

Supervisão ética em sistemas inteligentes de apoio à saúde

Ricardo Rheingantz Abuchaim^{1,2}, Daniel Brito de Araujo²

1. Advocacia-Geral da União, Pelotas/RS, Brasil. 2. Universidade Federal de Pelotas, Pelotas/RS, Brasil.

Resumo

O artigo tem como objetivo propor diretrizes éticas e um modelo de avaliação para o uso de inteligência artificial em diagnósticos médicos. Realiza-se análise crítica da literatura bioética e normativa, fundamentada nos princípios da autonomia, beneficência, não maleficência e justiça. Com base nessa análise, desenvolve-se uma hierarquia dos principais desafios éticos envolvidos, que abrange desde a qualidade e a governança dos dados até a sustentabilidade ética das tecnologias empregadas. Propõe-se, ainda, um modelo de avaliação baseado em métricas objetivas, voltado ao monitoramento contínuo da aplicação da inteligência artificial na prática clínica. Conclui-se que o equilíbrio entre inovação tecnológica e humanização do cuidado é essencial, cabendo às associações médicas um papel regulador estratégico na preservação dos valores fundamentais da medicina.

Palavras-chave: Inteligência artificial. Ética. Ética médica. Sistemas inteligentes.

Resumen

Supervisión ética en sistemas inteligentes de apoyo a la salud

El artículo tiene como objetivo proponer directrices éticas y un modelo de evaluación para el uso de la inteligencia artificial en diagnósticos médicos. Un análisis crítico de la literatura bioética y normativa, basado en los principios de autonomía, beneficencia, no maleficencia y justicia fue realizado. A partir de este análisis, se desarrolla una jerarquía de los principales retos éticos implicados, que abarca desde la calidad y la gobernanza de los datos hasta la sostenibilidad ética de las tecnologías empleadas. Además, se propone un modelo de evaluación basado en métricas objetivas, orientado al seguimiento continuo de la aplicación de la inteligencia artificial en la práctica clínica. Se concluye que el equilibrio entre la innovación tecnológica y la humanización de la atención es esencial, y que las asociaciones médicas tienen un papel regulador estratégico en la preservación de los valores fundamentales de la medicina.

Palabras clave: Inteligencia artificial. Ética. Ética médica. Sistemas inteligentes.

Abstract

Ethical monitoring in intelligent health support systems

This article aims to propose ethical guidelines and an evaluation model for the use of artificial intelligence in medical diagnoses. A critical analysis of bioethical and normative literature, based on the principles of autonomy, beneficence, non-maleficence, and justice was conducted. From this analysis, a hierarchy of the main ethical challenges involved is developed, ranging from data quality and governance to the ethical sustainability of the technologies employed. An evaluation model based on objective metrics is also proposed, aimed at the continuous monitoring of the application of artificial intelligence in clinical practice. It concludes that the balance between technological innovation and humanization of care is essential, with medical associations playing a strategic regulatory role in preserving the fundamental values of medicine.

Keywords: Artificial intelligence. Ethics. Ethics, medical. Intelligent systems.

Declararam não haver conflito de interesse.

A inteligência artificial (IA) revoluciona o setor da saúde por sua capacidade transformadora em diagnósticos médicos. Sua implementação proporciona avanços significativos em precisão diagnóstica, otimização de custos e aprimoramento dos resultados clínicos¹. Por meio de algoritmos avançados e técnicas de *machine learning*, os sistemas processam volumes expressivos de dados médicos, incluindo análise de imagens e interpretação de resultados laboratoriais, estabelecendo novos paradigmas na medicina diagnóstica.

O quadro tecnológico da IA na saúde demanda infraestrutura robusta de governança para garantir a segurança e a eficácia dos sistemas em ambiente clínico. A arquitetura dos algoritmos deve contemplar rigorosos protocolos de validação para mitigar riscos de diagnósticos imprecisos e consequentes impactos adversos no tratamento². A gestão da segurança da informação assume papel crítico, dada a sensibilidade dos dados de saúde processados, com destaque para a mitigação de vieses algorítmicos que possam perpetuar disparidades no atendimento entre diferentes grupos demográficos³. A introdução de metodologias de auditoria e validação contínua dos modelos é essencial para assegurar equidade nos diagnósticos baseados em IA.

O paradigma do consentimento informado requer adaptação ao contexto digital de modo a garantir transparência ao funcionamento dos algoritmos e seu papel no processo decisório⁴. A documentação técnica deve explicitar as capacidades e limitações dos sistemas, a fim de permitir adequada compreensão por parte de pacientes e profissionais de saúde.

A governança dos sistemas de IA demanda definição clara de tarefas e mecanismos de responsabilização⁵. O quadro regulatório deve acompanhar a evolução tecnológica, com diretrizes para desenvolvimento, validação e monitoramento contínuo das soluções. A transformação digital da saúde decorrente da IA requer atenção à equidade no acesso às inovações tecnológicas. Políticas específicas devem contemplar potenciais impactos socioeconômicos, assim como aspectos relacionados a capacitação profissional e reestruturação do mercado de trabalho no setor⁶.

Este trabalho examina o papel do profissional médico e do Conselho Federal de Medicina (CFM) na governança da IA em diagnósticos e propõe

diretrizes (hierarquia e métricas) para integração ética e segura da tecnologia. A análise contempla aspectos técnicos e regulatórios, com vista a estabelecer um modelo que maximize os benefícios da inovação tecnológica enquanto preserva princípios fundamentais da ética médica e autonomia do paciente. A sinergia entre IA e expertise humana representa um novo paradigma na medicina, que demanda quadro robusto de governança para assegurar desenvolvimento responsável e centrado no paciente⁷. O sucesso dessa transformação depende do alinhamento entre inovação tecnológica, princípios éticos e regulamentação adequada.

Este trabalho tem como objetivos: 1) propor diretrizes éticas para implementação de sistemas de IA em diagnósticos médicos, baseadas nos princípios bioéticos; 2) desenvolver uma hierarquia de desafios éticos para orientar a priorização de ações; e 3) estabelecer métricas objetivas para avaliação da efetividade ética desses sistemas. A análise visa definir o papel do profissional médico e do CFM na governança desses sistemas e, com isso, contribuir para integração segura e responsável da tecnologia na prática clínica.

