

PERSPECTIVAS E DESAFIOS ACERCA DA APLICAÇÃO DA INTELIGÊNCIA ARTIFICIAL COMO SISTEMA DE AUXÍLIO À TOMADA DE DECISÃO CLÍNICA
PERSPECTIVES AND CHALLENGES ABOUT THE APPLICATION OF ARTIFICIAL INTELLIGENCE AS A SYSTEM TO ASSIST CLINICAL DECISION-MAKING

Vinicius Vieira Queiroga¹, Carla Islene Holanda Moreira Coelho², Rafaela de Oliveira Nóbrega³ e José Guilherme Ferreira Marques Galvão⁴

ARTIGO

Recebido:

15/03/2023

Aprovado:

12/04/2023

Palavras-chave:

Inteligência Artificial. Medicina. Saúde Pública. Bioética.

RESUMO

A inteligência artificial (IA), consiste na junção de múltiplas áreas da ciência a fim de obtenção de agentes/máquinas inteligentes capazes de exercerem funções de aprendizado autônomo, reconhecimento de padrões, tomada de decisões e realização de ações. O aprimoramento dessas tecnologias trará revoluções ainda mais significativas, principalmente nas esferas de trabalho, incluindo a área de saúde como uma das que mais irá mudar. Em meio à era da informação, durante a gradativa ascensão da IA, debate-se as perspectivas futuras acerca da aplicação de tecnologias de auxílio à tomada de decisão clínica no cotidiano do profissional de saúde, tocando nas perspectivas éticas, legais e da própria tecnologia. Analisar a aplicabilidade da Inteligência Artificial no sistema de saúde e os desafios inerentes, sob a perspectiva da própria tecnologia, da bioética e da legislação. Trata-se de uma revisão integrativa da literatura, nas bases de dados PubMed e BVS, com a utilização dos Descritores em Ciências da Saúde (DeCS) Inteligência Artificial, Medicina, Saúde Pública e Bioética, associados ao operador *booleano* AND. Por meio da pesquisa, são buscados artigos entre os anos de 2017 a 2021, em português e inglês, cujos artigos compreendem os critérios de inclusão e exclusão propostos pelo estudo, bem como se enquadram dentro do objetivo traçado, a fim de se obter um apanhado perspectivo sobre o futuro da IA na saúde, bem como suas possíveis aplicações e resoluções para os desafios inerentes a essa tecnologia no contexto atual no ponto de vista bioético, legislativo e da composição da mesma.

ABSTRACT

Artificial intelligence (AI) consists of the combination of multiple areas of science in order to obtain intelligent agents/machines capable of performing autonomous learning, pattern recognition, decision-making and actions. The improvement of these technologies will bring even more significant revolutions, mainly in the spheres of work, including the health area as one of the areas that will change the most. In the midst of the information age, during the gradual rise of AI, future perspectives are being debated regarding the application of technologies to aid clinical decision-making in the daily life of health professionals, touching on ethical, legal and technology perspectives. To analyze the applicability of Artificial Intelligence in the health system and the inherent challenges, from the perspective of the technology itself, bioethics and legislation. This is an integrative literature review, in the PubMed and VHL databases, using the Descriptors in Health Sciences (DeCS) Artificial Intelligence, Medicine, Public Health and Bioethics, associated with the Boolean AND operator. Through the search, articles are sought between the years 2017 to 2021, in Portuguese and English, whose articles comprise the inclusion and exclusion criteria proposed by the study, as well as fit within the objective outlined, in order to obtain an overview perspective on the future of AI in health, as well as its possible applications and resolutions to the challenges inherent to this technology in the current context from the bioethical, legislative and composition point of view.

¹Graduando em Medicina pelo Centro Universitário Santa Maria;

²Docente do Centro Universitário Santa Maria;

³Docente do Centro Universitário Santa Maria;

⁴Docente do Centro Universitário Santa Maria.

