também pode ser usada para ajudar a **identificar e monitorizar surtos em tempo real**. Tal é particularmente importante na medida em que, atendendo à escassez de recursos no âmbito dos cuidados de saúde, a deteção precoce de um surto de uma doença pode contribuir para a prevenção da sua disseminação para outras áreas.

Com efeito, a administração e gestão da Saúde Pública é uma das áreas em que o investimento em IA mais se tem feito notar, exponenciado pela pandemia COVID-19. Neste contexto preciso, exige-se a possibilidade de deteção precoce de surtos em uma população determinada, além de uma estratificação do risco (de saúde) em si, tanto em

⁶⁹ Cf. a Estratégia Nacional da Inteligência Artificial. <https://portugaldigital.gov.pt/accelerar-atransicao-digital-em-portugal/conhecer-as-estrategias-para-a-transicao-digital/estrategianacional-de-inteligencia-artificial/>



termos temporais, como em termos geográficos. Daqui também a relevância da **vigilância epidemiológica**, para o que já existem aplicações de IA com capacidade de identificação de áreas geográficas de prevalência de doença e da construção de mapas da sua expansão, ou de comportamentos de risco, através da análise conjugada de fontes (sensores, telemóveis, redes sociais, *media*), em combinação com relatórios clínicos – aspetos que podem ter relevância em termos de saúde global, ainda que tais aplicações devam ser ponderadas com as questões da privacidade e autonomia individual dos cidadãos.

Um outro aspeto em que a IA pode fazer uma diferença decisiva é na compreensão das causas subjacentes à doença, através do reconhecimento de padrões e fatores de risco, permitindo o desenho de estratégias de prevenção mais eficazes. Assim, no patamar da Saúde Pública, de enorme mais-valia é a possibilidade de **implementação de análise preditiva** para identificação de pessoas em risco de desenvolvimento de doenças crónicas, permitindo uma intervenção precoce direcionada (desde o aconselhamento à medicação) e mais eficaz, em termos de custos, para os sistemas de saúde.

Os sistemas de IA podem ainda contribuir para a redução de iniquidades na prestação de cuidados de saúde, em termos de acesso à saúde – i.e., os algoritmos de IA podem contribuir para a identificação de áreas onde certas populações se encontram deficientemente servidas em termos de cuidados de saúde.

Contudo, em termos de Saúde Pública, o maior desafio, ultrapassada a questão da recolha de dados, não é o “se”, mas antes o “como” da sua codificação. Também aqui a IA se encontra a revolucionar o processo de codificação, na medida em que permite a **automação da codificação** e a diminuição de erros. As ferramentas de codificação movidas por IA (como, por exemplo, a CODIO) usam sistemas de processamento de linguagem e algoritmos de aprendizagem para analisar registos clínicos. Estas ferramentas executam a tarefa de forma bastante mais rápida e muito menos dispendiosa do que se feita apenas por humanos. Contudo, a automação da codificação acarreta alguns desafios, não apenas no patamar da privacidade e segurança, dada a utilização de dados de saúde, mas no contexto dos casos complexos, em que podem ocorrer dificuldades na interpretação dos dados que não seguem o padrão, pelo que deve ser sempre garantida a supervisão humana. Por outro lado, as normas de codificação – para lá da legislação aplicável – mudam, devendo a codificação necessariamente cumprir as normas aplicáveis, exigindo-se, assim, capacidade de adaptação e, também, auditorias e supervisão de cumprimento. Finalmente, talvez um dos desafios mais complexos possa ser o da **integração dos sistemas de codificação nas tecnologias de informação e comunicação da saúde**, para permitir a partilha de dados e a interoperabilidade dos sistemas. Daí que um dos elementos essenciais seja a uniformização de normas de codificação, dado que, não podendo a Saúde Pública continuar a ser pensada em termos territoriais num mundo globalizado, a informação que circula tem de ter o mesmo significado, ou seja, tem de ser padronizada – quer a nível nacional, quer ao nível da União Europeia, e mesmo no plano global, em estreita



articulação com a Organização Mundial da Saúde. Por outras palavras, sem sistemas de classificação internacionalmente reconhecidos, não pode ser garantida a fiabilidade da informação. Assim, sem um padrão global na codificação, como que um sistema internacional de classificação (como, p. ex., o *International Classification of Diseases*⁷⁰) que possa ser aplicável em sistemas de IA de codificação, a utilização dos dados por aplicações de IA pode conduzir a divergências de abordagem na prevenção e contenção, conduzindo também a possíveis questões de transparência e, mesmo, de discriminação e democraticidade na abordagem a crises de saúde global.

