TRABALHO FINAL
MESTRADO INTEGRADO EM MEDICINA

Ética Médica

Considerações Bioéticas sobre a aplicação de sistemas de Inteligência Artificial em Medicina

Luís Filipe do Carmo Arcângelo

Orientado por:
Professor Doutor Miguel Oliveira da Silva

Julho’2024

Abstract

Artificial Intelligence (AI) is today an inescapable topic often surrounded by speculative debates about existential risks and utopian futures. However, beyond these imaginings, it is crucial to focus on the realistic implications of AI, which promises to automate complex tasks across various sectors. Healthcare, with its infinite demand and limited resources, is a prime candidate for such automation. Unlike other sectors, healthcare's inherently human-centered purpose necessitates a strict ethical framework.

This study recognizes that perspectives on AI vary widely among the general public, healthcare professionals, IT specialists, and policymakers. This divergence, combined with the asymmetry of specialized knowledge—clinical, bioethical, and technological—poses challenges for the ethical and responsible implementation of AI in healthcare. Without multidisciplinary involvement, ensuring AI development and deployment align with superior moral values rather than materialistic goals becomes difficult.

This thesis aims to consolidate essential multidisciplinary information. Chapter I introduces Principlism as an ethical framework and its four guiding principles. Chapter II elucidates AI's theoretical and technical aspects, addressing misconceptions and detailing various AI subtypes and applications in healthcare. Chapter III leverages this foundational knowledge to explore bioethical considerations of AI in medicine from a principlist perspective.

The core objective of this thesis is to inform healthcare professionals about the bioethical and practical implications of AI, to educate AI developers and policymakers on the unique moral obligations in healthcare, and to underscore the importance of multidisciplinary collaboration throughout the AI lifecycle in healthcare.

Keywords: Artificial Intelligence; Principlism; Bioethics; Healthcare Automation;

Resumo

A Inteligência Artificial (IA) é hoje um tema incontornável, para além de discussões especulativas sobre ameaças existenciais e futuros utópicos, é crucial abordar as implicações realistas desta tecnologia, que promete automatizar tarefas complexas e dinâmicas em diversos setores. A Saúde, com a sua procura virtualmente infinita e recursos limitados, é uma candidata notória para esta automação. Contudo, distingue-se fundamentalmente de outros setores devido ao seu propósito intrinsecamente centrado em valores humanos superiores e a sua exigência de um código moral particularmente rigoroso.

Este trabalho parte da constatação de que as perspetivas sobre a IA são discrepantes entre o público em geral, profissionais de saúde, especialistas em TI e decisores políticos. A interseção desta divergência com a assimetria de conhecimento especializado—clínico, bioético e tecnológico—levanta preocupações sobre a implementação ética e responsável da IA na Saúde. A natureza altamente específica e exigente destas disciplinas dificulta que os profissionais de cada área individual tenham a visão multidisciplinar necessária para garantir o melhor desenvolvimento e implementação da IA, salvaguardando os superiores valores morais e que estes não são regidos exclusivamente por objetivos materialistas.

Este trabalho visa compilar informação multidisciplinar essencial, começando com o Capítulo I, que detalha o quadro ético do Principialismo e os seus quatro principais princípios. O Capítulo II clarifica conceitos teóricos e técnicos sobre a IA, abordando equívocos comuns e detalhando vários subtipos de IA e as suas diferentes aplicações na Saúde. O Capítulo III utiliza este conhecimento de base para explorar considerações bioéticas decorrentes da implementação da IA na Medicina, através de uma perspetiva Principialista.

O objetivo central desta tese é informar os profissionais de saúde sobre as implicações bioéticas e práticas da IA, familiarizar os desenvolvedores de IA e legisladores com as obrigações morais intrínsecas à Saúde e destacar a importância do envolvimento multidisciplinar ao longo do ciclo de vida da IA na Saúde.

Palavras-chave: Inteligência Artificial; Principialismo; Bioética; Automação na Saúde;

Índice

Introdução

A *Inteligência Artificial (IA)* é atualmente um tópico inescapável, cuja discussão alargada é frequentemente dominada por ideias fantasiosas de ameaças existenciais e promessas de um futuro pós-escassez. Depois de se render estas suas intrigantes vertentes imaginárias ao mundo literário, deve-se prosseguir com a importante discussão sobre as implicações realísticas desta tecnologia, que embora tenha sido conceptualizada nos anos 50 do século passado, apresenta aplicações importantíssimas nas mais diversas áreas da sociedade moderna, por prometer, pela primeira vez, automatizar tarefas complexas e dinâmicas, cuja concretização requer *alguma forma de inteligência*.

A Saúde é, então, um alvo evidente para a implementação da automatização deste tipo de tarefas, para que se possa aumentar a oferta numa área na qual a procura é virtualmente infinita e os recursos, humanos e materiais, necessários para a satisfazer são particularmente escassos. Contudo, a Saúde é uma área fundamentalmente díspar de outras áreas económicas e sociais, nas quais se poderá olhar com menos restrições para os benefícios trazidos pelo desenvolvimento e implementação da IA. A finalidade da Saúde é puramente humana, cuidar e promover o bem-estar dos pacientes, o que exige, entre muitas outras coisas, reger-se por um código moral excecional.

A motivação inicial deste trabalho partiu da realização de que existe uma enorme *divergência de pensamento relativamente à IA*, especialmente *no que motiva* o público geral, os profissionais de saúde, os especialistas das tecnologias de informação e os decisores políticos *a apoiarem, ou não*, o seu desenvolvimento, implementação e utilização, e nas *expectativas e exigências* que têm relativamente a esta.

A conjugação desta divergência de pensamento com a *assimetria de conhecimento especializado*, nomeadamente, *clínico*, *bioético* e *tecnológico*, entre os desenvolvedores e implementadores dos sistemas de IA, e os utilizadores finais, suscita sérias dúvidas sobre se a IA poderá ser implementada na Saúde de forma ética e responsável, não existindo envolvimento multidisciplinar em todo o seu ciclo de vida. O problema é agravado pela

constatação de que as diferentes disciplinas necessárias a este envolvimento são atualmente tão especializadas que os peritos nestas, por limitações de tempo e obrigações profissionais, acabam por não ter a visão abrangente necessária ao reconhecimento da importância basilar da multidisciplinariedade como mecanismo de controlo que assegura que o desenvolvimento e a implementação dos sistemas de IA não é guiado apenas por objetivos materialistas, garantindo a proteção e promoção dos superiores valores bioéticos que, caso contrário, correm o risco de ser relegados a considerações *a posteriori*, subjugadas a interesses não morais.

Este trabalho pretende, por isso, (1) reunir informação e conhecimento multidisciplinar essencial, começando pela apresentação sumária, no Capítulo I, de *conceitos bioéticos frequentemente aplicados na Saúde*, após a qual será explanado, no Capítulo II, *conhecimento técnico e teórico fundamental sobre sistemas de IA*; e por fim, no Capítulo III, (2) utilizar este substrato de informação para levantar e analisar diversas *considerações bioéticas resultantes da implementação de sistemas de IA na Medicina*.

Capítulo I - Principialismo

I.1 – Escolha do Modelo Principialista

Este trabalho propõe-se a identificar, explorar e avaliar, sistematicamente, considerações bioéticas decorrentes da implementação prática de *sistemas de Inteligência Artificial* na Medicina. Face a esta ambição, é de grande importância a utilização de um modelo ou sistema ético que que possua relevância prática, permitindo a análise detalhada de casos específicos. Apesar do indubitável valor filosófico de modelos éticos mais teóricos, a universalidade das suas premissas não se coaduna com esta metodologia analítica. (Flynn, 2022)

O *Principialismo*, conforme delineado em Beauchamp, T. L., & Childress, J. F. (2019). *Principles of Biomedical Ethics* (8th ed.). Oxford University Press., destaca-se como um modelo de particular robustez para este tipo de análise prática. Prova disso é a sua abrangente utilização enquanto referencial orientador para profissionais de saúde e em avaliações bioéticas no âmbito da saúde. Este modelo tem demonstrado a sua durabilidade e adaptabilidade ao longo do tempo, mesmo sob escrutínio académico e em contextos dinâmicos e desafiantes como é o caso da medicina moderna. (Gordon, s.d.)

Os *princípios* fundamentais do Principialismo - *Beneficência*, *Não Maleficência*, *Respeito pela Autonomia* e *Justiça* - fornecem uma estrutura sólida para a formulação de normas e regras éticas operacionalizáveis, constituindo um enquadramento metodológico sistemático e coerente adequado a este trabalho.

I.2 – Principialismo – Ética Normativa e Ética Aplicada

O Principialismo, alia a *ética normativa* e a *ética aplicada* à área da saúde e das ciências biomédicas. Este investiga quais são e por que razão existem as *normas morais* que devem guiar e avaliar a conduta profissional nas áreas dos cuidados de saúde, saúde pública, política de saúde e investigação biomédica. Estas normas são expressas através de *princípios*, abrangentes e fundacionais, por isso menos operacionalizáveis, e pelas suas derivativas *regras*, *obrigações*, *direitos* e *virtudes*, mais específicas e restritas, mas aplicáveis às questões concretas subjacentes às profissões, instituições e políticas públicas da saúde.

A justificação ética teórica deste modelo, segundo os autores, baseia-se na *Moralidade Comum*. Esta representa o conjunto de crenças morais universais, compartilhadas por todos os indivíduos dedicados a viver uma vida moral. Destas crenças constroem-se normas *universais*, enquanto produto da experiência e história humana conjunta. Estas normas são impérvias ao relativismo moral individual ou cultural, podendo ser exemplificadas pela proibição de causar danos injustificados aos outros, sejam físicos ou psicológicos e pela obrigação de respeitar os seus direitos.

Paralelamente à Moralidade Comum, existem *Moralidades Particulares*, mais específicas e por isso não universais, nas quais estão inseridos os *ideais morais* e os *códigos de conduta* que certas profissões exigem. As normas que advêm destas moralidades não são universais, mas são moralmente desejáveis e admiráveis. Estas poderão ser voluntariamente adotadas por indivíduos e grupos (profissionais, religiosos, ideológicos, etc.) que procuram um padrão mais elevado de moralidade, embora não passível de ser imposto universalmente. Estas diferenças explicam a existência da discordância e pluralismo moral, mas existindo na Moralidade Comum uma base universal para o julgamento ético, nega-se o relativismo absoluto.

A disposição para agir motivado por estes princípios, normas, obrigações e ideais morais constitui a *virtude moral*, que em conjunto com outras virtudes como a *compaixão*, *discernimento*, *confiança*, *integridade* e *consciência* são essenciais para a prática médica. Estas

não apenas promovem a qualidade dos cuidados, mas também ajudam a moldar a relação entre profissionais de saúde e pacientes de forma profundamente humana e moralmente enriquecedora.

É da Moralidade Comum que derivam os princípios enquanto estruturas analíticas e guias normativos generalistas, a partir dos quais se podem formular *regras*[1] e *juízos*, mais específicos e aplicáveis.

Ao contrário de alguns modelos mais teóricos, este modelo não apresenta um princípio unificador e absoluto, sobre qual toda a análise ética assenta. Não reconhecendo a superioridade *a priori* de um princípio perante o outro, a resolução de conflitos morais implica a sua especificação e ponderação.

- A *especificação* ajuda a concretizar uma norma, reduzindo o seu escopo e tornando a diretamente aplicável à situação particular.
- A *ponderação* envolve avaliar a força moral destas especificações e atribuir-lhe um peso relativo.

A conjugação destes dois "instrumentos" éticos permite a tomada de decisões moralmente justificadas.

Uma vez que nenhum dos quatro princípios é inerentemente superior, por serem todos igualmente justificados pela Moralidade Comum, as normas morais que deles derivam constituem, da mesma forma, obrigações *prima facie*[2]. Contudo, decisões e ações concretas exigem, frequentemente, a escolha de normas em detrimento de outras consigo inconciliáveis, o que poderá ser moralmente justificado partindo do juízo de que a obrigação escolhida é, na situação específica que motiva a ação, de igual ou maior peso moral que a preterida. Não

[1] Os autores assinalam três tipos: (a) *Regras Substantivas*: Guias específicos para ação que concretizam os princípios de forma prática. (b) *Regras Autoritárias*: Determinam quem tem o poder e a autoridade para tomar decisões em contextos específicos. (c) *Regras de Procedimento*: Estabelecem procedimentos a serem seguidos, utilizados quando as regras substantivas e autoritárias não são aplicáveis ou conclusivas.

[2] Ou seja, obrigações morais que devem ser sempre respeitadas, a não ser que exista uma obrigação moral de maior ou igual peso que justifique o seu incumprimento.

obstante, o importante carácter moral destas obrigações exige a proteção formal do ato deliberativa, evitando a parcialidade, arbitrariedade e dependência excessiva na intuição. As decisões morais devem, então, respeitar as seguintes condições justificativas: (a) São levantadas boas razões para agir segundo a norma que se sobrepõe, em detrimento da norma infringida. (b) O objetivo moral que justifica a infração tem uma probabilidade realística de ser atingido. (c) Não existe outra ação alternativa moralmente preferível. (d) A ação infringe o mínimo possível a norma. (e) Minimizámos todos os efeitos negativos provenientes da infração. (f) Todos os afetados pela decisão foram tratados de forma imparcial.

Após esta sucinta apresentação e explicação do Principialismo e das suas principais particularidades enquanto modelo ético analítico, justificado pela Moralidade Comum, irei explanar nos próximos subcapítulos, cada um dos quatro princípios normativos e as suas respetivas considerações especiais.

I.2.1 - Princípio da Beneficência

No Principialismo, a *beneficência* expressa, na generalidade, as normas, disposições e ações com o objetivo de beneficiar e promover o bem-estar de outrem[3]. Esta assume uma posição central na Medicina, sendo a justificação primordial da sua existência e o seu principal objetivo.

O *Princípio da Beneficência* afirma uma obrigação moral generalista de agir de forma a beneficiar os outros. Este assume duas principais formas:

- *Beneficência ativa*, que estipula que os profissionais de saúde devem agir proactivamente na promoção do bem-estar.
- *Princípio da Utilidade*, ditando que os benefícios das intervenções médicas devem ser maximizados, enquanto os danos devem ser minimizados.

Deste princípio derivam regras *prima facie*, incluindo (a) proteger os direitos dos outros; (b) prevenir e remover condições que lhes possam causar danos; (c) assistir pessoas em perigo. Algumas ações benevolentes, como atos de sacrifício extremo, embora moralmente *admiráveis*, não são *obrigatórias* segundo a moralidade comum. A diferença de obrigatoriedade é ilustrada pela distinção entre beneficência específica e geral:

- A *beneficência específica* é constituída pelas obrigações que certos sujeitos têm de agir em benefício e pela promoção doo bem-estar de indivíduos ou grupos específicos, impostas por relações morais e compromissos especiais, como as obrigações que os pais têm para com os seus filhos, ou os *deveres* que os médicos têm para com os seus pacientes.
- A *beneficência geral* abrange as obrigações morais que cada sujeito tem de agir em benefício e pela promoção do bem-estar de todas as pessoas, independentemente de possuírem ou não uma relação prévia.

[3] A sua virtude correspondente é a *Benevolência*.

O *dever de socorrer*, uma forma de beneficência obrigatória, estabelece os critérios pelos quais uma pessoa tem a obrigação moral de ajudar outra: (a) A outra pessoa está em risco de uma perda ou dano que irá afetar significativamente a sua vida, saúde ou interesses; (b) a sua ação é necessária e irá provavelmente prevenir essa perda ou dano; (c) a ação não implica riscos, custos ou encargos significativos para quem ajuda; (d) os benefícios expectáveis para quem requer ajuda são claramente superiores a qualquer potencial desvantagem para quem presta socorro.

Uma particularidade que pode ser vista à luz do Princípio da Beneficência é o *uso compassivo*. Este aplica-se em situações nas quais pacientes com doenças graves não têm alternativas terapêuticas disponíveis, sendo-lhes permitido o acesso a tratamentos experimentais, ainda não completamente aprovados pelos reguladores, mas já estando assegurado o seu perfil de segurança básico fundamental. Em circunstâncias que satisfaçam as condições supracitadas do dever de socorrer, podemos inclusive afirmar que existirá uma obrigação moral de garantir ao doente o *direito a tentar*.

O acesso a terapêuticas experimentais, em pacientes que já se encontram inseridos num programa de investigação biomédica, também pode ser justificado pela reciprocidade. O risco que estes aceitaram assumir, pelo progresso da Ciência e Medicina, deve ser retribuído através do acesso aos tratamentos em estudo, quando estes se revelam eficazes.

A reciprocidade pode impor que os pacientes igualem alguns dos encargos assumidos pelos pacientes antes deles, que permitiram o avanço das áreas Médico-científicas do qual beneficiam agora, por exemplo, permitindo aos profissionais obter informações, dados e aprendizagens essenciais para este avanço, do qual também os próximos pacientes poderão beneficiar.

Outra particularidade subjacente ao Princípio da Beneficência é o *paternalismo*, que surge quando um profissional de saúde sobrepõe os seus juízos aos de um paciente, com a intenção de o beneficiar ou proteger de danos. A sua permissibilidade moral depende apenas

das circunstâncias e da maneira como é aplicado, sendo eticamente neutro. O paternalismo pode ser classificado em *suave* ou *rígido*:

- O paternalismo suave ocorre quando o agente intervém para proteger um indivíduo das consequências negativas de ações e escolhas substancialmente não autónomas[4]; resultantes de alterações psicológicas ou psiquiátricas, vícios, ou até de crenças particularmente desinformadas e inequivocamente erradas.

Não sendo a determinação do grau de compromisso da autonomia completamente objetiva, as intervenções paternalísticas devem ser sempre feitas de forma transparente e limitada aos casos onde o paciente tem elevado risco de sofrer consequências significativamente prejudiciais.

- O paternalismo rígido manifesta-se pela ação profissional que contradiz escolhas aparentemente informadas e voluntárias do paciente (ainda que arriscadas).

O elevado grau de intrusão e de usurpação da autonomia das intervenções paternalísticas exige a sua restrição às seguintes condições: (a) O paciente tem um risco claro e previsível de sofrer danos ou perder benefícios significativos. (b) A intervenção tem alta probabilidade de prevenir os danos ou garantir os benefícios. (c) Os danos prevenidos ou os benefícios garantidos ultrapassam, muito provavelmente, os riscos da intervenção. (d) Não existe nenhuma alternativa moralmente superior, que não implique a limitação da autonomia. (e) A intervenção escolhida, das que previnem os danos ou garantem os benefícios, é a que menos restringe a autonomia do paciente. Uma sexta condição também pode ser considerada em certos casos: (f) As intervenções não deverão ir contra interesses substanciais e centrais ao indivíduo.[5]

[4] O conceito de autonomia será explorado em I.2.3 - Princípio do Respeito pela Autonomia

[5] Por exemplo, a recusa de transfusões sanguíneas por parte das testemunhas de Jeová

O paternalismo também pode ser *passivo,* quando o profissional de saúde opta por não realizar uma intervenção explicitamente desejada pelo paciente, por considerá-la medicamente fútil[6], ou seja, julga que esta não trará benefícios clínicos previsíveis e fundamentados.

O Subcapítulo I.2.1 focou-se até agora na *Beneficência Ativa*, deverá ser abordada agora a outra componente do Princípio da Beneficência, o *Princípio da Utilidade*. Agir proactivamente para o benefício dos outros é essencial, mas não suficiente para maximizar a beneficência geral, para a qual é indispensável a construção de sistemas, serviços e políticas de saúde.

A tomada de decisões que promovem estas construções sociais beneficentes requer ferramentas analíticas próprias, as *Análises de Utilidade*, que assumem várias formas que iremos, sucintamente, abordar. Estas determinam, avaliam e comparam os diversos *benefícios*[7], *custos*[8] e *riscos*[9] que advêm, por exemplo, de uma política de saúde pública ou da implementação de uma intervenção médica ou tecnológica, sendo, por isso, importantes aplicações práticas do Princípio da Beneficência.

- *Análises de Risco-Benefício* são amplamente utilizadas na regulamentação de medicamentos, intervenções e dispositivos médicos, visando a proteção dos doentes. Não são, no entanto, completamente livres de valores, pois a própria apreciação do que constituem prejuízos e benefícios é subjetiva, e não está acima de escrutínio. A relevância desta subjetividade é ilustrada pela *perceção de risco* que varia entre grupos e indivíduos. Esta perceção pode ser suficiente para originar uma distorção da política pública motivada pelo *excesso de zelo*, paralisando o desenvolvimento e impedindo a implementação de novos meios terapêuticos com imenso potencial benéfico, muitas vezes por medos vagos e infundados. A transparência, acesso a informação e

[6] O conceito de futilidade médica será explorado em I.2.2 – Princípio da Não Maleficência

[7] Resultado positivo face aos interesses individuais, como a vida, saúde ou bem-estar.

[8] Não apenas os recursos financeiros necessários, mas também potenciais efeitos negativos decorrentes da implementação de novas tecnologias ou tratamentos.

[9] Potencial de prejuízo aos interesses individuais. Termo probabilístico, composto pela *probabilidade* e pela *magnitude* do prejuízo antecipado.

comunicação clara, baseada na evidência, é fundamental para alinhar a perceção pública com a realidade científica.

- *Análises de Custo-Benefício* e *Análises de Custo-Efetividade* são empregues com intuito de reduzir subjetividades e o uso excessivo da intuição na criação de políticas públicas de Saúde. Todavia, estas não são isentas de valores e vieses, pressupondo elas próprias assunções éticas, de alinhamento primariamente utilitarista. O seu caráter redutor a números, impede-as muitas vezes de capturar de forma adequada a complexidade da situação real que pretende avaliar e os seus valores éticos subjacentes. Exemplificando, a métrica dos Anos de Vida Ajustados pela Qualidade (*QALYs*), ao combinar *anos de vida* com *qualidade de vida* num único índice, se aplicada de forma acrítica do ponto de vista moral, pode privilegiar o número de anos de vida face ao número de vidas individuais e a qualidade de vida à quantidade de vida.

I.2.2 - Princípio da Não Maleficência

O *Princípio da Não Maleficência* tem como principal enfoque *proteger* outras pessoas de sofrerem *dano*, sendo este definido como a afetação negativa dos interesses de outrem[10]. Embora seja considerado em certas situações sobreponível à beneficência, a abstinência de causar dano estabelece muitas vezes obrigações ainda mais rigorosas do que as obrigações ativas de beneficiar.

Ainda assim, a causação de dano pode ser justificada pelos benefícios expectáveis, mas enquanto obrigação *prima facie* deverá haver uma justificação criteriosa. A especificação e ponderação poderão ser empregues para caracterizar e distinguir os graus de dano[11] e benefícios expectáveis numa determinada situação, por exemplo, uma amputação será um dano justificado se for necessária para salvar a vida de um paciente.

A não maleficência também engloba a *não imposição de riscos* injustificados a outrem e as ações que contrariam esta obrigação são consideradas *negligência*, que se manifesta de duas formas principais:

- *Negligência intencional*, quando existe uma decisão e ação que conscientemente impõe riscos desnecessários aos pacientes.[12]
- *Negligência não intencional*, quando a conduta profissional revela falta de cuidado e zelo pelos outros e pelos seus direitos.[13]

[10] *Causar dano* deve ser distinguido de *fazer mal*, uma vez que que o primeiro não exige intenção maliciosa ou a violação intencional dos direitos de outrem.

[11] Atentados à propriedade, reputação, privacidade e liberdade, embora moralmente errados, podem ser distinguidos de atentados à integridade física ou vida de uma pessoa.

[12] Por exemplo, um profissional de saúde que opta por não refazer um penso sujo, sabendo que isso acarreta o um risco aumentado de infeção da ferida.

[13] Por exemplo, um profissional de saúde que divulga repetidamente informações confidenciais dos seus pacientes, não de forma premeditada, mas por ter uma postura imprudente que revela desconsideração pelo direito à confidencialidade.

Os cuidados de saúde exigem um código de conduta profissional que impõe responsabilidades específicas para com os pacientes, os *deveres profissionais*, nos quais se incluem critérios de não maleficência particularmente estritos. A ação médica negligente é por isso uma infração moral particularmente grave.

A área da saúde requer, simultaneamente, a tomada de decisões particularmente difíceis, como as de iniciar ou manter em tratamento[14] um doente com um prognóstico muito reservado, havendo a possibilidade de se estar a prolongar infrutiferamente o seu sofrimento. Situações de oposição entre obrigações *prima facie* de beneficência e de não maleficência levantam questões sobre a *obrigatoriedade de tratar*. Este tipo de ponderação é frequentemente influenciado por considerações baseadas em conceitos subjetivos como a *futilidade médica* e a *qualidade de vida*.

