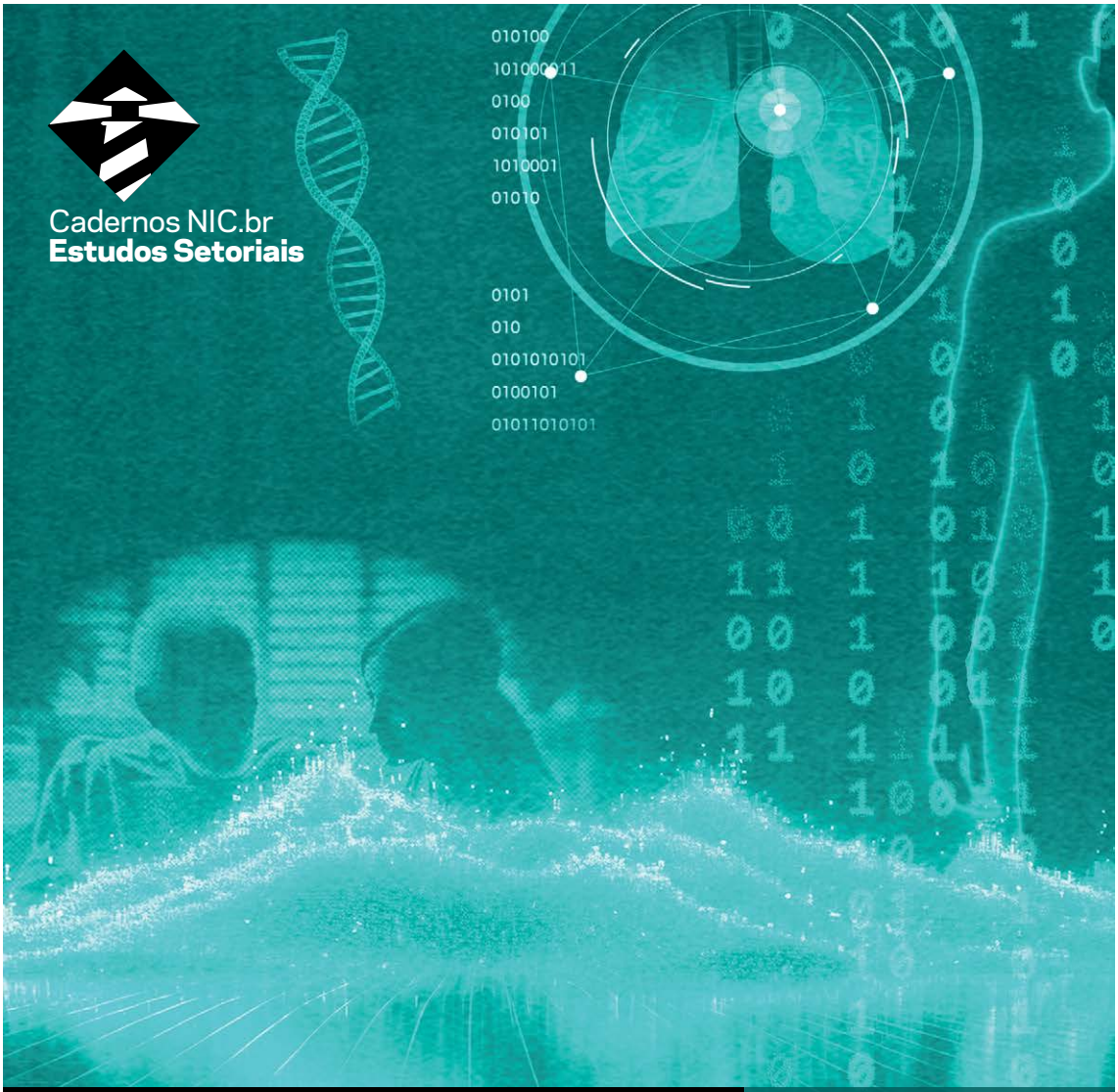


Cadernos NIC.br
Estudos Setoriais

INTELIGÊNCIA ARTIFICIAL NA SAÚDE:
potencialidades, riscos e perspectivas para o Brasil

nic.br
Núcleo de Informação e Coordenação do Ponto BR

Núcleo de Informação
e Coordenação do Ponto BR – NIC.br

Cadernos NIC.br
Estudos Setoriais

INTELIGÊNCIA ARTIFICIAL NA SAÚDE:
potencialidades, riscos e perspectivas para o Brasil

Comitê Gestor da Internet no Brasil – CGI.br
São Paulo 2024

CRÉDITOS DA EDIÇÃO

COORDENAÇÃO EXECUTIVA E EDITORIAL: Alexandre F. Barbosa (Cetic.br|NIC.br)

COORDENAÇÃO CIENTÍFICA: Heimar de Fátima Marin

COORDENAÇÃO TÉCNICA: Graziela Castello, Rodrigo Brandão de Andrade e Silva, Mariana Galhardo Oliveira e Javiera F. Medina Macaya (Cetic.br|NIC.br)

GESTÃO DA PESQUISA EM CAMPO: Monise Picanço, Priscila Vieira, Marina Castro de Oliveira e Florbela Ribeiro (Centro Brasileiro de Análise e Planejamento [CEBRAP])

APOIO À EDIÇÃO: Carolina Carvalho e Leandro Esmelardi Espindola (Comunicação|NIC.br)

PREPARAÇÃO E REVISÃO EM PORTUGUÊS: Érica Santos Soares de Freitas

TRADUÇÃO: Ana Zuleika Pinheiro Machado

PROJETO GRÁFICO E ILUSTRAÇÕES: Pilar Velloso

DIAGRAMAÇÃO: Milena Branco

FOTOS: Shutterstock

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Inteligência artificial na saúde [livro eletrônico] potencialidades, riscos e perspectivas para o Brasil / Núcleo de Informação e Coordenação do Ponto BR (NIC.br) ; tradução Ana Zuleika Pinheiro Machado. -- São Paulo : Comitê Gestor da Internet no Brasil, 2024.
PDF

Vários colaboradores.
Bibliografia.
ISBN 978-65-85417-55-6

1. Inteligência artificial 2. Pesquisa qualitativa 3. Saúde - Brasil 4. Saúde - Pesquisa I. Núcleo de Informação e Coordenação do Ponto BR (NIC.br).

24-218400 CDD-610.3

Índices para catálogo sistemático:

1. Ciências da saúde : Pesquisa 610.3
Tábata Alves da Silva - Bibliotecária - CRB-8/9253

Comitê Gestor da Internet no Brasil – CGI.br

SUMÁRIO

AGRADECIMENTOS

O Núcleo de Informação e Coordenação do Ponto BR (NIC.br), por meio do Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br), agradece a todos os profissionais envolvidos na elaboração da presente publicação. Especialmente, agradecemos a contribuição de Heimar Fátima Marin (International Medical Informatics Association [IMIA]), Organização Mundial da Saúde (OMS) e União Internacional de Telecomunicações (UIT), Fernando Aith e Daniel A. Dourado (Universidade de São Paulo [USP]), Monise Picanço e Priscila Viera (Centro Brasileiro de Análise e Planejamento [CEBRAP]), Glauco Arbix e João Paulo Cândia Veiga (USP).

APRESENTAÇÃO

A Internet tornou-se uma infraestrutura essencial para o desenvolvimento de novas aplicações digitais e para a produção e disseminação de dados em uma escala nunca vista. As aplicações baseadas em Inteligência Artificial (IA) beneficiaram-se dos componentes e do gigantesco volume de informações disponíveis na rede. Atualmente, a IA permeia diversos campos da vida social. Na área da saúde, as pesquisas e as discussões sobre essa tecnologia intensificam-se e tornam-se cada vez mais urgentes. Graças às técnicas de aprendizado de máquina (*machine learning* [AM]), a IA pôde promover avanços significativos em várias frentes: dos cuidados de saúde, incluindo a prevenção de pandemias, o desenvolvimento de novos medicamentos e a otimização da gestão de estabelecimentos de saúde, ao suporte na tomada de decisões por médicos, enfermeiros e outros profissionais da área. Em cada uma dessas frentes, inúmeros sistemas de IA estão em uso, ou em fases avançadas de desenvolvimento e teste, a fim de auxiliar os agentes humanos a superarem tanto desafios antigos, como a realização de diagnósticos precisos, quanto novos desafios que surgem, como o aumento dos custos de saúde e o envelhecimento populacional.

Existe uma clara promessa de a IA transformar a saúde, mas sua implementação cuidadosa será fundamental para maximizar os benefícios e minimizar os riscos dessa adoção. Paralelamente ao crescimento das iniciativas para tornar a IA uma aliada no aprimoramento dos cuidados à saúde, crescem também as discussões sobre riscos éticos envolvidos. As preocupações incluem a proteção de dados pessoais, a formação adequada dos profissionais de saúde para o uso dos sistemas de IA e o receio de que os benefícios potenciais da tecnologia, como tratamentos personalizados, sejam apenas acessíveis às parcelas mais privilegiadas da sociedade. Nesse contexto, multiplicam-se debates regulatórios, guiados pela necessidade de estabelecer diretrizes que minimizem os riscos da IA na saúde, sem desestimular seu desenvolvimento e sua experimentação.

Há mais de 20 anos, o Núcleo de Informação e Coordenação do Ponto BR (NIC.br) tem atuado, em colaboração com diferentes atores da sociedade, para a promoção de uma Internet

aberta e interoperável, contribuindo para que a rede seja segura, inclusiva e de qualidade. Complementarmente, o NIC.br desenvolveu mecanismos eficazes de gestão de segurança da rede, e possui um portfólio diversificado de produtos e serviços voltados à melhoria contínua da Internet. Nesse sentido, o NIC.br, por meio do Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br), tem produzido dados e análises para uma compreensão ampla dos efeitos da adoção das tecnologias digitais no Brasil, incluindo a IA. Dessa forma, desempenha um papel ativo tanto no mapeamento do cenário da adoção da IA no setor da saúde, quanto na produção de evidências que auxiliem os tomadores de decisão a navegarem pelo setor. Entre as iniciativas do NIC.br com esses fins, destaca-se também o Observatório Brasileiro de Inteligência Artificial (OBIA), que atua como ponto focal no Brasil para o monitoramento e a análise da evolução e impacto da IA. O OBIA é resultado de uma ação estratégica inserida no contexto da Estratégia Brasileira de Inteligência Artificial (EBIA) e do Plano Brasileiro de Inteligência Artificial (PBIA), e tem como missão consolidar e disseminar o conhecimento sobre os impactos da IA na sociedade por meio da coleta e análise de dados sobre sua adoção e seu uso.

Os indicadores produzidos pelo Cetic.br destacam-se entre as atividades desenvolvidas pelo NIC.br, pois evidenciam os avanços positivos trazidos pela expansão da Internet e das tecnologias digitais no Brasil, e apontam os desafios que ainda precisam ser superados para que a população possa usufruir das oportunidades de forma significativa. Entre os indicadores produzidos pelo Cetic.br|NIC.br, estão os da pesquisa TIC Saúde, que, em suas edições mais recentes, passou a divulgar dados sobre a adoção da IA na saúde no Brasil.

Os dados divulgados pelo Cetic.br|NIC.br são fundamentados na multissetorialidade, desde o planejamento da metodologia até a construção dos instrumentos de coleta de dados, contando com a colaboração de especialistas de diversas áreas. A disseminação desses dados para a sociedade subsidia a elaboração de políticas e iniciativas de aprimoramento, tanto nas camadas técnicas, quanto nas de conteúdo, além de promover a ampliação dos instrumentos a serviço da população e garantir direitos e um acesso crítico, responsável, seguro e produtivo à Internet.

A presente publicação é mais um esforço do Cetic.br|NIC.br para compreender o cenário da IA na saúde, e como ela pode ser um instrumento em prol da sociedade. Ela reúne um mapeamento cuidadoso dos principais debates sobre IA e saúde na literatura, com destaque para questões relacionadas a transparência e a explicabilidade, além das perspectivas de pesquisadores, responsáveis por equipamentos de saúde, representantes do poder público e do mercado, e profissionais de saúde sobre os desafios, riscos e oportunidades da tecnologia. Além disso, apresenta uma série de considerações regulatórias publicadas pela Organização Mundial da Saúde (OMS) e pela União Internacional de Telecomunicações (UIT), bem como notas para a elaboração de políticas públicas referentes à IA na área da saúde no Brasil.

Boa leitura!

Demi Getschko
Núcleo de Informação e Coordenação do Ponto BR — NIC.br

PRÓLOGO

Inteligência Artificial e saúde

Heimar de Fatima Marin[1]

1 Professora titular (aposentada) da Universidade Federal de São Paulo (Unifesp), coordenadora científica da pesquisa TIC Saúde do Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br) do Núcleo de Informação e Coordenação do Ponto BR (NIC.br), editora-chefe do International Journal of Medical Informatics (IJMI) e presidente eleita da International Academy of Health Sciences Informatics (IAHSI) da International Medical Informatics Association (IMIA).

O termo “Inteligência Artificial” (IA), criado há mais de 65 anos, tem sido usado cada vez mais em todas as áreas de atividade humana, inclusive na saúde. O frequente uso desse termo mostra a expansão nos estudos expressando progressos e aplicações na chamada saúde digital, um termo mais recente utilizado pela Organização Mundial da Saúde (OMS) para substituir e complementar o conceito de *e-Health*. Assim, o termo saúde digital é utilizado para enfatizar a importância de envolvimento do usuário de um sistema de saúde, ampliando-lhe o acesso aos recursos disponíveis.

O termo IA foi amplamente utilizado desde o final do ano de 1956 e teve o seu ápice entre os anos de 1960 até 1990, com diversos estudos desenvolvidos e publicados nos periódicos científicos da área de saúde e de computação. Após esse período, houve uma retração nítida do uso de técnicas de IA na área da saúde, inferindo-se que a dificuldade maior era a integração desses sistemas de apoio à decisão no registro eletrônico do paciente, que, embora ainda em estágio inicial, já continha dados clínicos importantes para auxiliar o diagnóstico, como sinais vitais, sintomas e queixas dos pacientes em busca de atendimento. Ou seja, para utilizar um sistema de apoio baseado em conhecimento ou sistema especialista, o profissional precisava interromper a atividade que estava realizando (como a anamnese ou o exame físico do paciente) e consultar o sistema, que frequentemente operava de forma independente *stand-alone*. Após obter o resultado apresentado por ele, o profissional decidiria se incorporaria tal resultado no registro e seguiria o diagnóstico ou tratamento recomendado. Ou seja, geralmente, eram sistemas isolados que dependiam da interação humana para serem consultados; ao aceitar a recomendação, o profissional assumia a responsabilidade por ter aceitado o resultado apresentado pelo sistema.

Durante esse período, iniciei meus estudos na área de IA na saúde, por ocasião de minha dissertação de mestrado (1991) e tese de doutorado (1994), quando a IA era entendida como um ramo da informática que empregava programas de computadores capazes de desempenhar tarefas normalmente associadas ao comportamento humano inteligente. Por volta de 1985 até o início dos anos de 1990, algumas aplicações também se destacaram, como a construção de sistemas de apoio à decisão

clínica ou sistemas especialistas, e os chamados sistemas baseados em conhecimento, já presentes no comportamento e na decisão humana. Ou seja, eram sistemas que usavam o raciocínio para inferir conclusões de fatos armazenados. Assim, nada do que se construía poderia superar a capacidade humana de pensar, pois tudo o que continha era resultado da capacidade cognitiva humana.

Em termos de arquitetura básica, esses sistemas eram compostos por uma base de conhecimento, motor de inferência e a interface com o usuário. Na base de conhecimento, estavam inseridas as regras, os fatos, os conceitos e as definições do que se conhecia sobre determinado tópico até o momento de sua criação. O motor de inferência continha estratégias de busca nessa base para ajudar a resolver um problema, auxiliar em um diagnóstico, recomendar um tratamento e comunicar seus achados ao usuário por meio da interface.

É fácil perceber que o maior desafio estava na construção da base de conhecimento, já que tão melhor o sistema poderia auxiliar quanto melhor e mais atualizada ela estivesse. Assim, algumas metodologias foram desenvolvidas em busca de simular o pensamento humano; entre elas, os algoritmos, os *frames*, as redes semânticas, as redes neurais e até mesmo os métodos híbridos que utilizavam técnicas de programação orientada por objeto com regras de produção, por exemplo, *backward chaining* ou *forward chaining*, e regras como "se... então" (*if-then*), árvores de decisão, reconhecimento de padrões, sistemas baseados em teorema de Bayes e lógica *fuzzy*, para citar apenas o que na época era mais utilizado pelos pesquisadores. Muitos estudos utilizavam as regras em *backward-chaining*, em que o conhecimento era representado por pares que expressavam condição-ação. Desta forma, condição é uma expressão que pode ser estabelecida como verdade para a regra a ser aplicada, e a ação é uma lista de comandos executada se a condição for verdadeira. O princípio da ciência em refutar hipóteses guiava o processo.

Em consequência, alguns sistemas desenvolvidos no início dos anos de 1970 até 1990 tornaram-se referências mundiais e muito contribuíram para o avanço das implementações. Alguns exemplos emblemáticos foram o MYCIN (indicação de antibiótico), o INTERNIST-1 (utilizado para o ensino

de medicina interna, o qual, mais tarde, se tornou o Quick Medical Reference [QMR]), o Leeds (baseado em teorema de Bayes, auxiliava o diagnóstico de dor abdominal aguda) e o HELP Sistemas (assistia nas funções hospitalares em gestão de dados clínicos).

Dentre todos esses sistemas precursores, vale destacar a grande influência do sistema MYCIN, que demonstrava poder dada sua forma de representação e abordagem de raciocínio. Com o aumento dos novos conhecimentos, novas regras eram incorporadas sem a necessidade de alterar o código do sistema como um todo. Sistemas baseados em regras em muitos domínios não-médicos foram desenvolvidos nos anos que se seguiram à introdução da abordagem pelo MYCIN. Porém, é fato que, embora esses sistemas fossem fáceis de entender, tinham limitação ao serem aplicados em grandes volumes de dados.

Após o citado período de retração ocorrido de 2000 a 2018, com diminuição importante em publicações científicas na saúde, um novo ápice começou a surgir a partir de 2018, em função do enorme avanço na capacidade de armazenamento e processamento em bases massivas de dados e técnicas de aprendizado profundo. Com o crescimento exponencial do conhecimento e da maior adoção de tecnologias digitais na área de saúde, os sistemas baseados em regras começaram a representar um desafio tanto em seu desenvolvimento quanto em sua manutenção. Embora as primeiras utilizações empregando técnicas de aprendizagem de máquina (*machine learning* [AM]) tenham começado a ser utilizadas nos anos de 1960, hoje ganharam espaço significativo em todos os desenvolvimentos de IA e quase se tornaram um sinônimo desta. Os tópicos explorados na época ainda continuam sendo muito semelhantes ao que se busca atualmente, porém com maior ênfase em diagnóstico, tratamento e estudo de sobrevida de todos os tipos de câncer, saúde mental e doenças neurológicas.

Essa evolução tecnológica tem sido impulsionada pela explosão massiva de dados em saúde, o que possibilita a criação de modelos avançados de AM conhecidos como aprendizagem profunda (*deep learning* [AP]). Essa disponibilidade de grandes volumes de dados (desde que haja qualidade) aumentou a capacidade de resolver problemas não tão claramente descritos e definidos. Entretanto, continua crítico o fato de realizar a

identificação específica do problema que pode ser tratado pela base de IA disponível e os algoritmos a serem empregados que trazem significado clínico.

A partir de 2018, conforme mencionado, houve um imenso uso e desenvolvimento de aplicações usando IA e suas técnicas mais avançadas. Atualmente, o termo domina a maioria dos estudos, das pesquisas e mesmo das investigações sobre como tais recursos podem ser aplicados na área de saúde e em todos os segmentos de atividade humana. Ademais, a questão de integração está mais bem resolvida com recursos de interoperabilidade sintática (formato e ordem do que é trocado) e semântica (significado daquilo que é trocado entre os sistemas). Com isso, a resistência ao uso de sistemas inteligentes passa a ser facilmente mitigada.

Na área de pesquisa, muitos estudos continuam apenas realizando comparações entre modelos preditivos e algoritmos, testando quais apresentam melhor desempenho para um determinado conjunto de dados de saúde. Apesar de tais estudos fornecerem inferências diagnósticas, recomendações de tratamento ou expectativa de sobrevida em determinadas patologias, o mais importante não acontece: o número de pesquisas realizadas e publicadas que evidenciam avaliação de impacto e significado clínico ainda é incipiente. Os estudos frequentemente usam uma base de dados grande o suficiente para que possa ser dividida em partes, para que haja limpeza dos dados, seguida de uma parte de dados usados para realizar o modelo e outra parte menor para a realização de testes internos de validação. Todavia, estudos que testam o modelo ou algoritmo em bases externas para a realização de validação externa ainda são raros na literatura.

No ensino e no treinamento de novas gerações, a defasagem é ainda maior. Poucos programas de formação de profissionais incluem temas de informática em saúde e saúde digital como parte integrante das competências necessárias do profissional a ser formado. Temas de IA, aprendizagem de máquina e modelos preditivos similares são mais raros ainda, até mesmo pelo déficit de massa crítica apresentado pelo Brasil. Isso é crítico, porque precisamos de profissionais capazes de usar tais recursos, uma vez que o desafio permanece: um uso de IA com comportamento ético, evitando emprego de *bias* que podem aumentar discriminação na formação e na seleção das bases de dados para estudo

com significado clínico, e que auxilie o profissional no ambiente cada vez mais complexo de atendimento à saúde da população.

É preciso saber traduzir as promessas de IA para transformar os modelos de atendimento e os programas de saúde realizados pelos sistemas de saúde nacionais e internacionais, usando todo seu potencial. No mundo real, esses recursos precisam estar integrados aos registros de saúde do cidadão. É fundamental perguntar: "em benefício de quem a IA tem sido aplicada?" e "quais são as consequências de sua aplicação?". Além disso, é necessário sempre estar ciente de que a inteligência humana permanece no controle do processo.

Motivados pelo tema e pelo cenário de novas possibilidades de uso, esta publicação oferece uma reflexão relevante sobre o estágio de adoção de IA na saúde no Brasil, abordando a exploração dos avanços e desafios documentados na literatura, com diretrizes específicas para a aplicação de IA no contexto brasileiro. Destaca-se também o tema mais importante e sensível do momento: a necessidade de regulamentação adequada para garantir segurança, eficácia e ética no uso de IA na saúde.

Parte 1

ARTIGOS

Inteligência Artificial na saúde: uma visão da literatura e diretrizes para o Brasil

Rodrigo Brandão[1]

1 Doutorando em Sociologia pela Universidade de São Paulo (USP), mestre em Ciência Política pela USP, com graduação em Ciências Sociais pela USP. É pesquisador na Coordenação de Métodos Qualitativos e Estudos Setoriais do Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br) do Núcleo de Informação e Coordenação do Ponto BR (NIC.br).

As tecnologias de informação e comunicação (TIC) emergentes e as possibilidades decorrentes da automação podem trazer grandes benefícios para o setor de saúde em suas diversas etapas: prevenção, diagnóstico, atendimento e tratamento dos indivíduos, assim como na gestão dos sistemas e equipamentos de saúde, otimizando o trabalho dos profissionais, os recursos, as operações e as rotinas estabelecidas. Ademais, a Inteligência Artificial (IA), por meio da utilização de dados de saúde provenientes de múltiplas fontes — como prontuários eletrônicos, aquisição e armazenamento de imagens, análises de perfis genômicos e de outros dados fisiológicos —, tem forte potencial para auxiliar o enfrentamento de desafios do setor de saúde, como o aumento contínuo dos custos, a falta de profissionais, e as mudanças epidemiológicas e demográficas em curso, a exemplo do envelhecimento populacional.

O uso de ferramentas de IA também é uma aposta promissora para desenvolver e implementar políticas públicas em larga escala, bem como alavancar e acelerar a produção de conhecimento científico do setor por meio de pesquisas. Todavia, o otimismo com a tecnologia demanda cautela, dado que o desenvolvimento e emprego de aplicações de IA na saúde não são livres de riscos, como o vazamento de dados pessoais sensíveis de pacientes, o aumento da opacidade de diagnósticos, a redução da *accountability* em decisões médicas tomadas a partir de tecnologias que utilizam técnicas de IA e oferecem sugestões de diagnósticos, tratamentos e prognósticos, e mesmo a ampliação da desigualdade no acesso a tratamentos de saúde de qualidade.

Frente a oportunidades e riscos tão significativos quanto esses, cabe indagar: como aproveitar os possíveis benefícios e mitigar os potenciais danos? A fim de mapear estratégias capazes de fazê-lo, o presente capítulo revisa a literatura referente ao emprego da IA na saúde.

Na Seção "IA na saúde: um breve histórico e conceitos fundamentais", apresenta-se uma breve história da IA, destacando os usos dessa tecnologia na saúde ao longo do tempo. Discute-se também o sentido polissêmico do termo "Inteligência Artificial" e aponta-se que, na saúde, assim como em outras áreas, ele vem sendo tomado como sinônimo de aprendizagem de máquina

(*machine learning* [AM]) em anos recentes. Na Seção "Panorama atual de usos da IA na saúde: oportunidades, desafios e riscos", traça-se o panorama atual do emprego de aplicações de IA/AM na saúde, tendo como referência a administração de equipamentos de saúde e a prática clínica, a fim de se explorarem as oportunidades e os desafios para o avanço da IA, assim como os riscos que essa tecnologia pode trazer.

Na Seção "Recomendações para o desenvolvimento e uso responsável da IA: um balanço", as recomendações da Organização Mundial da Saúde (OMS) para o desenvolvimento e uso responsável da IA na saúde no mundo são abordadas. Na Seção "Diretrizes para o desenvolvimento e uso responsável da IA na saúde no Brasil: o papel da ESD", avalia-se o alinhamento entre tais recomendações e a Estratégia de Saúde Digital para o Brasil 2020-2028 (ESD) (Ministério da Saúde [MS], 2020), e se discute como essa estratégia estatal pode ajudar a estruturar o desenvolvimento da IA no país. Por fim, as considerações finais (Seção "Conclusão") são apresentadas a partir dos principais pontos analisados nas seções anteriores.

IA NA SAÚDE: UM BREVE HISTÓRICO E CONCEITOS FUNDAMENTAIS

Nos anos 1950, dois eventos marcaram o nascimento da IA: o matemático Alan Turing lançou o artigo "Computing Machinery and Intelligence", em 1950, no qual o Teste de Turing foi mencionado pela primeira vez; já em 1956, durante o *workshop* Dartmouth Summer Research Project on Artificial Intelligence, o cientista da computação John McCarthy e seus colegas cunharam o termo "Inteligência Artificial" a fim de descrever "a ciência e a engenharia de fazer máquinas inteligentes" (McCarthy, 1956 como citado em Berryhill *et al.*, 2019, p. 11). Desse modo, a IA foi consolidada como nova área de pesquisa na engenharia e na ciência da computação.

Desde então, as definições de IA multiplicaram-se, não havendo uma que seja padrão e globalmente aceita (Miailhe & Hodes, 2017). Atentos a isso, Krafft *et al.* (2020) procuraram mapear como diferentes públicos entendem a IA, a partir de dois grupos: pesquisadores de áreas técnicas de IA e formuladores de políticas públicas. Com base na comparação das definições utilizadas por cada um deles em diferentes

documentos, como artigos acadêmicos e textos regulatórios, constataram que os primeiros costumam recorrer a definições que enfatizam funcionalidades técnicas dos sistemas de IA, ao passo que, entre os formuladores de políticas públicas, as definições utilizadas tendem a se basear em comparações entre aplicações de IA e os modos humanos de pensar e se comportar.

Krafft *et al.* (2020) observaram ainda que as definições utilizadas pelos pesquisadores de áreas técnicas contemplam aplicações da IA já em uso, enquanto as definições dadas pelos formuladores de políticas públicas são mais alinhadas a tecnologias que, hipoteticamente, podem vir a ser criadas. A partir de dois termos comuns às discussões sobre IA, as definições utilizadas pelo primeiro grupo se alinham à ideia de Inteligência Artificial Estreita (IAE), enquanto as do segundo grupo expressam a concepção de Inteligência Artificial Geral (IAG). A IAE — chamada muitas vezes de IA Fraca — opera apenas nos cenários em que foi programada para fazê-lo, executando tarefas específicas; a IAG — conhecida como IA Forte —, por sua vez, consegue realizar atividades intelectuais, como abstrair e generalizar (Miailhe & Hodes, 2017).

Krafft *et al.* (2020) concluíram que a definição oferecida para IA pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE) é, entre as diferentes definições encontradas por eles, a que se situa melhor entre as visões desses dois grupos:

> [...] um sistema baseado em máquina que pode, para um conjunto de objetivos definidos por humanos, realizar predições, recomendações ou decisões, influenciando ambientes reais ou virtuais. Os sistemas de IA são projetados para operar com níveis variáveis de autonomia. (OCDE, 2024 como citado em Krafft *et al.*, 2020, p.6)[2]

2 A OMS (2021) adota essa definição em seu relatório *Ethics and governance of Artificial Intelligence for health – WHO guidance*. Contudo, cabe notar que, em 2023, a OCDE atualizou essa definição para a seguinte redação: "Sistema de IA: um sistema de IA é um sistema baseado em máquina que, para objetivos explícitos ou implícitos, infere, a partir da entrada que recebe, como gerar *outputs* como previsões, conteúdo, recomendações ou decisões que podem influenciar ambientes físicos ou virtuais. Diferentes sistemas de IA variam em seus níveis de autonomia e adaptabilidade após a implementação". A nova definição está disponível em: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

Seria limitante, no entanto, afirmar que a profusão conceitual no campo da IA deve-se apenas a divergências de entendimento entre atores de campos distintos. Para a preparação de seu livro-texto sobre IA, Russel e Norvig (2016 como citado em Krafft *et al.*, 2020) analisaram livros-texto de ciência da computação ligados ao tema publicados entre 1978 e 1993 e constataram que os pesquisadores da área se mostravam interessados em construir sistemas de quatro tipos: (a) que pensam como humanos; (b) que agem como humanos; (c) que pensam racionalmente; e (d) que agem racionalmente. Sweeney (2003 como citado em Krafft *et al.*, 2020), por sua vez, analisou os 996 trabalhos citados no livro-texto preparado por Russel e Norvig (2016) e concluiu que, em 987 deles, as atenções dos pesquisadores — e, portanto, as definições utilizadas por eles — estavam concentradas no desenvolvimento de sistemas que pensam e agem racionalmente, ou, como observam Krafft *et al.* (2020), que pensam e agem "idealmente".

Ainda que girem em torno de um eixo comum ("raciocínio e comportamento racionais/ideais") na grande área de IA, as definições para o termo central ("Inteligência Artificial") não poderiam se resumir a uma formulação única e completamente consensual entre técnicos da área, porque a IA, assim como todo e qualquer campo do conhecimento científico, tem subdivisões e as diferentes definições dadas ao termo central são um reflexo esperado dessa fragmentação. Entre as subdivisões da grande área de IA, estão as áreas de AM, de representação de conhecimento e raciocínio, e de sistemas multiagentes, entre outras.

Entre todas elas, as pesquisas acadêmicas e comerciais têm se concentrado na área de AM desde o início dos anos 2000 (Arbix, 2020). Antes disso, a IA passava por um período que se convencionou chamar de "inverno da IA", quando o financiamento e os interesses sociais e acadêmicos pela área escassearam de modo expressivo, o que "congelou" seu desenvolvimento (Daugherty & Wilson, 2018; Kaul *et al.*, 2020). Não há consenso sobre quais seriam as datas exatas de início e de encerramento dessa fase, mas, em geral, entende-se que ela se estendeu do meio dos anos 1970 ao começo dos anos 2000.

Como discutido adiante, a relação entre IA e saúde sempre foi próxima, mesmo em momentos em que o desenvolvimento caminhou a passos lentos, pois a medicina foi considerada desde os primórdios da IA como uma área promissora à aplicação

da tecnologia (Yu *et al.*, 2018).De acordo com Kaul *et al.* (2020), a primeira fase da IA estendeu-se dos anos 1950 a meados dos anos 1970, época em que se destacou o desenvolvimento de sistemas capazes de realizar inferências ou decisões que, até então, eram realizadas apenas por humanos. Na área médica, o principal avanço nesse período foi a digitalização de dados, com a criação de sistemas de registros médicos e de bases de dados de informática clínica, com o desenvolvimento, em 1960, do Medical Literature Analysis and Retrieval System e do PubMed pela Biblioteca Nacional de Medicina Americana.

Nos anos subsequentes, os resultados obtidos pelas engenharias e pela ciência da computação em IA ficaram aquém do esperado. Essa frustração converteu-se no que ficou conhecido como o "primeiro inverno da IA", que se estendeu até os primeiros anos da década de 1980. O interesse pela área recobrou força quando alguns pesquisadores de ciência da computação conseguiram equipar sistemas computacionais com uma lógica de raciocínio do tipo *if-then*. Avanços como esse não foram suficientes, contudo, para reestabelecer a força da área, pois os custos envolvidos eram altos, sobretudo para a manutenção de bancos de dados especializados. Além disso, um outro evento que ocorria na mesma época atraía a atenção de financiadores e pesquisadores, qual seja, o desenvolvimento do *desktop* (Daugherty & Wilson, 2018). Somados, esses fatores deram margem ao "segundo inverno da IA", que se encerrou por volta dos anos 2000.

Na área médica, todavia, a IA observou desenvolvimentos importantes durante esses anos, desde a criação de articulações acadêmicas para fortalecer a relação entre IA e medicina até o desenvolvimento de sistemas médicos centrados na tecnologia em questão. Entre esses últimos, havia sistemas de apoio à decisão, como o MYCIN e o DXplain, desenvolvidos, respectivamente, no início dos anos 1970 e no final dos anos 1980[3]. Em anos posteriores, o desenvolvimento da IA seguiu em movimento, mas foi a partir de 2010 que ganhou tração, devido a três fatores: (a) os bancos de dados pessoais cresciam com velocidade, graças a dinâmicas econômicas e sociais centradas

3 Desenvolvido pelo professor de Stanford Edward Shortliffe, o MYCIN destacou-se, entre os sistemas de apoio à decisão, devido às contribuições aportadas tanto à medicina, quanto à ciência da computação (van Melle, 1978).

em aplicativos; (b) a infraestrutura tecnológica estava desenvolvida para armazenar essas informações; e (c) o campo da estatística estava preparado para analisá-las (Arbix, 2020).

Assim como outras áreas da IA, a área de AM dedica-se a construir algoritmos, isso é, "[...] procedimentos codificados para transformar dados de entrada em uma saída desejada, com base em cálculos especificados" (Gillespie, 2014, p. 167). No caso do AM, os pesquisadores desenvolvem procedimentos para a identificação de padrões em grandes volumes de dados e trabalham para que os sistemas computacionais aprendam esses procedimentos. Tendo aprendido o padrão de uma imagem de câncer, por exemplo, a tecnologia passa a ser capaz de estimar a probabilidade de uma nova imagem ser correspondente ao padrão aprendido. O Quadro 1 apresenta os três tipos mais comuns de algoritmos de AM.

QUADRO 1 - PRINCIPAIS TIPOS DE ALGORITMOS DE AM

AM supervisionada: um tipo de tarefa de AM que visa prever o resultado desejado (como a presença ou ausência de retinopatia diabética) com base nos dados de entrada (como fotografias de fundo de olho). Os métodos supervisionados de AM funcionam a partir da correlação de entrada-saída na fase de "treinamento" e da correlação identificada para prever o resultado correto dos novos casos.
AM não supervisionada: um tipo de tarefa de AM que visa inferir padrões subjacentes em dados não rotulados. Por exemplo, ele pode encontrar subgrupos dos dados originais, identificar valores discrepantes nos dados ou produzir representações de baixa dimensão dos dados.
Aprendizagem profunda (AP): um subcampo da disciplina mais ampla de AM. A AP emprega redes neurais artificiais com muitas camadas para identificar padrões nos dados.

FONTE: ADAPTADO DE YU *ET AL.* (2018).

Segundo Topol (2019), os algoritmos de AP distinguem-se por seu autodidatismo. Em linhas gerais, expostos a um conjunto de imagens anotadas, como câncer e não-câncer, os algoritmos de aprendizagem supervisionada aprendem a reconhecer características comuns dessas imagens, por exemplo o volume de um nódulo. Formados por camadas que remetem às redes neuronais do cérebro humano, os algoritmos de AP são capazes de avaliar inúmeros aspectos das imagens analisadas, sem que os responsáveis por elaborá-los estejam seguros sobre quais aspectos os algoritmos levaram em consideração para a

geração de resultados. Por essa razão, costuma-se dizer que muitos algoritmos de AP se assemelham a verdadeiras "caixas pretas", pois não permitem saber com precisão como *inputs* são convertidos em *outputs*. Na área de saúde, esse componente traz desafios importantes em termos de decisões éticas, transparentes e explicáveis.

Em virtude de sua lógica de funcionamento, os algoritmos de AP têm capacidade superior à de seus pares para reconhecer padrões em grandes volumes de dados, evidenciando que a nova geração de sistemas de IA é bastante distinta de gerações anteriores, as quais dependiam de curadorias de conhecimentos de saúde feitas por especialistas e regras robustas de decisão (Yu *et al.*, 2018). Expressiva, essa mudança pode ganhar força no médio prazo, dado que os investimentos privados em IA na área da saúde têm sido consistentes. O documento *Artificial Intelligence Index Report 2023* (Stanford University Human-Centered Artificial Intelligence [HAI], 2023) registra que as pesquisas médicas foram o principal destino desses recursos em 2022: foram dedicados US$ 6,1 bilhões a elas, contra US$ 5,9 bilhões para pesquisas sobre administração, processamento de dados e nuvem, e US$ 5,5 bilhões para desenvolvimentos ligados a *fintechs*. Frente ao estreitamento das relações entre IA e saúde, cabe entender quais são as oportunidades, os desafios e os riscos nelas embutidos. A Seção "Panorama atual de usos da IA na saúde: oportunidades, desafios e riscos" é dedicada a esse mapeamento.

PANORAMA ATUAL DE USOS DA IA NA SAÚDE: OPORTUNIDADES, DESAFIOS E RISCOS

Inicialmente, buscaram-se informações sobre a relação entre IA e saúde em periódicos científicos específicos, como *Nature Biomedical Engineering* e *Nature Medicine*, e recorreu-se a exemplares de literatura cinzenta, como relatórios governamentais e análises produzidas por consultorias. Foram analisados, por exemplo, o *Technology assessment – Artificial Intelligence in health Care: Benefits and challenges of machine learning technologies for medical diagnostics*[4], publicado em

4 Saiba mais: https://www.gao.gov/products/gao-22-104629

2022 pelo United States Government Accountability Office e pela Academia Nacional de Medicina dos Estados Unidos, e o *Artificial Intelligence in healthcare – Application, risks, and ethical and societal impacts*[5], publicado pelo Parlamento Europeu também em 2022.

Concluída a primeira rodada do levantamento bibliográfico, quatro temas a serem investigados em profundidade foram identificados: (a) a utilização de técnicas de IA em pesquisas biomédicas, sobretudo para o desenvolvimento de novos medicamentos; (b) o uso de sistemas de IA/AM na prática clínica para o aprimoramento de diagnósticos, prognósticos e tratamentos; (c) o emprego de aplicações de IA/AM na administração de equipamentos de saúde; e (d) os desafios éticos e regulatórios associados à IA nessas três frentes. Na base de dados Scopus[6], buscaram-se documentos acadêmicos referentes aos três últimos tópicos entre os dias dez e 15 de maio de 2023. Para cada um deles, foi realizada uma busca específica composta por quatro etapas.

Nos três casos, as seguintes palavras-chave e operadores booleanos foram utilizadas no campo "Article title, Abstract, Keywords": "artificial intelligence" OR "machine learning" OR "deep learning" AND "health" OR "public health" OR "healthcare" OR "health care". Ainda na primeira etapa da busca, essas palavras-chave foram combinadas por meio do operador booleano "AND" a três conjuntos distintos de palavras-chave: (a) "clinical practice" OR "diagnostic decision support" OR "diagnostic decision" OR "clinical decision support" OR "clinical decision"; (b) "healthcare management" OR "healthcare management system"; e (c) "ethics" OR "regulation".

Em um segundo momento, foram aplicados três filtros aos resultados encontrados na etapa anterior: (a) "ano" — o período selecionado foi "2019-2023"; (b) "tipo de documento" — no caso da "prática clínica", foram selecionadas apenas revisões, ao passo que, nos campos "administração de equipamentos de saúde" e "ética e regulação", selecionaram-se tanto revisões quanto artigos; e (c) "idioma" — foram selecionados os idiomas inglês, espanhol e português.

5 Saiba mais: https://www.europarl.europa.eu/thinktank/pt/document/EPRS_STU(2022)729512
6 Saiba mais: https://www.scopus.com/home.uri

O terceiro passo foi ordenar os resultados da segunda etapa de dois modos distintos: (a) do mais ao menos citado e (b) do mais ao menos relevante. Na etapa seguinte, foram lidos os títulos e os resumos dos 20 primeiros documentos ranqueados — seja pelo número de citações, seja pela relevância. Alguns deles eram comuns às duas listas. Foram avaliados 113 documentos relacionados à prática clínica; 94, à administração de equipamentos de saúde; e 219 documentos sobre desafios éticos e regulatórios.

Foram adotados três critérios de exclusão, a fim de permitirem a seleção de materiais que, juntos, oferecessem uma visão geral dos benefícios, desafios e riscos da IA na área da saúde: (a) documentos que relatam a testagem de modelos de IA; (b) documentos que abordam os benefícios, desafios e riscos da IA de maneira especulativa; e (c) documentos centrados em um único país (com exceção dos poucos documentos sobre o Brasil) ou em uma única doença (com exceção de dois textos que discutem o papel da IA em desafios de gestão ligados à pandemia COVID-19). A utilização desses critérios culminou na seleção de 21 documentos sobre prática clínica; cinco, sobre administração de equipamentos de saúde; e 32 documentos sobre questões éticas e regulatórias (Anexo 1[7]).

Nas Subseções "O uso de sistemas de IA/AM na prática clínica" e "O uso de sistemas de IA/AM na administração de equipamentos de saúde", apresenta-se o balanço das leituras realizadas nas duas rodadas do levantamento bibliográfico sobre prática clínica e administração de equipamentos de saúde, respectivamente. As referências sobre desafios éticos e regulatórios, por sua vez, são apresentadas ao longo dessas duas subseções e da Seção "Recomendações para o desenvolvimento e uso responsável da IA: um balanço", juntamente

7 O Anexo 1 apresenta um quadro que sumariza as informações gerais sobre os estudos selecionados para leitura. Em relação à listagem, cabem duas observações. A primeira delas é que o trabalho de Pap & Oniga (2022) apareceu nos resultados das buscas sobre prática clínica e sobre ética e regulação. Após a leitura do resumo, ele se revelou relevante às duas áreas; por essa razão, o Anexo 1 apresenta 57 documentos, ao invés de 58. A segunda observação é que, após a leitura dos textos completos, alguns se revelaram mais pertinentes à discussão de um tópico distinto daquele que ensejou sua busca. Esse é o caso de Hobensack *et al.* (2023), por exemplo: embora o trabalho das autoras tenha sido identificado durante as buscas por documentos relacionados ao papel de sistemas de IA na prática clínica, ele se revelou mais relevante à discussão sobre oportunidades e desafios ligados à administração de equipamentos de saúde. Apesar disso, o quadro lista os textos de acordo com os três temas que embasaram as buscas no Scopus.

com as recomendações da OMS (2021) para desenvolvimento e uso responsável da IA.

O USO DE SISTEMAS DE IA/AM NA PRÁTICA CLÍNICA

De modo geral, os documentos revisados apresentam resultados de estudos em que algoritmos de IA/AM mostraram bom desempenho na interpretação de quadros clínicos em diferentes áreas médicas. Topol (2019) cita exemplos do gênero em radiologia, patologia, dermatologia, oftalmologia, cardiologia, gastroenterologia, e saúde mental. O autor observa, contudo, que a "validação da performance de um algoritmo em termos de acurácia não é equivalente a demonstrar eficácia clínica" (Topol, 2019, p. 45).

Apesar disso, ao menos nos Estados Unidos, os algoritmos de IA/AM podem estar chegando à prática clínica sem o devido escrutínio da comunidade acadêmica e de autoridades governamentais (Topol, 2019), porque muitos estudos sobre a eficácia desses algoritmos não são publicados em periódicos pautados pela revisão por pares. Além disso, o Food and Drugs Administration (FDA) facilitou, em anos recentes, o processo de aprovação de algoritmos médicos, e programas governamentais, como o *Medicare* e o *Medicaid*, passaram a reembolsar organizações que adotam sistemas de IA, enquanto, em outros setores, as instituições interessadas em utilizar a tecnologia devem pagar por ela (Sahni *et al.*, 2023).

Atentos ao movimento do FDA, Benjamens *et al.* (2020) mapearam os algoritmos de IA/AM e dispositivos médicos neles baseados aprovados pela agência estadunidense. Os autores avaliaram 7.390 anúncios oficiais do FDA sobre produtos aprovados pela agência e encontraram 64 algoritmos e dispositivos que poderiam ser baseados em técnicas de IA/AM. Em 29 casos, Benjamens *et al.* (2020) observaram que o anúncio oficial contém a informação de que o produto aprovado é baseado em técnicas de IA/AM; nos outros 35 casos, outras fontes de informação (que não os anúncios oficiais) indicam que os produtos aprovados são baseados em técnicas de IA/AM. Desse modo, os autores incluíram em seu banco de dados apenas o primeiro conjunto. Em relação a ele, apontam que

> [...] as duas principais especialidades médicas com inovações baseadas em IA/AM são Radiologia e Cardiologia, com 21 (72,4%) e 4 (13,8%) dispositivos médicos e algoritmos aprovados pelo FDA, respectivamente. O restante dos dispositivos médicos e algoritmos focam medicina interna/endocrinologia, neurologia, oftalmologia, medicina de emergência e oncologia. O campo médico da radiologia é o criador de tendências em relação a dispositivos médicos e algoritmos aprovados pelo FDA [...]. (Benjamens *et al.*, 2020, p. 2)[8]

O trabalho de Benjamens *et al.* (2020) também traz indicações de que a disseminação de algoritmos e dispositivos baseados em IA/AM na prática clínica tem ganhado destaque na comunidade científica. Os autores observaram que, em 2010 e 2011, o FDA não aprovou nenhum produto embarcado com IA/AM. Em 2012, houve duas aprovações. Esse número saltou para 22 em 2019, chegando a 64 em 2020[9]. A seguir, são apresentados 12 documentos do levantamento bibliográfico que trazem pistas sobre desafios e riscos associados à penetração da tecnologia em pontos de cuidado.

O primeiro deles é o trabalho de Topol (2019). O autor analisou 27 publicações revisadas por pares que comparam a performance de algoritmos de IA e de médicos na interpretação de dados clínicos em diferentes áreas. A principal conclusão do autor é que "o campo está claramente longe de demonstrar uma precisão de máquina muito alta e reprodutível [...] para a maioria dos exames e imagens médicas no ambiente clínico do mundo real" (Topol, 2019, p. 45). O autor chegou à conclusão semelhante ao analisar algoritmos de IA/

---

8 Foram encontrados poucos trabalhos que comparam diferentes áreas médicas, a fim de identificar em quais delas os sistemas de IA/AM são mais utilizados. Um deles é o de Tran *et al.* (2019), que identificaram o número de artigos acadêmicos sobre técnicas de IA publicados para um conjunto de 25 doenças.

9 Benjamens et al. (2020) não apresentam informações técnicas sobre os algoritmos e os dispositivos mapeados, entretanto Ngiam & Khor (2019) e Ahmed *et al.* (2020) suprem parcialmente essa lacuna. Os primeiros mapearam ensaios clínicos em oncologia voltados a algoritmos de AM. Ahmed *et al.* (2020) mapearam exemplos de algoritmos de AM utilizados em diferentes áreas da saúde e identificaram qual método de análise é utilizado em cada um deles, como AP, regressão logística e regressão linear.

AM voltados à predição de resultados clínicos[10]. Ele mapeou 13 relatórios sobre algoritmos de AM e AP para a predição de resultados diversos, como suicídios e taxa de mortalidade após tratamento quimioterápico. Ao analisá-los, observou desafios técnicos importantes, como a heterogeneidade das coortes estudadas e da faixa de acurácia. Por essas razões, o autor acredita que "ainda não se sabe até que ponto a IA pode prever os principais resultados no ambiente de saúde, e isso não será determinado até que haja uma validação robusta em ambientes clínicos prospectivos do mundo real, com metodologia e análise estatística rigorosas" (Topol, 2019, p. 49).

Ao menos no curto prazo, as conclusões de Topol (2019) podem contrastar com o potencial da IA/AM para aprimorar o trabalho clínico. Portanto, resta saber como os atores envolvidos com a adoção da tecnologia no ponto de cuidado encaram as oportunidades e os desafios associadas a ela, uma vez que suas percepções podem impactar tanto o desenvolvimento quanto a utilização dos sistemas de IA/AM voltados à saúde.

Os trabalhos de Yang *et al.* (2021), Hogg *et al.* (2023) e Aquino *et al.* (2023) ajudam a suprir, ainda que parcialmente, essa lacuna. Os dois primeiros são revisões de literatura. Yang *et al.* (2021) procuraram identificar qual é a percepção de diferentes atores quanto ao impacto da IA na radiologia. A principal conclusão dos autores é que há otimismo de clínicos, cirurgiões, estudantes e pacientes com a tecnologia, com a ressalva de que a concretização de seu potencial depende de educação e treinamento para os profissionais com ela envolvidos.

Hogg *et al.* (2023), por seu turno, investigaram quais são os fatores que influenciam a adoção de sistemas de IA na prática clínica. Eles focaram cinco grupos de atores: desenvolvedores; profissionais da saúde; lideranças e gerentes na área da saúde; pacientes, cuidadores e o público em geral; e reguladores e formuladores de políticas públicas. Destacam-se as conclusões dos autores sobre os profissionais da saúde[11]. Os autores observaram que, para esse público, três fatores impactam a implementação da tecnologia ora em tela: (a) sensação de

10 De modo geral, os documentos revisados focam a interpretação, e não a previsão, de resultados clínicos. Topol (2019) é um dos poucos autores a abordar os dois tópicos.

11 Asan *et al.* (2020) oferecem reflexões adicionais sobre os desafios encontrados por clínicos ao recorrerem a sistemas de IA/AM.

capacidade para entender os sistemas de IA; (b) percepção de como a tecnologia pode alterar a relação entre eles e seus pacientes; e (c) possibilidade de alinhamento das mudanças trazidas pelos sistemas de IA a condutas atuais de cuidado.

Finalmente, Aquino *et al.* (2023) buscaram entender se o uso de tecnologias de IA na saúde pode prejudicar a capacidade técnica dos profissionais da área. Para tanto, realizaram 72 entrevistas semiestruturadas com diferentes profissionais ligados ao desenvolvimento, à utilização e à regulação de sistemas de IA. Os autores encontraram duas visões contrastantes. "A visão utópica era que a IA poderia melhorar as competências e sistemas clínicos existentes, enquanto a visão distópica era que ela poderia levar à substituição de tarefas ou funções pela automação" (Aquino *et al.*, 2023, p. 5).

Outros oito trabalhos selecionados discutem o uso da IA na prática clínica tendo como referência os Sistemas de Apoio à Decisão (SAD), complementando, assim, os estudos de percepção de Yang *et al.* (2021), Hogg *et al.* (2023) e Aquino *et al.* (2023). Eles têm duas características comuns. A primeira é o foco em SAD existentes nos dias de hoje, isso é, em sistemas que — por serem baseados em IAE — apresentam sugestões de cursos de ação a seus usuários, em vez de tomarem decisões de maneira autônoma. A segunda característica comum a esses trabalhos é o foco em SAD que auxiliam profissionais de saúde no ponto de cuidado. Ou seja, eles não se aprofundam nem em decisões clínicas de caráter administrativo, como a priorização de pacientes, nem em SAD cujo usuário primário é o próprio paciente, como aplicativos de saúde móvel voltados ao monitoramento e ao controle da *diabetes mellitus* (El-Sappagh *et al.,* 2019). Em resumo, os SAD abordados correspondem aos algoritmos de assistência do Quadro 2.

QUADRO 2 - NÍVEIS DE AUTOMAÇÃO DE ALGORITMOS DE IA/AM NA SAÚDE

	ALGORITMOS ASSISTIVOS DE IA		ALGORITMOS AUTÔNOMOS DE IA		
	NÍVEL 1 APRESENTAÇÃO DOS DADOS	NÍVEL 2 SUPORTE À DECISÃO CLÍNICA	NÍVEL 3 AUTOMAÇÃO CONDICIONAL	NÍVEL 4 AUTOMAÇÃO ALTA	NÍVEL 5 AUTOMAÇÃO COMPLETA
Monitoramento de evento	IA	IA	IA	IA	IA
Execução da resposta	Clínico	Clínico e IA	IA	IA	IA
Contingência	Não se aplica	Clínico	IA, com um clínico de retaguarda a partir de requisição da IA	IA	IA
Especificidade de domínio, sistema e população	Baixa	Baixa	Baixa	Baixa	Alta
Responsabilização	Clínico	Clínico	Depende do caso	Desenvolvedor da IA	Desenvolvedor da IA
Exemplo	IA analisa um exame de mamografia e indica regiões de alto risco	IA analisa um exame de mamografia e oferece uma pontuação de risco que é interpretada pelo clínico	IA analisa um exame de mamografia e apresenta uma recomendação para biópsia, com um clínico sempre à disposição como backup	IA analisa um exame de mamografia e apresenta recomendação para biópsia, sem um clínico sempre à disposição como retaguarda	Igual ao nível 4, mas planejado para uso em todas as populações e sistemas

FONTE: BITTERMAN *ET AL.* (2020, COMO CITADO EM ADLER-MILSTEIN *ET AL.*, 2022).

Os oito trabalhos podem ser divididos em dois grupos. O primeiro deles é composto por três documentos que analisam a utilização de SAD em áreas médicas específicas. Encomendado pelo *British Medical Bulletin*, o trabalho de Bishara *et al.* (2022) analisou 13 exemplos de algoritmos de AM utilizados em ambientes de cuidado intensivo. Tendo esse universo como referência, os autores discorrem sobre sete obstáculos à implementação da tecnologia na realidade investigada: (a) segurança/*accountability*/responsabilidade legal;

(b) interpretação para os pacientes dos achados feitos pelos algoritmos de AM; (c) privacidade/anonimato; (d) ética/justiça/equidade; (e) disponibilidade e acesso a dados e generalização dos resultados; (f) autorização regulatória; e (g) desafios econômicos.

Moazemi *et al.* (2023) realizaram uma revisão sistemática de literatura sobre SAD em unidades cardiovasculares de cuidado intensivo. Os autores avaliaram 21 artigos acadêmicos e concluíram que muitos resultados descritos não haviam sido validados em bancos de dados externos, revelando, portanto, fragilidades na generalização desses resultados. Segundo os autores, esse desafio torna-se mais crítico à medida que os bancos de dados primam cada vez mais pela confidencialidade. Moazemi *et al.* (2023) concluíram também que a interpretabilidade dos resultados gerados pelos sistemas de IA/AM é essencial para que as equipes médicas confiem neles.

Também uma revisão sistemática de literatura, o trabalho de Du *et al.* (2023) analisou o uso de SAD baseados em AM em diferentes frentes do cuidado gestacional. Os autores observaram que os algoritmos do tipo "caixa preta" têm sido cada vez mais utilizados, além de os clínicos apresentarem opiniões positivas sobre os SAD que utilizam, apesar da falta de explicabilidade comum aos algoritmos em questão.

Os outros cinco trabalhos discutem desafios associados aos SAD sem se pautar em áreas médicas específicas. O primeiro deles é o trabalho de Sutton *et al.* (2020), que discorre sobre sete pares de oportunidades e desafios dos SAD. Os autores listam tópicos, como a segurança do paciente *versus* a fadiga de profissionais de saúde a alertas gerados pelos SAD, e o aumento da capacidade de interpretação de exames *versus* a necessidade de interoperabilidade de dados entre sistemas de diferentes instituições. Sutton *et al.* (2020) não analisam os SAD baseados em IA/AM de maneira aprofundada: apenas afirmam que eles podem ser aliados importantes na definição de diagnósticos.

Encomendado pelo Escritório de Prestação de Contas do Governo dos Estados Unidos (*United States Government Accountability Office*) e pela Academia Nacional de Medicina dos Estados Unidos, o estudo de Adler-Milstein *et al.* (2022) discute oito fatores que podem impactar a adoção de SAD para definição de diagnósticos. Entre eles, estão a existência de

incentivos financeiros à adoção desses sistemas, como reembolsos governamentais; a infraestrutura necessária para que possam funcionar de forma adequada, como a interoperabilidade de dados e o poder de processamento dos computadores utilizados; a qualidade da interface dos sistemas diagnósticos, de modo que eles não causem fadiga nos profissionais que os utilizam; e a confiança dos médicos nos resultados gerados, a qual, segunda as autoras, depende da explicabilidade desses mesmos resultados.

Amann *et al.* (2020) atêm-se à explicabilidade dos resultados sugeridos pelos SAD. As autoras procuraram entender a relevância desse elemento sob quatro perspectivas distintas: técnica, legal, médica e dos pacientes. A partir de uma revisão de literatura, concluíram que a utilização de SAD baseados em algoritmos opacos, do tipo "caixa preta", pode ter ao menos duas consequências nefastas: (a) relegar o paciente a uma posição de observador no processo de decisão médica; e (b) fazer com que os médicos tenham que se ater estritamente aos resultados gerados pelos sistemas em discussão, a fim de evitar sanções legais e de seus pares.

Já Albahri *et al.* (2023) avaliaram diversos elementos necessários à construção de algoritmos confiáveis de IA na área da saúde. Uma das conclusões dos autores é que há poucos bancos de dados de qualidade publicamente disponíveis para esse fim, o que aponta o impacto da qualidade dos dados na qualidade dos resultados dos sistemas de IA.

Finalmente, o quinto e último trabalho é uma revisão sistemática de literatura conduzida por Xu *et al.* (2023). Tendo como referência 20 artigos acadêmicos extraídos de 16 periódicos distintos, os autores investigaram a interpretabilidade de resultados gerados por SAD pelas perspectivas tecnológica e médica[12]. Entre as conclusões, destacaram o mapeamento de métodos de análise que asseguram a interpretabilidade e constataram que o conceito varia entre médicos e pacientes. Para os primeiros, elementos técnicos,

12 Seguindo a maioria dos trabalhos revisados, Xu *et al.* (2023) não propõem uma distinção clara entre interpretabilidade e explicabilidade. Os autores entendem, contudo, que os resultados gerados por um SAD precisam ser interpretáveis para que, em seguida, possam ser explicáveis. Para uma discussão em profundidade sobre a relação entre os dois termos, ver Markus *et al.* (2021).

como os biomarcadores e a qualidade dos dados utilizados nos SAD, têm relevância não observada entre os pacientes; para estes, a interpretabilidade está associada à promoção do consentimento informado e à facilitação da participação deles nos processos de tratamento.

Sem se ater aos SAD, outros trabalhos revisados discutem os desafios da explicabilidade. Nesse grupo estão, entre outras pesquisas, as de Markus *et al.* (2021), Loh *et al.* (2022), e Liu *et al.* (2022). A primeira delas mapeia e discute diferentes métodos para assegurar a explicabilidade de sistemas de IA. Loh *et al.* (2022), por seu turno, realizaram uma revisão sistemática de literatura baseada em 99 artigos publicados em periódicos acadêmicos classificados como Q1 no *ranking* SCImago Journal & Country Rank (SJR), a partir da qual identificaram quais técnicas de explicabilidade de IA são mais utilizadas, sem se restringir a doenças ou bancos de dados específicos. Já o trabalho de Liu *et al.* (2022) é um artigo de opinião em que os autores apresentam uma estrutura para a auditoria de algoritmos médicos baseados em IA/AM.

Como se pode perceber, os textos revisados trazem indicações de que os profissionais da saúde precisam confiar em sistemas de IA/AM para que os adotem, e que a explicabilidade é essencial para isso. Mais precisamente, os trabalhos analisados sugerem que médicos, enfermeiros e outros profissionais no ponto de cuidado precisam estar seguros de que não serão induzidos a erro pela tecnologia, o que poderia gerar danos aos pacientes e levá-los a enfrentar processos judiciais. Além disso, devem se sentir capazes de demonstrar a seus pacientes por quais razões eles (os profissionais da saúde) optaram por determinados diagnósticos ou cursos de tratamento em detrimento de outros. Esses dois conjuntos de preocupações dialogam com os dois elementos que, segundo Floridi *et al.* (2018), constituem a explicabilidade, a saber, inteligibilidade ("como os sistemas de IA funcionam?") e *accountability* ("quem é responsável pelo modo como os sistemas de IA funcionam?")[13].

Os materiais consultados exploram pouco, contudo, o modo como os diferentes atores ligados aos sistemas de IA/AM

13 Os estudos revisados dedicam mais atenção à inteligibilidade do que à *accountability*.

podem se coordenar para promover a confiança nos resultados gerados. Entre os profissionais de saúde, há dúvidas, por exemplo, sobre quem seriam os principais responsáveis por lidar com desafios impostos pelos algoritmos do tipo "caixa preta": se eles mesmos — por meio de investimentos em educação — ou se os desenvolvedores da tecnologia (Hogg *et al.*, 2023).

Há outras questões em aberto. Topol (2019) problematiza, por exemplo, o fato de os médicos precisarem confiar nos algoritmos de IA/AM para adotá-los ao mesmo tempo em que prescrevem medicações cujo mecanismo de ação não é conhecido. Nessa mesma linha, Markus *et al.* (2021) observam que outros elementos podem ser essenciais para tornar os algoritmos de IA confiáveis na área da saúde, como a publicização da qualidade dos dados utilizados e a extensiva validação externa dos resultados obtidos. Considerações como essas sugerem que o papel da explicabilidade — ou, nos termos de Xu *et al.* (2023), dos tipos de interpretabilidade — na construção da confiança de profissionais da saúde em sistemas de IA/AM demanda novas rodadas de investigações acadêmicas.

Em conjunto, os trabalhos apresentados permitem duas conclusões. A primeira é que a análise do desempenho de sistemas de IA/AM na prática clínica demanda investigações que considerem tanto fatores técnicos, como a qualidade dos bancos de dados, quanto fatores humanos, como a confiança nos resultados gerados. A segunda é que os desafios e riscos para a adoção desses sistemas são diversos e de cunho sociotécnico. A seguir, apresenta-se uma listagem com os cinco desafios e riscos identificados como prioritários nos materiais selecionados. Antes de fazê-lo, duas observações são necessárias.

A primeira consideração é que, de modo geral, os trabalhos revisados se referem a desafios e riscos éticos, uma vez que podem causar prejuízos a dignidade tanto dos profissionais da saúde quanto de seus pacientes — seja em termos legais, morais ou físicos. Apesar da referência constante ao termo "ética", poucos autores explicitam em qual referencial teórico se baseiam. Amann *et al.* (2020) são uma das poucas exceções: as autoras apontam que suas considerações sobre desafios e riscos da IA na saúde são ancoradas nos quatro princípios da bioética: beneficência, não-maleficência, autonomia e justiça (*fairness*).

Floridi *et al.* (2018) analisaram como esses quatro princípios aparecem em documentos sobre ética em IA de diferentes áreas que não apenas a saúde. De acordo com os autores, o primeiro princípio pode ser entendido como o imperativo moral "faça o bem". Nos documentos estudados por eles, a beneficência costuma estar relacionada ao avanço do bem comum. A "não-maleficência" — entendida como o imperativo moral "não faça o mal" — costuma estar associada a cuidados necessários para a proteção da privacidade. Já a "autonomia", observam Floridi *et al.* (2018), aponta para a necessidade de que os seres humanos detenham o poder de escolher em quais situações querem delegar poderes de decisão a sistemas inteligentes e em quais situações desejam revogar essa decisão. O princípio da "justiça", por sua vez, refere-se à utilização da IA para a promoção da prosperidade e para a preservação da solidariedade. Por fim, os autores apontam que a explicabilidade é um princípio que aparece de maneira difusa entre os documentos que analisaram, girando em torno de ideias de inteligibilidade e transparência e de *accountability*.

A segunda observação sobre os desafios e riscos identificados é que os termos "desafios" e "riscos" tendem a ser usados de maneira intercambiável, visto que desafios podem ser entendidos como riscos em potencial. Todavia, para tornar mais clara a visualização de desafios e riscos ligados à adoção de sistemas de IA/AM na prática clínica, enfatiza-se, a seguir, os desafios, procurando explicitar tanto os riscos aos quais dão margem quanto os princípios éticos com os quais dialogam.

Desafio 1 – Preservar a privacidade dos pacientes. O desenvolvimento e o funcionamento de sistemas de IA/AM demandam grandes volumes de dados; logo, o primeiro desafio é assegurar coletas massivas de dados que respeitem a privacidade individual. Lacunas nessa área podem se traduzir em prejuízo à autonomia dos pacientes (Bishara *et al.*, 2022). Entretanto, em nome da beneficência, a privacidade deve ser compatibilizada com coletas de dados que possam trazer melhorias à saúde individual e coletiva. Uma vez coletados, os dados de saúde precisam ser protegidos contra perigos diversos, como vazamentos, compartilhamentos não autorizados por seus titulares e usos potencialmente discriminatórios baseados em técnicas de perfilamento, como a cobrança de

mensalidades de planos de saúde mais altas de pacientes cujos dados pessoais sugiram prováveis condições adversas de saúde no futuro (Bishara *et al.,* 2022).

Desafio 2 – Assegurar a qualidade e a representatividade dos dados utilizados. Autores como Moazemi *et al.* (2023) apontam para a necessidade de validação dos resultados de algoritmos de IA/AM na saúde em bancos de dados diferentes daqueles em que foram treinados, aumentando, assim, as chances de generalização de seus resultados. Albahri *et al.* (2023), por sua vez, identificaram a falta de bancos de dados para essa finalidade que sejam públicos e de qualidade. Por fim, Challen *et al.* (2019) mapearam os riscos associados a problemas nos dados utilizados para o desenvolvimento de sistemas de IA/AM (elencados e descritos no Quadro 3).

Nesse contexto, faz-se importante registrar que, embora os dados de saúde possam ser de diferentes tipos, parte significativa deles corresponde a Registros Eletrônicos de Saúde (RES), os quais, segundo Bishara *et al.* (2022), costumam ser fragmentados e divididos entre diferentes sistemas, padecer de má-formatação e de informações faltantes, e compreender informações não-estruturadas, além de nem sempre refletirem com precisão a situação clínica a qual se referem. Por essa razão, os dados associados a RES demandam tratamentos técnicos cuidadosos antes de serem utilizados em sistemas de IA/AM.

Os trabalhos de Topol (2019), Challen *et al.* (2019) e Schwalbe & Wahl (2020) sugerem outros dois desafios que potencializam os riscos do Quadro 3. O primeiro deles é a falta de consenso sobre como reportar ou mesmo comparar a acurácia de sistemas de IA/AM. O segundo é a falta de testes em ambientes clínicos reais. Em parte, ambos se devem ao fato de que muitos experimentos utilizando IA/AM não são submetidos a revisões por pares.

Além de volumosos e tratados, os dados a serem utilizados nos sistemas em análise precisam ser representativos, sobretudo em termos de raça, gênero, orientação sexual e idade, porque não é socialmente desejável, de acordo com o princípio da justiça (*fairness*), que a tecnologia seja pródiga em identificar, por exemplo, melanomas em pessoas de pele clara, mas não em pessoas de pele escura. Giovanola & Tiribelli (2023) elencam e discutem resultados enviesados e discriminatórios

dessa natureza; tendo como referência pensadores como John Rawls, as pesquisadoras abordam ainda a possibilidade de os sistemas de IA/AM acentuarem desigualdades socioeconômicas no acesso e na utilização de serviços de saúde de qualidade.

QUADRO 3 - RISCOS ASSOCIADOS AO VOLUME E À QUALIDADE DOS DADOS

TEMA	RESUMO	EXEMPLO
CURTO PRAZO		
Mudança distributiva	Uma divergência entre os dados ou entre o ambiente em que o sistema é treinado e o ambiente em que o sistema é usado pode resultar em uma previsão errônea "fora da amostra", devido a vieses no conjunto de treinamento, mudanças ao longo do tempo ou usos do sistema em populações diferentes	A precisão de um sistema que prevê lesão renal aguda iminente com base em outros dados de registros de saúde diminui ao longo do tempo, à medida que os padrões da doença mudam
Insensibilidade ao impacto	Um sistema faz previsões que não levam em conta o impacto de previsões do tipo falso positivo ou falso negativo no contexto clínico em que é utilizado	Um sistema inseguro de diagnóstico é treinado para ser extremamente preciso no diagnóstico correto de lesões benignas, às custas de lesões malignas ocasionalmente ausentes.
Tomada de decisão do tipo "caixa preta"	As previsões de um sistema não estão abertas à inspeção ou interpretação e só podem ser julgadas como corretas com base no resultado final	Um sistema de IA de análise de raios-X pode ser impreciso em certos cenários devido a problemas com os dados de treinamento, mas, como uma caixa preta, isso não pode ser previsto e se tornará aparente após um período prolongado de uso
Modo inseguro de falha	Um sistema produz uma previsão mesmo sem qualquer confiança no resultado gerado ou quando não possui informações suficientes para gerá-la	Um sistema inseguro de IA para apoio à decisão pode prever um baixo risco de doença mesmo quando faltam alguns dados relevantes. Sem qualquer informação sobre a confiança da previsão, um médico pode não perceber a falta de confiança do resultado gerado
MÉDIO PRAZO		
Reforço de práticas obsoletas	Um sistema é treinado em dados históricos que reforçam a prática existente, e não pode se adaptar a novos desenvolvimentos ou mudanças repentinas nas diretrizes de atuação	Um medicamento é retirado de circulação por questões de segurança, mas o sistema de apoio à decisão de IA não pode se adaptar, pois não tem dados históricos sobre alternativas

Predição autorrealizável	A implementação de um sistema reforça indiretamente o resultado que foi projetado para detectar	Um sistema treinado em dados de resultados passados prevê que certos pacientes com câncer têm um prognóstico ruim. Isto faz com que recebam tratamento paliativo, ao invés de curativo, reforçando, assim, o comportamento aprendido
LONGO PRAZO		
Efeitos colaterais negativos	O sistema aprende a desempenhar uma função restrita que não leva em conta algum contexto mais amplo, criando uma consequência perigosa e não intencional	Um ventilador autônomo deriva uma estratégia de ventilação que mantém com sucesso a oxigenação a curto prazo às custas de danos pulmonares a longo prazo
Recompensas falhas	Uma *proxy* para o objetivo pretendido é usada como uma "recompensa" e um sistema de aprendizagem contínua encontra uma maneira inesperada de alcançar a recompensa sem cumprir o objetivo pretendido	Uma infusão autônoma de heparina encontra uma maneira de controlar o Tempo de Tromboplastina Parcial Ativada (TTPa) no momento de medição sem alcançar controle de longo prazo
Exploração insegura	Um sistema de aprendizagem ativa começa a aprender novas estratégias testando condições limítrofes de forma insegura	Uma infusão autônoma de heparina com aprendizado contínuo começa a usar doses em *bolus* perigosamente grandes para obter um rápido controle do TTPa

FONTE: ADAPTADO DE CHALLEN *ET AL.* (2019).

A OMS (2021) também chama atenção para a representatividade de dados de saúde de Países de Renda Média e Baixa (PRMB). De acordo com a instituição, o uso de sistemas de IA acontece sobretudo nos Estados Unidos e na União Europeia, o que pode impactar negativamente os PRMB de duas maneiras. A primeira delas é que os sistemas de IA/AM tendem a ser pouco expostos a dados de saúde desses países, por isso a tecnologia pode deixar de beneficiá-los. A segunda consequência é que as discussões sobre o desenvolvimento e o uso responsável de sistemas de IA/AM na saúde tendem a levar em consideração os mecanismos socio-legais existentes em países ricos. Nos PRMB, os mecanismos podem ser distintos, observa a OMS (2021); além disso, muitos países têm baixa capacidade regulatória na área da saúde. Por essas razões, afirma a instituição, não se sabe ainda quais abordagens os PRMB utilizarão para enfrentar os desafios e riscos ligados à IA e também não está claro como

procurarão atuar para que a exclusão digital, já bastante acentuada em alguns deles, não prejudique o potencial da IA para melhorar resultados na área da saúde.

Desafio 3 – Desenvolver sistemas interoperáveis. Em virtude das características dos RES, é comum que tecnologias médicas operem em contextos muito específicos. Os sistemas de IA/AM adicionam camadas de complexidade a essa realidade. Uma delas é a necessidade de exposição dos algoritmos de IA/AM a novos dados de qualidade para que possam ser atualizados, a qual pode fazer com que seus usuários se percebam em uma situação de dependência tecnológica, ficando à mercê dos desenvolvedores da tecnologia ou de fornecedores específicos de dados (Centro da Quarta Revolução Industrial do Brasil [C4IR], 2022). Além disso, o bom funcionamento de alguns algoritmos de IA/AM pode exigir dados oriundos de sistemas diferentes, demandando, assim, a criação de interfaces que os tornem interoperáveis. Esses dados, por sua vez, precisam ser coletados de alguma fonte geradora de dados, como um equipamento de eletrocardiograma; guardados, seja em um "lago de dados" ou em um "armazém de dados"; e processados.

Desafio 4 – Desenvolver sistemas atentos à experiência dos usuários. Os trabalhos de Sutton *et al.* (2020) e de Adler-Milstein *et al.* (2022) apontam que os SAD não podem gerar um número excessivo de alertas e que as informações geradas devem ser facilmente acessíveis. Outro risco é a falta de inteligibilidade e de interpretabilidade, que dá margem a riscos físicos aos pacientes e a riscos profissionais e legais a médicos, enfermeiros e outros trabalhadores da saúde.

Desafio 5 – Assegurar formação e treinamento aos profissionais no ponto de cuidado. Médicos, enfermeiros e outros profissionais no ponto de cuidado não são os principais responsáveis por resolver os desafios apontados. Todavia, caso estejam preparados para interagir com sistemas de IA/AM, terão chances maiores não só de identificar situações de perigo para eles próprios ou para seus pacientes, mas também de fazer usos da tecnologia que aprimorem suas capacidades profissionais, beneficiando, assim, os pacientes atendidos. Em outros termos, a oferta de formação e treinamento aos profissionais no ponto de cuidado é essencial ao fortalecimento da supervisão humana sobre a tecnologia, de modo

que a experiência e o julgamento humanos sejam integrados ao funcionamento dos sistemas de IA. Conhecida como *human in the loop*, essa fórmula não é suficiente para sanar os riscos de sistemas de IA/AM, uma vez que ela própria pode apresentar falhas (Quadro 4). Ainda assim, pode ajudar a controlar os riscos do Quadro 3.

À guisa de conclusão, reforça-se que, na presente revisão bibliográfica, a existência de explicabilidade revelou-se um desafio crucial para que os sistemas de IA/AM possam ser considerados transparentes e, portanto, confiáveis. Na lista de desafios apresentada, ele se desdobra entre o "Desafio 2 – Assegurar a qualidade e a representatividade dos dados utilizados" e o "Desafio 4 – Desenvolver sistemas atentos à experiência dos usuários".

QUADRO 4 - PROBLEMAS RELACIONADOS À SUPERVISÃO HUMANA SOBRE SISTEMAS DE IA

VIÉS DE AUTOMAÇÃO	A confiança excessiva nos resultados gerados pelos algoritmos pode tornar inútil a supervisão humana e a atuação do *human-in-the-loop*, exigindo estratégias de mitigação e mecanismos adicionais de monitoramento da solução de IA.
COMPLACÊNCIA DE AUTOMAÇÃO	A preferência deliberada por "falsos positivos" ou "falsos negativos" nos *outputs* gerados por soluções de IA pode ser danosa sob o ponto de vista ético, exigindo justificativa adequada não só na etapa de desenvolvimento, mas também ao longo das fases de treinamento e execução.
COMPENSAÇÃO POR VIÉS CONHECIDO	A supervisão humana pode compensar em demasia os erros e vieses já percebidos nos resultados dos algoritmos, criando novas distorções que afetam o algoritmo sem efetiva correspondência nos bancos de dados.

FONTE: MULLIGAN & BAMBERGER (2019); RUBENSTEIN (2021, COMO CITADO EM C4IR, 2022).

O USO DE SISTEMAS DE IA/AM NA ADMINISTRAÇÃO DE EQUIPAMENTOS DE SAÚDE

Os documentos sobre o uso de sistemas de IA/AM na administração de equipamentos de saúde focam dois temas: (a) o aprimoramento de atividades administrativas e fluxos de trabalho, como o agendamento de consultas, a otimização de leitos e de salas de cirurgia, e a evolução de pacientes ao longo das diferentes etapas do processo de cuidado; e (b) a

utilização de dispositivos eletrônicos, como os dispositivos vestíveis, para a criação de modelos de cuidado centrados nos próprios pacientes.

Nas discussões sobre o primeiro tópico, um estudo a ser destacado é o de Sahni *et al.* (2023), os quais calcularam o volume de gastos em saúde que poderia ser economizado anualmente nos Estados Unidos até 2028 se tecnologias de IA existentes nos dias de hoje estivessem disseminadas entre hospitais, médicos e clientes particulares. A conclusão é que a economia seria de 5% a 10% dos gastos atuais ou, em valores de 2019, de US$ 200 bilhões a US$ 360 bilhões. Diante desse montante em potencial, os autores — que são ligados à Universidade de Harvard e à consultoria McKinsey — concluem que as tecnologias de IA não são tão utilizadas quanto poderiam e elencam dificuldades gerencias que poderiam explicar esse fenômeno, divididas entre dois grupos. O primeiro deles é formado por desafios enfrentados pelas organizações do setor de saúde de maneira individual, enquanto o segundo compraz dificuldades que, para serem superadas, demandam esforços de diferentes atores no campo da saúde. O Quadro 5 elenca esses desafios, apresentando de maneira concisa as considerações de Sahni *et al.* (2023) sobre eles.

QUADRO 5 - DESAFIOS À ADOÇÃO DE SISTEMAS DE IA/AM EM EQUIPAMENTOS DE SAÚDE

DESAFIOS DE GESTÃO (INTRAORGANIZACIONAIS)
Deixar claro como os sistemas agregarão valor às atividades da organização.
Atrair ou desenvolver profissionais com as habilidades necessárias à utilização dos sistemas.
Assegurar aos profissionais no ponto de cuidado que a adoção dos sistemas não lesará os pacientes.
Realizar investimentos em fatores críticos ao funcionamento dos sistemas, como bancos de dados.
Assegurar, desde o início da utilização dos sistemas, o gerenciamento dos dados utilizados, a fim de superar problemas, como a falta de interoperabilidade, a fragmentação e a preservação da privacidade.
Criar modelos de governança e de processo para a utilização dos sistemas.
DESAFIOS SETORIAIS (INTERORGANIZACIONAIS)
Diminuir a heterogeneidade dos RES.
Aumentar a confiança dos pacientes nos resultados gerados pelos sistemas.

Assegurar a exposição dos sistemas a novos conjuntos de dados, a fim de mantê-los atualizados.
Definir se o tempo "economizado" será empregado em novos atendimentos ou em atividades não-clínicas de trabalho, a fim de que os sistemas aliviem a carga de trabalho dos profissionais no ponto de cuidado.
Assegurar que os sistemas tenham aprovação regulatória.

FONTE: ELABORAÇÃO PRÓPRIA COM BASE EM SAHNI *ET AL.* (2023).