Trata-se de análise conceitual e revisão crítica da literatura científica e normativa sobre IA aplicada à saúde. O estudo se fundamenta nos princípios bioéticos clássicos (autonomia, beneficência, não maleficência e justiça), marcos regulatórios vigentes e referências bibliográficas recentes sobre ética em IA. Com base na análise, propõe-se uma hierarquia de desafios éticos e métricas para avaliação sistemática da implementação responsável de sistemas inteligentes em diagnósticos médicos.

Vieses algorítmicos e transparência algorítmica

É imprescindível esclarecer os conceitos que fundamentam as preocupações éticas associadas aos sistemas de IA. Os vieses algorítmicos se manifestam quando sistemas de IA produzem julgamentos ou decisões injustas devido a características discriminatórias presentes nos dados de treinamento ou na própria estrutura do algoritmo⁴. A discriminação algorítmica ocorre quando o sistema é influenciado por características efetivamente irrelevantes para a questão em análise, tipicamente relacionadas a preconceitos sobre membros de determinados

grupos. Um aspecto crítico dos vieses algorítmicos é que eles podem não apenas refletir, mas também amplificar e automatizar discriminações históricas presentes nos dados utilizados para treinar os sistemas, perpetuando ciclos de desigualdade por meio do processamento computacional³. Essa problemática é especialmente relevante em sistemas de suporte à decisão que impactam direitos e oportunidades individuais.

A transparência algorítmica, por sua vez, refere-se à possibilidade de compreender como um sistema de IA chega a suas conclusões ou *outputs*. Esse conceito está intrinsecamente relacionado à capacidade de explicar o funcionamento interno do sistema e suas decisões⁴. Um desafio fundamental da transparência algorítmica reside no fato de que, especialmente em sistemas que utilizam aprendizado de máquina e redes neurais, frequentemente é impossível para especialistas identificarem exatamente os padrões extraídos dos dados e como eles são utilizados pelo sistema para gerar resultados⁸. A “opacidade algorítmica” se torna ainda mais problemática quando combinada com a presença de vieses, uma vez que dificulta a identificação e a correção de discriminações potencialmente presentes no sistema, impedindo o médico de participar efetivamente da decisão.

Princípios éticos em diagnósticos em saúde por sistemas inteligentes

É inegável que a IA altera quantitativa e qualitativamente a atividade médica do diagnóstico; porém o uso de ferramentas inteligentes deve sempre respeitar o direito fundamental do paciente de tomar decisões sobre sua própria saúde⁴. Para garantir esse direito, é essencial que o paciente compreenda o processo e dele participe ativamente. Quando um sistema de IA é usado para auxiliar no diagnóstico, o paciente precisa ser informado de forma clara e simples: isso significa explicar não apenas como a tecnologia funciona, mas também suas limitações e possíveis falhas e vieses⁹. É importante ressaltar que a IA não substitui o médico; é uma ferramenta de apoio, de modo que a decisão final deve ser sempre tomada pelo profissional da ciência médica em diálogo com o paciente. O sistema apenas fornece informações adicionais para auxiliar nesse processo⁹.

Quanto às informações pessoais, o paciente deve manter total controle sobre seus dados, do que redundante que poderá autorizar ou não o uso dessas informações, verificar quais dados estão sendo utilizados e, sendo necessário, solicitar correções ou até mesmo a exclusão completa dos registros⁴.

O princípio ético da beneficência, que prioriza a maximização do benefício terapêutico, constitui um dos pilares na aplicação da IA em diagnósticos médicos. A arquitetura dos sistemas de IA potencializa a precisão diagnóstica graças ao processamento de volumes expressivos de dados clínicos com minimização de erros e otimização de resultados assistenciais¹. A seu turno, o princípio da não maleficência, frequentemente enunciado como “não causar dano”, demanda arcabouço robusto de salvaguardas técnicas e processuais e, portanto, requer metodologia rigorosa de validação dos sistemas, que contemple testes extensivos em diversas populações para mitigação de vieses algorítmicos e prevenção de desfechos adversos².

A concretização da IA na medicina diagnóstica estabelece um paradigma inovador fundamentado no princípio da maximização dos benefícios terapêuticos para o paciente. Entre as contribuições mais significativas dessa tecnologia, destaca-se sua capacidade de detecção precoce de condições patológicas por sistemas avançados de análise de dados clínicos – são sistemas que possibilitam a identificação de alterações fisiopatológicas sutis em estágios iniciais, anteriores à manifestação evidente da sintomatologia clínica¹. A capacidade preditiva proporciona aos profissionais de saúde a oportunidade de instituir intervenções terapêuticas em fases preliminares do desenvolvimento patológico, o que resulta em índices significativamente superiores de sucesso terapêutico e em melhorias substanciais na qualidade de vida dos pacientes. Também é relevante destacar que a disponibilização imediata de informações fundamentadas em evidências científicas e o suporte à tomada de decisões clínicas contribuem para o aperfeiçoamento das competências profissionais e, assim, para a elevação dos padrões qualitativos da assistência prestada.

A IA também confere nova dimensão à medicina personalizada. Pela análise sistemática e minuciosa de dados individualizados, sistemas computacionais avançados possibilitam a elaboração de protocolos terapêuticos específicos, fundamentados

nas características genéticas, no histórico clínico e nos determinantes comportamentais de cada paciente, estabelecendo um modelo de precisão terapêutica individualizada¹.

Uma característica distintiva dos sistemas de IA reside em sua capacidade de aprendizado progressivo. O processamento contínuo de dados clínicos possibilita o refinamento sistemático dos algoritmos e resulta no aprimoramento crescente da precisão e eficiência operacional. O processo iterativo estabelece um mecanismo de retroalimentação positiva, no qual a utilização progressiva do sistema resulta em incrementos sucessivos em sua capacidade de assistência aos pacientes.