1. INTRODUÇÃO

A Inteligência Artificial (IA) pode ser abordada como a composição de uma variedade de tecnologias, as quais se combinam para exercerem configurações que se assemelham à inteligência humana em tarefas de pensamento, aprendizado, reconhecimento, resolução de problemas, tomada de decisões, bem como o planejamento e realização de ações. Por meio do envolvimento dos mais diversos campos do conhecimento, desde a lógica, a estatística, até a psicologia cognitiva, neurociências e computação, forma-se essa unidade que atualmente já tem espaço significativo no cotidiano humano e vem crescendo exponencialmente com o decorrer do tempo (RUSSELL; NORVIG, 2004).

Tendo como marco de seu surgimento, após a segunda guerra mundial, na década de 50, o trabalho de Warren Macculloch e Walter Pitts (1943), que combinaram conhecimentos de fisiologia com a teoria computacional de Alan Turing para propor um modelo de "neurônios artificiais", deu-se início a esse campo da ciência que obteve uma visão mais geral a partir do artigo de Turing em 1950 denominado "*Computing Machinery and Intelligency*", no qual foi apresentado o Teste de Turing, que consiste em avaliar a equivalência do comportamento inteligente de uma máquina com o do homem, alegando que, para o computador inteligente ideal, o avaliador do teste não conseguiria distinguir se as respostas seriam realizadas ou não por um ser humano. Posteriormente, seguindo a linha para obtenção de um sistema autônomo na realização de atividades que exigem lógica, raciocínio e organização, em 1969, na Universidade de Stanford, foi criado o DENDRAL, primeiro programa inteligente capaz de tomar decisões autônomas. A partir desse ponto, cerca de 20 anos depois, a IA passou a entrar no mercado e na indústria por meio dos softwares comerciais especializados, tornando possível o gerenciamento automático dos mais diversos departamentos desse setor, iniciando pelo de pedidos, o que gerou um impacto positivo na economia dos principais conglomerados da época, que passaram a ser constituídos por grupos de inteligência artificial que usavam e desenvolviam essa tecnologia (GOMES, 2010).

Com o passar do tempo, eclodiram de forma crescente os estudos e avanços na área, que foi revolucionada rapidamente e atualmente se encontra com suas bases e fundamentos sólidos e bem definidos, permitindo um vislumbre do potencial que será alcançado. Estima-se que nas próximas décadas a IA irá revolucionar ainda mais a maioria dos setores da sociedade, iniciando uma nova era, na qual a vida humana se tornará completamente diferente da que já vivemos. Podemos encontrar essa tecnologia em nossa vivência, principalmente nas ferramentas eletrônicas que portamos e buscamos auxílio para obtenção e organização de

informações. Os mecanismos de busca, as preferências com as quais escolhemos configurar nossos smartphones, os tradutores, as agendas, os aplicativos de notícias, as redes sociais, os sites, entre outros, apresentam algoritmos em seu software que projetam o caminho para o processamento e detecção de padrões de uso, com os quais aprendem a se adaptar ao usuário, aprimorando cada vez mais o conteúdo e sugestões fornecidos, dependendo do interesse do mesmo. Essa é apenas uma das várias aplicações práticas da IA com as quais já convivemos. Sua expansão impactou nossa rotina, otimizando nossas atividades no que se refere ao reconhecimento, classificação e rotulagem de imagens estáticas, melhorias de desempenho de estratégia de negociação algorítmica, processamento eficiente e escalonável de dados de pacientes, manutenção preditiva, detecção e classificação de objetos, bem como distribuição de conteúdo. Graças à IA, poderemos obter previsões acerca de fatores importantes em nossas vidas, fomentando impacto positivo nas áreas da saúde, educação, do trabalho e relacionamentos interpessoais. Da mesma maneira, passará a ser possível a realização de atividades e metas até então inalcançáveis à humanidade (KULKARNI, 2020).

Tudo isso só é possível devido ao crescimento exponencial da capacidade computacional, que permite um aumento cada vez maior no potencial de armazenamento e processamento de dados, requisitos básicos para a consolidação dessa tecnologia. A partir dessa base, para ser criado um modelo de IA, há a necessidade da combinação de algoritmos de *machine learning* com bancos de dados extensos e específicos. Assim, quanto maior a complexidade do algoritmo e do objetivo alcançado, maior a demanda por análise e computação de dados, conseqüentemente por potência para esses atributos. Na era da informação representada pelo período atual, estabeleceu-se que nunca antes fora armazenado tantos dados e informações sobre tudo. Diante de bancos de dados, agora mais do que nunca, suficientemente extensos, a IA encontra terreno fértil para seu desenvolvimento nos mais diversos segmentos (GOMES, 2010).