Embora os custos dos sistemas de IA/AM também possam ser um impeditivo à adoção da tecnologia por equipamentos de saúde, há poucas estimativas de custos e de economia sobre o uso da IA na saúde, sendo que as existentes focam elementos específicos. De acordo com Adler-Milstein *et al.* (2022), as informações atuais indicam que o custo global de desenvolvimento e implementação de sistemas de IA/AM varia entre US$ 15 mil e US$ 1 milhão. As autoras apontam ainda que

> [...] outro desafio é a tensão entre contratar uma empresa de tecnologia de cuidados de saúde para desenvolver ou adaptar os algoritmos e ferramentas num ambiente de cuidados de saúde *versus* contratar e apoiar pessoal interno, o que pode custar entre US$ 600 e US$ 1.550 por dia. (Adler-Milstein *et al.,* 2022, p. 52-53)

Já a utilização de dispositivos eletrônicos, como os dispositivos vestíveis, para a criação de modelos de cuidado centrados nos próprios pacientes é abordada por Topol (2019) e Xie *et al.* (2021). O primeiro deles observa que o desenvolvimento de algoritmos de AP tem privilegiado usos pelos profissionais de saúde, deixando em segundo plano a utilização pelos próprios pacientes[14]. Já Xie *et al.* (2021) discutem como tecnologias de IA e *blockchain* podem ser integradas a dispositivos vestíveis no controle de doenças crônicas. Ainda que defendam que soluções tecnológicas desse tipo podem melhorar o bem-estar individual, os autores não deixam de reconhecer que elas estão associadas a diferentes desafios técnicos e sociais, como

14 Junaid *et al.* (2022) mapearam diferentes dispositivos vestíveis que utilizam técnicas de AM.

a segurança das taxas de acurácia e a interpretabilidade dos resultados, a interoperabilidade de dados necessária para que a tecnologia funcione adequadamente, a proteção dos usuários e o preço da tecnologia.

Nenhum dos dois trabalhos discute o risco identificado por Challen *et al.* (2019): a supervisão não-escalonável (*unscalable oversight*), definida pelos autores como a necessidade de atenção constante do usuário ao sistema de IA/AM. Como exemplo, citam bombas de insulina subcutânea autônomas que exigem do paciente informações exaustivas sobre sua alimentação para que consigam ajustar corrretamente o nível da substância a ser administrado antes de cada refeição.

Hobensack *et al.* (2023), por sua vez, revisaram a literatura sobre a utilização de técnicas de AM em dados eletrônicos de saúde gerados em contextos de *home healthcare* para a predição de resultados adversos, como hospitalização. Tanto esse trabalho quanto o de Xie *et al.* (2021) trazem menções à necessidade de se integrar os dados produzidos por dispositivos centrados nos pacientes ao funcionamento dos sistemas de saúde. Não exploram, todavia, as dificuldades para fazê-lo.

Caso não sejam endereçados, os desafios intra e interorganizacionais apresentados poderão manter os sistemas de IA/AM subutilizados na saúde, deixando de contribuir para a redução de custos no setor e, consequentemente, para a ampliação do acesso a serviços de qualidade.

CONSOLIDAÇÃO DOS RESULTADOS ENCONTRADOS NA REVISÃO DE LITERATURA

Os trabalhos discutidos indicam que inúmeros desafios sociotécnicos precisam ser superados para que as tecnologias de IA/AM possam ser úteis à promoção de tratamentos de saúde preditivos, preventivos, personalizados e participativos. Ignorá-los pode nos conduzir a situações extremas, seja a um novo "Inverno da IA", no qual a tecnologia será subaproveitada graças a temores sociais, seja a um cenário em que os sistemas de IA/AM serão utilizados de maneira indesejável, isso é, sem qualquer respeito a médicos e pacientes ou sem que se saiba quando e como eles são úteis aos processos de cuidado.

Procurando evitar quaisquer um desses dois desfechos, inúmeros pesquisadores e instituições propõem recomendações

para o desenvolvimento e o uso responsável da IA na saúde. Na revisão de literatura, foram identificadas inúmeras sugestões do gênero. Essa constatação encontra eco no trabalho de Jobin *et al.* (2019, como citado em Goirand, 2021): os pesquisadores averiguaram que, em anos recentes, diferentes atores ligados ao desenvolvimento e à utilização de tecnologias de IA publicaram, em conjunto, pelo menos 84 arcabouços de ética. A expansão de publicações desse tipo não significa, contudo, que suas recomendações estejam sendo internalizadas pelos atores da saúde. Evidência disso pode ser encontrada no trabalho de Goirand *et al.* (2021), no qual analisaram 33 documentos acadêmicos e não-acadêmicos publicados entre 2015 e 2020 sobre a implantação de arcabouços de ética em aplicações de IA na saúde. Uma das principais conclusões é que apenas oito documentos mencionam algum arcabouço específico de ética em IA.

As autoras afirmam que o resultado observado se deve, entre outros fatores, a dificuldades de implantação de um arcabouço de ética a situações específicas em que os sistemas de IA são utilizados. Essa constatação sinaliza que o enfrentamento dos obstáculos e a minimização dos riscos associados a tais sistemas depende, ainda que parcialmente, de esforços institucionais individuais. É certo, todavia, que o sucesso desses esforços depende de ao menos dois fatores: (a) a familiarização dos inúmeros atores ligados aos sistemas de IA na saúde com diretrizes para o desenvolvimento e o uso responsável dessa tecnologia; e (b) a existência de políticas públicas preocupadas em promover essa familiarização e em facilitar cooperações institucionais em torno das diretrizes em questão. Nas Seções "Recomendações para o desenvolvimento e uso responsável da IA: um balanço" e "Diretrizes para o desenvolvimento e uso responsável da IA na saúde no Brasil: o papel da ESD", discute-se como as recomendações da OMS (2021) podem ser úteis à formulação de políticas públicas voltadas a esses dois objetivos, debate especialmente relevante para o Brasil, pois as reflexões sobre a relação entre IA e saúde ainda estão em formação no país: no universo bibliográfico que embasa este capítulo, há poucos trabalhos referentes ao Brasil[15].

15 Dos trabalhos referentes ao Brasil, destacam-se os de Dourado & Aith (2022) e Nunes *et al.* (2022).

Apesar de o termo "ética" ser recorrente nos trabalhos revisados, optou-se por discutir o desenvolvimento e o uso "responsável" da IA na Seção "Recomendações para o desenvolvimento e uso responsável da IA: um balanço". Essa diferenciação semântica deve-se ao entendimento de que o segundo termo dá igual destaque aos desafios éticos e administrativos abordados, respectivamente, nas Subseções "O uso de sistemas de IA/AM na prática clínica" e "O uso de sistemas de IA/AM na administração de equipamentos de saúde".

RECOMENDAÇÕES PARA O DESENVOLVIMENTO E USO RESPONSÁVEL DA IA: UM BALANÇO

As recomendações da OMS (2021) para o desenvolvimento e o uso responsável da IA dialogam com o desafio geral identificado na seção anterior, qual seja, elaborar políticas públicas capazes de facilitar e estimular uma relação positiva entre IA e saúde. Por essa razão, apresenta-se, nesta seção, um balanço geral das recomendações feitas pela instituição, complementando-as com recomendações de outras instituições e autores identificadas na revisão bibliográfica. Na Seção "Diretrizes para o desenvolvimento e uso responsável da IA na saúde no Brasil: o papel da ESD", é feita uma comparação entre essas recomendações e o conteúdo da ESD, a fim de que diretrizes para o desenvolvimento das políticas públicas em questão possam ser identificadas.

RECOMENDAÇÕES DA OMS PARA O DESENVOLVIMENTO E USO RESPONSÁVEL DA IA

Publicado em 2021, o documento *Ethics and Governance of Artificial Intelligence for Health – WHO Guidance* (OMS, 2021) apresenta dois conjuntos de recomendações para o desenvolvimento e uso responsável de tecnologias de IA na saúde. O primeiro deles compreende elementos para a criação de arcabouços nacionais e internacionais de governança da IA capazes de proporcionar a essa tecnologia funcionar como uma aliada na construção da cobertura universal em saúde. O segundo conjunto de recomendações é voltado a desenvolvedores de sistemas de IA, ministérios de saúde, e instituições de saúde, e compreende orientações práticas para que essas três partes interessadas enfrentem os desafios e riscos da IA em saúde.

Em um esforço de síntese, pode-se dizer que os dois conjuntos de recomendações têm três elementos comuns. O primeiro deles é a proteção à privacidade, expressa nas ideias de privacidade desde a concepção *(privacy by design)* e privacidade por padrão *(privacy by default)*. Operacionalmente, elas significam que todos os esforços possíveis devem ser feitos para assegurar a privacidade e a confidencialidade das informações utilizadas no desenvolvimento, na validação e na utilização de tecnologias de IA.

A relevância de orientações públicas e privadas e de regulações oficiais que orientem auditorias e avaliações de risco é o segundo elemento que perpassa as diferentes considerações da OMS (2021). A instituição destaca a importância de legislações oficiais, como a *General Data Protection Regulation*, e de padrões internacionais e códigos de boas práticas, como os padrões ISO, as orientações do estadunidense National Institute of Standards and Technology, a série IEEE 7000, e a Health Level 7. A OMS (2021) observa que os quatro padrões são úteis à promoção de *compliance* e à minimização de desafios ligados à interoperabilidade, sendo que os três primeiros se referem à proteção da privacidade, e o quarto, à transferência de dados de saúde clínicos e administrativos.

Por fim, as recomendações da instituição deixam evidente que, tanto no desenvolvimento quanto na utilização de sistemas de IA/AM na saúde, articulações entre diferentes partes interessadas empenhadas em viabilizar avaliações do tipo *bottom-up* é fundamental à minimização de riscos e à potencialização de benefícios, sendo mais influentes do que avaliações conduzidas exclusivamente por autoridades públicas.

A seguir, os desafios e riscos discutidos nas Subseções "O uso de sistemas de IA/AM na prática clínica" e "O uso de sistemas de IA/AM na administração de equipamentos de saúde" são recuperados, e algumas sugestões específicas da OMS (2021) para enfrentá-los são apresentadas. As recomendações referentes à explicabilidade desdobram-se entre os itens "Assegurar a qualidade e a representatividade dos dados utilizados" e "Desenvolver sistemas atentos à experiência dos usuários".

Preservar a privacidade dos pacientes. O consentimento é um dos principais mecanismos para promover a proteção da privacidade e da confidencialidade. À luz da Lei Geral de

Proteção de Dados Pessoais (LGPD, 2018), ele pode ser entendido como uma manifestação livre, informada e inequívoca pela qual o indivíduo concorda com a coleta e o tratamento de seus dados pessoais para uma finalidade determinada. Todavia, inúmeros obstáculos têm desafiado seu exercício, como a alta frequência com que dados pessoais têm sido coletados, sobretudo na área da saúde, e a clareza dos termos de consentimento. Ciente desses desafios, a OMS (2021) cita três estratégias para assegurar o consentimento: (a) o consentimento informado eletrônico; (b) o consentimento dinâmico; e (c) o consentimento amplo.

Em situações específicas, o consentimento pode ser um obstáculo à realização de benefícios sociais e coletivos. Nesses casos, cabe às entidades governamentais não apenas esclarecer por que razões o consentimento é contrário ao interesse público, mas também articular mecanismos que assegurem compartilhamentos seguros de dados de saúde entre diferentes partes interessadas. A OMS (2021) menciona, por exemplo, o *hub* de dados do Departamento de Assuntos dos Veteranos e o Precision Medicine Initiative (*All of Us*) – ambos nos Estados Unidos.

A instituição aborda, ainda, estratégias para a preservação da privacidade e da confidencialidade quando o consentimento se mostra incapaz de fazê-lo, como a anonimização de dados de saúde e o uso de sistemas federados de dados, os quais permitem a diferentes instituições aplicar um mesmo modelo de AM a seus bancos de dados e comparar o que ele aprendeu em diferentes contextos. Dessa forma, os dados podem ser preservados nos locais em que estão, sem prejudicar o desenvolvimento da tecnologia[16]. Por essa razão, organismos internacionais, como o Fórum Econômico Mundial, consideram tais sistemas promissores (OMS, 2021).

Assegurar a qualidade e a representatividade dos dados utilizados. Os riscos descritos no Quadro 3 estão associados à programação dos sistemas de IA/AM e são tão mais prováveis quanto pior for a qualidade dos dados em que se baseiam. Por essa razão, é imperativo que avaliações de risco sejam

16 Rahman *et al.* (2022) analisam o aprendizado federado.

feitas em todas as etapas de desenvolvimento da tecnologia e aconteçam de maneira regular quando a tecnologia já estiver em uso. Além disso, a OMS (2021) afirma que "as tecnologias de IA devem ser testadas prospectivamente em ensaios randomizados e não contra conjuntos de dados laboratoriais existentes" (OMS, 2021, p. 141). A organização também observa que agências reguladoras podem contribuir não apenas para que essas avaliações aconteçam, mas também para que seus resultados sejam comunicados de forma clara.

Em relação à representatividade dos dados utilizados, a OMS (2021) apresenta uma série de recomendações bastante específicas, sobretudo para os desenvolvedores de sistemas de IA/AM. Em geral, elas giram em torno de um eixo comum: "examinar os efeitos de etnia, idade, raça, gênero e outras características, e garantir que as tecnologias de IA com resultados enviesados não tenham impactos negativos sobre indivíduos e grupos por conta dessas diferentes características" (OMS, 2021, p. 137)[17].

Por fim, os desafios para viabilizar a explicabilidade perpassam o "Desafio 2 – Assegurar a qualidade e a representatividade dos dados utilizados" e o "Desafio 4 – Desenvolver sistemas atentos à experiência dos usuários". De acordo com a OMS (2021, p. 141):

> [...] as regras de responsabilidade utilizadas nos cuidados clínicos e na medicina devem ser modificadas para avaliar e atribuir responsabilidade, incluindo a responsabilidade pelo produto, a responsabilidade pessoal dos decisores, a responsabilidade pelos insumos e a responsabilidade para com os fornecedores de dados. As regras devem incluir a responsabilidade causal, regimes objetivos de responsabilidade e responsabilidade por danos retroativos, bem como mecanismos para atribuir responsabilidade indireta, quando apropriado.

17 As considerações da OMS (2021) sobre resultados enviesados sugerem que vieses algorítmicos são negativos quando há implicações sociais deletérias.

Definições como essas demandam a coordenação dos ministérios de saúde.

Assegurar a infraestrutura (de dados) necessária ao bom funcionamento da tecnologia. A OMS (2021) sugere aos ministérios de saúde que avaliem se as estruturas de saúde existentes em seus países são suficientes para a operação, a manutenção e a supervisão dos sistemas de IA/AM. Caso não sejam, a organização aponta que a construção de articulações com a sociedade civil e com organismos internacionais pode ser essencial para aprimorá-las. Em relação à infraestrutura de dados, especificamente, vale destacar as recomendações de C4IR (2022): a exigência de licenças abertas e o uso de *softwares* livres podem evitar a dependência tecnológica a fornecedores específicos.

Desenvolver sistemas atentos à experiência dos usuários. A inteligibilidade dos *outputs* de sistemas de IA/AM é um dos principais desafios na prática clínica por duas razões. Primeiramente, os profissionais no ponto de cuidado precisam confiar nos resultados gerados pela tecnologia; para isso, precisam estar seguros de que podem interpretá-los corretamente. Em segundo lugar, os profissionais ora em tela precisam conseguir explicar aos pacientes como os resultados gerados pela tecnologia informaram suas condutas clínicas; logo, é imprescindível que os desenvolvedores de sistemas de IA/AM procurem assegurar tecnicamente a inteligibilidade, a fim de viabilizar a interpretabilidade, e a apresentação objetiva dos resultados gerados pela tecnologia. Para que tenham sucesso nessa dupla missão, a OMS (2021) recomenda a eles que esses desenvolvedores engajem diferentes partes interessadas no desenvolvimento dos sistemas de IA/AM e procurem entender os contextos nos quais serão utilizados, pois as tecnologias de IA voltadas à saúde são dependentes do contexto em que são empregadas.

Assegurar formação e treinamento aos profissionais no ponto de cuidado. No universo bibliográfico que embasa este capítulo, existem diferentes menções a esse desafio, apesar de não serem acompanhadas por indicações específicas de como os profissionais da saúde poderiam ser formados para que aprendam a interagir com sistemas de IA/AM. Tampouco há sinalizações sobre as competências

específicas que precisariam adquirir para que estejam aptos a utilizar os recursos tecnológicos em prol da melhoria do atendimento ao cidadão.

Tornar os sistemas de IA/AM financeiramente acessíveis. Os trabalhos revisados dedicam pouca atenção ao fato de que os custos dos sistemas de IA podem trazer prejuízos tanto à igualdade de oportunidades, acentuando, assim, desigualdades socioeconômicas, quanto ao potencial dos sistemas de IA/AM para reduzir os custos coletivos dos serviços de saúde.

Definir como a tecnologia pode aprimorar o trabalho clínico e administrativo. Entre os desafios de gestão do Quadro 5, um deles volta-se ao caráter acentuadamente administrativo, qual seja, apontar como os sistemas de IA/AM podem agregar valor às atividades das organizações que pretendem empregá-los. Essa missão pode ser conduzida por lideranças organizacionais que devem avaliar, para cada situação clínica e administrativa, se o uso de sistemas de IA/AM é necessário e apropriado. A OMS (2021) elenca uma série de ações para essas lideranças alcançarem o objetivo em discussão, como comparar riscos e benefícios dos sistemas de IA/AM com riscos e benefícios de sistemas existentes. É imprescindível, ainda, que as lideranças organizacionais averiguem se, no cenário local e específico em que atuam, o público atendido é favorável ao uso de sistemas de IA/AM no tratamento de seus problemas de saúde. A OMS (2021) refere-se a essa situação como "licença social" para o uso da IA.

DIRETRIZES PARA O DESENVOLVIMENTO E USO RESPONSÁVEL DA IA NA SAÚDE NO BRASIL: O PAPEL DA ESD

A ESD é um dos principais pilares da saúde digital brasileira. Ainda que não seja dedicada exclusivamente à IA, ela pode ser um bom ponto de partida para balizar o desenvolvimento e o uso responsável dessa tecnologia na área da saúde no Brasil, por se tratar de uma estratégia estatal discutida e negociada por diferentes partes interessadas. Publicada em 2020, ela atualiza o Plano de Ação, Monitoramento e Avaliação de Saúde Digital (PAM&A 2019-2023) para o Brasil.

A estratégia compreende sete prioridades, 18 subprioridades e 36 ações estratégicas. Elas correspondem a diretrizes, políticas, portarias, atos e iniciativas aprovados no âmbito do Sistema Único de Saúde (SUS), conforme apresentados no Quadro 6.

A seguir, os sete desafios discutidos na Seção "Recomendações para o desenvolvimento e uso responsável da IA: um balanço" são reapresentados e os pontos de contato entre as recomendações da OMS (2021) para cada um deles e o conteúdo da ESD são discutidos. Antes disso, cabe mencionar que, na ESD, existem eixos transversais, um dos quais é a implantação do Espaço de Colaboração, entendido como:

> [...] um espaço conceitual, virtual, distribuído, lógico e físico que viabilize a colaboração entre todos os atores em Saúde Digital, com claras definições de expectativas, papéis e responsabilidades. A colaboração proposta não é exclusivamente tecnológica e procura incluir modelos, serviços, métodos e conhecimentos que sejam viabilizados ou se tornem mais eficientes pelo uso da Saúde Digital. (Brasil, 2020, p. 13)

Tanto esse eixo transversal quanto a primeira prioridade da ESD, como revela sua descrição no Quadro 6, têm forte alinhamento com a indicação da OMS (2021) de que a articulação entre diferentes partes interessadas deve estar presente no desenvolvimento, na validação e na utilização de sistemas de IA/AM.

As recomendações da OMS (2021) possuem outros dois elementos comuns. O primeiro é a relevância de orientações públicas e privadas, e de regulações oficiais que orientem auditorias e avaliações de risco de sistemas de IA/AM. Esse aspecto não é alheio à ESD, visto que manifestações nesse sentido estão presentes tanto nas referências da estratégia à LGPD (2018), como discutido adiante, quanto em itens específicos das ações estratégicas. O segundo tópico comum às recomendações da organização internacional, tratado a seguir, refere-se à ideia de que a privacidade deve ser preservada a todo custo, refletindo-se nas ideias de privacidade desde a concepção (*privacy by design*) e privacidade por padrão (*privacy by default*).

QUADRO 6 - AS SETE PRIORIDADES DA ESTRATÉGIA DE SAÚDE DIGITAL PARA O BRASIL 2020-2028

Governança e liderança para a ESD	Garantir que a ESD28 seja desenvolvida sob a liderança do MS, mas que, ao mesmo tempo, seja capaz de incorporar a contribuição ativa de atores que participem das plataformas de colaboração
Informatização dos três níveis de atenção	Induzir a implementação de políticas de informatização dos sistemas de saúde, acelerando a adoção de sistemas de prontuários eletrônicos e de gestão hospitalar como parte integradora dos serviços e processos de saúde
Suporte à melhoria de atenção à saúde	Fazer com que a Rede Nacional de Dados em Saúde (RNDS) ofereça suporte às melhores práticas clínicas, por meio de serviços, como telessaúde, e aplicativos desenvolvidos no MS, bem como outras aplicações que sejam desenvolvidas pela plataforma de colaboração
O usuário como protagonista	Engajamento de pacientes e cidadãos, para promover a adoção de hábitos saudáveis e o gerenciamento de sua saúde, de sua família e de sua comunidade, além de auxiliar na construção dos sistemas de informação que irão utilizar
Formação e capacitação de recursos humanos	Capacitar profissionais de saúde em informática em saúde, garantir o reconhecimento da informática em saúde como área de pesquisa e o informata em saúde como profissão
Ambiente de interconectividade	Permitir que a RNDS potencialize o trabalho colaborativo em todos os setores da saúde para que tecnologias, conceitos, padrões, modelo de serviços, políticas e regulações sejam postos em prática
Ecossistema de inovação	Garantir que exista um ecossistema de inovação que aproveite ao máximo o ambiente de interconectividade em saúde, estabelecendo-se como um grande laboratório de inovação aberta, sujeito às diretrizes, normas e políticas estabelecidas por meio da prioridade um

FONTE: ESD (2020).

Preservar a privacidade dos pacientes. A ESD enfatiza a importância da preservação da privacidade, uma preocupação expressa de maneira clara em sua primeira prioridade, na qual há ações específicas para assegurar que as iniciativas da ESD sejam alinhadas à LGPD. Nesse sentido, ela destaca, por exemplo, a necessidade de fortalecimento do consentimento informado e de melhoria dos modelos de compartilhamento de dados. Nesse sentido, as discussões da OMS (2021) sobre modelos de consentimento e sistemas federados de dados podem ser referências úteis aos esforços da ESD em relação a essa prioridade.

Assegurar a qualidade e a representatividade dos dados utilizados. A ESD não aborda de maneira explícita os desafios e os riscos ligados à (falta de) representatividade dos dados em saúde. A atenção a esse tópico está presente, contudo, na ênfase da estratégia à qualidade dos dados em saúde, a qual se manifesta em suas inúmeras ações estratégicas voltadas ao fortalecimento da RNDS. Parte do programa Conecte SUS, ela tem como objetivo principal "promover a troca de informações entre os pontos da Rede de Atenção à Saúde (RAS), permitindo a transição e continuidade do cuidado nos setores público e privado" (MS, 2020 p. 20).

A RNDS é uma plataforma nacional que integra dados de saúde de todo o país, viabilizando a interoperabilidade entre sistemas de informação de saúde de todos os setores. Uma integração desse tipo é estratégica à acumulação de dados clínicos de diferentes fontes e, por consequência, à formação de um grande conjunto de informações individualizadas que podem ser usadas para entender a situação saúde-doença da população e programar esforços, fornecendo resultados que podem orientar melhor a tomada de decisão pelos gestores em benefício das comunidades e dos indivíduos. O cumprimento dessa missão depende, entre outros elementos, da conexão dos diferentes atores da saúde à RNDS. Mesmo que esse desafio seja superado, e a RNDS continue a se consolidar[18], não será suficiente para resolver outros dois desafios ligados à qualidade dos sistemas de IA/AM: (a) a falta de consenso sobre como reportar ou comparar a acurácia de sistemas de IA/AM; e (b) a falta de testes em ambientes clínicos reais.

18 É importante notar que, durante a pandemia COVID-19, a RNDS apresentou resultados importantes, em um claro sinal de que está se consolidando.

O primeiro deles pode ser superado, ao menos em parte, por alinhamentos entre os membros do Espaço de Colaboração. Diretamente ligado ao debate sobre *accountability*, o segundo desafio demanda atribuições claras de responsabilidade por eventuais erros da tecnologia. De natureza regulatória, essa ação demanda construções de entendimento entre os diferentes atores envolvidos com os sistemas de IA/AM; sem isso, a regulação pode ser pouco efetiva na proteção de pacientes e profissionais da saúde ou rígida a ponto de prejudicar a inovação. Por seu caráter multissetorial, o Espaço de Colaboração pode ser estratégico para evitar qualquer um desses dois desfechos indesejáveis.

Assegurar a infraestrutura (de dados) necessária ao bom funcionamento da tecnologia. A ESD tem diferentes ações voltadas à consolidação da RNDS, as quais se mostram estratégicas à construção e à manutenção da infraestrutura de dados necessária ao bom funcionamento dos sistemas de IA/AM. Três delas destacam-se: (a) a promoção da interoperabilidade com sistemas externos diversos, como os sistemas da Atenção Primária, de laboratórios, de farmácias, entre outros; (b) a implantação de um repositório de dados para armazenar informações em saúde; e (c) a adoção de padrões para a informação em saúde disponibilizados e internacionalmente reconhecidos e implantados. Enquanto as duas primeiras ações estratégicas podem evitar situações de dependência tecnológica, a terceira é essencial para enfrentar a falta de padronização que caracteriza muitos dos RES. Em contrapartida, é menos nítido quais elementos da ESD poderiam ser mobilizados para a superação de outro desafio de infraestrutura ligado ao funcionamento de sistemas de IA/AM: a necessidade de *hardwares* com significativo poder de processamento. Pouco explorado pelos documentos selecionados para este capítulo, esse desafio é discutido por Adler-Milstein *et al.* (2022).

Desenvolver sistemas atentos à experiência dos usuários. Desde sua criação, entre os anos 1960 e 1970, a área de informática em saúde sempre defendeu que diferentes atores sociais devem ser envolvidos no desenvolvimento de sistemas tecnológicos voltados à saúde, especialmente seus potenciais usuários. Atenta a isso, a ESD apresenta uma série de ações estratégicas que garantem centralidade aos cidadãos

e aos profissionais de saúde nos processos de cuidado. Ao se observar a disposição das prioridades, nota-se que a atenção aos cidadãos está propositadamente localizada no centro delas, dando a ideia de que as três primeiras e as três últimas prioridades devem estar direcionadas para eles. Todavia, por terem sido elaboradas antes da ampliação mundial no uso de sistemas de IA/AM na saúde, as ações listadas na estratégia contribuem de maneira restrita para que os profissionais de saúde e os cidadãos participem tanto do desenvolvimento quanto da avaliação do uso dos sistemas em questão. A OMS (2021) encoraja esse tipo de participação.

Assegurar formação e treinamento aos profissionais no ponto de cuidado. Os trabalhos revisados mencionam ser necessário que os profissionais de saúde estejam preparados para utilizar sistemas de IA/AM; porém, não apontam caminhos claros à consecução desse objetivo. A ESD tem ações estratégicas que, se adaptadas às particularidades da IA, podem ser úteis ao preenchimento dessa lacuna no caso brasileiro. Nesse contexto, destacam-se ações específicas ligadas à quinta prioridade ("Capacitação de Recursos Humanos"): (a) levantar e descrever competências, experiências, conhecimentos e habilidades associados a cada perfil funcional necessário para que profissionais e gestores de saúde, assim como profissionais da área de tecnologia da informação (TI), sejam ativos participantes da ESD; e (b) promover o reconhecimento da informática em saúde como profissão na Classificação Brasileira de Ocupações (CBO), o que inclui a definição de perfis profissionais e detalhamento de suas atribuições, deveres e limites éticos.

Tornar os sistemas de IA/AM financeiramente acessíveis. A ESD inclui ações como o mapeamento de fontes de financiamento público e o estabelecimento de mecanismos para o financiamento privado, as quais compreendem tarefas específicas, por exemplo, a preparação da documentação necessária para acessar os recursos privados. Como apontado, os documentos revisados trazem poucas indicações sobre maneiras de enfrentar os altos custos dos sistemas de IA/AM; desse modo, adensar as considerações da ESD sobre financiamento pode ser estratégico não só à construção de orientações nacionais de governança da IA em saúde, mas também internacionais. A experiência brasileira pode ser especialmente proveitosa a outros PRMB; portanto, para esse adensamento,

a subprioridade "saúde baseada em valor" – ligada à sétima prioridade ("Ecossistema de Inovação") – também pode ser útil, uma vez que seu objetivo central é encorajar a testagem de conceitos, modelos, métodos e conjuntos de dados que ajudem a superar o desafio de medir valor em saúde.

Definir como a tecnologia pode aprimorar o trabalho clínico e administrativo. Este desafio dialoga com os demais voltados à formação e ao treinamento de profissionais preparados para utilizar sistemas de IA/AM, ainda que seja mais centrado nos gestores de saúde. Um desafio adicional que devem enfrentar é a necessidade de aferir se as instituições pelas quais são responsáveis possuem uma "licença social" para utilizar sistemas de IA. Esse tipo de averiguação não é simples. Afinal, comunidades locais podem se opor ao uso da tecnologia, mesmo que possa render benefícios individuais e coletivos. Os gestores precisam estar preparados, portanto, para lidar com esse tipo de situação. Entre os mecanismos da ESD, as ações estratégicas do terceiro eixo ("Suporte à melhoria da atenção à saúde") podem ser instrumentais à condução de discussões comunitárias sobre potenciais benefícios e riscos dos sistemas de IA/AM.

CONCLUSÃO

Não raro, as discussões sobre IA são marcadas por ambiguidades temporais que misturam o que a tecnologia pode vir a fazer e o que ela realmente faz (Meadows *et al.,* 2020). Mais do que uma confusão discursiva, essa mistura de tempos verbais revela que ainda não se sabe ao certo qual é o estágio atual da IA. Sem qualquer pretensão de oferecer um diagnóstico completo e definitivo dessa situação, procurou-se mapear os desafios e as oportunidades vislumbrados para adoção das ferramentas de IA na área da saúde, a partir de uma análise de publicações sobre IA e saúde. Foram reunidos elementos que permitem afirmar que o potencial da IA ainda está sendo explorado no campo estudado, e a concretização desse potencial depende da superação de, pelo menos, sete desafios: (a) preservar a privacidade dos pacientes; (b) assegurar a qualidade e a representatividade dos dados utilizados; (c) assegurar a infraestrutura (de dados) necessária ao bom funcionamento da tecnologia; (d) desenvolver sistemas atentos à experiência dos usuários; (e) assegurar formação e treinamento aos

profissionais no ponto de cuidado; (f) tornar os sistemas de IA/AM financeiramente acessíveis; e (g) definir como a tecnologia pode aprimorar o trabalho clínico e administrativo.

Na sequência, procurou-se entender como os sete desafios podem ser superados, bem como identificar, no Brasil, possíveis instrumentos, políticas e diretrizes que auxiliem o país a enfrentá-los. Diante desse objetivo, a ESD foi analisada visando identificar se e como essa estratégia pode orientar usos responsáveis da IA no setor de saúde do país. Foram destacadas ações que viabilizam e facilitam o processo de adoção de técnicas de IA/AM e também potenciais adaptações na ESD que podem ser consideradas pelos profissionais envolvidos no atendimento em saúde populacional.

No primeiro grupo, estão a existência do Espaço de Colaboração e da RNDS, assim como as considerações da ESD sobre saúde baseada em valor, capacitação de recursos humanos, e atenção à saúde. Já no campo das adaptações, pelo menos cinco temas foram identificados: (a) o fortalecimento da proteção à privacidade dos pacientes por meio de medidas específicas, como a adoção de sistemas federados e modelos de consentimento próprios a coletas massivas de dados; (b) a falta de consenso sobre como reportar ou mesmo comparar a acurácia de sistemas de IA/AM; (c) o desenvolvimento de regulações que não sejam pouco efetivas na proteção de pacientes e profissionais da saúde, nem rígidas a ponto de prejudicar a inovação; (d) a necessidade de *hardwares* com significativo poder de processamento; e (e) a participação dos profissionais de saúde e dos cidadãos tanto do desenvolvimento, quanto da avaliação do uso de sistemas de IA/AM.

Para que as mudanças impostas à saúde pelos sistemas de IA/AM tenham resultados positivos, é necessária a formulação de políticas públicas que promovam o desenvolvimento e o uso responsável dessa tecnologia no país. No entando, tais elementos não são suficientes à consecução desse fim; para atingi-lo, é preciso compreender melhor qual é o estágio atual da relação entre IA e saúde no Brasil, o que demanda, necessariamente, investigações voltadas aos diferentes atores que atuam na intersecção entre esses dois temas. Os próximos capítulos desta publicação exploram aspectos aqui apresentados.

REFERÊNCIAS

Adler-Milstein, J., Aggarwal, N., Ahmed, M., Castner, J., Evans, B. J., Gonzalez, A. A., James, A. C., Lin, S., Mandl, K. D., Matheny, M. E., Sendak, M. P., Shachar, C., & Williams, A. (2022). Meeting the moment: Addressing barriers and facilitating clinical adoption of Artificial Intelligence in medical diagnosis. *NAM Perspectives, 2022*, 35-80. https://www.gao.gov/products/gao-22-104629

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., Satamaría, 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

Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. (2020). Explainability for Artificial Intelligence in healthcare: A multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, *20*(310), 1-9. https://doi.org/10.1186/s12911-020-01332-6

Aquino, Y. S. J., Rogers, W. A., Braunack-Mayer, A., Frazer, H., Win, K. T., Houssami, N., Degeling, C., Semsarian, C., & Carter, S. M. (2023). Utopia versus dystopia: Professional perspectives on the impact of healthcare Artificial Intelligence on clinical roles and skills. *International Journal of Medical Informatics, 169*, 1-10. https://doi.org/10.1016/j.ijmedinf.2022.104903

Arbix, G. (2020). A transparência no centro da construção de uma IA ética. *Novos Estudos, 39*(02), 395-413. https://doi.org/10.25091/s01013300202000020008

Asan, O., Bayrak, A. E., & Choudhury, A. (2020). Artificial Intelligence and human trust in healthcare: Focus on clinicians. *Journal of Medical Internet Research, 22*(6), 1-7. https://doi.org/10.2196/15154

Benjamens, S., Dhunnoo, P., & Meskó, B. (2020). The state of Artificial Intelligence-based FDA-approved medical devices and algorithms: An online database. *npj Digital Medicine, 3*(118), 1-8. https://doi.org/10.1038/s41746-020-00324-0

Berryhill, J., Heang, K. K., Clogher, R., & McBride, K. (2019). Hello, world: Artificial Intelligence and its use in the public sector. *OECD Working Papers on Public Governance, 36*, 1-185. https://dx.doi.org/10.1787/726fd39d-en

Bishara, A., Maze, E. H, & Maze, M. (2022). Considerations for the implementation of machine learning into acute care settings. *British Medical Bulletin, 141*, 15-32. https://doi.org/10.1093/bmb/ldac001

Centro da Quarta Revolução Industrial do Brasil. (2022). *Guia de contratações públicas de Inteligência Artificial*. https://c4ir.org.br/wp-content/uploads/2022/11/1648128585465GUIA-DE-CONTRATACOES-PUBLICAS-DE-AI_C4IR_v4.pdf

Challen, R., Denny, J., Pitt, M., Gompels, L., Edwards, T., & Tsaneva-Atanasova, K. (2019). Artificial Intelligence, bias and clinical safety. *BMJ Quality and Safety, 28*, 231-237. https://qualitysafety.bmj.com/content/28/3/231

Daugherty, P. R., & Wilson, J. (2018). *Human + machine: Reimagining work in the age of AI*. Harvard Business Review Press.

Dourado, D. A., & Aith, F. M. A. (2022). The regulation of Artificial Intelligence in health in Brazil begins with the General Personal Data Protection Law. *Revista de Saúde Pública, 56*(80), 1-7. https://doi.org/10.11606/s1518-8787.2022056004461

Du, Y., McNestry, C., Wei, L., Antoniadi, A. M., McAuliffe, F. M., & Mooney, C. (2023). Machine learning-based clinical decision support systems for pregnancy care: A systematic review. *International Journal of Medical Informatics, 173*, 1-9. https://doi.org/10.1016/j.ijmedinf.2023.105040

El-Sappagh, S., Ali, F., El-Masri, S., Kim, K., Ali, A., & Kwak, K-S. (2019). Mobile health technologies for diabetes mellitus: Current state and future challenges. *IEEE Access*, *7*, 21917-21947. https://ieeexplore.ieee.org/document/8534339

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People – An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*, 689-707. https://doi.org/10.1007/s11023-018-9482-5

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Ed.), *Media technologies: Essays on communication, materiality, and society* (pp. 167-193). The MIT Press.

Giovanola, B., & Tiribelli, S. (2023). Beyond bias and discrimination: Redefining the AI ethics principle of fairness in healthcare machine-learning algorithms. *AI & Society, 38*, 549-563. https://doi.org/10.1007/s00146-022-01455-6

Goirand, M., Austin, E., & Clay-Williams, R. (2021). Implementing ethics in healthcare AI-based applications: A scoping review. *Science and Engineering Ethics, 27*(61). https://doi.org/10.1007/s11948-021-00336-3

Hobensack, M., Song, J., Scharp, D., Bowles, K. H., & Topaz, M. (2023). Machine learning applied to electronic health record data in home healthcare: A scoping review. *International Journal of Medical Informatics, 170*. https://doi.org/10.1016/j.ijmedinf.2022.104978

Hogg, H. D. J., Al-Zubaidy, M., Technology Enhanced Macular Services Study Reference Group, Talks, J., Denniston, A. K., Kelly, C. J, Malawana, J., Papoutsi, C., Teare, M. D., Keane, P. A., Beyer, F. R., & Maniatopoulos, G. (2023). Stakeholder perspectives of clinical Artificial Intelligence implementation: Systematic review of qualitative evidence. *Journal of Medical Internet Research, 25*. https://doi.org/10.2196/39742

Junaid, S. B., Imam, A. A., Balogun, A. O., Silva, L. C., Surakat, Y. A., Kumar, G., Abdulkarim, M., Shuaibu, A. N., Garba, A., Sahalu, Y., Mohammed, A., Mohammed, T. Y., Abdulkadir, B. A., Abba, A. A., Kakumi, N. A. I., & Mahamad, S. (2022). Recent advancements in emerging technologies for healthcare management systems: A survey. *Healthcare (Switzerland), 10*(10), 1-45. https://doi.org/10.3390/healthcare10101940

Kaul, V., Enslin, S., & Gross, S. A. (2020). History of Artificial Intelligence in medicine. *Gastrointestinal Endoscopy, 92*(4), 807-812. https://doi.org/10.1016/j.gie.2020.06.040

Krafft, P. M., Young, M., Katell, M., Huang, K., & Bugingo, G. (2020, February). Defining AI in policy versus practice. *Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society (AIES '20)*. https://doi.org/10.1145/3375627.3375835

Lei Geral de Proteção de Dados Pessoais (LGPD). (2018). Lei n. 13.709, de 14 de agosto de 2018. Dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural. https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm

Liu, X., Glocker, B., McCradden, M. M., Ghassemi, M., Denniston, A. K., & Oakden-Rayner, L. (2022). The medical algorithmic audit. *The Lancet Digital Health*, *4*(5), e384-397. https://doi.org/10.1016/S2589-7500(22)00003-6

Loh, H. W., Ooi, C. P., Seoni, S., Barua, P. D., Molinari, F., & Acharya, U. R. (2022). Application of explainable Artificial Intelligence for healthcare: A systematic review of the last decade (2011-2022). *Computer Methods and Programs in Biomedicine*, *226*. https://doi.org/10.1016/j.cmpb.2022.107161

Markus, A. F., Kors, J. A., & Rijnbeek, P. R. (2021). The role of explainability in creating trustworthy Artificial Intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. *Journal of Biomedical Informatics*, *113*, 1-11. https://doi.org/10.1016/j.jbi.2020.103655

Meadows, R., Hine, C., & Suddaby, E. (2020). Conversational agents and the making of mental health recovery. *Digital Health*, *6*, 1-11. https://pubmed.ncbi.nlm.nih.gov/33282335/

Miailhe, N., & Hodes, C. (2017). Making the AI revolution work for everyone. *The Future Society, AI Initiative*, 1-29. https://thefuturesociety.org/wp-content/uploads/2019/08/Making-the-AI-Revolution-work-for-everyone.-Report-to-OECD.-MARCH-2017.pdf

Ministério da Saúde. (2020). *Estratégia de Saúde Digital para o Brasil 2020-2028*. https://bvsms.saude.gov.br/bvs/publicacoes/estrategia_saude_digital_Brasil.pdf

Moazemi, S., Vahdati, S., Li, J., Kalkhoff, S., Castano, L. J. V., Dewitz, B., Bibo, R., Sabouniaghdam, P., Tootooni, M. S., Bundschuh, R. A., Lichtenberg, A., Aubin, H., & Schmid, F. (2023). Artificial Intelligence for clinical decision support for monitoring patients in cardiovascular ICUs: A systematic review. *Frontiers in Medicine, 10*. https://doi.org/10.3389/fmed.2023.1109411

Ngiam, K. Y., & Khor, I. W. (2019). Big data and machine learning algorithms for health-care delivery. *Lancet Oncology, 20*(5), e262-e273. https://pubmed.ncbi.nlm.nih.gov/31044724/

Nunes, H. C., Guimarães, R. M. C., & Dadalto, L. (2022). Desafios bioéticos do uso da Inteligência Artificial em hospitais. *Revista de Bioética, 30*(1), 82-93. https://doi.org/10.1590/1983-80422022301509PT

Organização Mundial da Saúde. (2021). *Ethics and governance of Artificial Intelligence for health – WHO guidance*. https://www.who.int/publications/i/item/9789240029200

Organização para a Cooperação e Desenvolvimento Econômico. (2024). *Recommendation of the council on Artificial Intelligence*. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

Pap, I.A., & Oniga, S. (2022). A review of converging technologies in eHealth pertaining to Artificial Intelligence. *International Journal of Environmental Research and Public Health, 19*(18), 1-15. https://doi.org/10.3390/ijerph191811413

Rahman, A., Hossain, S., Muhammad, G., Kundu, D., Debnath, T., Rahman, M., Khan, S. I., Tiwari, P., & Band, S. (2022). Federated learning-based AI approaches in smart healthcare: concepts, taxonomies, challenges and open issues. *Cluster Computing,* 26, 2271-2311. https://doi.org/10.1007/s10586-022-03658-4

Sahni, N., Stein, G., Zemmel, R., & Cutler, D. (2023). The potential impact of Artificial Intelligence on healthcare spending. *National Bureau of Economic Research,* working paper 30857. https://www.nber.org/books-and-chapters/economics-artificial-intelligence-health-care-challenges/potential-impact-artificial-intelligence-health-care-spending#:~:text=Yet%20healthcare%20lags%20other%20industries,billion%20annually%20in%202019%20dollars.

Schwalbe, N., & Wahl, B. (2020). Artificial Intelligence and the future of global health. *The Lancet, 395,* 1579-1586. https://doi.org/10.1016/S0140-6736(20)30226-9

Stanford University Human-Centered Artificial Intelligence. (2023). *Artificial Intelligence index report 2023.* https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf

Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeke, K. I. (2020). An overview of clinical decision support systems: Benefits, risks, and strategies for success. *npj Digital Medicine, 3*(17), 1-10. https://doi.org/10.1038/s41746-020-0221-y

Topol, E. (2019). High-performance medicine: The convergence of human and Artificial Intelligence. *Nature Medicine, 25,* 44-56. https://doi.org/10.1038/s41591-018-0300-7

Tran, B. X., Vu, G. T., Ha, G. H., Vuong, Q-H., Ho, M-T., Vuong, T-T., La, V-P., Ho, M-T., Nghiem, K-C. P., Nguyen, H. L. T., Latkin, C. A., Tam, W. W. S., Cheung, N-M., Nguyen, H-K. T., Ho,

C. S. H., & Ho, R. C. M. (2019). Global evolution of research in Artificial Intelligence in health and medicine: A bibliometric study. *Journal of Clinical Medicine, 8*(3), 360-378. https://pubmed.ncbi.nlm.nih.gov/30875745/

van Melle, W. (1978). MYCIN: A knowledge-based consultation program for infectious disease diagnosis. *International Journal of Man-Machine Studies, 10*(3), 313-322. https://doi.org/10.1016/S0020-7373(78)80049-2

Xie, Y., Lu, L., Gao, F., He, S-J., Zhao, H-J., Fang, Y., Yang, J-M., An, Y., Ye, Z-W., & Dong, Z. (2021). Integration of Artificial Intelligence, blockchain, and wearable technology for chronic disease management: A new paradigm in smart healthcare. *Current Medical Science, 41*(6), 1123-1133. https://doi.org/10.1007/s11596-021-2485-0

Xu, Q., Xie, W., Liao, B., Hu, C., Qin, L., Yang, Z., Xiong, H., Lyu, Y., Zhou, Y., & Luo, A. (2023). Interpretability of clinical decision support systems based on Artificial Intelligence from technological and medical perspective: A systematic review. *Journal of Healthcare Engineering, 2023*, 1-13. https://doi.org/10.1155/2023/9919269

Yang, L., Ene, I. C., Belaghi, R. A., Koff, D., Stein, N., & Santaguida, P. L. (2021). Stakeholders' perspectives on the future of Artificial Intelligence in radiology: A scoping review. *European Radiology, 32*(3), 1477-1495. https://pubmed.ncbi.nlm.nih.gov/34545445/

Yu, K-H., Beam, A. L., & Kohane, I. S. (2018). Artificial Intelligence in healthcare. *Nature Biomedical Engineering, 2*, 719-731. https://doi.org/10.1038/s41551-018-0305-z

ANEXO 1 - INFORMAÇÕES GERAIS SOBRE OS ESTUDOS SELECIONADOS NA BASE DE DADOS SCOPUS

Nº	TEMA	ANO	AUTORES	TÍTULO	PERIÓDICO
1	Prática clínica	2019	Kelly, C. J. *et al.*	*Key challenges for delivering clinical impact with Artificial Intelligence*	BMC Medicine
2	Prática clínica	2019	Challen, R. *et al.*	*Artificial Intelligence, bias and clinical safety*	BMJ Quality and Safety
3	Prática clínica	2019	Sheikhalishahi, S. *et al.*	*Natural language processing of clinical notes on chronic diseases: Systematic review*	JMIR Medical Informatics
4	Prática clínica	2019	Noorbakhsh-Sabet, S. *et al.*	*Artificial Intelligence transforms the future of health care*	American Journal of Medicine
5	Prática clínica	2019	Reddy, S. *et al.*	*Artificial Intelligence-enabled healthcare delivery*	Journal of the Royal Society of Medicine
6	Prática clínica	2020	Sutton, R. T. *et al.*	*An overview of clinical decision support systems: Benefits, risks and strategies for success*	npj Digital Medicine
7	Prática clínica	2020	Goecks, J. *et al.*	*How machine learning will transform biomedicine*	Cell
8	Prática clínica	2020	Ahmed, Z. *et al.*	*Artificial Intelligence with multi-functional machine learning platform development for better healthcare and precision medicine*	Database
9	Prática clínica	2020	Asan, O. *et al.*	*Artificial Intelligence and human trust in healthcare: Focus on clinicians*	Journal of Medical Internet Research
10	Prática clínica	2022	Loh, H. W. *et al.*	*Application of explainable Artificial Intelligence for healthcare: A systematic review of the last decade (2011–2022)*	Computer Methods and Programs in Biomedicine

11	Prática clínica	2022	Liu, X. *et al.*	*The medical algorithmic audit*	The Lancet Digital Health
12	Prática clínica	2022	Busnatu, S. *et al.*	*Clinical applications of Artificial Intelligence — An updated overview*	Journal of Clinical Medicine
13	Prática clínica	2022	Alanazi, A.	*Using machine learning for healthcare challenges and opportunities*	Informatics in Medicine Unlocked
14	Prática clínica	2022	Yang, L. *et al.*	*Stakeholders' perspectives on the future of Artificial Intelligence in radiology: A scoping review*	European Radiology
15	Prática clínica; Ética e regulação	2022	Pap, I. A., & Oniga, S.	*A review of converging technologies in eHealth pertaining to Artificial Intelligence*	International Journal of Environmental Research and Public Health
16	Prática clínica	2022	Bishara, A. *et al.*	*Considerations for the implementation of machine learning into acute care settings*	British Medical Bulletin
17	Prática clínica	2022	Prakash, S. *et al.*	*Ethical conundrums in the application of Artificial Intelligence (AI) in healthcare—A* scoping review of reviews	Journal of Personalized Medicine
18	Prática clínica	2023	Moazemi, S. *et al.*	*Artificial Intelligence for clinical decision support for monitoring* patients in cardiovascular ICUs: A systematic review	Frontiers in Medicine
19	Prática clínica	2023	Xu, Q. *et al.*	*Interpretability of clinical decision support systems based on Artificial Intelligence from technological and medical perspective: A systematic review*	Journal of Healthcare Engineering
20	Prática clínica	2023	Hobensack, M. *et al.*	*Machine learning applied to electronic health record data in home healthcare: A scoping review*	International Journal of Medical Informatics

21	Prática clínica	2023	Du, Y. *et al.*	*Machine learning-based clinical decision support systems for pregnancy care: A systematic review*	International Journal of Medical Informatics
22	Administração	2022	Rahman, A. *et al.*	*Federated learning-based AI approaches in smart healthcare: concepts, taxonomies, challenges and open issues*	Cluster Computing
23	Administração	2022	Junaid, S. B. *et al.*	*Recent advancements in emerging technologies for healthcare management systems: A survey*	Healthcare (Switzerland)
24	Administração	2022	Albalawi, U., & Mustafa, M.	*Current Artificial Intelligence (AI) techniques, challenges, and approaches in controlling and fighting* COVID-19: A review	International Journal of Environmental Research and Public Health
25	Administração	2023	Luschi, A. *et al.*	*Semantic ontologies for complex healthcare structures: A scoping review*	IEEE Access
26	Administração	2023	Vargas, V. B. *et al.*	*Influential factors for hospital management maturity models in a post-Covid-19 scenario - Systematic literature review*	Journal of Information Systems Engineering and Management
27	Ética e regulação	2019	Wiens, J. *et al.*	*Do no harm: A roadmap for responsible machine learning for health care*	Nature Medicine
28	Ética e regulação	2019	Watson, D. S. *et al.*	*Clinical applications of machine learning algorithms: Beyond the black box*	BMJ (Online)
29	Ética e regulação	2019	Tran, B. X. *et al.*	*Global evolution of research in Artificial Intelligence in health and medicine: A bibliometric study*	Journal of Clinical Medicine
30	Ética e regulação	2019	Guan, J.	*Artificial Intelligence in healthcare and medicine: Promises ethical challenges and governance*	Chinese Medical Sciences Journal

31	Ética e regulação	2019	Ngiam, K. Y., & Khor, I. W.	*Big Data and machine learning algorithms for health-care delivery*	The Lancet Oncology
32	Ética e regulação	2020	Amann, J. *et al.*	*Explainability for Artificial Intelligence in healthcare: a multidisciplinary perspective*	BMC Medical Informatics and Decision Making
33	Ética e regulação	2020	Vollmer, S. *et al.*	*Machine learning and Artificial Intelligence research for patient benefit: 20 critical questions on transparency replicability ethics and effectiveness*	The BMJ
34	Ética e regulação	2020	Char, D. S. *et al.*	*Identifying ethical considerations for machine learning hHealthcare applications*	American Journal of Bioethics
35	Ética e regulação	2021	Goirand, M. *et al.*	*Implementing ethics in healthcare AI-based applications: A scoping review*	Science and Engineering Ethics
36	Ética e regulação	2021	Markus, A. F. *et al.*	*The role of explainability in creating trustworthy Artificial Intelligence for health care: A comprehensive survey of the terminology design choices and evaluation strategies*	Journal of Biomedical Informatics
37	Ética e regulação	2021	Saheb, T. *et al.*	*Mapping research strands of ethics of Artificial Intelligence in healthcare: A bibliometric and content analysis*	Computers in Biology and Medicine
38	Ética e regulação	2022	Quinn, T. P. *et al.*	*The three ghosts of medical AI: Can the black-box present deliver?*	Artificial Intelligence in Medicine
39	Ética e regulação	2022	Yoon, C. H. *et al.*	*Machine learning in medicine: Should the pursuit of enhanced interpretability be abandoned?*	Journal of Medical Ethics

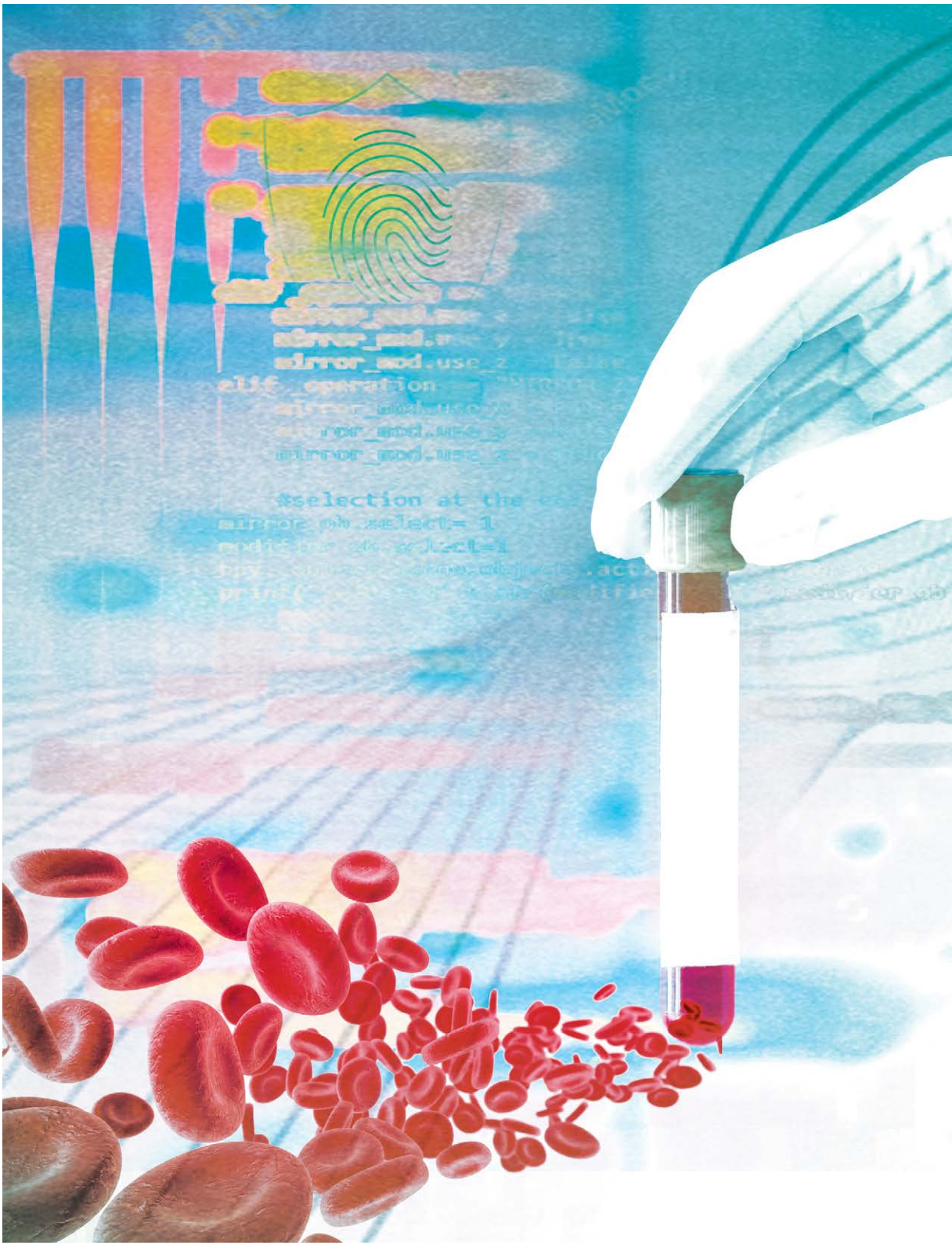
40	Ética e regulação	2022	Salazar, L. H. A. *et al.*	*Application of machine learning techniques to predict a patient's no-show in the healthcare sector*	Future Internet
41	Ética e regulação	2022	McLennan, S. *et al.*	*Embedded ethics: A proposal for integrating ethics into the development of medical AI*	BMC Medical Ethics
42	Ética e regulação	2022	Angerschmid, Al. *et al.*	*Fairness and explanation in AI-informed decision making*	Machine Learning and Knowledge Extraction
43	Ética e regulação	2022	Parviainen, J., & Rantala, J.	*Chatbot breakthrough in the 2020s? An ethical reflection on the trend of automated consultations in health care*	Medicine, Health Care and Philosophy
44	Ética e regulação	2022	Bærøe, K. *et al.*	*Can medical algorithms be fair? Three ethical quandaries and one dilemma*	BMJ Health and Care Informatics
45	Ética e regulação	2022	Siala, H., & Wang, Y.	*SHIFTing Artificial Intelligence to be responsible in healthcare: A systematic review*	Social Science and Medicine
46	Ética e regulação	2022	Dourado, D. A., & Aith, F. M. A.	*The regulation of Artificial Intelligence for health in Brazil begins with the General Personal Data Protection Law*	Revista de saúde pública
47	Ética e regulação	2022	Nunes, H. C. *et al.*	*Bioethical challenges related to the use of Artificial Intelligence in hospitals*	Revista Bioética
48	Ética e regulação	2022	Cartolovni, A. *et al.*	*Ethical, legal and social considerations of AI-based medical decision-support tools: A scoping review*	International Journal of Medical Informatics
49	Ética e regulação	2022	Okolo, C. T.	*Optimizing human-centered AI for healthcare in the Global South*	Patterns

50	Ética e regulação	2023	Giovanola, B., & Tiribelli, S.	*Beyond bias and discrimination: redefining the AI ethics principle of fairness in healthcare machine-learning algorithms*	AI and Society
51	Ética e regulação	2023	Aquino, Y. S. J. *et al.*	*Utopia versus dystopia: Professional perspectives on the impact of healthcare Artificial Intelligence on clinical roles and skills*	International Journal of Medical Informatics
52	Ética e regulação	2023	Albahri, A. S. *et al.*	*A systematic review of trustworthy and explainable Artificial Intelligence in healthcare: Assessment of quality bias risk and data fusion*	Information Fusion
53	Ética e regulação	2023	Wu. C. *et al.*	*Public perceptions on the application of Artificial Intelligence in healthcare: A qualitative meta-synthesis*	BMJ open
54	Ética e regulação	2023	Hogg, H. D. J. *et al.*	*Stakeholder perspectives of clinical Artificial Intelligence implementation: Systematic review of qualitative evidence*	Journal of Medical Internet Research
55	Ética e regulação	2023	Sallam, M.	*ChatGPT utility in healthcare education research and practice: Systematic review on the promising perspectives and valid concerns*	*Healthcare (Switzerland)*
56	Ética e regulação	2023	Harrer, S.	*Attention is not all you need: The complicated case of ethically using large language models in healthcare and medicine*	*eBioMedicine*
57	Ética e regulação	2023	Lehoux, P. *et al.*	*Tools to foster responsibility in digital solutions that operate with or without Artificial Intelligence: A scoping review for health and innovation policymakers*	*International Journal of Medical Informatics*

Considerações regulatórias sobre Inteligência Artificial para a saúde[1]

Organização Mundial da Saúde e União Internacional de Telecomunicações

1 Este capítulo foi adaptado da publicação *Regulatory considerations on Artificial Intelligence for health* e traduzido com a autorização da Organização Mundial da Saúde (OMS) e da União Internacional de Telecomunicações (UIT). A adaptação e tradução deste texto para o português não foi criada pela OMS. A OMS não é responsável pelo conteúdo ou precisão desta tradução. A edição original em inglês é a edição vinculativa e autêntica. A publicação original está disponível em: https://iris.who.int/bitstream/handle/10665/373421/9789240078871-eng.pdf?sequence=1


mirror_mod.use_z
elif operation
#selection at the
mirror_ob.select= 1

A missão da Organização Mundial da Saúde (OMS), definida em sua Estratégia Global de Saúde Digital 2020-2025 (OMS, 2021a), é promover a saúde, manter o mundo seguro e servir os vulneráveis. No centro dessa estratégia, a OMS visa melhorar a saúde para todos, em todos os lugares, acelerando o desenvolvimento e a adoção de soluções de saúde digital adequadas, acessíveis, de baixo custo, escaláveis e sustentáveis, centradas na pessoa, a fim de prevenir, detectar e responder a epidemias e pandemias, e desenvolvendo infraestruturas e aplicações. Muitas organizações internacionais e atores globais têm contribuído nessa área, juntamente com a OMS, incluindo o Fórum Internacional de Entidades Reguladoras de Dispositivos Médicos (International Medical Device Regulators Forum [IMDRF]), o Grupo de Trabalho de Harmonização Global (Global Harmonization Working Party [GHWP]), a Food and Drug Administration (FDA) dos Estados Unidos da América (EUA), a Health Canada, a Coalizão Internacional de Autoridades Reguladoras de Medicamentos (International Coalition of Medicines Regulatory Authorities [ICMRA]), a Organização Internacional de Padronização (ISO), a Organização para a Cooperação e Desenvolvimento Econômico (OCDE), a Agência Reguladora de Medicamentos e Produtos de Saúde (Medicines and Healthcare Products Regulatory Agency [MHRA]) do Reino Unido da Grã-Bretanha e da Irlanda do Norte, a Autoridade Reguladora de Produtos de Saúde da África do Sul (South African Health Products Regulatory Authority [SAHPRA]), a Comissão Europeia (CE), a Autoridade de Ciências da Saúde (Health Sciences Authority [HSA]) de Singapura, o Conselho Internacional para a Harmonização dos Requisitos Técnicos para Produtos Farmacêuticos para Uso Humano (International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use [ICH]), a Agência de Produtos Farmacêuticos e Dispositivos Médicos (Pharmaceuticals and Medical Devices Agency [PMDA]) do Japão, a Swissmedic e a Administração de Produtos Terapêuticos (Therapeutic Goods Administration [TGA]) da Austrália. Essas organizações internacionais e regionais e as autoridades nacionais reconhecem coletivamente o potencial da Inteligência Artificial (IA) para

melhorar os resultados de saúde ao aprimorar os testes clínicos, os diagnósticos e os tratamentos médicos, a autogestão dos cuidados e os cuidados personalizados, desenvolvendo mais conhecimentos, habilidades e competências baseadas em evidências para que os profissionais prestem cuidados de saúde. Além disso, com a crescente disponibilidade de dados sobre cuidados de saúde e o rápido progresso das técnicas de análise, a IA poderá, potencialmente, transformar o setor de saúde para atender a uma gama de necessidades dos atores interessados em cuidados de saúde e desenvolvimento terapêutico.

A fim de facilitar o uso seguro e apropriado de tecnologias de IA para o desenvolvimento de sistemas de IA na área da saúde, a OMS e a União Internacional de Telecomunicações (UIT) criaram um Grupo Focal em Inteligência Artificial para a Saúde (FG-AI4H). Para apoiar seu trabalho, a FG-AI4H formou vários grupos de trabalho, incluindo um Grupo de Trabalho sobre Considerações Regulatórias (GT-CR) sobre IA para a Saúde. O GT-CR é composto por membros que representam vários atores interessados – incluindo autoridades reguladoras, formuladores de políticas públicas, academia e indústria – que exploraram conceitos regulatórios e de avaliação de tecnologias de saúde e "boas práticas" emergentes para o desenvolvimento e o uso de IA em cuidados de saúde e desenvolvimento terapêutico. O trabalho do GT-CR representa um esforço multidisciplinar e internacional para aumentar o diálogo e examinar importantes conceitos para o uso da IA em cuidados de saúde.

Esta publicação, que se baseia no trabalho do GT-CR, tem como objetivo fornecer um panorama das considerações regulatórias sobre IA para a saúde que abrange as seguintes seis áreas temáticas gerais: documentação e transparência; abordagem do ciclo de vida total do produto e gestão dos riscos; uso pretendido e validação analítica e clínica; qualidade dos dados; privacidade e proteção dos dados; e engajamento e colaboração. Este panorama não pretende ser uma diretriz, um marco referencial regulatório ou uma política; ao contrário, trata-se de uma discussão sobre as principais considerações regulatórias e um recurso que pode ser considerado relevante por todos os atores interessados – incluindo desenvolvedores que exploram e desenvolvem sistemas de IA, reguladores e formuladores de

políticas públicas no processo de identificação de abordagens para gerenciar e facilitar sistemas de IA, fabricantes que projetam e desenvolvem dispositivos médicos habilitados para IA, e profissionais de saúde que implantam e utilizam esses dispositivos médicos e sistemas de IA. Consequentemente, o GT-CR recomenda que os atores interessados levem em conta as seguintes considerações à medida que continuarem a desenvolver quadros e melhores práticas para o uso de IA em cuidados de saúde e desenvolvimento terapêutico:

1. **Documentação e transparência**: A pré-especificação e a documentação da finalidade médica pretendida e do processo de desenvolvimento, como a seleção e o uso de conjuntos de dados, padrões de referência, parâmetros, métricas, desvios dos planos originais e atualizações durante as fases de desenvolvimento, deve ser considerada de forma a permitir o rastreamento dessas etapas, conforme apropriado. Uma abordagem baseada no risco deve ser considerada também para o nível de documentação e manutenção de registros utilizados para o desenvolvimento e a validação de sistemas de IA.
2. **Gestão de riscos e abordagens do ciclo de vida de desenvolvimento de sistemas de IA**: Uma abordagem do ciclo de vida total do produto deve ser considerada em todas as fases da vida de um sistema de IA, a saber: gestão de desenvolvimento pré-comercialização, vigilância pós-comercialização e gestão de mudanças. Além disso, é essencial considerar uma abordagem de gestão de riscos que envolva aqueles associados aos sistemas de IA, como ameaças e vulnerabilidades de cibersegurança, subadaptação, enviesamento algorítmico, etc.
3. **Uso pretendido e validação analítica e clínica**: Inicialmente, deve-se considerar o fornecimento de documentação transparente do uso pretendido do sistema de IA. As informações sobre a composição do conjunto de dados de treinamento subjacente a um sistema de IA, incluindo tamanho, configuração e população, dados de entrada e saída e composição demográfica, devem ser documentados de forma transparente e fornecidos aos usuários. Além disso, é fundamental considerar

a demonstração de desempenho, além dos dados de treinamento e teste, por meio de validação analítica externa em um conjunto de dados independente. Esse conjunto de dados de validação externa deve ser representativo da população e do contexto em que se pretende implantar o sistema de IA e deve ser independente do conjunto de dados usado para desenvolver o modelo de IA durante o treinamento e os testes. A documentação transparente do conjunto de dados externo e das métricas de desempenho deve ser fornecida. Além disso, é importante também considerar um conjunto graduado de requisitos para validação clínica baseada no risco. Os estudos clínicos randomizados são o padrão-ouro para a avaliação do desempenho clínico comparativo e podem ser apropriados para as ferramentas de maior risco ou quando for necessário o mais alto padrão de evidência. Em outras situações, a validação prospectiva pode ser considerada em um ensaio de implantação e implementação no mundo real, que inclua um elemento de comparação relevante que utilize grupos aceitos. Por último, é necessário considerar um período de monitoramento pós-implantação mais intenso, mediante vigilância pós-comercialização e fiscalização do mercado de sistemas de IA.

4. **Qualidade dos dados**: Os desenvolvedores devem levar em consideração se os dados disponíveis são de qualidade suficiente para suportar o desenvolvimento do sistema de IA, a fim de atingir a finalidade pretendida. Além disso, também devem ponderar a possibilidade de implantar avaliações rigorosas antes do lançamento dos sistemas de IA, para garantir que estes não amplifiquem nenhuma das questões discutidas na Seção "Qualidade dos dados" do presente documento, tais como enviesamentos e erros. O *design* cuidadoso ou a solução de problemas imediata podem ajudar a identificar problemas de qualidade de dados precocemente e prevenir ou mitigar possíveis danos resultantes. Os atores interessados devem pensar sobre a possibilidade de mitigar os problemas de qualidade dos dados e os riscos associados que surgem nos dados relativos aos cuidados de saúde, além de continuar a

trabalhar para criar ecossistemas de dados que facilitem o compartilhamento de fontes de dados de boa qualidade.

5. **Privacidade e proteção de dados**: A privacidade e a proteção de dados devem ser consideradas durante o *design* e a implantação de sistemas de IA. No início do processo de desenvolvimento, os desenvolvedores devem considerar a formação de uma boa compreensão dos regulamentos de proteção de dados aplicáveis e das leis de privacidade, garantindo que o processo de desenvolvimento atenda ou exceda esses requisitos legais. Também é importante considerar a implementação de um programa de conformidade que aborde os riscos e garanta que as práticas de privacidade e cibersegurança levem em conta os danos potenciais, bem como o ambiente de aplicação.
6. **Engajamento e colaboração**: Durante o desenvolvimento do roteiro de inovação e implantação de IA, é importante levar em consideração o desenvolvimento de plataformas acessíveis e informativas que facilitem o engajamento e a colaboração entre os principais atores interessados, quando aplicável e apropriado. É fundamental levar em conta a simplificação do processo de supervisão para a regulação da IA por meio desse envolvimento e colaboração, a fim de acelerar os avanços que alteram a prática da IA.

Finalmente, o GT-CR forneceu um fórum para reguladores e especialistas poderem discutir considerações regulatórias para o uso de tecnologias de IA e o desenvolvimento de sistemas de IA para fins médicos e de saúde. O grupo reconhece que o cenário da IA tem evoluído rapidamente e as considerações contidas neste documento podem exigir sua ampliação à medida que a tecnologia e seus usos se desenvolvem. O GT-CR recomenda que os atores interessados, incluindo reguladores e desenvolvedores, continuem a se engajar e que a comunidade em geral trabalhe para o entendimento compartilhado e o aprendizado mútuo. Além disso, grupos nacionais e internacionais estabelecidos, como o IMDRF e a ICMRA, devem continuar trabalhando em tópicos de IA, visando à potencial convergência e harmonização regulatórias.

PRINCIPAIS APLICAÇÕES DA IA NOS CUIDADOS DE SAÚDE E NO DESENVOLVIMENTO TERAPÊUTICO

A IA tem sido cada vez mais explorada para impulsionar os cuidados de saúde em várias frentes. A combinação de tecnologia e medicina em pesquisa e desenvolvimento facilita, atualmente, uma profusão de inovações que continuam sendo aprimoradas (Panesar, 2019). Muitos sistemas de IA relacionados à saúde existem ou estão em desenvolvimento para atender a uma variedade de necessidades dos atores interessados em cuidados de saúde e desenvolvimento terapêutico. Essas soluções têm usos abrangentes em todo o espectro da prestação de cuidados de saúde e desenvolvimento terapêutico. Por exemplo, os sistemas de IA estão sendo usados nos cuidados de saúde para apoiar os pacientes durante o diagnóstico e o tratamento de uma doença, aplicando soluções que favorecem a adesão à terapêutica e melhoram a capacidade de comunicação com os prestadores de cuidados.

Os cuidados de saúde têm se tornado mais centrados no paciente, com abordagens personalizadas para a tomada de decisões. Isso permite que os dados sejam usados para melhorar o bem-estar do paciente e da população, a educação e o engajamento do paciente, a prevenção e a previsão de doenças e riscos do tratamento, a adesão à medicação, a gestão da doença, a reversão/remissão da doença e a individualização e a personalização do tratamento e do cuidado. Para esses fins, a IA tem sido cada vez mais incorporada e utilizada no ambiente clínico. Por exemplo, dispositivos médicos habilitados para IA têm sido utilizados para apoiar a tomada de decisões clínicas, ao passo que os sistemas de IA podem facilitar a avaliação clínica de pacientes e a triagem de cuidados. Os sistemas de IA também estão sendo usados no desenvolvimento e na avaliação de produtos médicos, inclusive durante a descoberta de medicamentos para identificar potenciais candidatos terapêuticos e em pesquisas clínicas para o enriquecimento de pacientes. A Figura 1 ilustra as áreas de pesquisa e desenvolvimento de IA em todo o espectro da prestação de cuidados de saúde e desenvolvimento terapêutico. Embora não apresente uma lista exaustiva de todas as aplicações de IA, a figura fornece exemplos que se destinam a mostrar a ampla gama de usos atuais e potenciais de sistemas de IA.

FIGURA 1 - ESPECTRO GERAL DE PESQUISA E DESENVOLVIMENTO EM IA NA PRESTAÇÃO DE CUIDADOS DE SAÚDE E DESENVOLVIMENTO TERAPÊUTICO

FONTE: ELABORAÇÃO PRÓPRIA.

O espectro na Figura 1 ajuda a determinar quais considerações regulatórias podem ser aplicáveis e como elas podem ser implementadas. Este documento descreve uma seleção das principais considerações regulatórias e discute áreas temáticas relevantes para muitos atores interessados no atual ecossistema de IA para saúde.

ÁREAS TEMÁTICAS DE CONSIDERAÇÕES REGULATÓRIAS

Os sistemas de IA podem ser utilizados em todos os aspectos dos cuidados de saúde e do desenvolvimento terapêutico. Independentemente da categoria da aplicação desse sistema, os reguladores estão interessados em garantir não apenas que os sistemas de IA sejam seguros e eficazes para o uso pretendido, mas também que essas ferramentas promissoras cheguem a quem precisa delas o mais rápido possível. O diálogo entre todos os atores interessados que participam do ecossistema de IA para a saúde – especialmente desenvolvedores, fabricantes, reguladores, usuários e pacientes – é altamente aconselhável à medida que a comunidade de IA amadurece. Consequentemente, esta publicação visa estabelecer um entendimento comum do uso de sistemas de IA em saúde que possa ser relevante para os atores interessados.

Em 2020, os líderes dos subgrupos das áreas temáticas realizaram uma revisão sistemática da literatura de publicações científicas no PubMed e em outras bases de dados, que incluíram as diretrizes atuais e as boas práticas em cuidados de saúde e

desenvolvimento terapêutico. Essas fontes foram consideradas na definição da lista de áreas temáticas de considerações regulatórias para o uso de IA em cuidados de saúde e desenvolvimento terapêutico. Em sua primeira reunião, o GT-CR debateu as áreas temáticas propostas e buscou consenso para focar seu resultado nas seis principais áreas listadas na Tabela 1, ao mesmo tempo em que discutia as seções restantes desta publicação. O grupo de trabalho foi dividido em seis subgrupos, compostos por especialistas que elaboraram uma seção sobre cada área temática.

O GT-CR ressaltou que essa lista não inclui de maneira exaustiva todas as considerações importantes. O grupo de trabalho espera que a lista sirva de ponto de partida para futuras deliberações e atualizações subsequentes. Por exemplo, sistemas globais de proteção de propriedade intelectual (PI) podem ser uma área importante a ser discutida como parte de regulações de jurisdição cruzada para alguns atores interessados (principalmente desenvolvedores e fabricantes de sistemas de IA), e também em relação, por exemplo, à proteção de invenções relacionadas à IA por meio de leis sobre patentes e segredos comerciais. Embora não se tenha abordado neste relatório, a Organização Mundial da Propriedade Intelectual (OMPI) iniciou um diálogo sobre IA e PI (World Intellectual Property Organisation [WIPO], s.d.). Portanto, a OMS se engajará nesse trabalho, em conjunto com a OMPI e outros atores interessados relevantes.

TABELA 1 - SEIS PRINCIPAIS ÁREAS TEMÁTICAS DE CONSIDERAÇÕES REGULATÓRIAS

ÁREA TEMÁTICA 1	Documentação e transparência
ÁREA TEMÁTICA 2	Gestão de risco e abordagens do ciclo de vida de desenvolvimento de sistemas de IA
ÁREA TEMÁTICA 3	Uso pretendido e validação analítica e clínica
ÁREA TEMÁTICA 4	Qualidade dos dados
ÁREA TEMÁTICA 5	Privacidade e proteção de dados
ÁREA TEMÁTICA 6	Engajamento e colaboração

FONTE: ELABORAÇÃO PRÓPRIA.

DOCUMENTAÇÃO E TRANSPARÊNCIA

Documentação e transparência são conceitos críticos essenciais para facilitar avaliações científicas e regulatórias de sistemas de IA. Esses conceitos também ajudam a garantir a confiança não apenas no sistema de IA em si, mas também entre desenvolvedores, fabricantes e usuários finais. A documentação precisa e abrangente é essencial para permitir uma avaliação transparente dos sistemas de IA para a saúde. Isso inclui a adoção de uma abordagem do ciclo de vida total do produto para pré-especificar e documentar processos, métodos, recursos e decisões tomadas na concepção inicial, no desenvolvimento, no treinamento, na implantação, na validação (curadoria de dados ou ajuste de modelos) e na pós-implantação de sistemas de IA relacionados à saúde que possam exigir supervisão regulatória. A discussão a seguir concentra-se em alguns elementos relacionados à documentação e à transparência, mas não inclui a totalidade dos fatores relevantes para essa importante área.

A documentação eficaz e a transparência ajudam a estabelecer confiança e a proteger contra enviesamentos e exploração de dados. As mesmas expectativas e padrões regulatórios que garantem a segurança e a eficácia dos produtos regulados também se aplicam aos sistemas de IA usados em áreas reguladas. É importante que os reguladores sejam capazes de rastrear o processo de desenvolvimento e ter a documentação adequada das etapas essenciais e dos pontos de decisão. Por exemplo, os aspectos que exigem documentação cuidadosa incluem a especificação do problema que os desenvolvedores têm tentado resolver, o contexto no qual o sistema de IA foi concebido para funcionar e a seleção, a curadoria e o tratamento de conjuntos de dados de treinamento usados no processo de desenvolvimento.

A documentação deve permitir o rastreamento, o registro e a retenção de registros de etapas essenciais e decisões, incluindo justificativas e fundamentação para desvios dos planos pré-especificados. Uma documentação eficaz também pode ajudar a demonstrar que desenvolvedores e fabricantes estão levando em consideração toda a complexidade do contexto no qual se espera que o sistema de IA opere. Além

disso, desenvolvedores e fabricantes devem descrever como o sistema de IA está atendendo às necessidades dos usuários e o motivo pelo qual seria apropriado ampliar a base de usuários. Sem documentação transparente, torna-se difícil entender se as abordagens propostas serão generalizáveis a partir das evidências clínicas retrospectivas apresentadas na submissão regulatória para implantações no mundo real em novos cenários, o que pode reduzir acentuadamente o desempenho (Wu *et al.*, 2021). A Figura 2 mostra exemplos de etapas essenciais e pontos de decisão que desenvolvedores e fabricantes são encorajados a considerar para fins de documentação.