A capacidade preditiva da IA, existindo um **sistema de classificação harmonizado em termos internacionais** e que tenha preocupações de direitos humanos na sua formulação, poderá ajudar na preparação ou antecipação de futuras emergências de Saúde Pública. Pensando em atuais benefícios de aplicações de IA, como a conjugação de dados para envio de mensagens automatizadas para populações ou grupos populacionais (*micro-targeting*) que atuem em termos preventivos, de que já existiu o exemplo a recente experiência com a COVID-19, a já anunciada previsão de futuras pandemias depende, assim, de uma codificação de dados harmonizada, ao que deverá acrescer um quadro de governança comum, na qual seja incluído o "*human in the loop*". Acrescente-se, neste âmbito, a menção aos riscos de parcialidade por parte dos algoritmos, associados aos vieses que poderão estar subjacentes aos dados a que recorrem para promover as suas decisões. Um outro risco é o de decisões automatizadas baseadas na caracterização de perfis (*profiling*), que não têm em conta a informação contextual das pessoas doentes, não obstante existirem alguns instrumentos normativos (por exemplo, o RGPD ou o AI Act) que procuram mitigar esses riscos, na medida em que tais decisões não podem ser consideradas quando tenham impacto significativo na esfera do(s) cidadão(s)⁷¹.

Aliás, questões relativas à proteção e segurança dos dados de saúde também se colocam com muita acuidade no plano da administração pública, sobretudo quando eles aumentam exponencialmente e são cada vez mais cobijados pela indústria do *marketing*. Importa, então, rentabilizar os dados de saúde da pessoa em termos de Saúde Pública, ponderando as alternativas que se colocam para evitar graves conflitos de interesses que hoje se colocam, por exemplo, entre as empresas privadas que produzem ou tratam os dados, mas que visam o lucro, e o Estado, que é responsável pela sua utilização no melhor interesse dos seus legítimos proprietários, mas também da sociedade em geral. Tem-se proposto a criação de sindicatos de dados, geridos por pessoas eleitas democraticamente, ainda que seja difícil a sua operacionalização. Destacam-se, assim, as questões relacionadas com a privacidade, com a identidade e com a equidade e, também, com a segurança. Daí que um dos grandes desafios à IA a partir da perspetiva dos riscos seja, justamente, a **cibersegurança e a qualidade dos**

⁷⁰ Atualmente, o International Standard for Medical Coding segue o ICD-11, que conta com aproximadamente 55.000 códigos (<https://icd.who.int/en>); o anterior (ICD-10) tinha (apenas) 14.000 códigos.

⁷¹ Veja-se ainda o art. 9º, n.º 2, da Carta Portuguesa de Direitos Humanos na Era Digital.



dados. Com efeito, na prestação de cuidados de saúde com recurso a sistemas e aplicações de IA podem ocorrer diversas circunstâncias, tais como o acesso, disponibilização e transmissibilidade de dados de saúde que podem gerar preocupações de privacidade.

Ainda no patamar dos dados, o tratamento dos mesmos por sistemas de IA – atendendo ao facto de os conjuntos de dados usados para treinar, testar e validar modelos de IA serem muitas vezes insuficientemente representativos do público em geral – pode gerar resultados enviesados. Os riscos de enviesamento implicam que possam existir diagnósticos incorretos ou previsões inadequadas, afetando negativamente os resultados e a própria confiança na administração da saúde pública. Uma das formas de mitigar os riscos de viés – para além da garantia da qualidade dos dados – é a inclusão de pessoas doentes (e, eventualmente, outros interessados) no **processo de desenho e implementação de sistemas de IA**, o que poderá garantir que tais sistemas incluam tanto as necessidades, como os valores das pessoas doentes. Neste contexto, é de salientar a iniciativa da UE de um Espaço Europeu de Dados de Saúde (*European Health Data Space*), apresentado como um ecossistema específico para a saúde composto por regras, normas e práticas comuns, infraestruturas e um quadro de governação que visa capacitar os indivíduos através de um maior acesso digital e controlo dos seus dados pessoais (eletrónicos) de saúde, nacional e internacionalmente. Paralelamente, pretende-se o apoio à sua livre circulação, promovendo um verdadeiro⁷² mercado único para sistemas de registos de saúde eletrónicos (RSE) e funcionando de acordo com o princípio de dados FAIR (*Findable, Accessible, Interoperable, Reusable* - Localizáveis, Acessíveis, Interoperáveis, Reutilizáveis)⁷³. Um dos caminhos para a garantia do princípio de dados FAIR é, também, a encriptação de dados e a limitação de acesso.