A otimização da gestão de recursos assistenciais constitui outro benefício significativo da instalação da IA nos sistemas de saúde. A distribuição eficiente de equipamentos e profissionais, mediada por algoritmos avançados, proporciona substancial ampliação do acesso a serviços de saúde qualificados, com alocação criteriosa dos recursos disponíveis, em conformidade com as demandas assistenciais identificadas¹⁰.

Em todas as aplicações, o desenvolvimento tecnológico mantém como princípio norteador a promoção do bem-estar do paciente. Cada inovação implementada deve se orientar pela busca de otimização dos resultados clínicos, minimização do sofrimento e promoção integral da saúde. A observância ao princípio da beneficência assegura que o desenvolvimento da IA na medicina preserve seu objetivo primordial: promover a saúde e elevar a qualidade de vida da população, estabelecendo-se como instrumento tecnológico a serviço do progresso da assistência à saúde.

Quanto ao princípio da justiça na integração da IA em sistemas de diagnóstico médico, ele estabelece diretrizes fundamentais para a distribuição equitativa dos recursos e benefícios tecnológicos⁶. A efetivação desses sistemas requer estrutura metodológica abrangente, que transcenda as tradicionais barreiras socioeconômicas e geográficas, de modo a assegurar a democratização efetiva do acesso a recursos diagnósticos avançados em todas as esferas da sociedade.

O desenvolvimento e o aperfeiçoamento dos sistemas de IA demandam atenção metódica para a mitigação de vieses algorítmicos. A arquitetura dos modelos computacionais deve

fundamentar-se em bases de dados criteriosamente selecionadas, que contemplem adequadamente a diversidade populacional em suas múltiplas dimensões demográficas, étnicas e socioeconômicas⁴. A representatividade constitui elemento essencial para assegurar a uniformidade da precisão diagnóstica entre os diversos segmentos populacionais. A metodologia de coleta e processamento de dados representa componente crítico no processo de desenvolvimento. A concretização ética requer a incorporação sistemática de grupos minoritários nas bases de treinamento algorítmico, com procedimentos rigorosos que previnam discriminações sistemáticas e garantam equidade nos resultados diagnósticos para todos os segmentos populacionais³.

A viabilidade econômica emerge como aspecto fundamental na democratização do acesso a sistemas diagnósticos baseados em IA. A estruturação do modelo de implementação demanda políticas de precificação equilibradas e mecanismos de subsídio que assegurem a universalização do acesso a recursos diagnósticos, independentemente das condições socioeconômicas dos usuários do sistema⁶.

No âmbito internacional, essas tecnologias apresentam potencial significativo para reduzir disparidades na assistência à saúde, particularmente pela efetivação de recursos diagnósticos remotos. O processo de execução requer consideração criteriosa das especificidades regionais e das demandas particulares das áreas tradicionalmente subatendidas pelos sistemas de saúde convencionais¹.

A distribuição equitativa dos benefícios tecnológicos demanda uma estratégia de realização metodologicamente estruturada e abrangente. O planejamento estratégico deve priorizar áreas com elevado potencial de impacto social positivo, em vez de se concentrar exclusivamente em centros de alto volume assistencial, e assegurar que as otimizações na precisão diagnóstica e nos tempos de resposta beneficiem a totalidade da população⁴.

O processo de desenvolvimento necessita incorporar mecanismos efetivos de participação comunitária, integrando as diversas perspectivas culturais e sociais. O engajamento participativo assegura o alinhamento entre as funcionalidades tecnológicas implementadas e as necessidades específicas das comunidades atendidas, promovendo um empreendimento culturalmente apropriado e socialmente responsável⁶.

A estrutura normativa deve estabelecer garantias fundamentais de assistência não discriminatória, preservação da privacidade e acesso universal a serviços de saúde. A observância rigorosa do princípio da justiça possibilita que sistemas de IA atuem como instrumentos efetivos na promoção da equidade assistencial e contribuam significativamente para a mitigação das disparidades existentes no cenário global da saúde pública.

Governança de dados em saúde

A governança de dados em saúde no contexto de sistemas de IA diagnósticos representa um domínio crítico que demanda abordagem abrangente e rigorosa para assegurar tanto a efetividade clínica quanto a proteção dos direitos individuais⁴. O tema pode ser analisado sob diferentes dimensões fundamentais que se inter-relacionam. A privacidade dos dados de saúde constitui pilar essencial da governança, dada a natureza altamente sensível das informações médicas. Sistemas de IA diagnósticos processam volumes expressivos de dados clínicos, que contêm históricos médicos, resultados de exames, imagens diagnósticas e informações genéticas. A proteção dessa informação requer utilização de protocolos robustos de anonimização e pseudonimização, a fim de garantir que dados utilizados para treinar e operar sistemas de IA não permitam identificação individual dos pacientes⁶.

A segurança da informação em sistemas de IA diagnósticos demanda arquitetura tecnológica que contemple múltiplas camadas de proteção, incluindo criptografia avançada para dados em trânsito e em repouso, controles granulares de acesso baseados em papéis, autenticação multifatorial para usuários do sistema e registro detalhado de todas as operações realizadas. Deve-se levar a segurança em conta desde a concepção do sistema e adotar princípios de *privacy by design* e *security by design*⁴.

Cabe breve explanação: a granularidade se refere aos controles detalhados e específicos de acesso baseados em papéis, que permitem definir com precisão quais usuários podem acessar determinadas informações ou funcionalidades de acordo com sua função, e assim oferecem um

gerenciamento de acessos refinado com ajustes precisos de permissões para cada tipo de usuário.

Privacy by design é um princípio que estabelece que a privacidade deve ser considerada desde o início do desenvolvimento de um sistema, e não elemento adicionado posteriormente, ou seja, as medidas de proteção de dados são incorporadas na própria arquitetura do sistema de IA diagnóstico para garantir que a proteção de dados pessoais seja fundamental. Já *security by design*, similar ao anterior, significa que a segurança deve ser elemento fundacional integrado desde a fase de concepção, ou seja, parte integrante da arquitetura, dos processos e das funcionalidades, o que torna o sistema intrinsecamente mais seguro e menos vulnerável a ameaças.