A *futilidade médica* pressupõe que as opções terapêuticas existentes para um doente: (a) Não produzem o efeito fisiológico pretendido[15]. (b) São apenas especulativas, não tendo sido testadas para o efeito pretendido. (d) É altamente improvável desencadearem um efeito positivo. (e) Qualquer efeito que tivessem seria insignificante.

A *qualidade de vida*[16], quando suficientemente baixa, em concomitância com a futilidade da terapêutica disponível, poderá justificar moralmente as decisões de não iniciar ou suspender o tratamento. Estas decisões, aplicadas a doentes incapazes de decidir por si, devem basear-se no critério do *melhor interesse para o paciente*[17] e não no valor relativo da vida deste para os outros, como os possíveis encargos que poderá imputar à sua família ou à sociedade.

---

[14] A diferenciação entre *não iniciar* ou *interromper* uma linha de tratamento é moralmente ambígua. A *interrupção* pode ocorrer através de um *não reinício*, como não recarregar as baterias que alimentam um ventilador ou não colocar a fórmula de alimentação entérica numa sonda. Em tratamentos que envolvem várias fases, a decisão de não iniciar a fase seguinte de um plano de tratamento, pode ser equivalente à interrupção do tratamento.

[15] Por exemplo, a utilização de antibióticos (cujo princípio ativo afeta bactérias) para tratamento de uma infeção viral (=vírus).

[16] A definição exata da qual excede o objetivo deste trabalho, para o qual bastará reconhecê-la como ponto de desacordo moral, mas, no entanto, sendo utilizada como justificação para a tomada de todo o tipo de decisões na área da Saúde.

[17] Este critério será explorado em I.2.3 – Princípio da Autonomia.

Relacionadas à obrigatoriedade de tratar, estão as decisões de *deixar morrer*, nas quais a interrupção, ou não início, de intervenções de suporte de vida levará à morte do doente. Os autores argumentam que a que a manutenção incondicional da vida humana não é uma obrigação moral, é sempre necessário considerar a dor, sofrimento e desconforto do doente[18]. Assim, decisões que levam à morte do paciente podem ser moralmente aceitáveis, mas apenas na presença de uma das seguintes condições: (a) A tecnologia médica é fútil. (b) O paciente ou o *procurador de cuidados de saúde* autorizado recusou, *validamente*, o uso da tecnologia.

Esta *validação* deve ser feita pelos profissionais de saúde mediante certas exigências relativas ao decisor: (a) Ser considerado capaz de decidir. (b) Possuir o conhecimento e a informação adequada. (c) Apresentar estabilidade emocional. (d) No caso dos procuradores de cuidados de saúde, só não deverão ter conflitos de interesses com o paciente, como lhe deverão ser favoravelmente comprometidos e parciais.

Em adição às aplicações do Princípio da Não Maleficência na *prática clínica*, deverão ser abordadas duas ideias importantes, relacionados com a *proteção na investigação biomédica*.

A *subproteção* dos sujeitos estudados pela investigação científica foi e é uma grande e válida preocupação coletiva, que surge principalmente em resposta às execráveis experiências científicas realizadas em humanos até meados do século XX. No entanto, a *superproteção*, muitas vezes esquecida, também causa efeitos danosos. Regulamentações excessivamente rígidas, agravadas pela distância dos reguladores às atividades que regulam e alimentadas por vieses cognitivos como a aversão à perda, originam restrições à investigação científica que limitam a jusante o acesso a tratamentos e práticas médicas inovadoras, com graves consequências para os pacientes que delas poderiam beneficiar. (Warlow, 2005)

Outro tipo de dano, por vezes ignorado, advindo da investigação científica pouco cautelosa é o dano imposto a grupos de pessoas, que por serem mais difíceis de quantificar do que danos diretos a indivíduos, poderão passar mais despercebidos. Este pode ser induzido

[18] A análise que farei posteriormente não requer a aceitação ou recusa desta tese, cujo escopo ultrapassa a proposta deste trabalho.

por várias ações ou decisões moralmente questionáveis, como a utilização de amostras ou dados populacionais para objetivos distintos dos inicialmente divulgados, que mesmo anonimizados podem afetar adversamente interesses pessoais e coletivos, além de anular a validade de qualquer consentimento informado que tivesse sido previamente obtido. Um exemplo ilustrativo foi a recolha dos dados genéticos dos índios Havasupai, no contexto de um estudo sobre diabetes. Estes dados foram, posteriormente, utilizados para outras finalidades, sem autorização, como estudos sobre doenças psiquiátricas e prevalência de consanguinidade. Estes estudos desrespeitaram a autonomia dos indivíduos destes grupos, usando injustificadamente os seus dados. Os resultados destes estudos foram muito controversos, tendo sido considerado que contribuíram para a estigmatização desta tribo e tendo afetado negativamente a sua identidade e autorrepresentação. (Sterling, 2011) Destaca-se assim a necessidade de se ser criterioso e cuidadoso, protegendo do dano tanto indivíduos, como grupos.

O Princípio da Não Maleficência estabelece a obrigação *prima facie* de não causar dano. Nos casos em que este é inevitável, deverá haver uma justificação ética criteriosa, que pondere os benefícios significativos contra os possíveis prejuízos. Impõe também a não imposição de riscos injustificados, condenando a ação negligente. A sua correta aplicação requer a contínua vigilância e diligência por parte dos profissionais de saúde, para que se garanta que os cuidados prestados respeitam os elevados padrões da ética médica e protegendo a integridade dos pacientes.

I.2.3 - Princípio do Respeito pela Autonomia

O *Princípio do Respeito pela Autonomia* enfatiza a importância de respeitar a capacidade dos indivíduos de tomar decisões autónomas sobre a sua própria vida e saúde. A *autonomia* é sustentada por dois seus pilares essenciais: *liberdade* e *agência*:

- A liberdade diz respeito à independência de influências controladoras externas, permitindo que o indivíduo faça escolhas sem coerção ou manipulação.
- A agência refere-se à capacidade do indivíduo de agir intencionalmente, com base em sua própria compreensão, deliberação e razão.

No contexto da Saúde, além dos pilares conceptuais da autonomia, é essencial avaliar a *autonomia das escolhas*. Indivíduos geralmente autónomos podem enfrentar situações que limitam o seu julgamento, tal como a doença[19], o desconhecimento ou a coerção[20].

Respeitar agentes autónomos implica reconhecer e valorizar o direito de um sujeito de possuir as suas próprias ideias e pontos de vista e, tendo capacidade de tomar decisões, agir em conformidade com estas. O respeito pela autonomia é demonstrado *ativamente*, através de ações concretas que dignificam e destacam as opções dos pacientes, não apenas mantendo uma "atitude respeitosa". Opostamente, temos atitudes e ações que insultam, diminuem ou não reconhecem os direitos dos outros de escolher e agir autonomamente. O Princípio do Respeito pela Autonomia envolve então reconhecer o *valor intrínseco da pessoa autónoma* e *o seu direito de fazer escolhas*.

O respeito pela autonomia exige dois tipos de obrigações:

- *Obrigação negativa:* As ações autónomas não devem ser restringidas ou controladas por limitações impostas por terceiros. Isso implica eliminar proactivamente qualquer

[19] Por exemplo, um paciente cuja autonomia é temporariamente reduzida por um Acidente Isquémico Transitório.

[20] Por exemplo, uma pessoa autónoma incitada a assinar um consentimento escrito sem o ler.

forma de coação ou restrição externa que comprometa a capacidade do indivíduo de tomar as suas próprias decisões de maneira livre e informada.

- *Obrigação positiva*: A escolha autónoma deve ser promovida e facilitada, mantendo-se um compromisso com a transparência e a verdade, disponibilizando de informação e apoiando o processo de tomada de decisão e tratando os indivíduos como fins em si mesmos, auxiliando-os a alcançar os seus próprios objetivos e não como meios para alcançar fins de terceiros.

A *ação autónoma* pressupõe a existência das seguintes condições:

- *Intenção*: Inicialmente deverá existir um planeamento da ação, incluindo uma representação mental dos eventos futuros. A ação é, posteriormente, executada segundo este planeamento. Deve-se frisar que a materialização ou não dos resultados pretendidos, ou até mesmo a ocorrência de resultados indesejados pelo sujeito, não invalidam a intencionalidade da ação se estes eram previsíveis durante o planeamento.[21]

- *Compreensão*: A compreensão *plena* é muitas vezes inexequível e não é necessária à autonomia, será suficiente que o sujeito esteja *substancialmente* ciente do que a sua ação envolve e das suas potenciais consequências, possuindo informação suficiente para poder avaliar e decidir.

- *Ausência de influências controladoras*: O agente deverá estar livre de pressões externas ou estados internos que controlem ou restrinjam de forma significativa a liberdade da

[21] Uma pessoa que conduz negligentemente embriagada, sabendo os perigos que essa ação poderá acarretar, continua a ter agido de forma intencional se a sua ação resultar num dano ou prejuízo de outrem, mesmo que não o tenha desejado.

sua ação[22]. As diferentes magnitudes destas influências controladoras, condicionarão diferentes graus de autonomia da ação.[23]

Do Princípio do Respeito pela Autonomia são derivadas várias regras e normas orientadoras de conduta profissional, como: (a) Dizer a verdade, dando informações precisas e completas, necessárias à tomada de decisões informadas. (b) Respeitar a privacidade, evitando intromissões não consentidas. (c) Proteger a confidencialidade, não divulgado informações sensíveis confiadas no contexto das relações médico-paciente. (d) Obter consentimento informado, garantindo que os indivíduos entendem e aceitam voluntariamente intervenções propostas. (e) Apoiar a tomada de decisões importantes, quando solicitado pelo paciente.

Existem situações em que o consentimento pode não ser necessário ou possível de obter, tal como emergências médicas, questões de saúde pública, ou em investigações que utilizem dados adequadamente anonimizados e aleatorizados, impossibilitando a identificação dos pacientes. Ainda assim a possível interferência com um princípio *prima facie* deve motivar a minimização de qualquer impacto negativo sobre a autonomia dos sujeitos envolvidos. Logicamente que as normas e regras do respeito pela autonomia pressupõe que a pessoa é suficientemente autónoma para participar ativamente nas decisões que a afetam, o que pode não ser sempre o caso.

A *capacidade* é um conceito essencial na autonomia, expressando a aptidão de um indivíduo para fazer escolhas autónomas. *Juízos de capacidade* feitos por profissionais de saúde têm uma função normativa significativa, pois qualificam ou desqualificam indivíduos para tomar certas decisões ou realizar determinadas ações. O grau de capacidade de decisão

[22] Exemplos destas influências são a manipulação, mentira, desinformação ou coação.

[23] Um sujeito que age em prejuízo de outrem para mitigar o risco de sofrer uma perda financeira comete uma imoralidade superior à de outro que age da mesma forma sob uma ameaça grave à sua integridade física.

e ação de um individuo varia consoante a sua natureza, contexto e magnitude dos riscos que acarretam.[24]

A avaliação da capacidade é mais facilmente realizada pela negação do seu contrário, isto é, a ausência de sinais de *incapacidade*, que incluem a dificuldade ou impossibilidade de: (a) Expressar ou comunicar preferências ou escolhas. (b) Compreender a sua própria situação e as consequências das suas decisões. (c) Compreender informações relevantes para a tomada de decisão. (d) Fornecer razões lógicas para suas escolhas. (e) Manter um pensamento racional ao considerar opções. (f) Realizar análises de risco e benefício. (g) Chegar a uma decisão racional e justificada.

Para formular *juízos normativos de incapacidade*, os profissionais devem:

1. Identificar que competências físicas e mentais são relevantes para o tipo de capacidade pertinente à situação.
2. Determinar o grau mínimo necessário dessas competências para se considerar uma pessoa capaz.
3. Selecionar e aplicar testes empíricos adequados para avaliar essas competências.

Neste contexto, pode ser utilizar uma escala ajustável, exigindo tanto maior grau de evidência de que os testes empíricos são adequados para a declaração do sujeito como capaz, quanto maior for o risco associado à decisão que este deve tomar. Esta abordagem evita a estratificação rígida da capacidade em si, compatibilizando a sua determinação empírica para fins de qualificar ou desqualificar sujeitos da toma de certas decisões, mas reconhecendo que não sendo esta um atributo estático e totalmente objetivável, a restringência dos seus requisitos deverá reduzir quanto menos arriscada forem as decisões.

[24] Um doente severamente deprimido internado por elevado risco suicida é, paternalisticamente, privado da sua autonomia de ir para casa. Ainda assim, deverá ser respeitada a autonomia de outras suas escolhas de menor impacto, como o tipo de dieta que deseja comer no hospital ou que tipo intervenções, não vitais, está disposto a aceitar. A capacidade também varia ao longo do tempo, neste caso, se com a terapêutica instituída desaparecer o risco de vida eminente, retomando o paciente, comprovadamente, à sua plena capacidade, este poderá, autonomamente, optar por regressar a casa, mesmo contra o parecer médico.

Uma outra demonstração primária de respeito pela autonomia do paciente é o *consentimento informado*. Este assegura que decisões do paciente são tomadas de forma livre e informada.[25] Para que o processo de consentimento informado não seja apenas um procedimento formal, mas um processo moral, devemos objetivar a presença das suas componentes fundamentais, que podem ser estruturadas em três categorias principais:

1. Elementos Base:

   O consentimento informado começa com a verificação da *capacidade* do paciente para entender e decidir conscientemente sobre seu tratamento e cuidado. Segue-se a avaliação do *voluntarismo*, assegurando que suas decisões sejam tomadas livremente, sem a presença de qualquer forma de coerção ou manipulação.

2. Elementos de Informação

   A segunda categoria trata da *divulgação de informação material* necessária para que o paciente faça uma escolha informada, explicando os detalhes do tratamento de forma clara, acessível e personalizada para atender às necessidades individuais do paciente.[26] A *compreensão* deste conjunto de informações é vital. Não é necessário que o paciente compreenda todos os detalhes técnicos, mas é essencial que ele entenda os aspetos fundamentais para a decisão, como diagnósticos, prognósticos, natureza e propósito da intervenção, alternativas, e os riscos e benefícios envolvidos.

3. Elementos de Consentimento:

[25] Existem várias outras justificações para o consentimento informado: (a) *Transparência*, facilitando a clareza e a abertura na comunicação entre médicos e pacientes. (b) *Autorização*, dando aos pacientes controlo sobre os procedimentos médicos a que serão submetidos. (c) *Concordância com os valores do paciente*, assegurando que as intervenções médicas estejam alinhadas com os valores e preferências do paciente. (d) *Promoção do Bem-Estar*, salvaguardando os interesses do paciente, minimizando riscos e maximizando os benefícios. (e) *Promoção da confiança*, fortalecendo a relação médico-paciente através da demonstração ativa de respeito pelas escolhas do paciente. (f) *Promoção da integridade*, mantendo os padrões éticos superiores da prática médica.

[26] Incluindo: (a) Os factos que o paciente e o profissional consideram cruciais para a decisão. (b) As recomendações profissionais e a justificativa para o pedido de consentimento. (c) A natureza do consentimento e os limites da autorização concedida.

O paciente deve tomar uma *decisão* informada sobre prosseguir ou não com a intervenção, culminando na *autorização* formal do tratamento proposto.

Por fim, devemos abordar os casos nos quais os pacientes não são autónomos ou quando há dúvidas quanto à sua autonomia. Em casos de grande diminuição ou ausência de capacidade, haverá a necessidade de delegar as decisões de saúde relativas do paciente a um *representante*, o procurador de cuidados de saúde, que poderá ser um familiar, médico ou pessoa de confiança do paciente. Existem três principais padrões propostos, que regem moralmente o processo da tomada de decisão em representação do paciente não autónomo:

- *Padrão do Julgamento Substitutivo*. As escolhas do representante deverão ser baseadas no que o paciente teria decidido se estivesse em condições de exercer a sua autonomia. Para isso será essencial que o representante tenha um conhecimento profundo dos valores e preferências do paciente. Quando as vontades do paciente não são conhecidas ou são ambiguamente definidas, é exigida uma interpretação cuidadosa das evidências disponíveis sobre as suas preferências previamente expressas.

- *Padrão de Autonomia Pura*: Destina-se aos casos nos quais os pacientes, antes de perderem a capacidade de decisão, expressaram claramente as suas preferências quanto aos tratamentos que pretendem aceitar. Este padrão exige que se respeite essas decisões prévias, mesmo que o paciente não possa mais manifestá-las. Todavia, esta abordagem revela-se problemática quando o contexto pessoal do paciente se altera significativamente, levando a que se questione e se reavalie a aplicabilidade das diretivas antecipadas, que terão sido feitas num contexto totalmente díspar e poderão prejudicar irreversivelmente o bem-estar do paciente.

- *Padrão do Melhor Interesse*: Aplica-se quando não é possível determinar claramente quais seriam as preferências do paciente se fosse autónomo[27]. Este padrão requer que o representante avalie e decida com intuito de maximizar o benefício provável,

[27] Em pacientes que nunca foram verdadeiramente autónomos, por exemplo, por terem nascido com defeitos cognitivos graves.

considerando os interesses do paciente face aos riscos e os custos associados. É particularmente relevante em situações nas quais o paciente nunca foi capaz de expressar suas vontades de forma autónoma ou quando as preferências anteriores não são diretamente aplicáveis às circunstâncias atuais.

Em todos estes casos, a tomada de decisão em nome do paciente requer a análise meticulosa do seu estado de saúde e uma compreensão profunda dos seus interesses, valores éticos e vontades, primando unicamente pelo seu bem-estar e assegurando que as decisões se alinham, tanto quanto possível, com o que seria a vontade do paciente, caso este pudesse decidir por si próprio.

I.2.4 - Princípio da Justiça

O *Princípio Formal da Justiça*, tradicionalmente atribuído a Aristóteles, estipula que "tratar *igualmente* o que é *igual* e *desigualmente* o que é diferente". A *formalidade* deste princípio advém da sua ausência de substância, não dita critérios específicos para determinar o que são iguais, ou o que é o tratamento igual, mas tece um requerimento mínimo de justiça, generalista.[28] Será, por isto, necessário definir *iguais* e *desiguais* e determinar o que consiste em o *tratamento igual* ou o *tratamento desigual*.

Os *Princípios Materiais de Justiça* fornecem este conteúdo, que o Princípio Formal carece. Só desta forma poderemos encontrar normas e regras que orientem a Sociedade na prática, onde a *escassez* de recursos dita que se façam escolhas difíceis; muitas vezes existe a necessidade de sacrificar alguns benefícios para atender a outros mais prementes.[29] O *Princípio da Justiça* na Saúde, carece especialmente de conteúdo, pois num contexto onde as necessidades são virtualmente infinitas, e os recursos são tragicamente finitos, há uma necessidade acrescida de estabelecer especificações e ponderações, para que se possam fazer decisões distributivas justas.

Os Princípios Materiais de Justiça *não são universais*, variando significativamente de acordo com as diferentes *Teorias de Justiça* que os fundamentam. Essas teorias influenciam a forma como os recursos são distribuídos, determinando quem e em que medida os deve receber.

[28] A *Justiça* pode ser interpretada como o tratamento de indivíduos e grupos de forma equitativa e apropriada, em conformidade com o que lhes é devido. Já a *Justiça Distributiva* refere-se à distribuição equitativa e apropriada dos *benefícios* e *encargos* sociais, conforme determinado pelas normas que estruturam a cooperação entre os membros de uma sociedade.

[29] Um exemplo é o *Princípio da Necessidade*, que prioriza a atribuição de recursos indispensáveis às pessoas que, sem eles, enfrentariam efeitos negativos substanciais em áreas cruciais de suas vidas, como a sua saúde ou o seu bem-estar. Também constituem exemplos os critérios que determinam a *elegibilidade* para receber certos recursos. Como a cidadania enquanto critério para o acesso a tratamentos médicos específicos, como transplantes de órgãos, num determinado país.

Estes espelham diferentes prioridades e valores éticos que podem determinar a elegibilidade para receber recursos.[30]

A análise detalhada das diversas conceções de justiça distributiva excede largamente o escopo deste trabalho, mas será importante explorar o tópico da distribuição dos recursos de saúde. A acessibilidade e qualidade dos cuidados de saúde prestados aos cidadãos é uma das preocupações de muitos países. No entanto, não sendo a Saúde a única prioridade e necessidade de uma sociedade, deverá também haver a proteção dos recursos públicos, através da contenção de custos.

Os principais objetivos tendem a ser garantir o *acesso equitativo a saúde de qualidade* e a promoção de um bom estado de saúde geral, mas preservando a competitividade do mercado, para que haja *liberdade de escolha e eficiência*. Estes objetivos entram muitas vezes em conflito[31], exigindo uma reflexão e adaptação contínua das políticas que procuram equilibrar as diferentes necessidades da sociedade, de forma justa e equitativa

A maioria dos sistemas de saúde tenta equilibrar estes objetivos distinguindo dois níveis de cuidados: (a) O acesso universal a um mínimo essencial de cuidados de saúde, que implica uma cobertura social obrigatória para as necessidades de saúde de primeira necessidade. (b) Outras necessidades adicionais de saúde poderão ser procuradas individualmente, através da cobertura privada voluntária, como quartos de hospital mais luxuosos ou procedimentos estéticos não reconstrutivos.

Será também de relevo explorar os diferentes níveis de alocação de recursos na saúde, desde o nível macroeconómico, com a alocação dentro dos orçamentos de estado, até à alocação de recursos a programas e procedimentos específicos:

[30] Como no exemplo anterior, a exigência de cidadania para o acesso a transplantes de órgãos pode ser defendida sob algumas perspetivas, mas inaceitável sob outras.

[31] Por exemplo, a contenção de custos pode limitar o acesso a certos tratamentos, ou a liberdade de escolha dos pacientes e médicos pode, por vezes, entrar em conflito com a necessidade de eficiência social e de regulação de custos.

- *Repartição do Orçamento de Estado*: Decisões sobre quanto do orçamento nacional deve ser alocado à saúde versus outros bens sociais, como educação, segurança e infraestruturas. A saúde é um valor crucial, mas não é o único que a sociedade precisa defender e promover.

- *Alocação Dentro do Orçamento da Saúde*: Distribuição dos fundos dentro do setor da saúde, incluindo não apenas os cuidados médicos diretos, mas também políticas e programas de saúde pública, proteção civil, prevenção de acidentes e regulação de medicamentos e alimentos. Esta alocação requer uma análise cuidadosa das necessidades e dos benefícios potenciais de cada área.

- *Alocação Dentro do Orçamento Alvo*: Decisões sobre quais projetos específicos ou procedimentos devem receber financiamento. Por exemplo, determinação de que categorias de doenças e condições devem ser prioritárias baseadas em critérios como frequência, impacto social, dor e sofrimento associados, e influência na longevidade e qualidade de vida.

- *Alocar Tratamentos Escassos aos Doentes*: Sendo as necessidades em saúde virtualmente infinitas, mas os recursos limitados, torna-se necessário decidir que tratamentos fornecer e a quem. Este nível de alocação inclui as escolhas difíceis sobre quem receberá tratamentos possivelmente vitais, quando não é possível providenciá-los a todos que os necessitam.

Os dois últimos tipos de alocação, mais próximos da prestação de cuidados de saúde em si, são operacionalizados através de processos como a *priorização* e o *racionamento*.

A *priorização* emprega sobretudo as análises de utilidade, como a análise de custo-benefício e custo-eficácia.[32] Estas metodologias são essencialmente utilitaristas, pretendendo maximizar os benefícios de saúde, face ao com o investimento realizado. Estas análises podem evidenciar ineficiências na forma como se fazem as alocações. Um exemplo paradigmático é a

[32] Vide I.2.1 – Princípio da Beneficência.

tendência ineficiente de os serviços de saúde tenderem a alocar mais recursos ao *tratamento* das doenças, em detrimento da sua *prevenção*, pela dificuldade que requer tomar e justificar decisões menos intuitivas, nas quais se privilegiam vidas futuras, calculadas estatisticamente, ao invés de tratar indivíduos específicos, reais, que o necessitam de recursos de imediato, apesar das análises de utilidade revelarem que que a primeira abordagem é mais custo-eficaz.

As vantagens como a praticidade e aplicabilidade destas análises não devem, no entanto, esconder o seu potencial para infringir princípios éticos pela impossibilidade de abranger, de forma absolutamente precisa e objetiva, todos os fatores em jogo, que não se esgotam em objetivos clínicos. Análises que consideram a qualidade de vida, por exemplo, podem ser inadequadas para grupos onde esta é mais difícil de determinar como idosos, crianças ou pessoas com deficiências, originando decisões que os discriminam ou prejudicam. É então fulcral aliar às ferramentas analíticas um enquadramento bioético robusto ao definir prioridades na saúde de forma justa.