Todavia, um dos riscos na geração de dados prende-se com um fenómeno designado de **colonização de dados**, i.e., o desequilíbrio entre os que adquirem, acumulam, analisam e controlam os dados e os que fornecem os dados e têm pouco controlo sobre o seu uso. Em saúde global, a aquisição, codificação e gestão dos dados devem obedecer a um padrão (verificável) num contexto geral que permita o respeito pelos direitos fundamentais dos cidadãos⁷⁴.

Mas um dos aspetos mais delicados, para além das questões relacionadas com dados, prende-se com a antecipação de resultados face à concretização de resultados em termos de **contributo efetivo** para a Saúde Pública. Com efeito, as potencialidades dos sistemas de IA neste contexto dependem diretamente da capacidade e aptidão dos utilizadores desses mesmos sistemas. Existem já estudos que mostram que os

⁷² Comissão Europeia. European Health Data Space. https://health.ec.europa.eu/ehealth-digitalhealth-and-care/european-health-data-space_en

⁷³ No contexto dos dados, ainda fundamental pensar nos designados 10 V's dos dados: velocidade, vocabulário, validade, veracidade, volume, valor, variedade, variabilidade, vagueza/vaquidade, origem: Cfr. Panesar, A. Machine Learning and AI for Healthcare, Coventry, 2019.

⁷⁴ Comissão Europeia. (2022). Study on Health Data, Digital Health and Artificial Intelligence in Healthcare, 2022, doi: 10.2875/702007



parâmetros usados (em situações semelhantes) podem conduzir a resultados distintos. Nesse sentido, a aptidão de uso e de parametrização para a codificação são condições essenciais para o contributo positivo dos mesmos, realçando-se a importância do seu uso ético, bem como da formação.

Coloca-se ainda a questão da integração e interoperabilidade dos sistemas com as infraestruturas técnicas e com os respetivos fluxos de trabalho, que poderão apresentar diferenças entre si, carecendo de harmonização, não obstante os custos dessa mesma harmonização que se apresenta, aqui, como condição sem a qual a operacionalização dos sistemas não pode ter lugar. Assim, é exigível uma fase de investimento necessário e de disponibilidade de recursos, até que possa gerar-se o equilíbrio para a subsequente eficiência de custos.

Aspetos Éticos Relevantes da IA na Saúde Pública

- O uso de IA na administração da saúde pública pode ajudar a identificar e monitorizar surtos em tempo real e a efetivar vigilância epidemiológica através de análise preditiva, conquanto tal monitorização garanta a privacidade e a segurança dos dados;
- A automação da codificação de dados deve permitir a integração nas tecnologias de informação e comunicação de saúde, devendo ser garantida a cibersegurança dos sistemas;
- Devem ser prevenidos fenómenos de colonização de dados através da criação de sistemas de armazenamento que obedeçam ao princípio do interesse público;
- Os processos de automação apenas podem constituir um contributo efetivo para a saúde pública se for desenvolvida a literacia digital pelos prestadores.

5.5. Ensino e educação em saúde e (o impacto da) realidade virtual

A IA desempenha um papel cada vez mais relevante no plano da formação de profissionais de saúde e no da educação para a saúde dos cidadãos. O ensino de profissionais de saúde beneficia, hoje, de alguns recursos formatados pela IA que se vêm desenvolvendo rapidamente, como pode ser exemplo a realidade virtual (RV) desenhada para a prática clínica, promovendo a proficiência técnica em diversas áreas e o treino de diferentes formas de intervenção. A RV permite criar cenários de treino desenhados por computador, altamente realistas, que simulam a presença da pessoa num cenário específico, sentido como real, mas sem expor as pessoas doentes a riscos



e manipulações desnecessárias⁷⁵, o que é útil em qualquer área da saúde, mas ainda mais na cirurgia ou em outras áreas que impliquem intervenções mais invasivas.