O consentimento informado assume nova complexidade no contexto de IA em saúde. Os pacientes precisam compreender não apenas que seus dados serão utilizados para seu próprio diagnóstico, mas também que podem ser empregados para treinar e aperfeiçoar os sistemas de IA. O referido consentimento deve ser granular, para permitir que indivíduos exerçam controle sobre diferentes usos de suas informações médicas⁴.

A interoperabilidade segura entre sistemas representa outro aspecto crucial. Sistemas de IA diagnósticos frequentemente precisam integrar dados de múltiplas fontes e instituições de saúde, e nessa integração deve-se preservar a confidencialidade e integridade dos dados por meio de protocolos padronizados de intercâmbio de informações médicas que incorporem mecanismos robustos de segurança¹.

O armazenamento e a retenção de dados demandam políticas específicas que equilibrem necessidades clínicas, requisitos regulatórios e direitos individuais, para o que é incontornável definir períodos apropriados de retenção, procedimentos seguros de descarte de dados e mecanismos para que pacientes possam exercer direitos como acesso, retificação e exclusão de suas informações⁴.

A governança também deve contemplar aspectos de qualidade e integridade dos dados. Sistemas de IA diagnósticos dependem criticamente da acuidade e da completude das informações utilizadas, e por isso demandam processos rigorosos de validação de dados, detecção de inconsistências e

manutenção de trilhas de auditoria que documentem toda manipulação realizada².

O compartilhamento de dados para pesquisa e desenvolvimento representa área que demanda governança específica. Embora seja fundamental para o avanço científico e aperfeiçoamento dos sistemas, deve ocorrer por meio de protocolos que garantam anonimização efetiva e uso ético das informações, com especial atenção a riscos de reidentificação por cruzamento de bases de dados⁶.

A conformidade regulatória constitui elemento transversal da governança e leva em consideração múltiplos marcos normativos, como legislações de proteção de dados pessoais, regulamentações específicas do setor de saúde e diretrizes éticas para uso de IA, e sua conformidade deve ser continuamente monitorada e documentada⁴.

Por fim, a governança deve incorporar mecanismos de prestação de contas e transparência como: documentação clara de políticas e procedimentos, auditorias regulares, canais para questionamentos e reclamações e comunicação proativa com interessados sobre práticas de gestão de dados. A governança precisa ser dinâmica e adaptativa para evoluir continuamente de modo a assimilar novos desafios tecnológicos, requisitos regulatórios e expectativas sociais quanto à proteção de dados em saúde. O sucesso na implementação dessa governança é fundamental para construir e manter a confiança necessária para a adoção bem-sucedida de sistemas de IA diagnósticos⁶.

Estrutura de responsabilidade

A integração da IA em diagnósticos médicos estabelece novo paradigma de responsabilização profissional que contempla múltiplas dimensões críticas, devendo-se considerar três dimensões: o profissional de medicina, o CFM e os requisitos do sistema de IA. O quadro de capacitação emerge como requisito fundamental, que demanda programa abrangente de treinamento em funcionalidades algorítmicas, limitações sistêmicas e integração adequada das recomendações automatizadas no processo decisório⁵.

Quanto aos médicos, o processo decisório mantém a primazia do julgamento médico, e os sistemas de IA se posicionam como ferramentas consultivas complementares. A avaliação crítica

das recomendações automatizadas deve se fundamentar na *expertise* profissional e nas particularidades de cada caso clínico, de modo a evitar dependência exclusiva da automação⁷.

A governança de incidentes estabelece responsabilidade pela notificação sistemática de erros e vieses identificados nos sistemas. Trata-se de verdadeiro mecanismo de retroalimentação que constitui elemento essencial para refinamento algorítmico e otimização contínua da segurança e eficácia operacional².

O processo de validação do uso do sistema inteligente requer avaliação criteriosa de sua aplicabilidade em populações específicas, baseada em análise que contemple métricas de desempenho fundamentais como taxas de falsos positivos/negativos e monitoramento sistemático de potencial degradação da precisão diagnóstica².

A comunicação transparente com pacientes constitui imperativo ético e demanda explicação adequada da participação da IA no processo diagnóstico/terapêutico, de maneira a preservar a autonomia do paciente por meio do consentimento informado e da compreensão das implicações tecnológicas⁹.

A utilização responsável pressupõe aderência estrita ao escopo pretendido dos sistemas, para evitar aplicações inadequadas ou extrapolações que possam resultar em danos. O julgamento clínico deve orientar a mitigação de riscos potenciais por meio da compreensão das limitações sistêmicas⁸.

O desenvolvimento profissional contínuo emerge como requisito para acompanhamento da evolução tecnológica: é um processo que contempla atualização sobre novos sistemas, modificações algorítmicas e utilização de melhores práticas operacionais².

A prática ética demanda manutenção de padrões rigorosos na utilização da IA para se preservar os direitos dos pacientes, confidencialidade e equidade assistencial. A aderência a ditos princípios garante que sistemas inteligentes contribuam positivamente para a qualidade e segurança do processo diagnóstico e, assim, para a manutenção da integridade da responsabilização médica⁴.

Quanto ao CFM, o órgão assume função estratégica na orientação e regulamentação da implementação ética da IA no contexto assistencial,

com múltiplas atribuições fundamentais. Em primeiro lugar, o desenvolvimento de diretrizes constitui responsabilidade primordial, que demanda elaboração e atualização sistemática de normativas éticas para utilização da IA. Essa estrutura deve contemplar princípios essenciais de segurança, transparência, responsabilização e equidade, de modo a estabelecer modelo para realização ética⁸.

A função de supervisão profissional requer defesa ativa de políticas protetivas aos direitos dos pacientes, assegurando a utilização ética dos sistemas e dirimindo questões críticas como privacidade informacional, vieses algorítmicos e democratização do acesso, tudo em atuação a pressupor articulação com instâncias regulatórias para desenvolvimento de arcabouço legal apropriado⁴.