O *racionamento* de recursos é entendido e proposto de várias formas, consoante as diferentes teorias éticas e de justiça dos seus proponentes: (a) O acesso aos tratamentos é definido pela capacidade individual de pagar (sendo o mercado que o regula através de mecanismos de preço). (b) O governo estabelece o limite máximo de tratamentos ou recursos de saúde que cada individuo pode aceder. (c) O acesso a uma quantidade fixa de recursos de saúde considerados imprescindíveis é igualitário; outros recursos adicionais podem ser acedidos por indivíduos que os possam pagar.

Outras formas, mais específicas, de racionamento são: (d) A alocação de recursos baseada na *idade* dos indivíduos, considerando-se equitativo ajudar as pessoas a alcançar uma "idade de vida normal", em detrimento de gastar recursos para prolongar a vida além desse ponto. (e) A própria *triagem*, com o objetivo de alocar eficientemente os recursos médicos, baseando-se na utilidade médica, em vez de pela utilidade social.

Os profissionais de saúde e decisores políticos têm, claramente, um papel significativo na alocação de recursos escassos de saúde. A justiça da distribuição destes deve ser garantida estabelecendo critérios, que surgem na forma de *normas substantivas* e *regras processuais*:

Critérios de Elegibilidade: Definem o grupo de potenciais destinatários aptos a receber o recurso de saúde em questão, baseando-se em: (a) Fatores não médicos: Determinação de que grupos poderão receber os recursos, baseando-se, por exemplo, na cidadania ou capacidade financeira do individuo. (b) Progresso científico: A validade da investigação científica de terapêuticas, especialmente em fases experimentais, requer uma seleção criteriosa de que pacientes incluir nos estudos. (c) Perspetiva de Sucesso: A probabilidade de o tratamento efetivamente beneficiar o paciente deve ser determinada e dando-se prioridade aos casos nos quais esta tem maior probabilidade de sucesso, evitando o desperdício de recursos escassos.

Critérios e Processos para a Seleção Final: Uma vez definida a lista de pacientes elegíveis, é necessário determinar quem, dentro desta lista, receberá o tratamento. Os critérios incluem: (a) Utilidade médica: Procura maximizar o número de vidas salvas e a probabilidade de sucesso do tratamento. Este critério é depende de juízos de valor relativos à eficácia potencial dos tratamentos. (b) Necessidade médica: Dá prioridade a pacientes com condições mais graves. (c) Mecanismos impessoais: Incluem lotarias e listas de espera, que podem ser usadas quando a utilidade e necessidade médica é considerada equivalente em todos os pacientes elegíveis.

Outro importante conceito de justiça é o *princípio da oportunidade justa*. Este estipula, segundo John Rawls, que nenhum indivíduo deve ser beneficiado ou prejudicado socialmente devido a propriedades que não controla ou pelas quais não é responsável. Propriedades como género, raça, inteligência, etnicidade, nacionalidade e estatuto social são exemplos de

características que não podem ser eticamente utilizadas para criar princípios materiais de justiça.[33]

Subjacente a este princípio está a *lei da compensação*. Desvantagens moralmente arbitrárias, como deficiências, devem ser mitigadas pela sociedade, oferecendo suporte e ajuda para que os indivíduos prejudicados possam superá-las ou, pelo menos, reduzir o seu impacto, para que possam atingir um nível mais equitativo de bem-estar e participação social.

Infelizmente, também os serviços de saúde e a investigação biomédica podem perpetuar a desigualdade de oportunidade. Todo o tipo de decisão pouco criteriosa e mal fundamentada moralmente pode condicionar a não equidade no acesso de certas comunidades e grupos aos cuidados de saúde, ou resultar na sua subvalorização na investigação biomédica. Importa referir, no entanto, que as desigualdades não são todas inerentemente injustas, mas é imperativo compreendê-las e identificar as suas causas, para detetar e prevenir possíveis injustiças.[34] Este escrutínio é por vezes dificultado pela existência de *vieses implícitos*, que influenciam, injustamente, a alocação de recursos e oportunidades de maneira subtil e indireta, muitas vezes beneficiando desproporcionalmente aqueles que já são favorecidos socioeconomicamente. A identificação dos vieses explícitos e implícitos é crucial para colmatar desigualdades originárias de situações injustas.

Na investigação biomédica é possível encontrar vários destes vieses, explícitos e implícitos que poderão originar situações injustas, como as que ficam espelhada na própria escolha de sujeitos participantes em estudos de investigação. A *sobrerrepresentação* de participantes em condições de "vulnerabilidade"[35] socioeconómica deve ser analisada, particularmente os fenómenos de *indução injusta e lucro indevido*.

---

[33] Seria eticamente inaceitável existir um princípio como "Para cada um, consoante o seu QI/Género/Raça/etc.".

[34] A eficácia de um medicamento pode variar consoando diferenças biológicas e poderá condicionar justamente uma diferente abordagem terapêutica. O condicionamento da abordagem terapêutica por meras questões socioculturais, irrelevantes para a eficácia do mesmo já não será moralmente justificável.

[35] O próprio rótulo "vulnerável" pode, paradoxalmente, desqualificar a escolha autónoma de grupos inteiros, por mecanismos de sobreproteção, estereotipagem e paternalismo.

Indução injusta aplica-se quando indivíduos com dificuldades económicas são levados a aceitar riscos elevados pela oferta de compensações financeiras altamente atrativas, decisão que não tomariam se não tivessem essas dificuldades, o que pode ser considerado coercivo.

Lucro indevido, por outro lado, aplica-se quando os participantes recebem uma compensação financeira desproporcionalmente baixa, quando comparada com os retornos dos promotores da investigação. Participantes com carências económicas substanciais, poderão ter pouco poder negocial, sendo-lhes oferecido um valor que não reflete justamente os riscos que estes assumem.

O justo equilíbrio e proporcionalidade da compensação oferecida aos participantes exige que, por um lado os pagamentos não sejam tão atrativos que coagem os indivíduos a aceitar riscos não razoáveis, por outro deverão ser suficientemente altos para prevenir o aproveitamento e exploração dos participantes.

O princípio da justiça exige uma análise cuidadosa e contínua sobre como os recursos de saúde são alocados e como as decisões médicas afetam diversos grupos dentro da sociedade. A justiça em saúde não é apenas uma questão de distribuição de recursos, mas também de garantir que todos tenham a oportunidade de alcançar o melhor estado de saúde possível, independentemente de suas circunstâncias pessoais ou sociais. Este princípio desafia os profissionais de saúde a considerar não apenas o que é clinicamente eficaz, mas também o que é eticamente justo, garantindo que a prática médica promova uma sociedade mais justa e equitativa.

Capítulo II – Inteligência Artificial

A linguagem e os termos comummente utilizados na área da IA são fortemente envoltos em ambiguidade e até mistificação, em grande parte pela utilização de palavras, como inteligência, pensamento, raciocínio, aprendizagem, até aqui apenas aplicadas ao ser humano, às suas faculdades e ações. Estes termos provem mal-entendidos, desinformação e a construção de crenças falsas, adulterando a perceção pública perante o que constitui a IA e quais as consequências futuras das suas aplicações práticas nos diversos setores sociais.

Começar por definir adequadamente o que é a Inteligência Artificial torna-se premente à sua análise rigorosa. Neste capítulo serão explanados os principais conceitos e termos técnicos e detalhados os subtipos de IA, com enfoque no seu funcionamento de base e nas suas aplicações práticas, especialmente nas áreas médicas. Serão, também, levantados e desconstruídos mitos e crenças falsas frequentemente associados à IA. Constituir-se-á, desta forma, um substrato conceptual e técnico, tão detalhado e específico quanto a aplicação prática e específica de regras e normas derivadas dos princípios bioéticos exija, evitando a paralisia e imprecisão fomentada pela ambiguidade e falta de clareza conceptual.

II.1. - O que é?

Definir IA é uma tarefa exigente, em nada facilitada pela utilização da expressão "Inteligência", que nem na forma humana é totalmente compreendida, sendo estudada por áreas díspares como Ciências Cognitivas, Neurociências, Psicologia, Filosofia da Mente e Epistemologia, sem existir uma definição unificadora e plenamente satisfatória. No entanto, para esta análise bioética será suficiente assumir as formas mais restritas de Inteligência que a IA representa.

A exigência desta tarefa suscitou inúmeras tentativas de definir IA, a minha a análise partirá da adaptação da definição proposta pelo Grupo Independente de Peritos de Alto Nível criado pela Comissão Europeia em 2018, a qual complementarei e esclarecerei conforme necessário:

> Os sistemas de inteligência artificial (IA) são **sistemas** de *software* [e *hardware*] concebidos por seres humanos, que, tendo recebido um **objetivo complexo**, atuam na dimensão física ou digital **percecionando** o seu ambiente mediante a aquisição de dados, interpretando os **dados estruturados** ou **não estruturados** recolhidos, **raciocinando sobre o conhecimento** ou processando as informações resultantes desses dados e decidindo as melhores ações a adotar para atingir o objetivo estabelecido. Os sistemas de IA podem utilizar **regras simbólicas** [lógicas] ou aprender um modelo numérico, bem como adaptar o seu comportamento mediante uma análise do modo como o ambiente foi afetado pelas suas ações anteriores.
>
> Enquanto disciplina científica, a IA inclui diversas abordagens e técnicas, tais como a **aprendizagem automática** (de que a **aprendizagem profunda** e a **aprendizagem por reforço** são exemplos específicos), o **raciocínio automático** (que inclui o planeamento, a programação, a representação do conhecimento e o raciocínio, a pesquisa e a

otimização) e a **robótica** (que inclui o controlo, a perceção, os sensores e atuadores, bem como a integração de todas as outras técnicas em sistemas *ciberfísicos*).» (GIPAN2018, 2018)

Será necessário esclarecer os termos aqui utilizados e caracterizar os princípios técnico-científicos por detrás destes sistemas[36]. Começando pelo termo "**sistema**", que traduz a forma como a IA não existe num vácuo, mas como componente integrado noutros sistemas tecnológicos maiores, sejam eles *software* ou *hardware*. Estes sistemas são desenvolvidos para alcançarem **objetivos complexos**, definidos pelos seus desenvolvedores e, por isso, refletindo as suas motivações, crenças e valores.[37] Para alcançar os objetivos programados, os sistemas de IA devem realizar vários passos técnicos que importam agora detalhar:

Perceção. Capacidade dos sistemas de IA de extrair *inputs* do ambiente onde estão inseridos, seja ele digital ou físico, para que os possa posteriormente analisar, interpretar de acordo com a sua programação e treino, tomar decisões com base nesta interpretação e executar por fim a ação que melhor realiza o seu objetivo. Exemplificando, um sistema de IA em radiologia utiliza sensores, neste caso, os *scanners* de raios-X, para capturar as imagens que serão transformadas em dados digitais, para que possa posteriormente analisá-las e detetar sinais de fraturas ósseas.

Raciocínio. Após a recolha de dados, estes terão de ser processados e transformados em informação compreensível por um computador digital, isto é, transformar realidades físicas em valores numéricos, em última análise binários, isto é, zeros e uns. Esta informação, chamada de *conhecimento*, é o substrato utilizado para raciocinar, o que envolve realizar inferências através de regras lógicas, pesquisar um vasto conjunto de soluções possíveis e

[36] A terminologia da IA é continuamente atualizada, podendo haver variações nos termos ao longo do tempo, o Conselho de Comércio e Tecnologia EU-EUA mandatou um subgrupo para que atualize e categorizes esses termos. (EU-U.S. Trade and Technology Council, 2023, 2024)

[37] A sua orientação para a execução mais eficiente de objetivos pré-definidos, mesmo não resultantes de uma motivação própria, tem sérias implicações bioéticas, como veremos posteriormente.

selecionar a melhor. Por exemplo, um sistema de IA utilizado em diagnósticos de emergência recebe *inputs*, como sintomas, sinais vitais, dados de ECG, níveis de enzimas cardíacas, antecedentes do paciente, etc. e usando regras lógicas pode inferir se os *inputs* são consistentes com um enfarte agudo do miocárdio e sugerir a melhor abordagem terapêutica.

Aprendizagem. Será claro que existem problemas, que devido à sua grande complexidade, não poderão ser adequadamente especificados e traduzidos ao detalhe por regras lógicas, problemas tais como o processamento de linguagem, de fala, de imagens ou análises preditivas baseadas em grandes quantidades de dados. Estas capacidades aparentam ser fáceis para os seres humanos, mas capacidades essenciais a estas tarefas, como o *senso comum*[38] e a interpretação de *dados não estruturados*[39], demonstram-se tecnicamente difíceis de instituir na programação de sistemas de IA. Nestes casos utilizam-se técnicas de **aprendizagem automática** (*machine learning, ML*), incluindo a **aprendizagem supervisionada** (*supervised learning*), a **aprendizagem não supervisionada** (*unsupervised learning*) e a **aprendizagem por reforço** (*reinforced learning*)**.** [40]

Aprendizagem supervisionada. Nos sistemas que utilizam esta técnica, em vez de se detalharem todas as regras lógicas que o sistema deve seguir para chegar a uma decisão, são lhe fornecidos inúmeros exemplos de *inputs* e os respetivos *outputs* desejados, para que este estabeleça de alguma forma correlações probabilísticas que utiliza para generalizar, atribuindo *outputs* corretos a *inputs* novos, não utilizados no seu *treino*. Num exemplo médico, poderíamos fornecer várias radiografias de tórax onde se podem visualizar pneumotórax

[38] Uma pessoa ao se aperceber que a sua cozinha está a arder, irá parar prontamente de cozinhar e chamar os bombeiros. Se a polícia bater à porta da casa de alguém, esta sairá da cozinha para ir abrir a porta. O ser humano tem a capacidade de fazer uma análise prática, *in situ*, e estabelecer prioridades, mesmo em situações que nunca anteviu e por isso para as quais não se preparou. Este tipo de sabedoria prática, a *phronesis* aristotélica, é particularmente difícil de incutir numa IA, que tendo o objetivo de cozinhar, manter-se-ia na cozinha em chamas até inevitavelmente perecer. (Hasselberger & Lott, 2023)

[39] Dados que não apresentam uma organização pré-definida, como uma imagem ou um discurso, em oposição aos *dados estruturados*, por exemplo, obtidos através de formulários.

[40] Outras técnicas que fazem parte da aprendizagem automática, mas cujo funcionamento específico não apresenta relevância para esta análise bioética são: redes neuronais (*neural networks*); aprendizagem profunda (*deep learning*); visão computacional (*computer vision*); processamento de linguagem natural (*natural language processing*).

(*input*) e o respetivo relatório imagiológico que relata a presença de um pneumotórax (*output*); o que se pretende é que, depois do treino, a IA estabeleça correlações entre as informações que extrai das imagens e as informações que extrai dos respetivos relatórios, para que consiga ela própria relatar corretamente uma nova radiografia, não usada no seu treino.

Aprendizagem não supervisionada. Em circunstâncias onde se pretende a análise de inúmeros dados, mas sem se saber com certeza o que é que se deseja encontrar, pode-se usar este tipo de método, onde os dados introduzidos nos sistemas não são rotulados e como tal não existe fase de treino. Os desenvolvedores e investigadores poderão posteriormente interpretar os grupos de dados que o sistema formou. Em estudos epidemiológicos, por exemplo, sistemas de IA podem ser empregues para analisar extensos bancos de dados de saúde populacional, onde os dados não estão pré-rotulados. Nestes são identificados padrões e agrupamentos, por exemplo, subgrupos de pacientes com trajetórias similares na progressão de doenças crónicas, como a diabetes, podendo revelar novos subtipos da doença, que necessitem de abordagens terapêuticas diferenciadas.

Aprendizagem por reforço. Nesta técnica, dá-se liberdade ao sistema para que tome decisões, programando-se *recompensas* e *punições* conforme a adequação das decisões e outputs aos objetivos desejados. Com tempo suficiente, o sistema tende a privilegiar estratégias que maximizem o saldo positivo de recompensas, por exemplo, um sistema de IA pode aprender a ajustar dosagens de medicamentos ao receber *feedback* positivo quando os sintomas do paciente melhoram sem efeitos colaterais e *feedback* negativo quando ocorrem reações adversas. Com cada nova experiência, o sistema refina suas previsões e decisões para alcançar melhores resultados terapêuticos.

Robótica. Associa-se aos sistemas de IA um *robô* (máquina física), dando-lhe uma interface com o mundo físico. Exemplificando, um robô utilizado para entrega de suprimentos médicos, equipado com sensores de visão e um sistema de IA capaz de detetar obstáculos, identificar trajetórias seguras e reconhecer sinais de alerta, como a presença de pessoas ou equipamentos médicos, possibilitando que se mova de forma segura e eficiente, em contexto hospitalar.

Explicados e ilustrados os conceitos mais relevantes e com uma ideia mais concreta do que é a IA, será agora importante esclarecer o que esta não é. A discussão alargada sobre a IA é dominada por ideias erradas e mitos que ocultam muitas vezes problemas concretos e realísticos de sistemas já implementados ou prestes a sê-lo, em troca de teorias distópicas e utópicas. Estes mitos podem promover importantes conflitos éticos, especialmente em áreas como a Medicina, nas quais a desinformação pode levar a consequências desastrosas. Nesta secção pretende-se apresentar, esclarecer e ilustrar mitos comuns, utilizando como base o trabalho desenvolvido em (Haroon, Corien, & Erik, 2023).

II.2 - O que não é?

Terminologia. Muitos dos mitos e equívocos associados à IA decorrem da terminologia que esta disciplina emprega. A começar pelo termo "inteligência", que encerra uma significativa complexidade ontológica e epistemológica. A sua utilização para descrever entidades artificiais suscita múltiplas implicações filosóficas, religiosas/espirituais e sociais, especialmente porque a inteligência ocupa um papel central na nossa autorrepresentação e na ontologia humana. A secção anterior ilustrou a comum formulação de termos técnicos sobre IA por analogia a faculdades humanas, como "perceção", "raciocínio" e "aprendizagem". Esta abordagem promove o equívoco de que tais capacidades em sistemas de IA operam de modo semelhante às humanas, o que é, na realidade, uma presunção incorreta.

*Neutralidade***.** Há uma perceção comum de que, por não possuírem personalidade, emoções, motivações ou ideologias, os sistemas de IA estariam mais bem capacitados para tomar decisões objetivas. No entanto, existem várias fases do seu desenvolvimento que os podem permear de subjetividade, vieses e ideologias levando a que as suas decisões sejam tudo menos neutras. Como pudemos ver, o seu desenvolvimento envolve escolhas e deliberações, inerentemente permeadas pelo carácter, motivações, crenças e ideias de quem as faz. Os “dados”, ao contrário do que o nome indica, não surgem de interações neutras e passivas com o meio, as suas limitações seriam provavelmente mais bem expressas pelo termo “captados”, visto que apenas medimos ou captamos através de instrumentos um pequeno subconjunto dos infinitos aspetos do mundo (Greenfield, 2017). A escolha dos sensores é influenciada por considerações subjetivas, como o custo-eficácia, uma tecnologia mais barata pode, assim, ser preferida a uma mais precisa. Também os dados colhidos previamente e usados para treinar o sistema terão diferentes graus de subjetividade, dependente do seu tipo e de como, quando, onde, ao quê e por quem foram colhidos. Dados de má qualidade, contaminados, incompletos ou enviesados irão afetar a qualidade e objetividade do sistema

Mais racional que o ser humano. A perceção de que a IA é dotada de uma racionalidade superior à humana necessita de uma desconstrução meticulosa, particularmente no contexto médico, onde as implicações de tais suposições são profundas. Para esta contribui a

publicidade em torno dos sistemas de IA, que frequentemente desvaloriza a racionalidade humana com argumentos enganosos. Por exemplo, afirmar que "95% dos acidentes de condução são causados por erros humanos" ignora o facto de que, até recentemente, apenas humanos conduziam.

Falsas equivalências e utilizações falaciosas de tipos restritos de inteligência como se fossem a ideia platónica da mesma agravam este mito. O sucesso de programas de IA em jogos estratégicos como o xadrez é muitas vezes extrapolado para sugerir uma suposta superioridade cognitiva geral da tecnologia sobre o ser humano. No entanto, competência numa tarefa altamente específica e estruturada não equivale a uma compreensão e raciocínio verdadeiramente abrangentes e complexos.

Outra área onde a IA é insatisfatória é na correlação e causalidade, visto que (pelo menos) nas suas formas atuais é incapaz de estabelecer cadeias de causalidade. A interpretação errónea de correlações como causas em áreas críticas pode ser muito problemática. A desconsideração pela causalidade abre também um espaço perigoso à pseudociência. Atualmente existem sistemas de IA que tentam determinar emoções através da análise da mímica facial (Ballesteros et al., 2024) ou identificar a orientação sexual de um individuo por reconhecimento facial (Wang & Kosinski, 2018).), estas práticas, além de cientificamente infundadas (Barrett et al., 2019) podem ter implicações sociais e legais graves.

Caixa negra (black box). O termo "caixa negra" sugere que os processos de decisão dos sistemas de IA são opacos e indecifráveis, devido à dificuldade de interpretação dos mecanismos internos de certos modelos de aprendizagem automática, onde grandes volumes de dados são processados através de múltiplas camadas de processamento, podendo desenvolver representações internas que não são facilmente compreensíveis pelos humanos, nem mesmo pelos criadores dos modelos. A opacidade não é, contudo, uma inevitabilidade na IA, mas muitas vezes uma consequência de escolhas técnicas. Existe frequentemente um *trade-off* entre a precisão e a interpretabilidade dos modelos: sistemas mais complexos e menos transparentes podem oferecer uma performance superior, mas essa complexidade reduz a capacidade de os utilizadores entenderem como as decisões são feitas. (Jia et al., 2022) Por

exemplo, um algoritmo pode identificar com alta precisão a presença de tumores em imagens de radiologia usando padrões extremamente subtis nos dados, padrões esses que podem não ser percetíveis ou compreensíveis para um radiologista humano.

Malignidade. Uma ideia frequente no imaginário coletivo, principalmente em debates menos científicos, é a de que a IA pode desenvolver autonomamente intenções ou objetivos malévolos. No entanto, a preocupação subjacente sobre sistemas de IA que atuam de maneiras potencialmente prejudiciais não é totalmente infundada, especialmente quando consideramos o alinhamento dos objetivos da IA com valores humanos éticos e morais.

Os sistemas de IA, não tendo motivações ou vontades próprias, são programados por pessoas, estas sim com motivações e vontades próprias. A escolha, feita pelos desenvolvedores, dos parâmetros e objetivos, pode não estar alinhada com os valores e princípios bioéticos, seja por ignorância, negligência, segundas intenções, conflitos de interesse ou, no extremo, malignidade.

A tecnologia é a solução para tudo. De forma mais generalista, perdura a ideia de que a tecnologia constitui uma panaceia para todos os desafios humanos, levando-nos muitas vezes a negligenciar métodos, já estabelecidos, acessíveis e eficazes, ainda que menos avançados tecnologicamente. Por exemplo, a ressonância magnética e tomografia computadorizada, inegavelmente valiosas, mas em muitos casos, uma anamnese e exame objetivo bem realizados podem ser igualmente eficazes para diagnósticos de certas patologias, como a apendicite aguda, sem o aumento desnecessário nos custos de cuidados de saúde, e sem condicionarem atrasos no tratamento, neste caso, enquanto se espera pelos resultados dos exames de imagem.

II.3 – Sistemas de IA aplicados à Medicina

Finalmente, com uma melhor e mais fundamentada compreensão conceptual e técnica do funcionamento da IA, poderão ser apresentados exemplos concretos de formas como a IA poderá ser aplicada à Medicina. Estes exemplos serão agrupados em quatro principais categorias, consoante o seu objetivo principal:

1. *Apoio ao Diagnóstico.* Os sistemas de apoio ao diagnóstico baseados em IA empregam técnicas de raciocínio e aprendizagem automática para antecipar, monitorizar e diagnosticar uma vasta gama de doenças de forma precoce, precisa e abrangente. Estes sistemas combinam informações detalhadas sobre o paciente com conhecimento médico-científico avançado para promover diagnósticos eficazes e personalizados. Algumas das suas funcionalidades e aplicações são:

- *Interpretação de imagens médicas*. Oncologia: Interpretação de mamografias para a deteção precoce e precisa de neoplasias.[41] (Dembrower et al., 2023)
- *Análise de histórias clínicas e testes laboratoriais*. (a) Medicina Intensiva: Deteção precoce de sépsis. (Huat Goh et al., 2021) (b) Neurologia: Diagnóstico precoce de doenças neurodegenerativas e acidentes vasculares cerebrais (El-Assy et al., 2024), (Hassan et al., 2024).
- *Monitorização contínua*. Neurologia e Psiquiatria: Monitorização de sinais e comportamentos indicativos de declínio cognitivo ou perturbações mentais (Martínez-Nicolás et al., 2021).
- *Exames diagnósticos específicos*. (a) Cardiologia: Interpretação de Eletrocardiogramas (ECG). (Androulakis & Fielder, 2024) (b) Otorrinolaringologia: Rastreios auditivos neonatais (Kumar Panjiyar et al., 2019).