Complementarmente, a IA permite a personalização do ensino, promovendo uma adaptação às necessidades individuais dos estudantes no sentido de otimizar o processo de aprendizagem. Existe, contudo, um enorme desafio relacionado com a qualidade dos cuidados de saúde. É crucial garantir que a formação mediada por tecnologia não comprometa a capacidade dos profissionais de saúde em promover cuidados humanizados, que considerem as dimensões psicológicas e sociais da saúde. É reconhecida a importância da relação entre os profissionais de saúde e as pessoas doentes na promoção da confiança e da adesão terapêutica, assim como os efeitos positivos que essa mesma confiança, no profissional e na intervenção proposta, desempenha ao nível da satisfação com as intervenções. Deste modo, a associação de sistemas que promovem conversas automáticas com as pessoas (*chatbots*) a realidade virtual poderá constituir-se como uma forma de representar situações mais próximas da realidade, interativas, com muito maior controlo sobre as variáveis a utilizar, substituindo-se ao tradicional *role-playing* no treino de profissionais, isto nas mais variadas áreas de intervenção em saúde. Ainda que, até ao momento, tenham sido poucos os protocolos baseados em evidência científica que sustentem este tipo de treino e intervenção por parte dos profissionais⁷⁶, existem poucas dúvidas de que o recurso a este tipo de tecnologia tenderá a aumentar.

É, por isso, imperativo investir na investigação a este nível, facilitando os processos de aprendizagem e garantindo um muito maior número de experiências que sustentem a aquisição de competências úteis na promoção de relações de confiança entre os profissionais e as pessoas. Será ainda fundamental evitar o reforço de vieses. Os doentes virtuais deverão ser representativos da enorme diversidade humana e os cenários terão de ser baseados em evidências médicas sólidas. A transição entre a experiência tecnológica e a realidade deverá ainda ser sempre mediada por profissionais experientes, que promovam o espírito crítico dos estudantes em relação às próprias experiências adquiridas no treino inicial. Deste modo, aumentará a probabilidade de se conseguir uma melhoria no treino dos profissionais com um menor risco para as pessoas doentes, colocando a tecnologia, de facto, ao serviço das pessoas.

Com a evolução de uma filosofia de ensino que visa promover a autonomia do estudante e a personalização das aprendizagens, é desejável que o professor seja cada vez mais um mediador do que um detentor do conhecimento, uma vez que a informação está hoje facilmente disponível através dos algoritmos de busca que, bem desenhados,

⁷⁵ Maples-Keller, J. L., Bunnell, B. E., Kim, S. & Rothbaum, B., O. (2017). The use of virtual reality technology in the treatment of anxiety and other psychiatric disorders. *Harvard Review of Psychiatry*, 25(3), 103-113. <https://doi.org/10.1097/HRP.000000000000138>

⁷⁶ Rizzo, A. S., & Koenig, S. T. (2017). Is clinical virtual reality ready for primetime? *Neuropsychology*, 31(8), 877-899. <https://doi.org/10.1037/neu0000405>



poderão promover um mais fácil acesso e com maior diversidade. O papel do professor deverá ser o de promotor de um pensamento crítico e reflexivo em relação à qualidade da informação. Deste modo, e valorizando-se cada vez mais a disponibilidade do agente mediador, a utilização de *chatbots* e de outros assistentes digitais tenderá a crescer, uma vez que estes estarão sempre disponíveis e são imunes à fadiga. Do mesmo modo, sabendo que a imagem e a modelagem, com o seu papel relevante no envolvimento emocional, são fatores centrais nas aprendizagens e sedimentação de conceitos, a realidade virtual mediada por *chatbots* terá um papel privilegiado no ensino e na formação profissional.

Importa, por isso, promover a formação dos professores na utilização deste tipo de recursos, de modo que estes possam passar a desempenhar um papel efetivo de mediadores entre os estudantes e o acesso à informação e ao treino profissional. A adoção eficaz desta tecnologia, seja na formação seja na prática clínica, exige um certo nível de preparação tecnológica entre os profissionais de saúde. Infelizmente, a maioria dos profissionais da área de saúde não dispõe de formação adequada a este nível, pelo que podem não estar preparados para integrar plenamente estas ferramentas nas suas práticas. O desenvolvimento de competências a este nível é fundamental, seja no ensino pré-graduado, seja na formação contínua dos profissionais. A falta de competência tecnológica não só impede a eficácia da prática, como aumenta o risco de erros que podem ter consequências graves para os beneficiários dos serviços.