Promover transparência algorítmica emerge como imperativo, de sorte que deve incentivar desenvolvedores à concretização de sistemas interpretáveis. Paralelamente, patrocinar capacitação profissional em avaliação crítica das recomendações automatizadas também é medida que se impõe ao CFM, de modo que o desenvolvimento profissional continuado disponibilize recursos e oportunidades de habilitação em tecnologias emergentes, implicações éticas e integração responsável à prática clínica².

Intimamente conectado à promoção de instrução em novas tecnologias está o estabelecimento de padrões e certificações que visem assegurar concretização exclusiva de sistemas seguros, eficazes e aderentes a princípios éticos. Trata-se de processo que prevê definição de parâmetros objetivos de desempenho, governança de dados e integração assistencial⁵.

O processo de revisão ética constitui salvaguarda fundamental, no qual novas tecnologias são submetidas a escrutínio sistemático antes de serem adotadas em larga escala. É avaliação preventiva que identifica e endereça potenciais implicações éticas⁶.

O fomento à colaboração entre desenvolvedores, pesquisadores e clínicos potencializa o desenvolvimento de sistemas alinhados à ética médica. Trata-se de integração que vislumbra suporte à pesquisa em aplicações éticas e compartilhamento de melhores práticas¹.

Engajamento social visa promover confiança pública e alinhamento dos sistemas com valores

comunitários. É um processo que incorpora perspectivas dos pacientes no desenvolvimento tecnológico e responde a preocupações sociais⁶.

A derradeira medida do CFM está no monitoramento sistemático do impacto assistencial, com ajustes fundamentados em diretrizes e políticas e que respondam a questões éticas emergentes via retorno contínuo de profissionais e pacientes². A atuação abrangente do CFM visa assegurar aderência da IA em saúde a padrões éticos rigorosos, salvaguardar interesses dos pacientes, embasar a prática médica e promover avanço responsável da medicina tecnológica.

Considerando os requisitos do sistema inteligente, uma estrutura prática para avaliação ética de assistentes de IA em diagnósticos deve contemplar múltiplas dimensões de análise que permitam verificação sistemática da conformidade ética em todas as etapas do ciclo de vida do sistema. Referida avaliação se inicia pela análise da qualidade e da representatividade dos dados utilizados para treinamento do sistema, com verificação de potenciais vieses nos conjuntos de dados, garantia de diversidade demográfica adequada e assecuração de que os dados foram obtidos de forma ética e com consentimento apropriado dos pacientes³.

A transparência algorítmica constitui o segundo pilar na avaliação, e demanda verificação da capacidade do sistema em explicar seus processos decisórios de forma compreensível para profissionais de saúde e pacientes. O sistema deve permitir rastreabilidade das decisões e identificação de fatores que influenciaram cada diagnóstico sugerido. A transparência deve se estender às limitações conhecidas do sistema, que necessitam ser claramente documentadas e comunicadas⁴.

A avaliação deve contemplar aspectos de equidade e não discriminação, para o que deve verificar se o sistema mantém níveis consistentes de acuidade diagnóstica para diferentes grupos populacionais. A medida requer análise estatística rigorosa do desempenho do sistema estratificado por características demográficas relevantes, como idade, gênero, etnia e condições socioeconômicas. Quaisquer disparidades identificadas devem ser investigadas e adequadamente resolvidas³.

A privacidade e a segurança dos dados precisam ser avaliadas segundo critérios objetivos que verifiquem a previsão de controles técnicos e organizacionais apropriados, entre os quais análise dos

mecanismos de anonimização, protocolos de criptografia, controles de acesso, procedimentos de backup e recuperação, além de políticas de retenção e descarte de dados. A avaliação deve verificar também a existência de procedimentos claros para resposta a incidentes de segurança⁴.

O impacto na autonomia e no processo decisório médico representa aspecto crucial da avaliação ética. Deve-se analisar como o sistema se integra ao fluxo de trabalho clínico e verificar se preserva apropriadamente a autonomia dos profissionais de saúde e de seus pacientes. O sistema deve se posicionar claramente como ferramenta de suporte à decisão, sem substituir o julgamento médico ou comprometer a relação médico-paciente⁷.

É imperativa a existência de mecanismos claros de atribuição de responsabilidades em caso de erros ou falhas, devendo a avaliação examinar a atribuição de papéis e responsabilidades, processos de supervisão humana, procedimentos de auditoria e mecanismos para investigação e correção de problemas identificados⁵.

A sustentabilidade e a manutenção ética do sistema ao longo do tempo devem ser consideradas e concretizadas por meio de aferição da existência de processos de monitoramento contínuo de desempenho, atualização de modelos, validação de novos dados incorporados e adaptação a mudanças nos padrões éticos ou requisitos regulatórios. A avaliação deve examinar também planos de contingência para situações de degradação ou obsolescência do sistema¹.

O alinhamento com marcos regulatórios e diretrizes éticas estabelecidas precisa ser verificado de forma sistemática e estar em conformidade com legislações de proteção de dados, regulamentações específicas do setor de saúde e padrões éticos profissionais. A avaliação deve considerar também aderência a diretrizes internacionais relevantes para IA em saúde⁶.

Por fim, o processo de avaliação deve culminar em relatório detalhado que documente achados, recomendações específicas para correção de problemas e planos de ação para aplicação de melhorias. Tal relatório deve ser periodicamente atualizado, em um ciclo contínuo de avaliação e aperfeiçoamento ético do sistema. A estrutura de avaliação deve ser aplicada não apenas durante o desenvolvimento inicial do sistema, mas também periodicamente ao longo de sua operação,

a fim de assegurar manutenção contínua de altos padrões éticos. A frequência e a profundidade das reavaliações podem ser ajustadas com base em fatores como criticidade do sistema, velocidade de evolução tecnológica e mudanças no contexto regulatório².