Para um futuro próximo, aguarda-se grande influência e mudança no cotidiano de profissionais da maioria dos setores sociais e do mercado. Dentre esses, a área da saúde será uma das mais assistidas por essa tecnologia, sofrendo impactos que poderão mudar de forma integral a rotina dos profissionais e a forma como irão atuar. De forma paralela, já é realidade e com grande importância sua forte presença na gestão de dados nos sistemas de saúde, tanto no âmbito privado, quanto no público por meio do SUS no Brasil, por exemplo, no qual encontram-se as iniciativas como o projeto e-SUS na Atenção Básica, que visa reestruturar as informações da atenção básica em nível nacional no Sistema de Informação da Atenção

Básica (Siab), além disso, há o Conecte SUS, que é uma ação governamental voltada especialmente para a implementação da IA na gestão de saúde pública brasileira. Todos esses programas fazem parte da Rede Nacional de Dados em Saúde (RNDS) e possuem o intuito de facilitar e organizar o acesso a dados estatísticos e informações de saúde a qualquer profissional da área. Outrossim, tendo em vista o que já se encontra presente na realidade da saúde, sem se referir aos sistemas de instituições específicas que envolvem a gestão de prontuários, entre outros dados, espera-se que no futuro esse recurso se aprofunde ainda mais e atinja a esfera de tomada de decisão profissional, participando do dia a dia das consultas médicas, bem como da aferição e disposição de parâmetros dos pacientes (LEMES; LEMOS, 2020).

Dentro desse contexto, estudos já evidenciam e comprovam o potencial efetivo que a IA possui como ferramenta de auxílio à tomada de decisão clínica dentro dos mais diversos departamentos da medicina, desde os sistemas de triagem, ao diagnóstico e prognóstico com predições sobre o paciente, como também fazendo parte da área cirúrgica e de procedimentos por meio da robótica e dos sistemas de visualização. No entanto, juntamente com a característica promissora, promovendo excelentes melhorias no desempenho humano em diversas atividades da saúde, por ser uma nova ferramenta, também deve-se pensar o seu potencial para o mau uso, em como esse instrumento pode, da mesma forma, ser prejudicial. A partir disso, tange-se a perspectiva ética do assunto, envolvendo também os direitos e deveres tanto do profissional, quanto do paciente, bem como o contexto por trás da gestão de dados e privacidade, uma vez que os bancos de dados utilizados são alimentados com informações pessoais e muitas vezes sensíveis de cada pessoa. Da mesma maneira, outro desafio da IA na saúde se encontra na forma como a tomada de decisão advinda de um sistema de assistência à decisão clínica será esclarecida tanto para o profissional como para o paciente. A complexidade dos algoritmos utilizados para esses fins impede ou impossibilita uma explicação compreensível direcionada ao público leigo, urgindo meios de viabilizarem uma solução para uma tecnologia que está tão próxima de ser colocada em prática (WANG; PREININGER, 2019).

Em atendimento aos objetivos propostos, esta pesquisa fundamenta-se nos principais estudos da área de inteligência artificial e sua relação com a saúde, mais especificamente a medicina, possuindo, portanto, o aspecto de revisão integrativa da literatura, baseada em artigos e livros que abordam o referido tema.

Tendo em vista a importância do tema para o futuro do profissional de saúde, bem como das questões levantadas que interferem também no futuro paciente, o presente estudo propõe versar sobre a aplicabilidade da IA como sistema de auxílio na prática clínica, abordando os desafios para a sua implementação nas perspectivas éticas, legais e da própria tecnologia estudada, explanando os possíveis contextos e soluções futuras, incluindo o contexto específico do Brasil.

