

Capítulo

3

Representatividade, Soberania e Equidade Algorítmica como Desafios da Computação Aplicada à Saúde das Populações do Brasil Profundo

Ana C. S. Alcântara, Anselmo C. Paiva, Aristófanés C. Silva, Cecília C. C. Ribeiro, Elaine P. F. Costa, Erika B. A. F. Thomaz, Flávio S. Damos, Geraldo Braz Júnior, Harvey A. V. Vélez, João D. S. Almeida, João O. B. Diniz, Luciana S. B. Almeida, Rita C. S. Luz, Soraia de F. C. Souza, Tiago Bonini Borchart.

Abstract

Artificial intelligence (AI) has established itself as a transformative technology in healthcare; however, its benefits remain unevenly distributed due to biases in training data. In Brazil, this problem is particularly acute in “Deep Brazil,” which encompasses underrepresented indigenous, quilombola, and rural populations. This paper discusses challenges related to representativeness, data sovereignty, and algorithmic equity in the development of AI-based healthcare solutions. It analyzes the impacts of data scarcity on model performance, highlighting the role of multimodal data and electrochemical sensors in generating clinical data in resource-limited settings. Finally, guidelines are proposed to develop robust, decentralized AI systems tailored to these populations.

Resumo

A inteligência artificial (IA) tem se consolidado como uma tecnologia transformadora na saúde, porém, seus benefícios permanecem desigualmente distribuídos devido a vieses nos dados de treinamento. No Brasil, esse problema é crítico no “Brasil Profundo”, que abrange populações indígenas, quilombolas e rurais sub-representadas. Discutimos os desafios de representatividade, de soberania sobre os dados e de equidade algorítmica no desenvolvimento de soluções em saúde baseadas em IA. São analisados os impactos da escassez de dados no desempenho dos modelos, destacando o papel de dados multimodais e de sensores eletroquímicos na geração de dados clínicos em ambientes

com recursos limitados. Por fim, apontam-se diretrizes para o desenvolvimento de sistemas de IA robustos, descentralizados e adequados a essas populações.

3.1 Contextualização e relevance

A inteligência artificial (IA) apresenta-se como um importante vetor da transformação digital na saúde global. Modelos generativos e de aprendizado profundo apresentam precisão diagnóstica comparável à de especialistas humanos em diversas patologias. A Organização Mundial da Saúde (OMS), reforça o potencial transformador dessas tecnologias, alertando para os riscos de informações enviesadas decorrentes de dados de treinamento de baixa qualidade ou de representatividade insuficiente ([Topol, 2024] e [WHO, 2024]). O risco identificado pela OMS não é hipotético. O viés algorítmico é uma ameaça silenciosa e sistemática à equidade em saúde em contextos de recursos limitados, com impacto especialmente severo sobre populações não representadas nos conjuntos de dados globais [Joseph, 2025].

No Brasil, esse cenário encontra seu caso mais crítico no denominado Brasil Profundo. O conceito de Brasil Profundo empregado neste artigo fundamenta-se em [Ribeiro, 1995], que identifica cinco configurações culturais, os "Brasis". Essas configurações foram forjadas pela confluência de matrizes indígenas, africanas e europeias em contextos de exploração e de isolamento. Três dessas configurações estão associadas aqui ao conceito de Brasil profundo: o Brasil Caboclo amazônico, ligado à floresta e aos saberes dos povos originários; o Brasil Sertanejo, que vai do semiárido nordestino ao Cerrado do Centro-Oeste e concentra as maiores densidades de comunidades quilombolas e de povos originários do país; e o Brasil Crioulo, surgido nos engenhos da costa nordestina e marcado pela matriz negro-africana. Esses três Brasis compartilham uma herança estrutural de exclusão. São territórios de onde a riqueza foi extraída sem que as populações que a produziram dela se beneficiassem. Essa herança histórica é reconhecida aqui como a raiz do apagão de dados que hoje afeta, de modo mais severo, as comunidades não-urbanas, especialmente povos originários e quilombolas das regiões Norte, Nordeste e Centro-Oeste. É fato a exclusão estrutural das populações do Brasil profundo dos grandes conjuntos de dados que alimentam o desenvolvimento da IA em saúde.

Segundo [IBGE, 2023] e [IBGE, 2024], o Brasil abriga 1,6 milhão de pessoas indígenas (0,83% da população total) e 1,3 milhão de quilombolas (0,66%), presentes em 24 estados e no Distrito Federal. Somadas, essas populações superam 3 milhões de pessoas, um contingente maior do que o de diversas capitais brasileiras, concentradas majoritariamente nas regiões Norte, Nordeste e Centro-Oeste, as mesmas que apresentam os piores indicadores de saúde e os maiores déficits de infraestrutura digital do país.