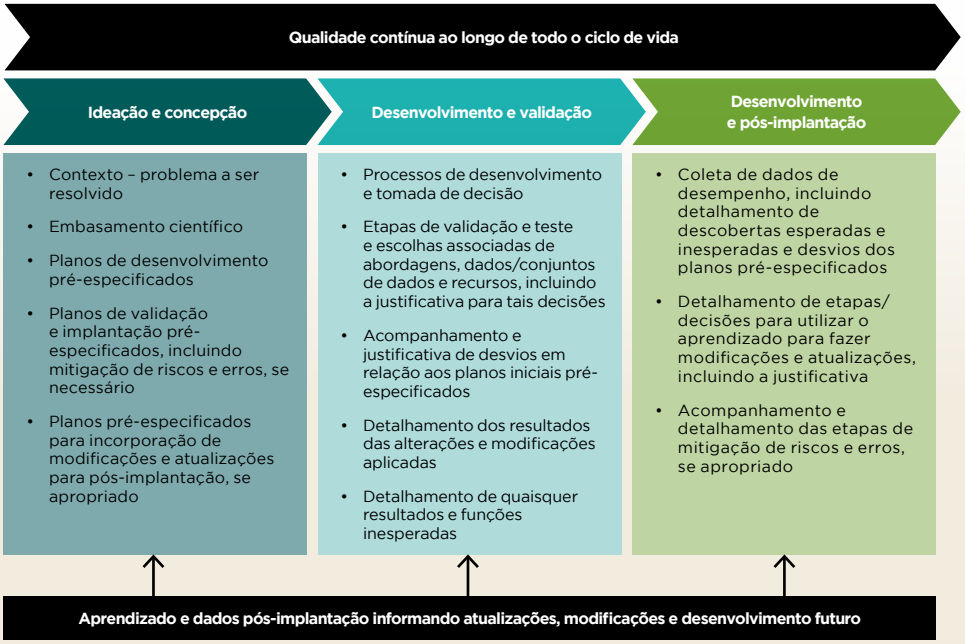
É provável que diferentes entidades com experiência multidisciplinar estejam envolvidas no desenvolvimento de sistemas de IA para a saúde e o desenvolvimento terapêutico. Logo, é necessário desenvolver uma compreensão compartilhada dos procedimentos necessários para uma documentação transparente e demonstrar que as decisões são cientificamente sólidas. Os sistemas usados para rastrear e documentar os processos de desenvolvimento e os principais pontos de decisão devem registrar o acesso e ser protegidos contra manipulação de dados e ataques adversários.

A documentação e a transparência não devem ser vistas como um fardo, mas como uma oportunidade para mostrar a força de um desenvolvimento baseado na ciência que considere todo o contexto em que se espera que o sistema de IA seja utilizado, incluindo as características dos usuários finais. Os instrumentos e os processos de documentação devem ser proporcionais aos riscos envolvidos e é aconselhado que sejam mantidas conversas com as autoridades reguladoras antes ou nos estágios iniciais de desenvolvimento, já que estas podem fornecer ajuda vital para informar as necessidades de documentação.

Além da perspectiva regulatória, é importante notar que a documentação eficaz e outras etapas que ajudam a garantir a transparência são maneiras importantes de estabelecer a confiança e uma compreensão compartilhada dos sistemas de IA em geral. Medidas para facilitar a transparência incluem: publicação em periódicos acadêmicos com revisão por pares, compartilhamento de dados e conjuntos de dados, e disponibilização de códigos para fomentar o aprendizado mútuo e facilitar estudos adicionais. Instituições acadêmicas,

periódicos médicos, organizações reguladoras e outros atores interessados têm trabalhado no avanço da transparência para o uso de IA no desenvolvimento diagnóstico e terapêutico.

FIGURA 2 - EXEMPLOS DOS PRINCIPAIS PONTOS DE DECISÃO NO DESENVOLVIMENTO DE SISTEMAS DE IA



FONTE: ELABORAÇÃO PRÓPRIA.

As colaborações, como a Padrões Consolidados de Relato de Ensaios para IA (Consolidated Standards of Reporting Trials for AI [CONSORT-AI]) (Liu *et al.*, 2020) e a Itens de Protocolo Padrão: Recomendações para Ensaios de Intervenção para IA (Standard Protocol Items: Recommendations for Interventional Trials for AI [SPIRIT-AI]) (Rivera *et al.*, 2020), forneceram orientações úteis sobre como projetar estudos para coletar evidências clínicas em que os sistemas de IA são usados e publicar os resultados. A transparência não é apenas uma consideração importante para a construção de confiança, mas também pode ser uma ferramenta útil para

educar os usuários finais. Isto pode ser alcançado, se for o caso, adaptando-se as estratégias de comunicação às necessidades dos usuários finais e de outros atores interessados, como pacientes e comunidades. Conforme descrito na Figura 2, o processo de desenvolvimento de um sistema de IA é multifacetado. Deve-se considerar uma abordagem metódica para a documentação durante todo o ciclo de desenvolvimento, incluindo a implantação e a pós-implantação.

A seguir, alguns elementos que podem ser úteis de se considerar para a documentação e a retenção de registros.

Documentação em todo o ciclo de vida do produto – garantindo uma qualidade contínua

Os desenvolvedores devem projetar, implementar e documentar abordagens e métodos para garantir a qualidade contínua em todas as fases de desenvolvimento. Uma documentação eficaz, que descreva todas as fases de desenvolvimento, aumentaria ainda mais a confiança no sistema de IA e mostraria como os desafios esperados e inesperados são identificados e administrados. Os processos de validação e *benchmarking* devem ser cuidadosamente documentados – incluindo as decisões para a seleção de conjuntos de dados específicos, padrões de referência, parâmetros e métricas para justificar tais processos. Por exemplo, deve-se ponderar cuidadosamente a documentação referente à forma e ao motivo pelos quais dados ou conjuntos de dados específicos são selecionados para treinar, validar externamente e retreinar o modelo (como o retreinamento pós-implantação).

Pré-especificação e documentação da finalidade médica, do contexto clínico e do desenvolvimento

A finalidade ou função médica pretendida dos sistemas de IA deve ser claramente documentada. Por exemplo, qual é o problema que o sistema de IA pretende resolver? Isso deve levar em consideração todos os contextos clínicos e de saúde nos quais se espera que uma ferramenta funcione. Os ambientes de cuidados clínicos podem ser muito complexos e envolver vários indivíduos com diferentes papéis e expectativas, por exemplo.

Assim, é necessário considerar e documentar como o sistema de IA deve funcionar em tais ambientes ativos. Como mostrado na Figura 3, há vários processos, etapas de teste/validação e protocolos que devem ser pré-especificados e documentados. Um dos elementos mais importantes que sustenta a confiança no processo de desenvolvimento é a pré-especificação. Isso evidenciará um processo de desenvolvimento coerente e embasará a justificativa para quaisquer mudanças futuras.

Implantação e pós-implantação

Os sistemas de IA podem ser projetados usando dados e conjuntos de dados de populações específicas. Como acontece com qualquer terapêutica, uma vez implantados, os sistemas de IA serão utilizados por uma população maior e usuários finais potencialmente variáveis. Deve-se considerar planos de implantação cuidadosos e uma justificativa para atingir diferentes usuários finais. Os fabricantes devem ser obrigados a realizar a vigilância pós-comercialização, que é o processo sistemático de coleta e análise da experiência adquirida com os sistemas de IA considerados dispositivos médicos que foram colocados no mercado (OMS, 2020). Desvios de planos pré-especificados, atualizações ou alterações no sistema de IA, desempenho pós-implantação, captura de dados e abordagens para avaliação contínua do sistema também devem ser documentados. Tais abordagens serão cada vez mais relevantes quando houver um entendimento mais amplo de que os sistemas de IA podem mudar após a implantação.

Abordagem baseada no risco e na proporcionalidade

Os marcos regulatórios recomendam uma abordagem baseada em riscos, com processos implementados para identificar e mitigar erros, vieses e outros riscos, de maneira proporcional à sua importância. Uma abordagem proporcional ao risco também deve ser levada em consideração para o nível de documentação e a manutenção de registros dos sistemas de IA. Os desenvolvedores de sistemas de IA devem ter em mente que as organizações regulatórias têm canais de diálogo e discussão que podem ser usados para lançar luz sobre os requisitos regulatórios.

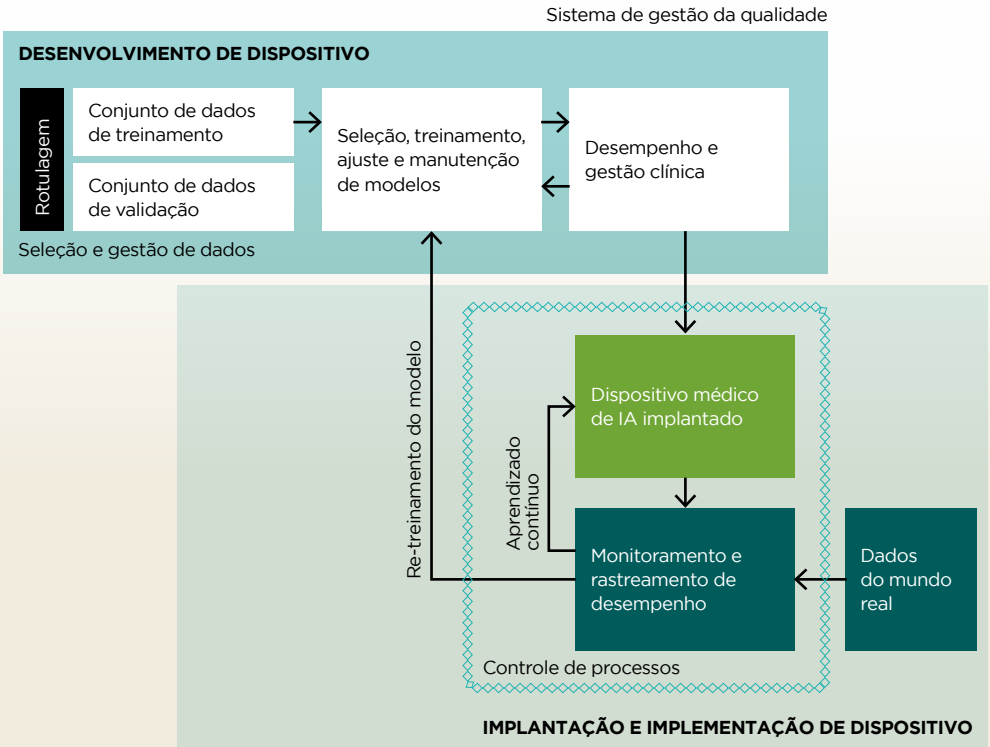
GESTÃO DE RISCOS E ABORDAGEM DO CICLO DE VIDA DE DESENVOLVIMENTO DE SISTEMAS DE IA

Os sistemas de IA se enquadram em muitas categorias, como dispositivos que dependem de IA e são usados como dispositivos médicos (comumente conhecidos como *Software* como Dispositivo Médico – *Software as a Medical Device* [SaMD]). Tais categorias de sistemas de IA são definidas pelo IMDRF como "*softwares* destinados a serem usados para um ou mais fins médicos que executam esses fins sem fazer parte de um dispositivo médico de *hardware*" (IMDRF, 2013). No entanto, as considerações regulatórias para tal categoria de sistemas de IA são semelhantes às de *softwares* tipicamente regulados como dispositivos médicos, com a adição de considerações que podem incluir capacidades de aprendizagem contínua, nível de intervenção humana, treinamento de modelos e reciclagem (IMDRF, 2013). Além disso, deve-se considerar uma abordagem holística de gestão de riscos que inclua a mitigação dos riscos associados às ameaças cibernéticas a um sistema de IA e às vulnerabilidades do sistema ao longo de todo o ciclo de vida do produto. Esta área temática visa apresentar uma abordagem holística baseada no risco para sistemas de IA em geral e aqueles usados como dispositivos médicos em particular, ao longo de seu ciclo de vida, inclusive durante a implantação pré e pós-comercialização.

Sistemas de IA durante o processo de desenvolvimento e implantação

A Figura 3 ilustra o processo de desenvolvimento e implantação de um sistema de IA. Desenvolvedores e implementadores devem estabelecer medidas para garantir o desenvolvimento responsável de sistemas de IA.

FIGURA 3 - PROCESSO DE DESENVOLVIMENTO E IMPLANTAÇÃO DO SISTEMA DE IA



FONTE: HSA (2022).

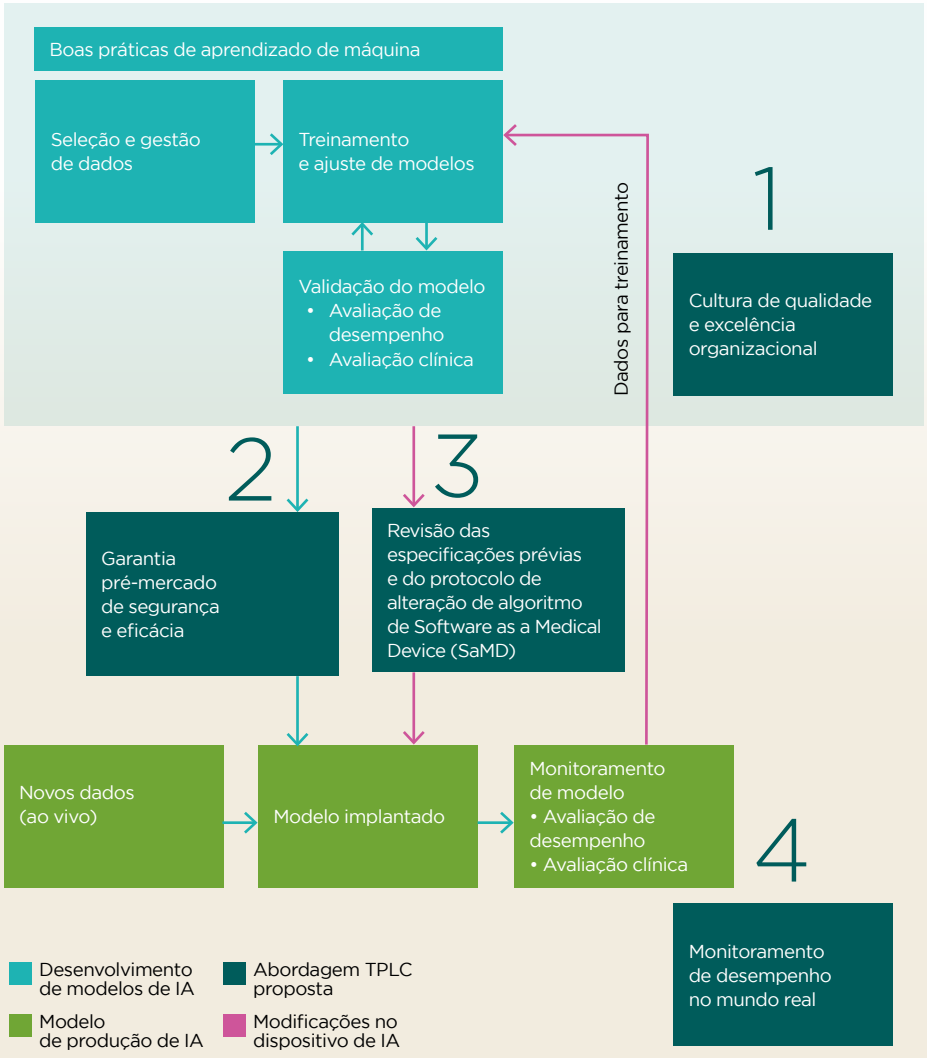
A Figura 3 mostra que todas as atividades relacionadas ao projeto, desenvolvimento, treinamento, validação, reciclagem e implantação de sistemas de IA devem ser realizadas e gerenciadas sob um sistema de gestão da qualidade baseado na ISO 13485 (HSA, 2022). Para desfechos clínicos, as dimensões de monitoramento específicas da IA incluem confiança (Oala *et al.*, 2021), viés e robustez (Oala *et al.*, 2022).

Ciclo de vida de desenvolvimento de sistemas de IA

Uma abordagem de ciclo de vida de desenvolvimento de sistemas de IA pode facilitar o aprendizado contínuo de IA e a melhoria do produto, ao mesmo tempo em que fornece salvaguardas eficazes. Isso pode ser alcançado se a abordagem do ciclo

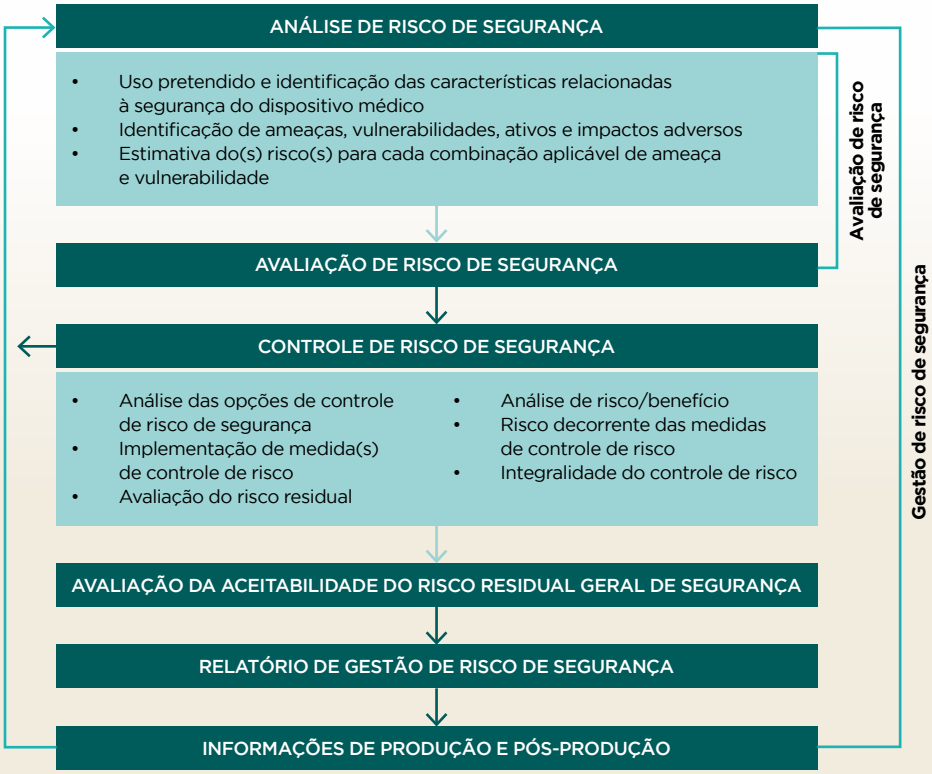
de vida de desenvolvimento envolver práticas apropriadas para o sistema de IA. Essa abordagem também poderia aumentar a confiabilidade e garantir o desempenho e a segurança do sistema de IA. Um exemplo é a abordagem do ciclo de vida total do produto (*total product lifecycle* [TPLC]) (FDA, 2019), que pode incluir os quatro componentes a seguir (conforme ilustrado na Figura 4):

FIGURA 4 - ABORDAGEM DO CICLO DE VIDA TOTAL DO PRODUTO DO SISTEMA DE IA NO FLUXO DE TRABALHO DE IA



FONTE: FDA (2019).

FIGURA 5 - REPRESENTAÇÃO ESQUEMÁTICA DO IMDRF DO PROCESSO DE GESTÃO DE RISCOS DE SEGURANÇA



FONTE: IMDRF (2019).

- demonstração de uma cultura de qualidade e excelência organizacional do fabricante dos sistemas de IA;
- garantia de segurança e desempenho pré-comercialização;
- revisão das pré-especificações dos sistemas de IA e do protocolo de mudança de algoritmo; e
- monitoramento de desempenho no mundo real.

Gestão holística de riscos

Deve-se considerar a avaliação e a gestão holísticas dos riscos, tendo em conta todo o contexto em que o sistema de IA pode ser

utilizado. Isso pode incluir não apenas o *software* ou sistema de IA em desenvolvimento, mas também outros *softwares* que podem ser usados no mesmo ambiente ou contexto. Outros riscos, como os associados a ameaças e vulnerabilidades de cibersegurança, devem ser considerados em todas as fases da vida de um dispositivo médico. Consequentemente, os fabricantes de sistemas de IA devem empregar uma abordagem baseada no risco para garantir que o *design* e o desenvolvimento de sistemas de IA usados como dispositivos médicos incluam proteções de cibersegurança adequadas. Isso exige que os fabricantes adotem uma abordagem holística para a cibersegurança do dispositivo, avaliando riscos e mitigações ao longo do ciclo de vida de desenvolvimento do sistema de IA. Para isso, o IMDRF publicou um processo de gestão de riscos de segurança, conforme ilustrado na Figura 5.

FIGURA 6 - ABORDAGEM GERAL DE GESTÃO DE RISCO DE DISPOSITIVOS MÉDICOS DE IA

FONTE: ELABORAÇÃO PRÓPRIA.

No entanto, para facilitar a gestão de risco de sistemas de IA, uma abordagem geral de gestão holística é introduzida nesta subseção com três grandes categorias de gestão: gestão de desenvolvimento pré-comercialização, gestão pós-comercialização e gestão de mudanças. Essas categorias estão ilustradas na Figura 6, e são debatidas a seguir:

- **Gestão de desenvolvimento pré-comercialização**: é necessário que haja transparência em relação ao funcionamento de quaisquer dispositivos fabricados e baseados em IA para garantir que os usuários possam ter uma melhor compreensão dos benefícios, riscos e limitações dos sistemas baseados em IA (FDA, 2021). Além disso, os controles e as medidas implementadas para garantir que esse sistema funcione conforme o esperado, minimizando o risco de danos, devem ser proporcionais aos riscos que poderiam ocorrer caso o sistema de IA funcionasse

mal. Por exemplo, a falha de um sistema de IA projetado para incentivar a adesão a uma dieta saudável é diferente daquela de um sistema de IA projetado para diagnosticar ou tratar certas doenças e patologias. Portanto, os desenvolvedores devem considerar uma abordagem baseada em risco em todos os processos, a fim de priorizar a segurança. É necessário que os desenvolvedores considerem tanto o uso pretendido do sistema de IA quanto o contexto clínico para avaliar o nível de risco. Por exemplo, a estrutura de risco do IMDRF para SaMD (IMDRF, 2014) identifica dois fatores principais que podem contribuir para o impacto ou risco de um sistema de IA. O primeiro fator é a importância das informações fornecidas pelo sistema de IA para a tomada de decisão em relação aos cuidados de saúde. A importância é determinada pelo uso pretendido da informação – para tratar ou diagnosticar, orientar a gestão clínica ou informar a gestão clínica. O segundo fator é a situação ou condição de saúde do paciente - que é determinada pelo tipo de usuário, doença ou condição e pela população em que se pretende utilizar o sistema de IA — ou seja, situações ou condições de saúde críticas, graves ou não graves. Em conjunto, esses fatores relacionados ao uso pretendido podem ser usados para classificar o sistema de IA em uma das quatro categorias, do menor risco (I) ao maior risco (IV), para refletir o risco associado à situação clínica e ao uso do dispositivo. O uso pretendido e a classificação de risco devem ser considerados ao testar diferentes modelos e equilibrar *trade-offs*, tais como transparência e precisão. Nos casos em que os conjuntos de dados de treinamento são limitados, modelos mais simples, como modelos de regressão ou de árvore de decisão, geralmente fornecem resultados equivalentes ou melhores do que modelos mais complexos e têm o benefício adicional de mais transparência e interpretabilidade. Por outro lado, nos casos em que conjuntos de dados são maiores e mais complexos, modelos complexos – como redes de aprendizado profundo (*deep learning* [AP]) – podem não ser facilmente explicáveis, mas podem fornecer maior precisão do que modelos mais simples. Entretanto, nos casos em que exista um maior risco de danos, os atores

interessados devem considerar a possibilidade de discutir os riscos e os benefícios da escolha de um modelo mais complexo e avaliar se existem formas de atenuar a falta de interpretabilidade e transparência e de criar confiança no modelo por meio de medidas de validação adicionais.

TABELA 2 - CLASSIFICAÇÃO DE RISCO DE SISTEMAS DE IA

Estado da situação ou condição de saúde	Importância das informações fornecidas pelo sistema de IA para a decisão em saúde		
	Tratar ou diagnosticar	Orientar a gestão clínica	Informar a gestão clínica
Crítico	IV	III	II
Grave	III	II	I
Não grave	II	I	I

FONTE: IMDRF (2014).

Além disso, dependendo do nível de risco, alguns sistemas de IA podem ser aprovados como disponíveis para implantação completa, enquanto outros podem ser inicialmente autorizados para implantação em instituições mais "prontas para IA". As instituições "prontas para IA" são aquelas certificadas com base em níveis rigorosos de vigilância, com sistemas de *backup* responsivos para lidar com qualquer falha do algoritmo, a fim de minimizar o risco de danos aos pacientes.

Em geral, é importante alcançar a transparência entre todos os atores interessados no sistema de IA, incluindo desenvolvedores, fabricantes, autoridades reguladoras e implementadores (ou seja, usuários em ambientes de cuidados de saúde, como médicos). A documentação adequada da gestão de riscos e os procedimentos de auditoria adequados são exemplos de maneiras que ajudam a garantir a transparência. De modo geral, a auditoria de componentes-chave específicos do dispositivo médico com IA deve ser considerada (como determinado *software*, *hardware*, dados de treinamento e casos de falha). Por exemplo, é importante fazer controle de versão com dados de treinamento, pois mais dados são adicionados a cada atualização. Se o desempenho de um algoritmo se deteriorar subitamente após uma atualização, pode ser desejável uma inspeção de tudo o que contribuiu para a atualização. Na maioria dos casos, o elemento que terá mudado é a adição

de novos dados de treinamento pelo desenvolvedor (em vez de alterações no *software* em si, como modificação nas redes neurais). Além disso, dado o quão imprevisíveis as mudanças no desempenho podem ser para a IA, recomenda-se ter relatórios ativos e investigação de casos de falha (nas diretrizes da CONSORT-AI) – embora não seja prescritivo, dada a ampla gama de opções de relatório e investigação disponíveis, desde uma auditoria clínica de senso comum (ou seja, inspeção humana) até soluções técnicas baseadas em inferência.

Embora não seja exclusivo da IA, existe uma teia cada vez mais densa de legislações e leis específicas de cada país, nação e jurisdição que os fabricantes e desenvolvedores podem precisar considerar para o desenvolvimento e a implantação de dispositivos médicos de IA para cuidados de saúde regulados. Esse conjunto de legislações inclui a Lei de Proteção de Dados Pessoais, a Lei de Pesquisa Biomédica Humana, a Lei de Hospitais Privados e Clínicas Médicas, a Lei de Portabilidade e Responsabilidade de Seguros de Saúde e o Regulamento Geral de Proteção de Dados (RGPD). A conformidade com legislações relevantes (leis locais, leis que abrangem várias jurisdições e leis de proteção de dados) precisa ser demonstrada pelos fabricantes e desenvolvedores de dispositivos médicos, independentemente de incorporarem um componente de IA ou não.

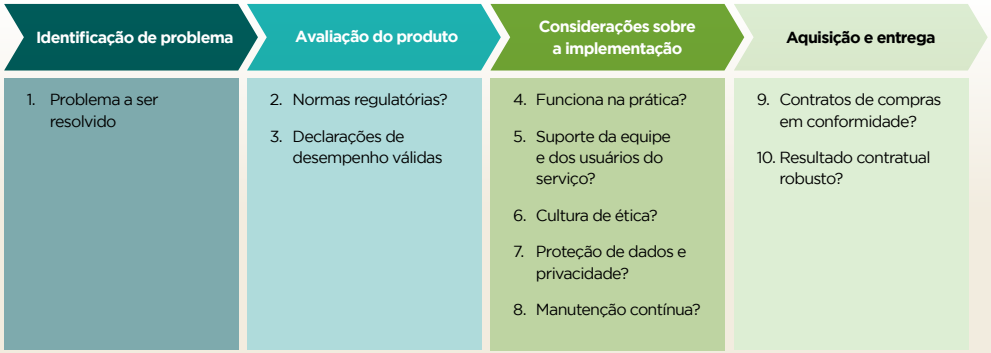
- **Gestão pós-comercialização**: o monitoramento e a vigilância pós-comercialização de dispositivos médicos de IA permitem a identificação oportuna de problemas relacionados a *software* e *hardware* que podem não ser observados durante o desenvolvimento, a validação e avaliação clínica do dispositivo. Novos riscos podem surgir quando o *software* é implementado em um contexto mais amplo no mundo real e usado por um espectro diversificado de usuários com diferentes conhecimentos. As empresas envolvidas na distribuição de dispositivos médicos com IA (fabricantes, importadores, atacadistas, representantes autorizados e registradores) são obrigadas a cumprir com seus deveres e obrigações pós-comercialização, que incluem relatórios às autoridades regulatórias relevantes em qualquer uma das seguintes circunstâncias (OMS, 2020; HSA, 2022):

- qualquer ameaça grave à saúde pública;
- morte, deterioração grave do estado de saúde do paciente, usuário ou outra pessoa;
- possibilidade de ocorrer morte, deterioração grave do estado de saúde do paciente, usuário ou outra pessoa;
- qualquer ação corretiva de segurança em campo (como a devolução de um tipo de dispositivo ao fabricante ou seu representante [também conhecido como *recall* em algumas jurisdições]; modificação do dispositivo; troca do dispositivo; destruição do dispositivo; conselhos dados pelo fabricante sobre o uso do dispositivo).

Além disso, os fabricantes devem coletar informações de forma proativa (por meio de literatura científica e outras fontes de informação, tais como bases de dados das autoridades reguladoras acessíveis ao público, treinamento dos usuários e pesquisas) como parte de seu plano de fiscalização pós-comercialização. O plano deve descrever a forma como os fabricantes devem monitorar e responder ativamente aos riscos, em evolução e recentemente identificados. As principais considerações para o plano de vigilância pós-comercialização incluem (HSA, 2022): divulgação de vulnerabilidades, gerenciamento de *patches* e atualizações, recuperação e compartilhamento de informações. Além disso, como parte dos deveres e obrigações pós-comercialização, as empresas envolvidas na distribuição de dispositivos médicos (fabricantes, importadores, atacadistas e registradores) são obrigadas a relatar eventos adversos associados ao uso de dispositivos médicos de *software* aos reguladores relevantes.

Em geral, há necessidade de acompanhamento do desempenho clínico pós-comercialização e verificações periódicas de segurança para relatar qualquer dano potencial. A intensidade da vigilância pós-comercialização pelo fabricante pode ser proporcional ao risco (em função de consequências da falha [criando um risco potencial de dano] e da probabilidade de detecção precoce dessa falha). Por último, a vigilância pós-comercialização requer um nível mínimo de avaliação para cada local, a fim de garantir que as potenciais vulnerabilidades dos algoritmos devido à variação nos ambientes locais possam ser detectadas.

FIGURA 7 - SERVIÇO NACIONAL DE SAÚDE (NATIONAL HEALTH SERVICE [NHS]) DO REINO UNIDO - *UM GUIA DO COMPRADOR DE IA NAS ÁREAS DE SAÚDE E CUIDADOS DE SAÚDE*



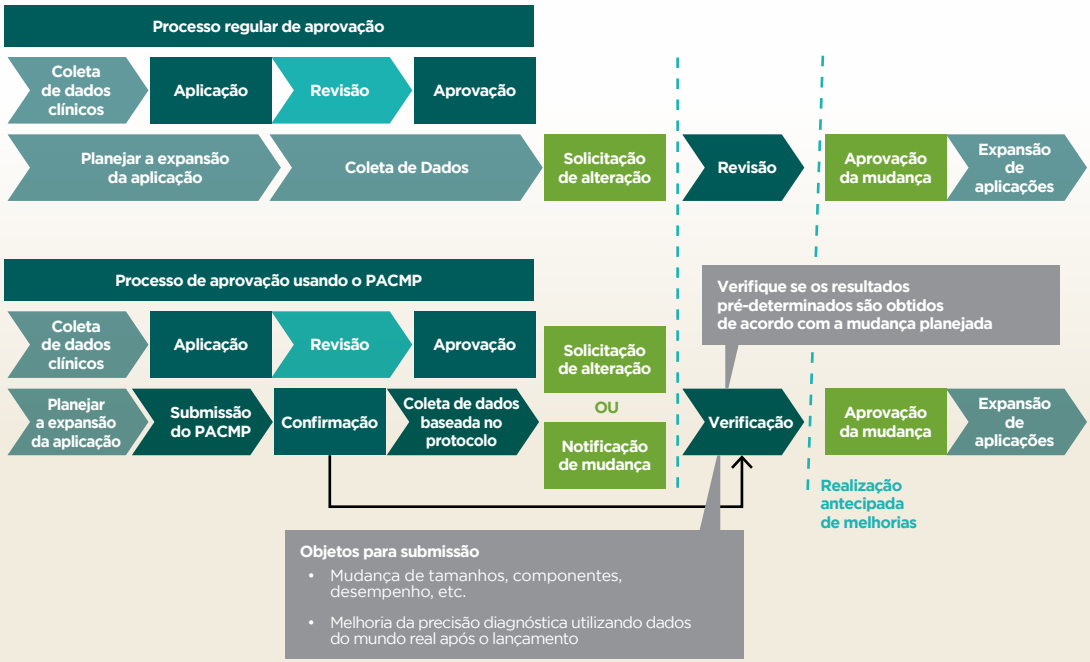
FONTE: NHS (2020).

Por exemplo, o Laboratório de IA do Serviço Nacional de Saúde do Reino Unido da Grã-Bretanha e Irlanda do Norte publicou orientações para acelerar uma adoção segura e eficaz de IA na saúde (NHS, 2022). O guia lista dez perguntas em quatro categorias para ajudar os compradores de produtos de IA a tomar decisões informadas, identificar problemas, avaliar produtos e considerar questões relacionadas à implementação, à aquisição e à entrega (Figura 7).

- **Gestão de mudanças**: Tendo em vista o caráter dos sistemas de IA, é importante que o sistema regulatório permita modificações contínuas para melhoria ao longo do ciclo de vida de desenvolvimento do sistema de IA. O termo "mudança" refere-se a tais modificações, incluindo aquelas realizadas durante a manutenção. Existem vários modelos e abordagens de gestão de mudanças propostos para sistemas baseados em IA. Alguns consideram a mudança como parte do ciclo de vida total do desenvolvimento (como na abordagem TPLC) (FDA, 2019) (Figura 4). Outros modelos concentram-se no processo de gestão de mudanças no ciclo de vida total de dispositivos médicos que podem ser continuamente melhorados. Um exemplo é a abordagem implementada pelo Ministério da Saúde, Trabalho e Bem-Estar do Japão e adaptada na Lei dos Produtos Farmacêuticos

e Dispositivos Médicos como Protocolo de Gestão de Mudanças Pós-Aprovação (PACMP) para dispositivos médicos (Notificação do Departamento de Dispositivos Médicos e Produtos Farmacêuticos n. 14/2021), (Figura 8).

FIGURA 8 - PROTOCOLO DE GESTÃO DE MUDANÇAS PÓS-APROVAÇÃO PARA DISPOSITIVOS MÉDICOS



FONTE: ELABORAÇÃO PRÓPRIA.

USO PRETENDIDO E VALIDAÇÃO ANALÍTICA E CLÍNICA

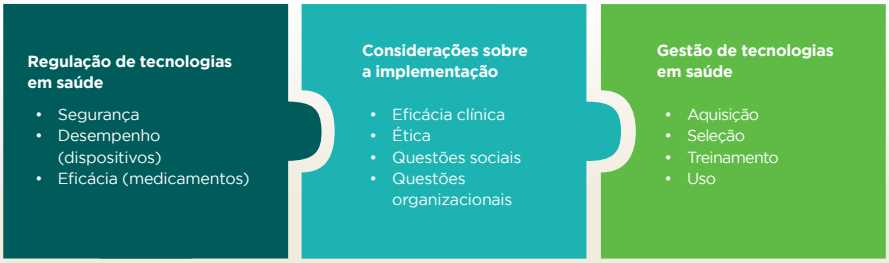
Em princípio, mecanismos regulatórios existem para responder à pergunta: "os dados disponíveis (incluídos na submissão regulatória) apoiam a conclusão de que um sistema de IA em fase de pesquisa ou experimental é seguro e funciona suficientemente bem para justificar a entrada no mercado e o acesso pelo público?". Além dos princípios discutidos em "Documentação e transparência" e "Gestão de riscos e abordagem do ciclo de vida de desenvolvimento de sistemas

de IA", deve-se avaliar também se o uso do sistema é seguro (ou seja, se não prejudicará o usuário, o paciente ou outras pessoas) e se as alegações feitas sobre seu desempenho podem ser verificadas (ver Figuras 9 e 10). A avaliação dessas alegações para sistemas de IA requer uma descrição clara do caso de uso, demonstração de validação analítica e clínica e avaliação do potencial de viés ou discriminação no sistema de IA.

Descrição do caso de uso, validação analítica e clínica

Avaliação clínica é a revisão de evidências que demonstram a segurança e o desempenho de um determinado produto para um uso pretendido específico. Para sistemas de IA (especialmente dispositivos que dependem de IA e são usados para fins médicos), a orientação é útil para coletar evidências de validação analítica e clínica. O desempenho dos sistemas de IA pode ser alterado rapidamente – não apenas como resultado de uma mudança de código, mas também para fornecer dados de treinamento/ajuste diferentes ou adicionais. Consequentemente, a avaliação clínica que leva em conta o TPLC desde o desenvolvimento até a validação analítica e clínica e a vigilância pós-comercialização deve ser considerada para os sistemas de IA.

FIGURA 9 - DOMÍNIOS DA REGULAÇÃO, AVALIAÇÃO E GESTÃO DE TECNOLOGIAS EM SAÚDE PARA MEDICAMENTOS E DISPOSITIVOS



FONTE: ELABORAÇÃO PRÓPRIA.

FIGURA 10 - DESCRIÇÃO DO IMDRF DOS COMPONENTES DA AVALIAÇÃO CLÍNICA

AVALIAÇÃO CLÍNICA		
Associação clínica válida	**Validação analítica**	**Validação clínica**
Existe uma associação clínica válida entre o resultado do seu SaMD e a condição clínica visada pelo seu SaMD?	O seu SaMD processa corretamente os dados de entrada para gerar dados de saída precisos, confiáveis e exatos?	O uso dos dados de saída precisos, confiáveis e exatos do seu SaMD atinge o propósito pretendido em sua população-alvo no contexto do atendimento clínico?

FONTE: FDA (2019).

Essa área temática abrange as considerações sobre as descrições de casos de utilização (incluindo declarações de uso pretendido) e a validação analítica e clínica. Essas considerações seguem o *framework* proposto pelo Grupo de Trabalho sobre Avaliação Clínica (GT-AC) do FG-AI4H da OMS/UIT (UIT, 2020). Uma descrição completa desse *framework* pode ser encontrada no produto final do GT-AC. A seção a seguir descreve as principais considerações e as melhores práticas e baseia-se no importante trabalho das autoridades reguladoras nacionais e regionais e dos organismos internacionais, como o IMDRF. Não se destina a substituir o trabalho desses órgãos. Ao delinear as principais considerações, este relatório chama a atenção para os desafios que permanecem nesse campo em rápida transformação. Por exemplo, é dada especial atenção a ambientes com poucos recursos, que podem ter capacidade regulatória limitada em nível nacional. O papel do *benchmarking* na avaliação de sistemas de IA em saúde também é explorado. Os princípios de avaliação são aplicados a essa área temática, bem como ao trabalho do FG-AI4H da OMS/UIT, no qual o *benchmarking* de avaliação é um componente importante (UIT, 2021).

Uso pretendido

Os sistemas de IA são complexos, dependentes não apenas dos códigos que os constituem, mas também dos dados de treinamento, do ambiente clínico e da interação com o usuário. Eles geralmente situam-se em um percurso clínico complexo ou estão sendo introduzidos em novos percursos clínicos (por

exemplo, em novos percursos de telemedicina ou como parte de novas ferramentas de triagem). Portanto, para sistemas de IA, a segurança e o desempenho podem ser altamente dependentes do contexto. A descrição do caso de uso tem um papel importante tanto para informar os usuários finais sobre onde a ferramenta pode ser utilizada de forma segura e adequada, como nos sistemas de IA regulados (a declaração de uso pretendido), para permitir que os reguladores avaliem se as evidências das etapas de validação analítica e clínica são adequadas e suficientes para o uso pretendido.

Ao desenvolver um sistema de IA relacionado à saúde, é importante descrever o caso de uso relevante. Essa consideração deve abranger o cenário (geografia, tipo de estabelecimento de saúde), a população (etnia, raça, sexo, idade, tipo de doença, gravidade da doença, comorbidades), o usuário pretendido (profissional de saúde ou paciente) e a situação clínica a que se destina. Muitas intervenções, testes e diretrizes são propensos a vieses, e essa é uma consideração particularmente importante para sistemas de IA, que são altamente sensíveis às características dos dados com os quais foram treinados e propensos a falhas com tipos de dados não vistos (como uma nova característica da doença ou um tipo de população ou contexto que não foi previamente encontrado). Os desenvolvedores e os fabricantes também devem fornecer uma explicação clínica e científica clara do desempenho pretendido de sua ferramenta, incluindo as populações e os contextos para os quais seu uso foi validado. Modelos de relatórios padronizados comuns a todos os atores interessados podem ajudar a comunicar o uso pretendido de forma mais eficaz (Sendak *et al.*, 2020; Verks & Oala, 2020; Oala *et al.*). Para alguns casos de uso pretendido, pode haver razões claras pelas quais o desempenho analítico da ferramenta difere em diversos contextos (Willis & Oala, 2021) (por exemplo, um verificador de sintomas pode ter um desempenho diferente em áreas com uma epidemiologia da doença diferente dos dados em que foi treinado). Se for esse o caso, as diferenças sistemáticas conhecidas no desempenho devem ser incluídas na declaração de uso pretendido. Para outros casos de uso pretendido, pode haver evidências emergentes de que a ferramenta em questão, ou outra muito semelhante, demonstrou

ter desempenho analítico semelhante em um cenário mais amplo do que aqueles em que a ferramenta foi inicialmente desenvolvida e validada (Calderon-Ramirez & Oala, 2021) (por exemplo, ferramentas retinianas demonstraram ter um desempenho semelhante em diferentes populações (Bellemo *et al.*, 2019)). A compreensão da generalização de ferramentas semelhantes pode ser levada em conta ao fornecer uma declaração do uso pretendido ou uma descrição do caso de uso (Mcdonald *et al.*, 2021).

Como parte do processo de gestão de riscos, os reguladores podem querer solicitar evidências de que os desenvolvedores consideraram se há situações em que uma ferramenta não deve ser usada (por exemplo, se houver dados de treinamento insuficientes para um grupo de pacientes específico ou ausência de validação em um ambiente específico), ou se há riscos potenciais de uso fora dos contextos pretendidos.

Validação analítica (também conhecida como validação técnica)

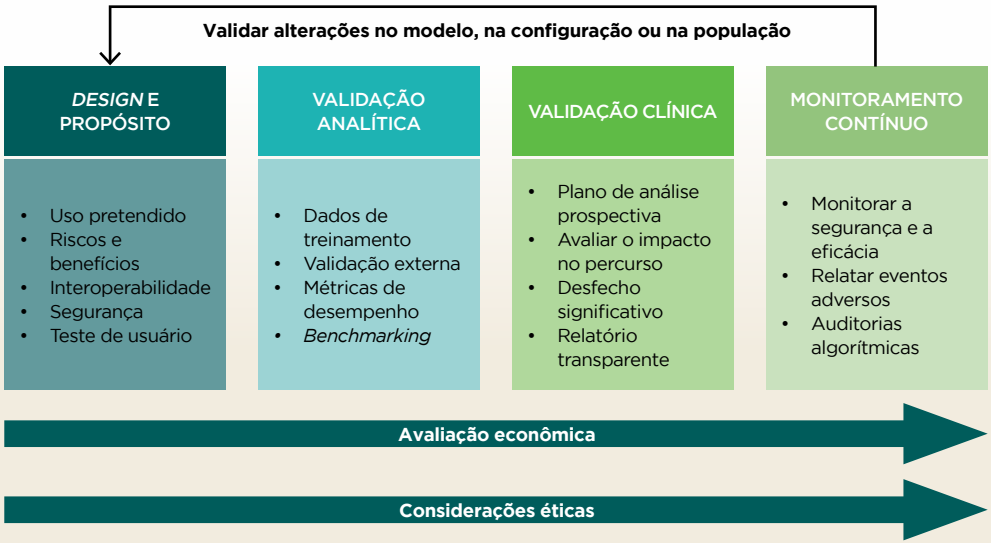
Para os fins do presente documento, validação analítica refere-se ao processo de validação do sistema de IA usando dados, mas sem realizar estudos clínicos ou de intervenção. Também pode-se chamar isso de validação técnica. A validação analítica apropriada demonstra que um modelo é robusto e tem um desempenho aceitável no contexto pretendido. Permite também compreender os potenciais enviesamentos e a generalização (e as medidas tomadas para compreendê-los).

Desenvolvedores e fabricantes devem fornecer uma descrição dos conjuntos de dados usados no treinamento, no ajuste, no teste e na validação interna do sistema de IA. A descrição do caso de uso pretendido (que pode estar em modelos padronizados de relatório) deve abranger o tamanho, o contexto, a demografia da população, o usuário pretendido e a situação clínica (com dados de entrada e saída). A transparência e a documentação sobre a seleção e as características do conjunto de dados são fundamentais para garantir que os sistemas de IA sejam usados adequadamente. Os desenvolvedores e os reguladores podem esperar que o sistema de IA tenha sido validado externamente em um conjunto de dados diferente daquele em que foi treinado e testado, a fim de demonstrar

a validade externa e a generalização do modelo além do conjunto de dados original. Espera-se que o conjunto de dados de validação externa seja representativo do contexto e da população descritos no uso pretendido (gênero, raça, etnia) para demonstrar um desempenho robusto no contexto pretendido. O conjunto de dados de validação deve ser de qualidade adequada, com robustez adequada dos rótulos. Como parte do processo de gestão de riscos, é importante identificar quaisquer casos que sejam, ou possam ser, de alto risco (Oala *et al.*, 2020).

Embora o enviesamento, os erros e a falta de dados não sejam exclusivos do desenvolvimento da IA, não deixam de ser preocupações sérias, que podem surgir por muitas razões – incluindo conjuntos de dados de treinamento ou de validação desiguais e não representativos, ou enviesamento estrutural nos sistemas em que os dados de treinamento são gerados (por exemplo, nos contextos dos cuidados de saúde). Informar gênero, raça e etnia das pessoas que fazem parte dos grupos de dados de treinamento e de validação, se possível, pode ajudar a resolver o potencial de enviesamento e evitar o seu impacto. Por exemplo, uma melhor compreensão do viés pode ajudar a identificar populações para as quais um sistema de IA pode não funcionar como esperado. A vigilância pós-comercialização também pode fornecer informações sobre o impacto de possíveis vieses.

FIGURA 11 - VISÃO GERAL DO QUADRO PARA AVALIAÇÃO CLÍNICA DE MODELOS DE IA EM SAÚDE DESENVOLVIDO PELO GT-AC



FONTE: ELABORAÇÃO PRÓPRIA.

Pode ser difícil obter conjuntos de dados para formação, ensaio e validação suficientemente representativos e de qualidade satisfatória. Organismos locais, regionais e nacionais interessados em adquirir sistemas de IA poderiam manter seu próprio conjunto de dados ocultos para permitir a validação externa (por exemplo, um programa recente do NHSX, do Reino Unido da Grã-Bretanha e da Irlanda do Norte, tem conjuntos de dados representativos em nível nacional para alguns casos comuns de uso). O acesso a conjuntos de dados representativos para validação é uma preocupação especial em muitos países de baixa e média renda. Quando os conjuntos de dados estão disponíveis em contextos com poucos recursos, também pode haver limitações na qualidade dos dados. A capacidade de produzir conjuntos de dados robustos com rótulos de verdade fundamental de alta qualidade provavelmente será afetada por limitações em outros aspectos do ambiente de saúde, onde podem existir barreiras que impeçam o acesso a diagnóstico e tratamento. Esses grandes desafios – com

potencial não apenas de propagar a desigualdade de acesso, mas também de comprometer a segurança e o desempenho de ferramentas baseadas em IA – são potenciais áreas para trabalhos futuros. A esse respeito, o recém-lançado International Digital Health & AI Research Collaborative (i-DAIR[2]) observa que o uso colaborativo, distribuído e responsável de dados está no centro de seu plano estratégico.

Embora a maioria das agências reguladoras tenha atribuições nacionais ou regionais, alguns países com capacidade regulatória limitada tendem a confiar em decisões tomadas por outros reguladores importantes. A disponibilidade de conjuntos de dados independentes, ocultos e representativos também oferece vantagens particulares aos países que não têm o próprio processo regulatório ou onde as decisões regulatórias podem ser informadas por dossiês fornecidos a outros órgãos. No entanto, o desempenho de sistemas baseados em IA é altamente dependente do contexto. A fim de confiar na revisão e nas decisões regulatórias, muitos reguladores (nacionais ou regionais) poderiam realizar a validação analítica como uma segunda etapa de validação local para garantir que as métricas de desempenho obtidas sejam consistentes com as demonstradas em outras jurisdições regulatórias. Isso poderia ser mais bem priorizado por meio de uma abordagem baseada nas necessidades – por exemplo, a identificação de áreas-chave nas quais as ferramentas baseadas em IA são promissoras e poderiam fornecer valor local – e a potencial criação prospectiva de conjuntos de dados para apoiar a validação.

Para entender o desempenho de um sistema de IA, deve-se avaliá-lo em relação a um padrão aceito. Embora o padrão mais adequado para comparação possa variar de acordo com o uso pretendido, os padrões comumente usados são o desempenho humano em uma tarefa semelhante ou outros modelos (por exemplo, derivados de regressão logística) com fortes padrões de precisão, sensibilidade e especificidade baseados em evidências ou obrigatórios (como em ferramentas de triagem). Dependendo do caso de uso pretendido, o requisito de desempenho comparativo pode ser mais ou menos rigoroso (por exemplo, quando

2 Saiba mais: http://i-dair.org

usado como ferramenta de triagem ou rastreamento, um nível diferente de desempenho comparativo pode ser aceitável em comparação com uma ferramenta usada para diagnóstico).

Foram realizados alguns *benchmarking*s comparativos limitados de sistemas de IA em um único contexto, mas estes podem se tornar mais comuns à medida que aumenta o número de ferramentas disponíveis (Salim *et al.*, 2020). No futuro, se um sistema de IA tiver eficácia e segurança clínicas comprovadas em um contexto específico, pode ser viável e apropriado comparar outras ferramentas mais recentes com esse sistema de IA para entender a potencial semelhança de desempenho. *Softwares* de *benchmarking* têm sido desenvolvidos como parte do trabalho da Open Code Initiative (UIT, 2022). Plataformas como essa também podem ser úteis como formas de realizar repetidas validações algorítmicas de modelos atualizados. Entretanto, isso não é o que acontece para nenhum caso de uso, e o *benchmarking* até agora tem sido usado apenas para entender o desempenho analítico comparativo. Além disso, o uso repetido dos mesmos dados para aferição comparativa de vários modelos atualizados (e, portanto, mesmo que inadvertidamente, para preparar o teste) corre o risco de introduzir vieses, e isso deve ser levado em conta ao se considerar o *benchmarking*.

Um grupo de trabalho do FG-AI4H designado para os métodos de avaliação de dados e soluções de IA[3] fornece orientações sobre os métodos, processos e desenvolvimento de *software* para a validação analítica dos sistemas de IA relacionados à saúde (Oala *et al.*, 2020).

Validação clínica

A validação analítica realizada retrospectivamente em um conjunto de dados existente fornece medidas de desempenho (precisão, valor preditivo negativo, valor preditivo positivo), porém não permite a avaliação de outros fatores que podem afetar o desempenho da ferramenta (por exemplo, interação do usuário, integração do fluxo de trabalho e consequências não intencionais da ferramenta dentro de um percurso clínico complexo).

3 Saiba mais: aiaudit.org

Tanto organismos nacionais quanto internacionais propuseram um conjunto graduado de requisitos com base no risco para ferramentas digitais de saúde (incluindo a importância das informações fornecidas pela ferramenta e o estado da condição de saúde) (IMDRF, 2016; National Institute for Health and Care Excellence [NICE], 2019). O documento do IMDRF sobre avaliação clínica de SaMD (Tabela 2) (IMDRF, 2014) propõe que os dispositivos da categoria I são as ferramentas de menor risco com evidências de validade analítica, e que uma nova ferramenta nessa categoria exigiria que os fabricantes coletassem dados de desempenho do mundo real e gerassem uma demonstração de validade científica. Relativamente aos SaMD de alto risco, espera-se que sejam apresentadas evidências de avaliação clínica com base em evidências de validade analítica. Não existe um consenso universal sobre o nível apropriado de evidências de desempenho clínico adequado para uma nova ferramenta de IA antes de sua implantação, sendo este assunto objeto de um grupo de trabalho separado no âmbito do FG-AI4H (GT-AC).

Dados de ensaios clínicos randomizados são o padrão-ouro na avaliação do desempenho clínico comparativo e podem ser apropriados para dispositivos de maior risco, nos quais uma ferramenta de IA não tem desempenho demonstrado nesse contexto, ou para grandes órgãos de aquisição nacional que buscam avaliar o desempenho antes de realizar uma despesa nacional. Espera-se que um ensaio que guie a prática clínica deva ter um desfecho primário clinicamente significativo (morbidade, mortalidade), mas, em certas situações, a taxa de eventos ou o intervalo de tempo entre o ensaio e o desfecho pode resultar em um desfecho substituto mais viável. Diretrizes para a elaboração de relatórios, apoiadas pela amplamente aceita rede EQUATOR, estão atualmente disponíveis para protocolos e ensaios clínicos usando sistemas de IA (Liu et al., 2020). No entanto, atualmente, ainda há um pequeno número de ensaios randomizados em andamento ou concluídos nesta área (Topol, 2020).

Os ensaios clínicos randomizados têm potenciais limitações que podem tornar preferíveis outras opções (podem ser lentos ou dispendiosos e avaliar o desempenho de grupos específicos em condições de ensaio). Quando evidências

randomizadas podem não ser necessárias (por exemplo, a evidência requerida pode ser proporcional ao risco ou o custo de uma ferramenta), a validação prospectiva em um ensaio de implantação e implementação no mundo real, com um grupo de comparação relevante que mostre melhoria em desfechos significativos usando ferramentas validadas ou desfechos amplamente aceitos e verificados e com relatórios sistemáticos de segurança, pode ser apropriada. O desempenho clínico deve ser considerado no contexto da capacidade dos profissionais de saúde, da largura de banda da Internet disponível e da infraestrutura de informática em saúde e dos *pipelines* de dados em tempo real. Os desenvolvedores devem fornecer uma descrição das etapas realizadas para conduzir a validação clínica em um contexto semelhante ao disponível no ambiente de uso pretendido.

Considerações adicionais sobre o nível ou tipo mais adequado de avaliação clínica para uma intervenção em saúde digital será fornecida pelo GT-AC.

Em algumas situações, conforme descrito abaixo, aplicam-se considerações especiais. Por exemplo:

Monitoramento pós-comercialização

O monitoramento pós-comercialização em alguns contextos regulatórios depende fortemente da notificação de eventos adversos. Orientações recentes da OMS recomendam que a vigilância proativa pós-comercialização deve ser realizada pelo fabricante.

Como parte de uma abordagem TPLC para a regulação nesse contexto, deve ser efetuado o acompanhamento clínico prospectivo pós-comercialização após a implantação. As entidades reguladoras devem ser notificadas sobre incidentes notificáveis (acontecimentos adversos), e os resultados de um monitoramento mais contínuo utilizando dados do mundo real podem ajudar os desenvolvedores e as entidades reguladoras a compreender melhor e a garantir a segurança e o desempenho desses dispositivos no uso no mundo real. Para o monitoramento prospectivo de dados do mundo real, será necessário um investimento significativo na curadoria prospectiva e na rotulagem de dados de validação. Um período definido de monitoramento próximo pode ser apropriado para ferramentas

baseadas em IA para aquelas com alto risco, devido a tendência a se ajustar excessivamente a características de dados errôneos e a produzir erros imprevisíveis a partir de características de dados não vistos, combinado com a falta de dados provenientes do uso em contextos do mundo real com resultados a longo prazo. As entidades reguladoras podem recomendar que os fabricantes desenvolvam medidas específicas de fiscalização do mercado que sejam adequadas aos sistemas de IA.

Mudanças na ferramenta de IA

Uma atualização de uma ferramenta de IA por uma mudança de código, mudança da interface do usuário ou fornecimento de dados de treinamento adicional podem alterar o desempenho analítico ou clínico de um sistema de IA. O grupo não tem conhecimento de sistemas médicos de IA atualmente aprovados que estejam "aprendendo continuamente", mas prevê que possam ser desenvolvidos. Tais sistemas de IA exigiriam uma avaliação de risco-benefício de acordo com os conceitos deste documento e com a avaliação clínica de sistemas de IA para a saúde. A adoção de "pontos de controle" – analisando a ferramenta, em intervalos regulares, no momento em que estiver funcionando – permite uma avaliação regular e pode assinalar alterações no desempenho. Dependendo do risco dos sistemas de IA e da extensão das mudanças, a validação apropriada deve ser acordada pelo desenvolvedor e pela entidade reguladora. A validação analítica em relação a conjuntos de dados não vistos – ou a avaliação comparativa em relação a conjuntos de dados aprovados representativos do contexto ou da população pretendidos – pode ser útil nesse cenário.

Países de baixa e média renda

Há uma variação considerável na regulação de aplicação dos dispositivos médicos e, por conseguinte, também nas tecnologias de IA implantadas e nos sistemas de IA desenvolvidos. Alguns países carecem de um órgão regulador nacional dedicado. As reuniões do GT-CR proporcionaram um fórum para o compartilhamento de conhecimentos e a discussão de problemas comuns, inclusive para órgãos reguladores e outros atores interessados, algumas das quais têm mandatos alinhados. Além disso, existem importantes considerações

regulatórias relacionadas ao uso pretendido e à validação analítica e clínica de sistemas de IA em saúde. Primeiro, em países de baixa e média renda, um dos usos potenciais das tecnologias de IA é trazer sistemas ou conhecimentos especializados baseados em IA para áreas que não dispõem de um especialista médico relevante (por exemplo, interpretação de exames da retina, lâminas histopatológicas ou imagens radiológicas). Em países de alta renda, os sistemas de IA são mais frequentemente posicionados como um complemento aos profissionais médicos. Usar uma avaliação realizada para apoiar a regulação em um contexto de alta renda para informar a utilização desses sistemas de IA em contextos de baixa ou média renda pode, portanto, não ser apropriado. Assim, todo o contexto de infraestrutura e recursos de saúde deve ser considerado. Em segundo lugar, alguns órgãos reguladores dependem de decisões de outros órgãos para apoiar seu trabalho regulatório. Dado que o desempenho dos sistemas de IA pode ser altamente dependente do contexto, poderão ser necessárias etapas adicionais. Existe o receio de que os desenvolvedores não garantam a adaptação ou a avaliação em contextos de recursos limitados se o mercado for menos atrativo. Agências reguladoras em países de alta renda poderiam apoiar essa adaptação, o que também aumentaria a generalização e a robustez dos sistemas de IA. No entanto, isso exigiria estudos adaptativos para garantir um uso mais amplo em países de baixa e média renda ou a utilização de incentivos para encorajar desenvolvimento, testes e validação adicionais. A disponibilidade de uma gama de conjuntos de dados representativos daria suporte à validação analítica local. Finalmente, os sistemas de IA para a saúde podem ser altamente sensíveis a mudanças na distribuição e nas características dos dados. Podem, portanto, ser sensíveis a diferenças na prevalência de doenças ao se deslocarem de países de alta renda para países de baixa renda, com a possibilidade de desempenho inferior sem avaliação ou ajuste adequados aos dados locais.

QUALIDADE DOS DADOS

Dados nos ecossistemas de saúde atuais

O setor de saúde tem sido muito receptivo aos benefícios da IA graças à explosão de dados e ao acesso ao poder computacional. Os dados são o ingrediente mais importante para o treinamento de algoritmos de IA/AM e podem ser classificados com base no formato, na estrutura, no volume e em muitos outros fatores. Os dados podem assumir qualquer forma, incluindo caracteres, texto, palavras, números, imagens, som ou vídeo. Ademais, esses dados podem ser estruturados, semiestruturados ou não estruturados (Panesar, 2019). Os dados estruturados são normalmente armazenados em bancos de dados que são estruturados de uma maneira que segue um modelo ou esquema específico – como dados armazenados em prontuários eletrônicos, dispositivos móveis e dispositivos de Internet das Coisas (IoT). Independentemente do formato, estrutura ou volume dos dados, uma classificação mais geral pode se basear nos dez "v" (em inglês) de dados (Panesar, 2019) (conforme ilustrado na Figura 12): volume, veracidade, validade, vocabulário, velocidade, imprecisão (*vagueness*), variabilidade, local (*venue*), variedade e valor.

Dados de boa qualidade em sistemas de IA de saúde

Todas as tarefas e as soluções de IA utilizam alguma forma de dados, independentemente de suas características, a fim de facilitar a aprendizagem, a adaptação e a melhoria da aprendizagem das máquinas. No entanto, a qualidade dos dados influencia muito o sucesso da segurança e eficácia dessas soluções. "Dados de boa qualidade" é um termo ambíguo e passível de interpretações equivocadas. Portanto, obter uma boa compreensão dos conjuntos de dados usados, a partir da perspectiva de dez "v", por exemplo, é crucial para avaliar a qualidade dos dados em sistemas de IA durante e até mesmo após o desenvolvimento. A Seção "Principais desafios e considerações de dados de qualidade para sistemas de IA de saúde" destaca os principais desafios e considerações para todos os atores interessados, incluindo desenvolvedores e entidades reguladoras, ao lidar com dados em sistemas de IA, a fim de alcançar uma boa qualidade de dados.

Principais desafios e considerações de dados de qualidade para sistemas de IA de saúde

A disponibilidade de conjuntos de dados de boa qualidade clinicamente relevantes é um dos principais desafios enfrentados pelos desenvolvedores. No entanto, dados de qualidade variável ainda podem ser usados dependendo da finalidade, portanto os desenvolvedores devem determinar se os dados disponíveis são de qualidade suficiente para apoiar o desenvolvimento de sistemas que possam atingir o objetivo pretendido. A falta de conjuntos de dados de boa qualidade para uso no desenvolvimento de sistemas de IA pode prejudicar sua eficácia e potenciais benefícios. Dados sem qualidade suficiente para a finalidade pretendida também podem levar a muitos problemas, enviesamentos e erros. Alguns problemas de qualidade de dados que frequentemente surgem durante o desenvolvimento de sistemas de IA e precisam ser considerados por todos os atores interessados são discutidos nesta seção e estão resumidos na Tabela 3. Essas questões e considerações podem estar diretamente relacionadas à gestão de conjuntos de dados, ao modelo de AM, à infraestrutura usada para gerenciar os dados ou a aspectos gerais de governança, da seguinte maneira:

- **Gestão de conjuntos de dados**: ao gerenciar conjuntos de dados para modelos de AM, um plano de gestão de dados claro deve ser pré-especificado e bem documentado. As abordagens de gestão de dados devem basear-se no risco e ser adequadas à finalidade. Isso pode incluir o volume de seleção de dados (incluindo volume de dados usados e volume de dados disponíveis), divisão, limpeza (incluindo quaisquer algoritmos de IA usados para limpar os dados), a usabilidade de dados (incluindo a forma como o conjunto de dados está estruturado em um formato legível por máquina), a rotulagem, as dependências, o aumento e o *streaming*. Se o aumento de dados for relevante, é importante desenvolver uma estratégia clara de aumento de dados. Os responsáveis pelo desenvolvimento devem também considerar a possibilidade de adotar boas práticas de *accountability* para aqueles que manuseiam os dados, a fim de garantir que a qualidade e a integridade dos dados sejam mantidas ao longo de toda

sua linhagem. Isso também é essencial para a gestão e a transferência de conhecimento em um campo altamente evolutivo. Outrossim, além do tratamento dos dados, a capacidade de planejar e realizar análises de dados também é importante.

FIGURA 12 - OS DEZ "V" DOS DADOS

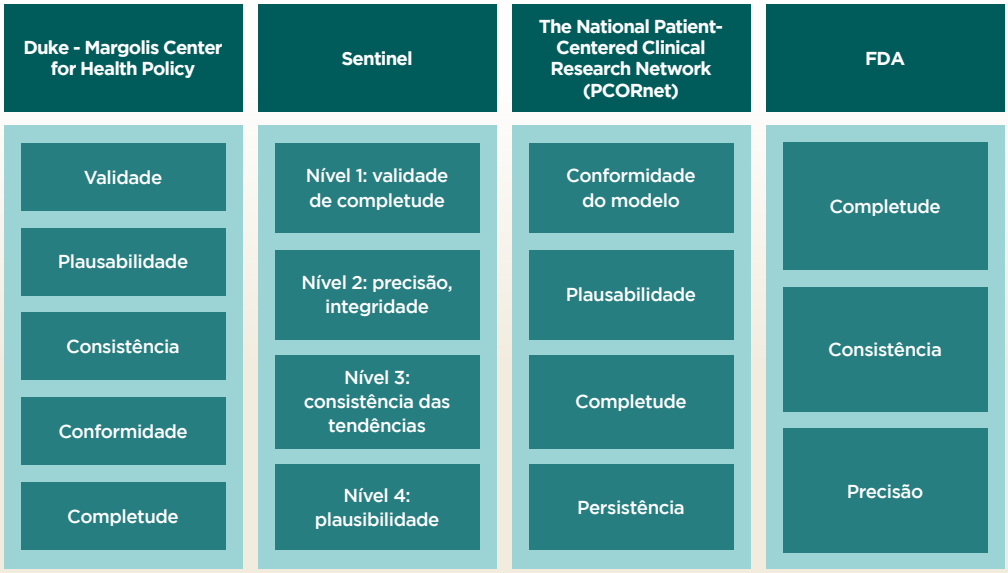
FONTE: PANESAR (2019).

- **Inconsistência de dados**: a alta heterogeneidade na sintaxe dos dados pode exigir harmonização para abordar questões relacionadas a múltiplas fontes de dados, com padrões, formatos, esquemas, estruturas e semântica ambígua variados, e gerar um conjunto de dados coerente para a finalidade de análise abrangente – o que é especialmente desafiador quando se usam dados de saúde. Por exemplo, muitos dados coletados de vários silos de informação são inconsistentes, incompatíveis ou não executáveis em formatos legíveis por máquina. Para diversas fontes de dados, pode haver variações na forma como os dados são capturados (por exemplo, definições de variáveis individuais).

- **Seleção e curadoria de conjuntos de dados**: conhecer a fonte dos dados e fazer uma avaliação inicial da qualidade dos dados pode ajudar a determinar o potencial de enviesamento da seleção e da informação. O viés de seleção resulta quando os dados utilizados para produzir o modelo não são totalmente representativos dos dados reais que o modelo pode receber ou do ambiente em que o modelo funcionará. Além do viés de seleção, o viés de mensuração é outra questão relevante que resulta quando o dispositivo de coleta de dados torna os dados sistematicamente enviesados em uma determinada direção. Consequentemente, os desenvolvedores devem estar cientes das limitações de qualidade de dados ao tentar organizar e utilizar esses conjuntos de dados em grande escala. Além disso, desenvolvedores e entidades reguladoras precisam saber a origem dos dados e como foram coletadas e selecionadas as informações. Isso é especialmente importante quando os conjuntos de dados são de um banco de dados de código aberto em que a fonte original e as especificações do conjunto de dados podem não estar disponíveis. Quando for difícil de estabelecer a origem dos dados, seria prudente que os desenvolvedores avaliassem os riscos do uso desses dados e os gerenciassem de acordo. Finalmente, mesmo que os conjuntos de dados sejam coletados de fontes confiáveis, a mitigação de vieses e a avaliação e mitigação de outros riscos à robustez dos dados continuam essenciais para um conjunto de dados heterogêneo.
- **Usabilidade de dados**: é essencial saber se os dados utilizados para o desenvolvimento do algoritmo se destinavam a esse treinamento, de modo que os desenvolvedores devem transmitir sua total compreensão do conjunto de dados e a razão pela qual ele era adequado para sua finalidade. Por exemplo, os dados provenientes de uma fonte de terceiros podem ser dados representativos destinados para fins de treinamento (como estudos de casos no ensino superior) e não ser adequados para o treinamento de um modelo de IA destinado a diagnosticar uma doença ou uma patologia.

- **Integridade dos dados**: a integridade dos dados pode ser definida como "integridade, consistência e precisão dos dados" (FDA, 2024). A falta de integridade dos dados é uma questão importante. Isso pode ser mais bem compreendido pela forma como a extração e a transformação foram executadas no conjunto de dados. Para manter a integridade dos dados, pode-se desenvolver verificações de dados. As verificações de dados são um componente essencial da garantia de qualidade dos dados quando utilizados no mundo real. Essas verificações também devem ser o primeiro passo na preparação de dados para qualquer fluxo de trabalho de AM.
- **Treinamento de modelos**: os algoritmos de IA são geralmente treinados em um conjunto de dados separado (chamado "conjunto de dados de treinamento") e validados em um conjunto de dados diferente, a fim de medir o desempenho do algoritmo de forma confiável. Os conjuntos de dados de treinamento devem ser bem representados (por exemplo, considerando a prevalência de uma doença/condição) para evitar o "desequilíbrio de classes". Os dados dos prontuários médicos são inerentemente tendenciosos e, portanto, é necessário incorporar dados não médicos, como os determinantes sociais da saúde (Obermeyer *et al.*, 2019). Além disso, a sub-representação de características diagnósticas importantes pode limitar o desempenho do modelo e causar enviesamento. Isso pode ser evitado garantindo que os critérios de inclusão e exclusão no nível do paciente e no nível de introdução de dados não criem um viés de seleção. Além disso, ao garantir que os conjuntos de dados reflitam o cenário em que o modelo será aplicado, a falta de dados diversos (idade, raça, áreas geográficas) pode limitar a generalização e a precisão de um sistema de IA desenvolvido. Isso é demonstrado por um estudo recente da Universidade de Stanford (Shana, 2020), o qual mostrou que 71% dos dados de pacientes de apenas três estados dos EUA treinam a maioria das ferramentas de diagnóstico de IA usadas nos EUA.

FIGURA 13 - EXEMPLOS DE PRINCÍPIOS DE CONTROLE DE QUALIDADE



FONTE: DUKE-MARGOLIS CENTER FOR HEALTH POLICY (2019).

- **Rotulagem de dados**: é importante assegurar uma rotulagem coerente, confiável e precisa dos conjuntos de dados para testes, em conformidade com as boas práticas. Nos casos em que padrões de referência subjetivos são usados, a qualidade será influenciada por muitos fatores – como a independência e as qualificações dos classificadores, o número de avaliadores por rótulo, se o padrão de referência é validado para se correlacionar com os resultados dos pacientes e se o padrão de referência segue métricas publicadas.
- **Documentação e transparência**: o algoritmo e os dados que o suportam muitas vezes não estão disponíveis ou não estão bem documentados para todos os atores interessados do sistema de IA. Isso dificulta a avaliação da qualidade dos dados subjacentes. A transparência e a documentação cuidadosa são importantes não apenas em relação à metodologia usada na coleta de dados, mas também para seleção e modificações dos conjuntos de

dados usados para treinamento, validação e testes. Uma boa documentação é fundamental para alcançar uma transparência que permita a verificação e a rastreabilidade. A transparência dos métodos deve garantir a qualidade dos dados. Além das diretrizes dos relatórios CONSORT-AI e SPIRIT-AI, listas de verificação foram criadas pela comunidade de AM para relatar representatividade, completude e outras características de qualidade de dados[4] (Gebru *et al.*, 2021).

Além disso, os desenvolvedores devem considerar a implantação de testes rigorosos de pré-lançamento para sistemas de IA a fim de garantir que eles não amplifiquem nenhum dos problemas discutidos – como vieses e erros nos dados de treinamento, algoritmos ou outros elementos do *design* do sistema. Ademais, o *design* cuidadoso ou a rápida solução de problemas podem ajudar a identificar precocemente problemas de qualidade de dados. Isso poderia potencialmente prevenir ou mitigar possíveis danos resultantes. Finalmente, para mitigar problemas de qualidade dos dados que surgem nos dados relativos aos cuidados de saúde e aos riscos associados, os atores interessados devem continuar a trabalhar para criar ecossistemas de dados para facilitar o compartilhamento de fontes de dados de boa qualidade.

A lista na Tabela 3 resume as principais considerações de qualidade de dados para a segurança e eficácia do sistema de IA[5].

4 Saiba mais: https://datanutrition.org/
5 Esta lista será atualizada e harmonizada com o trabalho do IMDRF.

TABELA 3 - CONSIDERAÇÕES GERAIS SOBRE A QUALIDADE DOS DADOS

CATEGORIA	ITEM DE AVALIAÇÃO DA QUALIDADE DOS DADOS
Conjunto de dados	• Divisão • Volume e tamanho da seleção • Viés de seleção • Definições de variáveis individuais em cada conjunto de dados • Dados brutos *versus* dados "limpos" • Tratamento e limpeza de dados • Parâmetros e hiperparâmetros • Usabilidade • Caracterização • Rotulagem • Dependências • Aumento • Manipulação • *Streaming* • Interfaces • Integridade • Requisitos exclusivos • Fonte de dados
Infraestrutura de dados	• Tamanho do armazenamento • Formato de armazenamento • Meio de transformação
Modelo de IA/AM	• Treinamento de dados • Ajuste de dados • Conjunto de verificação • Conjunto de validação • Testagem • Conjunto de desenvolvimento • IA estática *versus* IA dinâmica • IA aberta *versus* IA fechada
Gestão da governança	• Responsabilidade • Acesso aos dados • Gestão de riscos • Proteção de dados • Privacidade • Educação para adoção na prática clínica • Boas práticas • Padrões (de cuidado, governança, interoperabilidade, etc.) • Escopo de prática e uso do modelo de IA • Lista de verificação técnica • Documentação • Transparência

FONTE: ELABORAÇÃO PRÓPRIA.

PRIVACIDADE E PROTEÇÃO DE DADOS

A Estratégia Global da OMS para a Saúde Digital 2020-2025 classifica os dados de saúde como dados pessoais sensíveis ou informações de identificação pessoal, os quais requerem um

alto padrão de segurança e proteção. Portanto, a estratégia enfatiza a necessidade de um forte marco legal e regulatório para proteger a privacidade, a confidencialidade, a integridade, a disponibilidade e o tratamento de dados pessoais de saúde. Um quadro jurídico e regulatório responsivo pode também abordar questões de cibersegurança, criação de confiança, *accountability* e governança, ética, equidade, reforço de capacidades e literacia. Isso ajudará a garantir que dados de boa qualidade sejam coletados e, posteriormente, compartilhados para apoiar o planejamento, o comissionamento e a transformação dos serviços.

Para desenvolver e manter estratégias adequadas de segurança dos dados, é importante que os criadores, os implantadores e os fabricantes de sistemas de IA compreendam a teia cada vez mais densa de leis sobre privacidade e proteção de dados. Esta seção discute considerações de alto nível sobre privacidade e proteção de dados. Para outras considerações éticas, consulte o resultado final do Grupo de Trabalho sobre Considerações Éticas sobre IA para a Saúde[6] (OMS, 2021b).

Cenário atual

À medida que a demanda por dados relacionados à saúde aumenta, a proteção da privacidade está criando um desafio único para todos os atores interessados que desejam se beneficiar das muitas oportunidades criadas pelos sistemas e tecnologias de IA. Uma das principais razões para isso é que a alta dimensionalidade do *Big Data* pode dificultar a aplicação de métodos de anonimização e desidentificação. Além disso, garantir que conjuntos de dados em grande escala estejam protegidos contra acesso não autorizado em cada estágio do processo de desenvolvimento – coleta, armazenamento e gestão, transporte, análise, compartilhamento e destruição – é uma consideração importante.

Cerca de 145 países e regiões têm regulamentos de proteção de dados e leis de privacidade que regulam a coleta, o uso, a divulgação e a segurança de informações pessoais

6 Para uma discussão mais ampla sobre privacidade e outras considerações éticas para o uso de IA, consulte o documento final do Grupo de Trabalho sobre Considerações Éticas sobre IA para a Saúde do FG-AI4H e recomendações internacionais, regionais e nacionais.

(Greenleaf, 2021). Existem muitas definições e interpretações diferentes de "proteção de dados" e "privacidade". Em alguns casos, a proteção de dados e a privacidade são usadas de forma intercambiável. Entretanto, embora esses conceitos sejam semelhantes e muitas vezes se sobreponham, seus significados são diferentes, e os desenvolvedores devem estar cientes das implicações legais e éticas que resultam dessas diferenças.

As leis e os regulamentos que abrangem "a gestão de informações pessoais" são normalmente agrupados como "política de privacidade" nos EUA e como "política de proteção", na União Europeia (UE) e em outros locais. Essas leis são muitas vezes complexas e podem ter obrigações conflitantes. Ao desenvolver um sistema de IA para desenvolvimento terapêutico ou aplicações de cuidados de saúde, no início do processo de desenvolvimento, os desenvolvedores devem considerar obter uma compreensão dos regulamentos de proteção de dados aplicáveis e das leis de privacidade, incluindo disposições regulatórias especiais relacionadas a informações confidenciais, como dados genéticos. Os desenvolvedores também devem considerar as leis nacionais e regionais. Por exemplo, nos EUA, embora a Lei de Portabilidade e Responsabilidade de Seguros de Saúde (Health Insurance Portability and Accountability Act [HIPAA]) estabeleça uma linha de base para proteger dados de saúde, os estados têm poderes para promulgar leis de privacidade mais rígidas (como a Lei de Privacidade do Consumidor da Califórnia - California Consumer Privacy Act [CCPA], de 2018).

É importante entender o alcance jurisdicional das diversas leis. Por exemplo, como o âmbito de aplicação do RGPD é vasto e seu impacto é significativo, as empresas podem querer, pelo menos, avaliar em que medida estão sujeitas a ele. A maioria das leis de privacidade, incluindo a Lei de Proteção de Dados Pessoais de Singapura, aplica-se apenas aos dados pessoais tratados no próprio país, enquanto o RGPD[7] pode aplicar-se aos dados pessoais de cidadãos da UE, independentemente do

7 Ver também a proposta de Lei de Proteção de Dados Pessoais da Índia.

local onde os dados são tratados[8]. Consequentemente, as empresas estão sujeitas a obrigações de conformidade de acordo com o RGPD se estiverem localizadas na UE (inclusive se algum componente da organização estiver na UE), oferecerem bens e serviços a indivíduos na UE ou monitorarem o comportamento de pessoas na UE.

Também é importante que os desenvolvedores entendam os vários contextos legais e os requisitos para conceitos relacionados à privacidade, como "identificável", "anônimo" e "consentimento". Por exemplo, o Capítulo 1 do projeto de orientações do Reino Unido da Grã-Bretanha e da Irlanda do Norte sobre anonimização, pseudonimização e tecnologias de reforço da privacidade adverte que a referência a conjuntos de dados como "anonimizados", quando ainda podem conter dados pessoais sob uma forma pseudonimizada, representa o risco de violação da lei de proteção de dados do Reino Unido da Grã-Bretanha e da Irlanda do Norte, a partir de uma convicção equivocada de que o tratamento não envolve dados pessoais (Information Commissioner's Office [ICO], 2021). Os requisitos de consentimento também variam de acordo com a jurisdição. Por exemplo, várias jurisdições podem exigir "consentimento explícito", com requisitos de informação mais elevados para o tratamento de dados relacionados com a saúde (Artigo 9º do RGPD) (Regulamento (EU) 2016/679). Portanto, os desenvolvedores podem querer considerar os diversos contextos jurídicos ao documentar a forma como abordam os conceitos relacionados com a privacidade, incluindo as medidas tomadas para cumprir os requisitos de consentimento e o modo como definem as informações anônimas ou identificáveis.

Além disso, certas jurisdições têm *frameworks* regulatórios de proteção de dados que introduzem regras baseadas na reciprocidade e impõem restrições à circulação ou transferência de dados através das fronteiras. Isto pode ter um impacto significativo na forma como os dados são tratados e compartilhados entre países. Essas disposições servem para reduzir os fluxos transnacionais de dados para dentro e para fora de áreas que

8 Assim como o RGPD, o CCPA aplica-se a pessoas físicas residentes na Califórnia que estejam "domiciliadas no estado ou que estejam fora do estado para fins temporários ou transitórios". Cal. Código Regs. 18, para.17014.

são consideradas como não oferecendo um nível "adequado" de proteção de dados.