2. DESENVOLVIMENTO

2.1 Conceito e Funcionamento da Inteligência Artificial

Para conceituar a IA como ciência, RUSSEL e NORVIG (2004) a dividiram em quatro partes, cada uma com suas definições:

- **Sistemas que pensam como seres humanos:** “Consiste no esforço para fazer os computadores e máquinas pensarem como mentes, no sentido total e literal” (HAUGELAND, 1985)
- **Sistemas que pensam racionalmente:** “O estudo das faculdades mentais pelo seu uso de modelos computacionais” (CHARNIAK; MCDERMOTT, 1985)
- **Sistema que atuam como seres humanos:** “A arte de criar máquinas que executam funções que exigem inteligência quando são executadas por pessoas” (KURZWEIL, 1990)
- **Sistemas que atuam racionalmente:** “A inteligência computacional é o estudo do projeto de agentes inteligentes” (POOLE et al., 1998)

Os dois primeiros conceitos referem-se ao pensamento e raciocínio especificamente, enquanto os dois últimos ao comportamento. O primeiro e o terceiro pensamentos aferem a fidelidade com o desempenho humano, bem como o segundo e o quarto, mensuram o sucesso comparado a uma definição ideal de inteligência/racionalidade, sendo todos esses âmbitos estudados em conjunto historicamente no desenvolvimento dessa tecnologia, se completando e se corrigindo com diferentes abordagens. Ao centrar-se no ser humano, há uma prevalência da ciência empírica, abrangendo hipóteses e confirmações experimentais. Por outro lado, voltando-se para o pensamento e a racionalidade, há uma predominância das abordagens matemáticas e de engenharia (RUSSELL; NORVIG, 2004).

Dentro desse campo da ciência da computação, diversas áreas foram desenvolvidas, entre elas, a dos sistemas especialistas (que logo foram deixados de lado para dar lugar ao machine-learning), a robótica, os sistemas visuais, processamento de linguagem natural e planejamento e logística. Cada uma contribuindo de maneira individual ou em conjunto para a formação das mais diversas aplicações em inteligência artificial, incluindo a área da saúde (GOMES, 2010).

Os sistemas de IA atualmente fazem parte do espectro da Inteligência Artificial Fraca ou “estreita”, que engloba o conceito de pensamento simulado e já nos permitem desfrutar da disponibilidade de dispositivos como sensores, robôs, e aplicativos de software de auxílio à tomada de decisões inteligentes. Futuramente, prevê-se que, com os avanços nessa área, alcançaremos um patamar suficiente para usufruir da Inteligência Artificial Forte ou “Geral” (AGI), a qual engloba o conceito de pensamento real, no qual está atrelado à criação de uma espécie de “consciência artificial”, promovendo à possibilidade de autonomia da máquina no quesito pensamento, aprendizado e planejamento de ação. Apesar de ser algo para um futuro muito distante, especialista na área já preveem outra revolução no futuro (HOWARD, 2019).

Por outro lado, em se tratando do que atualmente é tangível para essa tecnologia, o meio mais utilizado para a consolidação da mesma é através de machine-learning (ML), uma subdisciplina da IA que permite a execução de vários métodos em computação a fim de se obter um insight cognitivo, previsões e apoiar tomada de decisões de uma máquina. Por meio dessa ferramenta associada a táticas de software que envolvem os algoritmos, os computadores se tornaram capazes de aprender e criarem suas próprias regras (HOWARD, 2019). Temos, portanto, como a base de execução do ML, os algoritmos, que possuem um conceito ou definição abstrata e incerta devido à sua elevada variedade e fins aos quais são utilizados. Gurevich Y. (2012), no entanto, entende os mesmos como “equações recursivas” que geram uma instrução computacional específica para determinado fim. Dessa forma, combinados com uma base de dados e programação objetiva, constrói-se uma tecnologia baseada em IA.

A recente era da informação, com a presente explosão no compartilhamento e armazenamento de dados digitais, que vão desde imagens, textos e transações, a dados de detecção humanos e ambientais, em grandes bancos (big-data), permitiu, além de um melhor manuseio desses componentes, a criação de algoritmos e sistemas mais complexos, capazes de elevar o nível das tecnologias em IA. Como consequência, além disso, houve um barateamento no custo para o processamento de dados, uma vez que estes passaram a ser processados a partir de nuvens e não nos locais físicos nos quais os bancos se encontram,

além de tornar possível “treinar” máquinas a realizarem tarefas para as quais não foram explicitamente programadas.