Um outro desafio que se coloca com a introdução da IA no contexto académico está associado ao aumento da complexidade no que diz respeito ao sistema de avaliação e à deteção de fraudes. Se o advento da Internet promoveu um aumento objetivo de recursos para os estudantes e para os docentes, a que se associou um incremento na capacidade na deteção de plágios e fraudes, o desenvolvimento de algoritmos e de *chatbots* de linguagem generativa vem colocar questões de maior complexidade. A possibilidade de fraude académica torna-se uma preocupação crescente. Esta tecnologia possibilita a elaboração de trabalhos académicos aparentemente originais que desafiam as ferramentas de deteção de fraude. Permite, inclusivamente, replicar um determinado estilo de escrita, plagiar documentos alterando a forma e o estilo e mesmo gerar conteúdos com base em trabalhos anteriores do mesmo autor, introduzindo uma forma híbrida de produção intelectual entre a pessoa e o algoritmo. Para além disso, possibilita a resposta a exames, bem como a elaboração de pesquisas e revisões, o que coloca em risco a integridade e mesmo a capacidade do sistema educativo. Na verdade, o que pode mesmo estar em causa é a organização do sistema educativo, desde o seu início, adaptando os modelos de lecionação e as formas de avaliação. De uma forma ou de outra, é crucial desenvolver ferramentas de deteção de fraude que, para além de robustas, sejam adaptáveis.

Será um desafio responder ao desenvolvimento dos algoritmos que permitam formas de fraude cada vez mais sofisticadas, devendo reconhecer-se que, associado à possível falibilidade destes sistemas, será sempre difícil fazer a prova da fraude. Por isso mesmo,



para além das ferramentas tecnológicas, é necessário desenvolver uma cultura de integridade académica que desencoraje práticas fraudulentas, nomeadamente ao desenvolver formas de avaliação que se centrem nos processos e não nos resultados. Aliás, a descoberta da fraude não deverá ser o problema mais preocupante, antes o facto de se poder estar a comprometer a criatividade e a reflexão crítica do estudante, bem como a promover uma maior desigualdade na educação, considerando que o acesso às tecnologias muito dificilmente será equitativo. A própria questão metodológica dos trabalhos deverá ser adaptada. A metodologia das revisões sistemáticas e das meta-análises sofrerá cada vez mais alterações, porque mais simples e rápidas com o recurso a algoritmos. O rigor metodológico na validação das saídas do modelo não será apenas um requisito técnico, mas também um imperativo ético, uma vez que se conhece o potencial efeito cascata de conteúdos académicos pouco rigorosos.

O paternalismo libertário, proposto por Thaler e Sunstein⁷⁷, é um modelo possível na apreciação desta problemática, ao recorrer a modelos de influência conscientes. Ou seja, neste caso, as pessoas são influenciadas em determinada direção pela IA, sempre orientadas por valores pró-sociais e de promoção da saúde, mas de uma forma consciente e com a possibilidade de recusar. Porém, no que concerne à IA e à sua falta de explicabilidade, será difícil aceitar modelos de influência que a pessoa não conhece, pois nem estará consciente de que o seu processo de decisão estará a ser condicionado pelos algoritmos, que podem orientar as suas formas de contactar com o mundo. A própria noção de consentimento informado em saúde pode ser posta em causa, na medida em que se, por um lado, a pessoa pode ficar com uma sensação de maior conhecimento em função da tecnologia, por outro, a forma de influenciar pode ser muito mais subtil e inconsciente, mas, ao mesmo tempo, mais eficaz.

Paralelamente, a personalização das informações pode levar à criação de filtros que limitam a exposição a perspetivas diversas, o que pode contribuir para o reforço de preconceitos e estereótipos, potenciando decisões de saúde mais polarizadas e discriminatórias. Neste contexto, torna-se essencial desenvolver algoritmos que sejam transparentes nas suas operações e que possam ser auditados por terceiros por forma a garantir a equidade e a imparcialidade. Devem ser introduzidas medidas de ética "*by design*", ou seja, os algoritmos devem ser desenvolvidos a partir dos princípios éticos como forma de mitigar os riscos associados ao viés e à discriminação.

Será, pois, necessário um grande investimento no desenvolvimento de estudos nesta área com o envolvimento da academia e das associações de profissionais de saúde. É fundamental que a interdisciplinaridade seja a norma, promovendo formação cooperativa nas áreas tecnológicas e de ciência de dados, da saúde, da ética e do direito. Do mesmo modo, é central que a formação de base dos profissionais de saúde seja enquadrada neste novo mundo tecnológico, para que possam ser mais do que meros

⁷⁷ Thaler, R. & Sunstein, C. (2008). *Nudge: Improving Decisions about Health, Wealth and Happiness*. London: Penguin Books. ISBN: 978-0-14-311526-7



utilizadores passivos das tecnologias e possam compreender o seu funcionamento, de forma a fazer uma utilização competente e eficaz, bem como estarem conscientes dos riscos associados.