[41] Esta funcionalidade beneficia virtualmente todas as especialidades que requeiram análise de imagens (Yamashita et al., 2018).

- *Testes genéticos e epidemiologia*. (a) Genética: Diagnóstico de doenças genéticas raras. (Abdallah, et al., 2023) (b) Saúde Pública e Infecciologia: Análise de dados populacionais e epidemiológicos para identificar padrões de doenças e responder rapidamente a surtos infecciosos (Raina MacIntyre et al., 2023).

2. *Apoio à Terapêutica.* Os sistemas de apoio à terapêutica baseados em IA recorrem às suas extensas bases de dados e capacidades analíticas para auxiliar na seleção, otimização e personalização dos tratamentos médicos. Estes sistemas são uma ferramenta valiosa tanto na prática clínica quanto na investigação, onde são utilizados para acelerar o desenvolvimento e a integração de novas modalidades terapêuticas. Aplicações destes sistemas incluem: (Alowais et al., 2023; Bajwa et al., 2021; Davenport & Kalakota., 2019)

- *Sistemas de apoio à decisão clínica*. Analisam dados médicos e científicos para recomendar as terapias mais adequadas, ajustadas às necessidades específicas de cada paciente.
- *Otimização de prescrições*. Utilizam dados farmacocinéticos, farmacodinâmicos e dados do paciente para maximizar os efeitos terapêuticos e minimizar os efeitos adversos.
- *Personalização da terapêutica*. Ajustam tratamentos baseando-se no perfil genético e biomarcadores específicos dos pacientes, aumentando a sua eficácia.
- *Simulação virtual de organismos*. Modelam virtualmente o organismo do paciente para prever os efeitos potenciais de um fármaco, melhorando a segurança e eficácia do tratamento proposto.
- *Análise prognóstica*. Utilizam algoritmos avançados para prever resultados como o risco de reinternamento, remissão de doenças ou o surgimento de infeções, permitindo intervenções mais precisas e informadas.
- *Robótica cirúrgica*. Empregam técnicas de alta precisão para auxiliar intervenções cirúrgicas, reduzindo os riscos de complicações.
- *Desenvolvimento de novas terapêuticas*. Analisam dados bioquímicos e biomédicos para identificar novos alvos terapêuticos e candidatos a medicamentos.
- *Intervenções de Saúde Pública*. Avaliam grandes conjuntos de dados epidemiológicos para otimizar a alocação de recursos e responder de forma mais eficaz a crises de saúde pública.

3. *Cuidados Automatizados.* Sistemas potenciados por IA que proporcionam cuidados médicos com um certo grau de autonomia. Estes sistemas são capazes de integrar e potenciar dispositivos médicos, abrindo caminho para novas modalidades terapêuticas e melhorando as existentes. Exemplos das suas funcionalidades e aplicações são: (Alowais et al., 2023; Bajwa et al., 2021; Bonmassar et al., 2024; Davenport & Kalakota., 2019)

- *Comunicação e interatividade*. Recorrendo às suas capacidades linguísticas, interagem diretamente com os pacientes, por exemplo, oferecendo conselhos, esclarecendo dúvidas, emitindo lembretes medicamentosos, e verificando a adesão às terapêuticas prescritas.

- *Monitorização contínua da saúde*. Mantêm uma vigilância constante sobre o estado de saúde dos pacientes, permitindo intervenções médicas imediatas quando necessário.

- *Reabilitação motora*. Integrados com interfaces Corpo-Máquina, estes sistemas auxiliam na reabilitação de pacientes com incapacidade motoras, temporárias ou permanentes.

- *Dispositivos de Acessibilidade*. Controlam próteses e cadeiras de rodas através da interpretação de sinais nervosos dos pacientes, melhorando a sua autonomia e qualidade de vida. (Shaima et al., 2024)

- *Intervenções psicológicas*. Aplicam técnicas de Terapia Cognitivo-Comportamentais adaptadas às necessidades individuais dos pacientes.

- *Apoio a toxicodependentes*. Auxiliam na manutenção da abstinência de substâncias, oferendo suporte contínuo, personalizado e acessível.

- *Proteção da saúde mental em idosos*. Atuam em contextos de isolamento social, estimulando cognitivamente os idosos e oferecendo suporte emocional e psicológico.

4. *Aplicações não clínicas.* O papel transformador da IA não se limita às suas aplicações clínicas. A análise de vastos conjuntos de dados estruturados e não estruturados permite melhorias nas mais diversas áreas sociais, económicas, políticas e científicas, beneficiando sinergicamente a área da saúde: (Agrawal & Goldfarb, 2022; Alowais et al., 2023; Bajwa et al., 2021; Davenport & Kalakota., 2019)

- *Estudos populacionais e epidemiológicos*: A IA amplia o escopo e a precisão dos estudos epidemiológicos, permitindo análises que abrangem grandes populações durante períodos mais prolongados. Estes estudos são cruciais para monitorizar tendências de saúde e antecipar surtos de doenças, proporcionando uma base sólida para políticas de saúde pública eficazes.

- *Gestão de recursos de Saúde*: A integração da IA na gestão de recursos de saúde revoluciona várias áreas, incluindo: (a) Melhoria da gestão hospitalar: Otimiza o orçamento e a eficiência operacional dos hospitais. (b) Otimização logística: Refina as ferramentas de análise e a distribuição de recursos. (c) Aprimoramento dos registos clínicos: Aumenta o tempo de contacto clínico dos médicos, minimizando trabalho burocrático não especializado. (d) Combate a fraudes e injustiças: Fortalece a equidade e justiça no acesso a cuidados de saúde.

- *Análise de políticas de Saúde*: Com a IA, é possível realizar uma avaliação contínua das políticas de saúde implementadas, analisando a sua eficácia e impacto ao longo do tempo. Isso inclui a monitorização de campanhas e intervenções de saúde pública, ajustando-as conforme necessário para maximizar os benefícios à população.

- *Deteção e resposta a emergências*: A identificação rápida de emergências de saúde pública ou de proteção civil, como epidemias e catástrofes naturais, expedita a resposta, permitindo minimizar o seu impacto.

Capítulo III – Considerações Bioéticas sobre a aplicação de sistemas de Inteligência Artificial na Medicina

III.1 – Sobre o Princípio da Beneficência

A introdução de sistemas de Inteligência Artificial na Medicina deve ser cuidadosamente planeada, assegurando que serve primeiramente o *bem-estar dos pacientes*, como justificação primordial da prática médica e em conformidade com o Princípio da Beneficência.

Iniciar o Capítulo III com o Princípio da Beneficência é deliberado, pretendendo-se evidenciar que as críticas e conflitos morais que serão inevitavelmente abordados ao longo desta análise bioética não devem obscurecer os benefícios da introdução de sistemas de IA na Medicina, deixando desde início claro que se rejeita o viés cognitivo que tolda a compreensão de uma negatividade irrazoável.

A inação motivada pela abstenção de causar dano[42] não deve sobrepor-se incondicionalmente à beneficência ativa. Todos os princípios éticos e as regras de si derivadas devem ser igualmente considerados, especificados e ponderados previamente à tomada das decisões e ações que culminarão na introdução destes sistemas na prática clínica, não esquecendo as condições necessárias à escolha entre obrigações *prima facie* abordadas no Capítulo I.

Como ilustrado no final do Capítulo II, as aplicações da IA têm demonstrado um enorme potencial para promover a *beneficência ativa*, pelas suas aplicações na prática médica, e o *princípio da utilidade*, otimizando a alocação e aumentando a quantidade de recursos disponíveis para promover os objetivos de saúde gerais.

O potencial de promoção do Princípio da Beneficência pela IA pode ser ilustrado por exemplos práticos de implicações nas seguintes regras morais, de si derivadas:

[42] Ilustrada pela popular expressão *Primum non nocere*, usada, por vezes, para justificar inação.

1. *Proteger os direitos dos outros*: Um sistema de IA que verifica a prescrição e administração de medicamentos em hospitais poderá alertar de imediato a equipa médica se detetar algum erro, para que este possa ser corrigido precocemente, prevenindo consequências graves e protegendo o direito de todos os pacientes a cuidados de saúde seguros e eficazes. (Bates et al., 2022)

2. *Prevenir e Remover Condições que Possam Causar Danos*: A monitorização contínua de sinais vitais em unidades de cuidados intensivos permite a deteção da presença de sépsis num paciente[43] de ser clinicamente evidente, o que permite alertar a equipa médica para que inicie precocemente o tratamento necessário[44], aumentando a probabilidade de recuperação do paciente e prevenindo complicações graves. (Huat Goh et al., 2021)

3. *Assistir Pessoas em Perigo*: Sistemas de IA integrados em veículos de emergência podem comunicar continuamente com os hospitais, fornecendo informações sobre o estado de saúde do doente transportado em tempo real, para que as equipas médicas melhor se preparem para a chegada do paciente, reduzindo o tempo de resposta em situações críticas, nas quais todos os minutos são essenciais para melhorar o prognóstico do doente. (Damaševičius & Bacanin & Misra, 2023)

No entanto, será demonstrado que o potencial dos sistemas de IA enquanto promotores das regras morais é igualado pelo potencial de as infringir. As implicações morais destes sistemas nos princípios bioéticos são complexas, sendo fundamental que a sua implementação seja feita de forma responsável. A mudança de paradigma trazida por tecnologias disruptivas, como a IA aparenta ser, originam consequências multifatoriais difíceis de prever totalmente, o que reforça a importância de se manter uma atitude crítica e apropriadamente cética relativamente à sua implementação e considerar não só os possíveis benefícios, mas também os riscos associados.

[43] Avaliando, por exemplo, a temperatura corporal, frequência cardíaca e pressão arterial.

[44] Por exemplo, a administração de antibióticos e fluidoterapia

A título exemplificativo, imagine-se um sistema de IA usado para marcação de consultas num centro de saúde, que tem como objetivo maximizar o número de consultas realizadas. Se não forem definidas condições que limitam o espectro de decisão e atuação da IA, protegendo a boa prática clínica e os valores éticos, o sistema poderá privilegiar tipos de consultas mais simples e rápidas, negligenciando pacientes com patologias mais graves e maior número de comorbilidades que necessitam previsivelmente de consultas de maior duração.

Qualquer pessoa razoável entenderá que a maior complexidade clínica destes pacientes, embora resulte em consultas de maior duração, justifica uma necessidade mais premente de os consultar. No entanto, este tipo de juízo implica uma série de *conhecimento implícito, faculdades e características humanas*, como a contextualização sociocultural, empatia, emoções, valores éticos e senso comum, que a IA simplesmente não possui, lembrando que esta é um executor eficiente, mas acrítico (Hasselberger & Lott, 2023). Assim como um ser humano otimiza uma linha de produção sem considerar as implicações morais para o produto, uma IA otimizará a gestão de recursos de saúde sem ponderar as implicações morais para os pacientes.

Este exemplo relativamente simples demonstra a importância da correta definição de regras e parâmetros operacionais dos sistemas de IA por parte dos seus desenvolvedores e implementadores. Este problema é amplificado quando as normas e orientações pertinentes à área de ação do Sistema são altamente especializadas e só um perito poderá defini-las adequadamente e assegurar a sua correta inclusão enquanto parâmetro ou regra, embora seja improvável que seja mesmo perito a programar os modelos de IA que compõem os sistemas. O hiato de conhecimento[45] entre quem programa os modelos de IA e quem conhece as necessidades e particularidades clínicas reforça a importância do envolvimento proativo dos profissionais de saúde nas diferentes fases de desenvolvimento, implementação e uso destes sistemas, empregando o seu valioso conhecimento especializado para guiar e auditar o

[45] Considere-se o fenómeno de *Déformation professionnelle*, a tendência de uma pessoa de olhar para todas as situações do ponto de vista da sua profissão, tomando decisões inapropriadamente simplistas que esquecem a importância de manter uma perspetiva mais abrangente ao analisar situações que têm implicações complexas em valores humanos.

processo. Advogar para que os objetivos específicos da IA, como qualquer outra tecnologia na Saúde, se alinhem com os princípios bioéticos e melhores práticas clínicas é um *dever profissional*. Mesmo que não sejam os profissionais de saúde que programam estes sistemas, poderão permear o seu desenvolvimento com o seu conhecimento especializado e importantemente com as suas considerações éticas, da mesma forma que um arquiteto, não construindo casas, pode permeá-las com as suas considerações estéticas.

A defesa do correto desenvolvimento dos sistemas de IA do ponto de vista da sua segurança, rigor científico, aplicabilidade clínica e respeito pelos princípios éticos não será tarefa fácil e trará novos desafios às *entidades reguladoras*. Em comparação com medicamentos e dispositivos médicos, a criação e desenvolvimento de modelos de IA requerem menos infraestruturas e a sua pesquisa e desenvolvimento, sendo baseados sobretudo em *software*, têm menos barreiras à entrada do que a indústria farmacêutica ou dos dispositivos médicos. Especialmente quando se consideram modelos mais restritos do ponto de vista técnico que, por sinal, também serão mais falíveis. Além disso, os modelos de IA são dinâmicos, podendo ser atualizados e alterados a qualquer momento, de formas mais ou menos profundas, o que requer um grau inédito de monitorização e regulação contínua. (Reddy et al. 2023)

Uma característica destes sistemas que demonstra particular dificuldade regulatória é a sua *opacidade*. Embora não tenham necessariamente de o ser, como já visto, muitos sistemas de IA são autênticas *caixas negras*, pela sua complexidade e dimensão. Os biliões ou triliões de parâmetros e dados fazem com que as razões, regras, ligações e inferências que permitem aos sistemas de IA transformar um substrato de dados em decisões e recomendações finais, não sejam facilmente acessíveis ou compreensíveis pelos reguladores ou utilizadores finais.

A *transparência* destes sistemas requer escolhas conscientes dos seus desenvolvedores[46], sem as quais dificilmente se poderá auditar estes sistemas, impedindo a identificação de possíveis vieses, erros e falhas, comprometendo a confiança que os profissionais de saúde e pacientes

[46] Os processos e métodos necessários para que os humanos consigam compreender e confiar num modelo de IA são estudados por uma disciplina de IA chamada *Explainable AI* ou *XAI*, "IA explicável". (Albahri et al., 2023).

poderão depositar nestes. Portanto, é essencial que o desenvolvimento de IA, especialmente na medicina, seja acompanhado por esforços ativos para manter a *compreensibilidade* dos sistemas.[47]

Um problema crítico que surge desta falta de compreensibilidade é a identificação das chamadas *alucinações* da IA. Estas ocorrem quando esta produz, de forma inesperada e incompreensível, informações falsas ou enganosas que são apresentadas da mesma forma que seriam se estivessem corretas[48]. As consequências devastadoras destas na prática clínica são evidentes, originando, por exemplo, diagnósticos errados e motivando opções terapêuticas inadequadas.

Ainda que os seres humanos também cometam erros, e, aplicando o *princípio da caridade*, assumindo que possam até cometê-los em números superiores a certos sistemas de IA, estes erros, embora também complexos e multidimensionais, são nos mais familiares e intuitivos[49]. (Kandul et al., 2023) As falhas dos sistemas de IA ocorrem de forma "poligénica" (Preetham, 2023), podendo permear, de forma latente e indetetável, todas as suas decisões até culminarem num evento de erro catastrófico e virtualmente imprevisível[50]. Estes problemas são inerentes aos modelos de IA e são diretamente proporcionais à sua complexidade, como tal *alcançar um perfil de segurança perfeito é impossível*. (Varshney, 2016) Deve, assim, ser destacada a importância de não se confiar cegamente nas suas recomendações, por melhor e mais cómoda que esta tecnologia se revele.

[47] Incluindo, por exemplo, o acesso às bases de dados, algoritmos e parâmetros utilizados por parte de auditores e reguladores.

[48] Embora o termo alucinações esteja amplamente disseminado, o termo *confabulações* aplicar-se-ia melhor a este fenómeno.

[49] Sendo inclusivamente considerados durante o planeamento de sistemas críticos, como centrais nucleares, através de *análises de fiabilidade humana*. (Reason et al., 2000)

[50] Uma proposta promissora para detetar e monitorizar erros complexos e multifatoriais é o uso de sistemas de IA para auditar outros sistemas de IA, através da análise do seu código, bases de dados e vieses algorítmicos, validação dos dados utilizados e parâmetros definidos e monitorizando continuamente as recomendações feitas, particularmente a sua segurança e eficácia. (Schwettmann et al., 2023)

Outro problema futuro poderá advir da *multiplicidade de sistemas de IA*, desenvolvidos por entidades com objetivos e valores díspares. Esta diversidade pode ser enriquecedora, permitindo um maior leque de escolha de entre as opções disponíveis, mas a utilização de métodos e bases de dados distintos, otimizados para diferentes finalidades e com diferentes níveis de validade científica, cria dificuldades para a padronização e regulação, essencial às tecnologias de Saúde.

Uma forma de atestar a *confiança* nos sistemas de IA pode ser através da certificação por parte de ordens e sociedades médicas, à semelhança do que é feito com as *guidelines* e recomendações que estas lançam periodicamente. A validação e recomendação por grupos de especialistas e ordens profissionais são cruciais para permitir aos médicos, enquanto utilizadores finais, saber que sistemas são seguros, eficazes, confiáveis e comprovadamente promotores do bem-estar dos pacientes.

Uma abordagem que pode parecer tentadora é a censura e proibição de certos modelos de IA por parte dos governos. Deve-se frisar, no entanto, que para além deste tipo de abordagem se revelar historicamente ineficaz e promotor de desconfiança nas instituições públicas[51], o atual contexto tecnológico permite o acesso e propagação de informação de forma descentralizada[52], o que facilita a criação e funcionamento de modelos de IA mesmo contra a vontade das autoridades governamentais. O combate à desinformação deve passar, então, principalmente pela facilitação do acesso a informações baseadas na evidência científica, de forma clara e correta. Esta postura é mais respeitadora dos direitos das pessoas, transparente e será mais eficaz a alinhar a perceção pública com a realidade científica, protegendo a confiança nos profissionais de saúde e nas instituições públicas.

Os médicos terão um papel fundamental relativo à avaliação da credibilidade destes Sistemas, como tal deverão manter sempre um espírito crítico quanto à sua legitimidade da sua

[51] Problema sério que parece estar novamente a ganhar tração, como demonstrado pela ascensão de sentimentos antissistema e preponderância de teorias da conspiração, especialmente dirigidas à área da Saúde, como as que surgiram durante e após a pandemia COVID-19.

[52] Por exemplo, através da tecnologia *blockchain* ou métodos de distribuição *peer-to-peer*.

proveniência, validando na linha da frente se promovem a precisão e segurança dos cuidados de saúde e que estes respeitam os superiores interesses clínicos e bioéticos. O seu *feedback*, juntamente a avaliação destes Sistemas através de estudos científicos será essencial para a constante atualização e melhoria destes.

Além do desenvolvimento e implementação dos sistemas de IA também a sua *utilização* deve ser feita da forma correta. Tal como outras ferramentas médicas que utilizam, os médicos devem entender apropriadamente o funcionamento e limitações destes sistemas, utilizando-os para as funções para as quais foram concebidos sob o risco de surgirem falhas do seu uso inapropriado.[53]

A médio/longo prazo, existe o risco de os sistemas de IA serem suficientemente eficazes para levar à complacência dos profissionais de saúde, mas ainda imprevisivelmente falíveis, originando erros que não são identificados e corrigidos antes de causarem danos aos pacientes. A IA deve ser uma ferramenta que *complementa* a prática médica, usada dentro das suas capacidades *para fornecer recomendações* valiosas que suscitam melhorias nos cuidados de saúde, mas *sempre sob a supervisão* e validação dos profissionais de saúde.[54]

A limitações, falhas e ajustes necessários a esta tecnologia podem fomentar abordagens de *superproteção*. É fundamental assegurar que não se colocam entraves e impedimentos arbitrários ou até mesmo mal-intencionados ao desenvolvimento da IA. O *excesso de zelo* e a total aversão ao risco pode privar-nos de inúmeros de benefícios desta tecnologia, que pode em certos casos salvar vidas. Quaisquer constrangimentos (regulatórios, legais, etc.) à sua implementação deverão ser devidamente justificado e deverá ser compreendido por quem os

[53] Por exemplo, sistemas de processamento de linguagem natural revelam-se incapazes de realizar operações aritméticas simples, como médias, apesar de conseguirem gerar texto rapidamente. Da mesma forma, um sistema que relate com grande precisão imagens de tomografias computadorizadas pode ser completamente incapaz de interpretar ressonâncias magnéticas.

[54] Por exemplo, uma IA que sugere planos de tratamento deve ser usada para apoiar à decisão clínica, e não como substituto do julgamento médico.

impõe que a *superproteção* pode levar também à violação dos Princípios da Beneficência e da Não Maleficência.

Existem situações em que a implementação dos sistemas de IA deve ser feita de maneira particularmente expedita, devido ao seu potencial benéfico significativo, podendo inclusive englobar-se na *beneficência obrigatória*, depois de comprovada a sua utilidade promissora e assegurado um perfil de segurança mínimo indispensável.

O *uso compassivo* de sistemas de IA deverá poder ser requisitado em casos graves, quando a sua utilização apresenta uma elevada probabilidade de beneficiar pacientes que enfrentam um risco de vida significativo. Nestes casos, a utilização destes Sistemas poderá justificar-se, mesmo que ainda não estejam aprovados para um uso mais generalizado. Imagine-se, por exemplo, um caso de um paciente com um quadro médico grave, no qual existe uma forte suspeição clínica deste ter um tumor, mas que ainda não foi identificado pela sua equipa médica. As capacidades da IA de interpretar imagens e exames médicos poderão ser utilizadas para identificar o local e tipo do tumor, para que se inicie o seu tratamento direcionado mais precocemente. O uso compassivo desta tecnologia não implica que este mesmo sistema já esteja aprovado para aplicações mais abrangentes, como interpretação de imagens médicas em massa, por exemplo, em rastreios populacionais (mamografias ou colonoscopias). O dinamismo destes sistemas é nestes casos um benefício, permitindo que se alargue a população alvo da sua utilização, consoante sejam lançadas versões mais atualizadas e comprovadamente seguras.

O *direito a tentar* também se pode aplicar ao acesso a tecnologias que utilizam IA, especialmente em casos nos quais as opções terapêuticas são limitadas ou inexistentes, após a devida ponderação dos valores envolvidos, como a probabilidade de benefício, a autonomia do paciente e a responsabilidade médica de evitar danos. Pacientes que voluntariamente cedem os seus dados médicos para o desenvolvimento de bases de dados usadas nos sistemas de IA podem ter uma expectativa legítima de beneficiar das inovações resultantes, havendo uma obrigação ética, baseada na reciprocidade, de permitir-lhes o acesso a essas tecnologias. A reciprocidade enquanto estrada de dois sentidos e pode, por outro lado, justificar que se os

doentes beneficiam dos sistemas de IA, deverão contribuir para o seu desenvolvimento, por exemplo, cedendo dados para o seu treino, evidentemente estando assegurada a sua anonimização e o seu direito à privacidade.

A IA terá um impacto profundo nas expectativas e exigências que os pacientes têm dos cuidados de saúde. A democratização da informação, viabilizada pelas capacidades linguísticas e de interpretação de vastas quantidades de dados da IA, permitirá aos pacientes um acesso facilitado a informações médicas compreensíveis e provenientes de diversas fontes, incluindo bases de dados científicas e a internet. O acesso amplo a informação de qualidade promove a autonomia dos pacientes e fortalece a sua capacidade de tomar decisões informadas sobre a sua saúde, levando, legitimamente, a uma maior exigência e escrutínio das opções de tratamento propostas pelos médicos[55].

No entanto, apesar das vantagens associadas ao acesso democratizado à informação de Saúde, também surgirão riscos significativos. Informações complexas ou descontextualizadas podem ser mal interpretadas e a disseminação de desinformação pode ocorrer com uma rapidez e profundidade sem precedentes.[56] O potencial destes Sistemas pode ser direcionado tanto para o benefício como para o prejuízo da população geral, dependendo de escolhas deliberadas ou influenciadas por fatores não conscientes dos seus desenvolvedores.

O paternalismo médico enfrenta desafios extraordinários neste contexto. Tradicionalmente utilizado para proteger os pacientes de informações inadequadas ou mal interpretadas, ou, nas suas formas mais severas, sobrepondo-se à autonomia do paciente pela presunção de que o seu relativo desconhecimento das ciências médicas invalida a sua tomada de decisão, o paternalismo terá de lidar no futuro com informações de diferente legitimidade e qualidade obtidas através da IA. A relação médico-doente sofrerá novas reestruturações, exigindo que os médicos considerem mais a opinião dos pacientes, reforçando a postura

[55] Será explorado em maior detalhe no Subcapítulo III.3.