Tecnologias como o reconhecimento de imagens e voz, sensores, dispositivos de auxílio à tomada de decisão, sistemas de pesquisa, etc, são construídos por meios de vários métodos de ML diferentes, são os chamados tipos de aprendizado. São eles:

- **Aprendizado Supervisionado:** Consiste em um modelo que é construído por meio de “treinamento” supervisionado de seu conjunto, no qual os dados são rotulados, dando exemplo de como são as entradas, bem como devem ser as saídas. Pode-se fazer uma analogia com um professor dando um problema a um aluno e ao mesmo tempo mostrando suas resoluções. A partir desses rótulos, o aluno (modelo) busca com autonomia os outros resultados. (ZHANG, 2020)
- **Aprendizado Não-supervisionado:** Nesse método não são treinados os rótulos e, por meio dos algoritmos, a máquina ou o programa buscam as soluções de forma independente, sem terem tido antes contato com as mesmas. Se assemelham a dar a um aluno um conjunto de padrões, pedindo que o mesmo descubra os motivos subjacentes que o geraram. (ZHANG, 2020)
- **Aprendizado Semi-supervisionado:** É um meio termo entre o aprendizado supervisionado e o não-supervisionado. Dessa forma, é ofertado ao modelo uma pequena parte de dados rotulada e uma grande parte de dados sem rótulos. Dependendo da forma como ocorre o aprendizado o mesmo ainda pode ser dividido em self-training, co-training e active learning (ZHANG, 2020). São utilizados quando há uma vasta quantidade de dados não etiquetados em forma de imagem, voz ou textos, por possuir um custo menor e mais praticidade quando comparado com o aprendizado supervisionado.
- **Aprendizado por Reforço:** Nesse modelo são fornecidos dados de feedback na forma de recompensas e punições com o intuito de treinar o agente de inteligência artificial em um ambiente dinâmico, sendo útil para aprimorar o desempenho na tarefa aprendida (ZHANG, 2020). É como se fosse uma adaptação da teoria psicológica do comportamento, voltando-a para as máquinas. Esse tipo de tecnologia pode ser usado em redes neurais, no treinamento de veículos autônomos e robôs, por exemplo (HOWARD, 2019).

Ademais, outros instrumentos importantes englobados pelo ML são as Redes Neurais, as quais se baseiam na imitação de conjuntos de neurônios para a obtenção de IA. Por meio de camadas matemáticas, de forma análoga às células do sistema nervoso em organização, o peso das probabilidades dos dados vai sendo reajustado e diminuído até que a saída desejada (o resultado) seja obtida, possibilitando o aprendizado a partir de dados. De maneira a utilizar as Redes Neurais, o Deep Learning se tornou um subconjunto dessa tecnologia que está começando a ser aplicado na prática e tendo resultados promissores. Por meio de várias camadas de processamento (neurônios computacionais) interconectadas, o dado entra e passa por cada uma, onde são atribuídos diferentes pesos matemáticos para os diferentes aspectos do dado, que é testado diversas vezes até a saída, na qual ocorre a reunião de todos os fatores obtidos nas camadas, o processamento final e o resultado. Caso a saída esteja incorreta, o erro é registrado e os neurônios/camadas são reajustadas para um novo processamento, diminuindo a margem de erro cada vez mais. Algoritmos utilizados nesse método obteram sucesso significativo com a utilização de imagens para reconhecimento e tomada de decisões (HOWARD, 2019).

2.2 Aplicações da Inteligência Artificial na Saúde

A IA encontra-se na rota para revolucionar a área da saúde e principalmente a prática do profissional médico. Nessa área, estudos envolvendo o desenvolvimento e aprimoramento de tecnologias de ML e Deep-learning, prometem resultados promissores que beneficiarão vários segmentos da medicina. Principalmente, tomando como ponto de vista o patamar atual desse campo da ciência, os que requerem capacidade de reconhecimento de padrões visuais para auxílio à tomada de decisões, tangendo a radiologia, oftalmologia, patologia e dermatologia como as mais estudadas até o momento, tendo também como motivo a disponibilidade de bancos de dados o suficiente para o treinamento de programas (KULKARNI et al, 2020).