Torna-se cada vez mais importante promover a avaliação deste tipo de tecnologia, para evitar a disseminação de falsos conceitos ou desenvolver meios que possibilitem a distinção entre o que pode ser verdadeiro ou falacioso. Trata-se de promover uma verdadeira ecologia comportamental, associada à compreensão da forma como se propagam as dinâmicas de opinião neste contexto. Além da necessária evidência científica, que implicará a realização de investigação ao nível da utilização e efetividade destas aplicações, será necessário promover a classificação a partir da avaliação da experiência do utilizador. Se o mercado é livre, e se o que se pretende é que a IA seja cada vez mais um instrumento acessível a todos, será cada vez mais difícil garantir a qualidade das aplicações a partir da indústria e da academia. A facilidade nos meios para produzir aplicações e outra tecnologia altamente impactante nas pessoas implicará que se invista na construção de uma avaliação independente, efetiva e de reconhecida qualidade, como forma de orientar e proteger os utilizadores nas suas escolhas e promover mais consequências positivas da utilização da tecnologia na educação para a saúde.

De uma forma ou de outra, o recurso a este tipo de tecnologia gera, quase inevitavelmente, a produção de mais dados pessoais e de saúde, sendo esse, tantas vezes, o objetivo principal da indústria que financia o desenvolvimento deste tipo de ferramentas. A privacidade dos dados é, pois, central também a este nível. Contudo, e por muito que se tenham tomado decisões importantes do ponto de vista jurídico, como seja o RGPD na União Europeia, e se faça um esforço maior em cibersegurança, os riscos são grandes. Em primeiro lugar porque a informação, para além de ser em maior quantidade e qualidade, está muito mais organizada do que no passado, o que faz com que cada quebra na privacidade possa representar um dano potencial maior. Acresce que a maior parte das pessoas tem competências ao nível da utilização das tecnologias, mas pouco conhecimento sobre o seu funcionamento, depositando confiança nos processos de uma forma que é inversa ao seu conhecimento. Deste modo, os cuidados sobre a disponibilização e privacidade dos dados são reduzidos, o que aumenta os riscos.

Torna-se cada vez mais relevante definir quem terá a responsabilidade pela guarda dos dados pessoais e de saúde. A enorme quantidade de dados obtidos nos cuidados de saúde e na investigação realizada coloca desafios assinaláveis, cuja resposta não é simples. Se parece pouco avisado que sejam entidades privadas (como seja as empresas tecnológicas que produzem o *software* e o *hardware*) a manterem o controlo dos dados em saúde, questiona-se se deverão ser entidades públicas. Importará discutir e promover soluções que garantam a privacidade dos dados em saúde, bem como a sua utilização, exclusivamente, no melhor interesse dos seus proprietários. Deste modo, a criação de agências reguladoras independentes, com a função exclusiva de gerir os



dados existentes em benefício dos seus proprietários, pode constituir-se num caminho a ser explorado. Devem ser promovidas medidas de transparência que permitam que as pessoas consigam compreender como são utilizados os seus dados, e que sejam empregues técnicas de encriptação e anonimização robustas a par de auditorias regulares que possam identificar dificuldades.

A incorporação da IA e da realidade virtual na formação em saúde representa muitas oportunidades e desafios. A sua aplicação responsável requer, necessariamente, uma abordagem multidisciplinar, em que a ética esteja presente desde o início, de forma a que os algoritmos sejam desenhados a partir dos próprios princípios éticos. No mais, é necessário desenvolver protocolos orientados por evidência científica para a melhor implementação das tecnologias e para a avaliação do impacto das mesmas. Paralelamente, a formação em literacia tecnológica deve ser integrada em todos os níveis de ensino, o que, certamente, incluirá os currículos de formação em saúde, pré e pós-graduada.

Importa ainda considerar que este desenvolvimento tecnológico exponencial traz consigo um conjunto de desafios que exige uma consideração cuidadosa, desde a privacidade dos dados ao risco de agravamento dos vieses e dos processos de discriminação. Uma gestão responsável dos dados, no sentido do melhor interesse das pessoas, terá de ser um dos pilares fundamentais de uma integração responsável de IA na educação em saúde. Além disso, o desenvolvimento de orientações sólidas que exijam a transparência dos processos, associado a uma formação ética abrangente dos profissionais de saúde, são condições mínimas para enfrentar os desafios éticos que esta nova era de formação digital representa.