[56] A confiança excessiva dos pacientes nas informações obtidas desta forma será exacerbada por crenças erróneas e por falácias cognitivas, como a crença irrazoável nas decisões automatizadas, o *cherry-picking*, o viés de confirmação e o efeito *Dunning-Kruger*, piorando o fenómeno atual conhecido como "Dr. Google".

humilde, transparente e respeitosa dos direitos dos pacientes, especialmente da sua autonomia, aproximando o papel do doente e do médico na tomada das decisões de saúde e o que contribui para a erosão de mais um dos pilares do autoritarismo médico.

Recomendações de boa qualidade, provenientes de Sistemas confiáveis e que são aparentemente benéficas para os pacientes devem ser cuidadosamente consideradas. Se, ainda assim, os médicos acreditarem que estas recomendações não são a melhor opção para o paciente, deverão usar a experiência profissional, julgamento clínico e capacidades de comunicação interpessoal para justificar ao paciente de forma clara as razões para preferir uma abordagem alternativa à consultada no Sistema.

O facto de os sistemas de IA serem desenvolvidos por diferentes entidades, que têm objetivos e valores distintos, que influenciam as escolhas das bases de dados, processos de aprendizagem e raciocínio, de parâmetros de funcionamento que irão resultar em recomendações totalmente diferentes só complica mais esta situação.

Não há, por exemplo, qualquer limitação conceptual à criação de um sistema de IA baseado na *frenologia*, que utiliza dados craniométricos e outros princípios frenológicos para fazer associações entre formas cranianas e patologias médicas, podendo em última análise sugerir a prescrição de medicação com base nas dimensões do crânio de uma pessoa.

Perante esta diversidade, o paternalismo médico continuará a ser necessário, empregando a capacidade dos médicos de integrar e interpretar informação clínica de forma holística, a experiência profissional, o conhecimento teórico, além das capacidades humanas, já referidas, que a IA ainda está longe de replicar, para salvaguardar a prática médica correta e moral, que promovam acima de tudo o bem-estar dos pacientes. Ainda assim, devem ser claramente preferidas as formas mais suaves de paternalismo, baseadas na negociação e partilha de decisões, assegurando que os pacientes se sintam ouvidos e respeitados, enquanto se protegem os benefícios que advêm do conhecimento e da experiência médica.

Não obstante, será fundamental a existência de proteções legais e sociais que permitam o *paternalismo passivo*, de forma que um médico possa recusar agir sobre recomendações que

considera serem prejudiciais para o seu paciente. Caso contrário, a Medicina corre o risco de cair em completa deslegitimação, resultando na aceitação passiva de todas as recomendações da IA devido ao medo de represálias, reduzindo os médicos a meras "interfaces físicas" coletoras de dados a pedido de decisores virtuais que seguidamente lhes transmitem ordens de execução que estes têm de realizar acriticamente, por mais erradas que sejam.

Haverá, indubitavelmente, a necessidade de uma adaptação das funções e capacidades dos médicos, sendo essencial que desenvolvam novas competências para trabalhar eficazmente com estas ferramentas e que compreendam as suas capacidades e limitações. Consequentemente, a formação médica deve incorporar a IA como tópico abordado, preparando os futuros médicos para utilizarem estas tecnologias de forma crítica e informada.

O caráter humanista da medicina será mais importante do que nunca. A relação médico-paciente é fundamental para a qualidade dos cuidados de saúde, e elementos como empatia, intuição e julgamento clínico são insubstituíveis. As tecnologias de IA devem ser vistas como ferramentas que complementam, e não substituem, o toque humano na medicina. As *virtudes* associadas aos cuidados médicos, a comunicação, o entendimento emocional e social, a habilidade de lidar com subjetividades são essenciais para estabelecer uma relação médico-doente que permita alcançar o melhor resultado para o paciente e não podem ser replicados por algoritmos ou modelos computacionais. Mais do que nunca, é crucial fomentar nos profissionais de saúde boas capacidades humanas e não apenas bom conhecimento teórico.

Abordando agora a outra componente do Princípio da Beneficência, o *Princípio da Utilidade*, poderá ser constatado o potencial que IA tem para este, pelas suas aplicações no aumento da eficiência e melhoria da distribuição de recursos na área da saúde. As *análises de utilidade* potenciadas pela tecnologia de IA poderão integrar um maior número de parâmetros e dados, tornando-se mais precisas e detalhadas. Este aumento de precisão resulta em decisões mais bem fundamentadas e numa alocação mais eficiente e equitativa dos recursos disponíveis, beneficiando um maior número de pacientes e maximizando resultados de saúde.

Os problemas inerentes estas análises, como a ocultação de conflitos morais e a simplificação de situações necessariamente complexas, continuarão a existir, mesmo que sejam potenciadas por funcionalidades de IA e resultados permanecerão enviesados pelas escolhas feitas durante a sua idealização, uma análise de utilidade baseada na qualidade de vida continuará a ser profundamente subjetiva, independentemente do número de parâmetros ou dados que usa para a definir.

Para ilustrar essa subjetividade, imagine-se o que aconteceria se levássemos um destes Sistema de IA que faz análises de utilidade baseadas na qualidade de vida e a tecnologia de base necessária ao seu funcionamento para diferentes alturas, culturas e locais ao longo da história, e explicássemos aos habitantes da altura toda a informação pertinente para que pudessem definir todos os parâmetros e dados que achassem adequados para determinar a qualidade de vida de alguém. Mesmo *mantendo a complexidade técnica* de um sistema, usando o mesmo número dados e parâmetros, usando as mesmas técnicas de raciocínio e aprendizagem automática, com recurso ao mesmo poder computacional, as *análises produzidas por este serão díspares*, por se basearem necessariamente nos dados e parâmetros a que tem acesso, que refletem diferentes conceções de qualidade de vida.[57]

As análises continuarão a apresentar as falhas inerentes aos modelos e métricas definidas. No caso de análises baseadas em *QALYs*, estas continuarão a não distinguir entre o número de anos de vida do número de vidas[58], e a qualidade da quantidade de vida[59], independentemente da quantidade de dados usados para definir o fator de qualidade de vida. A maior complexidade e magnitude destas análises poderá perpetuar a aplicação de técnicas analíticas com falhas conceptuais intrínsecas, ao disfarçá-las como uma sofisticação aparentemente superior.

[57] Como já abordado, embora a palavra "inteligência" sugira racionalidade e objetividade, em muitos casos, a IA apenas utiliza mais poder computacional e um maior número de dados e parâmetros do que algoritmos convencionais.

[58] Uma pessoa viver vinte anos seria, por esta ótica, preferível a duas pessoas viverem nove anos cada.

[59] Uma pessoa que vive com plena qualidade de vida durante cinco anos seria privilegiada face a uma pessoa que vive com um fator de qualidade de vida calculado em cinquenta porcento durante nove anos.

A IA também desempenha um papel crucial na *regulação de dispositivos médicos e fármacos*, cujos processos de avaliação de segurança e eficácia poderão ser melhorados pela IA através de *simulações virtuais* e *análises detalhadas de dados*.

- Simulações virtuais dos efeitos de fármacos num organismo permitirão uma avaliação mais rápida e controlada antes de avançar para os testes em animais e humanos. A aceleração do ciclo de desenvolvimento de novos fármacos, mantendo a sua segurança e eficácia, permite que novas terapêuticas possam ser introduzidas de forma mais célere, trazendo enormes benefícios aos pacientes que delas necessitam.

- A IA também poderá ser aplicada para avaliar de forma continuada e abrangente fármacos, intervenções e dispositivos médicos após a sua introdução no mercado, estudando maiores amostras de pacientes e por uma maior duração de tempo.

- A personalização de fármacos com base na análise detalhada do genoma e até do proteoma específicos de cada indivíduo pode aumentar a eficácia e reduzir os efeitos adversos associados aos tratamentos, permitindo o desenvolvimento de terapias mais precisas e adaptadas às necessidades individuais dos pacientes.[60]

A *perceção de risco* relativamente à implementação dos sistemas de IA deverá ser abordada. A IA, enquanto tecnologia disruptiva com aplicações sistémicas, promete trazer mudanças na sociedade universais e profundas, de maneiras ainda imprevisíveis. Promessas de novos paradigmas, inerentemente imprevisíveis, geram grande desconfiança e desconforto social, por levantarem nos indivíduos e grupos a dúvida de se estas alterações os irão efetivamente beneficiar, ou sairão prejudicados.

Uma perceção comum é o risco de *substituição profissional* e devemos considerar as suas implicações nos profissionais de Saúde, visto que o receio de perder o seu meio de subsistência

[60] Por exemplo, um indivíduo de raça negra com hipertensão que responda melhor a inibidores da enzima de conversão da angiotensina (IECAs) do que a bloqueadores dos canais de cálcio (BCC) devido à resposta específica do seu sistema renina-angiotensina-aldosterona (SRAA), será tratado com IECAs em vez de se basear o seu tratamento inicial, como se faz atualmente, em características fenotípicas generalistas, como a cor da pele, que esta tecnologia poderá vir a revelar como inadequado e ultrapassado.

altera a disponibilidade dos médicos para aceitarem e colaborarem com a implementação da IA na Medicina. O impacto desta possível substituição será sentido de forma diferente consoante as especialidades médicas. Especialidades que recorrem sobretudo à memorização e análise de grandes volumes de informação e identificação de padrões clínicos serão mais rapidamente transformadas, por serem as áreas nas quais as capacidades da IA demonstra maior potencial. Outras especialidades que envolvem interações mais humanas e subjetivas, que lidam com pacientes complexos, não só do ponto de vista clínico, ou, por outro lado, que envolvem manipulação fina do mundo exterior poderão ver uma transformação mais gradual.

Naturalmente que a reciprocidade dita que pessoas que dedicaram a sua vida a servir uma área de grande importância social não sejam desvalorizadas e tratadas como dispensáveis. Embora não se possa moralmente impedir que algumas das funções especializadas sejam substituídas por métodos comprovadamente mais eficazes, eficientes e benéficos[61], a perícia destes profissionais deve ser redirecionada, por exemplo, para situações mais particulares nas quais os sistemas de IA não têm dados suficientes para serem eficazes.

No entanto, é improvável que a IA possa substituir por completo as funções de uma especialidade, embora provavelmente altere o dia a dia destas. Deve-se recordar que as necessidades em saúde são virtualmente ilimitadas, as vacinas e antibióticos modernos não eliminaram a importância do trabalho médico com doenças infecciosas, mas permitiram que este se concentre em diagnósticos mais raros e específicos, em vez se esgotar no tratamento de pneumonias comuns. Há também que notar que nem sempre as novas tecnologias substituem as técnicas e práticas clínicas implementadas, com raízes profundas baseadas na experiência e conhecimento médico de longa data. Da mesma forma que a existência de Meios Complementares de Diagnóstico e Terapêutica sofisticados não substituíram o exame objetivo ou a informatização dos registos de saúde não substituiu a anamnese, Sistemas de IA que auxiliam a decisão são complementares, não substitutos à decisão.

[61] A existência de polícias sinaleiros não pode justificar moralmente a não implementação de semáforos sabendo-se que estes são mais eficazes, eficientes e seguros. Considerações corporativistas e tradicionalistas também não devem ser um entrave à função principal da Medicina, que é a promoção do bem-estar do paciente.

O acesso generalizado a informação sobre esta tecnologia, aliado à promoção de debate social alargado e à transparência durante o desenvolvimento, implementação e uso desta tecnologia, será fundamental para que se possa confiar nos benefícios trazidos pela implementação destes Sistemas e alinhar a perceção pública com a realidade técnico-científica. Uma das mais importantes razões para promover o envolvimento geral da sociedade em todas as fases do ciclo de vida da IA é assegurar que as *crenças e valores comuns* as possam permear integralmente, construindo-a diretamente sobre os fortes alicerces éticos da *Moralidade Comum.*

III.2 – Sobre o Princípio da Não Maleficência

O Princípio da Não Maleficência preconiza impedir ativamente a causação de danos aos outros e constitui uma parte essencial dos deveres dos profissionais de saúde para com os seus pacientes. Assim, estes profissionais deverão ter um papel crucial no desenvolvimento e implementação dos sistemas de IA, assegurando que esta tecnologia não causa danos nem coloca em risco os pacientes e os seus interesses. A sua experiência e conhecimento sobre a realidade clínica é preciosa para detetar e impedir que qualquer forma de negligência surja destes sistemas, para cumpram rigorosamente este princípio e as regras de si derivadas.

A introdução de qualquer tipo de tecnologia na Medicina deve ser feita com extrema cautela garantindo estas que não são prejudiciais a curto, médio e longo prazo. Exemplos históricos, como o caso da *talidomida*[62], revelam como terapêuticas com resultados benéficos iniciais provocaram danos irreparáveis a longo prazo. Embora os sistemas de IA não provoquem dano direto do ponto de vista farmacológico, têm o potencial de influenciar e difundir inúmeras más decisões, que podem variar de pequenas falhas a catástrofes. Um exemplo notório de falha tecnológica na medicina é o caso do *Therac-25,* um sistema utilizado para tratamentos de radioterapia que, devido a falhas de *software*, administrou níveis excessivos de radiação em vários pacientes, resultando em mortes e ferimentos graves. (Apgar & Prentice, n.d.) Esta tragédia ilustra como a confiança excessiva em sistemas automatizados, sem a devida supervisão e controlo rigoroso, pode levar a consequências desastrosas.

A complexidade inerente aos sistemas de IA e a infinidade de fatores que influenciam o seu funcionamento criam um ambiente propício a erro. Estes erros advêm de fatores conscientes e inconscientes que moldam as decisões dos desenvolvedores, implementadores, reguladores e utilizadores, bem como de falhas tecnológicas imprevisíveis e interações inesperadas entre os inúmeros parâmetros e componentes dos sistemas. Tal como o organismo humano pode ver a sua função fisiológica perturbada por uma infinidade de causas, os sistemas de IA com a sua

[62] Introduzida no final da década de 1950 como um sedativo e antiemético, foi rapidamente adotada para tratar enjoos matinais em grávidas. Inicialmente considerada segura, provocou uma tragédia global quando se descobriu que causava malformações congénitas graves nos fetos.

crescente complexidade podem tornam-se disfuncionais e prejudiciais por incontáveis motivos.

Não deverá ser esquecida, no entanto, a o potencial de sistemas de IA eficazes evitarem danos e reduzirem riscos para os doentes, identificando erros de diagnóstico, de escolhas terapêuticas, nos registos clínicos, aumentando a precisão de intervenções cirúrgicas, monitorizando continuamente pacientes em risco.

A análise dos passos e decisões subjacentes à criação dos sistemas de IA poderá ilustrar diversas formas de erros e falhas sejam introduzidas no sistema. A criação de sistemas de IA perfeitos é impossível, mas o escrutínio das diferentes fases do ciclo de vida da IA permitirá identificar e colmatar falhas no planeamento e execução que, na Saúde, poderão culminar em danos a pacientes.

Consideremos as principais fases do ciclo de vida da IA e como erros, vieses, conflitos de interesses, negligência e incompetência podem originar resultados prejudiciais para o paciente:

1. Conceptualização e Planeamento. Nesta fase são definidos *a priori* os objetivos e os moldes que estabelecem como é que as etapas subsequentes do desenvolvimento da IA se irão realizar. Más decisões e erros cometidos nesta fase têm o potencial de se amplificarem através de um "efeito dominó", originando a jusante graves falhas na segurança e na eficácia do sistema. Existem várias decisões nesta fase particularmente relevantes:

- *Definição e ponderação dos objetivos.* Na saúde existem inúmeros objetivos concorrenciais - clínicos, económicos, sociais, individuais, etc. - que dificilmente poderão ser traduzidos pela sua listagem e posterior atribuição de um peso relativo numérico. Realizá-lo de forma consensual em Sistemas a serem aplicados para propósitos generalistas e em larga escala é tarefa impossível, as deliberações em saúde são complexas, multifatoriais e, em parte, subjetivas.

- *Métricas de sucesso*. Como se não bastasse a dificuldade de definir estes objetivos, subjetivos e complexos, ainda deverão ser escolhidos indicadores que permitam medir

o seu cumprimento, as chamadas métricas de sucesso que se correlacionam com o objetivo, mas não são elas próprias o objetivo, usadas no seu lugar quando este não é numericamente mensurável. Estas podem correlacionar-se melhor ou pior com o cumprimento do objetivo e podem ter dificuldades díspares na sua determinação o que torna a sua escolha também controversa, por exemplo, uma métrica menos correlacionada pode ser preferida por ser mais barata de determinar.

- *Escolhas sobre os dados.* Durante esta fase irá será necessário tomar inúmeras decisões sobre os dados necessários ao treino dos modelos de IA, que devido à sua particular importância serão, posteriormente, analisadas individualmente.

O envolvimento de uma equipa multidisciplinar é essencial para que os peritos das áreas onde ser irá implementar sistemas de IA possam comunicar aos desenvolvedores as necessidades e nuances que os sistemas deverão ter em consideração. A falha em auscultar os peritos desde estas fases iniciais poderá motivar escolhas de base inapropriadas e irreversíveis, que põe em causa a eficácia e segurança dos sistemas que não poderão, por isso, ser utilizados em boa consciência nas áreas médicas.

Recuperando o exemplo do sistema de IA para a deteção precoce de sépsis e alarme à equipa médica:

- *Objetivos* possíveis incluem (a) a deteção e tratamento precoce de sépsis, (b) a minimização de casos falsos positivos, e (c) os custos associados à implementação do sistema.

- *Métricas* verosímeis para estes objetivos são (a) a sensibilidade e especificidade do sistema, (b) o tempo médio de deteção após os primeiros sinais clínicos, e (c) a taxa de redução de mortalidade associada à sépsis.

- Os *dados* serão usados para treinar o sistema de IA poderão ser (a) os sinais vitais dos pacientes, (b) resultados de exames laboratoriais, e (c) informações clínicas dos registos eletrónicos de saúde.

- A *equipa multidisciplinar* poderá incluir (a) médicos intensivistas, (b) enfermeiros, (c) especialistas em IA, e (d) especialistas em ética médica. Os grupos contribuem com *diferentes perspetivas e conhecimento*: (a) os médicos e enfermeiros compreendem os desafios clínicos e as necessidades dos pacientes, (b) os especialistas em IA desenvolvem os algoritmos e modelos, e (c) os especialistas em ética garantem que as decisões tomadas estão alinhadas com os valores e princípios éticos.

Um exemplo de falha parte do desenvolvido deste sistema sem o envolvimento de profissionais de saúde, resultando na seleção inadequada de que dados usar, resultando numa alta taxa de falsos positivos e subsequentes alarmes desnecessários constantes. A perda de confiança na eficácia do sistema motiva a equipa médica a ignorar os alarmes, eventualmente ignorando alarmes verdadeiros.

2. Dados. Os dados apresentam um ciclo de vida próprio incluído a sua conceptualização, colheita, manutenção, uso e eliminação. A colocação de questões durante este ciclo — "Quem? Como? Porquê? Onde? Quando?" —poderão ilustrar as inúmeras formas como estes dados podem ser contaminados por erros, vieses e conflitos de interesse, criando dados de má qualidade que, em última análise, resultam em decisões erradas.[63]

A qualidade dos dados é um mundo em si, estudado pela Ciência de Dados que se dedica a entender os fatores que os influenciam e a aperfeiçoá-los, mas assumindo que a sua magnitude e complexidade impossibilita alcançar dados perfeitos.

O grande problema que se coloca é que os modelos de IA necessitam de quantidades astronómicas de dados, cuja qualidade será determinante para a eficácia final do sistema e até falhas de qualidade relativamente pequenas podem amplificar-se em larga escala, culminando em decisões catastróficas que na sua saúde impõem riscos e danos inaceitáveis aos pacientes.

O debate ético sobre os dados de treino da IA centra-se maioritariamente na possibilidade dos vieses destes perpetuarem desigualdades nos cuidados de saúde. Esta questão é crucial e será

[63] Este fenómeno é explanado pela expressão, usada em ciência da computação, "*garbage-in, garbage-out*", literalmente, "*lixo entra, lixo sai*".

analisada no Subcapítulo III.4 – Sobre o Princípio da Justiça, mas não deverão ser abordadas outras formas como os dados de má qualidade afetam o Princípio da Não Maleficência.

- As *métricas de sucesso*, definidas durante a fase de conceptualização, são elas próprios dados. O objetivo que motiva a sua escolha tem um grande impacto no funcionamento dos modelos com eles treinados,

  Se o objetivo definido for a redução de custos será priorizada a colheita e monitorização de métricas como o *tempo de hospitalização*, em vez da necessidade individual de um paciente receber cuidados continuados, que é um objetivo clínico. Os modelos treinados com esta métrica podem preferir altas hospitalares prematuras, ainda que provoquem readmissões frequentes e agravamento das condições de saúde dos pacientes.

- A *recolha* de dados é realizada através de registos eletrónicos de saúde, sensores e inquéritos.

  Um problema possível nesta fase é estes dados fazerem uma má representação da população geral, por exemplo, unidades de saúde privadas ou em áreas centrais poderão ter mais recursos para recolher dados, resultando numa sobrerrepresentação das populações por si servidas, normalmente com um estatuto socioeconómico mais favorável. A menor representação de grupos nos dados de treino pode implicar uma eficácia inferior dos sistemas de IA perante representantes destes.

  Outro problema é a utilização de registos com erros ou imprecisões subjacentes para a recolha de dados, por exemplo, se num determinado local for usual registar os sintomas dos pacientes de forma exagerada, seja por razões relacionadas com seguros de saúde ou com a segurança social, os modelos de IA que forem treinados por dados baseados nestes registos, quando implementados noutros locais, poderão subestimar sistematicamente as queixas dos doentes, condicionando de forma importante os diagnósticos e planos de tratamento baseados nas suas recomendações.

- Os dados recolhidos terão de ser *armazenados*, usando sistemas de gestão de bases de dados, armazenamento em nuvem e centros de dados.

  A opção de utilizar soluções de armazenamento mais baratas, sem redundância e sem as medidas de segurança adequadas, torna os dados mais vulneráveis a ciberataques que divulgam os dados confidenciais dos pacientes, ou a falhas técnicas que resultam em dados corrompidos que não só afetam o treino dos modelos, como impossibilitam auditorias posteriores, o que compromete seriamente a confiança que se pode ter nestes.

- A *eliminação* de dados obsoletos terá de ocorrer para dar lugar a novos dados, considerados mais adequados. utilizando protocolos seguros e conformes os regulamentos.

  A ideia prevalente de que a quantidade de dados deve ser sempre maximizada, independentemente da sua qualidade, poderá perpetuar a manutenção e utilização de dados de má qualidade, que poderão ser reintroduzidos no treino de novos modelos de IA, reduzindo, novamente, a sua eficácia e segurança.

A título ilustrativo das potenciais falhas catastróficas causadas por más escolhas relativamente aos dados considere-se um sistema usado no auxílio ao diagnóstico precoce de doença cardíaca utilizado num homem de 80 anos com história de sensação de falta de ar e fadiga. Este sistema foi treinado com dados clínicos históricos que apresentavam vários erros, incluindo dados incompletos, desatualizados, inconsistentes e com um enviesamento para pacientes jovens. Estas falhas culminaram na interpretação por parte do Sistema de IA de que os sintomas deste homem seriam doença não cardíaca, provavelmente ansiedade ou stress, uma vez que os dados de treino refletiam, predominantemente, diagnósticos de pacientes mais jovens com sintomas semelhantes. Esta interpretação incorreta atrasou a gestão da doença cardíaca do paciente que se agravou, resultando numa pioria significativa do prognóstico deste.

3. Desenvolvimento e treino dos modelos de IA. Será nesta fase que os algoritmos são efetivamente desenvolvidos, ajustados e aperfeiçoados com base nos dados disponíveis, processos de aprendizagem automática e intervenção humana.

Além do problema da qualidade dos dados já abordado, surgem nesta fase problemas como o *overfitting*, que ocorre quando o modelo se ajusta demasiadamente aos dados de treino, integrando até o seu "ruído" e peculiaridades específicas desses dados, e tornando-se incapaz de generalizar os processos de aprendizagem estabelecidos no treino para novos dados.

A própria escolha de que algoritmos utilizar é crucial, algoritmos mais complexos, como *redes neurais profundas*, têm maior precisão, mas à custa de serem mais opacos e menos compreensíveis, dificultando a identificação de erros ou vieses no modelo. Recomendações que não são verificáveis e que não oferecem explicação dificilmente poderão ser aceites na saúde, por exemplo, um Sistema de IA utilizado para diagnosticar cancro deve fornecer explicações claras sobre como chegou a uma determinada conclusão, permitindo que os médicos compreendam e validem o processo de decisão, antes de tomarem decisões clínicas e comunicarem a má notícia ao paciente.