Para a radiologia, o exame mais utilizado como material de estudo no mundo é a radiografia de tórax. Isso se dá pelo fato de ser muito comum e dispor de um vasto banco de dados disponibilizado. Atualmente tem-se o ChestX-ray14 como um dos maiores conjuntos de informações e imagens de radiografia de tórax, disponibilizado gratuitamente, podendo ser aberto em qualquer aparelho eletrônico com acesso à internet. Dispondo de 112.120 radiografias de 30.805 pacientes únicos, as quais foram categorizadas em 14 condições clínicas distintas. Além disso, estudos em outros bancos de dados como o da Sociedade Norte

Americana de Radiologia (RSNA), quando dispostos a testes, os algoritmos utilizados com o intuito de se obter uma percepção e análise da imagem radiológica a fim de identificar a presença de fratura de punho em pacientes, obtiveram 98,1% de sensibilidade e especificidade, evidenciando a importância futura desses mecanismos no auxílio ao sistema de triagem no departamento de emergência, por exemplo (VOTH et al, 2017). De maneira paralela, exames de imagem mais complexos como as ressonâncias magnéticas e tomografias computadorizadas passam a ser cada vez mais utilizados para compor softwares auxílio à tomada de decisão a base de deep-learning.

No campo da oftalmologia, em busca da prevenção de cegueira por retinopatia diabética em países de baixa renda, a IA foi testada em modelos de análise à decisão clínica, a fim de permitir o diagnóstico precoce do paciente. A possibilidade de obtenção de imagens digitais da retina facilitou o manuseio e a pesquisa, que resultou em uma sensibilidade de 87,2% e especificidade de 90,7% a partir de 819 imagens analisadas pelo programa. (ABRÀMOFF, Michael D. et al, 2018) Por outro lado, no campo da patologia, a IA vem sendo aplicada a otimização de microscópios e da visualização de suas amostras biológicas, enquanto que na dermatologia está voltada para a detecção precoce de lesões e na predição de uma possível malignidade advinda da mesma (KULKARNI et al, 2020).

2.3 Desafios Acerca da Implantação da Inteligência Artificial

É fundamental frisar que, apesar dos avanços, diversos desafios e problemas continuam e surgem do ponto de vista da consolidação da própria tecnologia, tendo algum desses problemas influências em questões éticas que envolvem a dignidade e a privacidade do paciente, até mesmo o mesmo em risco de vida, dependendo do contexto. Dentre os principais obstáculos, estão a falta de explicabilidade (black box) por parte dos algoritmos complexos, o overfitting e a regulação dessa tecnologia e sua aprovação para uso na prática, a qual é dificultada, entre outros fatores, como consequência dos dois primeiros citados.

A problemática da black box consiste na incapacidade que os algoritmos de deep-learning possuem em deixar explícito como chegaram às suas conclusões. Ao chegar a um resultado ou decisão, tradicionalmente é impossível determinar quais recursos foram utilizados para tal, assim como foram analisados e até mesmo atingido determinado fim ao invés de outro. Isso ocorre com a maioria dos algoritmos dessa tecnologia, repercutindo para a não confiabilidade em relação a seus resultados. Em uma situação na qual um sistema de

tomada de decisões radiológicas trabalhasse, por exemplo, não seria possível afirmar quais imagens, e parâmetros de cada uma, o algoritmo tomou como base para chegar à decisão final. Alcançar a explicabilidade desse processo será um dos caminhos para estabelecer a confiança tanto do profissional de saúde quanto do paciente, bem como de órgãos reguladores, reforçando a IA para uma maior aplicabilidade na prática (KULKARNI et al, 2020).

O Overfitting (sobreajuste), por outro lado, ocorre quando um algoritmo treinado previamente se limita a determinados conjuntos de dados em oposição ao que traria a solução do problema a ser resolvido. Essa limitação se deve principalmente a práticas de treinamento excessivas dos algoritmos (overtraining) em um conjunto de dados específico. Fatores como o tamanho do conjunto de dados e a forma como esses são distribuídos dentro dele determinam a ocorrência desse viés. Quando há uma divergência muito grande entre os dados do conjunto provenientes dos momentos de treinamento e de teste, ou se foram adquiridos com parâmetros, ou até equipamentos diferentes, também se percebe sobreajuste. Tudo isso torna a efetividade da tecnologia mais baixa, gerando uma maior tendência ao erro. Por esses motivos, manobras específicas podem ser viáveis para que se possa driblar esse acontecimento e treinar o algoritmo na medida certa, seja apresentando novos rótulos ou mudando a complexidade da rede (parâmetros ou estruturas) (NAGENDRAN et al, 2020).