Essa ausência de representatividade dos dados tem consequências concretas. Em [Wu, 2024], após revisão de 91 conjuntos de dados clínicos de texto, foi verificado que 73% dos dados provinham das Américas e da Europa, regiões que representam apenas 22% da população global, e que mais da metade dos conjuntos de dados era em inglês, além de refletirem predominantemente sistemas de saúde altamente estruturados, com protocolos diagnósticos e padrões de registro que não se reproduzem em contextos de atenção primária em regiões de baixa cobertura assistencial. Quando modelos treinados nesse contexto são aplicados a populações sub-representadas, ocorre o fenômeno conhecido como 'covariate shift', ou seja, uma mudança na distribuição dos dados de entrada degrada sistematicamente o desempenho dos algoritmos. Isto acontece porque eles passam a

interpretar dados diferentes dos usados durante o treinamento. Por exemplo, biomarcadores laboratoriais (valores biológicos obtidos em exames de laboratório) variam entre etnias, padrões de imagem (características observadas em exames de imagem médica) diferem por fatores genéticos e ambientais, e intervalos de referência (faixas consideradas normais para exames) são etnicamente específicos, variáveis que os modelos globais simplesmente não incorporam. Modelos de diagnóstico por imagem apresentam viés de subdiagnóstico precisamente nos grupos populacionais mais vulneráveis [Seeyed-Kalamtari, 2021].

As desigualdades em saúde que motivam essa urgência no Brasil Profundo são documentadas e persistentes. A expectativa de vida nas regiões Norte e Nordeste é 3 anos menor que a média nacional e a taxa de mortalidade infantil é 3% superior à do Centro-Sul [Scheffer, 2023]. As regiões Norte e Nordeste mantêm níveis elevados de mortalidade materna, com risco três vezes maior no Nordeste do que no Centro-Oeste para causas diretas [Figueiredo, 2025]. As populações indígenas apresentam mortalidade infantil de até 97 óbitos por mil nascidos vivos [Coimbra, 2012]. Menos de 7% dos territórios quilombolas no Brasil estão titulados, e apenas 15% dos domicílios quilombolas têm acesso à rede pública de água [Zhou, 2024]. Essas condições tornam essas comunidades simultaneamente as mais vulneráveis e as menos representadas nos sistemas de informação em saúde. Esses números exigem modelos preditivos treinados com dados locais, não extrapolações de coortes do Sul e do Sudeste, nem de outros países.

A soberania de dados emerge, nesse contexto, não como questão técnica, mas como imperativo ético e estratégico. A dependência de plataformas de nuvem de grandes empresas globais, hospedadas no exterior, para processar dados sensíveis dessas populações, acrescenta um risco geopolítico. Dados coletados nessas comunidades podem retroalimentar modelos proprietários globais, sem qualquer benefício para as próprias comunidades. O resultado é uma forma contemporânea de extrativismo de dados.

Este trabalho argumenta que enfrentar a invisibilidade das características dessas populações do Brasil Profundo é mais do que um nicho de pesquisa: uma agenda de saúde pública e de soberania nacional. A janela para agir é estreita. Cada novo ciclo de desenvolvimento de IA, sem dados dessas populações, aprofunda o viés estrutural e torna a correção futura mais custosa. A Computação Aplicada à Saúde dispõe hoje de ferramentas tecnológicas para reverter esse quadro. O que ainda falta é o reconhecimento coletivo de que construir conjuntos de dados e modelos soberanos, desenvolver IA equitativa e garantir o acesso a funcionalidades computacionais sem internet no campo são, acima de tudo, atos de justiça social.

3.2 Descrição do Desafio

O desafio aqui proposto é desenvolver, validar e manter modelos de IA em saúde que sejam equitativos, precisos e culturalmente adequados às populações do Brasil Profundo, compreendendo os povos indígenas, quilombolas, ribeirinhos, sertanejos e as comunidades tradicionais, superando a ausência estrutural de dados representativos dessas populações nos fluxos globais de desenvolvimento de IA em saúde.

Esse desafio se desdobra em cinco dimensões interligadas que explicam tanto a profundidade do problema quanto a necessidade de enfrentá-lo de forma coordenada e urgente.

Dimensão 1: Viés algorítmico estrutural. O fenômeno de covariate shift produz algoritmos que erram de forma sistemática exatamente nas populações que mais dependem de apoio diagnóstico assistido por tecnologia. As variações étnicas em biomarcadores, padrões de imagem e intervalos de referência laboratorial, ignoradas pelos modelos globais, comprometem a validade clínica de qualquer ferramenta desenvolvida sem dados locais.

Dimensão 2: Soberania e segurança de dados. Os dados de saúde das populações do Brasil Profundo são simultaneamente valiosos do ponto de vista científico e vulneráveis do ponto de vista político e ético. A dependência de plataformas de nuvem comercial hospedadas no estrangeiro para o processamento dessas informações compromete a soberania nacional e expõe comunidades historicamente marginalizadas a novos vetores de exploração. A Lei Geral de Proteção de Dados e as Resoluções do Conselho Nacional de Saúde estabelecem requisitos que a infraestrutura tecnológica atual raramente consegue atender de forma satisfatória.