4. Implementação e Integração. É nesta fase que os sistemas são integrados na prática clínica, necessitando a sua adaptação às infraestruturas existentes e a formação dos profissionais de saúde para a sua utilização.

Os sistemas de IA deverão ser integrados com os sistemas informáticos das infraestruturas de TI que hospitais e clínicas já possuem, falhas de comunicação entre programas podem gerar erros no seu funcionamento. Se um sistema de IA de auxílio ao tratamento não conseguir extrair adequadamente os dados dos registos hospitalares já existentes, poderá não considerar todas as informações pertinentes à tomada de decisão, como as alergias medicamentosas ou medicações habituais de um paciente, resultando em sugestões de fármacos potencialmente prejudiciais para os pacientes.

A comunicação entre programas diversos também levanta preocupações significativas acerca da segurança e privacidade dos dados confidenciais dos pacientes, por criar possíveis

vulnerabilidades que podem ser utilizadas maliciosamente para obter informações sobre os pacientes.

A implementação de vários sistemas de IA simultaneamente poderá suscitar recomendações contraditórias, que causam confusão e entropia nos fluxos de trabalho, resultando em atrasos ou escolhas inadequadas de tratamento. Os fluxos de trabalho devem ser cuidadosamente considerados, minimizado interrupções e obstáculos ao dia a dia clínico, que já é *per se* complexo, para que não se crie resistência entre os profissionais de saúde e não prejudicar a eficiência e eficácia da prestação de cuidados.

Ações de formação são essenciais para que os profissionais de saúde saibam como integrar na sua prática clínica os novos sistemas de IA, como interpretar as recomendações destes e como agir, de acordo com estas, de forma segura e ponderada.

5. Validação e Teste. A fase de validação e teste dos sistemas de IA é crucial para assegurar que os modelos desenvolvidos funcionam corretamente antes de serem implementados em ambientes clínicos reais.

Esta fase envolve a verificação da precisão, robustez e segurança dos sistemas de IA, assim como a identificação e correção de possíveis falhas. Modelos que não são devidamente validados podem apresentar alto desempenho usando os dados de treino, mas falhar em dados reais.

A *robustez* refere-se à capacidade do modelo de manter um desempenho consistente em diferentes cenários e condições, por exemplo, um modelo de auxílio ao diagnóstico de doenças cardíacas deve ser testado com uma variedade de dados provenientes de diferentes populações e condições clínicas para assegurar que ele o seu correto funcionamento em todos os contextos.

A *validação cruzada* é uma técnica essencial que envolve a divisão dos dados em conjuntos de treino e testá-los múltiplas vezes, para assegurar que o modelo não está a memorizar os dados de treino e que é eficaz, consistente e confiável na presença de novos dados de entrada, por exemplo, um sistema com intuito de detetar precocemente neoplasias, a validação cruzada

pode assegurar que o modelo consegue identificar corretamente os casos de cancro em diferentes subconjuntos dos seus dados, minimizando o risco de falsos negativos e falsos positivos.

A existência de vieses nos modelos também deverá ser aferida, para que não se perpetuem decisões injustas ou discriminatórias, por exemplo, oferecendo diagnósticos menos precisos para grupos menos representadas nos dados de treino ou não devidamente considerados durante o desenvolvimento.

Os modelos de IA devem ser postos à prova em cenários desfavoráveis para avaliar o seu comportamento em situações extremas ou inesperadas, identificando fraquezas e pontos de falha que podem não ser aparentes em condições normais, por exemplo, testado cenários realísticos, nos quais os dados poderão estar incompletos ou incorretos, para garantir que mesmo nessas situações não exista compromisso da segurança do paciente.

O impacto do sistema pode ser avaliado em ensaios clínicos simulados, antes da aplicação em ambientes clínicos reais, permitindo observar como os sistemas de IA interagem com processos clínicos existentes e identificar potenciais áreas de melhoria, sem imputar riscos aos pacientes.

6. Uso e Monitorização. A fase de uso e monitorização é crucial para garantir que os sistemas de IA funcionam conforme o esperado em ambientes clínicos reais e que continuem a ser seguros e eficazes ao longo do tempo.

A avaliação contínua e feedback dos utilizadores será essencial para que os sistemas sejam atualizados e melhorados, através da identificação de problemas não detetados durante os testes iniciais, que certamente surgirão, especialmente em sistemas mais complexos. Isto inclui a verificação da precisão dos diagnósticos, a eficácia dos tratamentos recomendados e a identificação de quaisquer erros ou vieses que se possam evidenciar com o uso.

A monitorização contínua permite a deteção rápida de erros e falhas no sistema, que poderão não ter sido identificados nos testes, serem resultantes de atualizações malsucedidas ou de alterações nos dados de entrada, permitindo a correção antes que causem danos aos pacientes.

A atualização regular é essencial, para incorporar novos dados, melhorar os algoritmos, corrigir vulnerabilidades e assegurar a responda adequada a mudanças nos padrões de doenças, por exemplo, respondendo a novos surtos de doenças, como foi com a COVID, e garantindo que que os sistemas aderem a novos protocolos e recomendações clínicos, de acordo com as melhores práticas e evidência científica.

O feedback dos profissionais de saúde que utilizam os sistemas de IA é essencial, por estarem numa posição única que permite observar como o sistema funciona no dia-a-dia, o que permite fornecer insights valiosos sobre a sua eficácia e usabilidade, por exemplo, relatando se as recomendações do sistema são claras e fáceis de seguir, ou se há áreas onde o sistema poderia ser melhorado.

Incidente ocorridos durante o uso dos sistemas de IA, como erros de diagnóstico ou falhas de tratamento, devem ser documentados e analisados para entender as causas subjacentes e evitar recorrências, melhorando continuamente a segurança e a eficácia dos sistemas de IA.

7. Descontinuação e Eliminação. A fase de descontinuação e eliminação é a etapa final do ciclo de vida dos sistemas de IA, onde os sistemas são retirados de uso e os dados são eliminados de maneira segura e conforme as regulamentações, minimizando o impacto negativo nos cuidados de saúde.

É essencial desenvolver planos de descontinuação que incluam a transferência segura das funções críticas que estes asseguram para outros sistemas ou processos, garantindo que não haja interrupções, especialmente em serviços essenciais.

A comunicação com os utilizadores dos sistemas de IA é fundamental durante esta fase, devendo os profissionais de saúde ser informados com antecedência sobre a descontinuação do Sistema, as razões para tal e os passos que serão tomados para preparar a transição e garantir a continuidade dos cuidados.

Os dados deverão ser transferidos de forma segura e eficiente, conforme as regulamentações de privacidade e segurança, assegurando que os dados permaneçam acessíveis para auditorias e análises futuras, mas protegidos contra acessos não autorizados e eliminando dados

desatualizados ou desnecessários, utilizando protocolos seguros que garantam que não possam ser recuperados ou acedidos por terceiros não autorizados.

Estes passos deverão estar preconizados *a priori* fazendo-se uma avaliação de impacto para identificar e mitigar possíveis riscos que poderão afetar os pacientes, profissionais de saúde e processos clínicos, sendo também essencial procurar feedback dos utilizadores após a descontinuação, permitindo extrair importantes lições para projetos futuros.

A documentação de todo o processo é essencial para assegurar transparência e responsabilidade, relatando as etapas seguidas, os dados eliminados e as medidas de segurança adotadas para auditorias futuras.

A confiança crescente em sistemas de IA pode gerar um cenário futuro no qual os médicos, especialmente os que se formaram aquando da sua plena implementação, cedem ao facilitismo promovido pela automatização, o que prejudica o seu raciocínio clínico e a sua capacidade de agir adequadamente de forma independente, perdendo-se também o conhecimento e experiência necessária para que se verifique ou conteste as recomendações da IA. Tecnologias que se revelam incontornáveis não devem ser ignoradas nem a sua utilização desencorajada, mas a educação médica deve continuar a primar pelo desenvolvimento do raciocínio clínico e competências médicas fundamentais.

A *dependência excessiva na IA* também se pode manifestar pela obsolescência de meios complementares de diagnóstico até agora desenvolvidos para serem interpretados por humanos, sendo substituídos por meios apenas interpretáveis por IA, por explorarem dados que ultrapassam as limitações cognitivas e sensoriais humanas. A priorização deste tipo de tecnologia que desconsidera o operador humano pode a médio-longo prazo comprometer a continuidade e qualidade dos cuidados de saúde, se ocorrem falhas técnicas que neguem o acesso aos sistemas de IA.

Neste sentido, deve ser sublinhado que os sistemas de IA na saúde são significativamente mais vulneráveis do que a maioria das tecnologias atualmente utilizadas na saúde, devido à sua maior e mais complexa interdependência com outras infraestruturas, organizações e

tecnologias de suporte: (a) Falhas em servidores, redes de comunicação ou serviços de computação em nuvem podem paralisar sistemas críticos ao seu funcionamento. (b) Empresas tecnológicas estrangeiras dominam a produção de tecnologias essenciais, como semicondutores, o que juntamente com a dependência de componentes como sensores e atuadores que envolvem cadeias de abastecimento globais, materiais existentes em poucos países, tornam estes sistemas de IA particularmente vulneráveis a políticas económicas e a tensões geopolíticas. (c) A segurança e integridade dos sistemas pode ser comprometida por ataques cibernéticos e manipulação de dados. (d) Divergências nas políticas e regulamentações de dados, como o RGPD na Europa, complicam a partilha internacional essencial à aquisição de grandes volumes de dados diversificados e a cooperação internacional é essencial para o progresso e partilha de conhecimento, essencial para o desenvolvimento de novos algoritmos e técnicas de IA.

As tecnologias atuais[64] são menos vulneráveis devido à sua menor dependência na conectividade contínua e infraestruturas globais, tendo normalmente as infraestruturas de suporte, manutenção e suporte técnico local. A menor dependência de redes externas e dados internacionais, torna mais fácil garantir a sua segurança contra ataques cibernéticos e estabilidade face a influências geopolíticas.

As decisões de *deixar morrer*, uma componente importantíssima da não maleficência, são impactadas profundamente pela IA.

O primeiro problema advém da promessa de que a IA trará uma mudança de paradigma tecnológico e científico, o que obriga a que vários juízos de futilidade médica atuais sejam repensados. Embora haja ainda bastante especulação por detrás das capacidades destes sistemas, o leque completo das suas aplicações e efeitos ainda não é totalmente compreendido. A expectativa de que se trate de uma tecnologia revolucionária que se encontra

[64] Por exemplo, equipamentos de imagiologia (ressonância magnética e tomografia computorizada) e sistemas de gestão hospitalar.

na sua fase de crescimento e desenvolvimento exponencial torna impossível prever a magnitude das suas aplicações futuras, na ciência e na saúde.

A *futilidade médica* torna-se difícil de ajuizar a curto e sobretudo a médio prazo, pela possibilidade plausível de surgirão terapêuticas inovadoras e eficazes para condições de saúde até hoje irremediáveis. Interfaces cérebro-máquina potencializadas por IA demonstram-se capazes de descodificar sinais cerebrais de doentes, permitindo o desenvolvimento de dispositivos capazes de mitigar situações clínicas muito desfavoráveis, por exemplo, a doentes com síndrome de *locked-in* comunicar com o mundo exterior, ou a tetraplégicos controlarem próteses e dispositivos externos apenas através do pensamento. As melhorias profundas que na qualidade de vida destes pacientes deverão fazer, no mínimo, reconsiderar juízos de futilidade que eram válidos até há relativamente pouco tempo.

O segundo problema surge da tentação de utilizar a IA para a toma destas decisões, apesar das inúmeras formas em que estas podem ser incorretas e da impossibilidade de capturar satisfatoriamente grandezas humanas como a qualidade de vida na forma de dados. A utilização destes Sistemas em decisões relacionadas com o não início ou interrupção de tratamento deve fomentar particular preocupação e ceticismo.

Decisões subjetivas que lidam com dimensões profundamente humanas - como são as decisões de fim de vida - baseadas na ponderação de realidades inerentemente humanas - como a dor, sofrimento e desconforto - têm de ser tomadas usando todo o escopo das faculdades humanas - a racionalidade, as emoções, a intuição e a empatia - e por considerações éticas, espirituais e socioculturais. Este tipo de decisões não requer maior rapidez ou eficiência e não pode ser moralmente traduzido por algoritmos que maximizam métricas de utilidade, tratando a vida e a morte humana como um meio. Algoritmos impessoais não devem ser utilizados como um refúgio burocrático, ilusoriamente neutro e imparcial, para a tomada confortável de decisões que devem ser sempre desconfortáveis, pelas dimensões éticas que envolvem. *Aos humanos, o que é dos humanos.*

A abordagem que valoriza excessivamente os dados e análises quantitativas, é por vezes chamada *dataísmo*, que procura de forma subjacente através da categorização e medição reduzir a complexidade humana a valores objetivos e expressões algébricas. No entanto, a medicina não é uma mera área de atividade económica que se pode desapagar da dimensão humana na procura de maximização de eficiência e outras métricas de utilidade. Diagnósticos e tratamentos que ignoram as nuances e o contexto individual dos pacientes mecanizam a medicina, reduzindo o organismo humano doente a uma máquina defeituosa, com falhas objetiváveis e com uma estratégia clara de resolução. Esta lógica é totalmente contrária à experiência clínica, na qual se constata que duas pessoas com a mesma doença podem ter perspetivas totalmente diferentes de que necessidades procuram satisfazer ao procurar cuidados de saúde. A "algoritmização" não beneficia os pacientes porque pressupõe o contrassenso de querer tratar humanos desumanamente. A empatia, comunicação e disponibilidade para lidar com a incerteza e com o subjetivismo humano serão sempre cruciais para compreender a fundo as preocupações dos pacientes e poder ajudá-los de forma profundamente pessoal e humana.

As interações entre IA e as decisões dos pacientes ou dos decisores de substituição também devem consideradas. Com o disseminar de tecnologias de IA com capacidades linguísticas será cada vez mais frequente as pessoas fazerem-lhes perguntas e basearem-se nas suas respostas para tomar decisões.

Estes sistemas poderão capacitar os pacientes e os decisores, dando-lhes informações importantes de forma acessível e poderão constituir apoios de grande valor à tomada de decisão. O seu uso poderá ser direcionado para a proteção dos doentes, do ponto de vista ético e legal, advogando por estes e informando os seus decisores sobre os seus direitos.

No entanto, o contrário também se poderá verificar, respostas erradas, manipulativas, mal-intencionadas ou provenientes de confabulações podem prejudicar a capacidade de decidir racionalmente com base em informação adequada e o possível enviesamento e presença de conflitos de interesses põe em causa a parcialidade favorável aos interesses do paciente, essencial à tomada moral de decisões em substituição de doentes incapazes.

Os médicos deverão averiguar que sistemas de IA foram utilizados no processo de decisão, tentando perceber quais as limitações, vieses e conflitos de interesses destes, averiguando a sua legitimidade e a dos seus desenvolvedores. Enquanto *gatekeeper*, deverá questionar-se se consegue afirmar em boa consciência que os sistemas de IA usados neste sentido são de confiança e confrontar crenças erradas que estas possam ter promovido nos decisores, minando processos de decisão importantes, com implicações impactantes para os doentes e os seus direitos.

O problema da *superproteção* colocar-se-á certamente ao desenvolvimento de tecnologias de IA, sobre a forma de regulamentações excessivamente rígidas que impedem a inovação, investigação e progresso científico necessário ao desenvolvimento de novos tratamentos e sistemas médicos, que poderiam beneficiar significativamente os pacientes. Os doentes com doenças raras ou terminais são frequentemente os mais prejudicados, pois a demora na aprovação de novos tratamentos pode significar a diferença entre a vida e a morte. Um exemplo concreto são as terapias genéticas para o tratamento de doenças genéticas, nas quais a IA tem demonstra enorme potencial facilitando a análise, identificação e edição de alvos genéticos.

Relativamente aos danos causados a grupos, é previsível que a implementação de uma tecnologia tão centrada em dados agrave os problemas da utilização indevida dos mesmos. A utilização não autorizada de dados médicos para finalidades diferentes das inicialmente divulgadas, sem obter novo consentimento dos pacientes, além de comprometer a autonomia dos indivíduos, tem um potencial gravíssimo de impor danos a grupos através de generalizações erróneas dos seus dados.

Outro problema é o enviesamento dos sistemas de IA, que pode perpetuar ou mesmo exacerbar desigualdades existentes na Saúde. O treino de modelos usando dados não representativos pode determinar diagnósticos e tratamentos sistematicamente incorretos para certos grupos demográficos. Por exemplo, um algoritmo de IA desenvolvido para identificar doenças de pele pode ser treinado predominantemente com imagens de tons de pele semelhantes, resultando numa precisão significativamente menor para pessoas com outros

tons, levando a diagnósticos incorretos ou tardios e que prejudicando a saúde de grupos inteiros de pacientes. A mitigação deste risco exige, como já abordado, que os sistemas de IA sejam desenvolvidos e treinados com dados diversificados e representativos, e que sejam implementadas medidas de supervisão contínua para identificar e corrigir quaisquer preconceitos que possam surgir.

III.3 – Sobre o Princípio do Respeito pela Autonomia

À semelhança do que foi abordado nos subcapítulos anteriores, os sistemas de IA introduzem várias preocupações e promessas no Princípio do Respeito pela Autonomia, cuja análise poderá iniciar-se abordando os dois pilares da autonomia.

1. Os sistemas de IA na medicina têm o potencial de influenciar a *liberdade* dos pacientes de diversas maneiras:

A democratização do acesso a informações de saúde pode permitir aos pacientes obter conhecimento detalhado e atualizado acerca da sua saúde e opções de tratamento. Os pacientes são, assim, capacitados, fazendo escolhas mais informadas e sob menos influência de pressões externas, como coerção de profissionais de saúde ou de familiares. O conteúdo e a complexidade das informações podem ser ajustados conforme as suas necessidades e preferências individuais, garantindo a compreensibilidade e relevância para a sua situação específica. Esta abordagem permite que os pacientes façam escolhas livres, baseadas nos seus próprios valores e preferências.

No entanto, o potencial benéfico é igualado pelo potencial lesivo, como já estabelecido.

Erros, vieses e conflitos de interesse introduzidos durante o ciclo de vida da IA podem manipular as escolhas dos pacientes, de formas mais ou menos diretas, restringindo a sua liberdade. As redes sociais são uma demonstração atual de como algoritmos informáticos, associados a técnicas de manipulação psicológica podem influenciar profundamente o julgamento e decisão humana. Até escolhas aparentemente insignificantes, como a forma como os dados são organizados e exibidos pela IA podem direcionar as decisões dos pacientes. (Bar-Ilan et al., 2009) Outras influências poderão ser mais diretas, como a atribuição consciente de fatores de ponderação maiores a certas opções de tratamento, por parte dos desenvolvedores e implementadores, para que as as recomendações da IA promovam os seus objetivos e não necessariamente os valores e preferências pessoais do paciente.

A quantidade e a complexidade das informações fornecidas pelos sistemas de IA podem ser avassaladoras para alguns pacientes, dificultando, em vez de ajudar, a sua capacidade de

processar e compreender a sua própria situação, levando a que tomem decisões mal informadas ou a que fiquem ainda mais dependentes das interpretações de terceiros.

Os riscos da confiança excessiva nestes sistemas não se aplicam só aos profissionais que os utilizam, podendo também afetar os pacientes. Se estes começarem a depender exclusivamente das recomendações da IA para tomar decisões relativas à sua saúde, podem perder a capacidade de exercer um julgamento crítico independente, transformando a IA numa autoridade substituta em vez de um suporte à decisão. A opacidade dos algoritmos pode ser uma barreira significativa à liberdade dos pacientes, fomentando a confiança cega nas decisões da IA, pela impossibilidade de compreender os processos que as originaram, limitando o questionamento e a deliberação.

A desconfiança nos sistemas, por outro lado, alimentada pelo receio de que os dados pessoais possam ser acedidos ou utilizados sem consentimento, por exemplo, pode fazer com que os pacientes ocultem informações importantes para a deliberação do sistema, resultando em que este faça sugestões inadequadas. Tal como, recomendações baseadas em padrões históricos que refletem discriminação ou desigualdade podem influenciar negativamente as decisões de certos grupos de pacientes, diminuindo a sua liberdade de fazer escolhas informadas e justas, contribuindo para a sua marginalização.

2. A *agência* pode ser influenciada pela IA nas suas diferentes componentes, começando pela *compreensão* e *deliberação* e culminando na *ação intencional*.

Os sistemas de IA podem auxiliar significativamente a *compreensão*: (a) algoritmos de processamento de linguagem e técnicas de visualização de dados podem transformar informações médicas complexas em resumos acessíveis e visualmente compreensíveis, criando imagens simples e explicações em linguagem corrente. (b) Interfaces de conversação permitem aos pacientes fazer perguntas específicas e receber respostas detalhadas em tempo real, obtendo explicações interativas e personalizadas sobre diagnósticos, prognósticos e tratamentos.

A *deliberação* pode ser promovida, (a) ajudando os pacientes a considerar todas as opções disponíveis e a pesar os prós e contras de cada uma, (b) apresentando cenários de tratamento e simulando de forma personalizada, usando dados genéticos e outras informações pessoais relevantes, os resultados prováveis de cada opção.

A compreensão e deliberação adequada dos factos médicos sustentam a *intencionalidade das ações* dos pacientes, além disso: (a) As ações podem ser facilitadas por planos de ação personalizados que detalham os passos a seguir para alcançar objetivos de saúde específicos, por exemplo, um sistema de IA que auxilia a gestão da diabetes incluindo recomendações diárias de dieta, exercícios e medicação, ajustadas com base na monitorização dos dados de saúde e *feedback* contínuo do paciente. (b) Técnicas interativas de motivação e adesão comportamental, como o envio de lembretes personalizados, reforço positivo e ludificação dos objetivos, ajudam os pacientes a seguir os seus planos de ação e a comprometerem-se com as suas metas de saúde.

Por outro lado, a IA pode prejudicar a compreensão e deliberação dos pacientes (a) expondo-os a quantidades avassaladoras de informação, como já abordado; (b) Recomendações muito específicas e apresentadas de forma autoritária podem instigar ações sem a adequada reflexão crítica. (c) A longo prazo, o facilitismo, promovido pelas respostas rápidas para todas as perguntas, inibe o desenvolvimento das faculdades críticas de pensamento necessárias para avaliar de forma independente todo o tipo de informações, incluindo as médicas.

As obrigações éticas subjacentes ao respeito pela autonomia também se estendem aos sistemas de Inteligência Artificial:

- A *obrigação negativa* exige que as ações autónomas dos pacientes não sejam sujeitas a controlo ou coação por terceiros, esta é respeitada pela implementação de medidas que impeçam os sistemas de IA de exercer influências manipulativas ou controladoras nos seus utilizadores ou sujeitos visados pela sua utilização.

- A *obrigação positiva* é respeitada através de sistemas que promovam ativamente a escolha autónoma, dando prioridade à transparência, verdade e apoio contínuo, capacitando os pacientes para que tomem decisões informadas.

Além da influência que os sistemas de IA têm na *autonomia em si*, devemos analisar a forma como promovem, ou não, o *respeito pela autonomia* dos pacientes, reconhecendo e valorizando ativamente o direito dos indivíduos de tomar decisões informadas sobre a sua saúde. Neste sentido deverão ser cumpridas as regras e normas derivadas do princípio do respeito pela autonomia dos pacientes, como:

- *Dizer a verdade.* Fornecendo informações médicas precisas e atualizadas, de forma clara e compreensível, e explicando a lógica por trás das suas recomendações.

- *Respeitar a privacidade.* (a) Usando a mínima quantidade de dados pessoais possível. (b) Utilizando técnicas de anonimização e pseudonimização para proteger a identidade dos donos dos dados. (c) Oferecendo aos pacientes o controlo sobre os seus próprios dados, permitindo-lhes decidir que informações deseja partilhar e com quem.

- *Proteger a confidencialidade.* (a) Implementando medidas avançadas de segurança cibernética para proteger a confidencialidade dos dados dos pacientes. (b) Incluindo criptografia de dados e funcionalidades de auditoria que registam e analisam o acesso aos dados dos pacientes, garantindo que informações sensíveis só podem ser acedidas por pessoal autorizado e detetando rapidamente qualquer violação de segurança.

- *Obter consentimento informado.* (a) Usando processos interativos de consentimento informado, nos quais os pacientes recebem explicações detalhadas e personalizadas sobre os procedimentos propostos. (b) Aplicações interativas podem utilizar gráficos, vídeos e simulações para ajudar os pacientes a entenderem melhor os riscos e benefícios. (c) Documentando e armazenando eletronicamente os consentimentos informados, garantindo que todas as etapas do processo são devidamente registadas e que os pacientes podem aceder às suas declarações de consentimento a qualquer momento.