Aspetos Éticos Relevantes do Impacto da IA no Ensino e Educação em Saúde

- A IA e a realidade virtual oferecem oportunidades inovadoras para o treino e formação de profissionais de saúde;
- Estas tecnologias permitem a criação de cenários de treino realistas e adaptados às necessidades individuais dos estudantes, contribuindo para uma formação mais eficaz;
- O uso da IA e da realidade virtual levanta preocupações sobre a possível perda de capacidade dos profissionais de saúde em fornecer cuidados que considerem as dimensões psicológicas e sociais da saúde;
- Deve garantir-se que relação terapêutica e de confiança entre a pessoa doente e o profissional de saúde é preservada;
- A incorporação eficaz da IA e da realidade virtual requer um nível adequado de alfabetização tecnológica e de literacia digital entre os profissionais de saúde.



6. Visão prospetiva

A Inteligência Artificial (IA) é já hoje omnipresente no nosso quotidiano. Entretanto, continua a expandir-se a uma multiplicidade crescente de domínios da atividade humana em que reproduz diferentes funções tradicionalmente realizadas por pessoas de forma mais rápida, mais rigorosa, mais económica, isto é, de forma mais eficiente. Neste contexto, os benefícios que protagoniza para as sociedades e os cidadãos são meritórios, devendo ser identificados e implementados. Impõe-se, concomitantemente, uma ponderação cuidada relativamente aos eventuais riscos subjacentes, procurando discernir, na miríade de possibilidades que a tecnologia oferece, os caminhos que melhor prossigam os valores e princípios éticos em que nos revemos.

O Conselho Nacional de Ética para as Ciências da Vida (CNECV) sublinha que, qualquer que seja o domínio de aplicação da IA, no presente e futuro, se lhe impõem dois imperativos éticos estruturantes:

- alicerçar-se e desenvolver-se no mais estrito respeito pelos princípios éticos identitários do humano, isto é, pelos seus valores nucleares e direitos fundamentais;
- centrar-se no humano como finalidade última, isto é, rejeitando o pragmatismo da captura por interesses sectários e/ou particulares, mantendo-a como instrumento de realização do humano, individual e socialmente considerado.

Um dos domínios em que a IA vem evidenciando um mais forte impacto positivo é o das ciências da vida, especificamente no âmbito humano e em particular da saúde, em que tem vindo a ser introduzida a um ritmo acelerado, numa crescente diversidade de áreas de atuação, especialidades médicas e atos clínicos.

Estruturámos metodologicamente o impacto da IA nas ciências da vida em cinco planos ou grandes áreas de intervenção, distintas, mas intercetantes: investigação biomédica, assistência clínica, gestão hospitalar, administração da saúde pública, e ensino/educação em saúde. Cada uma requer ações particulares para o seu desenvolvimento, na observância de princípios de atuação ética específicos, como forma de garantir simultaneamente o progresso tecnológico e o benefício social.

Neste contexto, o CNECV destaca os desafios éticos que se colocam na aplicação da IA à:

(1) Investigação biomédica, em que importa

- a) salvaguardar a criatividade intelectual dos investigadores, admitindo e não penalizando o processo de conhecimento através da tentativa e erro
- b) preservar a replicabilidade e a explicabilidade dos processos de investigação e de produção de conhecimento, desenvolvendo medidas concretas que permitam ultrapassar a crise de replicação na ciência promovida pela IA
- c) reforçar a proteção do volume massivo de dados pessoais que a IA utiliza, através de novas modalidades de anonimização e de estratégias inovadoras



que garantam a privacidade individual e a confidencialidade dos dados pessoais, de que é exemplo o recurso a dados sintéticos gerados por IA

- d) rever a doutrina e a prática do consentimento informado de forma a que, na fidelidade ao seu desígnio de expressão da autonomia pessoal, se ajuste à profunda transição digital promovida pela IA
- e) reapreciar os critérios de atribuição da autoria científica e as políticas institucionais de integridade científica, na consideração das intervenções possíveis da IA generativa

(2) Assistência clínica digital, em que importa

- a) formular medidas para manter e aprofundar a confiança numa prática clínica associada à utilização de meios digitais
- b) prevenir uma diminuição da competência dos profissionais de saúde que poderia advir da perda de prática clínica associada a uma utilização generalizada da IA, bem como a uma excessiva dependência da robótica inteligente
- c) conferir a verificabilidade dos sistemas digitais no diagnóstico e na terapêutica, bem como a explicabilidade do processo de decisão profissional, procurando evitar uma possível diluição de responsabilidades
- d) adaptar os processos e modelos de registo da vontade manifesta da pessoa, nomeadamente a obtenção do consentimento informado, ao contexto da IA
- e) garantir o direito à privacidade individual e a proteção de dados pessoais em todas as aplicações de sistemas de IA à saúde