Tanto a ausência de explicabilidade, quanto a possibilidade de overfitting por parte de um sistema inteligente, são problemas que atingem negativamente a perspectiva bioética da consolidação da IA na prática clínica, bem como atrasa sua regulamentação por meio dos órgãos governamentais. Dessa forma, à luz dos princípios bioéticos da autonomia, beneficência, não maleficência e justiça (BEUCHAMP, 2001), pode-se inferir implicações tanto para os pacientes quanto para os profissionais de saúde. No tocante à autonomia do paciente, esta tem como um dos principais artifícios o consentimento informado, o qual, por meio de informações extensas, porém, compreensíveis, informa os riscos do procedimento e preserva a independência do indivíduo em permitir ou não determinado procedimento. No entanto, na ausência de um esclarecimento digno sobre o trajeto para obtenção de resultados por meio das tecnologias de IA na prática clínica, as bases do consentimento informado estariam quebradas, impactando a autonomia do paciente, como também a relação médico-paciente, violando o cumprimento de recomendações clínicas. Portanto, são imprescindíveis padrões éticos e explicativos apropriados para preservar as funções de autonomia do consentimento informado, assim como a confiança do paciente nos procedimentos ofertados (AMANN et al, 2020).

Da mesma maneira, esse raciocínio volta-se para os médicos quando se trata dos princípios da beneficência e não-maleficência, uma vez que é fundamental que o mesmo possua conhecimento pleno sobre a ferramenta com a qual atuará para garantir máximo benefício ao paciente, fazendo-se necessária a reflexão acerca da saída do sistema. Sendo assim, a explicabilidade de grande importância para a tomada de decisões clínicas confiáveis, para que o profissional não caia no viés da automação e julgue de forma crítica e adequada o quanto será relevante o auxílio dos sistemas de IA na tomada de uma decisão específica. Outrossim, é dever do médico também não prejudicar os pacientes, seja de forma incidental ou não. Nesse contexto, seria o desempenho e eficácia isolados o suficiente para garantir a confiabilidade nas tomadas de decisões específicas advindas de um computador inteligente? Não há como excluir a possibilidade de limitações mediante uma ferramenta tão recente e que possui previsão de melhores esclarecimentos no futuro (AMANN et al, 2020).

Considerando o princípio da justiça, o mesmo prega o igual acesso aos benefícios do progresso médico sem a discriminação de qualquer indivíduo ou grupo social. Entretanto, já foi evidenciado que alguns sistemas de IA ferem esse princípio, tendo como exemplo um algoritmo que discrimina levando em consideração dados relacionados à cor para quem mais se beneficiaria de um tratamento cirúrgico específico. Esse panorama leva à percepção de que sem a explicabilidade, não seria possível evitar esse tipo de acontecimento.

Do ponto de vista legal, essa tecnologia deve ser regulamentada tendo como base os princípios éticos e legais vigentes em cada nação/governo. Nos Estados Unidos, órgãos como o Food and Drug Administration (FDA) estão exigindo que explicabilidade e transparência para que dispositivos inteligentes tenham seu uso aprovado. A OMS, por sua vez, na mesma linha de raciocínio, exige também que haja explicações e recomenda que os órgãos governamentais e instituições sigam 6 princípios éticos (proteção da autonomia humana; promoção de bem estar, segurança humana e interesse público; garantia de transparência, explicabilidade e inteligibilidade; promoção de responsabilidade e prestação de contas; garantia de inclusão e equidade; e promoção de IA responsiva e sustentável) (WORLD HEALTH ORGANIZATION et al, 2021).