- *Apoiar a tomada de decisões importantes.* (a) Oferecendo aos pacientes e profissionais de saúde análises detalhadas sobre diferentes opções de tratamento, permitindo escolhas informadas e alinhadas com os valores pessoais. (b) Oferecendo aconselhamento automatizado e personalizado, com base no histórico de saúde individual e nas preferências do paciente.

A *avaliação da autonomia* das escolhas dos pacientes é essencial na Saúde. Como já defendido, a IA pode interferir com a capacidade de indivíduos normalmente autónomos tomarem decisões, podendo potenciá-las, mas também impor-lhes limitações. O respeito à autonomia dita que os sistemas de IA não só não constituam limitações, como também deverão ser empregues para colmatar situações de restrição das faculdades cognitivas, seja por estados emocionais, desconhecimento ou presença de pressões externas.

Os sistemas de IA poderão desempenhar um papel crucial na *avaliação da capacidade*, fornecendo ferramentas e processos avançados para avaliar se os pacientes apresentam sinais indicativos de limitação da sua aptidão para fazer escolhas autónomas.

A IA pode ser utilizada para monitorizar, de forma continuada, o comportamento e as respostas dos pacientes ao longo do tempo e em diferentes contextos. A consistência destas respostas e dados de múltiplas fontes, como sensores biométricos, gravações de vídeo e áudio, podem ser integrados para criar um perfil abrangente do comportamento do paciente e reconhecer padrões que sugiram alterações cognitivas ou emocionais que podem não ser identificáveis numa observação direta. Alterações subtis na cognição ao longo do tempo, pequenas mudanças nos padrões de resposta ou no tempo de reação podem ser indicativas de declínios cognitivos que afetam a capacidade. Dados de avaliações passadas podem ser comparados com avaliações atuais para identificar tendências de deterioração cognitiva.

Algoritmos de linguagem podem ser utilizados para analisar as capacidades linguísticas dos pacientes, identificando dificuldades na compreensão e uso da linguagem, falhas em manter uma conversação lógica, ou desvios significativos na expressão verbal. Inclusive pode identificar presença de vieses cognitivos nas respostas dos pacientes, como a tendência para

decisões baseadas em medo ou desinformação, ajudando a caracterizar barreiras à tomada de decisão informada.

Funções executivas, como memória de trabalho, atenção e raciocínio lógico poderão ser avaliadas através de simulações interativas que imitam cenários reais, permitindo que os pacientes demonstrem as suas capacidades de tomada de decisão em ambientes controlados.

A IA poderá potenciar ou colmatar limitações à capacidade de fazer escolhas autónomas:

- *Incapacidade de expressar ou comunicar preferências e escolhas.* Pacientes com dificuldades na expressão verbal ou escrita, por limitações físicas, cognitivas ou socioculturais, poderão ser ajudados a comunicar as suas preferências por tecnologias como assistentes de comunicação, reconhecimento de voz e conversão de texto-fala. Não obstante, a dependência nestas tecnologias poderá originar, no caso de falhas ou indisponibilidade dos sistemas, uma inaptidão para que sejam expressas eficazmente as preferências do paciente e que interfaces complexas ou mal ajustadas às capacidades cognitivas dos pacientes podem sobrecarregá-los, dificultando ainda mais a comunicação eficaz.

- *Compreender a própria situação e as consequências das suas decisões.* Materiais multimédia educativos interativos e personalizados, gerados por IA, poderão explicar de forma mais acessível e compreensível a situação de saúde do paciente e as potenciais consequências das suas decisões. A adaptação inadequada destes materiais ao nível de compreensão do paciente, pode aumentar a confusão ou desinformação, ao invés de esclarecer.

- *Compreender informações relevantes para a tomada de decisão.* Algoritmos de processamento de linguagem natural podem resumir informações médicas detalhadas em pontos chave que são mais fáceis de compreender e responder a perguntas dos pacientes, clarificando informações que possam não ter sido totalmente compreendidas inicialmente. Respostas erradas, pouco claras, irrelevantes ou excessivamente simplificadas poderão prejudicar a compreensão.

- *Fornecer razões lógicas para suas escolhas.* Guias de decisão, como questionários e árvores de decisão, podem ajudar os pacientes a refletir, estruturar e articular as razões lógicas para as suas opções de tratamento, baseando-se nos prós e contras de cada escolha e em dados concretos. A dependência excessiva desincentiva o desenvolvimento das capacidades analíticas e deliberativas baseadas em valores e juízos próprios e a complexidade acrescida destas análises pode impossibilitar a sua avaliação crítica eficaz, condicionando a confiança cega nestas, que poderão ser tendenciosas e incorretas.

- *Manter um pensamento racional ao considerar opções.* A sensação de imparcialidade e despersonalização da IA pode reduzir o stress associado à tomada de decisões em alguns pacientes, o que juntamente com a inclusão de funcionalidades de suporte emocional, como técnicas de terapia cognitivo-comportamental, pode ajudar os pacientes a manter a clareza mental durante o processo de decisão. O objetivo de minimizar o stress e ansiedade da decisão poderá simplificar em demasia as recomendações ou levar à omissão de informações importantes. Técnicas de suporte emocional e simulações de respostas emocionais fornecido pela IA não substituem de todo o apoio humano, além de parecerem muitas vezes insensíveis ou inapropriadas, o que pode agravar o sentimento de isolamento ou incompreensão.

- *Analisar riscos e benefícios.* Modelos preditivos associados a análises detalhadas de risco e benefício, poderão apresentar as implicações de cada opção de tratamento de forma que os pacientes as possam facilmente compreender. Estes modelos podem, contudo, não capturar adequadamente as especificidades individuais dos pacientes, resultando em análises de risco e benefício que não refletem verdadeiramente as suas circunstâncias individuais. A má compreensão dos pressupostos ou limitações desses modelos poderá provocar que se tomem decisões baseadas em previsões falaciosas.

- *Decidir de forma racional e justificada.* Os sistemas de IA podem apoiar todo o processo de decisão, como já visto, no entanto apresentam vários riscos subjacentes, como o encorajamento à postura passiva e acrítica perante as sugestões apresentadas e

possibilidade de enviesamento ou inadequação dos algoritmos, dados e recomendações. Uma outra aplicação útil poderá ser a documentação automática do processo de decisão, na qual se registam as justificações fornecidas pelos pacientes para as suas escolhas para referência futura, garantindo o respeito da autonomia a longo termo. A complacência poderá levar à assunção errada de que as decisões foram bem documentadas e justificadas, embora exista o risco das nuances e racionalizações individuais serem mal representadas pelos sistemas, distorcendo as verdadeiras decisões do paciente.

A IA pode também influenciar a *formulação de juízos normativos de incapacidade*:

1. *Identificação das capacidades relevantes.* A IA pode ser usada no mapeamento detalhado das capacidades cognitivas e emocionais, tomando proveito das suas grandes bases de dados para identificar padrões e correlacionar faculdades específicas, necessárias para se considerar um individuo capaz em diferentes contextos.

2. *Determinação do grau mínimo necessário de capacidades.* sistemas de IA podem analisar dados de diversos pacientes em diferentes contextos clínicos para construir *benchmarks* baseados na evidência que determinam qual o grau mínimo necessário de aptidões para se considerar um paciente capaz num determinado contexto.

3. *Seleção e aplicação de testes empíricos.* A IA pode ser utilizada para desenvolver e integrar novos testes empíricos que avaliam as capacidades cognitivas, emocionais e de julgamento.

Os sistemas de IA têm um impacto significativo na facilitação e melhoria do processo de *consentimento informado*, que assegura que as decisões dos pacientes são tomadas de forma livre e informada, integrando *elementos base*, *elementos de informação* e *elementos de consentimento*.

1. *Elementos Base.* A *avaliação da capacidade* do paciente pode ser realizada pela IA de forma automatizada e contínua, utilizando algoritmos que analisam as faculdades cognitivas e emocionais. O *voluntarismo* poderá ser protegido identificando sinais de

coerção ou manipulação durante o processo de consentimento, analisando padrões de linguagem e comportamento e alertando os profissionais de saúde para potenciais influências externas.

2. *Elementos de Informação.* O nível de detalhe e complexidade da *divulgação de informação material* pode ser personalizada com base nas necessidades individuais do paciente. Explicações interativas dos tratamentos, utilizando multimédia, gráficos e simulações, juntamente com respostas a qualquer momento às perguntas dos pacientes podem melhor ilustrar e explicar os diagnósticos, prognósticos, naturezas e propósitos das intervenções, bem como as suas alternativas e riscos, aprofundando o entendimento da informação. Justificações detalhadas para as opções de tratamento apresentadas podem ser fornecidas ao paciente, baseando-se na evidência científica e ajustadas ao perfil de saúde deste, garantindo que compreende a lógica por trás das sugestões médicas e capacitando a escolha informada.

3. *Elementos de Consentimento.* Guias automatizados podem ajudar os pacientes através do processo de tomada de decisão, oferecendo suporte estruturado, ajudando a clarificar as suas opções e motivando a reflexão sobre os seus valores e preferências, culminando numa decisão bem fundamentada. A digitalização, armazenamento e a documentação do consentimento informado, garante que todos os passos do processo são devidamente registados, acessíveis e em conformidade com as melhores práticas morais e legais.

Todavia, existem situações onde o consentimento informado pode não ser necessário ou possível de obter. Estas situações incluem emergências médicas, questões de saúde pública e investigações que utilizem dados anonimizados.

- *Emergências médicas.* Quando o consentimento explícito não pode ser obtido, o respeito pelos valores e preferências do paciente poderá ser promovido pela análise e acesso rápido a decisões previamente expressas pelo paciente, por exemplo, em

diretivas antecipadas de vontade e registos médicos anteriores, permitindo guiar as decisões de tratamento de acordo com a vontade presumida do paciente.

- *Questões de Saúde Pública.* A IA pode ser utilizada para analisar dados em larga escala e identificar tendências e riscos emergentes em saúde pública, ajudando na formulação de políticas e campanhas de saúde. No entanto, é essencial minimizar o impacto sobre a autonomia dos indivíduos, assegurando a privacidade e devida anonimização dos seus dados, e justificando sempre as decisões tomadas, priorizando a transparência e responsabilização.

- *Investigações com dados anonimizados.* As várias violações de bases de dados nas últimas décadas (Hammouchi et al., 2019) devem instigar um grau saudável de ceticismo na sua utilização maciça para treinar modelos. Os dados deverão, por isso, cingir-se ao mínimo imprescindível à tarefa em questão, e sempre adequadamente anonimizados e aleatorizados, impossibilitando a identificação dos pacientes. O grau de desconfiança e de exigência na proteção devem ser tanto maiores quanto mais críticos forem os dados. Embora o consentimento explícito possa não ser necessário para o uso de dados anonimizados, os investigadores devem ser transparentes sobre os métodos de anonimização e os propósitos da investigação, promovendo a transparência e responsibilização. Sobretudo porque existem métodos tradicionais de anonimização que, com a introdução da tecnologia de IA, pode torna-se insuficientes para assegurar a confidencialidade dos dados, pela possibilidade de se utilizar as potencialidades da IA para fazer *ataques adversariais* com o objetivo de extrair informações confidenciais das bases de dados. (Carlini et al., 2020)

Quando os pacientes não são autónomos ou existem dúvidas quanto à sua autonomia, as decisões de saúde devem ser delegadas a representantes. A IA pode desempenhar um papel significativo em apoiar este processo, garantindo que as decisões refletem os valores e os melhores interesses dos pacientes. Relativamente aos três principais padrões, *Padrão do Julgamento Substituído*, o *Padrão de Autonomia Pura* e o *Padrão do Melhor Interesse*:

- *Padrão do Julgamento Substituído.* A IA pode ajudar a recolher e analisar dados detalhados sobre as preferências passadas dos pacientes. A análise de registos médicos, notas de consultas anteriores e diretivas avançadas de vontade podem ajudar a construir um perfil compreensivo das preferências do paciente, que poderão servir de base para fornecer suporte aos representantes oferecendo informações contextuais e recomendações baseadas nestas. Um exemplo é a simulação de cenários, apresentado as opções que mais se alinham com os valores previamente expressos pelo paciente.

- *Padrão de Autonomia Pura.* A IA pode armazenar diretivas antecipadas de forma segura e garantir que estas são facilmente acessíveis quando necessário. Os profissionais de saúde e os representantes poderão ser alertados sobre a existência de tais diretivas, garantindo que são sempre consideradas nas decisões de tratamento. Em situações onde a condição do paciente se altera significativamente, a IA pode ajudar a reavaliar a aplicabilidade das diretivas antecipadas, utilizando dados atualizados sobre a saúde do paciente e algoritmos de previsão, para sugerir se uma diretiva antecipada ainda é aplicável ou se deve ser reconsiderada à luz das novas circunstâncias.

- *Padrão do Melhor Interesse.* Análises de risco-benefício detalhadas que refletem as condições de saúde específicas do paciente e as diferentes opções de tratamento ajudam os representantes a tomar decisões que maximizem o benefício provável para o paciente. Os diferentes cenários de tratamento poderão ser simulados, mostrando os potenciais resultados e impactos de cada opção e permitindo aos representantes visualizarem as consequências das suas decisões e a escolherem a opção que melhor se alinha com os interesses do paciente. Recomendações baseadas na análise e agregação de dados de casos semelhantes, poderão ajudá-los a fazer escolhas informadas e justificadas, mesmo na ausência de preferências previamente expressas pelo paciente.

III.4 – Sobre o Princípio da Justiça

As decisões dos sistemas de IA não são desprovidas de valor moral, refletindo as escolhas e ações (conscientes e inconscientes) motivadas e condicionadas, em diferentes proporções, pelas crenças e valores morais de todos os envolvidos no seu desenvolvimento, implementação e utilização. No entanto, a IA, à luz do que é o seu enquadramento tecnológico atual, é ela própria desmotivada e desinteressada no seu funcionamento, o que torna o seu processo decisório *formalmente* imparcial. Estas características são *teoricamente* promotoras do *Princípio Formal da Justiça*, sendo desejáveis para que se trate de forma igual casos iguais e de forma desigual casos diferentes. Todavia, assegurar que a IA protege a Justiça *na prática* requer que o *conteúdo* das suas decisões reflita e promova valores de justiça. Em circunstâncias moralmente complexas, nas quais não existe consenso sobre que valores morais implicados na decisão são superiores, estes sistemas não devem ser utilizados, sendo preferível remeterem a apreciação ética para decisores humanos, com a competência e responsabilidade necessária.

> A IA será sempre inadequada, por exemplo, para julgar legalmente um homicídio, cujo julgamento exige, para além de pressupostos legais, compreensão causal, juízo moral, raciocínios e pensamentos de alta ordem, faculdades emocionais e sociais. Estas capacidades humanas, ainda que fossem passíveis de ser programadas numa IA, implicariam formas de raciocínio automático de tal forma complexas, que os tornariam necessariamente em caixas-negras, que evidentemente não poderão ser aplicadas para propósitos de impacto tão elevado na vida da pessoa julgada, o que impõe que qualquer sentença terá de ser bem explicada, fundamentada e compreensível.
>
> Ainda assim, em linha com o exemplo legal anterior, decisões mais simples poderão beneficiar da imparcialidade da IA para assegurar o tratamento igual, independentemente do estatuto social de um individuo ou conflitos de interesse intrínsecos à situação. Exemplificando, um sistema de IA - devidamente desenvolvido, implementado, utilizado e auditado - poderia eventualmente detetar e punir infrações de trânsito, ignorando a posição social do infrator ou outros fatores moralmente

arbitrários como critério, imparcialidade considerada por vezes dúbia nos fiscais humanos.

Na saúde, a análise transversal de largas quantidades de dados clínicos facilita a identificação das necessidades de um segmento da população e se estas são atendidas de forma oferta equivalente à de outros subgrupos. Os padrões estabelecidos permitem averiguar se pacientes com condições semelhantes recebem tratamentos equivalentes e que pacientes com necessidades especiais recebem cuidados adaptados à mitigação desta diferença.

Contudo, a existência de sistemas de IA, construídos para este propósito, que sejam enviesados, de forma acidental ou intencional, poderá introduzir ou reforçar desigualdades de tratamento injustas. Decisões automáticas que favorecem certos grupos em detrimento de outros podem resultar, por exemplo, (a) da introdução intencional de critérios discriminatórios durante o desenvolvimento ou (b) do treino baseado em dados que representam desproporcionalmente certos grupos demográficos.

Recuperando a ideia ilustrada no Subcapítulo III.2 sobre transportar toda a tecnologia, informação e recursos necessários à construção de um sistema de IA sofisticado para outras alturas e locais históricos, permitindo à sua população faça as escolhas durante o ciclo de vida da IA, será evidente as injustiças e imoralidades subjacentes às hipotéticas recomendações feitas por um sistema de IA desenvolvido por cientistas Nazis e treinado com dados médicos por si determinados e colhidos, ainda que o sistema em si não tivesse motivação ou interesses próprios. Este exemplo extremo pretende ilustrar a possibilidade conceptual de uma tecnologia, em si amoral, perpetuar graves injustiças e imoralidades. No entanto, é expectável que os vieses dos sistemas enquadrados numa sociedade democrática moderna sejam menos graves, mas por isso, mais subtis, o que fomenta que se cometam incontáveis pequenas infrações ao Princípio da Justiça.

Os *Princípios Materiais de Justiça aplicados à Saúde* irão especificar os critérios para a distribuição de recursos de saúde, determinando quem deve receber que recursos e em que medida. Embora estes princípios detenham um caráter altamente ideológico e subjetivo,

algumas funcionalidades da IA podem ser empregues de forma ubíqua, auxiliando na operacionalização dos mais diversos princípios materiais:

- Aumentando a eficiência na distribuição de recursos ao prever que áreas ou grupos irão necessitar mais urgentemente de assistência, baseando-se em dados populacionais, económicos, geográficos, etc. Estas previsões permitem alocar, de forma mais informada, recursos limitados de maneira a maximizar os benefícios gerais para a sociedade, conforme estabelecidos pelas diferentes teorias de justiça.
- A monitorização contínua e detalhada da distribuição de recursos e dos resultados desta permite ajustes rápidos e informados nas políticas e práticas de alocação, permitindo decisões de justiça distributiva dinâmicas, mais bem-adaptadas às necessidades variáveis da população, consoantes os critérios de justiça vigente.

Algo importante a considerar é que a utilização de sistemas de IA construídos com base numa só conceção de justiça distributiva resultará na uniformização de decisões que não são necessariamente unânimes. Por exemplo, um Sistema de IA que defenda uma abordagem mais coletivista irá sempre priorizar a igualdade na distribuição dos recursos de saúde os direitos individuais, como o direito à autonomia. Esta uniformização de decisões não unânimes, associada à falta de transparência e compreensibilidade impõe, de forma antidemocrática, uma só conceção de justiça distributiva, definida pelas pessoas e/ou entidades que estabeleceram os critérios da IA. A mera perceção de que estes sistemas permitem a tomada de decisões incompreensíveis e por isso incontestáveis pelo público geral e pelos seus representantes eleitos, de forma tirânica, poderá minar a confiança nas instituições democráticas, incluindo de saúde, que os utilizam, mesmo se decisões até fossem baseadas em critérios justos. Elencar as profundas implicações sociais *realísticas* da IA seria um trabalho em si só, mas é importante sublinhar a implausibilidade de que seja esta tecnologia a promover união política e social e que a Saúde não se encontra num vácuo.

Independentemente das conceções sobre que critérios específicos definem a justiça distributiva, podemos constatar que a *acessibilidade* e *qualidade dos cuidados de saúde* são

duas áreas consideradas prioritárias para a maioria dos governos que estão investidos na promoção da saúde dos seus cidadãos. Já foram abordados vários exemplos de como as aplicações práticas da IA[65] poderão contribuir para estes dois grandes objetivos através das suas capacidades promotoras da eficiência, eficácia e precisão e pela natureza desmaterializada do seu acesso.

Não sendo a Saúde a única necessidade e prioridade de uma sociedade, a IA deve contribuir para a *contenção de custos e a proteção dos recursos públicos*, aumentando a eficiência operacional dos sistemas de saúde, por exemplo, otimizando a gestão de inventários de medicamentos, prevendo a necessidade de recursos com base em tendências epidemiológicas e melhorando a alocação de pessoal de saúde conforme a procura. Contudo, deve ser feita a ressalva que a implementação de sistemas de IA envolve custos iniciais elevados, não só pelas infraestruturas e tecnologias de suporte necessárias, como na sua aquisição, licenciamento, desenvolvimento e manutenção, e também na formação dos profissionais para a sua utilização. Se estes custos não forem geridos adequadamente, poderão constituir mais um peso significativo nos orçamentos de estado e da saúde, potencialmente desviando fundos de outras necessidades essenciais.

Do ponto de vista da *equidade de acesso à Saúde*, se os sistemas de IA forem apenas desenvolvidos e implementados por empresas privadas com interesses comerciais, para além da possibilidade de serem implementados modelos que menosprezam a segurança e o bem-estar dos pacientes priorizando o lucro, existe o risco de se agravar a desigualdade no acesso a serviços de saúde de alta tecnologia, que fica reservado apenas àqueles que os podem pagar. Poderá surgir, então, um sistema de saúde de duas velocidades, com tratamentos cada vez mais avançados e personalizados para uns, enquanto quem não os possa pagar fica limitado aos cuidados de saúde progressivamente mais desatualizados oferecidos pelo sistema público, prejudicando a coesão e justiça social.

[65] Por exemplo, (a) diagnósticos automáticos precoces e precisos, (b) personalização de tratamentos, (c) monitorização contínua do estado de saúde da população, (d) uso das capacidades preditivas em intervenções de prevenção e (e) automatização de tarefas administrativas.

A IA pode ter efeitos extremamente positivos e impactantes na economia e no orçamento de estado de um país, otimizando a alocação de recursos públicos através de análises preditivas precisas, e o aumento geral da eficiência e produtividade de um país promove o desenvolvimento socioeconómico e o aumento do conjunto de recursos disponíveis, incluindo para a Saúde. (Comissão Europeia, 2020)

- *Alocação dentro do orçamento alvo.* A IA também pode ser instrumental na definição de prioridades para o financiamento de projetos e procedimentos específicos, por exemplo, (a) permitindo análises de custo-benefício mais precisas para direcionar os investimentos para intervenções que ofereçam os maiores benefícios relativos aos custos envolvidos; ou (b) usando critérios como a frequência de doenças, escalas de impacto social, dor e sofrimento associados, para identificar áreas onde os recursos terão maior impacto.

- *Alocar tratamentos escassos aos doentes.* A IA pode facilitar a alocação justa de tratamentos escassos ao identificar os pacientes que mais beneficiarão de determinados tratamentos com base em análises detalhadas de dados clínicos. Esta abordagem pode ajudar a garantir que os recursos limitados são utilizados de forma a maximizar os benefícios para o maior número de pessoas. Sistemas de IA podem auxiliar na triagem e priorização de pacientes de forma mais objetiva e transparente, reduzindo o risco de decisões arbitrárias e aumentando a equidade no acesso aos cuidados.

A IA pode melhorar a *priorização* de recursos de saúde através de análises de custo-benefício e custo-eficácia avançadas que empregam grandes volumes de dados clínicos e epidemiológicos para determinar a alocação de recursos de maneira que maximize os benefícios de saúde. A análise de dados em tempo real também permite ajustes rápidos nas prioridades de saúde, refletindo prontamente mudanças nas necessidades de saúde da população.

No entanto, a fundamentação em metodologias utilitaristas, para além de negligenciar fatores importantes que não são facilmente quantificáveis, como a dignidade, autonomia e o valor intrínseco da vida humana, pode sustentar decisões que, embora eficientes, não são necessariamente justas. Poderão, por exemplo, negligenciar inadvertidamente as necessidades dos grupos sub-representados nos seus dados de treino como idosos, crianças, minorias étnicas/raciais ou pessoas com deficiências.

Os critérios de alocação são também construídos muitas vezes, mesmo que inconscientemente a pensar num segmento maioritário da população, o que resulta, por exemplo, na consideração dos tratamentos para doenças raras ou crónicas que afetam pequenas populações como menos prioritários, por não satisfazerem os mesmos critérios de benefícios agregados que outras intervenções mais abrangentes, colocando a eficiência económica acima da justiça.

Estes sistemas devem ser, por isso, enquadrados num modelo bioético robusto, e o seu desenvolvimento requer particular colaboração interdisciplinar, envolvendo não apenas engenheiros, cientistas de dados e economistas, mas também especialistas em bioética, profissionais de saúde e representantes políticos que assegurem uma abordagem moral na priorização dos cuidados de saúde.