Resultados

A análise desenvolvida permitiu identificar dois resultados principais: 1) hierarquia de desafios éticos estruturada em seis níveis (qualidade e integridade dos dados; transparência algorítmica e segurança; interface humano-máquina; equidade e justiça; responsabilização e governança; e sustentabilidade ética); 2) modelo de avaliação com métricas específicas para cada dimensão ética: equidade (disparidade de acuidade entre grupos), transparência (tempo para explicações compreensíveis), privacidade (taxa de sucesso em testes de reidentificação), autonomia (concordância médico-sistema), governança (tempo de resolução de incidentes) e sustentabilidade (frequência de atualizações éticas).

Proposta de hierarquia

A implementação ética de sistemas de IA em diagnósticos médicos apresenta uma hierarquia complexa de desafios que se inter-relacionam, a qual pode ser estruturada em níveis crescentes de complexidade e interdependência. No nível mais fundamental estão os desafios relativos a qualidade e integridade dos dados, porque constituem a base sobre a qual todo o sistema diagnóstico será construído. A qualidade dos dados impacta diretamente a acuidade diagnóstica e potenciais vieses do sistema e, portanto, estabelece relação de dependência fundamental com todos os níveis superiores da hierarquia³.

Em um segundo nível hierárquico se situam os desafios técnicos de execução, que compreendem transparência algorítmica e segurança do sistema. A primeira depende diretamente da qualidade dos dados e da arquitetura do sistema, enquanto segurança precisa ser considerada tanto no nível dos dados quanto nos processos algorítmicos – são aspectos técnicos que formam a infraestrutura necessária para enfrentar questões éticas mais complexas nos níveis superiores⁴.

O terceiro nível contempla desafios relacionados à interface humano-máquina e à integração do sistema no ambiente clínico e engloba aspectos como preservação da autonomia médica, adequada comunicação de capacidades e limitações do sistema e manutenção da relação médico-paciente. Este nível depende fundamentalmente da solidez dos anteriores, pois uma interface efetiva requer dados confiáveis e sistemas tecnicamente robustos⁹.

No quarto nível se encontram os desafios de equidade e justiça na distribuição dos benefícios da tecnologia. Isso envolve garantir acesso equitativo aos sistemas de IA diagnóstica e prevenir discriminações algorítmicas. Equidade só pode ser alcançada com dados representativos, sistemas transparentes e interfaces adequadas⁶.

O quinto nível aborda os desafios de responsabilização e governança, com estabelecimento de estruturas claras para atribuição de responsabilidades e gestão de riscos. Uma governança eficaz requer compreensão e controle de todos os aspectos do sistema, desde os dados até os impactos sociais⁵.

No nível mais alto da hierarquia se encontram desafios relativos a sustentabilidade e evolução ética do sistema ao longo do tempo. Abrangem-se aqui as capacidades de adaptação a novos padrões éticos, de incorporação de avanços tecnológicos e de resposta a mudanças nas necessidades sociais. Este nível depende do funcionamento harmônico de todos os níveis anteriores e requer mecanismos robustos de monitoramento e ajuste contínuo¹.

A estrutura hierárquica evidencia que falhas nos níveis mais fundamentais podem comprometer toda a implementação ética do sistema. Por exemplo, problemas na qualidade dos dados podem se propagar através de todos os níveis e afetar desde a precisão técnica até a equidade social. Similarmente, deficiências na transparência algorítmica podem impactar negativamente a confiança dos usuários e a efetividade da governança³.

A priorização de ações nessa hierarquia deve seguir uma abordagem *bottom-up*, isto é, assegurar primeiro a solidez dos níveis fundamentais e só então avançar para desafios mais complexos. No entanto, o planejamento deve adotar uma visão *top-down* e considerar como as decisões em cada nível impactarão os objetivos éticos mais amplos do sistema⁶.

O quadro hierárquico deve ser utilizado como guia dinâmico para a realização ética, com base no reconhecimento de que os diferentes níveis, embora distintos, operam de forma integrada e requerem atenção contínua para manutenção da integridade ética do sistema como um todo. A compreensão das relações de dependência e prioridade é fundamental para alocar efetivamente os recursos e desenvolver estratégias bem-sucedidas².

Métricas

A avaliação da efetividade ética dos sistemas de IA em diagnósticos médicos requer um conjunto abrangente de métricas que permitam monitoramento sistemático da aderência aos princípios éticos. As métricas podem ser organizadas em dimensões complementares que capturem diferentes aspectos da *performance* ética do sistema. No âmbito da equidade e não discriminação, métricas essenciais incluem a disparidade de acuidade diagnóstica entre diferentes grupos demográficos, calculada pela comparação estatística das taxas de verdadeiros positivos e negativos estratificadas por características como gênero, idade, etnia e *status* socioeconômico. O desvio-padrão dessas taxas entre grupos fornece indicador quantitativo da uniformidade de *performance* do sistema. Adicionalmente, deve-se mensurar a representatividade dos dados de treinamento pelo índice de similaridade populacional, que compara a distribuição demográfica dos dados com a população-alvo³.

Para avaliar a transparência algorítmica, métricas relevantes incluem o tempo médio necessário para geração de explicações compreensíveis das decisões diagnósticas, a taxa de completude das explicações medida pela proporção de fatores decisórios identificáveis e o índice de satisfação dos profissionais de saúde com as explicações fornecidas, coletado em pesquisas estruturadas. A rastreabilidade das decisões pode ser quantificada pelo percentual de diagnósticos para os quais é possível reconstruir completamente a cadeia decisória⁴.

No domínio de privacidade e segurança, métricas críticas envolvem a taxa de sucesso em testes de reidentificação realizados nos dados anonimizados, o tempo médio de detecção e resposta a tentativas de acesso não autorizado e o

índice de conformidade com protocolos de segurança estabelecidos. A efetividade dos controles de acesso pode ser medida pela proporção de acessos adequadamente autorizados em auditorias periódicas⁶.

A autonomia médica e do paciente pode ser avaliada por meio de métricas como a taxa de concordância entre recomendações do sistema e decisões finais dos médicos, o tempo médio dedicado à discussão das recomendações com pacientes e o índice de satisfação dos pacientes com seu envolvimento no processo decisório. A frequência de substituição das recomendações do sistema fornece indicador importante da preservação do julgamento clínico⁷.