(3) Administração da saúde, gestão hospitalar e intervenção à distância (telessaúde), em que importa

- a) beneficiar do contributo da IA para uma plena integração e coordenação da prestação de cuidados, e otimizar o acesso aos serviços e sistemas de saúde
- b) evitar o potencial agravamento das assimetrias existentes que podem decorrer da utilização da IA e proteger as pessoas ou grupos mais vulneráveis
- c) assegurar o respeito pela autonomia e privacidade das pessoas, na gestão operacional inteligente das unidades de saúde e manter uma relação ética e humanizante entre os profissionais de saúde e as pessoas
- d) manter o estrito cumprimento das regras da medicina baseada na evidência, sem perder de vista a necessidade do contacto humano e personalizado no desenvolvimento da telessaúde



- e) promover a gestão integrada do registo de saúde eletrónico, dado o seu enorme potencial de reforço da autonomia da pessoa, ao disponibilizar em tempo real e de forma consolidada a informação de saúde que se lhe refere
- f) afiançar que a prescrição médica resultante de uma gestão integrada na *interface* com a IA generativa seja fiável, verificável e que respeita a autonomia da pessoa.

(4) Administração da saúde pública e codificação de dados, em que importa

- a) dispor do uso de IA na administração da saúde pública, a qual pode ajudar a identificar e monitorizar surtos em tempo real e a efetivar vigilância epidemiológica através de análise preditiva, conquanto tal monitorização não coloque em causa a privacidade individual e a segurança dos dados
- b) desenvolver a integração da automação da codificação de dados nas tecnologias de informação e comunicação de saúde, fomentando a cibersegurança dos sistemas
- c) atender ao risco dos fenómenos de colonização de dados através da criação de sistemas de armazenamento que obedeçam ao princípio do interesse público, mantendo o respeito pela pessoa nas suas múltiplas dimensões
- d) investir na literacia digital dos cidadãos, capacitando-os, para que os processos de automação possam constituir um contributo efetivo e conscientemente assumido para a saúde pública

(5) Ensino e educação em saúde e a realidade virtual, em que importa

- a) implementar as oportunidades inovadoras que a IA e a realidade virtual oferecem para o treino e formação de profissionais de saúde, por exemplo através da criação de cenários realistas e adaptados às necessidades individuais dos estudantes, contribuindo para uma formação mais ética e mais eficaz
- b) desenvolver a investigação na utilização destas tecnologias no treino e na formação de profissionais, por forma a promover os processos de aprendizagem e a diminuir a probabilidade de vieses associados a uma menor representatividade da diversidade humana
- c) acautelar a possível perda de capacidade dos profissionais de saúde em fornecer cuidados que considerem as dimensões psicológicas e sociais de cada pessoa, eventualmente decorrente do uso da IA e da realidade virtual



- d) cultivar continuamente a relação terapêutica e de confiança entre o doente e o profissional de saúde, reforçando a formação e a sensibilização dos estudantes e dos profissionais nesse sentido
- e) considerar que a incorporação eficaz da IA e da realidade virtual requer um nível adequado de preparação tecnológica e de literacia digital entre os profissionais de saúde
- f) dotar os profissionais de competências de comunicação e de capacidade crítica no contexto do recurso às potencialidades e aplicações das novas tecnologias e promover uma cultura de integridade que desencoraje práticas fraudulentas.



Audições

A elaboração do Livro Branco realizadas no contexto do Livro Branco “Inteligência Artificial (IA): inquietações sociais, propostas éticas e orientações políticas” foi antecedida de uma sessão de audições e debate público, realizada no dia 16 de fevereiro de 2024 na Universidade do Minho, em Braga.

A iniciativa contou com uma mesa-redonda que teve como intervenientes os especialistas José Manuel Machado (UMinho), Maria Goreti Marreiros (Instituto Superior de Engenharia do Porto - ISEP) e José L. Vilaça (Instituto Politécnico do Cávado e do Ave - IPCA). O debate que se seguiu visou dinamizar a reflexão e recolher diferentes perspetivas sobre esta tecnologia em acelerado desenvolvimento.

O CNECV agradece reconhecidamente o valioso contributo dos especialistas, que muito enriqueceu o presente documento, bem como a participação do público. Agradece ainda o generoso acolhimento por parte da Universidade do Minho ao evento realizado.