Avaliações de adequação podem ser necessárias para determinar se o país destinatário tem normas de proteção de dados e salvaguardas "essencialmente equivalentes" ou "substancialmente similares" às da jurisdição de onde os dados foram transferidos. O RGPD, como um importante motor dos regimes emergentes de proteção de dados em nível mundial, prevê que a livre transferência de dados pessoais para países terceiros, Estados não-membros da UE, pode ocorrer principalmente quando o país terceiro é considerado, pela Comissão da UE, como tendo um nível de proteção "adequado"[9]. Até maio de 2023, a Comissão da UE havia reconhecido apenas 13 países como fornecendo proteção adequada (CE, s.d.).

Os desenvolvedores devem estar cientes das nuances dos regulamentos e das leis das diferentes jurisdições e considerar documentar suas práticas de proteção de dados de acordo com tais regulamentos e leis. Em geral, as empresas devem levar em consideração manter-se a par das novas leis e exigências, alavancando governança, análise de risco, políticas, treinamento e outras estratégias, de forma abrangente e coerente.

Documentação e transparência

A documentação e a transparência são fundamentais para facilitar a confiança no que diz respeito à privacidade e à proteção de dados. As divulgações detalhadas da política de privacidade fornecem às entidades reguladoras uma referência para examinar o tratamento de dados de uma empresa. Essas divulgações devem identificar os usos significativos de informações pessoais para decisões algorítmicas. Dependendo da jurisdição, essas divulgações podem exigir a inclusão de outras informações relevantes – por exemplo, os tipos e fontes de dados de saúde coletados e processados; as identidades das

9 Os fluxos de dados têm se tornado cada vez mais uma parte importante da interconexão global e do desenvolvimento de IA. Embora o caso Schrems II diga respeito à posição UE-EUA sobre transferências de dados, as implicações mais amplas informam as transferências globais de dados e a forma como devem ser compatíveis com os requisitos do RGPD, incluindo a validade das cláusulas contratuais padrão que dependem da existência de mecanismos eficazes para garantir a conformidade com o nível de proteção exigido pelo RGPD. *Data Protection Commissioner contra Facebook Ireland Limited, Maximillian Schrems* (Processo C-311/18, "Schrems II").

pessoas ou organizações que determinaram a finalidade ou os meios de tratamento de dados pessoais; a identidade da pessoa ou organização que processou os dados; as bases legais para o tratamento dos dados; a forma como os dados foram recolhidos (incluindo se foi fornecida uma notificação adequada ao titular dos dados e como os requisitos de consentimento foram cumpridos); e as informações técnicas e organizacionais sobre o armazenamento de dados, incluindo medidas de segurança.

Os desenvolvedores devem levar em conta a privacidade ao projetar e implantar sistemas de IA. Isso inclui projetar, implementar e documentar abordagens e métodos para garantir a qualidade contínua em todas as fases de desenvolvimento a fim de proteger a privacidade dos dados (Regulamento (EU) 2016/679)[10]. As proteções da privacidade não devem limitar-se apenas a fazer face aos riscos de cibersegurança, especialmente porque alguns riscos para a privacidade (por exemplo, danos à dignidade que podem causar constrangimento ou estigma, ou danos mais tangíveis, como discriminação, perdas econômicas ou danos físicos) (Instituto Nacional de Padrões e Tecnologia - National Institute of Standards and Technology [NIST], 2020) também podem surgir por meios não relacionados com incidentes de cibersegurança. Portanto, ao desenvolver soluções para lidar com os riscos, os desenvolvedores devem ter uma compreensão geral das diferentes origens dos riscos de cibersegurança e privacidade e desenvolver suas práticas de gestão de riscos de forma adequada (Figura 14).

Um programa de conformidade deve considerar os riscos e desenvolver prioridades de conformidade com a privacidade que levem em conta quaisquer danos potenciais específicos, bem como o ambiente de aplicação. Os desenvolvedores podem considerar incluir em sua documentação uma descrição das operações envolvidas no tratamento de dados pessoais, uma

10 Por exemplo, um dos pilares de qualidade contínua dos dados em algumas jurisdições, como a legislação da UE, é o princípio da *accountability*. De acordo com o Artigo 5 do RGPD, os controladores de dados devem obedecer aos cinco conjuntos de princípios consagrados no Artigo 5(1), por exemplo, minimização de dados. Os controladores de dados devem determinar medidas técnicas e organizacionais para atingir esses fins (Artigo 5, alínea 2), ao longo de todo o ciclo de tratamento de dados. Embora não seja mencionado, o princípio da *accountability* também está presente nos Artigo 24(1), Artigo 25(1) e Artigo 32 do regulamento com relação à responsabilidade do controlador, ao princípio da proteção de dados desde o design (e por padrão) e às medidas de segurança.

avaliação de risco e as medidas implementadas para mitigar riscos que levem em conta os interesses dos titulares dos dados.

FIGURA 14 - *FRAMEWORK* DE PRIVACIDADE DO NIST – RELAÇÃO ENTRE RISCO DE CIBERSEGURANÇA E PRIVACIDADE

FONTE: NIST (2020).

Certas regulamentações delineiam requisitos prescritivos de segurança para abordar riscos de cibersegurança e privacidade, como o princípio de proteção de dados por *design* e por padrão do RGPD (Artigos 25 e 32) (Regulamento (EU) 2016/679) e a política proposta de privacidade de dados por *design* da Índia (Lei n. 22/2023). Outras regulamentações incluem o dever de implementar e manter práticas de segurança razoáveis, apropriadas ao risco[11]. Os quadros de privacidade frequentemente incluem avaliações de impacto à privacidade, que são uma ferramenta amplamente utilizada de gestão de privacidade para avaliar proativamente e mitigar os riscos à privacidade. Algumas jurisdições, incluindo a UE (Artigo 35 da RGPD) (Regulamento (EU) 2016/679)[12], exigem

11 Por exemplo: CCPA § 1798.150(a)(1), Lei de Proteção de Informações Pessoais da África do Sul de 2013; Regulamentos Israelenses de Proteção à Privacidade (Segurança de Dados), 5777–2017 (implementando a Lei de Proteção da Privacidade, 5741–1981 de 1981); Lei Federal n. 2, de 2019; Lei de Comércio Eletrônico do Reino da Arábia Saudita de 2019 e suas Regras de Implementação.

12 "Uma avaliação do impacto da proteção de dados deve ser realizada se o tratamento puder resultar em alto risco para os direitos e liberdades das pessoas naturais".

que as empresas realizem essas avaliações[13]. Embora a legislação dos EUA não exija avaliações de impacto à privacidade, a estrutura de privacidade do NIST, do Departamento de Comércio dos EUA, recomenda que os desenvolvedores as realizem. De acordo com o NIST, "identificar se o tratamento de dados pode criar problemas para os indivíduos, mesmo quando uma organização pode estar totalmente em conformidade com as leis ou regulamentos aplicáveis, pode ajudar na tomada de decisões éticas no projeto ou na implantação de sistemas, produtos e serviços" (NIST, 2020). Isso, por sua vez, pode aumentar a confiança no sistema.

Os desenvolvedores também podem considerar a anotação de seus sistemas de IA e a criação de trilhas de auditoria que explicam quais tipos de decisões são tomadas durante o processo de desenvolvimento. As anotações fornecem transparência "após o fato" a terceiros e podem ajudar a explicar como a privacidade foi incorporada, se aplicável (West & Allen, 2020). Essas explicações e documentações devem estar disponíveis em diferentes níveis de detalhe e direcionadas a diferentes públicos-alvo – reguladores, gestores, desenvolvedores, operadores e usuários. A natureza das informações e das explicações necessárias pode variar, mas todas as suposições, as restrições, as fontes de dados, entrada e saída esperadas, e os principais riscos e limitações em cada nível devem ser claramente documentados. Além disso, uma trilha de auditoria mostra não apenas que os controles foram aplicados, mas também pode mostrar como os danos foram mitigados, no caso de uma violação de dados.

Muitas jurisdições publicam orientações ou aplicam determinados requisitos de cibersegurança para serem considerados pelos desenvolvedores de dispositivos médicos. Embora uma discussão aprofundada sobre os requisitos de cibersegurança esteja fora do escopo desta subseção, é importante entender o papel fundamental que a cibersegurança desempenha na proteção de informações pessoais de saúde. A cibersegurança concentra-se em implementações técnicas específicas

13 Embora as avaliações de risco sejam bastante comuns em padrões e requisitos de segurança da informação, elas raramente são vistas nas regras de privacidade nos EUA. O RGPD, no entanto, exige que uma organização que trata dados pessoais realize uma Avaliação de Impacto na Privacidade de Dados (*Data Protection Impact Assessment* [DPIA]) específica antes de iniciar o tratamento.

necessárias para proteger sistemas e redes contra os ataques cibernéticos, que podem comprometer tanto a segurança de sistemas e dados relacionados à saúde quanto a privacidade de um indivíduo, o que pode resultar em danos. Para fornecer transparência sobre as práticas de cibersegurança, os desenvolvedores podem considerar a documentação de práticas e abordagens para a segurança de dados, incluindo políticas públicas que ajudem a proteger a confidencialidade, a integridade e a disponibilidade de dados pessoais ao longo de seu ciclo de vida – como criptografia apropriada, controles de acesso, métodos de registro, monitoramento de risco e métodos de destruição segura. Os desenvolvedores também podem documentar os sistemas e as abordagens utilizadas como proteção contra a manipulação de dados e os ataques adversários (NIST, 2018). Por exemplo, tecnologias baseadas em *blockchain* podem ser um mecanismo para proteger a privacidade, a segurança e a integridade dos dados para a IA em um ecossistema tradicionalmente fragmentado de sistemas de informação em saúde, em contextos nacionais e regionais (Alsalamah *et al.*, 2021).

Sandboxes *regulatórios de IA*

Os desafios regulatórios acima são reconhecidos por autoridades reguladoras e formuladores de políticas em todo o mundo (Attrey *et al.*, 2020). Como resultado, mais de 50 países atualmente experimentam *sandboxes* em uma ampla gama de setores de alta tecnologia – notadamente no setor financeiro, mas *sandboxes* também ganharam popularidade para serviços de saúde e jurídicos (Matiega & van de Pol, 2022). A abordagem de *sandbox* regulatório obteve força considerável como um meio de ajudar os reguladores a abordarem o desenvolvimento e o uso de IA e outras tecnologias emergentes (Matiega & van de Pol, 2022). Os *sandboxes* regulatórios são geralmente ferramentas regulatórias que permitem a flexibilidade de testar produtos ou serviços inovadores com requisitos regulatórios mínimos (Matiega & van de Pol, 2022). Consequentemente, os *sandboxes* regulatórios são considerados uma abordagem ágil para inovação e regulação, por isso as autoridades reguladoras os têm cada vez mais favorecido. Na UE, foram propostos *sandboxes* regulatórios para testar soluções de vigilância na luta contra COVID-19 e estabelecer um *framework* para o acesso a

dados em toda a UE. Em relação especificamente aos regulamentos de IA, o primeiro *sandbox* regulatório-piloto de IA será presumivelmente lançado em 2023 pelo Governo da Espanha[14] com o objetivo de fornecer um guia a todos os Estados-Membros da UE e à CE (CE, 2022). Embora os *sandboxes* regulatórios de IA tenham levantado algumas preocupações, eles têm o potencial de trazer muitos benefícios importantes para entidades reguladoras de sistemas de IA, desenvolvedores, fabricantes e até mesmo pacientes (Matiega & van de Pol, 2022). Isso ocorre porque podem: (a) ajudar a permitir uma melhor compreensão dos sistemas de IA durante a fase de desenvolvimento e antes de serem colocados no mercado; (b) facilitar o desenvolvimento de políticas de aplicação adequadas e orientação técnica a fim de mitigar riscos e consequências não intencionais; e (c) promover a inovação em IA, estabelecendo um ambiente controlado de experimentação e teste para tecnologias, produtos e serviços inovadores de IA para sistemas de IA novos e potencialmente mais seguros.

ENGAJAMENTO E COLABORAÇÃO

Quando aplicável e apropriado, o envolvimento e a colaboração entre desenvolvedores, fabricantes, profissionais de saúde, pacientes, defensores dos pacientes, formuladores de políticas, órgãos reguladores e outros atores interessados podem melhorar a segurança e a qualidade de um sistema de IA. Muitos órgãos reguladores adotaram abordagens colaborativas e de engajamento nessa área, e esta seção discute as abordagens de cinco deles: a MHRA do Reino Unido da Grã-Bretanha e da Irlanda do Norte, a SAHPRA da África do Sul, a CE, a HSA de Singapura e a FDA dos EUA. A Tabela 4 lista exemplos de "com quem", "por que" e "como" essas entidades reguladoras promovem o engajamento e a colaboração. Os exemplos não pretendem ser abrangentes, mas sim destacar abordagens gerais. A Tabela 4 é seguida de uma análise que discute as

14 Em 9 de novembro de 2023, o Ministério de Assuntos Econômicos e Transformação Digital (Ministerio De Asuntos Económicos y Transformación Digital), da Espanha, publicou o "Real Decreto 817/2023, de 8 de novembro, que estabelece um ambiente controlado de testes para a avaliação da conformidade da proposta de Regulamento do Parlamento Europeu e do Conselho que estabelece normas harmonizadas em matéria de Inteligência Artificial". Ver Real Decreto 817/2023.

semelhanças e diferenças nas abordagens.

A Subseção "Dois casos bem-sucedidos de engajamento" examina dois exemplos de engajamento e comunicação entre reguladores e desenvolvedores de IA, culminando em resultados clínicos positivos (CURATE.AI e IDentif.AI). As últimas subseções consideram as implicações práticas para o engajamento e a colaboração em ambientes com recursos limitados e recomendam maneiras pelas quais os órgãos reguladores podem iniciar esse processo, mesmo em países sem experiência anterior em engajamento e colaboração. Isso é complementado por várias narrativas: como aplicar ferramentas de engajamento (com base na experiência), posicionar o regulador como um parceiro no contexto de um diálogo acessível, e orientações e recomendações durante o processo de desenvolvimento.

TABELA 4 – EXEMPLOS DE ABORDAGENS DE ENTIDADES REGULADORAS PARA O ENGAJAMENTO E A COLABORAÇÃO COM ATORES INTERESSADOS SOBRE O USO DE IA EM CUIDADOS DE SAÚDE E DESENVOLVIMENTO TERAPÊUTICO

1. MHRA, Reino Unido da Grã-Bretanha e Irlanda do Norte	
Com quem?	**Exemplos de partes interessadas com as quais a MHRA se envolve e colabora:** • **Pacientes/defensores dos direitos dos pacientes** • **Academia** • **Profissionais de saúde** por exemplo, prestadores do Serviço Nacional de Saúde (SNS) e prestadores privados de cuidados de saúde • **Indústria** por exemplo, indústria de dispositivos médicos e diagnóstico *in vitro*, indústria de tecnologia de saúde • **Parceiros do governo nacional** por exemplo, Departamento de Saúde e Assistência Social (DHSC), NHS England and Improvement, NICE e Comissão da Qualidade do Cuidado (Care Quality Comission [CQC])
Por quê?	**Exemplos de razões pelas quais a MHRA se envolve e colabora com as partes interessadas:** • Alertar os usuários sobre problemas com dispositivos médicos e medicamentos • Responder perguntas sobre funções na regulação ou aumentar a conscientização sobre questões de segurança • Buscar *feedback* sobre o desenvolvimento da política regulatória, gerenciando incidentes e riscos adversos • Interface com o governo do Reino Unido da Grã-Bretanha e Irlanda do Norte e o NHS, incluindo partes interessadas alinhadas a atividades digitais e relacionadas à IA

Como?	**Exemplos de maneiras como a MHRA se envolve e colabora com as partes interessadas:** • Sistema central de alerta para o SNS e prestadores de cuidados de saúde ou através de grupos profissionais • Consultas da mídia, do público e de outras partes interessadas por meio do centro de atendimento ao cliente da MHRA, e-mails dedicados e assessoria de imprensa • Conexão com grupos consultivos de especialistas, redes e grupos de partes interessadas em questões específicas • Consulta sobre o engajamento com pacientes e com o público • Reuniões de trabalho com as partes interessadas nacionais, reuniões bilaterais com outras partes do SNS, governo e homólogos internacionais

2. SAHPRA, África do Sul	
Com quem?	**Exemplos de partes interessadas com as quais a SAHPRA se envolve e colabora:** • **Pacientes/defensores dos direitos dos pacientes** • **Academia** • **Profissionais de saúde** • **Indústria** (por exemplo, fabricantes/distribuidores, associações comerciais). • **Parceiros do governo nacional** (por exemplo, Departamento Nacional de Saúde, Departamento Nacional de Comércio e Indústria, Serviço Nacional de Acreditação da África do Sul).
Por quê?	**Exemplos de razões pelas quais a SAHPRA se envolve e colabora com as partes interessadas:** • Facilitar a aprovação de sistemas inovadores de IA. • Sistema Nacional de Acreditação da África do Sul (SANAS) para garantir que a rede de Organismos de Avaliação da Conformidade seja estabelecida no país para certificar o sistema de gestão da qualidade (SGQ).
Como?	**Exemplos de maneiras como a SAHPRA se envolve e colabora com as partes interessadas:** • A estrutura para engajamento e colaboração ainda não foi formalizada • Recomendou que o engajamento de stakeholders adotasse o modelo de engajamento em cinco etapas desenvolvido pela TGA

3. CE, União Europeia	
Com quem?	**Exemplos de partes interessadas com as quais a CE se envolve e colabora:** • Pacientes/defensores dos direitos dos pacientes • Academia • Profissionais de saúde

Por quê?	**Exemplos de razões pelas quais a CE se envolve e colabora com as partes interessadas:** • "Apoiar a Comissão no desenvolvimento de ações para a transformação digital da saúde e dos cuidados de saúde na UE"
Como?	**Exemplos de maneiras como a CE se envolve e colabora com as partes interessadas:** • Fornecendo aconselhamento e conhecimentos especializados à Comissão, em particular sobre os temas definidos na comunicação[15] sobre permitir a transformação digital da saúde e dos cuidados de saúde no Mercado Único Digital, adotada em abril de 2018. Em particular, esses tópicos dizem respeito à interoperabilidade de dados de saúde e formatos de troca de registros, serviços de saúde digital, proteção de dados e privacidade, IA e "outros elementos transversais ligados à transformação digital da saúde e dos cuidados, como propostas de financiamento e investimento e tecnologias facilitadoras"

4. HSA, Singapura	
Com quem?	**Exemplos de partes interessadas com as quais a HSA se envolve e colabora:** • **Academia** (por exemplo, instituições de pesquisa) • **Profissionais de saúde** • **Indústria** (por exemplo, desenvolvedores de *software* e de IA, associações comerciais) • **Órgãos governamentais nacionais**
Por quê?	**Exemplos de razões pelas quais a HSA se envolve e colabora com as partes interessadas**: • Envolvimento precoce e suporte aos inovadores para facilitar a conformidade regulatória, facilitando, assim, o acesso oportuno a inovações seguras para os pacientes • Consultar de maneira ativa sobre políticas e diretrizes relacionadas à IA e aos dispositivos médicos com *software* para receber e incorporar as contribuições e perspectivas das partes interessadas (Diretrizes regulatórias para dispositivos médicos de *software* — uma abordagem de ciclo de vida) (HSA, 2022) • Trabalhar com outros órgãos responsáveis pela implementação e implantação de dispositivos médicos de *software* e IA no sistema de saúde para facilitar a adoção de tecnologias inovadoras no sistema de saúde

15 Saiba mais: https://digital-strategy.ec.europa.eu/pt/node/3067

Como?	**Exemplos de maneiras como a HSA se envolve e colabora com as partes interessadas:** • Disponibilizando portais de engajamento rápidos e simplificados para várias facetas da regulação dos produtos[16] • Os processos específicos que podem ser abordados de forma simples incluem o Sistema de Comunicação de Informação de Dispositivos Médicos (para apresentação de pedidos de licenças, autorizações, registros, etc.) • Ferramentas de autoajuda online para determinar a classificação de produtos e classificação de risco para dispositivos médicos e formulários simples para solicitar aconselhamento e confirmação do HSA • Consulta de Desenvolvimento de Dispositivos Médicos: sistema de agendamento de consultas online que permite que inovadores e desenvolvedores busquem aconselhamento científico e regulatório durante a fase de desenvolvimento de dispositivos médicos para facilitar a conformidade regulatória • Processo de consulta online das partes interessadas para todas as políticas e diretrizes novas e revisadas • Discussões e compromissos regulares em grupos focais com associações e empresas do setor

5. FDA, EUA	
Com quem?	**Exemplos de partes interessadas com as quais a FDA se envolve e colabora:** • **Pacientes/cuidadores/defensores dos direitos de pacientes** • **Academia** (por exemplo, instituições de pesquisa) • **Profissionais de saúde** • **Indústria** (por exemplo, desenvolvedores, fabricantes de dispositivos, empresas farmacêuticas, associações comerciais) • **Parceiros governamentais nacionais** (por exemplo, Institutos Nacionais de Saúde [NIH], Escritório do Coordenador Nacional de Tecnologia da Informação em Saúde [ONC], Comissão Federal de Comunicações [FCC]) • **Parceiros de governos estrangeiros** • **Organizações internacionais** (por exemplo, IMDRF, ICH) • **Consumidores/público em geral**
Por quê?	**Exemplos de razões pelas quais a FDA se envolve e colabora com as partes interessadas:** • Facilitar o acesso do paciente a tecnologias que possam beneficiá-lo em tempo hábil • Apoiar o desenvolvimento de produtos médicos novos e inovadores por meio de interações iniciais com as partes interessadas • Fornecer *feedback* oportuno sobre as políticas da FDA para reduzir a incerteza • Informar o público sobre dispositivos de IA/AM • Receber *feedback* sobre políticas, orientações e documentos de discussão

16 Saiba mais: https://www.hsa.gov.sg/e-services

Como?	**Exemplos de maneiras como a FDA se envolve e colabora com as partes interessadas:** • Promovendo diferentes tipos de reuniões de pré-submissão para fornecer *feedback* antecipado aos patrocinadores. • Participando e liderando esforços de harmonização internacional (por exemplo, IMDRF, ICH). • Engajando-se como membro de parcerias público-privadas e de comunidades colaborativas. • Colaborando em um ambiente pré-competitivo na pesquisa científica regulatória para avançar o entendimento da comunidade científica. • Recebendo comentários formais sobre políticas e orientações por meio do Registro Federal. • Realizando workshops e outros eventos de engajamento para obter feedback de pacientes, indústria e outras partes interessadas.

Discussão sobre estratégias de órgãos reguladores selecionados

A Tabela 4 mostra as abordagens de quatro organismos reguladores nacionais e um regional (no caso da CE) para promover o engajamento e a colaboração. Na primeira categoria ("com quem?"), há semelhanças consideráveis entre esses corpos. As metas compartilhadas para engajamento e colaboração incluem profissionais de saúde (indicados pela FDA, SAHPRA, MHRA, EC e HSA), academia (FDA, SAHPRA, MHRA, EC e HSA), indústria (FDA, SAHPRA, MHRA, EC e HSA), pacientes ou defensores dos direitos de pacientes (FDA, SAHPRA, MHRA e EC), órgãos governamentais nacionais (FDA, SAHPRA e MHRA), mídia (imprensa nacional e comercial; FDA e MHRA), provedores de saúde (FDA e MHRA) e consumidores (FDA e MHRA). Curiosamente, o documento de estratégia do NIST, do Departamento de Comércio dos EUA, também se refere à academia e aos órgãos governamentais domésticos como alvos de engajamento e colaboração.

Na segunda categoria ("por quê?"), a SAHPRA observa a importância de comunicar os benefícios e o uso pretendido dos dispositivos, presumivelmente para proteger e promover a saúde pública (listados pela FDA e sugeridos implicitamente pela MHRA). A FDA também enfatiza a importância da comunicação bilateral com os atores interessados para que as entidades reguladoras estejam cientes dos desenvolvimentos na indústria (ou academia) e para que esses atores interessados, por sua vez, estejam cientes dos desenvolvimentos na

regulação. Da mesma forma, a MHRA indica a importância de se obter dos atores interessados *feedback* sobre dispositivos médicos. Isto apoia os objetivos assumidos tanto pela SAHPRA como pela CE, a fim de facilitar a aprovação de soluções inovadoras e apoiar a transformação digital da saúde e dos cuidados. A HSA reconhece a importância do engajamento precoce com inovadores e desenvolvedores para fornecer mais clareza nos requisitos regulatórios e melhorar a transparência nos processos regulatórios.

Para a terceira categoria ("como?"), a FDA lista as medidas que são tomadas para promover o engajamento (por exemplo, a realização de *workshops*, a produção de material digital e impresso e a oferta de módulos de treinamento ou outros tipos de educação). A MHRA também observa a importância de realizar reuniões com os atores interessados (incluindo institutos governamentais nacionais e homólogos internacionais). A HSA introduziu um esquema de consulta pré-comercialização para apoiar a inovação e o desenvolvimento de dispositivos, fornecendo aconselhamento científico e regulatório para permitir a conformidade regulatória por desenvolvedores de *software* e IA, os quais, ao contrário dos fabricantes tradicionais de dispositivos médicos, não estão familiarizados com os requisitos regulatórios[17] (Department of Health and Age Care, 2017).

Dois casos bem-sucedidos de engajamento

Para entender o valor do engajamento e da colaboração entre órgãos reguladores e atores interessados, dois exemplos do mundo real (Caso 1 e Caso 2) são descritos. Caminhos coerentes para o engajamento entre órgãos reguladores e desenvolvedores de IA desempenham um papel importante para garantir que a avaliação rigorosa e a entrega acelerada de modalidades impactantes possam ser realizadas sem problemas. Um aspecto diz respeito à área de IA intervencionista/medicina digital, que envolve a aplicação de *softwares*/dispositivos (por exemplo, desenvolvimento de medicamentos baseados em IA e/ou plataformas de dosagem) e/ou a aplicação de compostos de medicamentos resultantes e/ou combinações

17 Saiba mais: https://www.iap2.org/

recomendadas por essas plataformas (Ho, 2020a; Ho, 2020b; Blasiak *et al.*, 2020). Nesse contexto, a integração da acessibilidade das entidades reguladoras com a inovação emergente, por vezes em circunstâncias urgentes, terá resultados que salvam vidas. É importante ressaltar que esses resultados não se limitarão ao tratamento pós-aprovação, mas também ao benefício substancial do paciente durante os estágios experimentais de validação.

No Caso 1, o roteiro de desenvolvimento e validação do CURATE.AI e a tecnologia fundamental da IDentif.AI foram discutidos com a Divisão de Dispositivos Médicos (HSA, 2022) da HSA em Singapura. Esta sessão interativa incluiu uma revisão aprofundada das principais descobertas das plataformas tecnológicas, o processo de implementação de ambas as plataformas, as estratégias de análise estatística emergentes para avaliar eficazmente os resultados do tratamento de medicina personalizada e as rotas regulatórias. Também foi realizada uma discussão mais ampla sobre como esses *design*s de ensaios clínicos podem evoluir devido ao surgimento da IA (Ho *et al.*, 2020; Shah *et al.*, 2019; Harrer *et al.*, 2019). Foi estabelecido um percurso claro para os pedidos de informação subsequentes, uma vez que se esperavam pedidos de orientação múltiplos e frequentes devido à natureza dos projetos de ensaios clínicos previstos. Estes incluíram *design*s de estudo N-de-1 para uma ampla gama de indicações concebidas para cada paciente. Especificamente, esses projetos foram personalizados com base em, por exemplo, calibrações de dosagem individualizadas do regime de medicamentos (regime selecionado pelo médico), medições de eficácia e toxicidade em série, protocolos de tratamento orientados pela eficácia e parâmetros de segurança. Submissões subsequentes incluíram engajamento com reguladores para classificações de risco associadas ao dispositivo para cada ensaio e discussão subsequente para submissão de Rotas de Acesso Especial (Special Access Routes [SAR]) (HSA, 2019) para a potencial implementação rápida de ensaios e fins de tratamento, se necessário. Respostas rápidas e informativas, e engajamento ativo dos membros da equipe regulatória da HSA resultaram em tempos de resposta eficientes para o início do estudo, o que ofereceu um resultado positivo para um paciente oncológico refratário. Um histórico sustentado de engajamento

com a comunidade regulatória tem desempenhado um papel fundamental no auxílio ao desenvolvimento de um fluxo de processos claro para solicitações de orientação *downstream*.

O Caso 2 foi desenvolvido em resposta à pandemia COVID-19. Especificamente, uma cepa de vírus vivo derivada de paciente foi utilizada para a otimização de terapia combinada, impulsionada pelo IDentif.AI, para servir como um sistema de apoio à decisão clínica. Ao contrário das abordagens tradicionais baseadas em IA, essa estratégia não usou conjuntos de dados de pacientes existentes. Em vez disso, a experimentação prospectiva foi usada juntamente com uma estratégia de análise de dados pequenos derivada de IA para identificar classificações prospectivas, respaldadas por dados, de combinações para consideração clínica futura e, potencialmente, eliminação de certas combinações de consideração clínica adicional. A tecnologia fundamental da IDentif.AI foi discutida anteriormente em detalhes com a Divisão de Dispositivos Médicos da HSA, e informações adicionais do estudo IDentif.AI SARS-CoV-2 foram fornecidas no contexto do suporte à decisão clínica, desenvolvendo combinações otimizadas identificadas por IDentif.AI e com ensaios potenciais sendo projetados com parceiros clínicos. Com relação ao engajamento do órgão regulador, a Divisão de Dispositivos Médicos do HSA foi contatada para fornecer orientação de classificação de risco de dispositivo para o envio de uma Notificação de Materiais de Pesquisa Clínica (Clinical Research Material Notification [CRM-N]) para fins de estudo. A obtenção de um CRM-N é parte obrigatória da apresentação de um programa de validação clínica, pois estipula o pré-requisito de uma avaliação inicial do risco do dispositivo a partir do HSA (HSA, 2023). O portal de submissão e a interação do portal foram particularmente simples de navegar e foram integrados com um portal de acesso uniforme, simplificado para supervisão e monitoramento eficientes com os órgãos reguladores. Isso demonstra ainda mais o processo simples de interação com o HSA. Esse caso foi um exemplo da importância crítica da acessibilidade direta do órgão regulador e de seu impacto profundamente positivo no avanço de tecnologias promissoras para uma maior avaliação clínica e validação.

Abordagens recomendadas para países sem experiência prévia

Para países com experiência limitada em engajamento e colaboração (e/ou recursos limitados), é importante estabelecer: (a) quais níveis de engajamento e colaboração são desejados; (b) quais medidas podem e devem ser tomadas para atingir esses níveis; e (c) quais desafios são apresentados pela tecnologia (por exemplo, explicabilidade da IA).

Em muitos casos, é desejável adotar modelos regulatórios adaptáveis, flexíveis, modulares e escaláveis, a fim de dar conta das incertezas da inovação por meio de supervisão e coordenação adequadas. Essas características enquadram-se não apenas nos desafios específicos das tecnologias emergentes, mas também na abordagem regulatória de países sem experiência passada nesse campo ou com recursos econômicos escassos. Por um lado, as prioridades devem ser escaláveis para que quantidades crescentes de trabalho possam ser adequadamente abordadas, adicionando recursos ao modelo regulatório. Por outro lado, no entanto, as prioridades devem ser determinadas de acordo com a adaptabilidade modular das etapas e níveis de engajamento. Em ecologia, a adaptabilidade aplica-se à capacidade de lidar com perturbações inesperadas no ambiente. Em engenharia, modularidade refere-se à inter-relação das partes separadas de um pacote de *software* ou ao particionamento do projeto para torná-lo gerenciável. Em Sistemas Multiagentes (SMA), refere-se ao uso eficiente de recursos computacionais. Podemos tirar proveito dessa noção para criar políticas adaptáveis que possam ser combinadas em sistemas regulatórios de governança jurídica. O objetivo deve ser abordar as incertezas da inovação e se alinhar com as preferências da sociedade em relação à inovação emergente, permitindo, simultaneamente, que os reguladores obtenham uma compreensão crescente dos desafios tecnológicos com crescente granularidade normativa (Pagallo *et al.*, 2019).

Narrativa sobre a utilização de ferramentas de engajamento com base na experiência prática

Para todos os países – desde aqueles com experiência limitada em engajamento e colaboração (e/ou recursos limitados)

até aqueles no outro extremo do espectro – as ferramentas de gestão de projetos e programas podem ajudar as organizações (incluindo reguladores) a estruturar e executar seu engajamento com atores interessados e usuários. Não importa qual ferramenta seja escolhida, a chave para um engajamento valioso é investir tempo, energia e pensar na melhor forma de envolver os atores interessados e, depois, dar seguimento a esse engajamento durante a duração de um projeto ou programa. O engajamento geralmente falha se o investimento for visto como um relacionamento de curto prazo, em vez de longo prazo.

O modelo de cinco etapas recomendado pelo governo australiano para engajamento (Department of Health and Age Care, 2017) é um bom ponto de partida para considerar como um órgão regulador poderia se envolver com desenvolvedores de produtos e serviços de saúde de IA. Nesse modelo, o engajamento começa com a reflexão sobre o propósito do engajamento (com base no que se espera alcançar) e identificação dos atores interessados relevantes. Ao planejar os diferentes níveis de engajamento com os atores interessados, recomenda-se mapear os relacionamentos existentes e definir o tipo de engajamento e relacionamento necessário com a parte interessada (e que tipo de relacionamento essa parte estaria disposta a ter). Por exemplo, um desenvolvedor de saúde digital que constrói um aplicativo (App) para apoiar pais com filhos acima de um peso considerado saudável pode descobrir que o órgão de saúde primário em questão é uma parte interessada influente que define políticas sobre a gestão do peso das crianças. No entanto, como este não é um órgão com o qual o desenvolvedor do aplicativo precisa se envolver regularmente, o desenvolvedor pode apenas "informar" o órgão de saúde do projeto. No entanto, um desenvolvedor pode querer trabalhar com pais de crianças acima de um peso saudável para co-projetar o aplicativo e garantir que ele atenda a suas necessidades. Seria, portanto, importante que o desenvolvedor "colaborasse" com um grupo representativo de pais e estabelecesse uma comunicação bidirecional ou multidirecional e o aprendizado compartilhado e a tomada de decisões ao longo do projeto.

Uma abordagem semelhante para garantir que os atores interessados recebam as informações certas no momento certo e estejam usando canais de comunicação ideais é delineada por

uma das principais empresas de *software* de desenvolvimento de produtos (Atlassian, s.d.). No "jogo" de comunicação com os atores interessados, dá-se importância a quem são esses atores, o método de comunicação desejado e a frequência da comunicação. Por exemplo, um projeto interno do governo que desenvolva um produto de saúde digital terá atores interessados internos (como financiadores do projeto e líderes de políticas) e atores interessados externos (como acadêmicos importantes). O plano de comunicação deve descrever como cada grupo de atores interessados será abordado (*e-mail*, conversa cara a cara, videochamada e/ou mídias sociais) e com que frequência haverá contato com o grupo de interessados (diário, quinzenal e/ou anual), com base no que o relacionamento com os atores interessados traz para os objetivos gerais (ou seja, compartilhamento de informações, *codesign* e/ou garantia de qualidade). Esse plano poderá, então, ser mapeado em uma tabela simples (cujos exemplos de títulos podem ser: o método, o público/parte interessada, o conteúdo a ser compartilhado, o porquê e a frequência) para ser seguido por toda a equipe de desenvolvimento.

Posicionamento narrativo do órgão regulador como parceiro no processo de desenvolvimento

Como demonstrado na Tabela 4 e discutido no texto subsequente, vários órgãos reguladores enfatizam a importância da comunicação aberta (bilateral) com os atores interessados para que os reguladores estejam cientes dos desenvolvimentos na tecnologia baseada em IA e esses atores interessados, por sua vez, estejam cientes das mudanças na regulação. Isso ocorre porque a tecnologia baseada em IA está em constante mudança, e a regulação precisa ser capaz de acompanhar o ritmo. O desenvolvimento, a implantação, a vigilância pós-comercialização e a iteração de produtos e serviços de IA na área da saúde devem, portanto, ser uma conversa contínua entre desenvolvedores e reguladores.

Recomenda-se que os reguladores observem a tecnologia baseada em IA na área da saúde a partir de uma mentalidade de engajamento acessível que, potencialmente, quando aplicável, facilite o trabalho ao lado do desenvolvedor para garantir a

conformidade com os requisitos regulatórios durante todo o processo de desenvolvimento e implementação. Uma abordagem de mentalidade de engajamento para a regulação consiste em construir relacionamentos colaborativos e de confiança entre desenvolvedores e o(s) órgão(s) regulador(es), e um diálogo bidirecional que permita aos desenvolvedores aprenderem com os reguladores e vice-versa.

Além disso, dependendo dos arranjos regulatórios de um país, um ou mais reguladores podem ser responsáveis por produtos e serviços de saúde baseados em IA. Isso significa que um desenvolvedor geralmente precisa trabalhar com mais de um órgão regulador (e atender a seus padrões). Para garantir que seja uma experiência suave e positiva para os desenvolvedores de IA, recomenda-se novamente que os reguladores adotem uma abordagem de serviço. Isso significa que um caminho único e claramente sinalizado deve ser estabelecido e seguido por um desenvolvedor de IA ao garantir a conformidade de um produto ou serviço. Os reguladores precisam colaborar uns com os outros em questões como mensagens claras para os desenvolvedores e níveis consistentes de engajamento com os desenvolvedores no ponto certo e compartilhando o que aprenderam com diferentes compromissos com os desenvolvedores.

Se um país deseja adotar uma abordagem de engajamento acessível para a regulação de produtos e serviços de IA, pode-se explorar a corregulação. Tal como referido por Clarke (Clarke, 2019), numa abordagem de corregulação, os reguladores delinearam um *framework* regulatório baseado no cumprimento exigido do(s) ato(s) legislativo(s). Os detalhes de como é aplicado na prática são desenvolvidos em conjunto por reguladores e uma amostra representativa de desenvolvedores (Clarke, 2019). Da mesma forma, ao considerar a regulação a partir de uma mentalidade de serviço, uma abordagem de corregulação, quando apropriada e com quaisquer potenciais conflitos de interesse adequadamente gerenciados, trata de gerar a adesão dos desenvolvedores, envolvendo-os no *design* e na implementação do processo regulatório. A abordagem envolve a concepção de um processo regulatório que reflita e reconheça as necessidades dos desenvolvedores e não apenas as do órgão regulador e grupos associados. Em

última análise, entretanto, os reguladores devem permanecer totalmente independentes dos desenvolvedores para tomar decisões que coloquem a segurança do público em primeiro lugar, bem como garantir que os recursos de saúde públicos e privados sejam usados apenas para tecnologias que atendam aos padrões de qualidade, segurança e eficácia desenvolvidos de forma independente.

RECOMENDAÇÕES PARA O CAMINHO A SEGUIR

Com base em seu trabalho, o GT-CR recomenda que os atores interessados examinem as 18 principais considerações discutidas na seção anterior e resumidas na Tabela 5 abaixo, à medida que continuam a desenvolver quadros e melhores práticas para o uso de IA em cuidados de saúde e desenvolvimento terapêutico.

TABELA 5 - PRINCIPAIS RECOMENDAÇÕES PARA CONSIDERAÇÕES REGULATÓRIAS SOBRE O USO DE IA PARA A SAÚDE COM BASE EM CADA UMA DAS SEIS ÁREAS TEMÁTICAS

1. Documentação e recomendações de transparência
1.1 Considere pré-especificar e documentar a finalidade médica pretendida e o processo de desenvolvimento, como a seleção e o uso de conjuntos de dados, padrões de referência, parâmetros, métricas, desvios dos planos originais e atualizações/alterações durante as fases de desenvolvimento. Estes devem ser considerados de forma a permitir o rastreamento das etapas de desenvolvimento, conforme apropriado. 1.2 Considere uma abordagem baseada em risco também para o nível de documentação e manutenção de registros utilizados para o desenvolvimento e a validação de sistemas de IA.
2. Recomendações para a abordagem de gestão de risco e ciclo de vida de desenvolvimento de sistemas de IA
2.1 Considere uma abordagem de ciclo de vida total do produto em todas as fases da vida de um dispositivo médico: gestão de desenvolvimento pré-comercialização, gestão/vigilância pós-comercialização e gestão de alterações. 2.2 Considere uma abordagem de gestão de riscos que aborde os riscos associados aos sistemas de IA, como ameaças e vulnerabilidades de cibersegurança, subadaptação, viés algorítmico, entre outros.

3. Uso pretendido e recomendações de validação analítica e clínica

3.1 Considere fornecer documentação transparente do uso pretendido do sistema de IA. Os detalhes da composição do conjunto de dados de treinamento subjacente a um sistema de IA — incluindo tamanho, configuração e população, dados de entrada e saída e composição demográfica — devem ser documentados de forma transparente e fornecidos aos usuários.

3.2 Considere demonstrar o desempenho além do conjunto de dados de treinamento por meio de validação analítica externa em um conjunto de dados independente. Esse conjunto de dados de validação externa deve ser representativo da população e do ambiente em que o sistema de IA deve ser implantado, e deve-se fornecer uma documentação transparente do conjunto de dados de validação externa e das métricas de desempenho. Esse conjunto de dados de validação externa deve ser adequadamente independente do conjunto de dados usado para o desenvolvimento do modelo de IA durante o treinamento e o teste.

3.3 Considere um conjunto de requisitos graduado para validação clínica com base no risco. Os ensaios clínicos randomizados são o padrão-ouro para a avaliação do desempenho clínico comparativo e podem ser apropriados para as ferramentas de maior risco ou quando for necessário o mais alto padrão de evidência. Em outras situações, considere a validação prospectiva em um ensaio de implantação e implementação no mundo real, que inclua um comparador relevante usando grupos aceitos e pertinentes

3.4 Considere um período de monitoramento pós-implantação mais intenso por meio de gestão pós-comercialização e vigilância do mercado para sistemas de IA de alto risco.

4. Recomendações de qualidade dos dados

4.1 Considere se os dados disponíveis são de qualidade suficiente para suportar o desenvolvimento do sistema de IA que pode atingir a finalidade pretendida.

4.2 Considere a implantação de avaliações rigorosas de pré-lançamento para sistemas de IA, a fim de garantir que eles não ampliarão nenhum dos problemas relevantes, como enviesamentos e erros.

4.3 Considere um design cuidadoso ou uma solução de problemas precoce para ajudar na identificação antecipada de problemas de qualidade de dados, o que poderia potencialmente prevenir ou mitigar possíveis danos resultantes.

4.4 Considere mitigar os problemas de qualidade de dados que surgem nos dados de saúde e os riscos associados.

4.5 Considere trabalhar com outras partes interessadas para criar ecossistemas de dados que possam facilitar o compartilhamento de fontes de dados de boa qualidade.

5. Recomendações de privacidade e proteção de dados

5.1 Considere a privacidade e a proteção de dados durante o design e a implantação de sistemas de IA.

5.2 Considere obter um bom entendimento das regulações de proteção de dados aplicáveis e das leis de privacidade desde o início do processo de desenvolvimento e assegurar que o processo de desenvolvimento atenda ou supere tais requisitos legais.

5.3 Considere a implementação de um programa de conformidade que aborde os riscos e desenvolva práticas e prioridades de privacidade e cibersegurança que levem em conta os danos potenciais e o ambiente de aplicação.

6. Recomendações de engajamento e colaboração
6.1 Considere o desenvolvimento de plataformas acessíveis e informativas que facilitem o engajamento e a colaboração, quando aplicável e apropriado, entre as principais partes interessadas do roteiro de inovação, implantação de IA, e colaboração 6.2 Considere simplificar o processo de supervisão para a regulação da IA por meio do engajamento e colaboração, visando potencialmente acelerar avanços que mudem práticas na área de IA.

CONCLUSÃO

A OMS reconhece o potencial da IA para melhorar os resultados de saúde ao aprimorar ensaios clínicos, diagnósticos médicos, tratamento, autogestão dos cuidados e cuidados centrados na pessoa, além de criar um conhecimento mais baseado em evidências, habilidades e competências para profissionais que apoiam os cuidados de saúde. Além disso, com a crescente disponibilidade de dados de cuidados de saúde e o rápido progresso das técnicas de análise, a IA tem o potencial de transformar o setor da saúde para atender a uma gama de necessidades dos atores interessados em cuidados de saúde e no desenvolvimento terapêutico. Por essa razão, a OMS e a UIT estão colaborando, por meio do FG-AI4H, para facilitar o desenvolvimento e o uso seguros e apropriados de sistemas de IA para cuidados de saúde. O GT-CR do FG-AI4H é constituído por membros que representam vários atores interessados – incluindo órgãos reguladores, formuladores de políticas públicas, academia e indústria – que exploraram considerações regulatórias e de avaliação de tecnologias de saúde e "boas práticas" emergentes para o desenvolvimento e uso de IA em cuidados de saúde e desenvolvimento terapêutico. Esta publicação, baseada no trabalho do GT-CR, tem como objetivo fornecer uma visão geral das considerações regulatórias sobre IA para a saúde, que abrange as seguintes seis áreas temáticas: documentação e transparência; abordagem do ciclo de vida total do produto e gestão de riscos; uso pretendido e validação analítica e clínica; qualidade dos dados; privacidade e proteção dos dados; e engajamento e colaboração. Esse panorama não pretende ser uma diretriz, um regulamento ou uma política. Ao contrário, trata-se de uma discussão sobre as principais considerações regulatórias e um recurso a ser considerado por todos os atores

interessados relevantes, incluindo desenvolvedores que estão explorando e desenvolvendo sistemas de IA, reguladores que estejam no processo de identificação de abordagens para gerenciar e facilitar sistemas de IA, fabricantes que projetam e desenvolvem dispositivos médicos habilitados para IA, profissionais de saúde que implantam e usam esses dispositivos médicos e sistemas de IA, e aqueles que trabalham nesta área. O GT-CR recomenda que os atores interessados examinem essas principais considerações e outras potenciais considerações à medida que continuam a desenvolver *frameworks* e melhores práticas para a utilização da IA nos cuidados de saúde e no desenvolvimento terapêutico em relação às seis áreas temáticas.

O GT-CR reconhece que a IA tem sido fundamental para o rápido avanço da pesquisa em cuidados de saúde e desenvolvimento terapêutico. Entretanto, reconhece, também, a crescente complexidade do cenário de IA e a necessidade de colaboração internacional para facilitar o desenvolvimento e o uso seguros e apropriados de sistemas de IA. Assim, a colaboração internacional em regulações e padrões de IA é importante por três razões. Em primeiro lugar, o compartilhamento de conhecimento e melhores práticas de considerações regulatórias em evolução poderia aumentar a velocidade de desenvolvimento desse cenário regulatório e reduzir a lacuna entre o avanço da tecnologia e a regulação. Em segundo lugar, a colaboração internacional melhora a consistência das regulações, o que é importante, pois muitas ferramentas provavelmente cruzarão fronteiras. A consistência das considerações regulatórias para sistemas e tecnologias de IA poderia melhorar os padrões e permitir uma implantação mais rápida. Em terceiro lugar, a colaboração internacional apoia os países com menor capacidade regulatória, garantindo que esses países também possam usar ferramentas com altos padrões, reduzindo o potencial de disparidade na introdução dessas ferramentas. Por fim, o GT-CR entende que o cenário de IA tem evoluído rapidamente e que as considerações contidas neste documento podem precisar ser expandidas à medida que a tecnologia e seus usos se desenvolvem. O grupo de trabalho recomenda que os atores interessados, incluindo reguladores, desenvolvedores e fabricantes, continuem a se envolver e a comunidade em geral trabalhe para o entendimento compartilhado e o aprendizado mútuo. Além disso, os grupos

nacionais e internacionais reconhecidos, como o IMDRF, o GHWP, o AMDF e o ICMRA, devem continuar trabalhando em temas relacionados com a IA para uma potencial convergência e harmonização regularórias.

REFERÊNCIAS

Alsalamah, S. A., Alsalamah, H. A., Nouh, T., & Alsalamah, S. A. (2021). HealthyBlockchain for global patients. *Computers, Materials & Continua*, *68*(2),2431–2449. https://www.researchgate.net/publication/351028022_HealthyBlockchain_for_Global_Patients

Atlassian. (s.d.). *Stakeholder communications*. https://www.atlassian.com/team-playbook/plays/ stakeholder-communications-plan

Attrey, A., Lesher, M., & Lomax, C. (2020). The role of sandboxes in promoting flexibility and innovation in the digital age. *Going Digital Toolkit Note, 2*. https://goingdigital.oecd.org/data/notes/No2_ToolkitNote_Sandboxes.pdf

Bellemo, V., Lim, Z. W., Lim, G., Nguyen, Q. D., Xie, Y., Yip, M. Y. T., Hamzah, H., Ho, J., Lee, X. Q., Hsu, W., Lee, M. L., Musonda, L., Chandran, M., Chipalo-Mutati, G., Muma, M., Tan, G. S. W., Sivaprasad, S., Menon, G., Wong, T. Y., & Ting, D. S. W. (2019). Artificial Intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: A clinical validation study. *Lancet Digit Health*, *1*(1), e35–e44. https://www.thelancet.com/journals/landig/article/PIIS2589-7500(19)30004-4/fulltext

Blasiak, A., Lim, J. J., Seah, S. G. K., Kee, T., Remo, A., Chye, D. H., Wong, P. S., Hooi, L., Truong, A. T. L., Le, N., Chan, C. E. Z., Desai, R., Din, X., Hanson, B. J., Chow, E. K.-H.,Ho, D. (2020). IDentif.AI: Rapidly optimizing combination therapy design against severe acute respiratory syndrome Coronavirus 2 (SARS-Cov-2) with digital drug development. *Bioengineering & Translational Medicine*, *6*(1), 1-16. https://aiche.onlinelibrary.wiley.com/doi/10.1002/ btm2.10196

Calderon-Ramirez, S., & Oala, L. (2021). More than meets the eye: Semi-supervised learning under non-IID data. In *The International Conference on Learning Representations*. https://arxiv.org/abs/2104.10223

Clarke, R. (2019). Regulatory alternatives for AI. *Computer Law & Security Review, 35*(4), 398–409.

Comissão Europeia. (2022). *Apresentação do primeiro ambiente de testagem da regulamentação sobre Inteligência Artificial*. https://digital-strategy.ec.europa.eu/pt/news/first-regulatory-sandbox-artificial-intelligence-presented

Comissão Europeia. (s.d.). *Decisões de adequação: como a UE determina se um país não-membro dispõe de um nível adequado de proteção de dados*. https://ec.europa.eu/info/law/law-topic/data-protection/ international-dimension-data-protection/adequacy-decisions_en

Department of Health and Age Care. (2017). *Stakeholder Engagement Framework*. https://www.health.gov.au/resources/publications/stakeholder-engagement-framework

Duke-Margolis Center for Health Policy. (2019). *Determining real-world data's fitness for use and the role of reliability*. https://healthpolicy.duke.edu/sites/default/files/2019-11/rwd_reliability.pdf

Food and Drug Administration. (2019). *Proposed regulatory framework for modifications to Artificial Intelligence/machine learning (AI/ML)-based Software as Medical Device (SaMD) - Discussion paper and request for feedback*. https://www.fda.gov/files/medical%20devices/ published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf

Food and Drug Administration. (2021). *Artificial Intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) Action plan*. https://www.fda.gov/media/145022/download

Food and Drug Administration. (2024). *Real-world data: assessing electronic health records and medical claims data to support regulatory decision-making for drug and biological products*. https://www.fda.gov/regulatory-information/search-fda- guidance-documents/real-world-data-assessing-electronic-health-records-and-medical-claims-data- support-regulatory

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., & Crawford, K. (2021). Datasheets for datasets. *ACM Communications, 64*(12), 86–92. https://arxiv.org/abs/1803.09010

Greenleaf, G. (2021). Global tables of data privacy laws and bills. *Proceedings of the Privacy Laws & Business International Report*, 169, 6–19. http://dx.doi.org/10.2139/ssrn.3836261

Harrer, S., Shah, P., Antony, B., & Hu, J. (2019). Artificial Intelligence for clinical trial design. *Trends in Pharmacological Sciences, 40*(8), 577–591. https://www.sciencedirect.com/science/article/pii/S0165614719301300#:~:text=AI%20techniques%20have %20advanced%20to,to%20assist%20human%20decision%2Dmakers.|&text=We%20explain%20how%20recent%20advances,towards%20 increasing%20trial%20success%20rates

Health Sciences Authority. (2019). *Import and supply of unregistered medical devices by request of qualified practitioners*. https://www.hsa.gov.sg/ medical-devices/registration/special-access-routes/qualified-practitioner-request

Health Sciences Authority. (2023). *Complementary health products (CHP) classification tool*. https://www.hsa.gov.sg/CHP-classification-tool

Health Sciences Authority. (2022). *Regulatory guidelines for software medical devices – A lifecycle approach* (*online*). https://www.hsa.gov.sg/docs/default-source/hprg-mdb/guidance- documents-for-medical-devices/regulatory-guidelines-for-software-medical-devices---a-life-cycle- approach_r2-(2022-abr)-pub.pdf

Ho, D., Quake, S. R., McCabe, E. R. B., Chng, W. J., Chow, E. K., Ding, X., Gelb, B. D., Ginsburg, Hassenstab, J., Ho, C.-M., Mobley, W. C., Nolan, G. P., Rosen, S. T., Tan, P., Yen, Y., & Zarrimpar, A. (2020). Enabling technologies for personalized and precision medicine. *Trends in Biotechnology*, *38*(5), 497–518. https://www.cell.com/trends/biotechnology/fulltext/S0167-7799(19)30316-6

Ho, D. (2020a). Artificial Intelligence in cancer therapy. *Science*, *367*(6481), 982–983. https://science. sciencemag.org/content/367/6481/982

Ho, D. (2020b). Addressing COVID-19 drug development with Artificial Intelligence. *Advanced Intelligent Systems*, *2*(5). https://onlinelibrary.wiley.com/doi/full/10.1002/aisy.202000070

Information Commissioner's Office. (2021). *Introduction to anonymisation: draft anonymisation, pseudonymisation, and privacy enhancing technologies guidance*. https://ico.org.uk/Media/About-the-ICO/Consultas/2619862/anonymisation-intro-and-first-chapter.pdf

International Medical Device Regulators Forum. (2013). *Software as a Medical Device (SaMD): Key definitions*. http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-131209-samd-key-definitions-140901.pdf

International Medical Device Regulators Forum. (2014). *Software as a medical device: Possible framework for risk categorization and corresponding considerations*. https://www.imdrf. org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk- categorization-141013.pdf

International Medical Device Regulators Forum. (2016). *Software as a medical device (SaMD): Clinical evaluation*. http://www.imdrf.org/docs/imdrf/final/consultations/imdrf-cons-samd-ce.pdf

International Medical Device Regulators Forum. (2019). *Principles and practices for medical device cybersecurity*. http://www.imdrf.org/docs/imdrf/final/consultations/imdrf-cons-ppmdc.pdf

Lei de Proteção de Dados Pessoais Digitais. (2023). Lei n. 22, de 11 de agosto de 2023. Dispõe sobre o tratamento de dados pessoais digitais de maneira que reconheça tanto o direito dos indivíduos de proteger seus dados pessoais quanto a necessidade de tratar tais dados pessoais para fins legais e para assuntos relacionados ou incidentais a isso. https://www.meity.gov.in/writereaddata/files/Digital%20Personal%20Data%20Protection%20Act%202023.pdf

Liu, X., Rivera, S. C., Moher, D., Calvert, M. J., Denniston, A. K.; & Grupo de Trabalho Standar Protocol Items: Recommendations for International Trials for AI e Consolidates Standars of Reporting Trials for AI. (2020). Reporting guidelines for clinical trial reports for interventions involving Artificial Intelligence: The CONSORT-AI extension. *Nature Medicine, 26*, 1364–1374. https://www.nature.com/articles/s41591-020-1034-x#citeas

Macdonald, J., März, M., Oala, L., & Samek, W. (2021). Interval neural networks as instability detectors for image reconstructions. In: Palm, C., Deserno, T. M., Handels, H., Maier, A., Maier-Hein, K., Tolxdorff, T. (Eds.), *Bildverarbeitung für die Medizin*, . Informatik aktuell (Image processing for medicine. IT update). Springer Vieweg, Wiesbaden. https://doi.org/10.1007/978-3-658-33198-6_79

Madiega, T., & van de Pol, A. L. (2022). Artificial Intelligence act and regulatory sandboxes. *European Parliamentary Research Service*. https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI(2022)733544_EN.pdf

National Institute for Health and Care Excellence. (2019). *Evidence standards framework for digital health technologies*. https://www.nice.org.uk/Media/Default/About/what-we-do/our- programmes/evidence-standards-framework/digital-evidence-standards-framework.pdf

National Institute of Standards and Technology. (2018). *Framework for improving critical infrastructure cybersecurity*. https://www.nist.gov/cyberframework

National Institute of Standards and Technology. (2020). *NIST privacy framework: A tool for improving privacy through enterprise risk management*. https://www.nist.gov/system/files/documents/2020/01/16/NIST%20Privacy%20Framework_V1.0.pdf

National Health Service. (2020). *A buyer's guide to AI in health and care*. https://www.nhsx.nhs.uk/ai-lab/explore- all-resources/adopt-ai/a-buyers-guide-to-ai-in-health-and-care/

Notificação do Departamento de Dispositivos Médicos e Produtos Farmacêuticos n. 14, emitida em 31 de agosto de 2021. (2012). Notificação n. 0831-14, de 31 de agosto de 2020 (chinês). Manuseio de solicitações para confirmação de PACMP para dispositivos médicos, PSEHB/SD (em japonês). Tóquio: Ministério da Saúde, Trabalho e Bem-Estar; 2020. https://www.mhlw.go.jp/content/11120000/000665757.pdf

Oala, L., Fehr, J., Gilli, L., Balachandran, P., Leite, A. W., Calderon-Ramirez, S., Li, D. X., Nobis, G., Alvarado, E. A. M., Jaramillo-Gutierrez, G., Matek, C., Shroff, A., Kherif, F., Sanguinetti, B., & Wiegand, T. (2020). ML4H Auditing: From paper to practice. *Proceedings of Machine Learning for Health (ML4H) NeurIPS Workshop*, 136, 280-317. https://proceedings.mlr.press/v136/oala20a.html

Oala, L., Johner, C., Goldschmidt, P. G., & Balachandran, P. (2022). Good Practices for Health Applications of Machine Learning: Considerations for Manufacturers and Regulators. *Proceedings of the Focus Group on Artificial Intelligence for Health*. https://www.itu.int/dms_pub/ itu-t/opb/fg/T-FG-AI4H-2022-2-PDF-E.pdf

Oala, L., Heiß, C., Macdonald, J., März, M., Kutyniok, G., & Samek, W. (2021). Detecting failure modes in image reconstructions with interval neural network uncertainty. *International Journal of Computer Assisted Radiology and Surgery*, *16*, 2089–2097. https://link.springer.com/article/10.1007/s11548-021-02482-2

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. https://www.science.org/doi/10.1126/science.aax2342

Organização Mundial da Saúde. (2020). *Guidance for post-market surveillance and market surveillance of medical devices, including in vitro diagnostics*. https://apps.who.int/iris/handle/10665/337551

Organização Mundial da Saúde. (2021a). *Global Strategy on Digital Health 2020–2025*. https://apps. who.int/iris/handle/10665/34424923).

Organização Mundial da Saúde. (2021b). *Ethics and governance of Artificial Intelligence for health: WHO guidance*.https://apps.who.int/iris/handle/10665/341996

Pagallo, U., Casanovas, P., & Madelin, R. (2019). The middle-out approach: Assessing models of legal governance in data protection, Artificial Intelligence, and the Web of Data. *The Theory and Practice of Legislation, 7*(1),1–25.

Panesar, A. (2019). *Machine learning and AI for healthcare: Big data for improved health outcomes*. Apress.

Real Decreto 817, de 8 de novembro de 2023. (2023). Estabelece um ambiente controlado de testes para a avaliação da conformidade da proposta de Regulamento do Parlamento Europeu e do Conselho que estabelece normas harmonizadas em matéria de Inteligência Artificial. *Boletim Oficial do Estado*. https://www.boe.es/boe/dias/2023/11/09/pdfs/BOE-A-2023-22767.pdf

Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016. (2016). Relativo à proteção das pessoas físicas no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE (Regulamento Geral sobre a Proteção de Dados) (Texto relevante para efeitos do EEE). https://eur-lex.europa.eu/eli/reg/2016/679/oj?locale=pt

Rivera, S. C., Liu, X., Chan, A., Denniston, A. K., & Calvert, M. J. (2020). Guidelines for clinical trial protocols for interventions involving Artificial Intelligence: The SPIRIT-AI extension. *BMJ*, 1351-1363.

Salim, M., Wåhlin, E., Dembrower, K., Azavedo, E., Foukakis, T., Liu, Y., Smith, K., Eklund, M., & Strand, F. (2020). External evaluation of 3 commercial Artificial Intelligence algorithms for independent assessment of screening mammograms. *JAMA Oncology, 6*(10), 1581–1588. https://pubmed.ncbi.nlm.nih.gov/32852536/

Sendak, M. P., Gao, M., Brajer, N., & Balu, S. (2020). Presenting machine learning model information to clinical end users with model facts labels. *npj Digital Medicine, 3*(1), 1–4.

Shah, P., Kendall, F., Khozin, S., Goosen, R., Hu, J., Laramie, J., Ringel, M., & Schork, N. (2019) Artificial Intelligence and machine learning in clinical development: A translational perspective. *npj, 2*(69), 1-5. www.nature.com/ artigos/ s41746-019-0148-3

Shana, L. (2020). The geographic bias in medical AI tools. Ethics and Justice, Healthcare, Machine Learning. *Stanford University Human-Centered Artificial Intelligence*. https://hai.stanford.edu/news/geographic-bias-medical-ai-tools

Topol, E. J. (2020). Welcoming new guidelines for AI clinical research. *Nature Medicine, 26,* 1318–1320. https://www.nature.com/articles/s41591-020-1042-x#:~:text=A%20new%20era%20needs%20new,in%20conjunction%20with%20the%20algorithm.

União Internacional das Telecomunicações. (2020). *Workshop on clinical evaluation of AI for health*. https://www.itu.int/en/ITU-T/focusgroups/ai4h/Pages/ws/2010.aspx

União Internacional das Telecomunicações. (2021). *FG-AI4H Open Code Initiative – evaluation and reporting package*.

União Internacional de Telecomunicações. (2022). *Iniciativa de Código Aberto FG-AI4H (OCI)*. https://www.itu. int/en/ITU-T/focusgroups/ai4h/Pages/opencode.aspx

Verks, B., & Oala, L. (2020)Data and Artificial Intelligence assessment methods (DAISAM) Audit Reporting Template. In *Proceedings of the ITU/WHO Focus Group on Artificial Intelligence for Health*.

West, D. M., & Allen, J. R. (2020). *Turning point: Policymaking in the era of Artificial Intelligence*. Brookings Institution Press.

Willis, K., Oala, L. (2021). Adaptação de domínio post-hoc via homogeneização guiada de dados. *arXiv*, 1-4. https://arxiv.org/abs/2104.03624

World Intellectual Property Organisation. (s.d.). *Artificial Intelligence and intellectual property policy*. https://www.wipo.int/about-ip/en/artificial_intelligence/policy.html

Wu, E., Wu, K., Daneshjou, R., Ouyang, D., Ho, D. E., & Zou, J. (2021). How medical AI devices are evaluated: Limitations and recommendations from an analysis of FDA approvals. *Nature Medicine, 27*, 582–584. https://www.nature.com/articles/s41591-021-01312-x

Transparência e explicabilidade: perspectivas para a regulação da Inteligência Artificial na saúde no Brasil

Daniel A. Dourado[1] *e Fernando Aith*[2]

1 Médico psiquiatra, advogado de saúde, mestre e doutor em Medicina Preventiva/Saúde Coletiva pela Faculdade de Medicina da Universidade de São Paulo (FMUSP), também é pesquisador associado e docente do Centro de Pesquisa em Direito Sanitário da Universidade de São Paulo (Cepedisa/USP).
2 Advogado; Professor Titular do Departamento de Política, Gestão e Saúde da Faculdade de Saúde Pública da Universidade de São Paulo; (HSP/FSP); Coordenador Científico do Centro de Pesquisas em Direito Sanitário da USP (Cepedisa/USP); Editor-chefe da Revista de Direito Sanitário da USP.

INTRODUÇÃO

A área da saúde encontra-se em um momento de transformação impulsionada pela Inteligência Artificial (IA), que tem o potencial de revolucionar práticas e serviços de saúde em todo o mundo. A era digital introduziu tecnologias inovadoras com capacidade de alterar substancialmente a estrutura dos sistemas de saúde, abrindo um leque de oportunidades inéditas para aprimorar a qualidade do cuidado individual e populacional, expandir o acesso, reduzir custos e explorar novas fronteiras na prevenção, no diagnóstico e no tratamento de doenças (Matheny *et al.*, 2022). Embora a integração da IA na saúde tenha progredido a um ritmo mais lento em comparação com outros setores, há um consenso crescente sobre sua inevitável integração em diversos domínios da medicina e da saúde pública (Sahni & Carrus, 2023). Para que essa tecnologia seja usada de forma segura, transparente, responsável e justa, o desenvolvimento de marcos regulatórios específicos para a IA na saúde é considerado essencial (Organização Mundial da Saúde [OMS], 2021).

Nesse contexto, a regulação da IA na saúde assume importância central. Os sistemas de IA precisam ter sua qualidade e segurança verificadas, reconhecendo que as ações e os serviços de saúde, tradicionalmente realizados por pessoas, têm sido significativamente influenciados e executados por sistemas automatizados. Portanto, ao mesmo tempo em que a adoção dessas novas tecnologias deve ser incentivada, é essencial que se estabeleça uma estrutura regulatória capaz de garantir que seu uso seja sempre em benefício dos seres humanos.

As ferramentas desenvolvidas para a saúde a partir de modelos de aprendizado de máquina (*machine learning* [AM]) representam reconhecidamente os maiores desafios nesse cenário, em função da capacidade de aprendizado com a experiência do mundo real e de adaptação para melhorar continuamente o desempenho (Bates, 2023). Diferentemente dos objetos da regulação em saúde tradicional, como medicamentos e dispositivos médicos, sistemas de IA podem mudar continuamente mesmo depois de implementados (Gottlieb & Silvis, 2023).

Uma agenda para a regulação da IA na saúde vem sendo implementada no nível global na primeira metade da década de 2020 com contribuições de diversas entidades, como a OMS, o

Fórum Internacional de Reguladores de Dispositivos Médicos (International Medical Device Regulators Forum [IMDRF]), a Organização Internacional para Padronização (International Organization for Standardization [ISO]), a Organização para a Cooperação e Desenvolvimento Econômico (OCDE), entre outras. Tem-se buscado definir princípios gerais que possam ser aplicados na regulação da IA na saúde em diferentes contextos: por países de baixa, média e alta renda, setores público e privado, e governos e organizações. As primeiras abordagens regulatórias relevantes para a IA na saúde relacionam-se às dimensões da governança de dados de saúde e da segurança e eficácia de dispositivos médicos baseados em IA. A evolução muito acelerada de seu panorama tem deixado essa atividade cada vez mais complexa (OMS, 2021, 2023, 2024).

Este artigo tem o propósito de analisar outra dimensão fundamental para a regulação da IA na saúde: a transparência. Para isso, faremos uma breve exposição sobre o conceito geral da IA, suas aplicações na saúde e na medicina, e apresentaremos os princípios éticos que servem de fundamento para a regulação da IA na saúde, destacando a importância da transparência. Em seguida, exploraremos os conceitos de interpretabilidade e explicabilidade dos sistemas de IA, frequentemente associados às expressões normativas de transparência. Abordaremos o direito à explicação sobre decisões automatizadas como mecanismo legal de transparência, examinando o atual estágio de discussão sobre essa matéria no cenário internacional e no Brasil. Por fim, analisaremos os limites das atuais técnicas de explicação em IA, a fim de propormos possíveis estratégias regulatórias a partir dos elementos identificados.

APLICAÇÕES DA IA NA SAÚDE E NA MEDICINA

Desde o surgimento da IA como campo de estudo, a área da saúde estabeleceu-se como uma das mais propícias para sua aplicação. Na década de 1950, os primeiros sistemas de apoio à atividade clínica foram concebidos e desenvolvidos por pesquisadores pioneiros. Na década de 1970, foram criados diversos programas que visavam simular o raciocínio humano especializado, com o objetivo de auxiliar médicos na formulação de hipóteses diagnósticas em casos complexos,

na interpretação de exames clínicos e na seleção de tratamentos adequados. Esses programas, denominados "sistemas de suporte à decisão clínica" (*clinical decision support systems* [CDSS]), representaram aplicações significativas da chamada "IA simbólica", paradigma dominante na época, e foram usados durante as décadas de 1980 e 1990 (Sutton *et al.*, 2020). No entanto, esses sistemas baseados em regras evidenciaram as limitações dessa abordagem tecnológica, destacando-se principalmente o elevado custo de manutenção e a necessidade de atualizações constantes, que demandavam revisões frequentes conduzidas por especialistas. Além disso, o desempenho desses sistemas estava restrito pela precisão do conhecimento médico previamente existente (Yu *et al.*, 2018).

Nesse cenário, o surgimento do campo de aprendizado de máquina foi recebido com grande entusiasmo no âmbito da saúde e da medicina. Essa técnica é um subtipo de IA que oferece ferramentas para o desenvolvimento de sistemas capazes de identificar padrões, até então desconhecidos, em conjuntos de dados, sem a necessidade de pré-definir regras de decisão para cada tarefa específica. A crescente disponibilidade de grandes volumes de dados na área da saúde, combinada com o aumento exponencial da capacidade computacional, impulsionou uma expectativa cada vez maior em relação à incorporação substancial da IA no setor, expectativa que se intensificou especialmente a partir da década de 2000 (Rajkomar *et al.*, 2019). A recente expansão no interesse pela IA na área da saúde deve-se, ainda, à implementação bem-sucedida de técnicas de um subtipo de aprendizado de máquina, o aprendizado profundo (*deep learning* [AP]), em diversos domínios, principalmente a partir de meados da década de 2010 (Hinton, 2018).

A aplicação da IA na área da saúde e medicina tem apresentado um crescimento acelerado. Começa pelas áreas baseadas na identificação de padrões em imagens – como radiologia, patologia e dermatologia – e expande-se para além da visão computacional, abrangendo áreas como: processamento de linguagem natural para análise de dados em registros eletrônicos de saúde, análise de informações genéticas em medicina de precisão e aprendizado por reforço (*reinforcement learning*) em cirurgia assistida por robótica. Essas técnicas

têm demonstrado desempenho promissor na realização de previsões úteis e precisas em diferentes cenários clínicos (Topol, 2019) e podem ser muito úteis também no nível coletivo, na tomada de decisões em saúde pública e no planejamento para alocação de recursos humanos e financeiros. Um relevante exemplo disso é o uso de ferramentas de IA para identificar e rastrear surtos de doenças infecciosas, além de monitorar estratégias de mitigação (Brownstein *et al.*, 2023). Ademais, novas ferramentas de IA têm potencial para reduzir significativamente o tempo e o custo associados aos métodos tradicionais de descoberta de medicamentos (Nature, 2023).

Nos últimos anos, a área da saúde presenciou o surgimento de uma nova fronteira com a rápida ascensão dos modelos de fundação (*foundation models*) a partir da década de 2020, com destaque para os grandes modelos de linguagem (*large language models* [LLM]). As aplicações dos modelos de fundação na área da saúde são vastas e apresentam um potencial transformador. Os modelos desenvolvidos nesse paradigma demonstram a capacidade de interpretar diversos tipos de informações médicas, por exemplo: imagens, registros de prontuários eletrônicos, resultados de exames laboratoriais, dados genômicos e textos médicos. Além disso, esses modelos podem oferecer diversos resultados, como explicações em linguagem natural, recomendações, anotações e até mesmo interagir com o interlocutor humano (Moor *et al.*, 2023). A expansão desses modelos para se tornarem multimodais – ou seja, capazes de interpretar não apenas texto, mas também imagens, áudio e vídeo – amplia exponencialmente suas possibilidades de aplicação na saúde individual e coletiva (Acosta *et al.*, 2022).

PRINCÍPIOS ÉTICOS PARA A REGULAÇÃO DA IA NA SAÚDE

A regulação da IA tornou-se uma prioridade global a partir dos anos 2020 (G7 Hiroshima Summit, 2023). A abordagem regulatória começou, então, a se estruturar por meio de códigos de conduta e diretrizes não vinculantes (*soft law*), produzidos por entidades governamentais, conselhos de especialistas para assessoramento de entes públicos, institutos de pesquisa e empresas privadas. Os principais fundamentos derivam do campo interdisciplinar da ética em IA, que aborda

as implicações morais, legais e sociais dessa tecnologia com intuito de orientar seu desenvolvimento e seu uso em harmonia com os valores humanos e normas sociais (Dubber *et al.*, 2020). Nos últimos anos, tem-se se formado uma convergência global em torno de cinco princípios éticos: (a) transparência; (b) justiça e equidade; (c) não-maleficência; (d) responsabilidade; e (e) privacidade. No entanto, ainda existe significativa divergência em relação ao sentido desses princípios, como devem ser interpretados e qual o caminho para implementá-los (Jobin *et al.*, 2019).

A importância de definir princípios éticos específicos para regulação da IA na saúde começou a ser debatida na segunda metade da década de 2010, quando surgiram as primeiras evidências de que o campo de aprendizado de máquina seria altamente promissor e revolucionário para a área da saúde (International Bioethics Committee, 2017). Em 2021, a OMS publicou o guia inaugural, com diretrizes sobre ética e governança da IA na saúde com pretensão de alcance global (OMS, 2021). Esse documento consolida os primeiros princípios básicos considerados consensuais no campo da ética em IA especificamente para a área saúde: (a) proteger a autonomia; (b) promover o bem-estar humano, a segurança humana e o interesse público; (c) garantir transparência, explicabilidade e inteligibilidade; (d) promover responsabilidade e prestação de contas; (e) assegurar inclusão e equidade; e (f) promover uma IA responsiva e sustentável.

A OMS tem liderado o processo para definição de uma estratégia global de saúde digital a fim de apoiar os sistemas nacionais de saúde, na qual se insere a preparação de uma estrutura de governança e regulação da IA. Nesse contexto, os princípios da OMS para ética da IA na saúde têm o propósito de orientar desenvolvedores, usuários e reguladores no desenvolvimento, na avaliação, na implementação e na avaliação contínua de tecnologias de IA voltadas à saúde. As diretrizes baseiam-se na conjugação dos princípios bioéticos básicos: autonomia, não-maleficência, beneficência e justiça (Beauchamp & Childress, 1979), com os atualmente reconhecidos princípios gerais de ética em IA: privacidade, transparência, responsabilidade (Floridi *et al.*, 2018). A OMS pretende, entre outras medidas, trabalhar de maneira

coordenada com entidades intergovernamentais a fim de identificar e formular leis e políticas, considerando a iniciativa de elaborar uma legislação modelo para ser usada como referência por governos que pretendam criar regulações próprias para IA na saúde (OMS, 2021, 2023).

A transparência destaca-se como o princípio ético mais frequentemente observado nos códigos de conduta que delineiam diretrizes gerais para a utilização da IA (Jobin *et al.*, 2019). No contexto da ética da IA aplicada à saúde, a transparência está vinculada à recomendação de que os sistemas sejam claros e compreensíveis para desenvolvedores, reguladores e usuários, incluindo profissionais de saúde e pacientes. Para isso, é essencial que as informações pertinentes sobre os sistemas de IA sejam documentadas antes de sua implementação e continuem a ser divulgadas de forma regular após a aprovação para uso. Nesse contexto, é imprescindível facilitar a consulta pública e a compreensão acerca do funcionamento dos modelos de IA no mundo real (OMS, 2021). Além disso, espera-se que essas tecnologias sejam explicáveis de acordo com a capacidade de entendimento daqueles a quem a explicação for direcionada, fornecendo esclarecimentos a respeito do funcionamento e das condições das decisões dos algoritmos para profissionais de saúde, pacientes e demais usuários dos sistemas (Watson *et al.*, 2019).

TRANSPARÊNCIA COMO FUNDAMENTO PARA A REGULAÇÃO DA IA NA SAÚDE

O reconhecimento da transparência como uma dimensão regulatória para a aplicação da IA na área da saúde decorre da ampla aceitação desse princípio ético fundamental. É importante ressaltar que o conceito abrangente de transparência implica a disponibilização de todas as informações que justifiquem as decisões baseadas nos resultados gerados por sistemas de IA para todas as partes envolvidas. Esse aspecto é considerado essencial para estabelecer a confiança da sociedade nas tecnologias de IA (Comissão Europeia, 2019; Floridi *et al.*, 2018). Portanto, a transparência em sentido amplo refere-se à explicação do contexto institucional no qual os sistemas de IA são concebidos, implementados e gerenciados, cujo foco está na compreensão global das pessoas

e das organizações responsáveis pelo desenvolvimento, pelo uso e pela regulação da IA.

Nesse sentido, a transparência pode ser promovida por meio de mecanismos como a adoção de documentação padronizada sobre a criação, o treinamento e a implementação de sistemas e modelos de IA, bem como por meio de processos de avaliação do impacto desses sistemas em diferentes contextos de aplicação (Mittelstadt, 2022). A documentação precisa e abrangente deve ser o principal mecanismo para assegurar transparência nas aplicações de IA na saúde; desse modo, o registro adequado de informações claras e detalhadas sobre métodos, recursos e decisões tomadas ao longo de todo o ciclo de vida dos sistemas de IA torna-se um requisito regulatório essencial. Reguladores devem ter acesso a uma documentação adequada que contemple desde a concepção, o desenvolvimento, o treinamento e a validação de modelos até a implementação e o período pós-implementação (OMS, 2023). Isso significa exigir o fornecimento de informações precisas sobre os pressupostos e as limitações dos sistemas de IA, de modo a abarcar os protocolos de operação e os métodos de seleção, processamento e rotulagem dos dados, bem como as condições de desenvolvimento e validação dos modelos de aprendizado de máquina.

Entretanto, o termo transparência é também amplamente empregado no campo da IA, particularmente em aprendizado de máquina, em sentidos mais específicos. A noção de transparência pode estar vinculada à compreensão do funcionamento de um algoritmo, sem necessariamente analisar os dados de treinamento ou as previsões individuais de um modelo (OCDE, 2019), perspectiva usualmente referida como transparência algorítmica ou "transparência do algoritmo" (Association for Computing Machinery US [USACM] Public Policy Council, 2017). Alternativamente, pode-se entender transparência como o acesso aos elementos essenciais para compreender como um modelo toma decisões, o que implica não apenas conhecer o algoritmo, mas também o modelo treinado e os dados utilizados no treinamento. Para atingir esse objetivo, é imprescindível uma compreensão abrangente dos recursos do modelo e de cada um de seus componentes aprendidos, como pesos e parâmetros (Lipton, 2018; Molnar, 2022). Essa

concepção de transparência relaciona-se com os conceitos de interpretabilidade e explicabilidade dos sistemas de IA, cuja compreensão é crucial para delimitar de que modo tais aspectos devem ser contemplados na regulação da IA na saúde.