Para governança e responsabilização, métricas relevantes incluem o tempo médio de investigação e resolução de incidentes reportados, a taxa de implementação de recomendações de auditorias éticas e o índice de completude da documentação de decisões críticas. A efetividade dos mecanismos de supervisão pode ser medida pela proporção de problemas identificados proativamente versus aqueles reportados externamente⁵.

A sustentabilidade ética do sistema pode ser avaliada por métricas como a frequência de atualizações do modelo para incorporar novos padrões éticos, o tempo médio de adaptação a mudanças regulatórias e a taxa de obsolescência de componentes críticos do sistema. O investimento contínuo em aprimoramento ético pode ser quantificado pela proporção do orçamento dedicada a iniciativas de ética e conformidade¹.

A efetividade das métricas deve ser periodicamente reavaliada por metanálises que examinem sua capacidade de prever e prevenir problemas éticos, albergando análise da correlação entre indicadores e ocorrência de incidentes éticos, bem como avaliação da completude da cobertura dos princípios éticos pelas métricas estabelecidas².

O arcabouço de métricas deve ser implementado via painel integrado que permita monitoramento contínuo e identificação precoce de tendências preocupantes. Limiares de alerta devem ser estabelecidos para cada métrica, os quais, quando ultrapassados, desencadeiem análises aprofundadas. A agregação ponderada das métricas pode fornecer indicadores compostos do status ético global do sistema, o que facilita

a comunicação com stakeholders e a tomada de decisão estratégica⁶.

As métricas devem ser documentadas em “quadro de resultados” (*scorecard*) ético que inclua definição precisa de cada indicador, método de cálculo, frequência de medição, responsáveis pelo monitoramento e ações corretivas associadas a desvios identificados. O *scorecard* deve ser regularmente atualizado para refletir a evolução das preocupações éticas e os aprendizados obtidos com a operação do sistema².

Delimitação de responsabilidade

A intersecção entre responsabilidade profissional e limitações tecnológicas em sistemas de IA diagnósticos constitui domínio complexo que demanda estruturação clara dos limites de responsabilização. A delimitação necessita considerar tanto as capacidades e as restrições inerentes aos sistemas de IA quanto as obrigações éticas e legais dos profissionais de saúde. A responsabilidade profissional do médico se mantém como elemento central no processo diagnóstico, mesmo com a introdução de sistemas de IA. O profissional de saúde tem obrigação fundamental de exercer julgamento clínico independente e tomar decisões finais considerando o contexto específico de cada paciente. A responsabilidade abarca a avaliação crítica das recomendações fornecidas pelos sistemas de IA, isto é, deve considerar suas limitações conhecidas e potenciais vieses⁵.

As limitações tecnológicas dos sistemas de IA devem ser claramente documentadas e comunicadas aos médicos, com especificação precisa do escopo de aplicação do sistema, descrição das populações representadas nos dados de treinamento, identificação de situações nas quais o sistema pode apresentar performance reduzida e explicitação das margens de erro esperadas. A responsabilidade por falhas decorrentes de limitações não documentadas ou inadequadamente comunicadas recai sobre os desenvolvedores e fornecedores do sistema⁸.

A responsabilização por erros diagnósticos deve considerar a cadeia de eventos e decisões que levaram ao resultado adverso. Quando o profissional de saúde segue adequadamente os protocolos estabelecidos para uso do sistema, incluindo consideração apropriada de suas limitações conhecidas,

sua responsabilidade deve ser avaliada primariamente em relação à adequação do julgamento clínico exercido, e não pertinentemente às limitações intrínsecas do sistema tecnológico⁵.

Os desenvolvedores e fornecedores de sistemas de IA diagnósticos têm responsabilidade específica relacionada à qualidade técnica do sistema, que compreende acurácia das predições dentro do escopo especificado, robustez dos mecanismos de segurança e efetividade dos protocolos de monitoramento e manutenção. Desvios decorrentes de deficiências nessas áreas devem ser atribuídas primariamente a referidas entidades, desde que o sistema tenha sido utilizado conforme especificado⁵.

As instituições de saúde que implementam sistemas de IA diagnósticos assumem responsabilidade pela adequação da infraestrutura de suporte, treinamento dos profissionais e estabelecimento de protocolos claros de uso. Erros decorrentes de deficiências nos aspectos organizacionais devem ser atribuídos primariamente à instituição, mesmo quando resultam em erros diagnósticos específicos⁸.

Um elemento inarredável na delimitação de responsabilidades é o estabelecimento de protocolos claros para investigação de incidentes. Tais protocolos devem permitir identificação objetiva da origem das falhas, com diferenciação entre limitações tecnológicas conhecidas, erros de execução, desacertos de procedimento e equívocos de julgamento clínico – a diferenciação é fundamental para atribuir apropriadamente responsabilidades e desenvolver medidas corretivas efetivas⁵.

A gestão de riscos em sistemas de IA diagnósticos deve incorporar mecanismos de compartilhamento de responsabilidade quando apropriado, de modo a alcançar estruturas de seguro específicas para cobrir diferentes tipos de incorreções, acordos contratuais que especifiquem limites de responsabilidade e protocolos de cooperação entre diferentes atores para resolução de problemas identificados⁵.

O monitoramento contínuo da performance do sistema e dos padrões de uso é responsabilidade compartilhada entre desenvolvedores, instituições de saúde e profissionais. A identificação precoce de desvios ou degradação de performance permite intervenção preventiva e ajuda a estabelecer

limites claros de responsabilização quando ocorrerem falhas².

A evolução tecnológica e o aprendizado obtido com a operação dos sistemas devem servir de base para atualizações periódicas no âmbito de responsabilização, as quais englobam refinamento dos protocolos de uso, ajuste dos limites de responsabilidade e desenvolvimento de novas salvaguardas conforme novas limitações ou riscos são identificados¹.