A IA afeta também os diferentes tipos de *racionamento de recursos* de saúde:

- No contexto em que o acesso aos tratamentos é regulado pela *capacidade individual de pagar*, a IA pode oferecer ferramentas para uma melhor gestão e transparência dos custos associados aos tratamentos. Algoritmos de IA podem ajudar a prever e gerir custos, facilitando a tomada de decisões informadas por parte dos pacientes e prestadores de serviços de saúde. Além disso, a IA pode identificar ineficiências e áreas onde os custos podem ser reduzidos, potencialmente tornando os tratamentos mais acessíveis a uma maior faixa da população. Por outro lado, a alocação de recursos orientada pelo poder de compra é propícia a que apenas as pessoas mais ricas e influentes tenham acesso aos cuidados de ponta proporcionados pela IA, agravando as disparidades de saúde.

- A IA pode ser um aliado valioso na monitorização e gestão dos limites máximos de recursos que cada indivíduo pode aceder *impostos por um governo*, otimizando a distribuição dos recursos disponíveis, garantindo que os limites estabelecidos sejam respeitados, facilitando a deteção de abusos e fraudes no sistema, para que se maximize os benefícios clínicos para o maior número de pessoas. No entanto a rigidez dos limites estabelecidos por IA aliado à opacidade dos processos de decisão algorítmica origina um sistema tirânico que desumaniza os pacientes, vendo-os como meros números e desconsiderando as suas necessidades individuais e circunstâncias específicas.

- No modelo de *acesso igualitário a recursos de saúde considerados imprescindíveis*, sistemas de IA cruzar dados demográficos e médicos para avaliar o acesso aos cuidados básicos necessários, independentemente da condição socioeconómica dos pacientes. A identificação e eliminação de ineficiências na distribuição dos recursos aumenta a sua disponibilidade relativa, podendo este diferencial positivo ser alocado para o benefício de quem mais precisa. As ferramentas de apoio clínico e de prestação de cuidados da IA são muitas vezes computadas remotamente, não requerendo a presença localizada de infraestruturas complexas e caras. Esta incorporeidade alarga a oferta de certos cuidados de saúde, como diagnósticos e tratamentos, considerados adequadamente administráveis por IA ou facilitados por esta, a populações que carecem atualmente destes cuidados, desde que se assegure a disponibilidade generalizada de tecnologias digitais, de informação e de rede. Estas soluções não devem ser utilizadas como pensos rápidos que se tornam permanentes, reduzindo o investimento nos cuidados de saúde humanizados em áreas marginalizadas em prol de cuidados automatizados. A dualidade de sistemas de saúde poderá ocorrer tanto da iniquidade do acesso a cuidados de saúde de ponta impulsionados por IA, como da utilização da automatização como solução rápida, ainda que não plenamente eficaz, para problemas socioeconómicos de base.

- A IA permite teoricamente qualquer *outro tipo de racionamento* usando os dados, critérios e métricas que lhe forem impostos para determinar a forma como os recursos são distribuídos:

  No caso da *alocação baseada na idade*, priorizando a maximização da longevidade e qualidade de vida até ao paciente alcançar uma "idade de vida normal". A perceção errónea da neutralidade da IA, aliado ao seu potencial funcionamento como caixa negra pode permitir que se utilizem critérios discriminatórios inaceitáveis e os algoritmos devem ser auditáveis e constantemente escrutinados para assegurar que não disseminam decisões imorais de forma oculta, prejudicando indivíduos e grupos por características pelas quais não são moralmente responsáveis.

  A *triagem* de pacientes é uma das funcionalidades da IA mais bem estabelecidas atualmente, demonstrando ser eficaz na alocação de recursos com base na utilidade médica. O acesso remoto, através do telemóvel, a estas triagens automatizadas permite evitar idas desnecessárias às urgências hospitalares, poupado tempo e carga de trabalho especializado que pode ser mais bem aplicado a casos mais complexos. No entanto, é fundamental assegurar que as triagens automáticas não constituem barreiras indevidas ao acesso a cuidados de saúde, causadas por critérios demasiado restritos ou por falhas técnicas dos sistemas de IA, atrasando o início de tratamento de pacientes com doenças agudas graves ou até mesmo desmotivando a procura de cuidados de saúde, levando a consequências desastrosas. A abordagem cética e racional é a de não assumir a infalibilidade dos sistemas automáticos e consequentemente exigir que ao errarem pequem por excesso de zelo, sendo preferível permitir demasiada acessibilidade e desperdiçar recursos, do que exigir demasiada restringência resultando em perdas de vida humana.

Sistema de IA também poderão influenciar a alocação de recursos escassos de saúde através do estabelecimento e aplicação de *critérios de elegibilidade e de seleção final*:

(a) Através da análise de *fatores não médicos*, como a cidadania e a capacidade financeira dos indivíduos. (b) Ao identificar que pacientes são mais adequados para aumentar a *validação científica* de determinados estudos, baseando-se em múltiplas variáveis e dados complexos, dificilmente aplicáveis manualmente. (c) Otimizando a alocação de recursos com base na *maior probabilidade de sucesso terapêutico*, determinada pela análise de dados clínico, reduzindo o desperdício de tratamentos escassos em casos nos quais estes não produziriam efeito. (d) Ao facilitar a aplicação do critério de *necessidade médica*, monitorizando continuamente o estado de saúde dos pacientes e identificar aqueles com condições mais graves e emergentes, assegurando que os recursos são direcionados para aqueles em maior risco. (e) Para auditar continuamente *mecanismos impessoais* como lotarias e listas de espera, reduzindo o risco de erros ou manipulação humana, aumentando a transparência e assegurando que os processos são justos.

Os sistemas de IA podem ser programados para identificar vieses atuais que determinam o acesso a tratamentos de saúde, para que se os possa corrigir. Por outro lado, podem perpetuar ou até exacerbar as desigualdades existentes através das correlações identificadas nos seus dados, que se a causalidade subjacente não for devidamente analisada e entendida, podem motivar decisões que mantêm ou até intensificam relações distributivas injustamente estabelecidas, transformando vieses, por vezes inconscientes, em normas algorítmicas pervasivas que reforçam injustiças sistémicas.

A *algoritmização* de decisões complexas e dependentes do contexto podem prejudicar a alocação justa de recursos em circunstâncias específicas, pouco ou nada refletidas nos dados pela sua raridade, por exemplo, pacientes com múltiplas comorbilidades ou com doenças raras podem perder a oportunidade de receber tratamentos simplesmente pela maior imprecisão das perspetivas de sucesso, que pode nem ser necessariamente inferior, só mais difícil de calcular.

A seleção por perspetivas de sucesso é, só por si, controversa. Mesmo que existisse um sistema que conseguisse calcular com enorme precisão a probabilidade de um tratamento trazer benefícios para o doente, o número calculado não seria plenamente esclarecedor de qual a

decisão mais justa a tomar, por exemplo, um tratamento que tenha 80% de probabilidade de surtir um efeito benéfico num paciente de 90 anos, que já tendo ultrapassado bastante a esperança média de vida, não irá realisticamente usufruir deste durante muito tempo, pode ser justificadamente atribuído a um paciente de 20 anos, mesmo que a sua probabilidade de sucesso seja menor, exemplificando, sendo 70%. Por mais precisos que sejam os cálculos, os pacientes continuarão a não ser números e não poderão ser tratados como tal, os profissionais de saúde e decisores políticos têm também a responsabilidade de escrutinar que os critérios de elegibilidade definidos e aplicados são justos.

O conceito de *oportunidade justa*, conforme delineado por John Rawls, tem várias implicações na implementação dos sistemas de IA na saúde. Este princípio estipula que ninguém deve ser beneficiado ou prejudicado devido a características sobre as quais não têm controlo, como género, raça, inteligência, etnicidade, nacionalidade e estatuto social. Além disso, a sociedade deve mitigar desvantagens moralmente arbitrárias, como deficiências, para promover o bem-estar mais equitativo e participação social das pessoas com elas afligidas.

- A IA tem o potencial de *detetar vieses subjacentes à alocação* de recursos médicos, detetando de forma abrangente se as decisões são tomadas com base em critérios médico-científicos ou com base em características comprovadamente arbitrárias como género, raça, etnicidade ou estatuto social. A identificação de padrões de desigualdade na prestação de cuidados de saúde permite a criação de políticas e intervenções que corrijam disparidades injustas no acesso e nos resultados dos tratamentos entre diferentes grupos populacionais. Por exemplo, se a IA identificar que certos grupos raciais ou étnicos estão sub-representados em tratamentos eficazes, medidas corretivas podem ser implementadas para garantir que esses grupos recebam atenção adequada.
- A *personalização dos cuidados médicos* pode ajudar a mitigar desvantagens moralmente arbitrárias, melhor atendendo às necessidades individuais, para que os pacientes desafortunados, por exemplo com deficiências ou doenças crónicas, possam alcançar um nível mais equitativo de bem-estar e participação social.

A IA viola o Princípio da Oportunidade Justa através de todas as formas abordadas em que promove desigualdades sociais com base nos vieses introduzidos, conscientemente ou inconscientemente, ao longo de todo o seu ciclo de vida. Os modelos poderão, pelas suas características técnicas, obscurecer decisões injustas, tornando mais difícil a sua correção e a responsabilização de quem para elas contribuiu.

A *lei da compensação*, que dita que se deva mitigar desvantagens morais arbitrárias, pode ter na IA uma grande mais-valia, pelas mais-valias trazidas a pacientes desafortunados por causa das suas inovadoras aplicações reabilitativas e de cuidados personalizados ajudando indivíduos com deficiências a melhorar sua integração, participação social e independência através de tecnologias como próteses inteligentes, sistemas de comunicação e interfaces cérebro-máquina. Todavia, existe o risco do acesso a estas tecnologias avançadas seja limitado por fatores não médicos, como o poder económico ou a instrução digital, e é por isso crucial que se criem políticas públicas que procuram aumentar o alcance destas terapias inovadoras.

No contexto da investigação científica deverá ser analisada a sua influência na *indução injusta* e *lucro indevido*, problemas que surgem na escolha de participantes em estudos.

A IA pode ser utilizada para identificar estudos científicos que têm sobrerrepresentação de participantes em situações socioeconómicas desfavoráveis, ajudando a identificar e prevenir a *indução injusta*. A análise de dados demográficos e socioeconómicos pode detetar padrões de recrutamento que aparentam induzir a participação desproporcional de indivíduos com dificuldades económicas, com esta informação, os investigadores podem ajustar os seus critérios de seleção para assegurar uma representação mais justa e equitativa dos participantes, reduzindo a probabilidade de que sejam as compensações financeiras a persuadir os indivíduos a aceitar riscos elevados que não aceitariam de outra forma.

Neste caso também poderá surgir o problema da correlação, algoritmos de IA encarregues de selecionar participantes, treinados com dados enviesados que refletem práticas históricas de recrutamento, podem originar decisões que favorecem a inclusão de indivíduos economicamente vulneráveis em estudos de risco, por vez de uma forma ainda mais opaca.

Sistemas de IA podem ser usados para regular se as compensações oferecidas aos participantes em estudos de investigação são justas e proporcionais aos riscos assumidos, calculando compensações financeiras que reflitam de forma mais precisa os riscos e esforços envolvidos, baseando-se em dados detalhados sobre diversos estudos e os impactos que tiveram nos seus participantes. Esta abordagem pode ajudar a prevenir a exploração financeira de participantes desfavorecidos, assegurando que as compensações são suficientemente altas para ser justas, mas não tão altas que se tornem coercitivas. Por outro lado, os próprios sistemas de IA podem ser utilizados pelos patrocinadores dos estudos para maximizar a eficiência económica destes em detrimento dos direitos dos participantes, agravando injustiças na investigação biomédica.

Conclusão

As propostas principais deste trabalho de (1) reunir informação e conhecimento multidisciplinar essencial, e (2) utilizar este substrato de informação para levantar e analisar diversas considerações bioéticas resultantes da implementação de sistemas de IA na Medicina, foram atendidas ao longo dos seus três capítulos.

No capítulo I, foi explanado o modelo bioético, comumente designado de *Principialismo*. A apresentação dos seus quatro Princípios orientadores – da Beneficência, da Não Maleficência, do Respeito pela Autonomia e da Justiça –, das regras e normas destes derivadas, e de particularidades aplicadas à prática clínica, permitiu analisar, posteriormente, em detalhe as aplicações dos sistemas de IA na Medicina, com base num modelo ético coerente e relevante para a realidade clínica.

O capítulo II partiu da definição de IA proposta pelo Grupo Independente de Peritos de Alto Nível, criado pela Comissão Europeia, para clarificar os seus principais conceitos teóricos e técnicos e desmistificar algumas das ideias mais prevalentes relativas a esta. Foram igualmente detalhados vários subtipos de IA e o seu funcionamento de base, com especial foco nas suas aplicações na Medicina.

Uma vez satisfeita a primeira proposta deste trabalho e em resposta à segunda, o Capítulo III focou-se na enumeração de várias considerações importantes que se levantam da aplicação destes Sistemas na Medicina e na análise, do ponto de vista Principialista, das mesmas.

A finalidade última deste trabalho é satisfeita se (a) a sua leitura elucidar profissionais de saúde em relação ao funcionamento da IA e as implicações bioéticas e práticas que esta poderá ter na sua área profissional; (b) se demonstrar a atuais e futuros desenvolvedores e legisladores destes Sistemas as inúmeras obrigações morais e humanas que a Saúde implica, mais exigentes do que grande parte das outras áreas de possível implementação desta tecnologia; e principalmente (c) se esta informação demonstrar a importância do envolvimento multidisciplinar ao longo de todo o ciclo de vida dos sistemas de IA aplicados à Saúde, para que

se possa colher os benefícios desta tecnologia promissora, sem causar qualquer prejuízo aos superiores valores morais e humanos subjacentes à prestação de cuidados de saúde.

Bibliografia

Abdallah, S., Sharifa, M., Khaleel IKH ALMADHOUN, M., Muneeb Khawar Sr, M., Shaikh, U., Balabel, K. M., Saleh, I., Manzoor, A., Kumar Mandal, A., Ekomwereren, O., Mon Khine, W., Oyelaja, O. T., & al Ahmad Al Jaber Al, J. (2023). *The Impact of Artificial Intelligence on Optimizing Diagnosis and Treatment Plans for Rare Genetic Disorders*. https://doi.org/10.7759/cureus.46860

Agrawal, A., Gans, J., & Goldfarb, A. (2022). *Prediction Machines* (Updated and expanded). Harvard Business Review Press.

Albahri, A. S., Duhaim, A. M., Fadhel, M. A., Alnoor, A., Baqer, N. S., Alzubaidi, L., Albahri, O. S., Alamoodi, A. H., Bai, J., Salhi, A., Santamaría, J., Ouyang, C., Gupta, A., Gu, Y., & Deveci, M. (2023). A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. *Information Fusion*, *96*, 156–191. https://doi.org/10.1016/j.inffus.2023.03.008

Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., Aldairem, A., Alrashed, M., bin Saleh, K., Badreldin, H. A., al Yami, M. S., al Harbi, S., & Albekairy, A. M. (2023). Revolutionizing healthcare: the role of artificial intelligence in clinical practice. In *BMC Medical Education* (Vol. 23, Issue 1). https://doi.org/10.1186/s12909-023-04698-z

Androulakis, E., & Fielder, C. (2024, April 2). *Artificial intelligence in ECG diagnostics - where are we now?* https://www.escardio.org/Councils/Council-for-Cardiology-Practice-(CCP)/Cardiopractice/artificial-intelligence-in-ecg-diagnostics-where-are-we-now

Apgar, C., & Prentice, R. (n.d.). *Therac-25*. The University of Texas at Austin. Retrieved July 4, 2024, from https://ethicsunwrapped.utexas.edu/wp-content/uploads/2022/10/Therac-25-1.pdf

Bajwa, J., Munir, U., Nori, A., & Williams, B. (2021). Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthcare Journal*, *8*(2). https://doi.org/10.7861/fhj.2021-0095

Ballesteros, J. A., Ramírez V., G. M., Moreira, F., Solano, A., & Pelaez, C. A. (2024). *Facial emotion recognition through artificial intelligence*. https://doi.org/10.3389/fcomp.2024.1359471

Bar-Ilan, J., Keenoy, K., Levene, M. and Yaari, E. (2009), Presentation bias is significant in determining user preference for search results—A user study. J. Am. Soc. Inf. Sci., 60: 135-149. https://doi.org/10.1002/asi.20941

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. Psychological Science in the Public Interest, 20(1), 1-68. https://doi.org/10.1177/1529100619832930

Bates, D. W., Cheng, H.-Y., Cheung, N. T., Jew, R., Mir, F., Tamblyn, R., & Li, Y.-C. (2022). "Improving smart medication management": an online expert discussion. *BMJ Health Care Inform*, *29*, 100540. https://doi.org/10.1136/bmjhci-2021-100540

Beauchamp, T. L., & Childress, J. F. (2019). *Principles of Biomedical Ethics* (8th ed.). Oxford University Press.

Bonmassar, G., Apollonio, F., Patil, P. G., Johnson, K. A., Nuf, D., W-j, N., F Dosenbach, N. U., Gordon, E. M., Welle, C. G., Wilkins, K. B., Bronte-Stewart, H. M., Voon, V., Morishita, T., Sakai, Y., Merner, A. R., Williamson, T., Horn, A., Gilron, ee, Gittis, A. H., … Wong, J. K. (2024). *Proceedings of the eeth Annual Deep Brain Stimulation Think Tank: pushing the forefront of neuromodulation with functional network mapping, biomarkers for adaptive DBS, bioethical dilemmas, AI-guided neuromodulation, and translational advancements*. https://doi.org/10.3389/fnhum.2024.1320806

Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., Roberts, A., Brown, T., Song, D., Erlingsson, U., Oprea, A., & Raffel, C. (2020). *Extracting Training Data from Large Language Models*.

Comissão Europeia. (2020). *LIVRO BRANCO sobre a inteligência artificial - Uma abordagem europeia virada para a excelência e a confiança*. (2020).

https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_pt.pdf.

Damaševičius, R., Bacanin, N., & Misra, S. (2023). *Actuator Networks Sensor and Review From Sensors to Safety: Internet of Emergency Services (IoES) for Emergency Response and Disaster Management*. https://doi.org/10.3390/jsan12030041

Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, *6*(2). https://doi.org/10.7861/futurehosp.6-2-94

Dembrower, K., Crippa, A., Colón, E., Eklund, M., & Strand, F. (2023). Artificial intelligence for breast cancer detection in screening mammography in Sweden: a prospective, population-based, paired-reader, non-inferiority study. *The Lancet Digital Health*, *5*(10), e703–e711. https://doi.org/10.1016/S2589-7500(23)00153-X

El-Assy, A. M., Amer, H. M., Ibrahim, H. M., & Mohamed, M. A. (2024). A novel CNN architecture for accurate early detection and classification of Alzheimer's disease using MRI data. *Scientific Reports*, *14*, 3463. https://doi.org/10.1038/s41598-024-53733-6

EU-U.S. Trade and Technology Council (2023). *EU-U.S. Terminology and Taxonomy for Artificial Intelligence - First Edition*.

EU-U.S. Trade and Technology Council (2024). *EU-U.S. Terminology and Taxonomy for Artificial Intelligence - Second Edition*.

Flynn, J. (2022). *Theory and Bioethics*. Obtido de Stanford Encyclopedia of Philosophy: https://plato.stanford.edu/entries/theory-bioethics/

Gordon, J.-S. (s.d.). *Bioethics*. Obtido em 10 de 05 de 2024, de Internet Encyclopedia of Philosophy: https://iep.utm.edu/bioethics/#H6

Greenfield, A. (2017). *Radical technologies: the design of everyday*. Brooklyn, NY: Verso.

Grupo Independente de Peritos de Alto Nível criado pela Comissão Europeia em 2018. (2018) *Uma Definição de IA: principais capacidades e disciplinas científicas*. Bruxelas: Comissão Europeia.

Hammouchi, H., Cherqi, O., Mezzour, G., Ghogho, M., & el Koutbi, M. (2019). Digging Deeper into Data Breaches: An Exploratory Data Analysis of Hacking Breaches Over Time. *Procedia Computer Science*, *151*, 1004–1009. https://doi.org/10.1016/J.PROCS.2019.04.141

Haroon, S., Corien, P., & Erik, S. (2023). *Mission AI: The New System Technology*. Den Haag: Springer Cham.

Hassan, A., Ahmad, S. G., Ullah Munir, E., Khan, I. A., & Ramzan, N. (2024). Predictive modelling and identification of key risk factors for stroke using machine learning. *Scientific Reports*, *14*, 11498. https://doi.org/10.1038/s41598-024-61665-4

Hasselberger, W., & Lott, M. (2023). "Where lies the grail? AI, common sense, and human practical intelligence." Phenomenology and the Cognitive Sciences. https://doi.org/10.1007/s11097-023-09942-x

Huat Goh, K., Wang, L., Yong Kwang Yeow, A., Poh, H., Li, K., Jie Lin Yeow, J., & Yu Heng Tan, G. (2021). Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. *Nature Communications*, *12*(711). https://doi.org/10.1038/s41467-021-20910-4

Jia, Y., Mcdermid, J., Lawton, T., & Habli, I. (2022). The Role of Explainability in Assuring Safety of Machine Learning in Healthcare. *IEEE Transactions on Emerging Topics in Computing*, *10*(4), 1746–1760. https://doi.org/10.1109/TETC.2022.3171314

Kandul, S., Micheli, V., Beck, J., Burri, T., Fleuret, F., Kneer, M., & Christen, M. (2023). Human control redressed: Comparing AI and human predictability in a real-effort task. *Computers in Human Behavior Reports*, *10*, 100290. https://doi.org/10.1016/J.CHBR.2023.100290

Kumar Panjiyar, A., Bhandari, S., Kumar Srivastava, R., Vashishta Rinkoo, A., Panjiyar, A., Songara, D., Sharma, A., Pareek, M., & Ranjan Singh, R. (2019). Evaluating the Feasibility of Rolling out Universal Hearing Screening for Infants in India using Sohum, an Artificial

Intelligence-driven Low Cost Innovative Diagnostic Solution. *J Child Dev Disord*, *5*(4), 11. http://childhood-developmental-disorders.imedpub.com/

Martínez-Nicolás, I., Llorente, T. E., Martínez-Sánchez, F., José, J., & Meilán, G. (2021). Ten Years of Research on Automatic Voice and Speech Analysis of People With Alzheimer's Disease and Mild Cognitive Impairment: A Systematic Review Article. *Front. Psychol*, *12*, 620251. https://doi.org/10.3389/fpsyg.2021.620251

Preetham, F. (2023, December). *Unpredictable Latent Errors in AI can be Catastrophic — Mathematical Explanation | Medium*. https://medium.com/autonomous-agents/unpredictable-latent-errors-in-ai-can-be-catastrophic-mathematical-explanation-95aa8f952d24

Raina MacIntyre, C., Chen, X., Kunasekaran, M., Quigley, A., Lim, S., Stone, H., Paik, H., Yao, L., Heslop, D., Wei, W., Sarmiento, I., & Gurdasani, D. (2023). Artificial intelligence in public health: the potential of epidemic early warning systems. *Journal of International Medical Research*, *51*(3), 1–18. https://doi.org/10.1177/03000605231159335

Reason J. (2000). Human error: models and management. *BMJ (Clinical research ed.)*, *320*(7237), 768–770. https://doi.org/10.1136/bmj.320.7237.768

Reddy S. (2023). Navigating the AI Revolution: The Case for Precise Regulation in Health Care. *Journal of medical Internet research*, *25*, e49989. https://doi.org/10.2196/49989

Schwettmann, S., Shaham, T. R., Materzynska, J., Chowdhury, N., Li, S., Andreas, J., Bau, D., & Torralba, A. (2023). *FIND: A Function Description Benchmark for Evaluating Interpretability* Methods.

Shaima, M., Nabi, N., Md Nasir Uddin Rana, Md Tanvir Islam, Estak Ahmed, Mazharul Islam Tusher, Mousumi Hasan Mukti, & Quazi Saad-Ul-Mosaher. (2024). Elon Musk's Neuralink Brain Chip: A Review on 'Brain-Reading' Device. *Journal of Computer Science and Technology Studies*, *6*(1), 200–203. https://doi.org/10.32996/jcsts.2024.6.1.22

Sterling, R. L. (2011). *Genetic Research among the Havasupai: A Cautionary Tale*. Virtual Mentor, 2(13), 113-117. doi: 10.1001/virtualmentor.2011.13.2.hlaw1-1102.

Varshney, K. R. (2016). *Engineering Safety in Machine Learning*.

Wang, Y., & Kosinski, M. (2018). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *Journal of Personality and Social Psychology*, *114*(2), 246–257. https://doi.org/10.1037/PSPA0000098

Warlow, C. (2005). *Over-regulation of clinical research: A threat to public health*. Clinical medicine (5), 33-38.

Yamashita, R., Nishio, M., Do, R. K., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, *9*(4), 611–629. https://doi.org/10.1007/s13244-018-0639-9