IA INTERPRETÁVEL E IA EXPLICÁVEL

Na dimensão da transparência na IA, os termos interpretabilidade e explicabilidade estão estreitamente interligados e frequentemente sobrepostos, pois ainda não há definições e limites de sentido amplamente acordados nesse campo (Mittelstadt, 2022). Uma definição sólida de interpretabilidade aponta para o grau em que a causa de uma decisão tomada por um modelo de IA é compreensível para um observador humano (Miller, 2019). Dessa forma, um modelo é considerado totalmente interpretável quando um ser humano é capaz de discernir o conjunto completo de causas que originaram um determinado resultado; em contraposição, um modelo opaco ou de "caixa-preta" caracteriza-se pela escassez de informações concretas sobre como ou por que um resultado específico foi obtido a partir das entradas (Burrell, 2016). Logo, a transparência pode ser considerada tanto no nível do modelo como um todo quanto no nível de seus componentes. Um modelo interpretável pode ser um modelo simples, no qual um humano é capaz de conectar os dados de entrada com os parâmetros para realizar os cálculos e gerar uma previsão, ou um modelo decomposto, de forma que cada parte (entrada, parâmetro, cálculo) permite um entendimento mais ou menos intuitivo (Lipton, 2018).

A ideia de "explicação" está frequentemente associada à maneira pela qual um observador humano pode adquirir compreensão (Miller, 2019). Dessa forma, a explicação é entendida como uma interpretação *post hoc*, na qual informações dos resultados algorítmicos são derivadas após decisões ou previsões terem sido feitas, processo que pode ocorrer mesmo em modelos opacos (Lipton, 2018). Sob essa perspectiva, a explicabilidade pode ser considerada uma característica independente da interpretabilidade intrínseca do modelo, visto que a explicação pode ser obtida por técnicas específicas nos modelos que não sejam naturalmente interpretáveis. Consequentemente, o termo explicabilidade

tem sido comumente utilizado tanto para se referir a modelos intrinsecamente interpretáveis quanto para as explicações obtidas para os modelos considerados como caixas-pretas (Amann *et al.*, 2022). Apesar da frequente sobreposição de significados, a partir desse referencial é possível adotar definições de IA interpretável e IA explicável, particularmente úteis para a aplicação no campo da saúde (Babic *et al.*, 2021).

A IA interpretável diz respeito aos sistemas de aprendizado de máquina construídos com base em modelos passíveis de interpretação por seres humanos. Embora não sejam necessariamente imediatamente compreensíveis por todos, esses modelos devem permitir a interpretação por pessoas com conhecimento na área. É importante notar que alcançar uma interpretabilidade global dos modelos na prática é bastante desafiador, já que, a partir de um certo número de características, um modelo se torna inconcebível para a mente humana – por exemplo, imaginar um espaço com mais de três dimensões. Portanto, uma abordagem de IA interpretável requer o uso de modelos com um número limitado de parâmetros e pesos, de modo que sejam inteligíveis. Também deve ser fundamentada em algoritmos de "caixa-branca", como funções lineares, em que os parâmetros correspondem a pesos que relacionam entradas e saídas, ou árvores de decisão, que produzem mapas intuitivos com base em regras claramente estabelecidas e compreensíveis (Molnar, 2022).

A IA explicável representa uma abordagem fundamentalmente distinta, pois envolve modelos caixa-preta, os quais são intrinsecamente incompreensíveis para os seres humanos. Em vez de tentar compreender diretamente esses modelos opacos, essa abordagem adota um método alternativo: um segundo algoritmo é treinado para replicar as previsões do modelo opaco, isto é, cria-se um modelo explicativo o qual procura uma função interpretável que se assemelhe o máximo possível às saídas do modelo de caixa-preta. Esse modelo substituto é, então, utilizado para fornecer explicações *post hoc* para as previsões do modelo original, embora não seja capaz de realizar previsões reais com precisão, devido à necessidade de simplificar o número de características para torná-lo compreensível. As explicações podem destacar quais atributos dos dados de entrada no modelo opaco são

mais relevantes para uma previsão específica ou criar um modelo linear de fácil compreensão, cujos resultados se aproximem daqueles do modelo original. Em resumo, a IA explicável concentra-se na tarefa de encontrar um modelo de "caixa-branca" capaz de explicar as previsões feitas por um modelo de caixa-preta (Babic *et al.*, 2021).

FUNDAMENTOS DO DIREITO À EXPLICAÇÃO

O conceito do direito à explicação (*right to explanation*) sobre decisões automatizadas envolve o reconhecimento da necessidade de garantir a todas as pessoas o direito de saber como são tomadas as decisões baseadas em IA que afetam suas vidas. Esse conceito tem se consolidado sobretudo a partir das discussões na elaboração do Regulamento Geral sobre a Proteção de Dados (General Data Protection Regulation [GDPR]) europeu, aprovado em 2016, que entrou em vigor em maio de 2018 (Regulamento (UE) n. 2016/679).

Conforme o regulamento europeu, os titulares dos dados têm o direito de serem informados quando decisões automatizadas os afetam significativamente ou geram efeitos em sua esfera jurídica. Eles também têm o direito de receber "informações úteis sobre a lógica subjacente, bem como a importância e as consequências previstas desse processamento" (artigos 13º, 14º e 15º da GDPR). Além disso, todo titular de dados tem o "direito de não ser sujeito a uma decisão baseada exclusivamente no processamento automatizado, incluindo a elaboração de perfis", a menos que a decisão seja necessária para a celebração ou a execução de um contrato, ou baseada no consentimento explícito do titular; nesses casos, o GDPR concede ao titular dos dados o direito de solicitar intervenção humana e de contestar a decisão automatizada (artigo 22º). O propósito dessas garantias é delineado no "Considerando 71" do GDPR, que oferece orientação interpretativa sobre esses dispositivos. O titular dos dados deve ter "o direito de receber intervenção humana, de expressar seu ponto de vista, de obter uma explicação sobre a decisão resultante dessa avaliação e de contestar a decisão". Dessa forma, além de receber uma explicação compreensível, é estabelecido o direito à oportunidade de ser ouvido, questionar e requerer revisão da decisão automatizada, princípio denominado "devido processo algorítmico" (Kaminski, 2019).

Desde a promulgação do GDPR, antes e após sua entrada em vigor, tem havido debates acerca da existência e do alcance do direito à explicação sobre decisões automatizadas (Bygrave, 2020). Essa questão fundamentalmente surge da aceitação de que a representação matemática complexa dos modelos de aprendizado de máquina é, na maioria das vezes, incompreensível para os seres humanos – especialmente com o aumento do uso de algoritmos de caixa-preta. De maneira geral, o debate divide-se em duas perspectivas. Por um lado, existem aqueles que defendem a viabilidade e o escopo do direito à explicação somente sobre o que se refere à funcionalidade geral do sistema, em vez de sobre decisões individuais e circunstâncias específicas (Wachter *et al.*, 2017). Por outro lado, há a compreensão de que a explicação também deve abranger decisões específicas, com transparência limitada apenas pela intrínseca opacidade dos algoritmos, permitindo, assim, que o titular dos dados exerça seus direitos de acordo com o GDPR e em conformidade com os princípios e leis de direitos humanos (Selbst & Powles, 2017).

Um entendimento alternativo é o de que o GDPR estabelece um sistema de "transparência qualificada" sobre a tomada de decisões algorítmicas. Essa leitura sustenta que a lei define regras direcionadas de diferentes graus de profundidade e escopo voltadas a diversos destinatários, de modo que seja necessário fornecer um tipo de informação aos indivíduos e outro tipo aos especialistas e aos reguladores (Kaminski, 2019).

A importância do direito à explicação para a área da saúde é dar às pessoas humanas a possibilidade de compreender a lógica de decisões automatizadas que impactem condutas tomadas em seus cuidados assistenciais. Tal preocupação deve estar cada vez mais presente em diversas situações clínicas. Atualmente, existem, por exemplo, sistemas de IA capazes de definir critérios para transplantes de órgãos, como alocação, correspondência entre doador e receptor, e preverem a sobrevida de pacientes transplantados (Khorsandi *et al.*, 2021). É possível que sistemas desse tipo comecem a ser usados na prática e que haja diferenças de ordem nas filas de transplantes em comparação com as definidas por critérios clínicos feitas apenas por humanos. Logo, o direito à explicação relaciona-se à dignidade humana, uma vez que decisões dessa natureza não poderiam ser tomadas exclusivamente com base em sistemas do tipo caixa-preta.

DIREITO À EXPLICAÇÃO E TRANSPARÊNCIA DE IA NA LEGISLAÇÃO BRASILEIRA

O debate iniciado a partir do paradigma europeu vem influenciando outras jurisdições, como é o caso do Brasil, que aprovou a Lei Geral de Proteção de Dados Pessoais (LGPD) em 2018 (Lei n. 13.709/2018), em vigor desde agosto de 2020 (as normas sobre sanções administrativas entraram em vigor em agosto de 2021). A LGPD é inspirada no GDPR e espelha muitos dos institutos criados pelo padrão europeu (Aith & Dallari, 2022). Assim, embora não trate especificamente da regulação de IA (os termos "Inteligência Artificial" e "algoritmo" sequer aparecem no texto da lei), a LGPD introduz no ordenamento jurídico brasileiro as bases legais do direito à explicação e à revisão de decisões automatizadas.

O direito à revisão de decisões automatizadas está definido explicitamente no art. 20 da LGPD, que concede ao titular o direito a "solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses" (Lei no 13.709/2018). Diferentemente do GDPR, a lei brasileira não prevê o direito de não se sujeitar a decisão exclusivamente automatizada e nem de obter intervenção humana em caso de revisão (Regulamento (UE) n. 2016/679)[3]. O direito à explicação não aparece textualmente na lei brasileira (assim como no GDPR), mas decorre da interpretação sistemática da própria LGPD em conjunto com dispositivos constitucionais e da legislação de proteção ao consumidor (Monteiro *et al.*, 2021). A lei brasileira garante a todo aquele afetado por decisões automatizadas o direito a obter informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados. Essa expressão da transparência só pode ser garantida por meio de alguma forma de explicação.

Os direitos à explicação e à revisão de decisões de sistemas de IA estão necessariamente vinculados e precisam ser compreendidos conjuntamente. Assim como deve acontecer em

3 Na redação original aprovada pelo Congresso Nacional (de agosto de 2018), o Artigo 20 da LGPD previa o direito de o titular solicitar revisão de decisões automatizadas "por pessoa natural". Contudo, esse dispositivo foi alterado pela Medida Provisória n. 869, de dezembro de 2018, retirando a possibilidade de obter intervenção humana. Essa alteração foi mantida na lei de conversão da MP (Lei n. 13.853/2019), que define a redação da LGPD até o momento atual.

outros países, com base no modelo europeu, a configuração desses direitos no Brasil ainda necessita de regulamentação e de futura elaboração doutrinária e jurisprudencial. A LGPD ressalva os segredos comercial e industrial em diversos dispositivos, de modo que essa ponderação deverá ser feita na regulamentação infralegal e mesmo na análise de casos concretos. A proteção dos segredos de negócio na LGPD pode ser vista como uma forma de fomentar o modelo empresarial baseado em algoritmos, mas deve necessariamente ser sopesada com o direito à explicação sobre decisões automatizadas a fim de observar os princípios éticos de uso da IA em harmonia com os direitos humanos (Dourado & Aith, 2022).

O processo legislativo para a regulamentação da IA no Brasil está em curso. Em razão das rápidas mudanças no campo, já houve diferentes propostas e relevantes mudanças no texto em relação à versão original do Projeto de Lei que está mais maduro neste momento (Projeto de Lei n. 2338/2023)[4]. Vale notar que, nos debates em andamento no contexto brasileiro, a ideia de transparência como princípio para regulação da IA tem estado muito frequentemente vinculada às noções de explicabilidade, inteligibilidade e auditabilidade, mas também aparece em referência às estruturas de governança adotadas no desenvolvimento e implementação dos sistemas de IA. Portanto, o termo transparência vem sendo empregado tanto em sentido específico (associado à explicabilidade) como em sentido abrangente. É necessário abordar ambas as perspectivas para compreender a construção dessa dimensão do marco regulatório.

MECANISMOS E LIMITES PARA EXPLICAÇÃO EM IA NA SAÚDE

A explicabilidade, contemplando tanto sistemas de IA inerentemente interpretáveis como explicações *post hoc* em sistemas opacos, tem sido reconhecida como aspecto essencial de transparência na IA em geral e na IA aplicada à

4 No momento da elaboração do presente artigo, o Projeto de Lei n. 2338, de 2023, iniciado no Senado Federal, está em fase adiantada de debate no Congresso Nacional. Entretanto, ainda não foi apresentada a versão final. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233.

saúde. A saúde é um setor em que a busca pela explicabilidade é considerada particularmente necessária, sobretudo para permitir o uso de sistemas de IA em atividades de assistência clínica (Herzog, 2022; Holzinger *et al.*, 2017). Explicações sobre um sistema de IA podem ser buscadas para justificar decisões, melhorar o controle, aprimorar modelos ou adquirir novos conhecimentos; em todas essas situações, o objetivo do usuário (seja profissional de saúde, paciente ou regulador) é muito relevante para a explicabilidade. Por isso, projetar sistemas para fornecer explicações é bastante complexo, principalmente para obter explicabilidade *post hoc* (Roscher *et al.*, 2020). O campo científico e prático da IA explicável (*eXplainable Artificial Intelligence* [XAI]) está em expansão e, atualmente, há muitas pesquisas desenvolvidas por empresas, órgãos de normalização, organizações sem fins lucrativos e instituições públicas com intuito de criar sistemas de IA que possam explicar suas previsões (Gunning *et al.*, 2019).

Em geral, não é viável nem necessário que uma explicação forneça todo o processo de tomada de decisão de um modelo de aprendizado de máquina. A explicabilidade de um sistema de IA é essencial para situações em que haja alguma falha sobre a qual uma instância específica do sistema precisa ser determinada, o que é especialmente importante quando resultados algorítmicos são usados para fazer recomendações ou tomar decisões que estariam habitualmente sujeitas à discrição humana. Para isso, uma explicação precisa ser apta a responder a pelo menos um dos seguintes pontos (OCDE, 2019): (a) principais fatores de decisão – indicar os fatores importantes para uma previsão feita por IA, preferencialmente ordenados por significância; (b) fatores determinantes de decisão – esclarecer fatores que afetam decisivamente o resultado; e (c) resultados divergentes – esclarecer por que dois casos de aparência semelhante podem apresentar resultados diferentes. Assim, as explicações precisam trazer informações interpretáveis por humanos sobre os fatores utilizados pelos modelos de IA para se chegar a um resultado e o peso relativo de cada fator. Além disso, devem ser capazes de oferecer respostas a questões contrafactuais, para saber se um fator considerado numa decisão algorítmica foi determinante para um resultado específico (Doshi-Velez *et al.*, 2019).

A forma mais direta e acessível de obter explicações é utilizar apenas algoritmos que criem modelos interpretáveis, tais como regressão linear, regressão logística e árvores de decisão – ou seja, usar IA interpretável[5]. De uma forma geral, os modelos de aprendizado de máquina desenvolvidos com algoritmos como esses (e outros também baseados em abordagens estatísticas) podem ser diretamente interpretados por meio de técnicas e cálculos compreensíveis por humanos. Contudo, o campo da IA na saúde está, cada vez mais, impulsionado pelo uso de modelos desenvolvidos por algoritmos mais complexos de aprendizado de máquina e redes neurais (aprendizado profundo), que funcionam como caixas-pretas. Para esses sistemas opacos, só é possível obter explicações por meio de técnicas *post hoc* – ou seja, empregar a IA explicável. Dentre as diversas técnicas atualmente utilizadas nesse campo, as mais usadas são as técnicas Explicações Locais Interpretáveis e Independentes de Modelo (*Local Interpretable Model-agnostic Explanations* [LIME]) e Explicações Aditivas de Shapley (*SHapley Additive exPlanations* [SHAP]), desenvolvidas com o objetivo de explicar previsões individuais dos modelos[6]. Ambas consistem no uso de um modelo substituto interpretável que explica um modelo complexo (opaco), conforme o paradigma da IA explicável. Há ainda métodos específicos para explicar modelos de redes neurais, como a técnica de mapas de saliência (também chamada de "atribuição de *pixels*"), que destaca *pixels* relevantes para uma determinada classificação de imagem, e dezenas de outras técnicas para os diferentes tipos de algoritmos (Molnar, 2022).

Inicialmente considerada quase um consenso, a explicabilidade algorítmica tem sido alvo de crescente controvérsia a

5 Regressão linear é um algoritmo que resolve problemas de regressão, definindo uma função que faz a previsão pretendida como uma soma ponderada dos recursos de entrada. Regressão logística é uma extensão do algoritmo de regressão linear para resolver problemas de classificação, usando uma função que limita as saídas em resultados entre 0 e 1 (probabilidades). Árvores de decisão são algoritmos que dividem os dados várias vezes de acordo com determinados valores de corte nos recursos (entradas), criando diferentes subconjuntos (nós) do conjunto de dados (cada instância em um subconjunto), prevendo resultados a partir do resultado de cada nó. Árvores de decisão podem ser usadas tanto para problemas de regressão como para problemas de classificação (Molnar, 2022).

6 A técnica LIME cria um modelo substituto local gerando um novo conjunto de dados que busca aproximar as previsões correspondentes às do modelo de caixa-preta (Ribeiro et al., 2016). A técnica SHAP tem o mesmo intuito, buscando atingi-lo calculando a contribuição de cada recurso para a previsão, com base nos chamados "valores de Shapley": método da teoria dos jogos de coalizão que calcula a contribuição média de cada integrante em todas as coalizões possíveis (Lundberg & Lee, 2017).

partir da década de 2020, em razão do reconhecimento das limitações da IA explicável. Em primeiro lugar, existe um *trade-off* entre explicabilidade e desempenho de modelos de aprendizado de máquina. Por definição, para que um sistema de IA seja explicável, é necessário reduzir as variáveis da solução a um conjunto pequeno o suficiente para que fique acessível ao entendimento humano. Isso tende a inviabilizar o uso de alguns sistemas em problemas complexos; por isso, a perspectiva de exigir explicação minuciosa pode ser incompatível com o uso de sistemas de IA que buscam alta acurácia preditiva (London, 2019).

Além disso, a explicação de sistemas de IA é limitada pelas reais possibilidades oferecidas pelos mecanismos existentes. As técnicas disponíveis atualmente para explicabilidade, ainda que sejam capazes de oferecer descrições amplas sobre como um sistema de IA funciona em sentido geral, são muito superficiais ou não confiáveis para decisões individuais. Outro aspecto a ser considerado é a constatação de que usuários tendem a confiar excessivamente em explicações dadas pelas ferramentas de IA explicável, muitas vezes sem observar que elas não têm garantias de desempenho, um ponto particularmente preocupante na área da saúde diante da possível adoção em escala crescente de soluções de IA em ambientes clínicos. Como as explicações são apenas aproximações do processo de decisão do modelo opaco, cria-se uma fonte adicional de erro, já que tanto o modelo original como o modelo explicativo podem estar equivocados (Ghassemi *et al.*, 2021).

PERSPECTIVAS REGULATÓRIAS PARA TRANSPARÊNCIA E EXPLICABILIDADE EM IA NA SAÚDE

Diante do cenário apresentado, foram identificados elementos a serem contemplados para a elaboração de um marco regulatório para a IA na saúde que incorpore mecanismos de transparência de forma efetiva. Propõe-se uma abordagem que considere os limites atualmente presentes para explicabilidade dos sistemas de IA na saúde e reforce a concepção de transparência em sentido amplo no que se refere ao desenvolvimento e implementação de ferramentas de aprendizado de máquina.

O *trade-off* entre explicabilidade e desempenho não necessariamente existe em qualquer situação em que modelos de

aprendizado de máquina são empregados na prática. Em sistemas de IA desenvolvidos para resolver problemas de dados estruturados com uma boa representação de características significativas (como no contexto da saúde), muitas vezes têm sido observadas diferenças mínimas de desempenho entre os algoritmos complexos, como redes neurais, e os classificadores mais simples de IA interpretável, após adequado pré-processamento dos dados (Rudin, 2019). Por isso, é possível argumentar que o uso de algoritmos interpretáveis deve ser sempre priorizado, principalmente no desenvolvimento de modelos para uso em situações nas quais a explicação seja essencial, como pode acontecer em condições sensíveis ou críticas na área da saúde.

Uma possível abordagem para a regulação da IA na saúde poderia ser exigir que desenvolvedores de modelos caixa-preta relatassem o desempenho de modelos interpretáveis testados e validados para os mesmos usos pretendidos. Isso permitiria uma avaliação direta da existência ou não do *trade-off* entre explicabilidade e desempenho, provavelmente incentivando o uso de algoritmos de IA interpretável sempre que possível. Uma proposta mais forte seria não permitir o uso de sistemas opacos para usos de alto risco se existir um sistema interpretável com o mesmo nível de desempenho.

Além disso, do ponto de vista de explicabilidade, a atual falta de transparência em IA na saúde deve persistir e, inclusive, se intensificar ao longo do tempo. A evolução do campo tem acontecido na direção de sistemas cada vez mais complexos – consequentemente mais opacos –, como os grandes modelos de linguagem e os modelos de fundação (Moor *et al.*, 2023). Isso considerado, o caminho para a construção de uma regulação eficiente para a IA na saúde deve partir da admissão de que sistemas opacos serão utilizados, em vez de pressupor a busca por explicações humanamente compreensíveis para decisões algorítmicas individuais. O ponto central deve ser a definição de regras e condições para o uso de sistemas do tipo caixa-preta, em consonância com preceitos éticos e os Direitos Humanos. Tudo indica que haverá situações em que o resultado de modelos opacos será suficientemente ótimo para justificar decisões neles baseadas. Se houver uma estrutura regulatória robusta para assegurar segurança e eficácia desses sistemas,

é bem possível que sejam usados com confiança, apesar das limitações de explicabilidade. Em certa medida, opacidade é uma característica usual na atividade clínica: a medicina tradicionalmente adota práticas que envolvem mecanismos não totalmente compreendidos, mas que continuam sendo amplamente usados, devido aos efeitos comprovados – como o uso de muitos medicamentos. Logo, algo semelhante pode ocorrer com a IA na saúde.

Entretanto, é preciso reiterar que a documentação rigorosa e completa constitui o pilar fundamental para garantir a transparência da IA na área da saúde. O registro meticuloso de informações claras e detalhadas acerca dos métodos, recursos e decisões tomadas ao longo do ciclo de vida integral dos sistemas de IA, além de ser fundamental para assegurar a confiabilidade nessas ferramentas, deve ser um requisito regulatório essencial. Os órgãos reguladores precisam ter acesso à documentação adequada que abranja desde a concepção, o desenvolvimento, o treinamento e a validação dos modelos até a implementação e o período subsequente à sua implementação (OMS, 2023). Algumas propostas existentes podem servir de referência para definir quais informações devem ser fornecidas e de que forma essas informações precisam estar organizadas[7]. É necessário exigir que os sistemas de IA incluam informações sobre as populações usadas nos dados de treinamento (fontes de dados e seleção de corte) e a demografia desses dados, de forma a permitir a comparação com a população na qual os modelos serão implementados. Além disso, requisitos devem incluir informações detalhadas sobre arquitetura e

7 Uma das iniciativas mais relevantes atualmente é o Relatório Transparente de um Modelo de Previsão Multivariável para Prognóstico ou Diagnóstico Individual para IA (Transparent Reporting of a multivariable prediction model of Individual Prognosis Or Diagnosis for AI [TRIPOD-AI]), uma extensão específica para IA da declaração Relatório Transparente de um Modelo de Previsão Multivariável para Prognóstico ou Diagnóstico Individual (Transparent Reporting of a multivariable prediction model of Individual Prognosis Or Diagnosis [TRIPOD]), protocolo de notificação padronizado que inclui uma lista de verificação de 22 itens, projetada para melhorar a qualidade dos relatórios de estudos que desenvolvem, validam ou atualizam modelos preditivos com finalidade clínica (diagnóstica ou prognóstica). Está associada à PROBAST-AI, desenvolvida como extensão da Ferramenta de Avaliação do Risco de Viés do Modelo de Previsão (Prediction model Risk Of Bias Assessment Tool [PROBAST]), ferramenta que avalia risco de viés e aplicabilidade de modelos preditivos a partir de 20 questões (em quatro domínios: participantes, preditores, resultado e análise). Essas ferramentas têm sido concebidas para orientar pesquisadores e revisores a avaliarem a qualidade de estudos e interpretar achados cientificamente relevantes (Collins et al., 2015, 2021, 2024; Wolff et al., 2019) e podem ser também muito úteis para auxiliar reguladores.

desenvolvimento dos modelos, a fim de propiciar a interpretação da intenção de uso em comparação com sistemas de IA semelhantes. Tais medidas podem dar aos reguladores e aos usuários um melhor entendimento sobre o funcionamento dos sistemas de IA, considerando a transparência como abordagem para identificar melhores práticas de aprendizado de máquina (Hernandez-Boussard *et al.*, 2020).

CONSIDERAÇÕES FINAIS

A IA tem se consolidado como uma das principais tecnologias na área da saúde; por isso, estabelecer um marco regulatório da IA na saúde assume importância central na sociedade contemporânea. Dentre os fundamentos para a regulação da IA na saúde provenientes da ética em IA, a transparência é reconhecida como dimensão essencial.

A compreensão ampla do conceito de transparência implica a necessidade de tornar acessíveis todas as informações que fundamentam as decisões originadas de modelos de aprendizado de máquina. Refere-se à clareza sobre os contextos técnico e institucional nos quais os sistemas de IA são concebidos, implementados e administrados, que deve ser expressa por meio de documentação padronizada. Portanto, um mecanismo central na estratégia regulatória deve ser definir critérios para a padronização da documentação detalhada e completa de todas as etapas de desenvolvimento e implementação dos sistemas de IA considerando listas existentes como referências.

A transparência também diz respeito a interpretabilidade e a explicabilidade dos sistemas de IA. A identificação de mecanismos e limites para a explicação em IA na saúde é crucial para definir a abrangência do direito à explicação e os requisitos regulatórios a serem estabelecidos nessa dimensão. Os marcos regulatórios devem diferenciar sistemas de IA interpretável de sistemas de IA explicável, reconhecendo que os atuais mecanismos de explicação, embora sejam capazes de fornecer fatores determinantes de decisão e esclarecer resultados divergentes em modelos interpretáveis (IA interpretável), representam apenas aproximações em modelos opacos (IA explicável). Por isso, os requisitos regulatórios de explicabilidade podem exigir explicações sobre decisões específicas em modelos de IA interpretável, mas devem se restringir a esclarecimentos sobre

o funcionamento geral em sistemas de IA explicável. Deve ser considerada a exigência de comparação de desempenho com modelos interpretáveis como um requisito regulatório para sistemas baseados em modelos de IA opacos. Nesse sentido, é necessário aprender a lidar com a opacidade, definindo regras para uso de sistemas de IA caixa-preta que se mostrem benéficos, a fim de garantir que tenham segurança e eficácia devidamente comprovadas.

REFERÊNCIAS

Acosta, J. N., Falcone, G. J., Rajpurkar, P., & Topol, E. J. (2022). Multimodal biomedical AI. *Nature Medicine, 28*(9). https://doi.org/10.1038/s41591-022-01981-2

Aith, F., & Dallari, A. B. (Orgs.). (2022). *LGPD na saúde digital*. Thomson Reuters Brasil.

Amann, J., Vetter, D., Blomberg, S. N., Christensen, H. C., Coffee, M., Gerke, S., Gilbert, T. K., Hagendorff, T., Holm, S., Livne, M., Spezzatti, A., Strümke, I., Zicari, R. V., & Madai, V. I. (2022). To explain or not to explain? — Artificial Intelligence explainability in clinical decision support systems. *PLOS Digital Health, 1*(2), 1-18. https://doi.org/10.1371/journal.pdig.0000016

Association for Computing Machinery US Public Policy Council. (2017). *Statement on algorithmic transparency and accountability*. https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

Babic, B., Gerke, S., Evgeniou, T., & Cohen, I. G. (2021). Beware explanations from AI in health care. *Science, 373*(6552), 284-286. https://doi.org/10.1126/science.abg1834

Bates, D. W. (2023). How to regulate evolving AI health algorithms. *Nature Medicine, 29*(26). https://doi.org/10.1038/s41591-022-02165-8

Beauchamp, T. L., & Childress, J. F. (1979). *Principles of biomedical ethics*. Oxford University Press.

Brownstein, J. S., Rader, B., Astley, C. M., & Tian, H. (2023). Advances in Artificial Intelligence for infectious-disease surveillance. *New England Journal of Medicine, 388*(17), 1597-1607. https://doi.org/10.1056/NEJMra2119215

Burrell, J. (2016). How the machine "thinks": Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1), 1-12. https://doi.org/10.1177/2053951715622512

Bygrave, L. A. (2020). Automated individual decision-making, including profiling. In Christopher, K., Bygrave, L. A., Docksey, C., & Drechsler, L. (Eds.), *The EU General Data Protection Regulation (GDPR): A commentary* (Article 22, pp. 522-542). Oxford University Press. https://doi.org/10.1093/oso/9780198826491.003.0055

Collins, G. S., Dhiman, P., Navarro, C. L. A., Ma, J., Hooft, L., Reitsma, J. B., Logullo, P., Beam, A. L., Peng, L., Calster, B. V., van Smeden, M., Riley, R. D., & Moons, K. G. (2021). Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on Artificial Intelligence. *BMJ Open, 11*(7). https://doi.org/10.1136/bmjopen- 2020-048008

Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Calster, B. V., Ghassemi, M., Liu, X., Reitsma, J. B., van Smeden, M., Boulesteix, A-L., Camaradou, J. C., Celi, L. A., Denaxas, S., Denniston, A. K., Glocker, B., Golub, R. M., Harvey, H., Heinze, G., Hoffman, M. M., Kengne, A. P., Lam, E., Lee, N., Loder, E. W., Maier-Hein, L., Mateen, B. A., McMradden, M. D., Oakden-Rayner, L., Ordish, J., Parnell, R., Rose, S., Singh, K., Wynants, L., & Logullo, P. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ, 385*. https://doi.org/10.1136/bmj-2023-078378

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *BMJ, 350*. https://doi.org/10.1136/bmj.g7594

Comissão Europeia. (2019). *Orientações éticas para uma IA de confiança*. https://digital-strategy.ec.europa.eu/pt/library/ethics-guidelines-trustworthy-ai

Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S., O'Brien, D., Scott, K., Schieber, S., Waldo, J., Weinberger, D., Weller, A., & Wood, A. (2019). Accountability of AI under the law: The role of explanation. *arXiv*. http://arxiv.org/abs/1711.01134

Dourado, D. A., & Aith, F. M. A. (2022). A regulação da Inteligência Artificial na saúde no Brasil começa com a Lei Geral de Proteção de Dados Pessoais. *Revista de Saúde Pública, 56*(80). https://doi.org/10.11606/s1518-8787.2022056004461

Dubber, M. D., Pasquale, F., & Das, S. (Orgs.). (2020). *The Oxford handbook of ethics of AI*. Oxford University Press.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People – An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*, 689-707. https://doi.org/10.1007/s11023-018-9482-5

G7 Hiroshima Summit. (2023). *G7 Hiroshima Leaders' Communiqué*. https://www.g7hiroshima.go.jp/documents/pdf/Leaders_Communique_01_en.pdf

Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2021). The false hope of current approaches to explainable Artificial Intelligence in health care. *The Lancet Digital Health, 3*(11), e745-e750. https://doi.org/10.1016/S2589-7500(21)00208-9

Gottlieb, S., & Silvis, L. (2023). Regulators face novel challenges as Artificial Intelligence tools enter medical practice. *JAMA Health Forum*, *4*(6). https://doi.org/10.1001/jamahealthforum.2023.2300

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G-Z. (2019). XAI – Explainable Artificial Intelligence. *Science Robotics*, *4*(37). https://doi.org/10.1126/scirobotics.aay7120

Hernandez-Boussard, T., Bozkurt, S., Ioannidis, J. P. A., & Shah, N. H. (2020). MINIMAR (MINimum Information for Medical AI Reporting): Developing reporting standards for Artificial Intelligence in health care. *Journal of the American Medical Informatics Association*, *27*(12), 2011-2015. https://doi.org/10.1093/jamia/ocaa088

Herzog, C. (2022). On the ethical and epistemological utility of explicable AI in medicine. *Philosophy & Technology*, *35*(50). https://doi.org/10.1007/s13347-022-00546-y

Hinton, G. E. (2018). Deep learning-A technology with the potential to transform health care. *JAMA*, *320*(11), 1101-1102. https://doi.org/10.1001/jama.2018.11100

Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). What do we need to build explainable AI systems for the medical domain? *arXiv*. https://doi.org/10.48550/arXiv.1712.09923

International Bioethics Committee. (2017). Report of the IBC on Big Data and health. *United Nations Educational, Cultural and Scientific Organization*. https://unesdoc.unesco.org/ark:/48223/pf0000248724

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*, 389-399. https://doi.org/10.1038/s42256-019-0088-2

Kaminski, M. E. (2019). The right to explanation, explained. *Berkeley Technology Law Journal*, *34*(1), 189-218. https://doi.org/10.15779/Z38TD9N83H

Khorsandi, S. E., Hardgrave, H. J., Osborn, T., Klutts, G., Nigh, J., Spencer-Cole, R. T., Kakos, C. D., Anastasiou, I., Mavros, M. N., & Giorgakis, E. (2021). Artificial Intelligence in liver transplantation. *Transplantation Proceedings, 53*(10), 2939-2944. https://doi.org/10.1016/j.transproceed.2021.09.045

Lei Geral de Proteção de Dados Pessoais (LGPD). (2018). Lei n. 13.709, de 14 de agosto de 2018. Dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural. https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm

Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue, 16*(3), 31-57. https://doi.org/10.1145/3236386.3241340

London, A. J. (2019). Artificial Intelligence and black-box medical decisions: Accuracy versus explainability. *Hastings Center Report, 49*(1), 15-21. https://doi.org/10.1002/hast.973

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Proceedings of the International Conference on Neural Information Processing Systems 31*, 4768-4777. https://dl.acm.org/doi/10.5555/3295222.3295230

Matheny, M. E., Israni, S. T., Ahmed, M., & Whicher, D. (Orgs.). (2022). *Artificial Intelligence in health care: The hope, the hype, the promise, the peril*. National Academy of Medicine.

Miller, T. (2019). Explanation in Artificial Intelligence: Insights from the social sciences. *Artificial Intelligence, 267*, 1-38. https://doi.org/10.1016/j.artint.2018.07.007

Mittelstadt, B. (2022). Interpretability and transparency in Artificial Intelligence. In C. Véliz (Org.), *The Oxford handbook of digital ethics* (pp. 378-410). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198857815.013.20

Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (2ª ed.). https://christophm.github.io/interpretable-ml-book/

Monteiro, R., Machado, C. V., & Silva, L. (2021). The right to explanation in Brazilian Data Protection Law. *International Journal of Digital and Data Law, 7*(1), 119-136. https://ojs.imodev.org/?journal=RIDDN&page=article&op=view&path[]=406

Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., & Rajpurkar, P. (2023). Foundation models for generalist medical Artificial Intelligence. *Nature, 616*, 259-265. https://doi.org/10.1038/s41586-023-05881-4

Nature, E. (2023). AI's potential to accelerate drug discovery needs a reality check. *Nature, 622*, 217. https://doi.org/10.1038/d41586-023-03172-6

Organisation for Economic Co-operation and Development. (2019). *Artificial Intelligence in society*. https://www.oecd-ilibrary.org/science-and-technology/artificial-intelligence-in-society_eedfee77-en

Organização Mundial da Saúde. (2021). *Ethics and governance of Artificial Intelligence for health: WHO guidance*. https://apps.who.int/iris/handle/10665/341996

Organização Mundial da Saúde. (2023). *Regulatory considerations on Artificial Intelligence for health*. https://iris.who.int/handle/10665/373421

Organização Mundial da Saúde. (2024). *Ethics and governance of Artificial Intelligence for health: Guidance on large multi-modal models*. https://iris.who.int/handle/10665/375579

Projeto de Lei n. 2338/2023. (2023). Dispõe sobre o uso da Inteligência Artificial. https://www25.senado.leg.br/web/atividade/materias/-/materia/157233

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine, 380*(14), 1347-1358. https://doi.org/10.1056/NEJMra1814259

Regulamento Geral sobre a Proteção de Dados (GDPR). (2016). Regulamento (UE) n. 679 do Parlamento Europeu e do Conselho, de 23 de abril de 2016. Relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE. https://eur-lex.europa.eu/eli/reg/2016/679/oj

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Nova York, NY, Estados Unidos da América, *22*, 1135-1144. https://dl.acm.org/doi/10.1145/2939672.2939778

Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. (2020). Explainable machine learning for scientific insights and discoveries. *IEEE Access, 8*, 42200-42216. https://doi.org/10.1109/ACCESS.2020.2976199

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5). https://doi.org/10.1038/s42256-019-0048-x

Sahni, N., & Carrus, B. (2023). Artificial Intelligence in U.S. health care delivery. *New England Journal of Medicine, 389*(4), 348-358. https://doi.org/10.1056/NEJMra2204673

Selbst, A. D., & Powles, J. (2017). Meaningful information and the right to explanation. *International Data Privacy Law, 7*(4), 233-242. https://doi.org/10.1093/idpl/ipx022

Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020). An overview of clinical decision support systems: Benefits, risks, and strategies for success. *Npj Digital Medicine, 3*(1). https://doi.org/10.1038/s41746-020-0221-y

Topol, E. J. (2019). High-performance medicine: The convergence of human and Artificial Intelligence. *Nature Medicine, 25*(1), 44-56. https://doi.org/10.1038/s41591-018-0300-7

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law, 7*(2), 76-99. https://doi.org/10.1093/idpl/ipx005

Watson, D. S., Krutzinna, J., Bruce, I. N., Griffiths, C. E., McInnes, I. B., Barnes, M. R., & Floridi, L. (2019). Clinical applications of machine learning algorithms: Beyond the black box. *BMJ, 364*, 1-10. https://doi.org/10.1136/bmj.l886

Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., & Mallett, S. (2019). PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine, 170*(1), 51-58. https://doi.org/10.7326/M18-1376

Yu, K.-H., Beam, A. L., & Kohane, I. S. (2018). Artificial Intelligence in healthcare. *Nature Biomedical Engineering, 2*, 719-731. https://doi.org/10.1038/s41551-018-0305-z

Parte 2

PESQUISA QUALITATIVA

Notas metodológicas

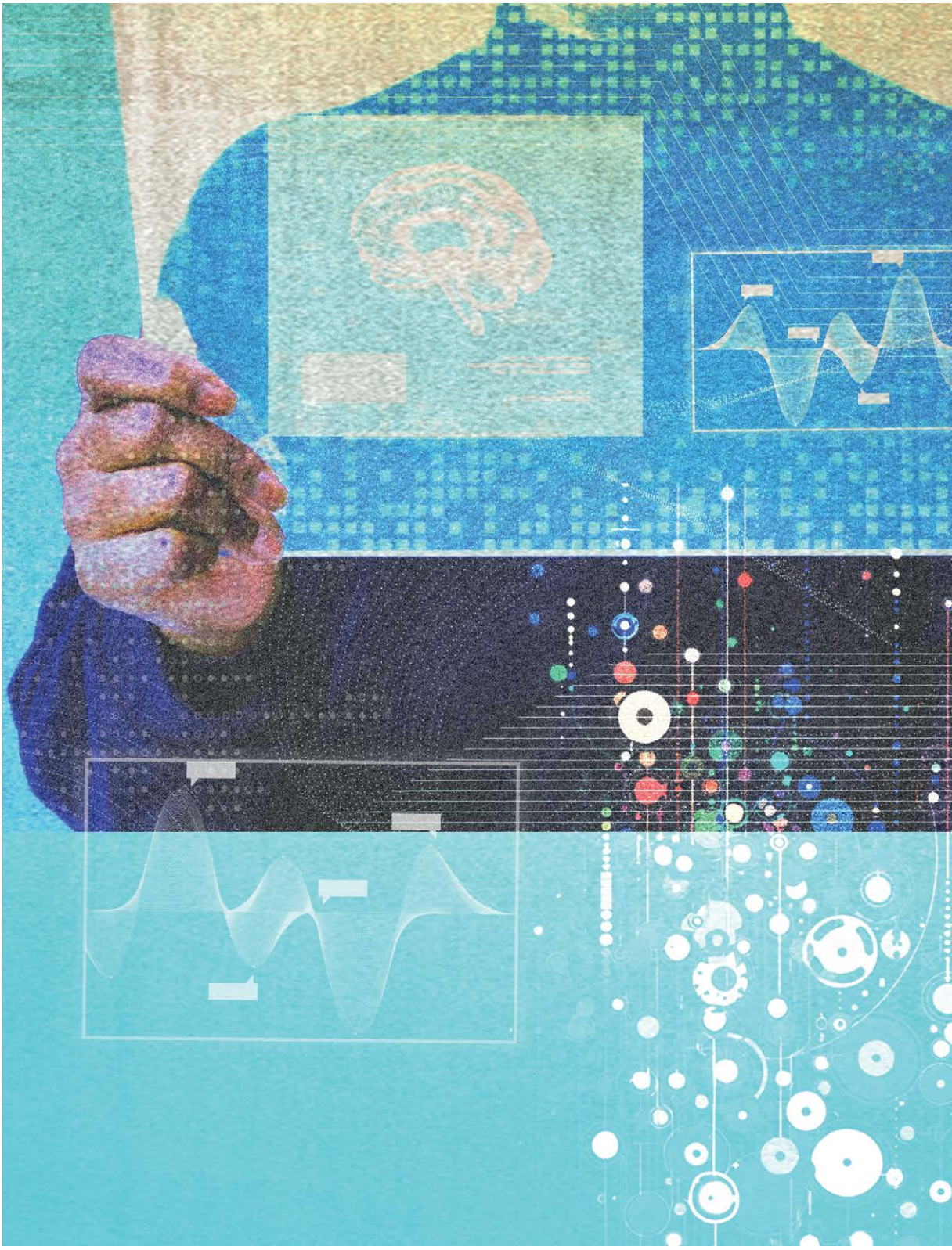
Graziela Castello[1], Monise Picanço[2], Priscila Vieira[3] e Rodrigo Brandão[4]

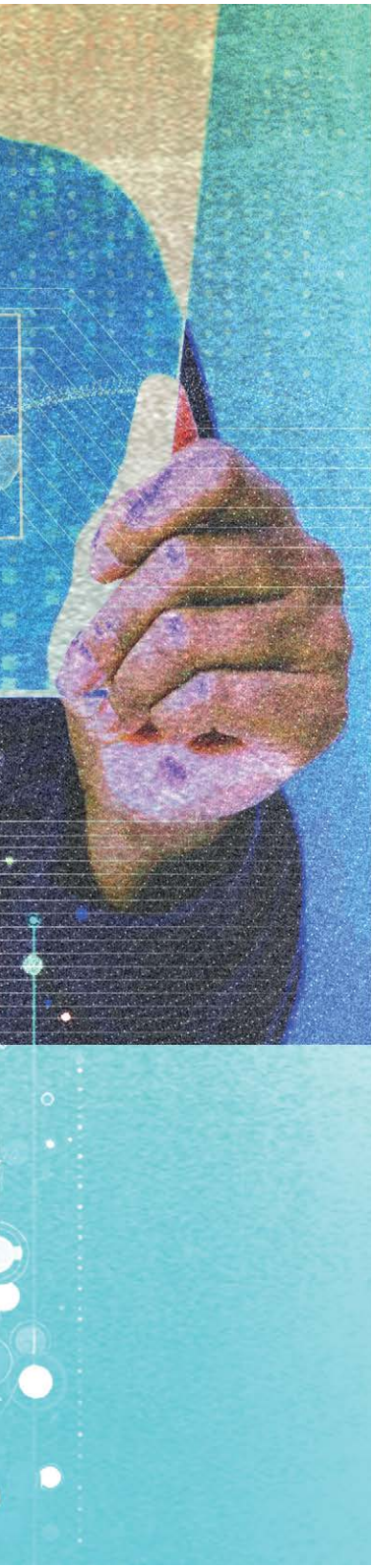
1 Cientista Social, é coordenadora de Métodos Qualitativos e Estudos Setoriais no Centro Regional para o Desenvolvimento da Sociedade da Informação (Cetic.br), departamento do Núcleo de Informação e Coordenação do Ponto BR (NIC.br).

2 Doutora e mestre em Sociologia pela Universidade de São Paulo (USP). Pesquisadora associada e coordenadora de projetos no Núcleo de Desenvolvimento do Centro Brasileiro de Análise e Planejamento (CEBRAP). Professora de metodologia e desenho de pesquisa na Faculdade de Informática e Administração Paulista (FIAP) e cebrap.lab.

3 Socióloga com mestrado e doutorado em Sociologia pela USP. Pesquisadora e coordenadora de projetos no Núcleo de Desenvolvimento do CEBRAP.

4 Doutorando em Sociologia pela Universidade de São Paulo (USP), mestre em Ciência Política pela USP, com graduação em Ciências Sociais pela USP. É pesquisador na Coordenação de Métodos Qualitativos e Estudos Setoriais do Cetic.br|NIC.br.





Este capítulo apresenta os procedimentos metodológicos adotados no estudo "Inteligência Artificial na saúde: diagnóstico qualitativo sobre o cenário brasileiro", cuja análise dos resultados encontra-se no próximo capítulo. O estudo consiste em uma investigação qualitativa exploratória com atores-chave no setor de saúde no Brasil, considerados na fronteira do conhecimento sobre o tema, de diferentes campos de atuação. A investigação visou mapear os debates atuais na agenda pública, coletar percepções sobre as oportunidades, os desafios e os riscos para o desenvolvimento e a adoção de ferramentas e soluções baseadas em IA na área da saúde, bem como identificar iniciativas e práticas em andamento no contexto brasileiro. Este capítulo também apresenta os objetivos gerais do estudo, o desenho da pesquisa, os instrumentos de coleta e a metodologia adotada para a análise dos resultados.

OBJETIVOS GERAIS DO ESTUDO E DESENHO PLANEJADO

A Inteligência Artificial (IA), por meio da utilização de dados de saúde complexos e múltiplas fontes – por exemplo: prontuários eletrônicos, estudos de imagem, dados genômicos e fisiológicos –, tem amplo potencial para enfrentar os crescentes desafios do setor de saúde, como o aumento contínuo dos custos, a falta de profissionais especialistas, as mudanças epidemiológicas e demográficas em curso, o envelhecimento populacional, entre outros. Nesse sentido, o uso de ferramentas baseadas em IA configura uma aposta promissora para alavancar e acelerar a produção de conhecimento científico, bem como desenvolver e implementar políticas públicas de saúde em larga escala (Academia Brasileira de Ciências [ABC], 2023).

Diante desse cenário, o estudo teve por objetivo mapear o estágio atual de desenvolvimento da IA no setor de saúde no Brasil (onde estamos e como estamos), a partir de um diagnóstico exploratório cujos objetivos específicos foram levantar informações sobre: (a) as oportunidades para o país com o advento dessa tecnologia na área de saúde; (b) os desafios colocados para o aproveitamento dessas oportunidades; (c) os riscos potenciais a serem gerenciados; (d) as agendas de pesquisa, políticas e inciativas em andamento no país; e (e) os

impactos percebidos a partir do uso das ferramentas de IA na prática clínica.

Pelo caráter exploratório, com o objetivo de problematizar potencialidades, riscos e perspectivas para o Brasil com o advento da IA na área de saúde, o estudo, de natureza qualitativa, foi implementado a partir da realização de entrevistas em profundidade (EP) realizadas com atores estratégicos do país na área de IA e saúde, considerados, neste estudo, informantes privilegiados e formadores de opinião sobre o tema (*stakeholders*).

No contexto de pesquisa qualitativa, "informante" refere-se a uma pessoa selecionada para fornecer informações detalhadas e *insights* sobre o tema de IA na saúde o qual, ao longo da análise dos dados coletados, será referenciado como "entrevistado". A escolha de informantes privilegiados (na fronteira do debate sobre o tema atualmente) teve como premissa explorar as experiências existentes no país, com o objetivo de investigar de maneira substantiva as oportunidades, os desafios e as barreiras a partir de situações concretas – o estágio atual de desenvolvimento da IA no setor de saúde no Brasil. Nesse sentido, o estudo qualitativo não levantou informações com informantes do setor de saúde de localidades e instituições ainda sem atividades associadas a IA. O estudo qualitativo, portanto, é um retrato do cenário brasileiro de quem atua com IA na saúde e tem, nesse sentido, o viés proposital de investigar as oportunidades e os desafios a partir de percepções e opiniões de indivíduos com forte relação com essa agenda. Como escolha metodológica, esse recorte permite aprofundar os temas investigados a partir de situações e experiências vivenciadas de atores sensibilizados sobre o tema. Com essa estratégia, é possível, além de mapear oportunidades e desafios, qualificar os diferentes graus de prioridade entre eles para o desenvolvimento oportuno da IA no setor de saúde do país, considerando a expertise daqueles imbuídos no tema.

Diante do desenho proposto, os critérios para seleção dos entrevistados se tornaram parte central do estudo e levaram em consideração as seguintes características: (a) diversificação dos segmentos que representam na área de saúde; (b) exercício de cargo de liderança e/ou gestão na organização em que atuam ou ser especialista; (c) forte incidência no debate

público, considerando produção cientifica, participação em eventos públicos sobre o tema e exposição na mídia associada a temática.

Considerando esses critérios, uma etapa necessária foi o mapeamento dos potenciais entrevistados, realizada a partir de levantamento de conteúdos públicos disponíveis, produção científica, presença em eventos e exposição na mídia sobre o tema, busca ativa em fóruns e *sites* institucionais, além de indicações de atores reconhecidos publicamente. O mapeamento resultou na identificação de uma lista inicial de 93 nomes de potenciais entrevistados de diversos segmentos da área de saúde com atuação em IA. A segunda etapa para seleção considerou, então, a diversidade de segmentos representados no estudo, com a alocação de um número mínimo de cinco entrevistas por segmento.

Foram definidos cinco segmentos para seleção dos entrevistados, a partir da natureza das principais atividades desenvolvidas pelas organizações dos possíveis entrevistados na área de saúde, tais como: pesquisa, prática clínica, desenvolvimento de soluções e promoção de políticas públicas, e convergentes com a definição presente na Estratégia de Saúde Digital para o Brasil 2020-2028 (ESD) sobre os atores da saúde digital (Ministério da Saúde [MS], 2020)[5]. Em cada segmento, a seleção dos respondentes também priorizou a diversidade de organizações; nesse sentido, foi adotada, como critério, a realização de apenas uma entrevista por organização. A partir da lista inicial de 93 nomes de potenciais entrevistados, chegou-se a um total de 28 especialistas conforme os critérios definidos. Dessa forma, no total foram realizadas 28 entrevistas, distribuídas na Tabela 1.

Vale uma ressalva importante: apenas no caso dos profissionais no ponto de cuidado, a seleção dos entrevistados foi feita a partir de indicações solicitadas aos entrevistados especialistas (*stakeholders*), uma vez que as ferramentas de IA em uso no setor de saúde no Brasil ainda são pouco difundidas e não há informações públicas suficientes sobre o tema. Nesse sentido, essa subamostra foi construída por conveniência.

5 É importante observar que "associações de pacientes" e "cidadãos", dois grupos de atores da saúde digital, definidos pela ESD, não foram objeto de investigação nesse estudo.

TABELA 1 - DESCRIÇÃO DOS SEGMENTOS INVESTIGADOS E TOTAL DE ENTREVISTAS

Segmentos	Descrição dos Segmentos	Equivalência na ESD[6] (atores da saúde digital)	Total de entrevistas
Academia	Universidades, centros de pesquisas, think tanks, centros de referência	Sociedades técnico-científicas; universidades e centros de formação	6
Equipamentos de saúde (públicos e privados)	Hospitais, clínicas, laboratórios	Prestadores de serviços do sistema de Saúde	6
Poder público	Ministérios, secretarias de governo e agências	Ministério da Saúde; agências; secretarias estaduais e municipais de saúde	5
Mercado	Empresas de tecnologia, *startups*, planos de saúde	Indústria e setor de tecnologia; fontes pagadoras do sistema de saúde	6
Profissionais de saúde (no ponto de cuidado)	Médicos e enfermeiros que usam ferramentas de IA no ponto de cuidado	–	5
Total			**28**

FONTE: ELABORAÇÃO PRÓPRIA.

O trabalho de campo foi realizado entre os meses de agosto e dezembro de 2023. As entrevistas tiveram duração média de uma hora e foram realizadas remotamente, a partir de videochamadas previamente agendadas com os atores recrutados. Todas as entrevistas foram gravadas e transcritas para posterior codificação do material.

Todos os participantes recrutados foram informados a respeito da natureza e objetivos do estudo e do uso e tratamento dos dados coletados. Antes da realização das entrevistas, os informantes assinaram um Termo de Consentimento com o detalhamento dessas informações. A pesquisa seguiu as diretrizes da Lei Geral de Proteção de Dados (LGPD) (Lei n. 13.709/2018) e os protocolos de ética na pesquisa, a fim de possibilitar desistência a qualquer momento e assegurar o anonimato e o sigilo das informações. A adesão foi voluntária, sem pagamento ou qualquer outro tipo de incentivo material aos entrevistados por sua participação.

6 Para mais informações sobre os atores de saúde digital, definidos pela ESD, ver MS (2020).

INSTRUMENTOS DE COLETA

Para a condução das entrevistas com os *stakeholders* dos grupos academia, mercado, poder público e equipamentos de saúde, foi utilizado o mesmo roteiro semiestruturado, dividido em dois módulos. O primeiro módulo estimulou questões sobre o cenário atual da IA aplicada à saúde no país e buscou coletar percepções sobre estágio atual, oportunidades, desafios, riscos e prioridades. Além disso, propôs um levantamento de informações sobre as iniciativas de IA na saúde em andamento nas instituições a que estavam vinculados, para entender a natureza das práticas, o estágio de desenvolvimento, as potencialidades e as dificuldades enfrentadas em sua concepção e implementação. O segundo módulo contemplou questões oriundas das temáticas das sete prioridades elencadas pela ESD, consideradas prioridades essenciais para saúde digital (MS, 2020), entretanto direcionadas especificamente à IA. As perguntas do segundo módulo foram aplicadas apenas quando o tema não apareceu espontaneamente durante a realização do primeiro módulo do roteiro (Anexo 1).

Já os informantes do grupo "profissionais nos pontos de cuidado" foram entrevistados a partir de outro roteiro, também semiestruturado, mas com diferenças substantivas, cujo foco estava na compreensão da incorporação das ferramentas de IA no dia a dia de trabalho. Assim, o roteiro estimulava questões sobre aplicação da IA na prática clínica, processo de adoção, treinamento, confiança, interpretabilidade, relação com paciente, desafios e riscos, além de perspectivas de futuro. Como a pesquisa com esse grupo tinha outro objetivo de análise e foi conduzida a partir de outro instrumento de coleta, os dados provenientes dessas entrevistas foram analisados separadamente. O roteiro completo aplicado aos profissionais nos pontos de cuidado encontra-se no Anexo 2 deste capítulo.

METODOLOGIA ADOTADA PARA ANÁLISE DOS RESULTADOS

Após realização das entrevistas e suas respectivas transcrições, foi conduzida a etapa analítica por meio da metodologia de análise de conteúdo para sistematizar os dados das EP por meio do processo de codificação. Esse processo consiste em

categorizar sistematicamente dados qualitativos para identificar padrões e temas relevantes a partir das transcrições das entrevistas.

A análise de conteúdo é uma metodologia de análise qualitativa que busca desenvolver uma descrição sistemática e objetiva de mensagens manifestas[7], ou seja, sobre o conteúdo que está apresentado. Seu propósito é realizar uma análise crítica dos materiais sem perder de vista a necessidade de desenvolver uma análise com inferências válidas e procedimentos replicáveis. Trata-se de um método originário dos anos 1970 que utiliza uma lógica de análise frequentemente dedutiva (verifica e prova a partir de testes de hipóteses) ou abdutiva, (investiga causas e/ou mecanismos para explicar a ocorrência de certo fenômeno) (Krippendorff, 1980; Neuendorf, 2002).

Ao usar a codificação como procedimento de sistematização e análise, essa metodologia segue os seguintes passos para sua operacionalização: pré-análise dos materiais, desenvolvimento de plano de codificação, codificação de uma amostra dos dados para validação, revisão dos códigos, codificação do material e descrição analítica dos principais resultados (Bardin, 1977). É especialmente importante para essa análise a constituição de um plano de codificação robusto *a priori*, fixado antes da análise sistemática de todo o material.

Para a pré-análise dos materiais deste estudo, foi realizada a leitura transversal de todo o material, com uma sistematização inicial dos principais destaques. Em seguida, foi feita uma leitura sistemática de uma amostra de entrevistas, selecionando-se uma entrevista de cada um dos segmentos definidos no estudo. Esse passo foi crucial para o desenvolvimento do plano de codificação, envolvendo a organização e a categorização das informações coletadas nas entrevistas, construído a partir de uma discussão coletiva levando em conta os objetivos da pesquisa.

O plano de codificação desenvolvido tem duas camadas de sistematização de informação. A primeira é baseada no roteiro de entrevista e suas questões, a fim de localizar os estímulos utilizados durante as entrevistas. A segunda camada de

7 Essa metodologia diferencia-se da análise de discurso, por exemplo, que se focaliza frequentemente nos sentidos do enunciado e no que ele elucida de seu contexto (Orlandi, 1999).

informação é formada por códigos analíticos, ou seja, por temas e subtemas identificados na pré-análise dos materiais.

Após a construção do plano de codificação preliminar, foi realizado um teste a partir da codificação de um número reduzido de entrevistas. A revisão do plano de codificação após o teste incorporou dois novos códigos extras, acréscimo que validou e mostrou a robustez do plano inicial aplicado ao restante do material.

A codificação foi realizada a partir do *software* Atlas.ti[8], que permitiu a organização e a análise dos dados qualitativos, além da segmentação dos textos das entrevistas em trechos específicos, facilitando a atribuição de códigos a esses segmentos de acordo com as dimensões e os temas identificados. Ademais, o Atlas.ti oferece diversas funcionalidades que aprimoram o processo de codificação e análise, como a criação e o gerenciamento de códigos, visualização das relações entre diferentes temas e a composição de relatórios que sintetizam as informações codificadas. Logo, a utilização deste *software* proporcionou uma sistematização robusta dos dados coletados, permitindo que a análise dos resultados fosse realizada de maneira organizada e consistente.

Como resultado, foi gerado um banco de dados detalhado com identificação dos trechos de entrevistas correspondentes a cada uma das dimensões identificadas, bem como a suas características de perfil. Essa sistematização do material por meio de codificação de temas subsidia a análise dos resultados apresentada no próximo capítulo desta publicação.

8 Saiba mais: https://atlasti.com/

REFERÊNCIAS

Academia Brasileira de Ciências. (2023). *Recomendações para o avanço da Inteligência Artificial no Brasil: GT-IA da Academia Brasileira de Ciências*. https://www.abc.org.br/wp-content/uploads/2023/11/recomendacoes-para-o-avanco-da-inteligencia-artificial-no-brasil-abc-novembro-2023-GT-IA.pdf

Bardin, L. (1977). *Análise de conteúdo*. Edições 70.

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology* (2nd ed.). SAGE.

Lei Geral de Proteção de Dados Pessoais. (2018). Lei n. 13.709, de 14 de agosto de 2018. Dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural. https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm

Ministério da Saúde. (2020). *Estratégia de Saúde Digital para o Brasil 2020-2028*. https://bvsms.saude.gov.br/bvs/publicacoes/estrategia_saude_digital_Brasil.pdf

Neuendorf, K. A. (2002). Defining content analysis. In *The Content Analysis Guidebook* (pp. 1-26). SAGE.

Orlandi, E. P. (1999). *Análise de Discurso: princípios e procedimentos*. Pontes.

ANEXO 1 – ROTEIRO IA NA SAÚDE NO BRASIL, APLICADO AOS STAKEHOLDERS (ACADEMIA, EQUIPAMENTOS DE SAÚDE, PODER PÚBLICO E MERCADO)

MÓDULO 1.1: CENÁRIO BRASILEIRO

Para começar a conversa sobre IA em Saúde, vamos considerar o cenário brasileiro.

1. Em sua opinião, qual o estágio atual da Inteligência Artificial aplicada a saúde no Brasil?
2. Em sua opinião, quais são as oportunidades que a Inteligência Artificial gera para otimização do Setor de Saúde no Brasil? (EXPLORAR: Quando digo otimização do Setor de Saúde, quero dizer as oportunidades para melhorar serviços, ampliar escala de atendimento e diminuir custos do setor)
3. E o quão preparado você acha que o país está para aproveitar as oportunidades, bem como para gerenciar os riscos contidos no avanço da IA no setor de saúde? Por quê?
4. Ainda no mesmo sentido, em sua opinião, quais deveriam ser as apostas prioritárias do país para aproveitar as oportunidades que a IA pode trazer para o setor de Saúde? Quais são as áreas, setores, segmentos que teriam maior potencial? Onde deveríamos investir?
5. E quais são os possíveis riscos associados ao uso de Inteligência Artificial no setor de Saúde?
6. E quais são, em sua opinião, os grandes gargalos, desafios e barreiras para o desenvolvimento da IA no setor de Saúde no Brasil? Quais são desafios comuns a diferentes países e contextos, em sua opinião? E quais são desafios específicos no caso brasileiro?

MÓDULO 1.2: PRÁTICAS EM ANDAMENTO

Agora, vamos falar em práticas e usos de IA em Saúde que já estão em andamento em sua instituição, pensando especificamente na sua rotina e de sua instituição com aplicação de IA na área da Saúde.

7. (CENTROS DE PESQUISA/ EQUIPAMENTOS DE SAÚDE/ EMPRESAS) Atualmente, quais são as agendas de pesquisa, projetos, tecnologias e/ou soluções na área

de IA em Saúde que sua instituição tem desenvolvido? Com qual finalidade?

a. Qual o estágio atual de desenvolvimento desses projetos na área de IA e saúde?
    i. No caso específico da sua instituição, vocês têm buscado ou gerado incentivos para o desenvolvimento da área? Como?
b. Em sua opinião, quais são as potencialidades/ oportunidades que esses projetos geram?
c. Quais são as principais dificuldades/ empecilhos que vocês têm enfrentado nesses campos de atuação?
d. Por fim, como vocês lidam com os riscos que a IA impõe?

8. (GESTORES PÚBLICOS) Quais são as principais propostas, agendas e ações que vocês têm debatido atualmente pensando nos possíveis avanços da IA na área de Saúde?
    a. Qual o estágio atual de desenvolvimento desses projetos na área de IA e saúde?
        i. No caso específico da sua instituição, vocês têm buscado ou gerado incentivos para o desenvolvimento da área? Como?
    b. Em sua opinião, quais são as potencialidades/ oportunidades que esses projetos geram?
    c. Quais são as principais dificuldades/ empecilhos que vocês têm enfrentado nesses campos de atuação?
    d. Por fim, como vocês lidam com os riscos que a IA impõe?

MÓDULO 2: 07 PRIORIDADES ESD

Considerando que as potencialidades vislumbradas com o advento da Inteligência Artificial no setor de Saúde dependem da organização, processamento e gestão de uma vasta gama de dados e informações, gostaríamos de explorar sua percepção em alguns temas específicos.

M2.1 Em sua opinião, quais são os principais desafios de infraestrutura que o Brasil precisará enfrentar para maximizar as

oportunidades possíveis da IA no setor de saúde? Pensando em equipamentos, tecnologias, recursos, quais deveriam ser as prioridades de investimento do país em termos de infraestrutura?

M2.2 E em relação aos recursos humanos necessários para o desenvolvimento e bom aproveitamento das aplicações de IA em Saúde, você acha que os profissionais do setor de saúde no país estão preparados para lidar com essas novas tecnologias? Quais qualificações e/ou competências serão necessárias? Como preparar os profissionais para esse possível futuro?

M2.3. Aproveitar as potencialidades da IA passa por garantir uso amplo, diverso e em grande escala de diferentes fontes de dados. Pensando nisso, quais seriam possíveis estratégias para promover a interoperabilidade de sistemas de informação? Quais cuidados deveriam ser tomados nessa promoção?

a. Quais estímulos poderiam ser criados para garantir a integração e boa governança desses dados?
b. Como a sua instituição tem atuado nessas questões, visando inclusive maior interoperabilidade de dados com outras instituições?

M2.4. Pensando ainda no ambiente necessário para o desenvolvimento da IA na área de saúde, poderia me dizer, em sua opinião, quais são os principais desafios para o país em relação a:

a. Regulamentação. O que é necessário? Como o país está em relação a isso?
b. Ética. Quais são as principais questões éticas que devem ser endereçadas para o desenvolvimento seguro e proveitoso da IA na área de saúde no país?

M2.5. O quanto você conhece a RNDS – Rede Nacional de Dados em Saúde (*plataforma nacional de interoperabilidade em saúde do Programa Conecte SUS - MS*)? O que você conhece? (PARA OS QUE CONHECEM) Em sua opinião, a RNDS teria potencial, como plataforma de integração dos dados do país, para atender a demanda que a IA em saúde gera? Como?

M2.6. Por fim, vamos falar rapidamente sobre os usuários de saúde, os cidadãos.

a. Quais direitos dos usuários devem ser cuidados com maior atenção quando falamos de IA?

(ESTIMULAR APENAS SE NECESSÁRIO) Como lidar com as questões de privacidade? Como garantir a transparência e a ética nos processos?

b. Quais são as políticas atuais de sua instituição para lidar com essas questões? São suficientes? O que poderia ser aprimorado? O que funciona bem?

ANEXO 2 – ROTEIRO IA NA SAÚDE NO BRASIL, APLICADO AOS PROFISSIONAIS NO PONTO DE CUIDADO

APRESENTAÇÃO GERAL

1. Para começar, você pode se apresentar brevemente e contar um pouco sobre o seu trabalho atual (onde trabalha, qual especialidade médica, que tipo de atendimento clínico)?

APLICAÇÃO IA NA PRÁTICA CLÍNICA

2. Agora você pode contar como usa/tem usado ferramentas de IA na sua prática clínica? [Explorar: Quais ferramentas, quem desenvolveu, quais objetivos, qual a frequência usa]
3. Pode nos contar/descrever em detalhes como funciona essa ferramenta e como você aplica no seu dia a dia de trabalho? Em qual momento ou situação específica aplica? Como usa para tomar decisões clínicas? [disponibilidade de informações, relatório clínico, diagnóstico]

PROCESSO DE ADOÇÃO IA/INÍCIO DO USO NA PRÁTICA CLÍNICA

4. Como foi o início do processo de adoção dessas ferramentas na sua prática clínica? E algo mudou em relação a aplicação que você faz hoje?
5. Por que decidiu adotar essas ferramentas? [Explorar: Teve algum tipo de incentivo ou imposição da instituição em que trabalha? Foi uma iniciativa própria? Se inspirou na prática de alguém/algum profissional/alguma instituição?]
6. Teve alguma dificuldade no começo? Qual?
7. Teve algum tipo de treinamento? Como aprendeu a usar a ferramenta?
8. Considerando o que era a sua prática clínica antes, o que mudou com a adoção das ferramentas de IA?
    a. Quais são os principais impactos dessa adoção em sua prática clínica?

PACIENTE

9. Agora pensando na sua relação com seus pacientes, mudou alguma coisa em sua relação com eles após a adoção dessas ferramentas de IA?
10. O/A paciente é informado/a sobre o uso dessa ferramenta no seu atendimento/tratamento? Que tipo de informação recebe e em qual momento do atendimento/tratamento? Se não recebe, por que não é informado/a? [explorar: Qual a diretriz da organização em que você trabalha para essa questão?]
11. Tem alguma possibilidade do/a paciente recusar o uso da ferramenta no seu tratamento/atendimento? Como funciona? [explorar: Qual a diretriz da organização em que você trabalha para essa questão?]
12. Que tipo de dado do/a paciente é registrado? O/a paciente está ciente do uso desses dados? [Explorar: Qual a diretriz da organização em que você trabalha para essa questão?]
13. Onde é registrado e qual uso desses dados? Há compartilhamento desses dados? Se sim, como acontece?

FERRAMENTA: POTENCIALIDADES, MELHORIAS, DESAFIOS E RISCOS

14. Pensando na sua prática clínica, quais as vantagens dessa ferramenta? Que tipo de problemas ela ajuda a resolver/minimizar?
15. E do ponto de vista do paciente, qual a vantagem dessa ferramenta para a promoção da sua saúde e bem-estar? E, em sua opinião, existe algum malefício para o paciente pela adoção dessa ferramenta?
16. Em sua opinião, quais são fatores determinantes para o bom uso desse tipo de ferramenta? E quais seriam as principais barreiras, empecilhos atuais, para esse bom uso?
17. Essa ferramenta já lhe causou dificuldades/problemas? Quais? Como?
18. E quais aperfeiçoamentos/melhorias poderiam ser feitos nas ferramentas que você usa para se tornarem um recurso ainda mais benéfico?

19. Você identifica algum risco do uso dessa ferramenta de IA? Que riscos? Para quem?
20. Consegue imaginar estratégias para eliminar ou reduzir esses riscos? Quais?

CONFIANÇA:

21. Você confia nos resultados gerados pela tecnologia de IA?
22. Você sente receio de que ela lhe leve a tomar decisões equivocadas?
23. Em sua opinião, quem deveria ser responsabilizado em caso de possíveis erros ou equívocos em diagnósticos e tratamentos recomendados por essas tecnologias? Por quê?

INTERPRETABILIDADE:

24. Para você, o modo de funcionamento dessa tecnologia é claro?
25. Algum paciente já lhe indagou sobre como a tecnologia funciona?
    a. Se "sim", você se sentiu preparado para dar essa explicação? Quais estratégias utilizou para dá-la?
    b. Se "não", você se sentiria preparado para dar essa explicação? Como procuraria fazê-lo?

PERSPECTIVAS PARA O FUTURO

26. Por fim, em sua opinião, qual o futuro para os profissionais de saúde, especialmente para os profissionais no ponto de cuidado, considerando o desenvolvimento das tecnologias de Inteligência Artificial da área de Saúde?
27. Que tipos de políticas ou ações deveriam ser priorizadas no país para o uso adequado das tecnologias de IA por profissionais de saúde no ponto de cuidado?

Inteligência Artificial na saúde: diagnóstico qualitativo sobre o cenário brasileiro

Graziela Castello[1], Monise Picanço[2], Priscila Faria[3] e Rodrigo Brandão[4]

1 Cientista Social, é coordenadora de Métodos Qualitativos e Estudos Setoriais no Centro Regional para o Desenvolvimento da Sociedade da Informação (Cetic.br), departamento do Núcleo de Informação e Coordenação do Ponto BR (NIC.br).

2 Doutora e mestre em Sociologia pela Universidade de São Paulo (USP). Pesquisadora associada e coordenadora de projetos no Núcleo de Desenvolvimento do Centro Brasileiro de Análise e Planejamento (CEBRAP). Professora de metodologia e desenho de pesquisa na Faculdade de Informática e Administração Paulista (FIAP) e cebrap.lab.

3 Socióloga com mestrado e doutorado em Sociologia pela USP. Pesquisadora e coordenadora de projetos no Núcleo de Desenvolvimento do CEBRAP.

4 Doutorando em Sociologia pela USP, mestre em Ciência Política pela USP, com graduação em Ciências Sociais pela USP. É pesquisador na Coordenação de Métodos Qualitativos e Estudos Setoriais do Cetic.br|NIC.br.

INTRODUÇÃO

As tecnologias digitais emergentes e as possibilidades decorrentes da automação de processos têm o potencial de revolucionar e trazer grandes benefícios para o setor de saúde em suas múltiplas etapas: prevenção, diagnóstico, prognóstico e atendimento e tratamento dos indivíduos, bem como a gestão dos sistemas e equipamentos de saúde, otimizando o trabalho dos profissionais, dos recursos e das operações e rotinas estabelecidas. Neste sentido, a Inteligência Artificial (IA), por meio da utilização de dados de saúde complexos e de múltiplas fontes – como dados de prontuários eletrônicos, estudos de imagem, dados genômicos e fisiológicos –, tem amplo potencial para enfrentar crescentes desafios na transformação digital no setor de saúde. Esses desafios incluem aumento contínuo dos custos, falta de profissionais, mudanças epidemiológicas e demográficas em curso (como o envelhecimento populacional) entre outros (Academia Brasileira de Ciências [ABC], 2023, p. 7-8).

Nesse sentido, o uso de ferramentas de IA configura uma estratégia promissora para desenvolver e implementar políticas públicas de saúde em larga escala, assim como alavancar e acelerar a produção de conhecimento científico. Além das potencialidades, o desenvolvimento e a adoção de tecnologias de IA também impõem riscos a serem mitigados, especialmente no setor de saúde[5].

Diante desse cenário, faz-se relevante compreender o "estado da arte" da adoção da IA na área de saúde no Brasil – em que estágio de desenvolvimento e de uso estão essas ferramentas em relação aos estabelecimentos de saúde e aos profissionais do setor –, encarando as oportunidades e os desafios a serem endereçados no Brasil. Isso inclui questões sobre (a) os avanços da produção científica nacional sobre o tema, os principais debates e as agendas; (b) os limites e as possibilidades ofertadas pela adoção de ferramentas de IA para gestão e aprimoramento dos sistemas de saúde (públicos e privados); (c) e, fundamentalmente, os motivadores de adoção, o significado clínico e as implicações do uso dessas ferramentas para os profissionais.

---

5 Para exemplos de discussões sobre os possíveis riscos da IA na área de saúde, ver Challen *et al.* (2019) e Adler-Milstein *et al.* (2022).

Este capítulo apresenta resultados do estudo "Inteligência Artificial na saúde: diagnóstico qualitativo sobre o cenário brasileiro" realizado pelo Cetic.br, departamento do NIC.br, ligado ao Comitê Gestor da Internet no Brasil (CGI.br), em parceira com o CEBRAP. Trata-se de um estudo exploratório com objetivo de mapear experiências em andamento, coletar percepções de atores estratégicos e levantar informações sobre o estágio atual de desenvolvimento da IA na área da saúde no contexto brasileiro, em particular investigar as oportunidades, os riscos e os desafios para o aproveitamento do potencial dessas tecnologias no setor de saúde no Brasil.

A pesquisa consistiu em uma investigação qualitativa com diferentes perfis de *stakeholders* envolvidos no desenvolvimento e na implementação de IA na saúde no país. Foram realizadas 28 entrevistas em profundidade (EP), de aproximadamente uma hora de duração cada, com gestores e especialistas no tema da academia e de centros de pesquisa, do poder público, de empresas e estabelecimentos de saúde (públicos e privados), além de profissionais de saúde que usam essas tecnologias no atendimento a pacientes nos pontos de cuidado. Nas "Notas Metodológicas" desta publicação, são detalhados: (a) o desenho do estudo; (b) as definições metodológicas; (c) os perfis dos respondentes, os critérios de seleção e o total de entrevistas realizadas por perfil; (d) os instrumentos de coleta dos dados; e (e) as técnicas utilizadas para a codificação e a análise das entrevistas.

Buscando uma diversidade de perspectivas sobre o tema, o estudo teve como objetivo reconhecer as principais questões na agenda atual da IA aplicada à saúde no Brasil e, dessa forma, mapear as ideias emergentes, as preocupações iminentes e as expectativas futuras. Assim, a pesquisa explorou as percepções sobre o estágio atual de desenvolvimento da IA no setor de saúde no país, bem como oportunidades, desafios e riscos particulares à realidade brasileira. Também buscou-se coletar informações sobre iniciativas e práticas em andamento em diferentes áreas da saúde e compreensões sobre gestão de dados, segurança da informação, ética, regulação, entre outros temas relevantes para essa discussão. Desse modo, este estudo identificou que há um intenso debate sobre a IA no campo da saúde e que o ecossistema desfruta de

muito otimismo em relação aos benefícios potenciais dessa tecnologia, além de visões complexas e divergências entre os diferentes *stakeholders*, mas, principalmente, muitas sobreposições e convergências, as quais serão destacadas.

Este capítulo descreve os principais resultados dessa investigação empírica e conta com 4 seções, além desta breve apresentação. Nas próximas seções, são apresentados os resultados das entrevistas com *stakeholders*, sistematizados a partir de três direções, conforme ilustra a Figura 1. Na Seção "Cenário brasileiro", são apresentadas as percepções sobre estágio atual da IA na saúde no país e visões sobre oportunidades, dificuldades e riscos associados ao uso dessas ferramentas na realidade brasileira. Na Seção "Práticas em andamento", são descritas as iniciativas de formulação, desenvolvimento e implementação de IA na saúde no país, tal como mapeadas no estudo em diferentes contextos: academia, mercado, poder público e equipamentos de saúde. Já a Seção "IA nos pontos de cuidado" traz dados sobre os usos dessas tecnologias por profissionais de saúde em situações de atendimento a pacientes (nos pontos de cuidado). O texto encerra-se com a Seção "Considerações finais", em que são destacados os principais achados do estudo.

FIGURA 1 – ESTRUTURA DE ANÁLISE: PERFIL DOS ATORES ENTREVISTADOS E TEMAS INVESTIGADOS

FONTE: ELABORAÇÃO PRÓPRIA.

CENÁRIO BRASILEIRO

Nesta seção, são apresentadas as percepções dos *stakeholders* sobre a IA aplicada à saúde no contexto brasileiro. O objetivo é apresentar um panorama multifacetado desse tema e identificar as ideias em debate, os anseios e as perspectivas formuladas por um conjunto de atores-chave do país nessa agenda.

Estão descritas as compreensões sobre o estágio atual de adoção da IA na saúde no país, bem como as visões sobre oportunidades, dificuldades e riscos associados ao uso de ferramentas baseadas em IA na realidade brasileira. Cada um desses temas constitui um tópico da presente seção, que busca retratar as percepções sobre a IA na saúde no cenário brasileiro. Para concluí-la, apresenta-se um último tópico composto pela análise de perguntas específicas sobre tema da

Estratégia de Saúde Digital (ESD) aderentes ao debate sobre IA. Ao final de cada tópico, estão sistematizados os principais destaques analíticos.

Inicialmente, foi dada ênfase a consensos, temas e questões mais frequentes e vocalizados nas entrevistas, ou seja, as percepções mais transversais a todos os grupos de *stakeholders*, considerando que, entre os diferentes grupos, elas foram majoritariamente convergentes; entretanto, particularidades observadas em grupos específicos serão destacadas. Por fim, foram exploradas também possíveis "faltas", ou seja, questões ou temas que apareceram com menor intensidade do que o esperado, de acordo com a literatura pertinente.

ESTÁGIO ATUAL DA IA NA SAÚDE: AVANÇOS E LACUNAS

Esta seção apresenta uma visão geral do atual estágio da IA aplicada à área da saúde no Brasil. No início de cada entrevista, buscamos coletar percepções sobre o cenário atual, entender os avanços e as lacunas particulares à realidade brasileira. Também investigamos as percepções sobre a capacidade e a condição de preparo do país para absorver esses avanços e superar as lacunas.

Esses questionamentos iniciais trouxeram um panorama formado por muitos temas e assuntos relevantes, explorados e detalhados ao longo das entrevistas. Em resumo, foram identificadas as seguintes questões: (a) otimismo generalizado em relação às possibilidades da IA para a saúde no Brasil; (b) alguma preocupação de que a IA seja um "modismo", a ponto de muitas tecnologias que não são IA serem classificadas dessa forma; (c) percepção comum de que, na área da saúde, a IA ainda está em estágio inicial no Brasil; (d) existência de poucas iniciativas, as quais estão fragmentadas e desarticuladas; (e) iniciativas geralmente originadas no mercado, já que o poder público apresenta pouca capacidade de influenciar a agenda devido à burocracia ou à necessidade de maior cuidado na prática pública; (f) por fim, áreas específicas com um estágio de desenvolvimento mais avançado, tal como a radiologia, apesar da incipiência geral. Em um primeiro momento, esses aspectos serão apresentados de modo mais abrangente para serem aprofundados nos tópicos subsequentes.