Considerações finais

Os resultados centrais deste trabalho, a hierarquia de desafios éticos e o modelo de métricas para avaliação, fornecem instrumentos práticos para a implementação responsável de sistemas de IA em diagnósticos médicos. Esses resultados respondem diretamente aos objetivos propostos na medida em que estabelecem diretrizes concretas fundamentadas nos princípios bioéticos e oferecem estrutura organizacional para priorização de ações e mecanismos objetivos de monitoramento contínuo⁶.

A estruturação proposta – fundamentada nos pilares da autonomia, beneficência, não maleficência e justiça – oferece arranjo robusto para orientar a utilização responsável dos sistemas inteligentes de apoio diagnóstico. A abordagem reconhece tanto o potencial transformador da IA para aprimoramento da precisão diagnóstica e democratização do acesso à saúde quanto a necessidade crítica de salvaguardas éticas que preservem a centralidade do paciente e a integridade da prática médica⁴.

O papel estratégico do CFM se mostra inafastável para a concretização dessa visão, por meio do desenvolvimento de diretrizes, da supervisão regulatória e do fomento à capacitação profissional. Sua atuação deve assegurar que a evolução tecnológica ocorra em alinhamento com valores fundamentais da medicina, assim como promover a integração harmoniosa entre *expertise* humana e capacidades computacionais⁸.

A delimitação clara de responsabilidades e o estabelecimento de métricas objetivas para avaliação da efetividade ética constituem contribuições práticas essenciais deste trabalho. São ferramentas que oferecem mecanismos concretos para operacionalização dos princípios discutidos e assim

facilitam a efetivação e o monitoramento contínuo de sistemas de IA diagnósticos⁵.

A governança de dados em saúde e a hierarquização dos desafios éticos apresentadas fornecem arrumação organizacional que permite abordagem sistemática e priorizada dos múltiplos aspectos envolvidos. A organização facilita a alocação eficiente de recursos e o desenvolvimento de estratégias efetivas para enfrentamento das questões identificadas⁶.

Por derradeiro, ressalta-se que o sucesso da transformação tecnológica dependerá fundamentalmente

da capacidade de manter o equilíbrio dinâmico entre inovação e humanização do cuidado. O compromisso inabalável com princípios éticos, associado à compreensão profunda das potencialidades e das limitações dos sistemas de IA, permitirá a construção de uma medicina mais precisa e acessível, sem comprometer a essência humanística que caracteriza a relação médico-paciente. O novo paradigma de prática médica, potencializado pela IA mas firmemente ancorado em valores éticos, representa promissora evolução para o futuro da assistência em saúde.

Referências

1. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med* [Internet]. 2022 [acesso 7 jan 2025];28(1):31-8. DOI: 10.1038/s41591-021-01614-0
2. Faes L, Liu X, Wagner SK, Fu DJ, Balaskas K, Sim DA *et al*. A Clinician's guide to artificial intelligence: how to critically appraise machine learning studies. *Transl Vis Sci Technol* [Internet]. 2020 [acesso 7 jan 2025];9(2):7. DOI: 10.1167/tvst.9.2.7
3. Parikh RB, Teeple S, Navathe AS. Addressing bias in artificial intelligence in health care. *JAMA* [Internet]. 2019 [acesso 7 jan 2025];322(24):2377-8. DOI: 10.1001/jama.2019.18058
4. Schönberger D. Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications. *International Journal of Law and Information Technology* [Internet]. 2019 [acesso 7 jan 2025];27(2):171-203. DOI: 10.1093/ijlit/eaz004
5. Smith H, Fotheringham K. Artificial intelligence in clinical decision-making: Rethinking liability. *Medical Law International* [Internet]. 2020 [acesso 7 jan 2025];20(2):131-54. DOI: 10.1177/0968533220945766
6. Murphy K, Di Ruggiero E, Upshur R, Willison DJ, Malhotra N, Cai JC *et al*. Artificial intelligence for good health: a scoping review of the ethics literature. *BMC Med Ethics* [Internet]. 2021 [acesso 3 abr 2020];22(1):14. DOI: 10.1186/s12910-021-00577-8
7. Desai AN. Artificial intelligence: promise, pitfalls, and perspective. *JAMA* [Internet]. 2020 [acesso 3 abr 2020];323(24):2448-9. DOI: 10.1001/jama.2020.8737
8. Matsuzaki T. Ethical issues of artificial intelligence in medicine. *California Western Law Review* [Internet]. 2018 [acesso 3 abr 2020]; 55(1):19. Disponível: <https://bit.ly/4saPVkk>
9. Soellner M, Koenigstorfer J. Compliance with medical recommendations depending on the use of artificial intelligence as a diagnostic method. *BMC Med Inform Decis Mak* [Internet]. 2021 [acesso 3 abr 2020];21(1):236. DOI: 10.1186/s12911-021-01596-6
10. Marchiori C, Dykeman D, Girardi I, Ivankay A, Thandiackal K, Zusag M *et al*. Artificial intelligence decision support for medical triage. *AMIA Annu Symp Proc* [Internet]. 2020 [acesso 3 abr 2020];793-802. Disponível: <https://bit.ly/4aQmME1>

Ricardo Rheingantz Abuchaim – Doutor – ricardoabuchaim@hotmail.com

 0009-0006-9802-6648

Daniel Brito de Araujo – Doutor – araujodb@gmail.com

 0000-0002-4840-945X

Correspondência

Ricardo Rheingantz Abuchaim – Rua Félix da Cunha, 722, Centro. 96010-000. Pelotas/RS, Brasil

Contribuições dos autores

Ambos os autores colaboraram da mesma forma na redação do manuscrito, revisão crítica do conteúdo intelectual e aprovação da versão final para publicação. Ricardo Rheingantz Abuchaim foi responsável pelo delineamento e pela concepção do estudo, com a qual Daniel Brito de Araujo colaborou.

Disponibilidade de dados: Todos os dados utilizados ou gerados na pesquisa estão integralmente descritos e apresentados no corpo do artigo.

Editora responsável: Dilza Teresinha Ambrós Ribeiro

Recebido: 16.4.2025

Revisado: 1.9.2025

Aprovado: 14.1.2026