3. CONSIDERAÇÕES FINAIS

Diante do exposto, percebe-se que acerca da privacidade dos pacientes por meio dos dados manipulados para a obtenção de resultados, o que vai de encontro com os regulamentos

de proteção de dados e terá o envolvimento do direito contratual e da responsabilidade civil para que se tenha uma resposta final mais cedo ou mais tarde.

Por fim, o conflito entre os benefícios, a inovação e a falta de determinadas informações tornam necessária uma construção cuidadosa da consolidação legal dessa tecnologia no futuro.

REFERÊNCIAS

ABRÀMOFF, M. D. et al. **Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices**. NPJ digital medicine, v. 1, n. 1, p. 1-8, 2018.

AMANN, J.; BLASIMME, A.; VAYENA, E.; FREY, D.; MADAI, V. I. **Explainability for artificial intelligence in healthcare: a multidisciplinary perspective**. BMC Medical Informatics And Decision Making, [S.L.], v. 20, n. 1, p. 1-9, 30 nov. 2020. Springer Science and Business Media LLC. <http://dx.doi.org/10.1186/s12911-020-01332-6>.

BEAUCHAMP, T. L. et al. **Principles of biomedical ethics**. Oxford University Press, USA, 2001.

GOMES, D. dos S. **Inteligência Artificial: conceitos e aplicações**. Olhar Científico. v1, n. 2, p. 234-246, 2010.

GUREVICH, Y. **What Is an Algorithm?** In: Bieliková M, Friedrich G, Gottlob G, Katzenbeisser S, Turán G, organizadores. SOFSEM 2012: Theory and Practice of Computer Science. Lecture Notes in Computer Science, vol 7147. Springer, Berlin, heidelberg: Springer-Verlag Berlin Heidelberg; 2012. p. 31-42.

HOWARD, J. **Artificial intelligence: implications for the future of work**. American Journal Of Industrial Medicine, [S.L.], v. 62, n. 11, p. 917-926, 22 ago. 2019. Wiley. <http://dx.doi.org/10.1002/ajim.23037>.

KULKARNI, S. et al. **Artificial Intelligence in Medicine: where are we now?**. Academic Radiology, [S.L.], v. 27, n. 1, p. 62-70, jan. 2020. Elsevier BV. <http://dx.doi.org/10.1016/j.acra.2019.10.001>.

LEMES, M. M.; LEMOS, A. N. L. E. **O uso da inteligência artificial na saúde pela Administração Pública brasileira**. Cadernos Ibero-Americanos de Direito Sanitário, [S.L.], v. 9, n. 3, p. 166-182, 29 set. 2020. Cadernos Ibero-Americanos de Direito Sanitario. <http://dx.doi.org/10.17566/ciads.v9i3.684>.

MOSCHOVAKIS, Y. N. **What is an algorithm?** In: Mathematics unlimited—2001 and beyond. Springer, Berlin, Heidelberg, 2001. p. 919-936.

NAGENDRAN, M. et al. **Artificial intelligence versus clinicians**: systematic review of design, reporting standards, and claims of deep learning studies. *Bmj*, [S.L.], 25 mar. 2020. BMJ. <http://dx.doi.org/10.1136/bmj.m689>.

ROUHIAINEN, L. **Inteligencia artificial**. Madrid: Alienta Editorial, 2018.

VERGARA, S. C. **Métodos de pesquisa em administração**. São Paulo: Atlas, 2005.

VOTH, M.; LUSTENBERGER, T.; AUNER, B.; FRANK, J.; MARZI, I. **What injuries should we expect in the emergency room?** *Injury*, [S.L.], v. 48, n. 10, p. 2119-2124, out. 2017. Elsevier BV. <http://dx.doi.org/10.1016/j.injury.2017.07.027>.

WORLD HEALTH ORGANIZATION et al. **Ethics and governance of artificial intelligence for health**: WHO guidance. 2021

WANG, F.; PREININGER, A. **AI in Health**: state of the art, challenges, and future directions. *Yearbook Of Medical Informatics*, [S.L.], v. 28, n. 01, p. 016-026, ago. 2019. Georg Thieme Verlag KG. <http://dx.doi.org/10.1055/s-0039-1677908>

ZHANG, X.-D. M. I. **A Matrix Algebra Approach to Artificial Intelligence**. Springer, Singapore, 2020. p. 223-440.