Ao abordar questões sobre o estágio atual da IA aplicada à saúde no Brasil, os entrevistados do estudo expressaram muitas expectativas e grande entusiasmo pelo que pode ser desenvolvido no futuro. "Eu consigo olhar para frente e é um futuro bem promissor", sintetizou um *stakeholder* que gere um equipamento de saúde. De modo geral, há um grande otimismo e uma percepção generalizada de que o uso dessa tecnologia tende a trazer inúmeros benefícios para os sistemas de saúde público e privado do país. Ademais, há um consenso de que o uso da IA na área da saúde tem um grande potencial, tanto na gestão quanto na assistência médica, principalmente no apoio às tomadas de decisão pelos profissionais competentes. Além da redução de custos, a eficiência e a ampliação da escala dos serviços de saúde também são esperadas de ferramentas que aplicam IA da forma mais oportuna e promissora.

As pessoas em si estão muito empolgadas. A gente vê muitos alunos empolgadíssimos com a IA. [...] O interesse da população é muito grande, dos trabalhadores [da área] é muito grande.

(STAKEHOLDER *ACADEMIA*)

Sobre o otimismo observado de maneira praticamente generalizada, cabe registrar que alguns entrevistados advertem que ela tende a produzir uma banalização ou vulgarização do termo "Inteligência Artificial" nas ferramentas de saúde. Assim, embora soluções que utilizam algoritmos ou mesmo IA em processos simples de automação sejam apresentados como novidade, não representam necessariamente usos inovadores de IA com base em aprendizagem de máquina (*machine learning* [AM]) ou aprendizagem profundo (*deep learning* [AP]) para fornecer suporte à gestão ou à assistência médica. Um informante expressou a percepção de que o entusiasmo atual leva a um "modismo" de forma que todo mundo precisa dizer que usa IA em sua prática profissional, mesmo que, para isso, faça um uso equivocado ou superestimado do termo.

As pessoas inserem algoritmos ou inserem novas tecnologias e virou comum falar "eu uso IA" e eles vulgarizaram, na minha visão, o termo IA. Há uma grande diferença do uso de algoritmos preditivos, da adoção de algoritmos que aprendam sozinho e possam realmente entregar algo ou tomar decisões para poder aprimorar uma entrega qualquer. Eu vejo que a gente ainda está patinando na descoberta da melhor forma de encaixar o conceito e o uso da IA nesta jornada de saúde.

(STAKEHOLDER *MERCADO*)

A IA, hoje, ela está assumindo uma proporção muito grande, então tem a questão do modismo. Está todo mundo falando disso.

(STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

Eu diria que tem bastante gente engajada na adoção da IA dentro de suas ferramentas, porque há uma percepção de quem constrói as coisas, os *builders*, de "olha, eu preciso encaixar a IA". Então, eu diria que está todo mundo engajado a usar IA, mesmo que não saiba ainda para quê e por quê. Esse é o receio que eu vejo no mercado.

(STAKEHOLDER *MERCADO*)

Eu só quero fazer um parêntese: há muita confusão do que é IA e o que é estatística avançada, do que é *bot*, do que é uma mera automação. Então, a gente precisa primeiro separar isso aqui.

(STAKEHOLDER *MERCADO*)

A grande expectativa futura em relação ao desenvolvimento de soluções tecnologias baseadas em IA voltadas ao setor da saúde no país está associada a uma avaliação do cenário presente. Há um consenso de que a aplicação de IA na saúde no Brasil ainda está em um estágio inicial. Termos como "inicial", "começo", "embrionário" e "experimental" foram frequentemente mencionados, além de metáforas como "um bebê engatinhando" ou "aprendendo a andar".

Está bem embrionário, tá? [...] E, mesmo no estágio embrionário, já tem uma entrega de valor bem interessante. [...] Eu acho que, apesar de estar na fase embrionária, a gente tem muita coisa bacana que já foi feita, muita coisa em desenvolvimento e, nossa, vai ser um *boom* para o futuro!

(STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

Em uma escala de 0 a 10, sendo bem pragmático, eu diria que o Brasil está indo pelo 2 ou 3: não é nem engatinhando ainda, é observando como se anda.

(STAKEHOLDER *ACADEMIA*)

A gente viu a primeira onda, que é uma onda de experimentação, uma série de experimentos espalhados em todo país. Você vê pessoas na área da saúde testando algumas aplicações e algumas *startups* experimentando aplicações em fins clínicos para ajudar, de fato, no diagnóstico e no monitoramento de pacientes, e, até mesmo *backoffice* para melhorar a automação de ciclo de receita e de tarefas de relatórios. Eu acho que a segunda onda foi aí, a aplicação dessas soluções nas operações usando a IA para melhorar a detecção de nódulos pulmonares em ressonância, em tomografia de tórax, por exemplo. Você tem também aplicações para redução de perdas e desperdícios na área de fármacos e na parte de procedimentos. Então, eu diria que agora a gente sai desse estágio para a implementação, de fato, dessas tecnologias no âmbito Brasil.

(STAKEHOLDER *MERCADO*)

Apesar da visão consensual de que o país está em um estágio inicial do desenvolvimento e aplicação de IA na saúde, alguns entrevistados destacam que o *gap* entre o Brasil e outros países mais avançados nessa área não é tão grande, porque,

de modo geral, a aplicação de IA na saúde tem ocorrido em um ritmo mais lento do que em outras áreas. A incorporação dessas ferramentas na prática clínica e no apoio ao diagnóstico enfrenta mais restrições, precisa passar por mais testes e se submeter a regulações mais rígidas. Assim, mesmo países mais avançados no desenvolvimento tecnológico dessas ferramentas estariam em fase inicial de implementação, aprendendo a lidar com as implicações médicas e regulatórias do uso dessas tecnologias na prática. As citações a seguir ilustram essa interpretação comparativa:

> No Brasil, assim como nos outros países no mundo, está ainda em fase bastante inicial.
>
> (STAKEHOLDER *ACADEMIA*)

> Na área da saúde, os mesmos desafios que a gente tem em um grande hospital aqui de São Paulo é o mesmo que a gente tem lá em Houston por causa das regulações. Então, acho que, de todos os setores [em que há aplicação de IA], o da saúde é o menor *gap*, é o que tem o menor *gap* de IA entre o Brasil e outros países.
>
> (STAKEHOLDER *MERCADO*)

> Acho que, de forma global, a gente está bem no começo em saúde. [...] Eu não acho que o Brasil é tão atrasado em IA. Em saúde, talvez a gente esteja ainda, mas eu acho que está todo mundo ali meio que engatinhando assim [na aplicação de IA à área da saúde].
>
> (STAKEHOLDER *MERCADO*)

> Na saúde, eu acho que já existem muitas experiências na prática acontecendo, tanto no setor público quanto no privado. Mas eu acho que nós não temos ainda um olhar diagnóstico em relação ao inventário [de iniciativas] pra saber em que ponto nós estamos de maturidade em relação a outros países internacionalmente. A minha impressão é que não estamos na dianteira; [...] mas, para um país que está na situação que nós estamos, em um contexto mais geral, a gente está avançando.
>
> (STAKEHOLDER *PODER PÚBLICO*)

Em relação ao caso brasileiro, a percepção geral é de que o desenvolvimento das aplicações de IA tem ocorrido de forma fragmentada, ou seja, por meio de ações particulares desenvolvidas por diferentes atores, sem articulação. Essa fragmentação pode prejudicar o aproveitamento das vantagens da IA, limitando seu impacto positivo. Isso pode estar relacionado à falta de uma estratégia nacional, à diversidade de atores envolvidos (mas focados em experiências particulares) e à ausência de governança de grandes bases de dados na área da saúde. A integração eficiente de dados e a coordenação entre iniciativas são essenciais para maximizar os benefícios da IA na saúde. Assim, os entrevistados entendem que, no Brasil, atualmente há muitos projetos em desenvolvimento nas

universidades, nas empresas e mesmo na gestão pública, mas com dificuldades para implementação prática. Quando isso acontece, a aplicação é feita em contextos muito particulares, sem escala e capilaridade; geralmente são iniciativas localizadas e isoladas. A passagem da fase de projetos e protótipos para a fase de aplicação prática é ainda mais incipiente para ferramentas de IA de suporte à prática clínica; nesse sentido, soluções de apoio à gestão se disseminam mais facilmente. Os entrevistados entendem que o desenvolvimento dessa agenda no país ocorre por meio de uma dinâmica *bottom-up*, a partir de práticas localizadas em algumas instituições e sem diretrizes gerais. Mesmo em contextos organizacionais (como empresas e equipamentos de saúde públicos e privados), esses movimentos são liderados por um ou poucos grupos de profissionais mais do que por políticas institucionais abrangentes.

Eu diria que há uma expectativa muito grande do mercado de que a IA possa ser uma bala de prata na entrega de saúde, na questão primária, na questão especializada, na questão farmacêutica. E o pessoal já coloca muita expectativa, mas na prática eu vejo muita pouca atuação. Na prática, eu vejo que tem uma dificuldade de materializar a entrega da IA.

(STAKEHOLDER *MERCADO*)

Ah, em um estágio bem inicial, a gente ainda está na fase de organização de dados, de levantamento de dados, e pensando em como interoperar. Acho que a gente tem alguns cases muito pontuais, se a gente for falar em grande escala assim.

(STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

Quem está ativando muitas vezes isso são as *startups*, os médicos, ainda é muito *bottom-up,* né? Poucas empresas declararam, poucas empresas de saúde declararam assim "olha, a partir dessa data agora, o *top-down* vai ser assim".

(STAKEHOLDER *MERCADO*)

Acho que tem já algumas coisas muito bem estruturadas e funcionais, porém, não são muitas iniciativas. [...] Tem muito uso de IA em imagem e muito IA pronto para usar na área de gestão, mas tem pouca gente que usa. E acho que agora está entrando aí a IA generativa que o pessoal está querendo usar nessa parte de atendimento.

(STAKEHOLDER *ACADEMIA*)

Eu acho que são ilhas de excelência, eu acho que existem iniciativas hoje no setor privado bastante robustas. [...] Enfim, existem várias iniciativas pontuais que não são muito semelhantes, mas que tem uma certa sinergia sobre o tema.

(STAKEHOLDER *MERCADO*)

Entre os entrevistados, há um entendimento comum de que o mercado – empresas, *startups* e planos de saúde –, muitas vezes em parceria com a academia e centros de pesquisa, tem liderado o movimento recente de desenvolvimento e

disseminação de soluções de IA na área da saúde no Brasil. Os *stakeholders* de diferentes segmentos indicam também que o poder público estaria atrás nessa corrida: o tema estaria começando a entrar agora na agenda pública e a ter seu grande potencial reconhecido. Mas, de acordo com parcela significativa de entrevistados, o desenvolvimento dessa agenda no setor público aconteceria em ritmo mais lento em razão de questões burocráticas e regulatórias, além de falta de recursos. As entrevistas revelam que, no setor público, as iniciativas de desenvolvimento e adoção de IA na saúde também acontecem de forma descentralizada e territorializada, sem escala nem diretrizes especificas para seu uso: faltaria uma estratégia nacional organizada e centralizada para incentivar o uso da IA na saúde.

Apesar dessa lacuna destacada nas entrevistas, é muito forte a percepção de que há um grande potencial de uso da IA no sistema público de saúde brasileiro, graças ao Sistema Único de Saúde (SUS) e às suas bases de dados (de abrangência nacional). Os dados coletados pelo SUS são de enorme valor, considerando a capilaridade e a escala do sistema, e a diversidade da população brasileira, e representam um passo importante em direção à integração de dados, caso sejam tratados e preparados para essa finalidade.

A integração de dados foi apontada, na maioria das entrevistas, como ponto crucial para o desenvolvimento das ferramentas de IA. O tema foi apresentado como elemento-chave tanto nas oportunidades quanto nos desafios para o desenvolvimento de IA na Saúde no país, o qual será retomado nos tópicos seguintes. Por ora, é importante reiterar a percepção de que o sistema público de saúde brasileiro e suas bases de informação são estratégicos para a integração de dados em nível nacional.

As citações a seguir ilustram o retrato atual da IA aplicada à saúde no poder público brasileiro. As falas dos *stakeholders* destacadas abordam a dificuldade governamental de absorver as inovações tecnológicas, conforme percepção dos atores dos diferentes segmentos. Diversas razões são apontadas para isso, como a falta de recursos e a existência de questões burocráticas. As citações também evidenciam que não existem incentivos coordenados e diretrizes claras para a aplicação de ferramentas da IA nas políticas governamentais de saúde. Em

suma, as entrevistas apontam que, apesar do grande potencial do SUS para a articulação de iniciativas e a integração de dados, a inserção da IA no setor público de saúde enfrenta, hoje, grande fragmentação e desarticulação.

Um dos elementos principais é uma grande desigualdade – iniquidade até – entre o público e o privado. O privado tem já uma maturidade com relação a algoritmos de análise para a saúde populacional, identificação de riscos, vulnerabilidades para atuação em população. Está sendo utilizado, por exemplo, em alguns planos de saúde e em grandes corporações com um impacto incrível, um impacto grande utilizando nessas grandes bases e prontuários eletrônicos, por exemplo. Mas a gente tem no país as frontais do SUS, um dos maiores prontuários do mundo, senão o maior. Mas faltam soluções, [falta] o próprio sistema de regulação em si... A gente tem um grande desafio como política nacional da atenção especializada.

(STAKEHOLDER *PODER PÚBLICO*)

O setor público acaba que demora até um pouco mais absorver essas alterações e essas tecnologias. E não é problema de falta de conhecimento, é mais por questões burocráticas, do que é permitido, do que não é permitido. Quando você está no setor privado, você tem um pouco mais de liberdade para explorar as novidades. Mas eu vejo que, embora o setor público possa estar um pouco mais cauteloso, não está deixando passar não, está fazendo projetos, está buscando informações

(STAKEHOLDER *PODER PÚBLICO*)

No setor público, tudo que você tem hoje em relação a IA é uma questão muito mais momentânea, política do prefeito, do secretário de saúde em âmbito estadual ou municipal, do que, de fato, dentro da estratégia ou dentro de um plano de governo isso você não tem. [...] Hoje, você tem uma desconexão do Ministério da Saúde (MS) em relação a uso da IA com os estados e municípios. Não tem hoje política clara, você não tem sentidos claros de porque um prefeito deveria incorporar IA no Programa de Saúde da Família para ajudar o agente comunitário de saúde porque falta letramento digital.

(STAKEHOLDER *MERCADO*)

A gente precisaria de um incentivo de investimento mais apropriado do setor público. Vou falar especialmente a questão do SUS. Nessa área, para a gente ter soluções de impacto mais global, a gente tem que trabalhar mais cooperação de rede entre municípios e estados e nacional, do que experiências localizadas em um ponto que a gente vai escalar depois. E SUS, na verdade, é a cooperação, né?

(STAKEHOLDER *PODER PÚBLICO*)

O Brasil não tem um projeto de IA nacional, ao contrário da grande maioria dos outros países do mesmo tamanho, com economias semelhantes. [...] Projetos para incentivar a gente ainda não tem, não tem uma estratégia nacional para a IA.

(STAKEHOLDER *ACADEMIA*)

Na gestão [pública] nós ainda estamos entrando na camada de *analytics*, estamos ainda montando o que seria as antigas salas de circulação ou sala de comando, indo para tecnologias de BI, *Big Data*. Ainda é muito pouco. [...] E na parte clínica, começamos a ver alguma coisa com sistema de apoio a decisão, porém ainda muito incipiente.

(STAKEHOLDER *PODER PÚBLICO*)

Hoje, eu acho que, se a gente olhar o ambiente da saúde privada, a gente tem coisas sendo feitas e utilizadas que são iguais a países como Estados Unidos e países da Europa. No SUS, a gente tem alguns nichos de excelência, mas se a gente olhar o volume geral, ainda tem muita carência.

(STAKEHOLDER *ACADEMIA*)

Os entrevistados também percebem que a evolução da IA não se dá de forma homogênea entre os diferentes setores e áreas da saúde. Como antecipado, o uso de IA no setor de gestão e administração estaria mais avançado que na prática clínica. Entre as áreas médicas em que há maior avanço no desenvolvimento e aplicação de IA na saúde no cenário nacional, destaca-se a radiologia. O uso de IA nos exames e diagnósticos por imagem está mais consolidado e incorporado às práticas cotidianas, inclusive porque, na visão dos entrevistados, os equipamentos começam a ter esse tipo de tecnologia embutida; por isso, nos cursos universitários e de especialização, os profissionais têm sido preparados para lidarem com essas ferramentas. A seguir, apresentamos um conjunto de falas que sublinham os avanços da aplicação de IA na área de radiologia no país:

Em imagem está mais funcional. Nessa área, já é possível usar.

(STAKEHOLDER *ACADEMIA*)

Claro que tem algumas áreas mais evoluídas, que são as áreas da Radiologia. Eu vejo bastante coisa que já funciona, mas também vindo embutida dentro dos *softwares* e equipamentos né.

(STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

É importante ressaltar que, claro, que muitos projetos giram em torno da radiologia porque, mesmo nos EUA, dos algoritmos que foram aprovados pelo Food and Drug Administration (FDA), 75% são relacionados de alguma forma a radiologia. Então, realmente a radiologia, né? O diagnóstico foi uma área que, por ser uma área tecnológica, a IA passou a ser utilizada de uma forma mais ampla, desde o começo.

(STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

Algumas áreas estão mais desenvolvidas que outras, então, na saúde, eu entendo que a área da radiologia é uma área que que vai se beneficiar e está se beneficiando da contribuição da IA. Então, eu entendo que a radiologia é a principal área que está se beneficiando. Aí, você tem também aquela área ou setores de interesse no prontuário eletrônico que também podem se beneficiar.

(STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

Ao abordar os avanços da IA aplicada à saúde no país, alguns entrevistados identificaram a crescente busca de

profissionais de saúde por conhecimento e qualificação relacionados ao tema. Nesse sentido, alguns citaram o aumento substantivo do interesse por parte de diferentes profissionais da saúde, inclusive entre estudantes e médicos recém-formados. Também foi mencionada a percepção de aumento da interlocução entre profissionais de tecnologia da informação (TI) e da área médica na construção, aprimoramento e manutenção de ferramentas de IA aplicadas ao contexto de saúde. Esse movimento é considerado positivo e significaria, na visão dos entrevistados, um avanço na construção de equipes multidisciplinares e qualificadas para construir e operar essas ferramentas.

> A percepção que eu tenho, até por estar dentro de uma escola médica, é de um grande interesse dos profissionais médicos. Então, hoje a gente já tem, inclusive de forma mais comum na área da radiologia, temos médicos radiologistas que aprendem a mexer com esses algoritmos e até desenvolvê-los. Hoje em dia, uma boa parte da implementação já está feita, a gente acaba fazendo muito mais uma customização, ajuste de parâmetros. Mesmo assim, você precisa ter um certo domínio de programação, precisa saber mexer naquele ambiente. E hoje, a gente tem médicos mais jovens, que estão se formando e já tem interesse e familiaridade com as questões computacionais e que começam a mexer com isso e avançam.
>
> *(STAKEHOLDER ACADEMIA)*

> De um lado tem um interesse muito crescente nos últimos anos da comunidade da área das exatas, engenharia da computação e física. E por outro lado, os próprios profissionais da saúde, os médicos e os outros profissionais, como enfermagem, costumam ser muito ativos na área de saúde digital. A gente vê uma aproximação dos dois lados. Então, eu vejo o pessoal ali da computação e engenharia se aproximando da saúde porque é uma área de grande impacto. Por outro lado, o pessoal da saúde também começando a trabalhar e entender esses modelos.
>
> *(STAKEHOLDER ACADEMIA)*

Ainda considerando esse panorama sobre o estágio atual da IA na saúde no país, destaca-se, a partir das entrevistas, a percepção de algumas lacunas e fragilidades. Os aspectos apontados pelos entrevistados como principais lacunas para a superação desse estágio inicial e incipiente das tecnologias de IA na saúde no Brasil são principalmente relacionados à disponibilização e à governança de dados, e à regulação.

No que se refere ao tema dos dados, explorado em profundidade em outras seções deste capítulo, cabe registrar que os entrevistados do estudo entendem que a falta de dados de qualidade e/ou a dificuldade de integração de grandes bases de dados dificultam o avanço da IA na saúde no país. A

ausência de uma arquitetura adequada para disponibilização e uso de dados é considerada uma fragilidade marcante do cenário brasileiro.

> Então, a gente precisa ter um pouco mais de cuidado e ir acompanhando as mudanças e se preparando para o momento em que o Brasil esteja pronto para colocar os algoritmos, de fato, na prática clínica, o que vai depender da qualidade de dados. Vai depender de superar alguns dos grandes desafios de colocar os algoritmos na prática clínica que é a questão em primeiro lugar da qualidade desses algoritmos, a performance deles em cada local específico. A gente vai ver se eles estão funcionando nas diferentes realidades brasileiras.
> (STAKEHOLDER *ACADEMIA*)

> Hoje, o maior problema aqui nos hospitais é a qualidade dos dados. Para eu aplicar qualquer produto de IA, eu preciso ter uma camada de dados de qualidade, porque senão de nada vai valer o meu produto, ele não vai atingir o máximo potencial.
> (STAKEHOLDER *MERCADO*)

Quanto à regulação, observa-se uma percepção generalizada de que há uma lacuna relevante no contexto atual brasileiro. Os entrevistados de diferentes segmentos partilham da percepção de que faltam normativas e diretrizes para desenvolvimento e aplicação de ferramentas de IA na saúde. Além disso, apontam a ausência de estruturas de governança para acompanhar, validar e fiscalizar essas aplicações na prática.

> A gente vê as soluções sendo mais rapidamente implementadas [no Brasil]. O ambiente é muito fértil e com muitas soluções sendo implementadas e desenvolvidas. Mas, por outro lado, uma preocupação minha é que isso abre um outro viés que eu acho complicado e delicado. Pelo fato de conseguir fazer a prova de conceito mais rapidamente, mais rapidamente você pode colocar isso em prática como ferramenta clínica efetiva. E, aí, eu acho que é um ponto onde a gente ainda tem uma certa dificuldade porque falta um pouco de organização e talvez de governança, além de maturidade de quem faz. Uma coisa é a gente estar no laboratório e mesmo no ambiente clínico, mas trabalhando com estratégias de investigação. Outra coisa é você partir para o ambiente clínico com uma ferramenta de uso e produção, e que exija cuidados de manutenção e acompanhamento que eu acho que ainda não são muito claros.
> (STAKEHOLDER *ACADEMIA*)

> Falta uma regulação a nível de país que incentive isso, porque parte dessas tecnologias e parte dessas ferramentas ou plataformas de IA precisam ter Agência Nacional de Vigilância Sanitária (Anvisa), e a gente tem uma agência que não tem a menor capacidade, essa é a verdade, de avaliar ou de ter uma política com critérios claros e a homologação de ferramentas de IA. Então, o regulatório ainda não está correspondendo em relação a isso.
> (STAKEHOLDER *MERCADO*)

De maneira geral, os entrevistados acreditam que o Brasil precisa endereçar principalmente essas duas grandes questões – arquitetura para disponibilização e governança de dados, e

regulação – para vencer o estágio atual de desenvolvimento e se preparar para as potencialidades e riscos que a IA traz para a saúde. Problemas de infraestrutura e culturais, como resistência à inovação, também foram mencionados, mas com menor frequência e relevância na fala das pessoas entrevistadas

BOX DESTAQUES - ESTÁGIO ATUAL DA IA NA SAÚDE

- Prevalência de expectativas muito positivas sobre os benefícios da IA para o sistema de saúde público e privado no país. Perspectivas de redução de custos, aumento da eficiência e escalabilidade dos serviços de saúde.
- Percepção geral de haver grande potencial para o uso de IA na área da Saúde no país, tanto para a gestão como na assistência médica. Existência de áreas específicas com um estágio de desenvolvimento mais avançado, como a radiologia.
- Aumento do interesse e do entusiasmo por parte de profissionais e estudantes da área da saúde proporcionados pelo cenário positivo, além de uma interação crescente entre profissionais de TI e da área médica na construção e na operação de ferramentas de IA na saúde.
- Consenso de que o país está em estágio inicial e experimental do uso de IA na saúde. Soluções desenvolvidas de forma fragmentada, com pouca articulação.
- Percepção de que as iniciativas geralmente são originadas no mercado devido a diferentes restrições próprias das organizações do estado, tais como: maior burocracia, falta de recursos e a necessidade de maior zelo na prática pública.
- Existência de uma dificuldade na passagem da fase de protótipos para a implementação prática em relação à operacionalização, especialmente no caso de ferramentas de suporte à prática clínica, que dependem da formação dos profissionais e das dinâmicas da rotina hospitalar.
- Ausência de normativas e diretrizes regulatórias para o desenvolvimento e a aplicação de ferramentas de IA na saúde, apontada como uma lacuna significativa, assim como a falta de estruturas de governança para validar e fiscalizar o uso das ferramentas de IA no setor.
- Falta de arquiteturas para disponibilização e integração de dados de qualidade.

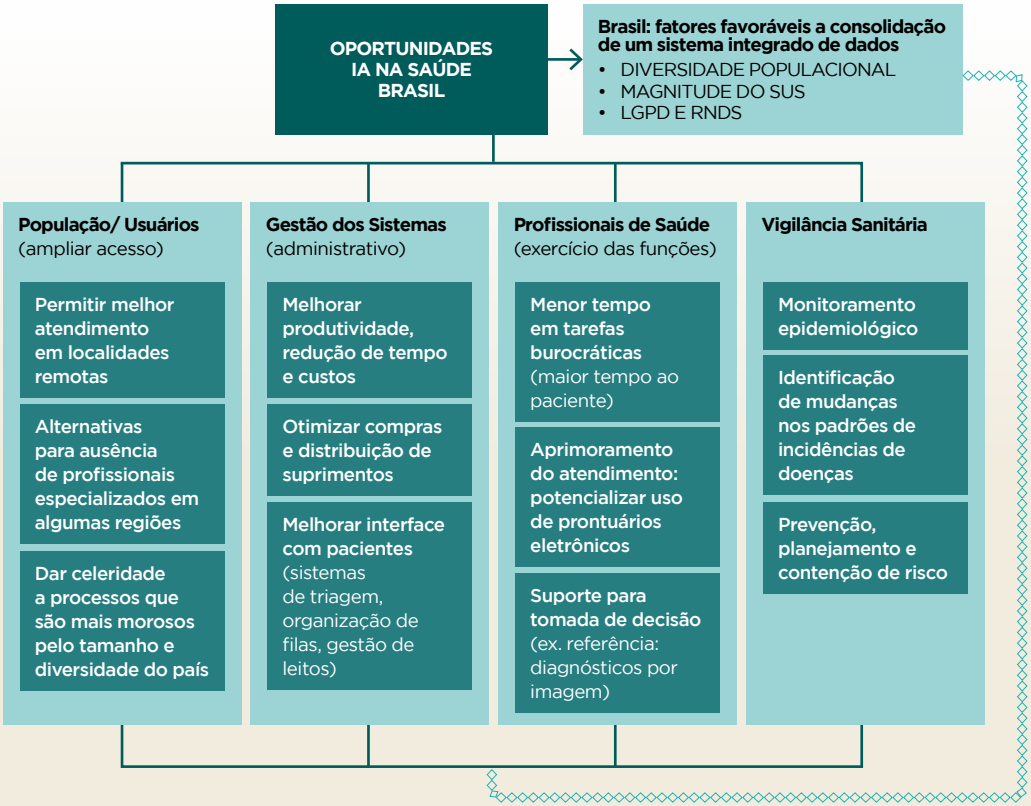
OPORTUNIDADES PARA IA NA SAÚDE

Esta subseção é dedicada a apresentar as percepções dos *stakeholders* a respeito das oportunidades para a IA aplicada à saúde, considerando as especificidades do contexto brasileiro. As entrevistas buscaram captar as visões sobre as potencialidades desse tipo de tecnologia: quais setores, áreas da saúde, tipos de ferramentas e aplicações têm maior potencial e representam as melhores oportunidades para a realidade brasileira. Também buscou-se compreender que tipos de gargalos na área da saúde do país essa tecnologia poderia ajudar a resolver.

Como mencionado, o otimismo é grande entre os entrevistados do estudo, os quais visualizam muitas oportunidades para o uso da IA em saúde no Brasil. São apontadas potencialidades em diversos setores e processos da gestão de saúde e do atendimento clínico, com melhorias vislumbradas para diferentes instituições e atores envolvidos nesse ecossistema, e benefícios para a coletividade como um todo. A Figura 2 sintetiza uma visão geral das oportunidades identificadas nas entrevistas.

> Acho que [com IA] é possível ampliar a oferta de serviços, é possível escalar, é possível melhorar o acesso, é possível melhorar a precisão de diagnósticos, de ter informações para tomada de decisão, tanto no nível de um atendimento clínico – exames de imagem é uma coisa que comprovadamente a IA é superior ao olho humano – como também na predição de futuras emergências sanitárias, por exemplo. Na área de vigilância e predição, acho que também nós temos um potencial muito grande e que o impacto já vem sendo aproveitado pelo Brasil.
>
> (STAKEHOLDER *PODER PÚBLICO*)

FIGURA 2 – OPORTUNIDADES DA IA NA SAÚDE NO BRASIL



FONTE: ELABORAÇÃO PRÓPRIA.

População/usuários

Considerando as particularidades do Brasil e o reconhecimento de uma ampla desigualdade no acesso à saúde de qualidade, muitos entrevistados apontaram a IA como uma tecnologia capaz de minimizar os efeitos desse problema social. Partindo do diagnóstico de uma significativa disparidade territorial no sistema de saúde que resulta na escassez de serviços, equipamentos e profissionais em muitas regiões do país, ferramentas de IA poderiam conectar pacientes e profissionais dessas localidades a conhecimento especializado para apoiar decisões operacionais e clínicas.

Em cenários em que há escassez de médicos especialistas, por exemplo, as soluções de IA podem ajudar os profissionais disponíveis a preverem desfechos clínicos e agilizar encaminhamentos. Os sistemas de apoio à decisão médica são considerados pelos entrevistados boas oportunidades para reduzir a desigualdade no acesso a atendimento qualificado, além de agilizar fluxos e processos. Esse avanço é particularmente relevante no contexto brasileiro, em que a distribuição inadequada de recursos e profissionais resulta na falta de serviços e na lentidão em atendimentos, diagnósticos e tratamentos.

Aí está uma certeza: da mesma forma que os algoritmos transformaram outras áreas, vão também transformar a área da saúde e garantir um atendimento de qualidade, principalmente em regiões mais remotas do Brasil. Regiões onde você tem um único médico, por exemplo. A gente tem muitas cidades brasileiras onde você tem um único médico; ele tem que tomar decisões em todas as especialidades, não tem um cardiologista, não tem um pneumologista para encaminhar... ele tem que tomar essas decisões. Com a chegada desses algoritmos, ele vai ter o auxílio dos melhores médicos do mundo, diminuindo assim a imensa desigualdade que a gente tem no atendimento à saúde nas diferentes regiões brasileiras. A gente vai trazer atendimento especializado de qualidade nas regiões mais remotas do Brasil com a ascensão desse algoritmo de IA.

(STAKEHOLDER *ACADEMIA*)

Outra oportunidade é que, através da IA, você conseguiria oferecer serviços para a população que hoje são muito caros de serem providos. Por exemplo, vamos imaginar um paciente que é um cardiopata ou um diabético de terceira idade que mora sozinho. [...] através do IA e da nuvem você vai estar acompanhando esse paciente que hoje está na casa dele e às vezes não tem o dinheiro para o transporte ou o lugar que ele tem que ir é longe. Então, você conseguiria dar o mesmo médico que está aqui no [Centro hospitalar A][6] para atender e ele poderia, por exemplo, automatizar alguns protocolos clínicos. E o seu José que mora na Amazônia e é um cardiopata ou um diabético e o posto mais próximo está a 100 km da casa dele, ele poderia preencher umas perguntas e o mesmo formulário, o mesmo *bot* que atende São Paulo atenderia aquele cara lá. E se a IA detecta que o seu João que está no meio da Amazônia, cujo posto de saúde mais próximo está a 100 km, o médico conseguir saber que esse cara precisa de atendimento de um profissional.

(STAKEHOLDER *MERCADO*)

A gente sabe que tem vários Brasis no Brasil. Então, ao você pensar que, no Norte, no Nordeste, ou mesmo aqui no estado de São Paulo, você precisa de médico qualificado, sistema de apoio à decisão e triagem... Então, existem oportunidades nessas áreas de apoiar locais onde há uma restrição de médicos, por exemplo, radiologista, ou até mesmo patologista. [...] Então, assim, certamente um sistema de apoio à decisão apoiaria com maior acurácia. São oportunidades que eu vejo, tanto na triagem quanto no apoio à decisão para locais onde precisa de uma *expertise* maior. Então, imagina um sistema treinado com milhões de imagens e um residente, um médico que não é especialista da área num local assim, eu acho que tem uma contribuição valiosíssima.

(STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

---

6 Centro hospitalar com sede em São Paulo e filial em Brasília. Considerado como referência em atendimento na América Latina, integra uma rede de especialistas e formações em saúde.

Gestão dos sistemas

Em relação a questões administrativas e operacionais, os entrevistados destacaram amplamente o potencial da IA para o aumento da produtividade e da eficiência em gestão, operação e logística de sistemas, serviços e equipamentos de saúde. Citaram particularmente o aperfeiçoamento de fluxos e processos, redução de tempo e otimização de recursos financeiros e humanos, potenciais vislumbrados para a melhoria tanto do sistema público quanto privado de saúde.

Tem uma parte de melhoras de fluxo financeiro, convênio, solicitações de compras... Então, acho que tudo isso da parte administrativa tende a evoluir [com a IA].

(STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

Diminuir custos. Na hora que eu faço uma predição, uma previsão que um cidadão pode ter uma necessidade de um atendimento de alto custo e na atenção primária eu evito isso, são recursos [economizados] e foi um benefício para o cidadão.

(STAKEHOLDER *PODER PÚBLICO*)

Os entrevistados mostraram-se otimistas quanto ao uso de ferramentas de IA para organização de logística e processos internos de trabalho, como compras e distribuição de suprimentos, equipamentos e medicamentos, além da alocação de recursos humanos. Também é grande a expectativa em relação à melhoria de fluxos e operações com interface com os pacientes, tais como triagem, organização de filas, priorização de atendimento e gestão de leitos e de agendamento de exames.

Existem muitas aplicações hoje voltadas mais para parte de gestão que são muito importantes também. Otimizar fluxo de atendimento, otimizar fluxo de uso de equipamentos, fluxo de ocupação de leitos. Eu sou otimista no sentido de potencial dessas ferramentas.

(STAKEHOLDER *ACADEMIA*)

Impacto da IA para eficiência operacional. Centro cirúrgico rodar mais rápido, algoritmo de aceleração de ressonância, então acaba diminuindo a fila desses exames, de rotatividade de sala, giro de leito, segurança do paciente dentro do hospital. Isso tem um impacto muito grande.

(STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

Outra frente importante também que a gente tem que existem em todos os hospitais é gestão de fila, por exemplo, fila cirúrgica ou fila de procedimentos, qual o critério que você vai adotar para priorizar a ordem de uma fila. Então, a gente tem utilizado com base nos históricos de atendimento dos nossos pacientes, métodos automatizados para sugerir reordenamento de filas, quando o critério já está bem estabelecido. Isso também a gente está fazendo utilizando métodos de aprendizado de máquina, IA, mas especificamente aprendizado de máquina.

(STAKEHOLDER *ACADEMIA*)

> Então, fluxo de trabalho, identificação de prioridades e gestão de filas de casos graves, eu acho que isso tem muito a ser melhorado [no Brasil]. Por exemplo, essa ferramenta de IA que eu falei. Se eu tenho um pronto-socorro que faz um monte de exames, se eu tenho uma IA que toda vez que apareceu um caso muito grave ele já soltar um alerta para a equipe... Você não vai depender do processo do humano olhar, ver que aquilo é grave para poder tomar uma conduta, mas isso já é um baita de um ganho de tempo, os modelos preditivos em geral. Eu acho que a modelagem preditiva tem muito a beneficiar.
>
> (STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

A melhoria desses aspectos de gestão pode resultar na otimização dos sistemas de saúde como um todo, aumentando suas capacidades de absorção de demandas e atendimento, a fim de expandir o acesso da população à saúde. Nesse sentido, para os entrevistados, aprimorar os fluxos operacionais dos sistemas e equipamentos de saúde é uma das principais potencialidades da IA no Brasil.

As entrevistas destacaram que esses tipos de aplicação de IA na gestão e na operação já são realidade em alguns contextos, principalmente no sistema privado de saúde, no qual seus efeitos positivos já são percebidos na prática.

Profissionais de saúde

Como mencionado, as entrevistas também apontam grandes potenciais para IA na área clínica e na frente de cuidado. Parte substantiva dos *stakeholders* mencionou o potencial de ferramentas de IA para o suporte ao trabalho cotidiano dos profissionais de saúde nos pontos de cuidado, como médicos, enfermeiros, fisioterapeutas, etc. Os entrevistados vislumbram que essas ferramentas podem reduzir o tempo do trabalho desses profissionais em tarefas burocráticas e administrativas, como preenchimento de formulários, documentos e prontuários. Com esses recursos em prática, esses profissionais poderão dedicar mais tempo à escuta, à atenção e ao cuidado humanizado. Além disso, a otimização do tempo de trabalho dos profissionais de saúde trará mais produtividade nos atendimentos, melhorias no cotidiano de trabalho dos profissionais de saúde a fim de tornar a atuação clínica mais efetiva e focada, beneficiando o cuidado do paciente em última instância.

Aumentar bastante a produtividade dos médicos e diminuir muito o tempo perdido com burocracias. Hoje, boa parte do tempo do médico é perdido preenchendo papelada. Os algoritmos conseguem preencher isso automaticamente via conversa do médico, via identificação dos procedimentos que foram realizados, aumentando o potencial dos médicos de usarem o seu tempo da forma mais produtiva e da forma que eles gostariam de utilizar que é atendendo paciente e não preenchendo burocracias internas.

(STAKEHOLDER *ACADEMIA*)

Vamos para o lado do profissional [de saúde]. [...] Quando a gente olha o trabalho do profissional de saúde hoje, tem vários artigos falando disso, é impressionante que mais de 30% do tempo dele ele está ali atrás de um computador preenchendo papel. [...] Então, eu acho que o primeiro ponto é que a IA vem para auxiliar muito nesse processo de preenchimento de papéis.

(STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

Nesse sentido, as potencialidades de ferramentas relacionadas ao prontuário médico foram amplamente mencionadas pelos entrevistados, sobretudo pelo potencial de melhorar a eficiência no acesso e na gestão das informações de pacientes. Foram citadas aplicações que registram automaticamente dados no sistema de prontuário a partir da transcrição de falas do atendimento e de outras situações, que aprimoram e sintetizam as informações-chave do histórico do paciente. As entrevistas também destacam as grandes vantagens dos sistemas de prontuários eletrônicos que integram e organizam informações valiosas sobre o histórico dos pacientes ao longo da vida. Ademais, o desenvolvimento dessas aplicações também tem potencial para produção de análises mais aprofundadas sobre o estado clínico presente e visualizações sobre o futuro dos pacientes, por meio de prognósticos e diagnósticos precoces. De acordo com as entrevistas, as ferramentas de IA para registro, otimização e visualização de prontuário também trariam maior produtividade e eficiência aos fluxos de atendimento, benefícios para o trabalho dos profissionais de saúde e um cuidado mais efetivo.

Hoje em dia, a equipe de saúde não tem tempo para cuidar de um paciente, não dá tempo. Você vê aquele portuário eletrônico cheio de informação trabalhosa, com 50 exames anteriores, ele não tem meia hora para olhar tudo isso, uma hora para olhar tudo isso. Ele tem 5 minutos para ler o histórico e 10 [minutos] para atender ao paciente. Então, como que eu faço para que, naqueles 5 minutos, ele consiga digerir todos os dados que já têm lá? O sistema pode digerir toda essa massa de dados, refinar ela, mais que minerar, e gerar um *lead* ultraqualificado de quem são as pessoas que precisam de atendimento agora.

(STAKEHOLDER *MERCADO*)

Eu vejo no prontuário [uma oportunidade]. [...] O *burnout* [dos profissionais] associado a utilização do prontuário é um dos principais problemas na área de saúde. [Com IA] você consegue entregar a pessoa para cuidar da pessoa, em vez de ficar preenchendo prontuário. A segunda coisa é como que a gente analisa esse prontuário? [...] Então, ao mesmo tempo que a gente pode transformar o prontuário e analisar esse prontuário melhor, eu posso gerar visualizações diferentes para cada usuário baseado no perfil. Eu consigo pegar dados totalmente desestruturados e estruturar, mas eu consigo também devolver para pessoa, eu consigo devolver para gente poder cuidar daquela pessoa.

(STAKEHOLDER *MERCADO*)

Outra coisa que eu vejo [como potencial] é você orientar o profissional de saúde que é outro grande campo. [...] Fazer um resumo do histórico de utilização da pessoa e fazer um resumo do histórico de doenças, que na medicina a gente chama de história patológica pregressa. Isso já acelera para o médico tomar uma decisão. Exemplo: é um paciente idoso que teve um Acidente Vascular Cerebral (AVC) há quatro anos atrás, recebeu alta e fez tratamento em clínica de reabilitação e foi classificado como apto após a reabilitação. Então, se eu boto um resumo desse para o profissional, eu ajudo ele a tomar uma decisão mais rápida sobre os problemas.

(STAKEHOLDER *MERCADO*)

Os entrevistados também mencionaram muitas potencialidades da aplicação de IA no suporte à tomada de decisão na prática clínica. Foram mencionadas ferramentas que podem ajudar na formulação dos diagnósticos e sistemas baseados em modelos preditivos para visualização de desfechos clínicos no futuro e estratégias de prevenção.

A oportunidade principal no contexto da saúde é o apoio à decisão médica.

(STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

A grande expectativa é você começar a ter os assistentes pessoais para poder ajudar o médico. Então, ajudar o médico a encerrar uma consulta, transcrever coisas que o médico fala durante o processo para os sistemas eletrônicos, fazer comparações dos padrões e dizer que aquilo está fora e sugerir terapêuticas, essas coisas todas. Acho que tem um campo grande [de oportunidades] que a gente ainda não sabe tudo. [...] Você teria mais quantidade [de atendimentos] e, logo, mais acesso. Melhorar qualidade também, você conseguiria ir em áreas que não têm essa qualidade e transferir isso, então melhorar a qualidade do serviço.

(STAKEHOLDER *ACADEMIA*)

Tem ali uma aplicação [de IA] que é muito direta com o profissional, você ter um algoritmo ou um modelo que ajuda a diagnosticar, ajuda a prever, por exemplo, evolução, ajuda a prever desfechos clínicos. Então, baseado no que a gente tem no prontuário e num modelo treinado com um conjunto muito grande de dados prévios, a gente consegue ter uma referência de um desfecho clínico nos próximos seis meses, um ano, dois anos. Isso é uma coisa fenomenal. [...] Você conseguir prever evolução, pensando na saúde pública da comunidade, de populações, com determinados desfechos mais prováveis, população que está envelhecendo, conseguir prever tendência... Eu já vi trabalhos com idosos para prever tendência de queda de idoso, essas aplicações diretas para o profissional da saúde podem melhorar o cuidado.

(STAKEHOLDER *ACADEMIA*)

Na parte clínica, pensando em diagnósticos precoces, pensando em prevenção. Acho que a gente vai trabalhar muito com IA em termos preventivos do paciente, do indivíduo... ainda antes de ele ser paciente, já ter as informações dele capturada em algum lugar e fazer *insights* para ele, para ele melhorar o cuidado com a saúde, para que ele não venha a precisar de internação.

(STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

As ferramentas de IA aplicadas aos exames por imagem são apontadas como bons exemplos de soluções que apoiam a tomada de decisão e reduzem o tempo envolvido em laudos e diagnósticos, como expressam as citações seguintes:

Outro exemplo é o diagnóstico por imagem. Então, a gente hoje já vê vários diagnósticos por imagem usando IA e acelerando o tempo. Você faz uma tomografia, você faz uma ressonância e até aquilo ali receber um laudo, às vezes leva dez dias, 15 dias. Com a IA, o laudo é praticamente imediato, obviamente que é preciso que um humano, um médico confira aquilo ali e, se houver uma divergência entre o que a IA apontou e aquilo que o médico acha, uma outra opinião deve ser ouvida, mas isso também acelera. Ao invés de ter um laudo que vai demorar uma semana ou mais, esse laudo é imediato.

(STAKEHOLDER *PODER PÚBLICO*)

Se eu fosse escolher algum terreno fértil em oportunidades em que a gente pode colher resultados rápidos de largo impacto, é imagem. É muito bom a gente ter um algoritmo de identificação de imagem radiográfica para laudar mais rapidamente. Isso é muito importante e está avançando muito com relação a isso.

(STAKEHOLDER *PODER PÚBLICO*)

Vigilância sanitária

No que ser refere ao campo da saúde pública e da vigilância sanitária, também foi mencionado o potencial da IA para auxiliar no processo de monitoramento epidemiológico e identificação de mudanças nos padrões de incidências de doenças na população. As ferramentas de IA aplicadas a esses fins podem melhorar a capacidade de prevenção, planejamento e contenção de riscos e emergências sanitárias.

Então, pensando no nível do SUS... Você ter alguma coisa que não seja uma ferramenta muito sofisticada, mas que te ajuda a organizar fluxo de pacientes e identificar determinados padrões na população que ajudem a fazer um rastreamento e uma estratificação, por exemplo, de alguns tipos de câncer, ou de doenças crônicas. [...] Então, essa possibilidade de ter ferramentas que ajudam nesse processo de identificar dentro de um contexto muito grande as coisas mais críticas, acho que essas ferramentas são muito úteis. [...] Quem está na gestão e tem que tomar decisões, o dinheiro sempre falta, a população é muito grande. Então, para onde que a gente vai? [...] As possibilidades [oportunidades] são enormes.

(STAKEHOLDER *ACADEMIA*)

É [o potencial da] visão populacional, visão que a gente pode estar alcançando maior quantidade de pessoas num tempo reduzido.

(STAKEHOLDER *PODER PÚBLICO*)

[...] é a oportunidade de detecção de surtos e epidemias de patógenos com potencial epidêmico ou pandêmico; acho que não vai ser a primeira vez que você vai ouvir hoje, mas o Brasil produz dado para todo lado. O Ministério da Saúde tem mais de 800 bancos de dados diferentes de aspectos, que vão de clínica e laboratório, passando por logística, insumos que a sua grande maioria não se comunica entre eles. Nesses dados, a gente tem uma riqueza muito grande de informações que nos permitem detectar, por exemplo, o início de uma epidemia, de um patógeno, antes mesmo que os sistemas oficiais de vigilância sejam capazes de fazer [...].

(STAKEHOLDER *ACADEMIA*)

Fatores favoráveis para consolidação de um sistema integrado de dados no Brasil

Vários entrevistados destacaram que o Brasil apresenta um cenário especialmente favorável para o uso da IA na saúde em razão do grande volume de dados, pois permitiria a construção de ferramentas baseadas em uma massa muito robusta e diversa de informações. Essa percepção está conectada à grandeza populacional e territorial do país, à diversidade genômica que nos é característica e à existência de um sistema único de saúde pública que atende um volume muito grande de usuários e registra sistematicamente suas informações. Essas características do país seriam favoráveis ao desenvolvimento de algoritmos consistentes de aprendizado de máquina. Além disso, como mencionado no tópico anterior, os sistemas e repositórios de dados de saúde do SUS são avaliados com grande potencial de integração visando a interoperabilidade.

O Brasil é um país de 210 milhões de habitantes e é o único país com mais de 100 milhões em um sistema único de saúde subfinanciado, mas estruturado. Ninguém tem isso. Nós somos um laboratório para o mundo. Não é só por causa da nossa genética, é porque nós temos esse sistema de dimensões continentais. [...] Para essas ferramentas, tem que ter integração, então integração para nós é oportunidade.

(STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

A saúde [no Brasil] é muito rica em dados, dados da vida do cidadão, tanto no SUS quanto na saúde complementar, que é a saúde privada. Esses dados, com as devidas pesquisas e com os devidos olhares, não só da TI, mas o olhar do profissional de saúde, eles podem fazer muito pelo cidadão. Eles podem antecipar situações de saúde, evitando realmente maiores problemas e reduzindo custos. [...] Então, são muitos casos que a gente pode usar na saúde, muito nesse sentido de predição mesmo, do que o cidadão pode vir a ter para que ele evite que tenha.

(STAKEHOLDER *PODER PÚBLICO*)

> Uma vantagem é que a gente tem um SUS, que é o nosso Sistema Único, é unificado e ele é muito grande e tem muito dado. Então, eu acho que, perto de outros países, a gente tem um potencial de ter o banco de dados de saúde no SUS, tem um potencial de ser gigantesco, um dos maiores do mundo em qualidade e em quantidade de dados, então eu acho que isso e eu sei que está avançando bastante, tem umas iniciativas estratégicas de organização dos dados do SUS, do DATASUS que estão avançando bastante. Então, eu acho que isso é uma oportunidade. [...] O SUS tem esse potencial de enxergar tudo, desde o dado assistencial até dados administrativos, mas ainda tem muito chão para a gente tornar esse dado acionável, utilizável.
>
> (STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

> O [nosso] sistema de saúde é o maior gerador dados do mundo, só que a gente não usa esses dados. Então, não precisa gerar dados novos, pegar o dado que já está lá e transforma.
>
> (STAKEHOLDER *MERCADO*)

Ainda na temática dos dados, dois outros aspectos foram apontados pelos entrevistados como potencialidades particulares da realidade brasileira, mesmo que de forma menos frequente e enfática. São eles a existência da Rede Nacional de Dados em Saúde (RNDS) (MS, s.d.) e da Lei Geral de Proteção de Dados (LGPD) (Lei n. 13.709/2018).

BOX 1 – REDE NACIONAL DE DADOS EM SAÚDE

A RNDS é a plataforma nacional de interoperabilidade em saúde, instituída em 2020, com o objetivo de promover a troca de informações entre os pontos da rede de atenção à saúde, permitindo a transição e a continuidade do cuidado nos setores públicos e privados. Sua constituição é parte estruturante do Conecte SUS, programa voltado a transformação digital da saúde do país, e está conectada à ESD. No website da RNDS do MS, é possível acessar mais detalhes sobre as funcionalidades e os potenciais benefícios da rede (https://www.gov.br/saude/pt-br/composicao/seidigi/rnds).

RNDS

BOX 2 – LEI GERAL DE PROTEÇÃO DE DADOS

Instituída pela Lei n.13.709, de 14 de agosto de 2018, a LGPD apresenta as diretrizes de tratamento de dados pessoais no Brasil, inclusive nos meios digitais, com o objetivo de proteger os direitos de liberdade e privacidade dos cidadãos e instituições.

LGPD

A LGPD foi apontada pelos entrevistados como um avanço regulatório que permitiria um bom desenvolvimento e aplicação das ferramentas de IA no Brasil. Com esse arcabouço regulatório inicial estabelecido, acredita-se que as aplicações podem começar a ser desenvolvidas, testadas e aplicadas. Por outro lado, em contextos nacionais em que esse tipo de regulação sobre uso de dados não foi instituído ainda, a evolução da IA ficaria prejudicada.

Já a RNDS foi apontada como uma estratégia positiva, com potencial para promover a interoperabilidade de dados no futuro. A perspectiva de integração de dados e sistemas no horizonte, tal como colocada pela rede, pode ser considerada um potencial do país para a evolução da IA aplicada à saúde.

> Olhando para frente, a gente está em um cenário bastante promissor [...] porque estamos finalmente trazendo a área do conhecimento da informática em saúde. [...] Porque, hoje, a gente tem a perspectiva da RNDS, e a interoperabilidade entre o público privado e entre todas as esferas de gestão.
>
> *(STAKEHOLDER PODER PÚBLICO)*

Além dessas potencialidades mais relacionadas a particularidades do contexto brasileiro, as entrevistas coletaram muitas percepções sobre oportunidades da IA aplicada à saúde no geral. Assim, foram recorrentes as falas sobre IA como tecnologia que pode promover a solução de problemas e a melhoria de diversos aspectos da saúde no país, mas não exclusivos à realidade brasileira. De modo geral, essas falas apontam os grandes potenciais da IA na administração e na parte operacional, por um lado, e na realidade clínica e de cuidado ao paciente, por outro.

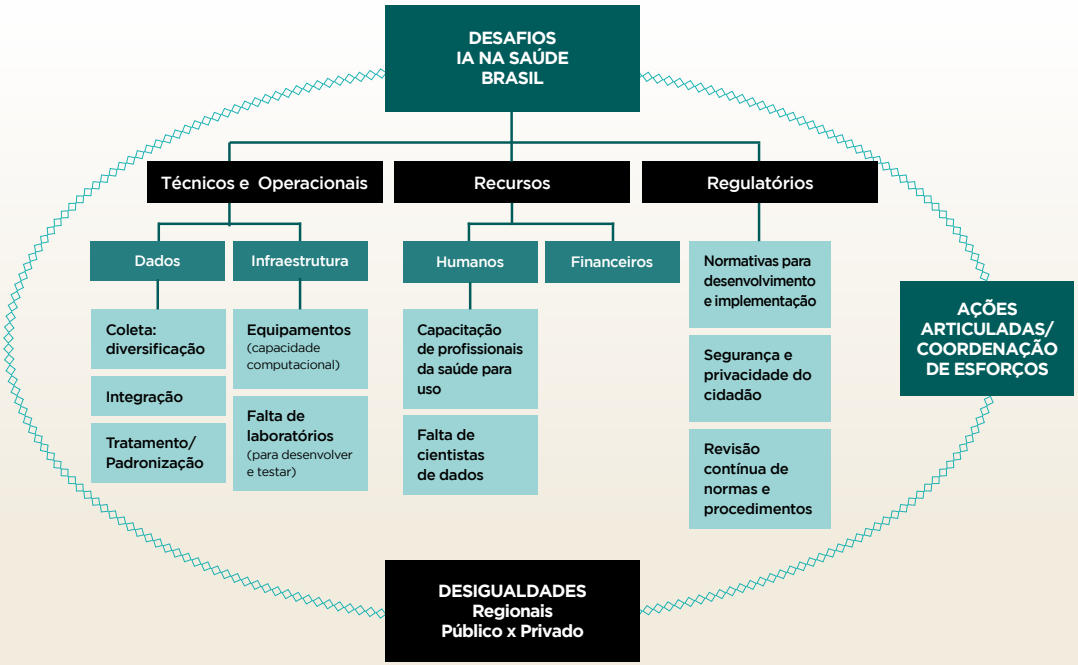
BOX DESTAQUES - OPORTUNIDADES PARA IA NA SAÚDE

- Otimismo quanto ao uso da IA na saúde brasileira: identificação de oportunidades em diversos setores e processos da gestão e do atendimento clínico.
- Expectativa de que a IA possa minimizar as desigualdades territoriais de acesso à saúde de qualidade, típicas do contexto nacional.
- Oportunidades para o paciente: ampliação da oferta de serviços de saúde, melhoria no acesso e precisão de diagnósticos, e conexão de pacientes a conhecimento especializado, especialmente em regiões remotas.
- Oportunidades para prestadores de saúde: otimização de fluxos de trabalho internos; melhoria da gestão de leitos, filas de espera e logística hospitalar; aceleração de laudos de exames por imagem.
- Oportunidades para profissionais de saúde: redução do tempo gasto em tarefas burocráticas, ferramentas de IA para análise e interpretação de prontuários médicos, suporte à tomada de decisão médica.
- Oportunidades para saúde pública e vigilância sanitária: monitoramento epidemiológico; identificação de mudanças nos padrões de incidências de doenças; prevenção, planejamento e contenção de riscos.
- Características únicas do contexto brasileiro que produzem um contexto rico em oportunidades para a IA na saúde: grande volume de dados disponíveis, existência de sistemas de saúde pública integrados, regulação existente sobre proteção de dados (LGPD) e potencial integração de dados via RNDS.

DESAFIOS PARA O DESENVOLVIMENTO DA IA NA SAÚDE

Nesta seção, serão apresentadas as percepções dos *stakeholders* a respeito de gargalos, desafios e barreiras para o desenvolvimento da IA no setor da saúde no Brasil. Além disso, buscamos explorar as especificidades brasileiras em torno desse tema, comparando-as com outros países e contextos. Serão discutidos os desafios relativos a aspectos técnicos e operacionais (dados e infraestrutura), recursos humanos e financeiros, questões regulatórias e ações de coordenação, que foram majoritariamente destacados nas entrevistas como os principais desafios para o país. Ao analisar os desafios, é possível identificar áreas prioritárias para intervenção e desenvolvimento de estratégias que possam acelerar a adoção de tecnologias de IA, adaptadas a particularidades e necessidades do sistema de saúde brasileiro. A Figura 3 apresenta uma visão geral dos desafios identificados.

FIGURA 3 – DESAFIOS PARA O DESENVOLVIMENTO DA IA NA SAÚDE NO BRASIL



FONTE: ELABORAÇÃO PRÓPRIA.

Técnicos e operacionais

De acordo com entrevistados de diferentes perfis, e como mencionado nos tópicos anteriores, uma das principais barreiras para o avanço da IA na área da Saúde é a falta de dados de qualidade e/ou a dificuldade de integração de grandes bases de dados, uma vez que os dados são essenciais ao treinamento eficaz de algoritmos capazes de proporcionar o bom desempenho das ferramentas de IA. Contudo, para além do volume, o fator imperativo, nesse caso, trata-se da qualidade dos dados.

De modo geral, os entrevistados foram enfáticos ao afirmarem que o país carece de uma coleta sistematizada que forneça dados estruturados e integrados. Para que isso ocorra, é

necessário o estabelecimento de protocolos bem definidos de coleta e armazenamento, levando em conta, inclusive, a dimensão territorial do Brasil.

O nosso grande gargalo é ter dados para treinar esses algoritmos.

(STAKEHOLDER *ACADEMIA*)

Com os dados abertos eu sei que a pessoa tomou um medicamento e teve um exame bioquímico pago para aquilo, só que eu não sei o resultado daquele exame bioquímico, o que seria fantástico se eu soubesse... Eu não sei se a pessoa morreu, se ela parou de tomar a medicação porque ela parou ou porque ela morreu. Só que, na verdade, esses dados estão todos integrados. Enfim, integração de dados é o primeiro [desafio].

(STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

Outro ponto crítico mencionado em relação a esse tema estratégico é a falta de uniformidade e padronização nos dados de saúde do país. Os *stakeholders* entendem que a enorme diversidade nos sistemas de informação, inclusive nas terminologias empregadas, dificulta a integração eficiente de dados, comprometendo a interoperabilidade e a capacidade de utilizá-los para treinar modelos de IA. Apesar de algumas iniciativas citadas, como a RNDS e o e-SUS Atenção Básica, é consensual a percepção de que a efetiva integração de dados clínicos ainda é um grande desafio, comprometendo a utilidade desses dados para aplicações de IA.

Hoje, para mim, é a coleta e tratamento de dados. Isso é o maior gargalo. É você conseguir juntar uma quantidade de dados grande e com qualidade, porque não basta você ter o dado. Você tem que ter o dado, digamos, anotado, muito bem-organizado do jeito com o que você quer predizer. Vamos supor, eu quero fazer uma IA para diagnosticar câncer de mama em mamografias. Não basta eu pegar lá no SUS e falar: "me dá todas as mamografias que estão armazenadas no SUS". Eu só vou ter um monte de foto de mamografia! Eu não sei quem tem câncer, quem não tem câncer. Eu não sei se aquela imagem corresponde ao câncer que foi do tipo "a" na biópsia, do tipo "b" na biópsia. Então, para eu treinar o modelo, não é só ter as imagens, eu tenho que ter a imagem e saber: essa é uma imagem normal, essa é uma imagem com câncer, esse foi o câncer do tipo tal, esse foi um câncer de quem morreu rápido, esse aqui teve uma sobrevida muito alta, muito grande até o prognóstico. Então, para eu criar os modelos de IA, eu preciso ter o dado e um monte de informação de toda a jornada daquela coisa e isso não é fácil, porque na saúde os dados estão todos em silos, em caixinhas. O dado de imagem está no sistema de armazenamento. O prontuário do paciente está em outro. O dado de prontuário não é estruturado, é um texto corrido. As pessoas anotam de um jeito diferente. Como é que eu transformo isso numa tabelinha super estruturada, um dado organizado? Isso daí é, de longe, o maior gargalo na saúde para você ter ferramentas de fato robustas.

(STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

O tema da infraestrutura também foi mencionado pelos entrevistados como um obstáculo para o avanço da IA no Brasil. Entre os desafios citados, destacou-se o elevado custo computacional das soluções, que as tornam financeiramente inacessíveis em muitos casos. Analisando o país como um todo, a infraestrutura existente é considerada inadequada e insuficiente para suportar as novas demandas dessas tecnologias.

Há também, nesse ponto, a percepção de desigualdade, pois os recursos tecnológicos não são distribuídos de forma homogênea entre diversos contextos locais e entre setores público e privado. A falta de laboratórios especializados para desenvolver e testar as aplicações de IA na saúde foi apontada como uma barreira, especialmente no sistema público de saúde brasileiro. Na área privada, existem alguns centros de excelência, com laboratórios que, embora se destaquem em termos de infraestrutura, enfrentam dificuldades para implementação dessas tecnologias em escala.

A escassez de recursos financeiros, humanos e tecnológicos é identificada como um desafio significativo da realidade brasileira, ao mesmo tempo que também pode ser observada em outros contextos nacionais. Assim, as entrevistas sinalizam que a implementação bem-sucedida da IA na área da saúde no Brasil demandará esforços coordenados e investimentos em diversas frentes, incluindo dados, infraestrutura e recursos humanos.

Recursos

Outro grande desafio relaciona-se à escassez de recursos humanos preparados e com habilidades digitais necessárias para implementar e utilizar a IA na área da saúde. Essa lacuna apresenta múltiplos aspectos: algumas entrevistas mencionam a dificuldade dos profissionais de saúde na familiarização com as ferramentas de IA, enquanto outras ressaltam a complexidade para atrair cientistas de dados especializados. Algumas entrevistas abordaram ambas as dificuldades da realidade brasileira, sublinhando a carência de profissionais capacitados como um ponto crítico na adoção efetiva da IA na saúde brasileira.

Eu acho que nós temos profissionais de TI bons, mas infelizmente tem caído o padrão. [...] A pandemia trouxe uma inflação nos salários dos profissionais de TI, e aí, hoje, você vê jovens com pouco conhecimento e experiência, e com altos salários. E aí, eles não querem aprender mais, eles acham que aquilo já é suficiente, que eles sabem por que eles estão bem remunerados por agora. A dificuldade que a gente tem aqui para encontrar profissionais, engenheiro de dados, cientistas de dados é grande. E a gente não restringe geograficamente, porque hoje eu já consigo, obviamente, trabalhar com muitos profissionais remotos. Então, eu falo de Brasil, por exemplo, e a dificuldade de encontrar profissionais, cientistas de dados, engenheiro de dados é gritante.

(STAKEHOLDER *PODER PÚBLICO*)

Várias entrevistas abordaram a questão da escassez de mão de obra técnica qualificada como um gargalo relevante, salientando, particularmente, que o mesmo profissional necessário ao desenvolvimento da IA na área de saúde pode atuar em outros segmentos, como o financeiro, que oferece melhor remuneração. Assim, a atração e a retenção de profissionais especializados no desenvolvimento de IA na área da saúde são consideradas pelos entrevistados um desafio no contexto brasileiro.

Você precisa de recursos computacionais e pessoas especializadas. Talvez a gente tenha muitos gargalos para isso no país. Hoje, as faculdades, principalmente matemática e computação que formam cientista de dados, o número de profissionais formados, mesmo na engenharia, formados com competência para isto, não atendem à demanda do mercado. O mercado financeiro absorve uma boa parte desse pessoal, então falta mão de obra para poder trabalhar com essas coisas de uma maneira crítica. De novo, não é consumir um enlatado, uma tecnologia que alguém treinou lá fora e simplesmente ligar e começar a usar no seu hospital. Você precisa validar isto, precisa ter pessoal que critique e faça a avaliação de maneira consistente.

(STAKEHOLDER *ACADEMIA*)

É importante registar que a falta de cientistas de dados não é uma exclusividade brasileira. Esse ponto foi destacado como um problema mundial. A demanda aumentou muito, de forma muito rápida, sem que os países conseguissem formar pessoas para supri-la. De toda forma, os entrevistados acreditam que há países enfrentando essa questão de forma mais estratégica, com alocação sistemática de recursos em IA. No Brasil, os entrevistados identificam apenas ações pontuais de agências de fomento à pesquisa tecnológica – como a Financiadora de Estudos e Projetos (FINEP) e a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP); contudo, essas ações, por si só, não configuraram uma estratégia abrangente e de caráter nacional.

> Nós não temos uma infraestrutura nacional para processamento de alto desempenho para treinar algoritmos. As fundações de apoio e amparo a pesquisa nos estados têm colocado editais para que grupos de pesquisa possam desenvolver, a própria FINEP tem colocado alguns editais. Mas isso, para a demanda que a gente tem, talvez seja muito pequeno. A gente ficou com alguns gargalos importantes que certamente vão reduzir a nossa competitividade, a gente vai ter que correr para poder recuperar esse possível atraso. A falta de recursos para isso acho que é muito importante, o país não tem uma estratégia em IA. Vários países, por exemplo, os Tigres Asiáticos, lá eles têm claramente onde eles vão colocar os recursos para IA. Então, tem um planejamento estratégico, eu não vejo isso ainda no país, a menos dessas chamadas das fundações de amparo a pesquisa, FINEP, mas são coisas muito pontuais. Não é uma estratégia. Acho que o país deveria ter uma estratégia nacional para isto, como por exemplo, os dois grandes, Estados Unidos e China, eles têm uma estratégia muito clara em IA.
>
> (STAKEHOLDER *ACADEMIA*)

Porém, as entrevistas também revelam que o Brasil enfrenta obstáculos particulares à superação do problema de escassez de mão de obra qualificada. Cabe reiterar que, no país, os desafios enfrentados no âmbito dos dados e dos recursos humanos também estão relacionados a algumas especificidades da realidade brasileira. Uma delas seria a desigualdade territorial, que tornaria ainda mais complexa a implementação de sistemas de IA na área de saúde. Os entrevistados percebem que há uma substantiva disparidade entre contextos estaduais e municipais, tanto em recursos disponíveis quanto em termos de prioridade para a inovação. Contextos municipais e estaduais mais desfavorecidos enfrentam desafios substantivos para a contratação de profissionais de saúde elementares, como médicos e enfermeiros, o que tornaria inviável priorizar a contratação onerosa de profissionais de TI. A discrepância nos salários oferecidos a engenheiros de dados em outros setores econômicos, mais atrativos do que na saúde, ilustra o desafio desse segmento na atração de profissionais especializados em TI. Na gestão governamental e na saúde pública brasileira, esse desafio seria ainda maior. Os entrevistados acreditam que, devido à escassez de recursos, o setor público tende a não priorizar a contratação de especialistas para implementar soluções baseadas em IA. Assim, a falta de profissionais qualificados nessa área estaria intrinsecamente ligada às desigualdades existentes no país, abrangendo questões salariais e prioridades na alocação de recursos públicos.

A citação a seguir ilustra claramente como a carência de recursos humanos qualificados está intimamente ligada à desigualdade entre municípios. Essa disparidade é típica

do contexto brasileiro, em que a distribuição desigual de profissionais de saúde e outras áreas essenciais resulta em grandes diferenças na qualidade e na disponibilidade dos serviços oferecidos. Municípios menores e mais distantes frequentemente enfrentam maiores desafios, exacerbando as disparidades regionais e limitando o acesso igualitário às inovações tecnológicas e aos cuidados de saúde de qualidade.

O secretário de saúde de um município do entorno de São Paulo me mostrou uma ferramenta e falou "poxa, eu queria desenvolver essas coisas lá no município, como eu faço?". E eu falei "o instituto desenvolve gratuitamente e leva lá para você". Ele falou "nossa adorei, eu queria ter um cabra que fizesse essas coisas lá, mas, quando eu tentei contratar, ele me custaria 25 mil reais". Por 25 mil reais, eu contrato 5 enfermeiras, e, aí, eu contrato um cara para ficar na frente do computador fazendo essas coisas ou 5 enfermeiras que vão ajudar os pacientes?

(STAKEHOLDER *ACADEMIA*)

Regulatórios

Além dos aspectos apresentados, a questão da regulação para o uso de IA no setor da saúde também surgiu nas entrevistas como um desafio de forma recorrente. A percepção de que faltam normativas e diretrizes para o desenvolvimento e a aplicação de ferramentas de IA na saúde é uma preocupação compartilhada por muitos especialistas e profissionais da área. A rápida evolução da tecnologia de IA e seu potencial impacto na prestação de cuidados de saúde levantam questões éticas, legais e de segurança que ainda não foram enfrentadas no Brasil. Nesse sentido, a ausência de regulação pode resultar em lacunas em qualidade, segurança, confiança e eficácia das aplicações de IA na saúde. Ademais, a falta de diretrizes específicas pode prejudicar a consistência e a confiabilidade dos resultados dessas ferramentas, aumentar o risco de ocorrência de vieses algorítmicos que podem ter implicações sociais negativas, a falta de transparência e a iniquidade no acesso aos cuidados de saúde. Essas preocupações refletem a necessidade contínua de desenvolver e atualizar marcos regulatórios, à medida que a IA na saúde continua a evoluir, garantindo que seus benefícios sejam maximizados de maneira ética e responsável.

A tecnologia que está começando, então ela ainda vai ser muito regulada para frente, né? Acho que a regulação principal é essa, é você não deixar usar os dados contra o cidadão. Você garantir a anonimização de dados, essas coisas que precisa ser feito. Eu acho que era importante criar uma regra para você mandar dados para fora do país. Isso é uma coisa que precisava ser mais estruturado. Você pode mandar, mas precisa ser anonimizado, precisa ter um contrato, precisa saber o que vai. Precisa ter um mínimo de estruturação.

(STAKEHOLDER *ACADEMIA*)

Existe um movimento, existe discussão [sobre regulação de IA], existe discussão inclusive em esferas governamentais, de fato. Então, não é que tem nada acontecendo, não, tem bastante coisa acontecendo, mas tem várias coisas que estão sendo discutidas novamente em ilhas. Então, tem ilhas de movimentos que estão discutindo, seja dentro da universidade ou outros grupos de pesquisa, né, outras instituições não governamentais fazendo essas discussões. Mas está sendo todo mundo chamado para as discussões aonde, de fato, a decisão vai ser tomada? Não está.

(STAKEHOLDER *ACADEMIA*)

Eu preciso ter uma gestão mínima do que se pode fazer com esses dados, principalmente os dados que são gerados [pela IA], novos dados baseados nos dados que você já tem.

(STAKEHOLDER *MERCADO*)

Ainda no tema da regulação, a LGPD foi citada em diferentes perspectivas. Conforme observamos na subseção anterior, ela foi apontada por alguns entrevistados como um avanço regulatório que permitiria um bom desenvolvimento e a aplicação das ferramentas de IA no Brasil. Todavia, um grupo diminuto de entrevistados, principalmente do mercado, consideram-na uma preocupação: acreditam que a LGPD é um arcabouço normativo que pode gerar interpretações variadas e limitar os processos de inovação; além disso, o desconhecimento da população sobre a LGPD pode gerar resistência ao compartilhamento de dados, o que, por sua vez, prejudicaria inovações na utilização de IA.

A falta ou a preocupação do compartilhamento do dado de saúde. Você pode fazer uma pesquisa, pegar 10 pessoas e falar sobre o que cada uma acha da LGPD. Dependendo do nível dessas pessoas, muitas não vão saber nem do que se trata. Um outro grupo vai saber do que se trata, mas não liga. E um outro grupo vai ser extremamente preocupado, achando que estamos com o Big Brother, que o cara vai ter acesso à vida toda, a vida dele inteira. Então, você tem isso aqui ainda porque as pessoas não conhecem. [...] Então, eu diria que as diferentes interpretações da LGPD, para mim, é um gargalo para inovações, para poder acelerar e entregar mais coisas, entende?

(STAKEHOLDER *MERCADO*)

Em resumo, de acordo com a maior parte das entrevistas, a ausência de normativas claras para o desenvolvimento e a aplicação de IA na saúde contribui para a incerteza legal e

ética, exigindo uma governança mais efetiva para validação, acompanhamento e fiscalização dessas aplicações.

Articulação

Entre os desafios mencionados pelos entrevistados sobre a realidade brasileira, destacam-se ainda a descentralização e a desarticulação de ações e políticas sobre o tema. A falta de estratégias abrangentes, de caráter nacional e com articulação entre atores, segmentos e setores público e privado seria um gargalo importante no contexto brasileiro. Esse aspecto foi mencionado em diferentes momentos das entrevistas, apontando que as articulações necessárias para inclusão de novas tecnologias na prestação de serviços de saúde se tornam um desafio significativo no Brasil. A falta de uma visão estratégica clara e de uma atuação conjunta eficiente, portanto, pode resultar em atrasos consideráveis na superação dessas barreiras, especialmente na saúde pública.

Enquanto serviços privados podem incorporar tecnologia de forma mais ágil, os serviços públicos de saúde frequentemente enfrentam dificuldades devido a procedimentos excessivamente rigorosos, falta de recursos e arranjos complexos de governança e operação, envolvendo os governos federal, estadual e municipal. Desse modo, a organização tripartite da saúde pública implica a necessidade de integração e articulação entre diferentes atores e níveis de governo para a adoção transversal de novas tecnologias, considerando elementos como prioridades locais e divergências políticas.

Assim, os entrevistados entendem que a superação desses desafios requer não apenas investimentos financeiros, mas também a construção de uma abordagem estratégica para esse tema, a qual deve considerar os arranjos federativos, as desigualdades territoriais, as diferenças regionais e a diversidade de prioridades locais. A implementação efetiva de tecnologias avançadas na área de saúde no Brasil exige, por conseguinte, uma visão abrangente, que considere não apenas a tecnologia em si, mas também o contexto social, político e econômico de cada região de um país de proporções continentais.

> Como é que você implanta uma ferramenta altamente tecnológica em municípios que têm situações graves, que o cara vai estar muito mais preocupado com a água que as pessoas estão bebendo do que essas novas tecnologias. Então, eu vejo assim: articular nessa configuração tripartite [do SUS] a inclusão dessas novas tecnologias é um desafio muito grande. Mas isso é em composição da nossa estrutura de estado que eu acho que é boa, republicana, tem essa independência, né? Essa atuação conjunta, mas isso, se ela não tiver uma visão estratégica clara de futuro é um gargalo, coisas que poderiam ser, assim, serem superadas com muita velocidade, demora muito, muito tempo, entende? Você vê, às vezes você pega um município pequeno que é um polo de saúde, lá um hospital privado ele tem a mesma qualidade de um hospital privado de um grande centro urbano, mas o serviço público de saúde não. Entende? Porque o cara do privado ele consegue colocar tecnologia lá, e o público não consegue porque isso requer uma série, um esforço conjunto, recurso financeiro e, aí, uma pactuação bipartite, uma situação com o estado e tem a questão das triangulações partidárias, enfim aí, essa composição não gera.
>
> (STAKEHOLDER *PODER PÚBLICO*)

Esta seção apresentou as percepções destacadas nas entrevistas a respeito dos gargalos, desafios e barreiras para o desenvolvimento da IA no setor de saúde no Brasil. Pontos como a falta de dados de qualidade, a escassez de recursos humanos especializados e a ausência de normativas claras são obstáculos substanciais. Além disso, o contexto brasileiro é marcado por desigualdades territoriais que dificultam a aplicação capilarizada e integrada dessa tecnologia. Ademais, a falta de uma visão estratégica nacional e de uma atuação conjunta entre os setores público e privado retarda a incorporação dessas tecnologias na prestação de serviços de saúde. Por fim, a ausência de normativas e diretrizes para o desenvolvimento e aplicação da IA na saúde é uma preocupação compartilhada, a qual reflete a necessidade de marcos regulatórios mais robustos que, ao mesmo tempo, incentivem e regulem as iniciativas e as aplicações dessas ferramentas.

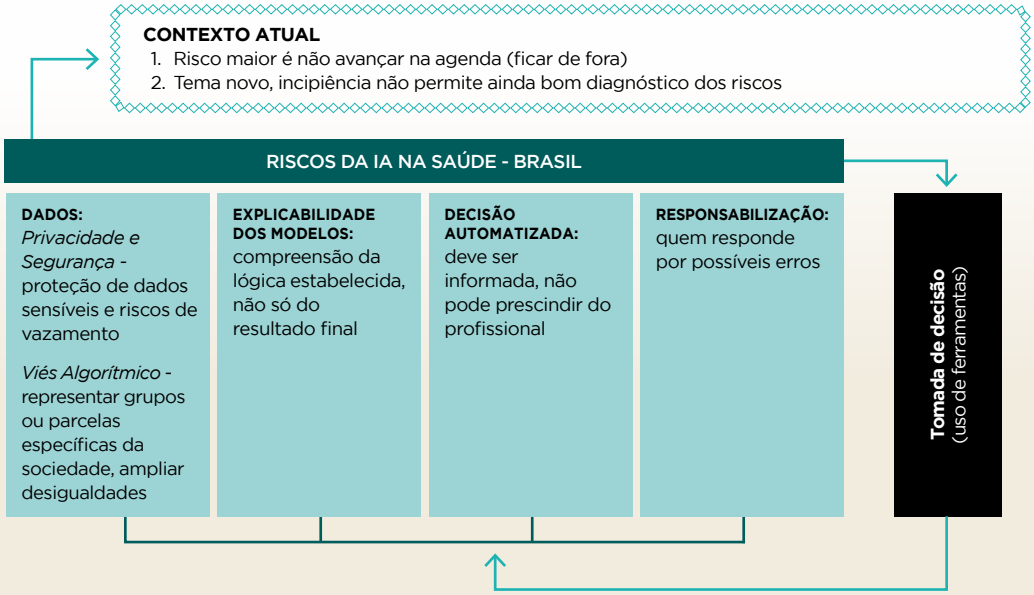
BOX DESTAQUES – DESAFIOS PARA DESENVOLVIMENTO DA IA NA SAÚDE

- Disponibilidade e qualidade dos dados: uma barreira significativa para o avanço da IA na saúde no Brasil. Falta de uniformidade e padronização nos dados de saúde dificultam a integração e a interoperabilidade. Entrevistados destacam a necessidade de coleta sistematizada e padronizada de dados, especialmente em regiões mais isoladas.
- Infraestrutura inadequada e insuficiente. Elevado custo computacional das soluções de IA, tornando-as financeiramente inacessíveis. Falta de laboratórios especializados, especialmente no sistema público de saúde brasileiro.
- Escassez de recursos humanos qualificados. Dificuldade na familiarização dos profissionais de saúde com ferramentas de IA. Dificuldades para atrair cientistas de dados especializados. Carência de profissionais capacitados como um ponto crítico na adoção efetiva da IA no setor de saúde.
- Desigualdades territoriais e de recursos: dificuldade na implementação de IA devido às diferenças e desigualdades entre contextos estaduais e municipais. Disparidade entre serviços privados e públicos na incorporação das tecnologias de IA.
- Regulação e governança. Ausência de normativas claras para o desenvolvimento e a aplicação de IA na saúde criam incerteza legal e ética e exigem uma governança mais ativa e efetiva.

RISCOS DO USO DE IA NA SAÚDE

Até este ponto, tratamos de avanços, lacunas, oportunidades e dificuldades relacionadas ao uso da IA no setor da saúde no Brasil. A partir desta seção, apresentaremos as visões sobre possíveis riscos para o uso da IA no setor da saúde. Buscamos identificar os principais riscos percebidos pelos entrevistados e as questões que, para eles, deveriam ser antecipadas e enfrentadas para que o país possa desfrutar dos benefícios dessa tecnologia. A Figura 4 apresenta uma visão geral dos elementos identificados.

FIGURA 4 – RISCOS DA IA NA SAÚDE NO BRASIL



FONTE: ELABORAÇÃO PRÓPRIA.

É importante registrar, de início, que os possíveis riscos não foram um tema de grande destaque nas narrativas. O assunto só surgia quando explicitamente estimulado, entretanto sem grandes aprofundamentos. Entre os entrevistados, há uma compreensão geral de que os benefícios da IA superam os possíveis riscos associados a seu uso; portanto, há maior atenção aos potenciais benefícios do que preocupação com os potenciais riscos.

É possível que essa visão menos atenta aos riscos tenha relação com o estágio inicial de desenvolvimento e adoção das tecnologias de IA na saúde. Em razão de o desenvolvimento dessa agenda ainda ser muito recente no Brasil, como expressaram alguns entrevistados: “estamos arranhando a superfície”, “estamos engatinhando”. Assim, pode haver dificuldade em antecipar os riscos futuros. Conforme ilustra a fala destacada a seguir, o desenvolvimento e a aplicação das tecnologias de IA no Brasil ocorrem sem um mapeamento

adequado dos riscos e, consequentemente, sem uma estratégia de contenção para esses mesmos riscos.

> Eu vejo que a gente está engatinhando ainda em analisar risco. É muito difícil, não é comum você encontrar alguém que tome decisões baseadas em riscos. O Brasil ainda está engatinhando. Assim com a LGPD, tem um alinhamento... Do que eu entendo da LGPD, ainda tem um alinhamento de qual, o quão eu vou realmente vou usar IA, qual é o paciente e o impacto dela [IA] na minha rotina. Então, eu diria que a gente ainda está engatinhando nessa questão do risco, porque o Brasil não é uma cultura baseada em risco. [...] A gente não tem uma escala que possa nos orientar se aquela aplicação é ou não alto nível de risco para o que ela se propõe a fazer. Eu não tenho escala, eu diria para você que não tem uma trilha ainda. A gente não tem um *workflow*.
>
> (STAKEHOLDER *MERCADO*)

Quando perguntados explicitamente sobre possíveis riscos associados ao uso da IA na saúde, alguns entrevistados discorriam novamente sobre barreiras para o desenvolvimento da tecnologia no país, partindo de uma compreensão de que o principal risco é não avançar nessa área, ou seja, o principal risco é "ficar para trás" nessa corrida. Assim, citavam a regulação para contenção de riscos como um entrave para o desenvolvimento e a aplicação da IA no Brasil. Para exemplificar a ideia de que a regulação pode representar um risco, um dos entrevistados da pesquisa mencionou o Projeto de Lei n. 2338/2023. que classificaria a IA na área da saúde como um sistema de alto risco, independente da aplicação.

> Essa definição de sistemas de alto risco, ele leva uma série de obrigações, de responsabilidades para quem passa a utilizar a IA. Ou seja, é um fator inibitório para a utilização, considerando que, se algo ocorrer, a instituição acaba sendo responsável. Então, sendo considerada de alto risco, isso vai levar a um tempo maior no desenvolvimento de sistemas, de algoritmos e também maiores custos, porque vai implicar em mais burocracia, mais supervisão, mais testagens, tudo... Ou seja, vai lentificar todo o processo de incorporação de IA. E realmente nós não enxergamos a IA como um sistema de alto risco, né? Nós enxergamos a IA como uma nova tecnologia, que tem que ser regulamentada, mas não com uma tecnologia de alto risco, ao contrário, pode trazer enormes benefícios para a área da saúde.
>
> (STAKEHOLDER *MERCADO*)

Dados: privacidade, segurança e viés algorítmico.

Apesar da dificuldade de visualização de riscos futuros, alguns pontos foram levantados de forma recorrente. Um deles foi a preocupação com dados, que apareceu com frequência entre as menções a riscos. As pessoas entrevistadas apontaram preocupação com mecanismos de proteção de dados e garantia

da privacidade de dados sensíveis. Assim, o risco de vazamento de dados é apontado como relevante.

O grande risco, na verdade, é o risco da privacidade, o risco da divulgação dos dados individuais de pacientes, principalmente em questões da presença de doenças mais delicadas e que não seria interessante a gente divulgar e esses dados vazarem. Mas isso já é algo que já é feito hoje, a coleta de dados já existe independente da IA, são coisas que tem evoluído de uma forma independente. A gente tem pouquíssima IA na prática clínica, porém você tem coleta de dados gigantesca desses pacientes. E é essa coleta de dados que é um dos grandes problemas que a gente tem hoje na saúde, que é o risco de vazamento desses dados individualizados, como eu falei, [que] independe da IA.

(STAKEHOLDER *ACADEMIA*)

Alguns entrevistados relataram que, embora as instituições geralmente atuem em conformidade com a LGPD na coleta e registro de dados, ainda existem brechas regulatórias que podem fragilizar a segurança das informações. Nesse sentido, lembraram que as ferramentas de IA que operam com dados coletados de acordo com a LGPD podem gerar novos dados, os quais podem ser utilizados em outros contextos. Há uma visão convergente entre os entrevistados de que o tratamento de dados coletados está razoavelmente bem regulado no país, principalmente pelo dispositivo da LGPD, mas a disponibilidade e os usos secundários desses dados precisam de regras mais claras. Enquanto isso não ocorre, haverá o risco de vazamento e/ou uso indevido dos dados.

A partir do momento que uma empresa terceira, ela, através dos teus dados, ela gera novos dados, de quem são aqueles dados? Eu sei que os dados do paciente estão bem regulamentados, mas esses novos dados que foram gerados a partir do paciente [na ferramenta de IA], de quem são esses dados? Então, a gente ainda não viu nenhuma estrutura que balize isso para a gente

(STAKEHOLDER *MERCADO*)

É melhor você ter uma regra restritiva, mas que ela seja clara do que não ter regra nenhuma. [...] Com uma solução [de IA], eu consigo gerar dados, entregar valor, aqueles dados podem ser úteis para outras soluções. Mas quem garante que eu vou utilizar da melhor maneira, né? Então, será que esse é um processo que deve ser realizado pelo hospital, pelo meu cliente? Eu não acho que deva, que cada cliente vai ter um processo diferente, uma gestão diferente desses dados. Então, será que pelo menos uma base a gente consegue regular para os meus próprios clientes se basearem nessa régua, né?

(STAKEHOLDER *MERCADO*)

O uso de banco de dados, essa busca frenética por banco de dados, né, pode muitas vezes envolver questões éticas muito sensíveis. Por exemplo, por quanto tempo esses dados vão ser guardados e reusados? De que forma vão ser reusados? Por que vão ser reusados? Você pode mudar a imagem, você anonimiza e depois você faz tantos cruzamentos que você acaba desanominizando e alguém pode passar a ser identificado e, aí, você pode ter acesso a informações delicadas, que podem eventualmente ter um caráter de privacidade e que podem mudar a imagem de uma pessoa, vamos dizer, após a morte dela, como é que a gente lida eticamente com isso?

(STAKEHOLDER *PODER PÚBLICO*)

De acordo com as entrevistas, o processo de desenvolvimento de algoritmos também envolve riscos que precisam ser enfrentados e mitigados. Embora o desenvolvimento de algoritmos precise ser robusto, com base em dados confiáveis e volumosos, é fundamental haver atenção nas etapas de testes e validações. Os vieses algorítmicos que podem surgir no desenvolvimento também representam um risco, devido à indução de interpretações equivocadas e excludentes.

Um dos informantes da pesquisa tratou os vieses algorítmicos como "risco malicioso", que se refere à possibilidade de modelos preditivos replicarem vieses sociais, de classe, etnia ou gênero presentes nos dados de treinamento. A correção desses vieses durante a fase de teste é fundamental para garantir que os sistemas de IA não ampliem as desigualdades na saúde.

Pode ser que o modelo para detectar, por exemplo, um determinado tumor, ele possa ir super bem em população branca e na população negra ele erre um pouquinho mais. Isso é uma inequidade, é como se fosse uma discriminação velada sem você perceber, porque os dados históricos de base em treinamento refletem isso. Existe essa discussão muito forte de discriminação histórica ou de representatividade populacional dentro de bases de dados para treinamento de IA e esse, por exemplo, é um potencial risco e você tem que ter mecanismos de conseguir checar e estar ativamente olhando. Ah, isso significa que a gente tem que parar e não podem mais fazer IA? Não, lógico que não. Na verdade, você tem que continuar fazendo, só que você tem que ter todo um time, uma maturidade de ciência de dados para conseguir mitigar e diminuir esses riscos. E, aí, a gente volta naquela história, como é que você mitiga esse risco? Se eu tenho dado representando todo mundo. Se o SUS, por exemplo, para representatividade não for uma fonte de dados forte para treinamento de IA, se só a iniciativa privada for organizando seus dados, você vai ver que as ferramentas vão refletir a distribuição e o comportamento das doenças na população rica, isso vai acontecer.

(STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

Os entrevistados destacaram que, se os dados utilizados para treinar a IA refletirem desigualdades preexistentes no sistema de saúde, a ferramenta de IA pode perpetuar essas disparidades. Por exemplo, se certos grupos populacionais

têm históricos defasados de atendimento, as tecnologias de IA dificilmente poderão auxiliar em diagnósticos e prognósticos sobre eles, pois, em decorrência da falta de dados desses grupos, não poderão ser treinadas sobre suas características deles. Desse modo, poderão funcionar como um instrumento que aprofunda situações de negligência previamente existentes. Por outro lado, se a IA for treinada com dados que destacam determinados grupos beneficiados, ela pode aprimorar ainda mais o atendimento a essas populações, intensificando a desigualdade entre diferentes segmentos da sociedade e acentuando também as desigualdades territoriais e regionais no Brasil. Portanto, enquanto parte da população pode experimentar benefícios substanciais, outros grupos podem ver acentuadas as lacunas existentes em seus cuidados de saúde. A seguir, observamos um conjunto de citações que discorre sobre os riscos de reprodução e/ou ampliação de desigualdades e iniquidades:

A gente pode aprofundar a iniquidade com relação ao acesso. A gente tem, por exemplo, a diferença entre o público e privado com o uso de um atendimento, vamos dizer assim, de uma medicina personalizada ou preditiva de maneira muito mais dentro do privado do que dentro do público, onde poderia ter uma coleta de informação muito maior. E como é que esses modelos, por exemplo, estão de fato sendo treinados numa população que não é a população brasileira, é um recorte, existe um recorte de classe com relação a essas bases da assistência privada.

(STAKEHOLDER *PODER PÚBLICO*)

Se os dados não forem representativos, de fato, onde aquilo vai ser aplicado, e quando falo que ele vai ser aplicado, eu não estou dizendo que tem que funcionar para tudo, não. Eu brinco dizendo que eu enxergo esses modelos como bula de remédio, ter eles como bula de remédio, onde a gente fala bem assim: "olha, isso aqui é indicado para isso, isso aqui é indicado para aquilo outro, então se for contraindicado para você, não usa". Então, a gente precisa saber quais são as contraindicações e as indicações em cada um desses sistemas, porque eles definitivamente não são para todo mundo, mas muitas vezes são colocados dessa forma. Então, tem uns riscos muito altos, dele estar sendo aplicado onde ele não deveria ser aplicado porque não vai funcionar e o sistema não vai responder e, aí, a gente aumenta as desigualdades que já existem nesse país.

(STAKEHOLDER *ACADEMIA*)

Garantir que vão ser treinados com dados em regiões mais remotas do Brasil que é onde se coleta menos dados, onde que se fazem menos exames, então assim, ter essa qualidade de dados porque senão você só treina o teu algoritmo em hospitais de regiões ricas, que mais coletam dados mesmo. Você corre o risco desse algoritmo aprender só auxiliar diagnóstico e prognóstico desses pacientes e aí, quando você vai para regiões com outras características, ele não funciona muito bem. Então, é fundamental que a gente tenha uma coleta sistemática desses dados, principalmente onde eles vão ser mais necessários, que é nas regiões mais remotas do Brasil.

(STAKEHOLDER *ACADEMIA*)

Explicabilidade dos modelos

Os entrevistados também destacaram como risco a ausência de protocolos obrigatórios para testes e validações das ferramentas. Nesse sentido, alguns *stakeholders* defenderam que o processo de desenvolvimento de algoritmos precisa ser transparente, passar obrigatoriamente por um conjunto de testes e ser validado por alguma instância neutra para ter um certificado ou selo e não funcionar como "remédio sem bula", nas palavras de um entrevistado.

> Eu acho que os sistemas de IA, eles precisam ser muito testados e validados. Não pode simplesmente você construir um sistema e dizer: está pronto e usa, e todo mundo simplesmente acreditar. Ele precisa ser testado, validado, passar por várias utilizações para adquirir essa confiança. Porque senão eu posso estar olhando uma informação que está errada. Então, eu penso isso, que essa credibilidade, ela vem realmente de muitos e muitos testes e validações de qualquer projeto de IA que seja.
>
> (STAKEHOLDER *PODER PÚBLICO*)

> Respondendo a sua pergunta em relação ao risco, é que falta uma meta ou um benchmark para que eu possa medir erros e acertos. Para mim, isso é um risco no uso de ferramentas de IA. Vamos imaginar que hoje, sem o uso de IA, eu faço uma conduta com aquele paciente e essa conduta tem se mostrado assertiva. Como é que eu sei que ela é assertiva? Porque eu faço um estudo e digo o seguinte, olha, toda vez que eu encaminho o paciente por essa trilha A, a chance de melhora é de 90%, enquanto a do B é de 80% ou de 50%. Agora vou voltar no uso da IA, pergunto para você, quem está medindo o uso da IA? Que a partir do uso da IA naquele direcionamento, eu realmente tive 90% de acerto? Acerto da melhora do paciente, do desfecho do paciente, não no algoritmo. [...] Então, é a falta de um *benchmark* e a falta de um modelo de medição de desfecho no uso. Para mim, é um risco. Quando você coloca o medicamento no mercado, qualquer que seja o medicamento, você tem uma série de fases para que aquele medicamento possa entrar no mercado. Na IA, o cara bota o negócio lá no computador ou no sistema dele e fica por isso mesmo, não tem como avaliar, não passa por nenhum tracing e ninguém segue nenhum modelo.
>
> (STAKEHOLDER *MERCADO*)

As lacunas de explicabilidade das ferramentas de IA também foram enfatizadas como risco em potencial, ainda que para um número mais limitado de entrevistados. Os que abordaram o tema entendem que o desenvolvimento e implementação das aplicações de IA na saúde não estão sendo acompanhadas de esforços de transparência e disseminação das decisões e conhecimentos básicos que embasam a construção dessas ferramentas. Para eles, sem compreender os processos e as decisões por trás da construção dos algoritmos, os usuários dessas tecnologias não teriam capacidade de avaliar e criticar as decisões pautadas no uso das ferramentas.

Acho que a governança dos dados é um risco, é um item que pode virar um risco para a ética. Essas tecnologias que hoje são desenvolvidas, como é que você tem a explicabilidade disso? Ou seja, como é que você: "ah, esse algoritmo foi treinado", mas quem treinou esse algoritmo? Com base no quê? Eu vejo assim, na ética inclusive tem uma metáfora, é como se ele tivesse um remédio que não tem bula, aí ele cura a dor de cabeça. Agora foi testado onde, por quem, quais são os *papers* que demonstram? Você tem na farmacêutica isso, mas hoje você não tem na IA. A gente foi treinado por tantos pacientes em tais lugares, por tal universidade, aqui está toda a rastreabilidade de quem testou ele, nome, pesquisador e tudo. Tudo isso não tem. Eu acho que isso é uma coisa muito, muito importante fazer isso.

(STAKEHOLDER *ACADEMIA*)

Mostrar o porquê a IA chegou a uma conclusão, isso ajuda muito, todo mundo que já fez escola, você vai na aula de matemática, se tem aquele problema enorme e a professora nunca aceita está aqui a resposta. Não, não, mostra como você chegou no resultado. Então, a gente está no momento onde todo mundo acha que a IA ela vai te dar 100% a melhor resposta que nem uma calculadora, mas na verdade ela vai ter que explicar como chegou lá, para quando errar você entender. [...] Você precisa mostrar o fluxo de pensamento e, a partir do momento que o mundo entender que a IA não te dá só o resultado, te dá o fluxo, você mitiga muito risco, porque você está acompanhando o racional da IA.

(STAKEHOLDER *MERCADO*)

Decisão automatizada

Outro risco apontado pelos entrevistados é o uso inapropriado das informações geradas pelos sistemas de IA de apoio à decisão. Eles destacam que é importante assegurar entre os usuários o entendimento de que esses dados não prescindem da avaliação do profissional de saúde, pois o benefício da IA nessa área não seria o de substituição de humanos, mas sim o de auxílio a tomada de decisão. A maioria dos sistemas deve gerar respostas que precisam ser avaliadas pelo profissional de saúde antes do fechamento de um diagnóstico ou encaminhamento. Assim, os profissionais precisariam de preparo para aproveitar os benefícios das ferramentas sem negligenciar suas limitações.

O risco maior é na transição para aplicação efetiva como ferramenta. Na minha percepção nós não temos maturidade no ambiente de saúde hoje para garantir que isso vai ser implementado de uma maneira adequada, uso acompanhado. Então, ali o risco que eu vejo é mais de mau uso mesmo, no sentido assim de você não garantir o controle de qualidade. A gente precisa criar um mecanismo de entendimento que esse tipo de ferramenta não é algo que você coloca e deixa lá. Ele tem que entrar dentro desse contexto que a gente entende que o que é ferramental e tecnologia para saúde tem que ter um acompanhamento. Implantação de controle de qualidade e manutenção.

(STAKEHOLDER *ACADEMIA*)

Responsabilização

Por fim, a ética, ainda que com pouca frequência, também foi um tema citado nos questionamentos sobre os riscos do uso de IA na saúde. Porém, quando aparece nas narrativas sobre riscos, observa-se que ela está relacionada ao indivíduo, à ética do profissional de saúde que utiliza uma ferramenta de IA. Logo, as balizas de ética para o uso dessa tecnologia não são consideradas construções coletivas, fruto de normativas e de discussões da sociedade, mas princípios de ética individual.

A discussão sobre ética e riscos suscita o debate sobre responsabilização, pois, no atual estágio de desenvolvimento da IA na saúde, ainda não está claro qual ator ou organização se responsabiliza por cada parte do processo de desenvolvimento e aplicação das ferramentas de IA. Assim, fica em aberto a pergunta de um dos entrevistados do estudo: "quem vai ser responsável por algum erro?". Tal questão complexa carece de problematizações e encaminhamentos assertivos da sociedade e da comunidade médica, mas, como as entrevistas sugerem, ainda é uma discussão pouco aprofundada.

Para te falar bem a verdade eu acho que depende muito da ética individual. [...] Eu acho que tomada de decisão individual ou na saúde pública com base em achados e evidências produzidas por máquinas – nada contra máquinas, elas nos servem –, mas elas precisam ser supervisionadas.

(STAKEHOLDER *ACADEMIA*)

O risco ético. Se a IA toma uma decisão errada, quem é que vai ser o responsável? Quem fez a IA? Vai ter que sempre um médico assinar, ou um profissional da saúde assinar. Isso tem que ser discutido antes desse dilema ético. Aí é a ética até de quem está treinando essa IA. Como é que você está treinando essa IA? Se você treina essa IA com resultados, por exemplo, você treina uma IA de raio-X de pulmão: os laudos que não são confiáveis, ela vai dar respostas que não são confiáveis. Então, como é que a gente faz uma avaliação prévia disso? Tem que ter uma regulamentação. Você está treinando com quem? Quem está fazendo esses laudos? Qual o nível de confiabilidade desses laudos? E aí, quem vai ser responsável por algum erro?

(STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

BOX DESTAQUES – RISCOS DO USO DE IA NA SAÚDE

- Consenso entre os *stakeholders*: os benefícios da IA na saúde superam os possíveis riscos, mas ainda há preocupações importantes que precisam ser abordadas para garantir um uso responsável e eficaz dessa tecnologia.
- Regulação em torno do uso da IA na saúde: considerada uma área de preocupação pelos diferentes perfis, mas com pluralidade de visões. Alguns entrevistados entendem ser necessário criar novos dispositivos de regulação para garantir a segurança e a confiança das ferramentas de IA na área da saúde, enquanto acreditam que ampliar o escopo de mecanismos regulatórios pode aumentar os custos e o número de exigências legais, associados ao desenvolvimento e à aplicação dessa tecnologia.
- Privacidade e segurança dos dados: preocupações com a proteção de dados sensíveis e o risco de vazamento de informações são destacados. Mesmo com a LGPD, acredita-se haver brechas regulatórias que podem comprometer a segurança dos dados, especialmente quando novos dados são gerados pelo uso de ferramentas de IA.
- Viés algorítmico: percepção de que existe o risco de algoritmos de IA reproduzirem e ampliarem desigualdades preexistentes no sistema de saúde. Se os dados utilizados para treinar a IA refletirem desigualdades, a ferramenta pode perpetuar essas disparidades, impactando negativamente certos grupos populacionais.
- Transparência e explicabilidade: poucos entrevistados mencionaram de maneira explícita que os resultados gerados por sistemas de IA devem ser explicáveis. Todavia, muitos apontaram que a falta de transparência nos processos de desenvolvimento de algoritmos de IA é um risco. A compreensão dos processos e decisões por trás das ferramentas de IA também foi lembrada como essencial para avaliar e criticar suas decisões.
- Uso inapropriado das ferramentas de IA: risco de os profissionais de saúde não compreenderem totalmente as limitações das ferramentas de IA e confiarem excessivamente em suas recomendações. É importante garantir que os profissionais sejam adequadamente treinados para usar essas ferramentas como auxílio à tomada de decisão, não como substituto.
- Responsabilidade ética: a falta de um arcabouço ético com dispositivos claros sobre responsabilização e orientações de como agir é vista como um risco para o uso da IA na saúde. Apesar de mencionarem esse risco, os entrevistados tenderam a se referir à ética em IA como resultado de práticas individuais, e não de construções coletivas.

TEMAS PRIORITÁRIOS PARA A AGENDA DE IA NA SAÚDE

Ao longo das subseções anteriores, apresentamos elementos que, juntos, constituem um panorama inicial do estado da IA na saúde brasileira. A fim de completar esse quadro, retomamos e sistematizamos, nesta subseção, como alguns temas-chave apareceram nas entrevistas realizadas. O termo "temas-chave", vale esclarecer, refere-se a assuntos que se destacam no debate[7] acadêmico e governamental sobre IA e Saúde e, principalmente, estruturam a ESD (MS, 2020) e são aderentes à agenda de IA. São eles: interoperabilidade, infraestrutura, recursos humanos, regulação, ética e direitos do usuário.

Dada a relevância desses temas, faz-se pertinente detalhar os procedimentos metodológicos utilizados para aferir a percepção dos entrevistados sobre eles. Nesse sentido, relembramos que o roteiro da entrevista foi composto por dois blocos, como exposto no capítulo "Notas metodológicas" desta publicação. O primeiro bloco previa perguntas sobre o contexto brasileiro de desenvolvimento da IA na saúde; nesse momento, buscou-se coletar percepções sobre o estágio atual, as oportunidades, as potencialidades, os desafios e os riscos do uso de ferramentas de IA. Nesses estímulos iniciais das entrevistas, surgiram diversas falas espontâneas sobre os seis temas-chave da ESD na maioria das entrevistas. Já em um segundo bloco de perguntas, o roteiro explorou explicitamente perguntas sobre temas relacionados a diretrizes da ESD.

Para otimizar o tempo das entrevistas e priorizar a exploração de temas não abordados pelos entrevistados, estabeleceu-se a seguinte estratégia de coleta: os temas infraestrutura, recursos humanos e direito dos usuários foram estimulados apenas quando não citados espontaneamente na entrevista até então, enquanto os temas interoperabilidade, regulação e ética foram estimulados, com o objetivo de aprofundamento, mesmo quando já citados ao longo da entrevista. A partir dessa dinâmica, muitas vezes, novas reflexões foram formuladas;

7 Para discussão sobre os temas mais discutidos na literatura, ver o artigo "Inteligência Artificial na saúde: uma visão da literatura e diretrizes para o Brasil", de Rodrigo Brandão, presente na Parte 1 – Artigos desta publicação.

assim, os temas abordados nesse tópico podem retomar alguns assuntos dos tópicos anteriores. O intuito é sistematizar brevemente as percepções dos *stakeholders* sobre esses temas-chave para o debate acadêmico sobre IA e Saúde e para a ESD.

BOX 3 - ESTRATÉGIA DE SAÚDE DIGITAL

A ESD tem como objetivo organizar e fortalecer as ações na saúde digital. Seu propósito é orientar tanto iniciativas públicas quanto privadas a impulsionarem a transformação digital na saúde nacional. São delineados três grandes eixos de ação para alcançar o Plano de Ação da Estratégia de Saúde Digital: (a) ações do MS para o SUS, destacando o programa Conecte SUS como peça-chave para a visão de saúde digital; (b) definição de diretrizes para colaboração e inovação em saúde digital, com ênfase na expansão e na consolidação da governança e dos recursos organizacionais necessários; e (c) estabelecimento do Espaço de Colaboração da ESD, buscando uma troca eficiente entre todos os atores do setor, com papéis e responsabilidades definidos.

Adicionalmente, o plano apresenta sete prioridades: (a) governança e liderança na saúde digital; (b) informatização dos três níveis de atenção à saúde; (c) apoio à melhoria da atenção à saúde; (d) empoderamento do usuário da saúde digital; (e) capacitação de recursos humanos para a área; (f) estabelecimento de um ambiente de interconectividade, e (g) desenvolvimento de um ecossistema de inovação na saúde digital (MS, 2020).

Interoperabilidade

A interoperabilidade é um tema central na agenda do uso de IA na área da saúde. A temática apareceu com muita força e de modo espontâneo na maioria das entrevistas. No cenário brasileiro, a questão da interoperabilidade foi enquadrada pelos *stakeholders* tanto como oportunidade quanto como desafio. No que diz respeito às oportunidades específicas para o Brasil, destaca-se o fato de o país possuir um sistema único de saúde pública com grande volume de dados.

Promover a interoperabilidade entre os dados desse sistema é considerado uma oportunidade única pelo volume e diversidade de dados. Os entrevistados acreditam que a integração das bases de dados do SUS pode propiciar o desenvolvimento de algoritmos robustos e impulsionar o desenvolvimento dessa tecnologia no país.

A RNDS é percebida como uma estratégia positiva para promover a integração de dados e sistemas de saúde no futuro. Há

um entendimento de que a rede pode operar não só como um articulador importante para a integralização de dados no setor público, mas também para incluir os dados do setor privado, já que a concorrência e a falta de diretrizes gerais dificultam a integração de informações neste setor.

A despeito dessas potencialidades, a falta de dados de qualidade e a dificuldade para constituir a interoperabilidade estão entre as principais lacunas identificadas pelos *stakeholders*. Eles apontam a ausência de conexão entre sistemas, a dificuldade de desenvolver protocolos para a padronização de dados e a desigualdade de recursos para realização de registros nos diferentes territórios do país. As entrevistas enfatizam que a falta de uniformidade e padronização nos dados de saúde compromete a interoperabilidade, dificultando o treinamento eficaz de algoritmos. Essas deficiências trazem riscos, pois potencializam os vieses algorítmicos. Bases de dados limitadas ou com concentração de informações para certos perfis da população aumentam as chances de análises enviesadas, fragilizando a qualidade do que é produzido pela IA e, por fim, podendo prejudicar o acesso à saúde e a qualidade dos diagnósticos e encaminhamentos clínicos.

Infraestrutura

A infraestrutura adequada ao desenvolvimento e aplicação de IA está distribuída de maneira desigual no país, com maior presença nos grandes centros urbanos. As entrevistas destacam que muitos municípios enfrentam situações de precariedade, com falta de laboratórios especializados para desenvolver e testar aplicações de IA na saúde, além da carência de equipamentos para digitalização dos dados nos equipamentos de saúde. Essa situação de precariedade constitui, portanto, uma barreira para o avanço do uso da IA na saúde nos mais diferentes territórios e contextos locais.

Recursos Humanos

A falta de recursos humanos com preparo para implementar e utilizar a IA na área da saúde é apontada como desafio significativo no cenário brasileiro. As entrevistas revelam a existência de uma grande dificuldade no recrutamento de profissionais que entendam minimamente de IA e de saúde, por exemplo cientistas

de dados que trabalhem com dados de saúde ou profissionais de saúde com formação em tecnologia e programação. Além disso, apontam que o aumento da demanda de profissionais de TI aumentou os salários desses profissionais em outros setores (como o financeiro), o que dificultou a retenção de profissionais especializados no país.

Regulação

A discussão sobre a regulação da IA aplicada à saúde apresenta-se de forma complexa entre os entrevistados. De um lado, há demanda por mais clareza legal sobre o que pode e o que não pode ser feito em relação a diversos tópicos envolvendo a IA na área da saúde. Por outro lado, a criação de regras voltadas aos sistemas de IA é observada com temor, uma vez que podem desestimular, de acordo com diferentes entrevistados, o avanço desses sistemas no país, fazendo com que o Brasil "fique para trás" na corrida do desenvolvimento e da implementação de IA na saúde.

Ainda que não tenha sido construída com essa intenção, para os entrevistados do estudo, a LGPD é compreendida como uma ferramenta relevante que permite bom desenvolvimento e aplicação das ferramentas de IA no Brasil. Ela é considerada um arcabouço regulatório inicial estabelecido, a partir do qual aplicações podem ser desenvolvidas, testadas e aplicadas. No entanto, como ela não foi criada considerando as particularidades da tecnologia de IA, seu arcabouço regulatório deixa brechas para interpretações variadas sobre aplicações práticas e limites legais. Embora a LGPD regule o uso dos dados pessoais, ela não incide sobre novos dados criados a partir de dados pessoais existentes, nem sobre o que acontece com eles depois da morte das pessoas. Alguns poucos entrevistados apontam que isso pode configurar um risco à proteção e à segurança dos dados na implementação da tecnologia. Assim, a ausência de uma regulação específica para o desenvolvimento e a aplicação da IA na saúde pode ter implicações éticas, legais e de segurança de dados.

Ética

A discussão sobre regulação dos usos da IA estar associada ao debate sobre ética e IA; porém, como adiantado, não apareceu

espontaneamente e com força no conjunto das entrevistas: com algumas exceções, a questão da ética surgiu apenas quando estimulada explicitamente. Todos os entrevistados foram questionados sobre as implicações éticas do uso dessa tecnologia, mas, de maneira geral, não desenvolveram reflexões aprofundadas sobre a questão, além de muitos não deixarem claro o que entendem por "ética" no debate sobre IA. Via de regra, essa discussão é vinculada aos valores dos indivíduos que desenvolvem ou operam ferramentas de IA, ou seja, o arcabouço ético para lidar com as implicações do uso de IA na saúde depende da ética individual dos profissionais. Quando conseguem vislumbrar mecanismos coletivos ou sociais de ordenamento ético para o uso dessas tecnologias, apontam o fortalecimento da ética nos cursos de formação profissional e dos códigos de conduta profissionais existentes.

Considerando que a IA seria apenas uma ferramenta de trabalho como outras disponíveis, é suficiente para os profissionais uma formação geral sobre ética na saúde, tal como consta nos códigos profissionais de médicos, enfermeiros e outros profissionais da saúde. Parte significativa dos *stakeholders* entende que o trabalho com IA se assemelha a qualquer tipo de estudo empírico ou clínico e deve seguir os mesmos preceitos de ética na pesquisa e na medicina, sem necessidade de orientações éticas específicas. Assim, seria suficiente assegurar a privacidade dos dados e a aplicação dos princípios médicos. A validação dos procedimentos éticos para uso de IA na saúde também poderia seguir a mesma dinâmica de validação de estudos empíricos, os quais se pautam, em geral, em debates promovidos por fóruns de profissionais da saúde e na avaliação de pareceristas de trabalhos acadêmicos.

> Eu, de fato, não acho que [para] trabalhar com ética e regulamentação em IA a gente precise investir, assim, modificar radicalmente o jeito que isso é feito, criar algo. Não, são as mesmas regras de sempre, essas regras têm que ficar claras, então existem vários manuais ali de boas condutas, de como se fazer uma pesquisa e ajustar viés em IA, só com um time capacitado para fazer esse *checklist* das pesquisas. Eu acho que é educação, educação mesmo, tem que trabalhar bastante isso com os pesquisadores e na indústria.
>
> (STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

Se você trata o dado do paciente e dá para a ferramenta do paciente, do mesmo jeito que o código de conduta que os médicos aderem quando viram um médico, as enfermeiras aderem quando viram enfermeiras, que os hospitais aderem quando eles abrem o serviço, deveria ser o mesmo, isso deveria ser suficiente, porque é a mesma coisa, é só a ferramenta que você está usando. Um advogado que está regido pela OAB, ninguém vai lá e fala assim, "ah, quero saber qual é o código de conduta da Dell, da Lenovo que fez o computador que você escreveu essa lei" ou "quem escreveu essa defesa". Ninguém faz isso, é a ferramenta que ele está usando, já existe OAB para cuidar do uso final e do resultado final e não da ferramenta. "ah, mas a ferramenta tem que ser cuidada". Logico, tem que passar por todos os critérios de segurança que vai ter, de qualquer *laptop* vai ter, então cada mercado tem o seu, você não precisa criar uma coisa enorme.

(STAKEHOLDER *MERCADO*)

Eu acho que ética é uma questão de formação. Então, a gente precisa trabalhar na formação dos profissionais da área da saúde, da área de exatas, engenharias, de quem trabalha nesse ambiente. [...] Então, isso, eu acho que essa questão de ética precisa trabalhar com os dois lados [TI e Saúde]. Então, isso, eu acho que é formação profissional de quem trabalha nesse ponto.

(STAKEHOLDER *ACADEMIA*)

Primeiro ponto é passar pelos fóruns competentes para decidir sobre isso; o segundo ponto nesse processo você vai ter que garantir anonimização de dados, para ser aprovado um projeto de pesquisa nessa área certamente o [revisor] *ad hoc* vai olhar para a questão da distribuição equânime entre representantes da população.

(STAKEHOLDER *ACADEMIA*)

São as pessoas porque não é o sistema que vai ser [ético], são as pessoas que estão desenvolvendo para questionar se aquilo está infringindo algum ponto ético no desenvolvimento e na implantação do modelo. Então, acho que os aspectos éticos têm que estar nas pessoas e não esperando que ele vai estar no modelo.

(STAKEHOLDER *ACADEMIA*)

Entre alguns *stakeholders* atuantes na academia e nos equipamentos de saúde, a discussão sobre as implicações éticas passa também por uma preocupação com os princípios que orientam o desenvolvimento de ferramentas de IA. Assim, foi reforçado por eles que o uso da IA será ético se for equitativo, ou seja, se os algoritmos e sistemas de IA forem desenvolvidos de maneira justa e imparcial, e se seus resultados não contribuírem ao aprofundamento de desigualdades sociais previamente existentes. Além disso, também foi colocada a importância de a preocupação ética estar em todos os processos de desenvolvimento do algoritmo, devendo-se identificar se os critérios utilizados pelos modelos estão ou não infringindo a ética.

[Tratar a ética é implementar um] esquema da equidade, dar mais para quem está precisando mais, então fazer algoritmo lá na ponta que vai ajudar dez pessoas, acho que temos que trocar cuidados e dar mais para quem está precisando mais e fazer o bem e outro também não causar maior danos.

(STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

Direitos dos Usuários

Por fim, importa destacar que, embora a importância de garantir a privacidade e a segurança dos dados dos pacientes tenha sido mencionada pela maioria dos entrevistados, a discussão sobre os direitos dos beneficiários finais das ferramentas tecnológicas na saúde (pacientes) não apareceu de forma espontânea na maioria das entrevistas. Assim, apesar de o direito dos usuários e beneficiários das ferramentas ser um tema central na ESD, ele apareceu de maneira superficial na fala das poucas pessoas entrevistadas que abordaram o assunto. Quando estimuladas a discorrer sobre esse tema, as falas eram curtas, pouco aprofundadas e frequentemente retornavam ao assunto da segurança de dados.

Parte significativa dos entrevistados entende que os direitos dos usuários dos sistemas de saúde estão circunscritos à privacidade e à proteção de dados. Discorrem sobre a importância de a população ter mais conhecimento sobre o fornecimento de dados pessoais para a saúde a fim de diminuir receios e resistências, além de compreenderem melhor o uso da LGPD e também defenderem a importância de estratégias de educação digital para a população entender os benefícios da tecnologia e os potenciais usos dos dados fornecidos. No questionamento sobre os direitos dos usuários dos sistemas de saúde, um *stakeholder* sintetizou sua percepção como "o direito a ter os benefícios do uso da IA" e outro respondeu ser "dever do cidadão ceder seus dados para que ele possa se beneficiar". Além dos direitos à privacidade e proteção de seus dados, outros direitos dos usuários foram pouco explorados nas entrevistas.

> É um negócio que vai ser feito para beneficiar ele em algum momento da vida dele, entendeu? Então, acho que ele [usuário] tem mais dever de dar os dados mesmo do que o direito, eu acho que tem que compartilhar mesmo, precisa de dado para fazer o algoritmo funcionar bem. O que não pode de jeito nenhum é vazar dado, garantir a anonimidade da pessoa, mas posso garantir que essa pessoa nunca vai ser reconhecida, nunca, vai ser impossível chegar na pessoa e que esse negócio só vai ser usado para fazer um negócio benéfico para a rede de saúde. Eu acho que é mais dever da população. E você só garantir o anonimato.
>
> (STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

> Eu sou do ponto de vista de que se é para causar um benefício coletivo e individual, a gente não deveria ter muitos limites. Eu acho que a gente não pode. Nosso limite individual não pode passar um benefício coletivo.
>
> (STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

Por fim, registramos que o tema da explicabilidade das ferramentas de IA, tão relevante na discussão acadêmica sobre IA e Saúde, também não foi apresentado com uma grande preocupação para a maior parte dos entrevistados do estudo. A falta de transparência e compreensão dos processos de construção de decisões e encaminhamentos das aplicações de IA aparece como preocupação para uma parcela pequena dos entrevistados. Isso indica que parte dos temas prioritários nesse campo de estudos – como explicabilidade, confiabilidade e direito do usuário – não tem ecoado entre os *stakeholders* desse campo no Brasil.

BOX DESTAQUES – TEMAS PRIORITÁRIOS PARA A AGENDA DE IA NA SAÚDE

- A interoperabilidade na saúde brasileira é considerada uma oportunidade, devido ao grande volume de dados do SUS, assim como o potencial da RNDS para promover a integração de dados públicos e privados.
- Ao mesmo tempo, a interoperabilidade é considerada um desafio, pela dificuldade em colocar em prática o que é necessário para que ela aconteça. Uniformidade, padronização e igualdade de registro nos diferentes territórios do país são pontos essenciais para que os resultados das tecnologias de IA sejam confiáveis.
- Desigualdades econômicas e sociais geram lacunas na infraestrutura e nos recursos humanos para IA na saúde. Há uma carência de recursos em municípios e contextos locais, como laboratórios e equipamentos para digitalização de dados, em comparação ao que ocorre em grandes cidades do país.
- Há um grande desafio em recrutar profissionais capacitados para implementar e utilizar IA na saúde no Brasil, tanto cientistas de dados como profissionais de saúde com competência em tecnologia.
- A regulação da IA na saúde é complexa. Há demanda por diretrizes para tratamento de dados, mas há também o receio de que medidas que venham a ser criadas restrinjam e, portanto, atrasem o país na corrida tecnológica. Por mais que a LGPD seja considerada um arcabouço regulatório inicial relevante, sua falta de consideração às particularidades da IA levanta preocupações éticas, legais e de segurança de dados.
- A discussão sobre ética na regulação da IA na saúde surgiu apenas

quando explicitamente estimulada nas entrevistas. Nesses casos, a visão mais recorrente é que a ética depende dos valores individuais dos profissionais, sugerindo que a formação profissional e os códigos de conduta existentes são suficientes. Além disso, a maioria dos entrevistados não deixou claro o que entende por "ética" nos debates sobre IA. Os poucos que o fizeram enfatizaram a importância da equidade em todo o desenvolvimento de ferramentas de IA.

- O tema "direito dos usuários" não foi central nas entrevistas, com destaque apenas para privacidade e proteção de dados, e falta de educação digital da população para compreender seus usos. A questão da explicabilidade das ferramentas de IA também não foi amplamente discutida entre os entrevistados.

PRÁTICAS EM ANDAMENTO

As entrevistas realizadas também buscaram identificar exemplos de usos de sistemas de IA na saúde no Brasil e ações voltadas ao desenvolvimento dessa tecnologia no país.

Inicialmente, são descritos e analisados os apontamentos dos entrevistados sobre o estágio atual de desenvolvimento e aplicação das iniciativas de IA em suas organizações, destacando as estratégias iniciais mais comumente adotadas, considerando as barreiras iniciais mais frequentemente percebidas.

Em seguida, serão descritas as iniciativas em andamento a partir de suas finalidades. Foram mapeadas majoritariamente iniciativas com quatro tipos de finalidade: (a) promoção da interoperabilidade, (b) soluções para o aprimoramento da gestão, (c) ferramentas de diagnóstico e (d) modelos de predição.

Por fim, a seção encerra-se, destacando os desafios e os riscos apontados pelos diferentes *stakeholders* para desenvolver e/ou para implementar as iniciativas citadas.

CENÁRIO DE IMPLEMENTAÇÃO E ESTÁGIO ATUAL DAS INICIATIVAS NO BRASIL

A análise das entrevistas revela que as iniciativas de IA existentes nos diferentes segmentos investigados (academia, poder público, mercado e equipamentos de saúde) estão em estágio inicial de desenvolvimento ou aplicação e enfrentam desafios estruturais semelhantes.

Independentemente do segmento de atuação dos entrevistados, os relatados revelaram que as iniciativas de IA existentes têm sido implementadas inicialmente a partir de estruturas próprias das instituições, como laboratórios, centros de inteligência e departamentos específicos. As organizações preocupam-se em desenvolver estruturas mínimas e um ambiente de implementação direcionado especificamente para iniciativas de IA em saúde.

Por outro lado, os entrevistados do estudo também enfatizaram a importância do desenvolvimento de parcerias com atores de outros segmentos para o desenvolvimento de soluções de IA e o fomento de um ecossistema de inovação. As parcerias mais mencionadas foram entre equipamentos de saúde, academia e empresas de tecnologia. No caso do poder público, as parcerias foram menos mencionadas; quando citadas, em geral, tratavam de relações estabelecidas entre os diferentes níveis governamentais ou com universidades e, eventualmente, *startups*.

Tem uma variedade muito grande, tem uma possibilidade das *startups* e empresas fazerem associação com laboratórios da universidade. Então, tem empresas ali que são da área de Bioinformática, que trabalha nessa parte de medicina genômica, que tem parceria com laboratórios com o departamento de genética, por exemplo.

(STAKEHOLDER *ACADEMIA*)

A necessidade de recursos financeiros e humanos foi apontada pelos entrevistados como um vetor fundamental de estímulo a parcerias para a busca de financiamento. A busca de recursos financeiros de agências de fomento, por exemplo, é uma prática bastante sedimentada não somente na academia, mas também entre empresas de equipamentos de saúde ligados a universidades e entre *startups*.

A gente tem várias empresas para as quais a gente está desenvolvendo tecnologia para a área da saúde, desde empresas da área de equipamentos médicos que produzem no país essa tecnologia e nosso laboratório entra com o desenvolvimento de *software* para melhorar a eficiência desse equipamento. Então, o financiamento de tudo isso do laboratório vem majoritariamente de projetos da lei de informática, mas a gente tem também linha regular de pesquisa junto a [Agência de fomento], especificamente em IA.

(STAKEHOLDER *ACADEMIA*)

Então, na parte financeira, que é um *back-up* muito importante em termos de investimento [...] é a [Agência de Fomento], onde eles têm um programa que é incrível, [...] financiamento que a empresa precisa para tirar aquela ideia do papel ou validar aquela ideia junto ao mercado e construir um protótipo, uma solução, né?

(STAKEHOLDER *MERCADO*)

No que diz respeito a especificidades na atuação, não foi possível diferenciar tipos específicos de solução de IA mais desenvolvidos em cada um dos segmentos entrevistados. Entretanto, foi possível identificar diferentes estágios de implementação conectados às áreas de atuação dos entrevistados. Entre os equipamentos de saúde e atores de mercado, um número maior de iniciativas implementadas ou em fases mais avançadas de teste foram mencionadas. A partir das entrevistas com esses segmentos, também foi possível mapear o uso de soluções de IA implementadas em outros países e inseridas no cotidiano do trabalho de alguns equipamentos de saúde do Brasil.

PROMOÇÃO DA INTEROPERABILIDADE

Em todas as entrevistas, a interoperabilidade aparece como um tema central, por ser um processo chave para o desenvolvimento da IA. A maior integração entre diferentes fontes de dados contribui para a maior qualidade das soluções. Como mencionado na seção anterior, embora seja um processo prioritário para todos os setores e necessário para o avanço da IA no país, seu desenvolvimento é bastante complexo, como relatado nas entrevistas, pois demanda a articulação de atores públicos e privados, o investimento em infraestrutura, e, em alguns casos, a cessão de dados entre concorrentes, o que pode gerar conflitos de interesse.

Os entrevistados mencionaram diferentes esforços e medidas para garantir a interoperabilidade. Nessa frente de iniciativas, foram destacados os esforços das instâncias governamentais para digitalizar e integrar sistemas de saúde por meio da RNDS, que conta ainda com desafios relacionados à adoção de padrões nacionais e ao engajamento de todas as partes interessadas. Além desses esforços, as iniciativas relativas à interoperabilidade lideradas pelos diversos segmentos entrevistados mostram diferentes níveis de maturidade, recursos e limitações para implementação.

O Governo Federal tem delineado mudanças institucionais alinhadas às diretrizes da ESDe à Política Nacional de Informação e Informática que contribuem para a interoperabilidade. Entre as mudanças, está a reformulação de um departamento do MS para monitorar e avaliar a digitalização e uso de IA na saúde no Brasil. Além disso, conselhos e órgãos conectados a esse ministério têm dedicado esforços para

desenvolver a troca de informações e as redes institucionais entre entes governamentais, com o objetivo de estabelecer uma base para o diálogo sobre o avanço do uso de IA e inovação no SUS.

Tem o histórico da ESD. Esse documento que é um marco fundamental, ele é atrelado à Política Nacional de Informação e Informática, [...] tem uma oportunidade fantástica. Institucionalmente está acontecendo que é a reestruturação do [departamento de avaliação], também a nova Secretaria que é também um marco do país importante, porque estamos resgatando coisas antigas, mas com um olhar novo que são a Sala de Apoio à Gestão Estratégica e a Rede Interagencial de Informações para a Saúde, ou seja, são pontes de diálogos, mais do que tudo.

(STAKEHOLDER *PODER PÚBLICO*)

O MS também tem investido na interoperabilidade, preparando e desenvolvendo mecanismos de integração entre os dados de diferentes sistemas e bancos públicos de informações sobre os pacientes. Esse processo está por trás da plataforma do Conecte SUS[8], que disponibiliza para pacientes, profissionais e gestores informações sobre os pacientes e seus atendimentos.

No MS, a gente está adotando um modelo de arquitetura de interoperabilidade que é a rede nacional de dados em saúde, padrão Fast Healthcare Interoperability Resources (FHIR)[9] e a gente está trabalhando progressivamente no tratamento, na descarga e no enriquecimento de dados. Vamos agora começar a tentar devolver o processado para estados e municípios para também poderem consumir esses dados de diferentes níveis de agregação, né? Isso está nos permitindo ter e produzir para esses cidadãos, profissionais e gestores o Conecte SUS, o Conecte do cidadão, o Conecte SUS profissional, que é o próprio prontuário eletrônico, o Conecte SUS gestor, que são as informações agregadas na forma de indicadores para monitoramento, avaliação das políticas públicas, né? Então, é um passo importante que o Brasil deu por esse modelo de interoperabilidade da RNDS, e a gente está em processo. Estamos agora no processo de descarga da regulação, que já tem modelo informacional e computacional, dados mínimos estabelecidos, o registro de atendimento também já tem dados de vacinação, laboratórios, tudo isso já está sendo descarregado na RNDS.

(STAKEHOLDER *PODER PÚBLICO*)

Ainda que em desenvolvimento e enfrentando as dificuldades de operacionalização em um país de dimensões continentais, marcado por desigualdades de acesso diversas, os esforços de âmbito federal podem ser considerados um passo importante para a construção de repositórios de informação de maior qualidade e com menos riscos de vieses sobre a

8 Saiba mais: https://www.gov.br/saude/pt-br/composicao/seidigi/conecte-sus/conecte-sus
9 Trata-se de um padrão para a troca de informações em saúde.

população. Entretanto, em virtude de não haver um repositório consolidado com regulação de acesso e construção conjunta de diferentes atores no momento, esforços para conexões de dados de pacientes aparecem nas falas dos entrevistados como uma ação necessária para a construção de seus projetos de IA. Assim, os entrevistados de todos os setores, mesmo do setor governamental, citam diferentes estratégias para o desenvolvimento de repositórios a partir dos dados aos quais têm acesso.

Academia, Mercado, governos municipais e gestores de equipamentos de saúde relataram o uso de diferentes formatos de repositórios e de operacionalização da integração entre diferentes fontes de dados. O uso de *data lakes*, ou seja, de um repositório central que pode combinar dados estruturados, dados semiestruturados e não estruturados, com ou sem chaves identificadoras, foi citado por entrevistados de governos municipais, equipamentos de saúde e mercado como um instrumento utilizado para desenvolver a interoperabilidade dos dados.

> Então, a gente começa a Saúde Digital em 2021 tendo esse norte, a gente precisa de um registro eletrônico de saúde, integrar essas bases. Conseguimos implantar [...] um "data lake", a gente tem um registro nacional de saúde de todo o sistema de saúde da prefeitura fazendo conectores, inclusive com privado, conectando outras redes.
>
> (STAKEHOLDER *PODER PÚBLICO*)

> Uma outra coisa que a gente está fazendo também já, há vários anos, né? É um projeto estratégico justamente de organização, estruturação dos nossos dados. Por exemplo, a gente tem hoje um *data lake* com muitos dados de saúde, dados laboratoriais da população brasileira, e o nosso grupo é um grupo de abrangência nacional, então, e ele tem um sistema de dados unificado há muitos anos.
>
> (STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

Para além do formato de repositório, os entrevistados relataram diferentes operacionalizações para a integração dos dados. O barramento, por exemplo, foi implementado por atores do poder público e de equipamentos de saúde, o qual implica um investimento de infraestrutura e de construção de chaves que permitem a interconexão entre dados utilizados em diferentes sistemas de *software*. Esse procedimento é utilizado nas organizações de diversos entrevistados para integrar e centralizar as informações em uma mesma base de dados.

Quando a gente trabalha ESD no [município], a nossa base precisa conectar essas informações. A gente está produzindo e coletando muito, mas se não organizar isso a partir de um padrão de interoperabilidade com barramento, né, junto a Rede Nacional de Saúde, vai refletir também uma capacidade de poder integrar as bases que estão aqui em modelos informacionais, que eles não foram produzidos pela RNDS, esse é o ponto de partida.

(STAKEHOLDER *PODER PÚBLICO*)

A gente está no processo agora de barramentos, criar nosso barramento que a gente tem um repositório principal (RP) e tem em torno de 50 *softwares* que circundam esse RP e não se conversam né. O nosso primeiro movimento, grande movimento, foi criar esse barramento para interoperar todos esses *softwares* e a gente pode ter os dados concentrados em base única.

(STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

Os entrevistados da academia também relataram a construção de sistemas de integração de bases de dados, tanto aqueles que contam com um identificador comum, quanto casos que desenvolvem algoritmos que conectam bases de dados distintas, sem identificadores únicos e por meio de códigos criptografados e não reversíveis, a fim de garantir a anonimização, privacidade e proteção dos dados pessoais.

A gente tem usado uma ferramenta, é um algoritmo probabilístico, cuja função é conectar essas bases de dados que não têm identificadores únicos, usando dados nome, nome da mãe, data do nascimento, mas na seguinte condição: ele cria uma criptografia, ele transforma esse nome, ele codifica em trincas de DNA de nucleotídeo. Então, ele vai converter em cada letra numa trinca de DNA, como se fosse uma sequência, e cada vez que o paciente entrar, ele é codificado com um conjunto de nucleotídeos diferentes, não reversíveis, você não consegue retroconverter. Então, eles [algoritmos] criptografam os dados sensíveis. Então, em tese, você resolveu parte do problema de anonimização e também nessas sequências a gente consegue comparar com a outra base de dados usando o algoritmo do Basic Local Alignment Search Tool (BLAST). Isso trata para a gente o problema de interoperabilidade, para conectar sistemas, bases, que não são conectados, para que a gente possa ter minimamente essa interoperabilidade. Isso a gente tem tentado oferecer para o MS, secretarias, para eles usarem sem custo.

(STAKEHOLDER *ACADEMIA*)

FERRAMENTAS DE GESTÃO

O uso de IA para o desenvolvimento de ferramentas de gestão é abrangente. Entrevistados do mercado, de equipamentos de saúde e acadêmicos citaram o investimento nesse tipo de iniciativa[10]. Contudo, a partir das entrevistas, foi possível ve-

10 Entre os entrevistados do poder público, não foram mapeadas iniciativas que contribuem para a gestão de saúde. Houve, no entanto, a menção de uso de ferramentas de categorização de dados para busca de informações sobre projetos de lei.

rificar que iniciativas de gestão são uma área de investimento especialmente para *startups* e empresas mais consolidadas no ramo de tecnologia. Diferentemente das ferramentas para diagnóstico, identificamos, entre as ferramentas de gestão, soluções que estão sendo implementadas de maneira cotidiana nos hospitais, laboratórios e demais estabelecimentos de saúde.

As iniciativas em andamento mencionadas pelos entrevistados envolvem o atendimento ao paciente, a otimização de processos operacionais internos e diferentes dispositivos que contribuem para a triagem e organização de filas. No que diz respeito ao atendimento, observa-se o investimento em ferramentas que usam IA generativa em *chatbots* para melhorar a adesão do paciente ao cuidado e para acompanhar o pós-operatório. Um entrevistado de um equipamento de saúde tem desenvolvido um modelo de IA que orienta o paciente de transplante de medula no pós-operatório, indicando o que ele deve fazer, que tipo de restrições precisa implementar na rotina e quando deve contatar o médico.

O uso de IA também tem sido empregado para registrar ou gerar dados para o atendente. Nesse sentido, os entrevistados relataram iniciativas que recuperam a jornada do usuário, apresentando, assim, um histórico da internação, por exemplo, de maneira automatizada para os diferentes atendentes do hospital que precisam ser envolvidos. Existem ainda ao menos duas iniciativas de registro de informação: uma delas utiliza o atendimento *online* escrito do paciente para construir um registro no prontuário, apoiando o profissional de saúde na decisão de encaminhamento para atendimento presencial ou telemedicina; outra, utiliza o relato oral não organizado da análise de uma imagem para registrar informações em laudos. A primeira tem sido utilizada em um plano de saúde e, por meio de técnicas de aprendizado de máquina, tem tido bons resultados de registro.

A gente tem [...] o que foi ativo foi o do prontuário, então a gente tem um resumo, a gente faz uma anotação de consulta, como chamar? faço um pré-registro, isso é melhor, faço um pré-registro de enfermagem, [...] eu tenho uma diminuição de tempo de uso do prontuário, mas não foi, e eu tenho um "quali" muito forte, uma experiência muito boa de utilização. Esse foi o primeiro protótipo que a gente conseguiu construir do ponto de vista de prontuário. [...] [O pré-registro] de todos os enfermeiros, de todos os atendimentos, 500 atendimentos por dia utilizam IA como pré-registro, no campo "Sintomas (Subjetivo)" e "Objetivo". A gente usa uma metodologia chamada SOAP [, cujo prontuário requer registro de Subjetivo-Sintomas, Objetivo, Avaliação e Plano], que sintomas e objetivos são autopreenchidos, [mas] o "A e P" [Avaliação e Plano] ainda não. [...] A gente atende pelo *chat*, então ele vai atendendo. Quando ele vai registrar, ele fala: "autorregistrar". E ele pega a conversa que o enfermeiro teve com você no *chat*, usa o *prompt* que a gente construiu, estrutura isso de forma técnica, separada na metodologia do *Swap* e te traz como uma sugestão de registro, o qual enfermeiro ajusta o texto e publica, ou gera outra vez ou começa do zero.

(STAKEHOLDER *MERCADO*)

Temos um projeto que visa melhorar a eficiência do laudo de radiologia, fazer com que o radiologista possa falar, ao invés dele redigir o laudo, né? Ele fala e a IA ela transforma as informações num [laudo] padronizado, né? Para evitar que as informações sejam diferentes de radiologista para radiologista, e depois o radiologista revê o laudo e ele assina, então isso ajudaria a padronização de laudos de radiologia.

(STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

Algumas *startups* também desenvolvem sistemas que utilizam IA para auxiliar no atendimento remoto, tais como ferramentas para análise de sintomas e diagnósticos preliminares a partir do relato do paciente. Um entrevistado relatou que a *startup* em que atua construiu uma ferramenta de IA para triagem voltada às consultas remotas que busca identificar a demanda do paciente e realizar a triagem para o tipo de atendimento necessário.

Além de aplicações em teste para realizar a triagem, há também o desenvolvimento de algoritmos para organização de filas de atendimento e otimização de processos operacionais internos. Nesse sentido, vale mencionar a iniciativa de uma universidade que está desenvolvendo programas em parceria com empresas para questões administrativas, como a gestão de filas, além de desenvolver Robotic Process Automation (RPA), ou seja, robôs para tarefas operacionais.

Também vale citar uma iniciativa empresarial que usa IA para mapear a jornada do paciente e comunicar aos diferentes profissionais envolvidos os problemas identificados, conectando informações entre diferentes setores, bem como para buscar a automatização de procedimentos padronizados de autorização.

A gente acredita muito também na IA para a parte de negócio mesmo, gestão de filas, gestão de área administrativa, RPA... essas coisas a gente vem fazendo testes em algumas regiões.

(STAKEHOLDER *ACADEMIA*)

Aqui na empresa, nós digitalizamos toda a jornada do paciente dentro de um hospital. Quem entra num hospital já passou pelo *check-in*, documentação, se foi para uma consulta ou se foi internado ou para uma cirurgia. Em cada ponto desse que eu citei, nós temos um módulo que digitaliza parte da operação. Qual é a operação que a gente digitaliza? Depende do ponto, vai ser sempre aquele processo que tem mais gargalo. Então, se eu estou no *check-in*, a parte de documentação, de checagem com os provedores de saúde, se é na internação... existe um grande problema na internação que é, na operação, no centro cirúrgico, que é você estar passando por uma cirurgia e falta algum insumo ali na hora e tem a figura de um enfermeiro circulante que tem que sair dessa sala e sair correndo para pode pegar um equipamento, uma bandagem, algo que faltou. Nossa empresa tem uma solução que coloca na sala para facilitar a comunicação com esse enfermeiro e as áreas do centro cirúrgico. Na internação, como eu falei antes, a gente vê que tem gargalos ali de eficiência no time da enfermagem, a gente vai, coloca a nossa solução para facilitar a comunicação do paciente e as áreas.

(STAKEHOLDER *MERCADO*)

Eu tenho uma frente [...] que é essa questão de facilitar o acesso a autorização mais rápido, com automação. [...] A gente tem um banco de dados organizado, então eu já conheço o comportamento de instituições e eu já sei conforme o tipo de risco da pessoa, o local onde ela pode ficar mais ou menos tempo internada. [...] Hoje, eu já tenho isso com o *analytics* que já fez o estudo desse risco. Então, eu indico para equipe de auditoria revisar, porque o dado não pode ser resolvido por si só, precisa de alguém na ponta para tomar a decisão com a informação e na maioria das vezes acerta, algumas vezes não. Ao mesmo tempo eu construí agora o mesmo tipo de informação para o médico, então o médico e o assistente também têm essa informação.

(STAKEHOLDER *MERCADO*)

Também conectado aos processos de gestão, um dos entrevistados citou um sistema de IA que está em desenvolvimento para facilitar a comunicação entre equipes e aumentar a eficiência, a partir da análise da jornada de atendimento. A IA registra todo o percurso do paciente, desde a chegada a um equipamento até o desfecho clínico, passando por documentação, exames, medicamentos e cirurgias, e utiliza essas informações para articular de forma mais automatizada os diferentes setores durante o atendimento.

FERRAMENTAS DE DIAGNÓSTICO

O trabalho de campo revelou que uma parte significativa das iniciativas em andamento é dedicada ao desenvolvimento de algoritmos para análise diagnóstica. Universidades, empresas de tecnologia, *startups* e equipamentos de saúde têm investido, sobretudo, em soluções de diagnóstico a partir de

imagens. Por exemplo, foi indicada uma universidade que utiliza ferramentas de IA para análise de imagens na formulação de laudos, priorização de atendimento e triagem. Essas iniciativas combinam desenvolvimento próprio e o uso de soluções implementadas em outros países. Um exemplo é uma empresa que introduziu em suas filiais brasileiras a análise automatizada de imagens para raios-X, tomografia de tórax e ressonância de próstata e de crânio. Essa iniciativa foi testada e já é comercializada no Brasil.

As iniciativas de análise automatizada de imagens abrangem diversos tipos de registros, incluindo tomografia, radiografia, ressonâncias e até histologia. Entrevistados de diferentes setores destacaram, sobretudo, o desenvolvimento de algoritmos voltados para patologias específicas. Um exemplo é uma iniciativa citada, ainda em teste em uma universidade pública, sobre o desenvolvimento de um algoritmo para diagnóstico de pacientes que sofreram AVC, capaz de classificar casos como hemorrágicos ou isquêmicos, que demandam rotinas de tratamento e medicação distintas. Além disso, foram relatadas iniciativas em equipamentos de saúde para detecção de tumores de fígado e alterações moleculares (biomarcadores) em imagens de tecido e células humanas (histológicas).

A gente também tem várias iniciativas que estão sendo feitas aqui, tem trabalhos voltados para essa parte de imagem. Então, a gente trabalha muito com imagens neurológicas, essa parte de neurologia é uma área muito forte aqui na faculdade e no hospital. Aí tem pesquisas com coisas locais, sistemas, projetos locais, mas tem também trabalhos de parceria com empresas que estão testando soluções dentro do hospital, por exemplo, para avaliar pacientes que tiveram AVC, para ver se é hemorrágico ou isquêmico para poder medicar. A unidade de emergência aqui tem um centro de AVC que é referência para a região, eles têm trabalhos voltados para aplicação de métodos para apoio de diagnóstico. Isso tem também em outras áreas, oncologia também tem, na hematologia.

(STAKEHOLDER *ACADEMIA*)

Nós temos um projeto muito interessante que foi divulgado recentemente aí na imprensa, aonde a IA é utilizada para detectar tumores de fígado em pacientes com doenças crônicas, paciente com cirrose, doença crônica do fígado. Eles são pacientes de alto risco para detecção de tumores de fígado, e nós desenvolvemos um algoritmo que ajuda a detectar esses tumores, ajuda o radiologista a detectar esses tumores em exame de imagem.

(STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

A gente está muito interessado em detecção de biomarcadores diretamente a partir da imagem histológica, então o câncer é uma doença genética impulsionada por mutações que acontecem no DNA das células. O manejo do paciente de uma maneira mais personalizada é baseado na detecção de biomarcadores, são medidas que são avaliadas, e esses biomarcadores eles são detectados por meio de análise de teste genético, né? De biologia molecular, só que nem sempre há uma disponibilidade desse ferramental, dessa esteira tecnológica nos laboratórios, na maioria dos hospitais. A ideia nossa é a gente detectar, a partir da imagem histológica, verificar se o paciente tem o biomarcador, se tem a deficiência, por exemplo.

(STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

O desenvolvimento de IA relacionado a doenças específicas resulta, em parte, de especialidades médicas com histórico sólido de pesquisas em equipamentos de saúde, mas também reflete o grau de maturidade e disponibilidade de bancos de dados, que são fundamentais para a evolução dos algoritmos.

Entre as iniciativas que buscam criar algoritmos mais abrangentes, destacamos o esforço de um laboratório universitário em desenvolver algoritmos voltados à análise e ao diagnóstico de imagens na radiologia de maneira mais ampla. Esse avanço foi possível graças a um banco de dados construído durante a pandemia, que reúne informações detalhadas de pacientes, incluindo imagens e exames laboratoriais. Esse repositório tem sido essencial para o desenvolvimento de soluções de IA que, por identificar padrões e características relevantes nas imagens, produzem laudos diagnósticos. Os algoritmos de imagem desenvolvidos foram testados, validados e estão sendo comercializados. Segundo a pessoa entrevistada, os resultados apresentam qualidade comparável aos das empresas líderes do mercado.

Nós criamos um laboratório. [...] Aqui, o nosso utiliza muito [IA] em imagem, radiologia. Então, a gente já testa, desenvolve o algoritmo, temos até o *marketplace*. Mas a gente tem apostas que a gente está colocando as fichas. Uma é o uso de IA na imagem. Então, para acelerar o laudo, aumentar a eficiência, tudo isso na parte de imagem.

(STAKEHOLDER *ACADEMIA*)

Por fim, a análise diagnóstica também tem sido usada para priorização de casos e triagem, combinando, assim, ferramentas de diagnósticos e soluções para gestão. Uma empresa descreveu o desenvolvimento de ferramentas que analisam imagens e históricos de pacientes para auxiliar na triagem e priorização de casos, além de identificarem e diagnosticarem precocemente doenças crônicas ou de alto risco.

> Hoje, a gente tem duas linhas de produtos. Uma de análise de imagens com visão computacional, onde a gente tria pacientes em pronto-socorro, em teleatendimentos, pré-analisando as imagens médicas. [...] Tem um outro produto onde a gente olha o paciente como um todo, olhando todo esses textos, você foi no medico, você foi só no nutricionista, foi no ortopedista e mencionou que você toma um remédio que só toma quem é diabético, não tem nada a ver com a dor no joelho, mas ele escreveu aquelas coisas lá, pô, tem um algoritmo que consegue identificar isso, assim: "oh, será que vale a pena agendar esse paciente com endócrino para cuidar da diabetes dele para ver se está em ordem, para ver se está tomando a dose certa, para fazer um comportamento ou só fazer uma reunião de acompanhamento com uma auxiliar de enfermagem para saber se ele estava aplicando, se está bem treinado em aplicar, a fazer medicação, se ele sabe medir certinho a glicose na hora de aplicar", sabe, coisas simples, né? E esse tem sido o nosso maior foco.
>
> *(STAKEHOLDER MERCADO)*

MODELOS DE PREDIÇÃO

Assim como no caso das soluções de IA para diagnósticos, os entrevistados relataram o uso de modelos de predição importados e de desenvolvimento próprio. No caso do primeiro, foi mencionada uma ferramenta que procura prever quais são as patologias que os pacientes podem desenvolver, gerando inclusive um ordenamento a partir das chances de ocorrerem. Ainda que o algoritmo tenha sido desenvolvido a partir de outras populações, ele foi implementado e é utilizado hoje por um equipamento de saúde no Brasil que faz uso de um repositório integrado aos dados de pacientes brasileiros.

No que diz respeito ao desenvolvimento próprio, os entrevistados citaram dois modelos de predição de eventos. Um deles mede a possibilidade de emergir uma epidemia, enquanto o outro afere o avanço de uma doença transmissível, procurando identificar quais serão seus próximos focos territoriais. Tais iniciativas foram desenvolvidas pela academia e por uma equipe do poder público, respectivamente. Em ambos os casos, os entrevistados frisaram a importância da predição para preparação e possível contenção do evento em questão, inclusive pensando a logística de operação para atendimento e otimização na alocação de recursos.

A gente está com uma iniciativa nova aqui, de fazer uso de prontuários eletrônicos que são produzidos por médicos, enfermeiros de atenção básica, nas Unidade de Pronto Atendimento (UPA), na rede de emergência, para fins de vigilância em saúde, detecção de surtos e epidemias rapidamente. Ninguém usa mais prontuário em papel, salvo lugares muito específicos, então é um dado riquíssimo que tem a descrição clínica do paciente, e esse dado é negligenciado, ninguém usa aquilo. Esse projeto tem como objetivo aplicar modelos sobre esse texto para classificá-los dentro de um conjunto de síndromes que tem potencial epidêmico. A ideia é aplicar os modelos, classificar e monitorar usando três categorias que é: pessoa, tempo e lugar. A ideia é consumir esse dado, classificar em síndrome para produzir uma informação de vigilância em tempo real e informar o gestor de saúde que tem algo acontecendo não há um mês atrás, mas hoje, ou ontem no máximo, e que é no bairro x, entre as ruas y e z, e parece que é dengue. Aí ele já sabe quais medidas tomar, é uma espécie de alerta epidemiológica através de dados clínicos.

(STAKEHOLDER *ACADEMIA*)

Então, uma das coisas que o [Centro de inteligência conectado à gestão da saúde] fez foi aquela curva epidemiológica que a gente fez com base em otimização espacial, para saber no caso, em [capital de um estado brasileiro], onde teria incidência em quatro semanas, em que território haveria mais casos, como se fosse uma migração assim, um movimento do caso de óbitos. Esse foi um trabalho legal de otimização espacial.

(STAKEHOLDER *PODER PÚBLICO*)

Foram citadas também iniciativas de predição de risco que envolvem de maneira mais direta os pacientes. Destacam-se aqui três casos mencionados pelos entrevistados: (a) um projeto de um equipamento de saúde, em estágio inicial, que prevê câncer de mama com base em exames laboratoriais; (b) um projeto do poder público estadual que prediz o risco de morte em recém-nascidos, ainda em fase piloto; (c) e um projeto em desenvolvimento por uma empresa de monitoramento e predição de queda de pacientes. Nos três casos, os entrevistados ressaltaram a importância da construção e da manutenção de uma base de dados extensa para desenvolverem o algoritmo; nos últimos dois, os entrevistados relataram que os sistemas utilizados apresentaram resultados positivos e contribuíram para a priorização e a alocação de recursos.

[...] focado muito na parte de exames de imagem laboratoriais, então a gente tem dados hoje muito grandes. Vou te dar um exemplo: a gente recentemente organizou um painel com dados de exames de rastreamento de câncer de mama. Então, de mulheres com mais de 20 anos de dados, né? Então, a gente tem dados, por exemplo, que mostram a prevalência do câncer de mama de acordo com a idade em Estados diferentes. [...] eu tenho liderado uma pesquisa para fazer um modelo preditivo de risco de câncer de mama baseado em exames laboratoriais, por exemplo. Laboratoriais comuns, como hemograma, colesterol, perfil lipídico, que tem uma relação de risco de câncer de mama, com esses exames de rotina.

(STAKEHOLDER *EQUIPAMENTO DE SAÚDE*)

Nós temos um projeto que foi construído no final de 2018, que só agora ele está ganhando corpo, sendo implantado e ganhando até prêmios. É um projeto que visa a redução da mortalidade infantil. [...] E o que é que ele faz? Ele captura várias características daquela criança que nasceu, desde a gestação, tempo de gestação, quantidade de consulta pré-natal e dados do nascimento, peso, se tem má formação, se é prematuro, se não é, então são várias características que são coletadas no nascimento. Aí, o que o projeto faz? Ela vai na base de dados dos nascidos vivos e encontra crianças com aquelas características. E, depois, eu pego essas crianças e vou para a base de mortalidade até um ano. Essas crianças que nasceram com essas condições, elas sobreviveram ou elas faleceram até um ano? Quantas faleceram? E, aí, a solução faz essa predição para uma criança que nasce, recebe essa avaliação da solução, e os técnicos de enfermagem identificam lá: "ah, essa criança precisa de 20% de atenção". Além desse número, tem uma série de cuidados: limpar o cordão umbilical, dar banho e não sei o quê, verificar a temperatura tantas vezes ao dia, tem várias recomendações, aplicar tal vacina, então tem várias recomendações que são feitas pela solução, de acordo com a condição de cada criança.

(STAKEHOLDER *PODER PÚBLICO*)

A grande solução que nós temos hoje são sistemas de monitoramento de paciente internado para risco de queda. Então, a gente sabe que risco de queda é um evento adverso, não deveria nunca acontecer, e quando acontece, além de impactar o paciente, ela impacta também a instituição na visão que ela tem com o mercado; você não consegue ganhar nenhuma creditação/certificação de nenhum órgão de fora se teu nível de queda é alto. Hoje, a nossa empresa tem câmeras nos leitos e, por visão computacional, nós conseguimos detectar a posição do paciente antes de ele levantar da cama, então detecta o risco do paciente em cair cruzando com os dados do prontuário. Então, se eu estou entre vinte e trinta anos, não tomei nenhum medicamento nas últimas seis horas, passei por um procedimento de, sei lá, de braço, eu posso sentar na cama, mas talvez o João que fez um procedimento no joelho, que tem a mesma faixa etária e tomou uma medicação x há três horas atrás ele não possa. Então, esse já é um risco de queda. E, aí, o legal porque hoje essa ferramenta ela não só é treinada com imagens reais, mas também ela gera as próprias imagens baseada em imagens reais.

(STAKEHOLDER *MERCADO*)

DESAFIOS NA IMPLEMENTAÇÃO DE IA NA SAÚDE

Quando nos debruçamos sobre os desafios enfrentados pelos diferentes setores engajados para implementar essas iniciativas de IA, é possível identificar alguns temas recorrentes.

No que diz respeito aos recursos financeiros, entrevistados da academia e do mercado apontaram para dificuldade de investimento em projetos que envolvem IA. Eles relatam a existência de poucas linhas de investimento estatal para desenvolvimento autônomo nesse tipo de tecnologia. No caso da academia, foi ainda relatada a dificuldade de as linhas existentes, por vezes, terem como condicionante a parceria com instituições privadas, o que poderia gerar alterações significativas no trabalho, como na autonomia para sua execução, bem como conflitos associados aos interesses colocados para o desenvolvimento das soluções.

Por sua vez, desafios relativos a recursos humanos foram mencionados por entrevistados de todos os segmentos.

Acadêmicos, gestores de equipamentos de saúde, empresários e gestores públicos apontam o recrutamento, a especialização e a manutenção da equipe no projeto como desafios recorrentes para a implementação de suas iniciativas.

Possuir recursos para manter as equipes é colocado como um desafio central por todos os tipos de perfis entrevistados. A disparidade salarial foi mencionada especialmente entre os gestores de equipamentos de saúde públicos e os profissionais das universidades. Os entrevistados desses grupos argumentaram que o pagamento das equipes, em geral, é feito exclusivamente pelas bolsas das agências de fomento, por isso se torna difícil manter profissionais bastante qualificados com uma remuneração frequentemente menor do que a oferecida pelo mercado.

Por outro lado, os entrevistados de equipamentos de saúde privados e gestores de organizações do mercado que trabalham com IA relataram que a disputa por profissionais e a falta de recursos estão relacionadas com outros setores da economia que utilizam esse tipo de tecnologia e com a demanda de outros países. Como mencionado entre os desafios para o desenvolvimento da IA no país, as ofertas de trabalho com IA em finanças e a possibilidade de trabalhar remoto em organizações internacionais, com salários mais competitivos aparecem como barreiras importantes para o recrutamento e a manutenção desses profissionais.

A rotatividade de profissionais é, para os *stakeholders*, consequência desse desafio e tem como implicações dificuldades de sustentação e descontinuidade de certos projetos, bem como necessidade de ambientação constante de novos profissionais e consequente dificuldade de construir avanços.

> [...] o meu laboratório... já estou na terceira equipe. Eu monto a equipe, vem um banco e leva todo mundo embora. Eu monto a equipe, vem um cara e leva todo mundo. Então, a mão de obra [...].
>
> (STAKEHOLDER *ACADEMIA*)

> Manter talentos é uma baita dificuldade, começa aqui e depois de um ano está na [empresa global de tecnologia]. É muito difícil manter, é muito difícil competir com dólar, sabe? São poucos que você acha que são bons e você não consegue segurar.
>
> (STAKEHOLDER *EQUIPAMENTOS DE SAÚDE*)

A formação e as capacidades dos profissionais envolvidos com as atividades necessárias para o desenvolvimento da IA também aparecem como um gargalo de recursos humanos para os entrevistados. Esse desafio também é apontado para o cenário brasileiro, mas, quando pensamos nas iniciativas, ele não aparece conectado exclusivamente às equipes que estão desenvolvendo os algoritmos. Para os entrevistados, especialmente aqueles lidados às universidades, há também um gargalo nas capacidades dos profissionais responsáveis, em primeira instância, pela construção das bases de dados e pelo uso das ferramentas. São profissionais de saúde que precisam ser treinados para registrar informações melhor especificadas e padronizadas, ou mesmo para ler e discriminar laudos feitos com aprendizado de máquina. Na visão dos entrevistados, a falta de capacitação para esses profissionais afeta principalmente a estruturação e tratamento dos dados utilizados nas possíveis iniciativas, mas também impactam negativamente as possibilidades de aprimorar o desenvolvimento dos algoritmos a partir da leitura humana dos resultados.

> Falta contratação ligeira de IA, difícil, nem as faculdades federais e nem as faculdades particulares estão formando esses profissionais.
>
> (STAKEHOLDER *EMPRESAS*)

> Quando a gente vai para dentro do hospital, o corpo profissional do hospital, fundamentalmente aquele que interage com os dados e tal, ou eles são os profissionais de saúde, ou eles são profissionais mais da área de gestão, da organização do hospital. Então, embora nada, se é que tem uma equipe, uma divisão de informática, eles são muito mais focados na questão mais de infraestrutura. [...] Os recursos humanos disponíveis estão com foco no cuidado, tem que dar conta do atendimento ali, mas abre um gap ali, uma limitação muito grande para essa etapa que seria para você evoluir e conseguir trazer essas coisas até mais rapidamente para dentro da rotina. O que o pessoal fala de Medicina Transnacional [...] a gente acaba gastando muito tempo e demorando para conseguir que as coisas possam ser mais bem aproveitadas [...], porque a hora que eu for colocar para rodar [...], eu tenho que organizar os dados, a hora em que ele for rodar, ele vai consumir dados que precisam estar também padronizados. Então, isso eu acho, eu diria que essa é a principal limitação, é que isso é uma limitação até recursos humanos, porque precisa, é um investimento, numa infraestrutura e pessoal, que não está direto ali no atendimento, hoje é difícil, a gente sempre tem escassez de recurso, enfim.
>
> (STAKEHOLDER *ACADEMIA*)

Entre entrevistados do poder público e do mercado, a resistência em investir em IA também foi citada como dificuldade para implementação das iniciativas, especialmente no âmbito governamental. Foram mencionados o desconhecimento e a

desconfiança do que é possível se desenvolver a partir da IA, e a dificuldade de priorização da agenda, dado que as ferramentas dificilmente são desenvolvidas a curto prazo e/ou tratam de problemas urgentes. Esse último cenário, combinado com a escassez de equipe, trouxe como consequência, para um dos entrevistados do poder público, a descontinuidade e a realocação de sua equipe para atender a outras prioridades da gestão.

Essa dificuldade não opera apenas no âmbito do desenvolvimento das ferramentas em si (ou algoritmos), mas também no que diz respeito à interoperabilidade. Nesse caso, entrevistados do governo e do mercado identificam tanto a resistência e a desconfiança entre os atores sobre a viabilidade de integrar os dados sem ferir qualquer regulação, como também os conflitos de interesse, os quais podem envolver desde engajamento e priorização do tema da agenda em diferentes instâncias do poder executivo até perspectivas conflitantes sobre a detenção dos dados a serem utilizados para desenvolver os algoritmos. Nesse sentido, merecem destaque dois relatos: o de um entrevistado do poder público, o qual sugere que a maior dificuldade para a interoperabilidade é a articulação entre os diferentes atores responsáveis pelos equipamentos detentores dos dados; e o de um entrevistado do mercado, que diz precisar pactuar com clientes a cessão dos dados para desenvolver a solução, haja vista que muitas das organizações consideram esse material um recurso estratégico por ser capaz de garantir, por exemplo, propriedade sobre o desenvolvimento do algoritmo.

> A rede [necessária para garantir a interoperabilidade] não é só na questão tecnológica: ela é uma questão humana, porque o maior problema que a gente está observando não é colocar parafernália para rodar, eu deixar uma rede neural rodando lá, não é esse o problema. O problema ainda é conectar as pessoas.
>
> (STAKEHOLDER *GOVERNO*)

> Entendimento do cliente em ceder os dados. Eu repito isso, meu cliente ele tem que saber que, ao ceder os dados, anonimizados, da forma como ele quiser estruturar, ele está fazendo isso para o bem dele, né? [...] A gente sofre bastante com isso quando a gente vai para um novo cliente. Ele acha que é como acontecia antigamente, que é, eu vou te dar o capital intelectual dos meus profissionais e, aí, por isso eu tenho que ter participação de *royalties* naquele produto. Com a IA, isso não funciona, porque quando tu me dás acesso, eu vou gerar não só esse produto, eu vou gerar uma série de produtos dentro do hospital, né? [...] Então, acho que isso parte dos *players* entenderem que o acesso de dados é importante e que ele beneficia o próprio *player*.
>
> (STAKEHOLDER *MERCADO*)

Mais uma vez, a regulação e, mais especificamente, a LGPD aparecem como uma das dificuldades enfrentadas pelos entrevistados para implementar suas iniciativas. Ela foi citada como um mecanismo que, a princípio, deveria garantir a manipulação segura dos dados, porém opera como barreira à inovação. A noção de desidentificar e anonimizar, e de como construir o acesso aos dados de maneira estruturada, garantindo essas prerrogativas, seriam temas em discussão e abertos à interpretação, o que pode implicar limitações para acessar os dados e desenvolver as soluções.

> [...] LGPD, de dados sensíveis é uma coisa que a saúde precisa estar muito bem estruturada, senão ela não vai chegar lá, porque se a gente deixar muito travado, você não vai conseguir testar nada.
>
> (STAKEHOLDER *ACADEMIA*)

> [...] LGPD, que deveria ser um mecanismo de abertura de portas, hoje é usada como fechadura para pessoas que não cooperam com a gente. Essa dificuldade para acesso a dado é essencial para desenvolver qualquer tipo de modelo dessa natureza e, sem ele, não tem como desenvolver.
>
> (STAKEHOLDER *ACADEMIA*)

Por fim, uma das dificuldades mais frequentemente citadas pelos entrevistados é a capacidade efetiva dos algoritmos. Esse problema aparece associado à dificuldade de se construir a interoperabilidade dos dados, seja pela inexistência de uma base nacional, seja pela desarticulação dos atores que compõem o ecossistema do setor, ou mesmo pelas dificuldades regulatórias de se combinar de maneira desidentificada informações de usuários dos sistemas de saúde em diferentes equipamentos (privados e públicos).

Boa parte dos *stakeholders* cita a busca de parcerias com outras instituições como forma de enfrentar esse desafio. Entretanto, entre os entrevistados, a maior parte citou iniciativas que acabam se apoiando em dados limitados ao equipamento envolvido, ou seja, dados de uma patologia específica em que o equipamento tenha especialidade ou dados dos pacientes de um hospital, por exemplo, ou ainda, no caso de governos, dos municípios e/ou dos estados em questão. Isso significa que o poder de explicação e predição está limitado e condicionado pelos perfis aos quais esses espaços têm acesso; em outras palavras, limitado às restrições dos bancos de dados

disponíveis. Essa limitação é mencionada por entrevistados de todos os segmentos e um dos principais riscos mencionados para implementação de ferramentas de IA no setor de saúde, como apresentamos na próxima seção.

RISCOS ASSOCIADOS À IMPLEMENTAÇÃO DE IA NA SAÚDE

Os *stakeholders* de diversas áreas de trabalho apontam três principais riscos na implementação de ferramentas de IA: viés algorítmico, experimentação e segurança dos dados. A maneira de lidar com esses riscos varia entre os entrevistados, aparentemente em função do setor de atuação, constatação especialmente válida para a segurança dos dados.

Quanto aos vieses algorítmicos, os entrevistados destacaram que esse risco está intrinsecamente ligado ao acesso a dados populacionais. Como não há interoperabilidade de dados dos pacientes em âmbito nacional, tanto na rede pública quanto na privada, o aprendizado dos algoritmos fica limitado às bases de dados acessíveis às equipes responsáveis pelas diferentes iniciativas.

A principal questão para os entrevistados é em razão de o algoritmo responder apenas "sobre a população cujos dados teve acesso", que podem não ser um espelho da população em geral, como também podem contar uma história fragmentada, sem todas as informações sobre o paciente. Isso significa que há, de um lado, um risco de viés, no qual o algoritmo só é capaz de fazer predições para perfis específicos, como também torna arriscado extrapolar as conclusões desenvolvidas para parte da população para os demais, podendo incorrer em "racismo algorítmico", como apontado por uma das pessoas entrevistadas. Para lidar com esse tipo de risco, os entrevistados citam uma cautela na generalização dos resultados e um cotejamento do diagnóstico com o conhecimento especializado de profissionais. Entretanto, para nenhum dos grupos essas ações aparecem como suficientes para mitigar o risco de viés algorítmico.

Em parte, a percepção dos *stakeholders* é de que o trabalho com as soluções de IA envolve risco porque está em um momento de experimentação. Os entrevistados apontam que as soluções de IA estão frequentemente em fases

experimentais, o que implica um risco de falhas aplicadas em cenários clínicos reais. Embora esse aspecto tenha sido menos mencionado por atores do mercado, os entrevistados observam essa situação na produção das iniciativas em diferentes campos. Para mitigar esse risco, recomendam o acompanhamento do desenvolvimento do algoritmo e a validação contínua dos resultados por profissionais.

Por fim, os entrevistados destacam os riscos relacionados à segurança de dados e ao cumprimento das diretrizes da LGPD. Todos os perfis de entrevistados citaram a importância da desidentificação dos dados para conformidade, ao mesmo tempo em que mencionaram a dificuldade técnica de desenvolver soluções baseadas em dados desidentificados, uma preocupação mais forte entre entrevistados do mercado e gestores de equipamentos de saúde. Para esses grupos, os riscos quanto à segurança e à privacidade dos dados são frequentemente considerados a partir da lógica de proteção da empresa. Nesse sentido, não se trata apenas de proteger os dados dos usuários de forma ética, mas também evitar riscos para o funcionamento da organização. Eles mencionam ações ligadas aos departamentos jurídicos das empresas para criar diretrizes e também como parte das equipes de acompanhamento das iniciativas.

IA NOS PONTOS DE CUIDADO

Em uma segunda etapa da pesquisa, foram conduzidas cinco entrevistas com profissionais de saúde que usam IA diretamente nos pontos de cuidado. Nessa etapa da coleta, foi utilizado um roteiro diferente, que buscou explorar algumas questões distintas em relação ao primeiro conjunto de entrevistas com *stakeholders*. O foco dessas entrevistas foi direcionado para temas específicos, tais como: os usos práticos das ferramentas de IA na prática clínica, o processo de adoção dessas ferramentas, as potencialidades e os desafios associados a seu uso cotidiano, as percepções sobre o novo cenário de trabalho, as potenciais mudanças na relação profissional-paciente, entre outros impactos cotidianos experimentados pelos profissionais de saúde. As experiências compartilhadas pelos entrevistados nessa etapa dão materialidade aos tópicos abordados anteriormente, enriquecendo a compreensão do papel da IA na prática clínica.

Antes de explorar cada um desses temas, serão descritas, de forma resumida, as ferramentas de IA citadas pelas pessoas entrevistadas. Duas entrevistadas atuam como enfermeiras de um plano de saúde, atendendo membros do plano por mensagens via *chat online*. Nesse processo, utilizam uma ferramenta de IA que acessa o registro das mensagens, extrai informações dessas interações e realiza uma síntese do diálogo com base em um comando pré-definido que organiza informações-chave, como data de início dos sintomas e sinais de alerta. Todas as informações fornecidas pelo paciente durante o atendimento no *chat* são convertidas em um resumo, que passa a compor seu prontuário.

Outras duas pessoas entrevistadas nessa etapa são da área de radiologia e atendem em hospitais. Por fim, a quinta pessoa entrevistada é da área de enfermagem intensivista, atualmente docente em uma universidade, e trabalha com ensino e pesquisa de IA na área da saúde. Esses três profissionais descreveram algumas ferramentas que fazem parte de seu dia a dia.

O primeiro é um sistema de IA que faz a detecção automática de lesões de esclerose múltipla (doença degenerativa do sistema nervoso central) em pacientes que fazem exames consecutivos. A detecção de lesões cerebrais é feita por meio de exames de ressonância magnética, cujo acompanhamento é realizado mediante exames repetidos, com um intervalo de alguns meses entre eles. No entanto, a avaliação é desafiadora para médicos e radiologistas, pois podem surgir milhares de lesões no cérebro ao longo do tempo. A comparação entre as imagens de exames sucessivos é necessária para identificar a presença de novas lesões, porém a complexidade dessa visualização aumenta significativamente; portanto, a aplicação de IA tem função auxiliar nesse processo, a fim de diminuir a possibilidade de erro.

Além desse sistema, relataram outros para detecção de doenças, como hemorragia intracraniana nas tomografias de crânio, especialmente úteis nos atendimentos de emergência. Se a hemorragia foi detectada pela IA, o médico recebe um alerta de que aquele laudo deve ser priorizado na fila. Como não é um sintoma visível, mas que demanda atendimento urgente, o tempo otimizado pode salvar a vida do paciente.

Às vezes, a gente tem aqui 60 exames para laudar. Cada exame desse de ressonância, às vezes a gente leva 15 a 20 minutos para laudar. Então, [com] 60 exames na fila, provavelmente esse exame vai ser laudado, aí, [em] um a dois dias, e o paciente vai embora para casa com um AVC. E isso não acontece mais por causa da ferramenta, então o tratamento de um AVC é muito mais agilizado com a ferramenta.

(PROFISSIONAL PONTO DE CUIDADO)

Na área de radiologia, algumas ferramentas de IA têm aprimorado a eficiência e a precisão do trabalho dos profissionais. O primeiro exemplo é um sistema de reconhecimento de voz para transcrição de laudos, implementado em 2013. Essa tecnologia otimiza o processo de documentação, pois permite que os radiologistas transcrevam laudos de maneira rápida e precisa, melhorando o fluxo de trabalho. Além disso, há ferramentas de aquisição e reconstrução de imagens integradas aos equipamentos de grandes fabricantes: são *softwares* incorporados aos aparelhos que otimizam a captura de imagens.

Quem trabalha com tomografia computadorizada hoje, a gente tem métodos de IA nos equipamentos que fazem com que a gente trabalhe com baixa dose de radiação e também aquisições mais rápidas dos exames.

(PROFISSIONAL PONTO DE CUIDADO)

Também foi citada uma solução de IA para otimização da gestão de insumos. Trata-se de um *software* em bomba injetora para reduzir o desperdício de iodo. A ferramenta identificou perdas significativas do insumo durante o abastecimento da bomba injetora em razão de técnicas inadequadas. O armazenamento e análise desses dados levaram a um treinamento da ferramenta que reduziu as perdas de 10-20 ml para, no máximo, 3 ml por abastecimento.

Além dessas ferramentas presentes na prática clínica ou na gestão hospitalar dos entrevistados, foram citadas algumas em fase de teste ou desenvolvimento, tais como: sistemas de AM para predição de mortalidade materna no puerpério, identificação de doenças crônicas não transmissíveis com metodologia de rede neural e um metaverso, simulação de ambientes com fins pedagógicos e formativos, em que alunos encenam personagens em atendimento fictício. Todas essas ações têm como objetivo a disponibilização gratuita pelo SUS.

Também destacamos o caso da universidade onde atua uma das pessoas entrevistadas no estudo, a qual tem introduzido ferramentas de IA na formação básica de profissionais de

saúde, formando um comitê de saúde digital para estimular o uso dessa tecnologia. Uma das ferramentas disponibilizadas é um aplicativo de apoio à decisão para paciente em condição de ventilação mecânica, especialmente projetado para estudantes. O aplicativo faz uma análise previa do paciente e oferece orientações sobre melhorias na oxigenação, ajustes na postura e posição, aspiração e parâmetros do ventilador.

Com a incorporação da IA na área da saúde, como nos exemplos anteriores, profissionais do setor têm experimentado uma transformação em suas práticas diárias. As ferramentas de IA estão desempenhando papéis variados, desde leituras de imagens mais eficientes até melhoria na gestão do atendimento ao paciente. No entanto, esse avanço tecnológico também levanta questionamentos a respeito dos desafios e dos limites de operacionalização das ferramentas, da necessidade de supervisão humana e das implicações éticas envolvidas. Nesse contexto, as percepções dos profissionais de saúde sobre o uso da IA trouxeram elementos relevantes para compreender como essas inovações têm sido recebidas no ambiente clínico.

IMPACTOS POSITIVOS PERCEBIDOS PELO USO DE IA NO PONTO DE CUIDADO

De maneira geral, as pessoas entrevistadas relataram experiências positivas do uso de sistemas de IA para o suporte às tarefas cotidianas de trabalhos:

> Otimiza muito o nosso tempo, porque, no dia a dia, era um tempo que a gente destinava muito grande para registrar, reler toda conversa que a gente teve, às vezes uma conversa super longa que a gente teve ali dentro do *chat*, então ela consegue resumir bastante e otimizar o nosso tempo.
>
> *(PROFISSIONAL PONTO DE CUIDADO)*

> Hoje, muitos dos médicos já não conseguem ficar sem as ferramentas [de IA]. Então, por exemplo, a esclerose múltipla, se dá algum problema que não processa no exame e vai ser laudado agora, as pessoas já ficam inseguras de laudar sem a ferramenta. Não que a ferramenta dispense a avaliação do médico: vai ter que abrir as duas imagens e comparar mesmo assim. A gente não aceita o que a máquina fala como verdade, o médico vai ter que avaliar. Mas como a máquina é extremamente sensível para detectar lesões novas, se ela não detectar, a chance de ter é quase zero. Então, quando a máquina não funciona e o exame não é processado [pela IA], as pessoas sentem falta, já criou como se fosse uma dependência entre aspas.
>
> *(PROFISSIONAL PONTO DE CUIDADO)*

Em algumas entrevistas, os impactos positivos dos sistemas

de IA apareceram de maneira detalhada. Nelas, foram feitas menções a contribuições positivas da tecnologia para a otimização do tempo dos processos de triagem ou tomada de decisão; para o aumento da segurança e da padronização dos registros, prontuários e documentação; para melhora da relação com os pacientes; e para o alívio da carga emocional que, por vezes, se impõe aos profissionais da saúde em decorrência de atendimentos complexos ou numerosos.

> Eu entendo que a IA minimiza a chance de falha de registro, de falta de registro mesmo. Então, ela traz segurança maior de deixar ali concreto tudo que a gente conversou com o paciente também, o que a gente direcionou para aquela pessoa.
>
> *(PROFISSIONAL PONTO DE CUIDADO)*

> Me permite ter um diálogo mais fluido com o paciente e desenvolver mais vínculo e conseguir fazer uma anamnese mais intensa e profunda, do que só ficar pensando o quanto eu vou ter que resumir tudo aquilo, que [...] IA já me traz esse resumo, independente do quanto eu conversei com aquela pessoa.
>
> *(PROFISSIONAL PONTO DE CUIDADO)*

> Se tornou um atendimento muito mais rápido porque a gente gastava muito mais tempo olhando para a mensagem do paciente e transcrevendo aquela mensagem com nossas palavras, do que a gente realmente interagindo com o paciente. Hoje em dia, como eu tenho a IA, eu posso focar mais no paciente do que no registro, porque eu sei que eu tenho um *backup* ali da IA para trazer todas as informações que eu trocar com paciente por mensagem, né? No *chat*.
>
> *(PROFISSIONAL PONTO DE CUIDADO)*

LIMITAÇÕES PERCEBIDAS E NOVAS DEMANDAS, A PARTIR DO USO DAS FERRAMENTAS DE IA

É importante registrar que as ferramentas de IA, embora sejam vistas pelos profissionais da saúde como um recurso que auxilia o trabalho, têm limitações e demandam supervisão constante. Como exemplo de limitação, a ferramenta de IA que sistematiza informações do paciente oferecidas no *chat* de atendimento não processa áudios nem lê imagens, por isso os profissionais no ponto de cuidado precisam interpretar esses dados, transformá-los em linguagem médica e fazer o registro manualmente.

Além disso, é consenso entre os entrevistados desse perfil (profissionais nos pontos de cuidado) que a aplicação de IA deve ser sempre um atalho, e não a produtora final de diagnóstico ou conteúdo. Esse ponto foi muito destacado, pois, como relataram os entrevistados, as ferramentas de IA generativa podem

"alucinar", ou seja, fornecer respostas sem lastro na realidade; assim, é sempre necessária a revisão de um profissional responsável. Entretanto, apesar da "alucinação" ser um ponto de atenção, ela não é suficiente para gerar desconfiança entre os profissionais entrevistados. Foi possível identificar entre os entrevistados que utilizam essas tecnologias na prática clínica haver um nível razoável de confiança nas ferramentas.

> Eu confio, porque ao mesmo tempo que ela pode alucinar, ela também pode trazer informações que eu esqueceria, então tem seu ônus e seu bônus. Acho que a gente não pode confiar 100%, por isso que estou aqui também, porque senão não existiria mais enfermeiros, não precisaria de enfermeiros, que bom que ainda estou aqui. Mas, aí, precisa passar por essa etapa com certeza, porque ela também, enfim, não traduz algumas coisas que um profissional de saúde traduziria, não é um profissional de saúde.
>
> *(PROFISSIONAL PONTO DE CUIDADO)*

> A gente precisa sempre revisar o fim do laudo, e ele também tem algumas palavras que são bem médicas, que a gente precisa treinar, né? Então, algumas palavras eles não sabem automaticamente, eles escrevem errado, no reconhecimento de voz, escreve errado, daí a gente arruma, ensina ele. Ele tem esse modelo de aprendizado, a aprendizagem do reconhecimento de voz, mas algumas palavras eles demoram para reconhecer e saber escrever de forma correta um termo que é muito médico.
>
> *(PROFISSIONAL PONTO DE CUIDADO)*

Para além das percepções iniciais e impactos na rotina de trabalho, outros temas foram questionados nas entrevistas, como potencialidades, melhorias, desafios e possíveis riscos do uso das ferramentas de IA na prática clínica, abordados a seguir.

POTENCIALIDADES, DESAFIOS E POSSÍVEIS RISCOS DO USO DE IA NA PRÁTICA CLÍNICA

Com relação aos potenciais dessa tecnologia, foi destacada pelos profissionais no ponto de cuidado a capacidade de melhoria da atenção primária com a prevenção de complicações dos sintomas dos pacientes e a redução de pessoas com sintomas leves no ambiente hospitalar. Acredita-se que IA pode ser útil nas estratégias de monitoramento dos pacientes em casos não graves.

Considerando o tema das possíveis melhorias, os entrevistados ressaltaram que as ferramentas de IA precisam de constantes ajustes e calibragens. Para que isso ocorra de forma eficiente, o profissional de saúde que a utiliza no cotidiano deve ter diálogo com o departamento de TI ou a equipe de desenvolvedores. Dessa forma, a ferramenta pode ser aprimorada constantemente

a partir do uso prático e de demandas concretas. De acordo com relatos, quanto mais manutenção e calibragem, menores as chances de "alucinação" das ferramentas de IA.

No começo, tinha esse problema de alucinar, então acontecia bastante essas alucinações e, aí, a gente sempre tinha que corrigir. A gente sempre ia dando também alguns *feedbacks*, por exemplo, para mudar o *prompt*, para melhorar ele para atender a nossa demanda.

(PROFISSIONAL PONTO DE CUIDADO)

Como a gente não pode confiar 100% nela por conta da alucinação, eu acho que talvez tenha esse risco de ter que passar por uma etapa humana ainda no final, e o humano errar, porque o humano erra, então acho que seria mais esse risco de não revisar e ela alucinou, aí acho que é um problema.

(PROFISSIONAL PONTO DE CUIDADO)

As entrevistas apontaram também desafios práticos na implementação efetiva de ferramentas de IA na área médica, tal como a resistência à adoção das novidades tecnológicas. Primeiro, destacaram que essas aplicações precisam estar muito bem integradas aos sistemas operacionais de hospitais e equipamentos de saúde, de forma que sua adoção não represente novos processos e etapas de trabalho. Quanto mais integradas aos sistemas e aos processos de trabalho vigentes, menor a chance de resistência à tecnologia e melhor sua adoção pelos profissionais de saúde.

Outro ponto importante de atenção é a facilidade do uso dessas aplicações e a oferta de interfaces intuitivas e descomplicadas. Tal elemento também influencia a adoção e a redução da resistência ao uso.

Vou falar uma experiência que eu tive em que não deu certo. Por exemplo, eu tentei fazer um *website* primeiro para a pessoa entrar e ver os resultados da ferramenta, ninguém usou, como eu falei, o pessoal precisa entrar no *site*, fazer o *login* e, aí, ver a informação lá e não vai fazer porque é muito difícil, então a maioria das ferramentas hoje, o processo já de entrega de resultados já é automático.

(PROFISSIONAL PONTO DE CUIDADO)

[Pode ser a] Melhor ferramenta de IA do mundo, mas se você não automatizar e integrar ela com o seu sistema do hospital, entregar as informações de forma fácil e rápida para os médicos usarem, ela não vai ser usada, porque os médicos, assim, a rotina é muito difícil, se adicionar alguns passos a mais, que seja alguns cliques na tela, vamos falar assim, a pessoa não vai usar.

(PROFISSIONAL PONTO DE CUIDADO)

A IA, tecnologia, é beleza, é ótima, se for alguma coisa intuitiva. Se não é intuitivo, tem uma resistência muito grande de começar a usar essa tecnologia [...] se aumenta o tempo de trabalho, pode esquecer, que a equipe não vai fazer aquilo, sabe? Se aumentar o tempo que a pessoa gasta, porque o médico, ele ganha por produção. Então, assim, se reduzir a produtividade do médico, a IA ou a tecnologia que está sendo implementada, eles não vão usar. [...] Se as pessoas não tiverem o treinamento e essa fase de adaptação, elas não vão utilizar a tecnologia.

(PROFISSIONAL PONTO DE CUIDADO)

Os médicos ainda têm uma resistência imensa, pelo menos aqui, no uso dessas ferramentas. Tem muito médico ainda que está no escrito, na receita escrita.

(PROFISSIONAL PONTO DE CUIDADO)

IMPLEMENTAÇÃO E ADOÇÃO NOS PONTOS DE CUIDADO

Além da resistência às novas ferramentas tecnológicas, as entrevistas trataram de outros aspectos envolvidos no processo de implementação e adoção pelos profissionais nos pontos de cuidado. De início, há sempre uma fase de adaptação, que pode ser mais ou menos fácil a depender do suporte técnico e do treinamento oferecido. Os informantes da pesquisa afirmaram que nem sempre há treinamento, embora acreditem que, se o sistema for simples, uma instrução inicial é suficiente, e o treinamento, desnecessário.

Mais uma instrução... Olha, vai aparecer assim, esse botão, vocês apertam, vocês revisam e apertam como concluído. Não tinha muito, não teve um treinamento.

(PROFISSIONAL PONTO DE CUIDADO)

Um exemplo da ferramenta de esclerose múltipla é assim... A informação já chega pronta para o médico, ele nem sabe como funciona a ferramenta, chega a informação pronta para ele. Assim que ele vai laudar, a informação está pronta, então é muito fácil a adoção.

(PROFISSIONAL PONTO DE CUIDADO)

Dependendo da ferramenta e das circunstâncias, entretanto, o processo de adaptação pode não ser tão simples. Uma pessoa entrevistada relatou um caso de transição entre duas versões de uma ferramenta de IA aplicada à radiologia que não ocorreu conforme o esperado e, em vez de aumentar a produtividade, gerou atrasos nos processos de produção de laudos.

Na troca da versão, nosso reconhecimento de voz ficou pior, as máscaras ficaram mais difíceis, então eles ficaram aqui mais uma semana nos ajudando e fazendo esse treinamento. E, assim, eu vou te dizer que a mudança do hábito ela é bem complicada, porque os médicos radiologistas eles são bem metódicos e são obsessivos, que não pode perder nenhuma lesão, não pode. E, aí, tu muda a maneira como a pessoa lauda, isso demora um tempo para ela conseguir se adaptar com a nova; mesmo que a ferramenta seja melhor, ele tem uma resistência inicial de trabalhar com essa nova tecnologia, porque não era aquela que ele já está acostumado e já é mais rápido. E, assim, foi bem doloroso para nós, esses 2 primeiros meses que era para eu aumentar a produtividade, teve uma redução de produtividade. Até todo mundo se adaptar com as mudanças da nova versão. Então, agora que já faz seis meses que a gente trocou a versão, agora que o reconhecimento de voz está ficando como era antes.

(PROFISSIONAL PONTO DE CUIDADO)

Também foi ressaltada a importância de um esforço de preparo mais abrangente dos profissionais de saúde para atuar nesse novo cenário de trabalho com a presença da IA. Seria importante treiná-los não apenas para operar os sistemas, mas também para compreender de modo mais aprofundado como eles são construídos e quais as implicações de seu uso. Ainda falta avançar em processos de transparência e explicabilidade dessas soluções até para que os profissionais usuários entendam as responsabilidades e as consequências de sua utilização.

Outro ponto importante é saber um pouco como funcionam essas ferramentas e saber avaliar se essa ferramenta foi bem-produzida. [...] Os médicos hoje eles não precisam saber a fundo programação, saber de IA, eles só precisam saber avaliar os estudos dessas ferramentas, se o *dataset* foi bem-feito, se as estatísticas foram bem-feitas, se o estudo foi bem desenhado, para ele poder avaliar se a ferramenta que daqui a pouco vai estar usando no hospital dele, contratada de uma empresa privada ou alguma outra instituição, se essa ferramenta realmente é boa e foi bem-feita. Acho que isso é uma coisa importante também, que muitos ainda não sabem avaliar um trabalho de IA, não sabem que o que avaliar para dizer se o trabalho foi bem-feito ou não.

(PROFISSIONAL PONTO DE CUIDADO)

RESPONSABILIDADE DAS DECISÕES E A RELAÇÃO MÉDICO-PACIENTE

As pessoas entrevistadas acreditam que, no caso de mau uso ou de decisões equivocadas baseadas em conteúdo gerado por IA, a responsabilidade é sempre humana, e os fatores determinantes para o bom uso dessas ferramentas são: comprometimento, ética profissional e diretrizes claras para o uso da IA.

Tem que treinar os profissionais que vão usar essas ferramentas para ter a conscientização de que as ferramentas não acertam 100% das vezes. Elas erram como um ser humano erra, às vezes menos ou mais, mas elas erram, e que não se dispensa a avaliação deles do exame.

(PROFISSIONAL PONTO DE CUIDADO)

A responsabilidade é sempre do médico. A gente sempre fala, você vê a ferramenta, mas olha o exame como se não existisse a ferramenta. Então, a ferramenta serve como um auxílio, mas não substitui a avaliação do médico.

(PROFISSIONAL PONTO DE CUIDADO)

No que diz respeito à relação médico-paciente, a percepção das pessoas entrevistadas é de que as ferramentas de IA são benéficas, pois, com a otimização do tempo do profissional de saúde, há mais possibilidade de interação de qualidade com o paciente.

A gente consegue destinar mais tempo para o paciente do que se concentrar em preencher papel digital. A gente tinha muito essa preocupação de "eu tenho que registrar de uma forma completa, não posso deixar passar as informações, tem que estar dentro daquela metodologia, vou gastar um tempo, estou atendendo três pessoas ao mesmo tempo". Ali, estou falando com três pessoas e estou tendo que garantir qualidade nos meus atendimentos, claro. Então, quando eu tiro um pouco essa carga mental e esse tempo de uma coisa burocrática, de uma etapa burocrática do meu trabalho, eu consigo concentrar mais na assistência ao paciente, com certeza. Consigo garantir mais tempo até para ler o prontuário dele, conhecer mais sobre ele, entender melhor o histórico clínico dele para eu atender ele naquele momento. Então, eu acho que sim, o paciente, talvez ele não saiba, mas nos bastidores a gente consegue otimizar bastante para ele.

(PROFISSIONAL PONTO DE CUIDADO)

As entrevistas revelam que nem sempre o paciente tem conhecimento de que uma ferramenta de IA apoia o profissional de saúde em seu atendimento ou na análise de seus dados. De acordo com os entrevistados, na maioria dos casos, os equipamentos de saúde não reportam, notificam ou consultam os pacientes, que são seus beneficiários finais, sobre essa questão. Como grande parte das ferramentas em uso hoje são mais voltadas para processos internos, como triagem, registro ou prontuário, os profissionais entrevistados acreditam que essa tecnologia não interfere diretamente no atendimento nem consideram necessário informar ao paciente sobre seu uso.

Quanto que ele sabe se eu estou usando IA ou não? Em todos esses métodos? Eu acho que ele não sabe. Que eu usei um reconhecimento de voz? Que eu usei uma técnica de redução de radiação? Não. Eles sabem que eu tenho um equipamento mais avançado, mas ele não tem essa precisão.

(PROFISSIONAL PONTO DE CUIDADO)

O paciente em si nem sabe que essas ferramentas estão funcionando. O que impacta é realmente no tratamento e na agilidade do tratamento pensando na detenção de um AVC precoce, na detenção da hemorragia precoce. Esses pacientes que têm essas alterações, eles são tratados mais rapidamente aqui no nosso hospital.

(PROFISSIONAL PONTO DE CUIDADO)

Embora a ausência de práticas de notificação e consulta sobre a aplicação de ferramentas de IA não tenha surgido como um tema de grande preocupação para os profissionais entrevistados, um deles apresentou uma reflexão mais aprofundada, acenando para uma possível mudança no cenário futuro. Isso pode vir a acontecer por meio de uma participação mais ativa do paciente no cuidado à saúde, como expressão de uma apropriação maior do uso da tecnologia em sua vida. Para tanto, a perspectiva do beneficiário final deveria ser levada em consideração no atendimento clínico e também no desenvolvimento e na implementação das ferramentas tecnológicas.

Não pode ser uma pessoa que só absorva. Vão ter aqueles que vão absorver mais. [...] Precisa ser um personagem da sua própria saúde, passa a exigir mais qualidade. Ele não é um ser passivo que fica ali só esperando o resultado, ele conhece o seu corpo, ele conhece saúde, ele sabe onde está muita coisa atrapalhando, precisa ser ouvido e as tecnologias elas precisam ter esse arcabouço.

(PROFISSIONAL PONTO DE CUIDADO)

As pessoas entrevistadas afirmam que os dados dos pacientes utilizados por essas ferramentas estão protegidos, cujas práticas são aprovadas pelos comitês de ética hospitalar. De acordo com os relatos, antes de serem processadas pela IA, todas as imagens e as informações passam por um processo de anonimização, a fim de garantir a proteção da identidade. Nenhuma informação pessoal é utilizada, e os exames não são empregados para o retrabalho da máquina, reforçando o cuidado na manipulação dos dados. Ademais, em todas as experiências relatadas, os bancos de dados são privados e não há compartilhamento deles, o que assegura a integridade e a confidencialidade dos dados dos pacientes envolvidos.

PERSPECTIVAS DE FUTURO PARA OS PROFISSIONAIS DE SAÚDE

Um último tópico abordado nas entrevistas com profissionais do ponto de cuidado tratou das perspectivas de futuro para

os profissionais de saúde, considerando o uso da tecnologia de IA na prática clínica e nos pontos de cuidado.

Em uma das entrevistas, um profissional da medicina, pensando no futuro dos profissionais, destacou a importância de desenhar uma reforma curricular que priorize o tema da tecnologia na saúde. Segundo ele, essa mudança é essencial para reduzir a resistência dos profissionais de saúde em incorporar a IA em seu cotidiano. Ele argumentou que, quanto mais cedo os profissionais forem preparados, mais rapidamente compreenderão os benefícios da IA em seu trabalho. Além disso, essa abordagem ajudaria a normalizar a visão sobre o tema, dissipando o temor de uma substituição completa do ser humano pela máquina com o advento das tecnologias de IA na área de saúde.

Os profissionais da enfermagem que incorporam a IA em suas práticas diárias não acreditam que serão substituídos e destacam dois fatores essenciais: o necessário acolhimento humano na relação com o paciente e a experiência na prática clínica.

> Eu não sou totalmente substituível, tem algumas coisas que eu faço e que eu realmente gostaria mesmo que fosse, porque são questões de trabalho muito burocráticas que a gente perde muito tempo e ninguém que faz enfermagem, medicina, estuda para preencher papel, isso é um fato. Quando a gente tem que fazer isso, a gente fica cansada daquela etapa, porque a gente estudou para atender a pessoa, para cuidar das pessoas, e não para fazer essas etapas burocráticas que uma máquina faz.
>
> *(PROFISSIONAL PONTO DE CUIDADO)*

> Eu tenho certeza que a tecnologia não vai resolver tudo, mas ao mesmo tempo os profissionais precisam estar mais abertos, mais dinâmicos, mais interessados e se preparando cada vez mais para viver uma nova era. Ela já está aí, ela já atropelou todo mundo, então ou a gente acorda e vê onde eu posso contribuir, o que aquilo contribui para mim, o que é melhor na minha profissão, qual o impacto, ou a gente acorda para isso ou realmente vai ir aí um trator.
>
> *(PROFISSIONAL PONTO DE CUIDADO)*

Essa etapa da pesquisa revelou, portanto, que a IA traz benefícios significativos para a prática de trabalho de profissionais de saúde nos pontos de cuidado, proporcionando otimização do tempo, redução da carga mental e melhoria na relação profissional-paciente. Contudo, os desafios operacionais, os custos e a necessidade de supervisão humana destacam a complexidade da implementação da IA na área da saúde. Logo, a questão da ética, da responsabilização médica e da consideração da perspectiva do paciente emergem como elementos ainda pouco explorados na adoção da IA na saúde.

APONTAMENTOS FINAIS

A pesquisa qualitativa coletou percepções e experiências de atores estratégicos da área de saúde, com atuação em IA de diferentes esferas: academia, poder público, mercado e equipamentos de saúde. Nosso objetivo foi compreender o estágio das iniciativas de IA aplicadas à saúde no país e identificar temas, anseios e expectativas presentes na agenda pública, com o intuito de apresentar um panorama dessa temática no cenário brasileiro.

As entrevistas revelaram que o país está em um estágio inicial de desenvolvimento e implementação de ferramentas de IA na saúde. Há um clima de otimismo e grande expectativa quanto ao potencial da IA, especialmente no setor privado, em que o desenvolvimento está mais avançado em comparação com o setor público. Os principais avanços mencionados incluem o uso da IA para melhorar processos de assistência médica e gestão, com ênfase na ampliação da eficiência e na redução de custos.

Apesar do entusiasmo, os entrevistados do estudo reconhecem que o país está apenas engatinhando nessa jornada e enfrenta desafios, como falta de estratégia nacional, questões regulatórias e deficiências na qualidade e na integração de dados. A evolução da IA na saúde brasileira é percebida como desigual, com o setor privado liderando o movimento, enquanto o setor público enfrenta desafios burocráticos e de escassez de recursos. As principais lacunas apontadas são relacionadas à qualidade dos dados e à falta de regulação específica para usos de IA na saúde. Segundo os entrevistados da pesquisa, é essencial enfrentar esses desafios para avançar de forma significativa no campo da IA aplicada à saúde no Brasil.

Em relação às potencialidades da IA, foram destacados três eixos principais de oportunidades: melhorias para pacientes, profissionais de saúde e prestadores de serviços. No âmbito dos pacientes, a IA tem potencial de ampliar o acesso aos serviços de saúde e melhorar a precisão diagnóstica, ao passo que, para os profissionais de saúde, as ferramentas de IA podem reduzir processos burocráticos, otimizar o tempo gasto em tarefas administrativas e oferecer apoio na tomada de decisões clínicas, promovendo diagnósticos mais rápidos e precisos, além de ampliar a capacidade de atendimento em

regiões com escassez de especialistas. Já para os prestadores de serviços de saúde, a IA pode trazer, tanto no setor público quanto no privado, eficiência operacional, otimização de recursos e melhoria nos processos de gestão, logística e atendimento.

Todo esse potencial também pode servir para o enfretamento das desigualdades envolvidas na oferta de saúde, ampliando acesso aos serviços e reduzindo o impacto da escassez de recursos em regiões mais desfavorecidas do país. Em resumo, a IA é percebida como uma ferramenta capaz de promover melhorias significativas em diversos aspectos da saúde no Brasil, desde a gestão operacional até o cuidado clínico, com benefícios tanto para os profissionais de saúde quanto para os pacientes e a sociedade como um todo.

Ainda com relação às oportunidades, o SUS foi apontado como o grande diferencial brasileiro em relação à obtenção de dados, volume e diversidade suficientes para o desenvolvimento de algoritmos e aplicações robustas e confiáveis. A RNDS, na visão dos entrevistados, tem um enorme potencial para promover interoperabilidade e criar um ambiente propício para o desenvolvimento e aplicação da IA no setor de saúde brasileiro. Para tanto, os dados precisam ser de qualidade, e não apenas em grande volume, o que implica a importância de protocolos e rotinas que assegurem, entre outros elementos, a padronização de informações de saúde.

Nesse clima de otimismo com a IA na área da saúde, foi identificada certa dificuldade de resposta dos entrevistados em relação aos possíveis riscos implicados no uso da IA. Contudo, destacaram-se preocupações relacionadas à privacidade e à segurança dos dados dos pacientes. O risco de vazamento de informações sensíveis foi apontado como uma questão relevante, considerando a possibilidade de uso inapropriado dos dados utilizados e/ou gerados pelas ferramentas de IA. Já a discussão sobre os direitos dos usuários dos sistemas de saúde apresentou limitações, com ênfase na necessidade de educação digital para promover a compreensão dos benefícios da tecnologia. Também foram mencionados riscos relacionados à ética, embora essa questão não tenha sido tão proeminente nas falas. Para mitigar os potenciais riscos do avanço da IA na saúde, é preciso que haja transparência nas decisões

que embasam a construção das ferramentas, estrutura clara de governança para avaliação de riscos e desenvolvimento de algoritmos livres de vieses, a fim de que produzam resultados confiáveis, de acordo com os entrevistados.

Ainda a respeito da questão dos riscos potenciais, a responsabilidade por possíveis equívocos da IA são, em consenso, do profissional que usa a ferramenta em sua prática de trabalho. Assim, a pesquisa revelou a dificuldade de elaboração sobre mecanismos sociais e coletivos de ordenamento ético, com prevalência da percepção de responsabilização individual diante de decisões baseadas em IA.

O estudo também mapeou ações, iniciativas e projetos em andamento no país de IA na saúde, nos quatro segmentos de *stakeholders* entrevistados. Relatos de iniciativas que têm ocorrido no âmbito acadêmico e nos setores público e privado foram coletados com o objetivo de apresentar um panorama amplo sobre o cenário atual brasileiro. De maneira geral, as iniciativas de IA aplicada à saúde em andamento no Brasil visam otimizar processos de gestão e aprimorar a qualidade do cuidado. Cada esfera tem seus focos específicos: na academia, destaca-se o desenvolvimento de algoritmos com ênfase na diversidade populacional brasileira e na correção de vieses algorítmicos, pois podem ter implicações sociais nocivas; no mercado, há um foco na IA generativa para criar produtos variados, como *chatbots* para atendimento ao paciente, além de ferramentas para otimizar diagnósticos por imagens; no poder público, o foco está na digitalização e na interoperabilidade de dados; e nos equipamentos de saúde, há uma ênfase em projetos diversos para assistência ao diagnóstico e à gestão. As similaridades e as divergências observadas entre as iniciativas desenvolvidas nos diferentes segmentos investigados ilustram as potencialidades das ferramentas e as dificuldades no desenvolvimento e implementação.

Por fim, como forma de dar materialidade os temas abordados, o estudo dedicou-se a ouvir a experiência de cinco profissionais de saúde que utilizam soluções de IA em suas tarefas cotidianas nos pontos de cuidado. As entrevistas exploraram os usos práticos das ferramentas de IA na prática clínica, o processo de adoção dessas ferramentas, as potencialidades, os desafios e os impactos cotidianos experimentados

por esses profissionais. Eles relataram experiências positivas, destacando a otimização do tempo, a diminuição da carga mental e, como decorrência, a melhoria na relação com os pacientes. No entanto, também mencionaram desafios na implementação efetiva das ferramentas de IA, incluindo a resistência à adoção pelos profissionais da área de saúde, tanto pela dificuldade de adaptação quanto pela falta de treinamento adequado, e a necessidade de integração das soluções de IA com os sistemas existentes nos equipamentos de saúde.

Houve ênfase na importância da supervisão humana na adoção de soluções de IA, na necessidade de constante calibragem das ferramentas utilizadas e na responsabilidade ética dos profissionais de saúde em relação aos diagnósticos apoiados por essa tecnologia. Quanto ao futuro, os entrevistados do ponto de cuidado destacaram a necessidade de reformas curriculares para incluir o tema da tecnologia na saúde na formação dos profissionais, de modo a prepará-los para a utilização da IA. Também refletiram sobre a importância do acolhimento na relação com o paciente e a valorização da experiência clínica dos profissionais, pois, em sua percepção, o atendimento humano sempre será imperativo em relação às tecnologias.

REFERÊNCIAS

Academia Brasileira de Ciências. (2023). *Recomendações para o avanço da Inteligência Artificial no Brasil: GT-IA da Academia Brasileira de Ciências*. https://www.abc.org.br/wp-content/uploads/2023/11/recomendacoes-para-o-avanco-da-inteligencia-artificial-no-brasil-abc-novembro-2023-GT-IA.pdf

Adler-Milstein, J., Aggarwal, N., Ahmed, M., Castner, J., Evans, B. J., Gonzalez, A. A., James, A. C., Lin, S., Mandl, K. D., Matheny, M. E., Sendak, M. P., Shachar, C., & Williams, A. (2022). Meeting the moment: Addressing barriers and facilitating clinical adoption of Artificial Intelligence in medical diagnosis. *NAM Perspectives, 2022*. https://www.gao.gov/products/gao-22-104629

Challen, R., Denny, J., Pitt, M., Gompels, L., Edwards, T., & Tsaneva-Atanasova, K. (2019). Artificial Intelligence, bias and clinical safety. *BMJ Quality and Safety, 28*(231-237). https://qualitysafety.bmj.com/content/28/3/231

Lei Geral de Proteção de Dados Pessoais (LGPD). (2018). Lei n. 13.709, de 14 de agosto de 2018. Dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural. https://www.planalto. gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm

Ministério da Saúde. (s.d.). *Rede Nacional de Dados em Saúde*. https://www.gov.br/saude/pt-br/composicao/seidigi/rnds

Ministério da Saúde. (2020). *Estratégia de Saúde Digital para o Brasil 2020-2028*. https://bvsms.saude.gov.br/bvs/publicacoes/estrategia_saude_digital_Brasil.pdf

CONSIDERAÇÕES FINAIS

Direcionadores para as políticas públicas de Inteligência Artificial na saúde

Glauco Arbix[1] *e João Paulo Cândia Veiga*[2]

1 Professor titular do Departamento de Sociologia da Universidade de São Paulo (USP).
2 Professor de Ciência Política e Relações Internacionais da USP e pesquisador do Centro de Inteligência Artificial (C4AI).

INTRODUÇÃO

A área da saúde é espaço econômico decisivo para o Brasil. Além de responder por 9% do Produto Interno Bruto (PIB) brasileiro e gerar cerca de 9 milhões de empregos diretos, a saúde oferece uma oportunidade enorme para o país entrar em sintonia com o novo ciclo tecnológico, que tem em seu coração uma das tecnologias mais poderosas que a humanidade criou: a Inteligência Artificial (IA).

A IA é profundamente transformadora da pesquisa, da economia e da sociedade. Por sua flexibilidade e potência, ela se consolida cada vez mais como uma tecnologia de propósito geral, da mesma linhagem que a eletricidade, o automóvel, a televisão, o computador e a Internet. Suas características diversificadas e multifacetadas facilitam a integração a outras disciplinas e ferramentas, que sustentam sua penetração em praticamente todos os âmbitos da economia e da sociedade.

Porém, mais do que seu alcance, a IA tornou-se imprescindível para a geração de inovações. Na saúde, protagonizou ou compartilhou enormes avanços na bioquímica, na biologia, na genética, nos materiais, na energia e em todos os domínios que formam e informam os cuidados modernos da população. O mesmo acontece na área da gestão, na integração entre ciência e infraestrutura médica, além de, potencialmente, contribuir com o aumento da transparência para usuários de serviços de saúde. E ocorre o mesmo na pesquisa avançada de novos fármacos e medicamentos, nos processos de fabricação, nos equipamentos, na elevação da qualidade e na eficiência dos serviços necessários para garantir um direito elementar da população: o direito à saúde. Por todas essas potencialidades, abordar estratégias para a saúde sem colocar ênfase e foco nos impactos das novas tecnologias de IA tornou-se praticamente impeditivo.

A preocupação central que inspirou as sugestões para a elaboração de uma estratégia pública para a saúde, apresentada neste texto, está informada pela necessidade de uma agenda de mudanças no Sistema Único de Saúde (SUS), a começar pela ampliação significativa do acesso sustentável aos serviços públicos e privados que cuidam do bem-estar da população.

As tecnologias de IA, se utilizadas de forma ética e de modo responsável, podem contribuir para a implementação de políticas de saúde mais eficientes e para elevar o patamar de acesso aos serviços de saúde, em especial da população mais vulnerável. Esse horizonte adquire especial relevância em um país como o Brasil, marcado por carências históricas e por desigualdades sociais profundas, que se expressam na renda, nas disparidades regionais e nas limitações atuais para o SUS oferecer um atendimento de maior qualidade a um amplo espectro de procedimentos médicos, do básico aos de maior complexidade[3].

EXPERIÊNCIAS DE IA NA SAÚDE

Em sintonia com as tendências mundiais, apenas em 2023, mais de 30 milhões de atendimentos médicos foram feitos à distância apenas no Brasil, segundo dados da Federação Nacional de Saúde Suplementar (FenaSaúde, 2023), um forte contraste com as 11 milhões de consultas remotas realizadas entre 2020 e o final de 2022[4]. Em 2024, o Ministério da Saúde (MS) espera mais de 50 milhões de atendimentos (Folha de São Paulo, 2024).

A telessaúde, além da medicina, engloba o atendimento remoto em várias áreas, a exemplo da enfermagem, da fisioterapia e da psicologia. Pelas regras brasileiras estabelecidas, o profissional de saúde tem autonomia para decidir sobre o uso dessa prática, inclusive em relação à primeira consulta. A legislação também garante o direito da recusa ao atendimento à distância pelo paciente, assim como a confidencialidade de seus dados. Segundo o MS, apenas em 2023, foram atendidos 1,2 mil municípios com eletrocardiogramas realizados à distância, com uma média de 6 mil laudos por dia. Os planos do MS para ampliar a digitalização da saúde incluíram vários procedimentos de telessaúde, como mecanismos que melhoram o acesso, o acompanhamento e a continuidade do atendimento ao paciente, além de aprimorar a gestão das filas de espera. Trata-se

3 Algumas recomendações oferecidas por este trabalho foram baseadas em análises contidas no documento *Proposta Saúde São Paulo 2022*, organizado pelo SindHosp (2022) e elaborado a partir de consultoria técnica do Centro Brasileiro de Análise e Planejamento (CEBRAP).

4 Apesar desse crescimento e de sua praticidade, o atendimento à distância foi regulamentado pelo Conselho Federal de Medicina (CFM) apenas em 2022 (Lei n. 14.510/2022), dado que a prática da telessaúde havia sido permitida em caráter emergencial durante a pandemia COVID-19 (Lei 13.989/2020).

de um recurso que não substitui todas as modalidades do cuidado médico, mas que complementa o atendimento presencial, com ganhos significativos no acesso aos cuidados básicos, no acompanhamento de pacientes com doenças crônicas, no atendimento às áreas distantes que convivem com um serviço ainda precário, além da redução de custos e praticidade. Dado esse elenco de vantagens, o número de clínicas e hospitais que usam a telessaúde aumentou rapidamente para englobar as áreas de gestão, planejamento, previsão de riscos de internação, entrega e busca de materiais para exame.

Os sistemas de telessaúde têm se mostrado cada vez mais dependentes das tecnologias de IA. Essa sinergia aumentou o espectro do atendimento à saúde a partir da inclusão de procedimentos de maior complexidade, como a análise de imagens e o inter-relacionamento entre sintomas e biomarcadores de dados clínicos para caracterizar e prognosticar doenças, viabilizar o uso de sensores, promover cirurgias e monitorar pacientes (Joshi *et al.*, 2021; Liu *et al.*, 2022; Weenk *et al.*, 2020). A incorporação dos recursos e metodologias de IA impulsionou aplicações dedicadas a medição de sinais vitais, reconhecimento de movimento e mesmo parâmetros de cognição capazes de indicar confusão, queda e alterações psicológicas (Shaik *et al.*, 2022). Há exemplos abundantes de algoritmos que alavancaram em diferentes países e regiões uma massa de atividades exploratórias que procuram mudar o volume e a qualidade do atendimento em saúde, com ganhos expressivos para a população, em especial quanto à parcela menos favorecida.

O mesmo ímpeto é observado na pesquisa médica, tanto no sequenciamento de genomas quanto no desenvolvimento de novas drogas e tratamentos, viabilizados por técnicas de aprendizado de máquina (AM). Médicos e cientistas têm sido cada vez mais assistidos por agentes artificiais na interpretação de grandes volumes de dados em curto espaço de tempo (Helm *et al.*, 2020; Krittanawong *et al.*, 2022), voltados para prever a deterioração precoce da saúde de pacientes que convivem com disfunções crônicas (Guo *et al.*, 2022; Liu *et al.*, 2022). Sistemas de IA, crescentemente, conseguem processar dados para reconhecer e identificar padrões que ajudam na tomada de decisões dos profissionais da saúde (Dean *et al.*, 2022). Na realidade, o aumento da capacidade e da velocidade

computacional em anos recentes levou ao desenvolvimento de redes neurais artificiais e algoritmos de *deep learning* (DL) que movimentam bancos de dados de alta complexidade (Kalfa *et al.*, 2020), automatizam e controlam tarefas, a fim de evitar erros humanos e aumentar a segurança do paciente (Tandel *et al.*, 2022).

Estudo com base nos bancos de dados da Web of Science, Scopus, Springer, Pub Med, Science Directy e ACM Digital Library (Shaik *et al.*, 2023) identificou uma ampla gama de domínios da saúde em que a IA impulsionou a telemedicina, com impacto em países avançados e emergentes. Os resultados da pesquisa apontaram que a telemedicina movida à IA progride aceleradamente nos Estados Unidos da América (EUA) e na China desde a pandemia, tendência registrada também no Brasil. Levantamento da Kaiser Family Foundation de 2023 mostrou que a quase totalidade dos planos de saúde nos EUA oferecia procedimentos de telemedicina no ano de 2022, enquanto no Japão a telemedicina passou a integrar o sistema de saúde público com uma estratégia própria e com foco na IA. O MS japonês implantou serviços de medicina à distância em mais de dois mil hospitais e clínicas em todo o país para tratar de doenças crônicas, apoiar atendimentos de emergência e realizar consultas remotas em regiões afastadas. No Reino Unido, a telemedicina é parte integrante do National Health Service (NHS) – sistema de saúde britânico que inspirou o SUS brasileiro –, que elegeu a IA para analisar dados dos pacientes, elaborar diagnósticos preliminares e viabilizar mecanismos que previnem a hospitalização. No mesmo sentido, a Organização Mundial da Saúde (OMS), em 2024, indicou que tecnologias de IA estavam atuantes nos sistemas de telessaúde em todos os países da Europa (OMS, 2024).

RISCOS DA IA NA SAÚDE

Esse rápido avanço da IA na saúde não ocorre sem riscos técnicos e éticos. Do ponto de vista jurídico e de proteção da sociedade, a regulamentação do uso de dados é essencial, a começar pela privacidade e pela proteção de dados de usuários de serviços de saúde, reconhecidamente dois direitos básicos distintos dedicados à proteção, à autonomia e à dignidade dos indivíduos. Em geral, esses dois direitos

estão lavrados em regulamentações gerais de proteção de dados, a exemplo da que existe na União Europeia (UE) e no Brasil. Isso significa que as ferramentas que processam dados de saúde são obrigadas a se adequarem aos requisitos legais, de modo a garantir sua licitude, transparência, limitação de finalidade, exatidão, limitações de armazenamento, integridade, confidencialidade, assim como as limitações decorrentes de sua governança.

A saúde é um campo especial em que as orientações e salvaguardas éticas são fundamentais, uma vez que cuidados de saúde impactam a vida dos pacientes e seus familiares. Isso significa que o esforço de regulação e padronização do uso de dados, assim como de elaboração de leis, deve estar fundamentado em princípios éticos que regem a medicina. O desafio, no entanto, é remodelar as referências éticas para dar conta de processos digitalizados, novos e interativos, realizados via *display* de telefone, *tablet* ou de computador, e configurados de um modo radicalmente diferente do que a medicina utiliza há séculos. Não à toa, a segurança do paciente, a transparência e a explicabilidade dos procedimentos utilizados são essenciais para que sejam aceitos e incorporados por usuários e profissionais às práticas médicas, o que nem sempre é fácil, dada a natureza dos novos recursos.

Muitos algoritmos são mais eficientes que os humanos no tratamento e na interpretação de dados complexos, assim como na previsão de resultados; contudo, não são capazes de demonstrar como essas conclusões foram alcançadas ou se houve falhas em seu percurso estatístico, pois a estrutura das redes neurais profundas não é autoexplicativa: reveste-se de grande opacidade e expressa uma dinâmica conhecida como *black box* (Yang *et al.*, 2022). Desse modo, é impossível prever o que as redes neurais aprendem efetivamente com os dados; a partir daí, podem mobilizar recursos que discriminam informações dos pacientes (Chen *et al.*, 2021).

Essa incerteza, que se torna maior quando há desequilíbrios e falhas na formação dos bancos de dados, compromete a rápida adoção dessas tecnologias na saúde. Ou seja, mesmo tendo consciência que as aplicações no campo da saúde têm sido profundamente transformadas pela IA, com avanços significativos na telessaúde, desafios como o adequado

tratamento dos dados, a explicabilidade e a privacidade precisam ser equacionados para diminuir a incerteza e aumentar sua confiabilidade.

Ainda que iniciais e limitados, há experimentos consistentes que tentam ampliar a transparência dos sistemas de IA com indicações do percurso trilhado para chegar às suas decisões que, sem carregar nos atributos técnicos, buscam aumentar a confiança de médicos e profissionais de saúde (Lauritsen *et al.*, 2020). A busca de critérios de confiabilidade também forma o pano de fundo quando os médicos são levados a se afastar dos diagnósticos convencionais, baseados em médias populacionais, e se esforçam para considerar a variabilidade individual dos pacientes e de suas respostas aos tratamentos. Nesse sentido, a articulação da telemedicina com recursos da IA permite um monitoramento centrado na vida clínica do paciente e de modo personalizado, o que tem se mostrado uma ferramenta especialmente útil no tratamento de doenças crônicas como perturbações de saúde mental, diabetes, cardiopatias e outras (Mukherjee *et al.*, 2020). Apesar de promissores, esses procedimentos precisam ser ancorados em plataformas de IA e armazenados em nuvem para que os dados sejam analisados, o que coloca, de imediato, preocupações de privacidade e segurança sobre os dados de saúde, além de seu alto custo, por mobilizar enormes recursos tecnológicos e consumo de energia.

Mesmo com esses riscos, a pesquisa médica intensifica o uso de IA devido às vantagens oferecidas quando comparadas a práticas científicas tradicionais e procedimentos diagnósticos, clínicos e cirúrgicos. O objetivo central dos pesquisadores e dos experimentos é construir um quadro personalizado e confiável dos pacientes, capaz de captar sinais de declínio de suas condições de saúde o mais cedo possível e, com isso, prever eventos adversos e melhorar a precisão dos tratamentos e de seus resultados[5].

Os trabalhos citados nos capítulos anteriores dão uma pequena amostra tanto do vigor das pesquisas em curso quanto da cautela necessária para que a IA se consolide como

5 Em especial no monitoramento de pacientes com doenças crônicas, há uma profusão de algoritmos e wearables voltados também para prescrever a dosagem e o tempo da medicação (Watts *et al.*, 2020; Wu *et al.*, 2022; Yu *et al.*, 2023; Zheng *et al.*, 2021).

um poderoso instrumento da prática médica cotidiana. E a confiabilidade dos procedimentos, recursos e ferramentas é um dos maiores desafios – senão o maior – que a pesquisa enfrenta atualmente na medicina (Mohanty & Mishra, 2022).

Os benefícios que a IA pode trazer para as áreas da saúde não apagam a alta complexidade dessa tecnologia. Nesse sentido, é fundamental que as instituições estejam comprometidas com o uso responsável da IA para minimizar riscos e maximizar oportunidades. A OMS (2021) divulgou diretrizes para nortear o uso ético da IA, sintetizadas em seis pontos: (a) O controle dos sistemas e as decisões sobre saúde devem ser tomados por humanos e não exclusivamente pela IA; (b) Os desenvolvedores e os responsáveis pelos sistemas de IA devem monitorar e garantir o pleno funcionamento de todas as ferramentas, de modo a garantir o atendimento de todas as regras de segurança; (c) Os desenvolvedores também são responsáveis por publicar dados e informações sobre os produtos e seu manuseio de modo totalmente transparente; (d) Sistemas de saúde que adotam a IA devem garantir treinamento adequado e capacitação dos profissionais responsáveis pelas ferramentas; (e) A IA deve ser treinada com dados de diversas nacionalidades, gêneros e etnias, com o objetivo de promover a diversidade e evitar algoritmos enviesados; (f) As ferramentas de IA devem ser continuamente avaliadas e aperfeiçoadas com base em seu desempenho.

Apesar das diretrizes da OMS, sabe-se que as novas tecnologias digitais se multiplicam muito mais rapidamente do que o debate e o esforço de regulação e disciplina dos agentes de saúde, com o estabelecimento de regras para circunscrever sua utilização responsável e ética. É sabido também que a relação entre pacientes e médicos é multidimensional e, quando saudável, é baseada na transparência sobre os procedimentos utilizados. Com a IA, essa preocupação deve ser redobrada, dada a novidade da tecnologia e a insegurança que ainda gera. Mais ainda, a prática médica responsável deve se afastar ao máximo do determinismo tecnológico que enseja visões acríticas e se concentrar na melhoria da qualidade das interações entre médicos e pacientes, que define o terreno do que é propriamente humano e que pode ser mapeado e codificado pela estatística, mas dificilmente compreendido

e sentido artificialmente. Ou seja, a IA deve ser estruturada para trabalhar com os médicos em prol dos usuários finais, e não para substituí-los.

PROFISSIONAIS QUALIFICADOS E A DIGITALIZAÇÃO DA SAÚDE NO BRASIL

Pelo mundo todo, grandes corporações privadas investem fortemente em aplicações e pesquisa de IA na saúde, especialmente aquelas sediadas nos EUA, China, Alemanha, Japão, Reino Unido e em países em desenvolvimento, como a Índia. Entre os centros de pesquisa, destacam-se as universidades de Stanford e Harvard, assim como o Massachusetts Institute of Technology (MIT), nos EUA; as universidades de Oxford e Cambridge e o Allan Turing Institute, no Reino Unido; e as universidades de Beijing e Tsinghua, na China. No Brasil, a Universidade de São Paulo (USP), a Universidade Estadual de Campinas (Unicamp) e as universidades Federais de Minas Gerais (UFMG), do Rio Grande do Sul (UFRGS) e do Rio de Janeiro (UFRJ) estão entre as que mais se dedicam a esse campo de pesquisa.

O Brasil conta com um corpo de direcionadores definidos pelo Ministério da Ciência, Tecnologia e Inovação (MCTI) em 2021, que compõem a Estratégia Brasileira de IA (EBIA) (MCTI, 2021; Portaria MCTI n. 4.979/2021). Também estão em processo de implantação 11 centros de pesquisa, financiados pelo Governo Federal, por fundações estaduais de amparo à pesquisa (como a Fundação de Amparo à Pesquisa do Estado de São Paulo [Fapesp]), por empresas privadas e pelo Comitê Gestor da Internet no Brasil (CGI.br). E para dotar o país de um marco regulatório próprio para a IA, há projetos de lei em plena discussão no Congresso Nacional e na sociedade brasileira.

A EBIA orienta atualmente os esforços do setor público no apoio à IA e estabeleceu, com base em recomendações da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), três eixos temáticos (transversais) e seis eixos verticais.

Os eixos temáticos são:

1. Legislação, regulação e uso ético: trata de parâmetros jurídicos, regulatórios e éticos para o desenvolvimento da IA.

2. Governança de IA: estrutura de governança que promova métodos e procedimentos para assegurar a observância aos princípios da IA no desenvolvimento de soluções com essa tecnologia.
3. Aspectos internacionais: trata de plataformas de cooperação e integração para trocas de informações, experiências, regulamentações e boas práticas na condução da IA no cenário mundial.

Os eixos verticais, que definem áreas prioritárias para o desenvolvimento da IA, são:

1. Educação;
2. Força de trabalho e capacitação;
3. Pesquisa, desenvolvimento, inovação e empreendedorismo;
4. Aplicações em setores produtivos;
5. Aplicações no Poder Público;
6. Segurança pública.

Com essas diretrizes gerais, a EBIA procura articular o governo com o setor privado, universidades e terceiro setor. Na saúde, essa articulação de amplo espectro é essencial para que as tecnologias de IA possam ajudar a fortalecer, integrar e tornar mais efetivo o SUS, o repositório de grandes bancos de dados, e impulsionar a pesquisa orientada para resultados, o que é central para um país que precisa atuar com extremo senso de urgência para superar seu histórico de desigualdade e déficit social.

Nesse sentido, as palavras-chave que sintetizam as dimensões mais relevantes que precisam ser dinamizadas no SUS são: formação de profissionais, digitalização, integração, eficiência, qualidade e financiamento. Este texto trata das novas tecnologias digitais, em especial a IA. Porém, os autores reconhecem que a tecnologia é direcionada e determinada pela sociedade, não o contrário.

A ampliação do acesso, especialmente para os mais carentes, a coleta e o tratamento dos dados, nem sempre capturados na base da pirâmide social, e o fortalecimento e a integração da rede de atenção básica são fundamentais para melhorar a eficiência do SUS. As tecnologias podem ajudar substantivamente

a aumentar a qualidade dos serviços, realizar um atendimento mais ágil e efetivo, reduzir o uso indiscriminado de medicamentos e diminuir as hospitalizações prematuras. A telessaúde é capaz de operar em todas essas dimensões e ajudar a elevar sua resolutividade. A amplitude das dimensões que podem ser impactadas pela melhoria da gestão significa que os esforços para a digitalização do SUS e a construção de sistemas de telessaúde movidos a IA devem ser abordados de forma integrada a uma estratégia nacional de fortalecimento e reconfiguração do SUS, que precisa incluir a inovação nas áreas de equipamentos e na farmacêutica.

A pandemia COVID-19 expôs várias fragilidades do complexo da saúde nas áreas de medicamentos, vacinas, testes, equipamentos e mesmo produtos básicos. Em outras palavras, a necessidade de o Brasil promover sinergias entre a pesquisa e o complexo industrial é chave para o sistema de saúde.

A urgência dessa estratégia está ligada à rapidez do desenvolvimento do novo ciclo tecnológico que sacode o planeta e que tem a IA como motor. É praticamente consenso entre especialistas que a ausência de políticas públicas voltadas para a inovação na saúde dificulta a evolução da indústria de fármacos, medicamentos e novos equipamentos, que só conseguirá superar seu atraso e suas deficiências com uma atuação articulada entre o setor privado e público: entre empresas e o MS, o Ministério do Desenvolvimento, Indústria, Comércio e Serviços (MDIC), o MCTI e as agências de fomento. Políticas públicas que aproximam a indústria da pesquisa, além da provisão de insumos básicos, contribuem para diminuir o peso das exportações na balança comercial brasileira e liberar recursos para serem alocados no próprio SUS. Isso significa que não há como contornar o uso das novas tecnologias, em especial da IA, na implementação de uma estratégia nacional de saúde, seja nos serviços ou na indústria.

A DIGITALIZAÇÃO COMO ESTRATÉGIA

O SUS caracteriza-se por um conglomerado de instituições públicas, privadas e semipúblicas – governamentais ou não –, que movimentam um universo de competências, nem sempre complementares, e envolvem diferentes estruturas de gestão e de comando, tanto nacionais quanto regionais, estaduais e municipais. A complexidade desse sistema, sua

diversidade, regulação, financiamento, formação e contratação de profissionais e serviços, assistência, avaliação e absorção de tecnologia colocam enormes desafios para sua gestão, para a qualidade de seus serviços e para a agilidade e eficiência do atendimento. Experiências internacionais de integração de sistemas permitem estimar o impacto que a IA pode gerar no controle e gestão do SUS, de modo a: (a) Viabilizar o avanço da informatização das mais de 50 mil equipes de saúde da família existentes em quase 100% dos municípios brasileiros e atuantes em mais de 35 mil Unidades Básicas de Saúde (UBS); (b) Potencializar os recursos digitais para melhor integrar os agentes comunitários de saúde (ACS) e, assim, facilitar a presença e a capacitação remota dos profissionais da saúde; (c) Implantar e multiplicar procedimentos de telediagnóstico; e (d) Ajudar na estruturação de equipes multidisciplinares de profissionais da saúde que atuam remotamente e atendem às demandas das UBS.

Para os serviços, assim como para a indústria, a confiabilidade da informação ganhou enorme relevância. Durante a pandemia COVID-19, houve uma série de experiências localizadas e setoriais que reuniram dados sobre internações e testes. O cotidiano do sistema de saúde, no entanto, é diferente, pois a coleta de informações é realizada de modo fragmentado, via Departamento de Informática do SUS (DATASUS), Agência Nacional de Saúde Suplementar (ANS) e órgãos estaduais e municipais. Apenas um exemplo, em 2021, o Senado brasileiro aprovou um projeto de lei voltado para criar uma plataforma digital capaz de unificar os prontuários dos pacientes das redes públicas e privadas de todo o país (Agência Senado, 2021). O SUS foi encarregado de centralizar essas informações, com dados sobre prescrições, encaminhamentos, prontuários médicos e laudos de exames. Mais ainda, a lei obrigou o SUS a adequar sua plataforma às exigências da Lei Geral de Proteção de Dados (LGPD) (Lei n. 13.709/2018), tendo o MS já anunciado uma versão do sistema Conecte SUS que contempla o prontuário eletrônico. Em 2021, a atualização da Política Nacional de Informação e Informática em Saúde (PNIIS) (Resolução n. 659/2021) e a Estratégia de Saúde Digital 2020-2028 (ESD) (MS, 2020) reafirmaram essa direção ao definirem a Rede Nacional de Dados em Saúde (RNDS) como ambiente de interoperabilidade e comunicação entre os atores do SUS.

Apesar da legislação favorável à integração e ao compartilhamento de dados, os obstáculos para sua efetivação continuam enormes e incluem a disponibilidade de infraestrutura para que essa rede se torne realidade, a disposição dos atores do SUS para compartilharem dados, e o aprofundamento das bases normativas e legais para o sistema, inclusive nos requisitos éticos de utilização das informações dos pacientes e custos. Do ponto de vista tecnológico, a interoperabilidade não pode prescindir da IA como ferramenta para realizar essa integração.

Passos nessa direção são mais do que necessários para aperfeiçoar a gestão do sistema de saúde, a começar pela preparação de bancos de dados anonimizados e que permitam análises agregadas capazes de gerar conhecimento sobre a saúde populacional. Esses bancos devem possibilitar análises de indicadores de saúde, bem como dos desfechos dos procedimentos realizados no âmbito de uma dada região do SUS. Essas informações são instrumentos essenciais para a prevenção e conscientização, a otimização do investimento em treinamentos e para uma alocação mais adequada de infraestrutura. Com esses dados, o SUS teria condições de acompanhar trajetórias regionais, realizar comparações, projeções e fortalecer de modo inédito todas as atividades de planejamento, com maior racionalização e rigor na utilização de recursos e no controle dos objetivos das políticas de saúde.

RECOMENDAÇÕES PARA A CONSTRUÇÃO DE UMA ESTRATÉGIA DA IA NA SAÚDE

A seguir, propõem-se algumas sugestões para estruturar políticas públicas de saúde com a IA como protagonista, facilitadora ou viabilizadora de inovações:

1. Priorizar as tecnologias de IA na telessaúde para ampliar o Acesso Sustentável, a Atenção Primária e sua integração com os procedimentos médicos e hospitalares de média e alta complexidade.
2. Introduzir sistemas avançados de gestão no SUS, com integração de dados e coleta de informações em tempo real.
3. Estimular modelos de governança para integrar os bancos de dados, favorecendo o avanço no compartilhamento dos prontuários eletrônicos.
4. Formar, capacitar e requalificar profissionais da saúde, inclusive remotamente, para que consigam acompanhar

as mudanças e aproveitar as oportunidades das tecnologias com a finalidade de elevar a qualidade do SUS e de seu trabalho.

5. Estabelecer salvaguardas de proteção da população via regramento ético e legislação específica para a IA.
6. Incentivar a pesquisa sobre aplicações de IA na saúde, seja para aperfeiçoar e ampliar o atendimento à população, seja para melhorar a gestão e a qualidade das políticas públicas.
7. Apoiar a pesquisa avançada em IA na saúde e a articulação universidade-empresa para executar projetos mensuráveis, monitorados, eticamente transparentes e capazes de oferecer resultados palpáveis para a população.
8. Impulsionar a inovação na saúde: o uso do poder de compra do Estado pode alavancar a transferência tecnológica e financiar grandes projetos com base na mobilização de competências em distintas áreas de pesquisa, seja nas universidades, nas empresas e na articulação de centros de pesquisa, a começar pelos institutos de IA.
9. Considerar a construção de um modelo de observatório da IA na saúde dedicado a captar tendências internacionais, acompanhar a evolução de tecnologias inteligentes no Brasil e desenvolver tecnologias de avaliação de desempenho de projetos e iniciativas públicas.
10. Dotar o SUS e as políticas de saúde de um regramento ético baseado na transparência e na confiabilidade dos sistemas tecnológicos, de modo a resguardar a população e a manter a dignidade das relações entre médico e paciente.

REFERÊNCIAS

Agência Senado. (2021). Aprovada criação de plataforma para unificar dados do SUS e da rede privada. *Senado Notícias*. https://www12.senado.leg.br/noticias/materias/2021/05/18/aprovada-criacao-de-plataforma-para-unificar-dados-do-sus-e-da-rede-privada

Chen, K., Zhang, D., Yao, L., Guo, B., Yu, Z., & Liu, Y. (2021). Deep learning for sensor-based human activity recognition. *ACM Computing Surveys, 54*(4), 1-40 https://dl.acm.org/doi/10.1145/3447744

Dean, N. C., Vines, C. G., Carr, J. R., Rubin, J. G., Webb, B. J., Jacobs, J. R., Butler, A. M., Lee, J., Jephson, A. R., Jenson, N., Walker, M., Brown, S. M., Irvin, J. A., Lungren, M. P., & Allen, T. L. (2022). A pragmatic, stepped wedge, cluster-controlled clinical trial of real-time pneumonia clinical decision support. *American Journal of Respiratory and Critical Care Medicine, 205*(11), 1330-1336. https://doi.org/10.1164/rccm.202109-2092OC

Federação Nacional de Saúde Suplementar. (2023). *Relatório anual de atividades*. https://fenasaude.org.br/publicacoes/relatorio-anual-de-atividades-fena-saude

Folha de São Paulo. (2024). *Atendimentos por telemedicina no país crescem 172% em 2023 após lei que regulamenta saúde digital*. https://www1.folha.uol.com.br/equilibrioesaude/2024/04/atendimentos-por-telemedicina-no-pais-crescem-172-em-2023-apos-lei-que-regulamenta-saude-digital.shtml

Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., Spitzer, A. I., & Ramkumar, P. N. (2020). Machine learning and Artificial Intelligence: Definitions, applications, and future directions. *Current Reviews in Musculoskeletal Medicine, 13*(1), 69-76. https://doi.org/10.1007/s12178-020-09600-8

Guo, J., Huang, X., Dou, L., Yan, M., Shen, T., Tang, W., & Li, J. (2022). Aging and aging-related diseases: From molecular mechanisms to interventions and treatments. *Signal Transduction and Targeted Therapy, 7*(391). https://doi.org/10.1038/s41392-022-01251-0

Joshi, M., Archer, S., Morbi, A., Arora, S., Kwasnicki, R., Ashrafian, H., Khan, S., Cooke, G., & Darzi, A. (2021). Short-term wearable sensors for in-hospital medical and surgical patients: Mixed methods analysis of patient perspectives. *JMIR Perioperative Medicine, 4*(1), e18836. https://doi.org/10.2196/18836

Kalfa, D., Agrawal, S., Goldshtrom, N., LaPar, D., & Bacha, E. (2020). Wireless monitoring and artificial intelligence: A bright future in cardiothoracic surgery. *The Journal of Thoracic and Cardiovascular Surgery, 160*(3), 809-812. https://doi.org/10.1016/j.jtcvs.2019.08.141

Krittanawong, C., Johnson, K. W., Choi, E., Kaplin, S., Venner, E., Murugan, M., Wang, Z., Glicksberg, B. S., Amos, C. I., Schatz, M. C., & Tang, W. W. (2022). Artificial intelligence and cardiovascular genetics. *Life, 12*(2), 279-307. https://doi.org/10.3390/life12020279

Lauritsen, S. M., Kristensen, M., Olsen, M. V., Larsen, M. S., Lauritsen, K. M., Jørgensen, M. J., Lange, J., & Thiesson, B. (2020). Explainable artificial intelligence model to predict acute critical illness from electronic health records. *Nature Communications, 11*(3852), 1-11. https://doi.org/10.1038/s41467-020-17431-x

Lei n. 14.510, de 27 de dezembro de 2022. (2022). Altera a Lei n. 8.080, de 19 de setembro de 1990, para autorizar e disciplinar a prática da telessaúde em todo o território nacional, e a Lei n. 13.146, de 6 de julho de 2015; e revoga a Lei n. 13.989, de 15 de abril de 2020. https://www.in.gov.br/web/dou/-/lei-n-14.510-de27-de-dezembro-de-2022-454029572

Lei n. 13.989, de 15 de abril de 2020. (2020). Dispõe sobre o uso da telemedicina durante a crise causada pelo novo coronavírus (SARS-CoV-2); revogado pela Lei nº 14.510, de 2022. http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/Lei/L13989.htm

Lei Geral de Proteção de Dados Pessoais (LGPD). (2018). Lei n. 13.709, de 14 de agosto de 2018. Dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural. https://www.planalto. gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm

Liu, H., Wang, L., Lin, G., & Feng, Y. (2022). Recent progress in the fabrication of flexible materials for wearable sensors. *Biomaterials Science, 10*(3), 614-632. https://doi.org/10.1039/d1bm01136g

Ministério da Ciência, Tecnologia e Inovação. (2021). *Estratégia Brasileira de Inteligência Artificial (EBIA)*. https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ebia-diagramacao_4-979_2021.pdf

Ministério da Saúde. (2020). *Estratégia de Saúde Digital para o Brasil 2020-2028*. https://bvsms.saude.gov.br/bvs/publicacoes/estrategia_saude_digital_Brasil.pdf

Mohanty, A., & Mishra, S. (2022). A comprehensive study of explainable Artificial Intelligence in healthcare. In S. Mishra, H. K. Tripathy, P. Mallick, & K. Shaalan (Eds.), *Augmented intelligence in healthcare: A pragmatic and integrated analysis* (pp. 475–502). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-1076-025

Mukherjee, A., Ghosh, S., Behere, A., Ghosh, S. K., & Buyya, R. (2020). Internet of Health Things (IoHT) for personalized health care using integrated edge-fog-cloud network. *Journal of Ambient Intelligence and Humanized Computing, 12*(1), 943-959. https://doi.org/10.1007/s12652-020-02113-9

Organização Mundial da Saúde. (2021). *Ethics and governance of Artificial Intelligence for health: WHO Guidance*. https://iris.who.int/handle/10665/341996

Organização Mundial da Saúde. (2024). *Exploring the digital health landscape in the WHO European Region: Digital health country profiles*. https://iris.who.int/handle/10665/376540

Portaria MCTI n. 4.979, de 13 de julho de 2021. (2021). Altera a Portaria n. 4.617, de 6 de abril de 2021, que institui a Estratégia Brasileira de Inteligência Artificial e seus eixos temáticos. *Diário Oficial da* União. https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ebia-portaria_mcti_4-979_2021_anexo1.pdf

Resolução n. 659, de 26 de julho de 2021. (2021). Dispõe sobre a Política Nacional de Informação e Informática em Saúde (PNIIS). https://bvsms.saude.gov.br/bvs/saudelegis/cns/2022/res0659_15_06_2022.html

Sindhosp. (2022). *Proposta Saúde São Paulo 2022*: rumo ao acesso sustentável. https://sindhosp.org.br/linha-do-tempo/#:~:text=O%20que%20%C3%A9%3F,com%20mais%20equidade%20e%20sustentabilidade

Shaik, T., Tao, X., Higgins, N., Gururajan, R., Li, Y., Zhou, X., & Acharya, U. R. (2022). Fedstack: Personalized activity monitoring using stacked federated learning. *Knowledge-Based Systems, 257(*12), 109929. https://doi.org/10.1016/j.knosys.2022.109929

Shaik, T., Tao, X., Higgins, N., Li, L., Gururajan, R., Zhou, X., & Acharya, U. R. (2023). Remote patient monitoring using Artificial Intelligence: Current state, applications, and challenges. *WIREs Data Mining and Knowledge Discovery, 13*(2), e1485. https://doi.org/10.1002/widm.1485

Tandel, S., Godbole, P., Malgaonkar, M., Gaikwad, R., & Padaya, R. (2022). An improved health monitoring system using iot. *Proceedings of the International Conferenc on Innovations and Research in Technology and Engineering, 7*, 1-5. https://doi.org/10.2139/ssrn.4109039

Weenk, M., Bredie, S. J., Koeneman, M., Hesselink, G., van Goor, H., & van de Belt, T. H. (2020). Continuous monitoring of vital signs in the general ward using wearable devices: Randomized controlled trial. *Journal of Medical Internet Research, 22*(6), 1-11. https://www.jmir.org/2020/6/e1547110.2196/15471

Wu, Q., Chen, X., Zhou, Z., & Zhang, J. (2022). FedHome: Cloud-edge based personalized federated learning for in-home health monitoring. *IEEE Transactions on Mobile Computing, 21*(8), 2818-2832. https://doi.org/10.1109/tmc.2020.3045266

Yang, G., Ye, Q., & Xia, J. (2022). Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. *Information Fusion, 77*, 29-52. https://doi.org/10.1016/j.inffus.2021.07.016

Zheng, X., Shah, S. B. H., Ren, X., Li, F., Nawaf, L., Chakraborty, C., & Fayaz, M. (2021). Mobile edge computing enabled efficient communication based on federated learning in internet of medical things. *Wireless Communications and Mobile Computing*, 1-10. https://doi.org/10.1155/2021/4410894


UNESCO
unesco
Centro
sob os auspícios
da UNESCO


cetic.br
Centro Regional
de Estudos para o
Desenvolvimento
da Sociedade
da Informação


nic.br
Núcleo de Informação
e Coordenação do
Ponto BR


cgi.br
Comitê Gestor da
Internet no Brasil


OBIA