Dimensão 3: Ausência de representatividade multimodal. Não existem, no Brasil, repositórios abertos de retinografias, imagens odontológicas, dados eletroquímicos de biossensores, sequências metagenômicas do microbioma oral ou cortes representativas das populações do Brasil Profundo, especialmente integrando múltiplas modalidades de aquisição em condições reais de campo, onde variáveis como interferentes químicos, variações de matriz biológica e limitações instrumentais impactam diretamente a qualidade do estado coletado. Essa lacuna põe em dúvida o desenvolvimento de qualquer modelo de IA com validade clínica para essas populações, independentemente do poder computacional disponível para o treinamento.

Adicionalmente, destaca-se que sensores e biossensores eletroquímicos representam uma tecnologia promissora para a geração de dados clínicos *in situ* em regiões com infraestrutura limitada, permitindo a detecção rápida e de baixo custo de biomarcadores relevantes (e.g. dopamina, glicose, ácido úrico e compostos relacionados a doenças infecciosas e crônicas). No entanto, a ausência de datasets eletroquímicos representativos dessas populações limita tanto o desenvolvimento quanto a validação de modelos de IA embarcados para diagnóstico descentralizado.

Dimensão 4: Limitações de conectividade e de hardware. As soluções de IA em saúde costumam pressupor uma infraestrutura de conectividade que simplesmente não existe em vastas regiões do Brasil Profundo. Comunidades indígenas, quilombolas e sertanejas não têm, em grande parte, acesso estável à internet, tornando inaplicáveis, em campo, os modelos que requerem inferência em nuvem. A computação de borda e a operação offline tornam-se, nesse contexto, não alternativas desejáveis, mas requisitos técnicos incontornáveis [Zhou 2024].

Dimensão 5: Barreiras linguísticas e culturais. As ferramentas de apoio à decisão clínica baseadas em processamento de linguagem natural não incorporam terminologias, cosmovisões e especificidades epidemiológicas das etnias indígenas, das comunidades quilombolas ou das populações sertanejas e tradicionais do Brasil Profundo. O ajuste fino de grandes modelos de linguagem para contextos de saúde do Brasil Profundo, incluindo línguas como o Guajajara/Tenetehara, dentre outras, é praticamente inexistente no cenário nacional, criando obstáculos culturais concretos à adoção dessas tecnologias.

O campo da IA em saúde atravessa um processo acelerado de consolidação de conjuntos de dados de referência, benchmarks e modelos fundamentais. Se o Brasil Profundo não

produzir seus próprios dados de treinamento nos próximos anos, os modelos de IA que chegarão a essas populações serão inadequados às suas realidades, perpetuando uma forma contemporânea de exclusão tecnológica, com consequências diretas para a saúde.

Importa reconhecer, contudo, que o Brasil dispõe hoje dos instrumentos necessários para enfrentar este desafio. Do ponto de vista legal, a Lei Geral de Proteção de Dados [BRASIL, 2018] e as Resoluções do Conselho Nacional de Saúde (CNS) constituem o arcabouço regulatório. Do ponto de vista institucional, as universidades e os institutos federais presentes nessas regiões, somados ao fomento governamental, constituem uma base científica capaz de sustentar iniciativas de longo prazo. O que ainda falta é o reconhecimento, pela comunidade de pesquisa e pelos formuladores de políticas públicas, de que a construção de conjuntos de dados soberanos constitui uma intervenção direta nas condições de equidade em saúde das populações do Brasil Profundo.

3.3 Estado da arte

A literatura tem documentado, de forma crescente e sistemática, o problema da lacuna de representatividade que se estende a um grupo diverso de populações, cujas realidades de saúde permanecem invisíveis ou distorcidas nos grandes repositórios que servem de base à IA global.

Trabalhos como [Obermeyer,2019] e [Seyyed-Kalantari,2021] apresentaram evidências de que algoritmos de IA em saúde reproduzem e amplificam desigualdades estruturais preexistentes, seja subestimando sistematicamente a gravidade clínica de pacientes negros ao utilizar custos históricos de saúde como proxy de necessidade, seja produzindo diagnósticos significativamente menos precisos para grupos sub-representados nos dados de treinamento, demonstrando que o viés algorítmico não é uma falha técnica isolada, mas o reflexo computacional de iniquidades sociais mais profundas.

Em [Joseph,2025] é proposta uma taxonomia de cinco tipos de vieses na IA em saúde pública: (1) viés histórico (injustiças anteriores ficam incorporadas nos conjuntos de dados de treinamento); (2) viés de representação (amostras de grupos urbanos ou conectados excluem populações sub-representadas); (3) viés de mensuração (desfechos de saúde são aproximados por variáveis *proxy* diferentes entre ambientes socioeconômicos); (4) viés de agregação (modelos assumem homogeneidade entre grupos heterogêneos); e (5) viés de implantação (ferramentas desenvolvidas em um contexto falham ao ser aplicadas em contextos diferentes).

Para mitigar o viés algorítmico, é necessária a utilização de dados abertos, de partes interessadas e de intervenções na fase de pré-processamento [Sasseville, 2025], condições que exigem a participação ativa das próprias comunidades no *design* e na validação das ferramentas. Em algumas iniciativas, as comunidades são posicionadas como parceiras iguais no processo de pesquisa, desde o *design* de algoritmos até sua implantação clínica, um exemplo é a *Community-Based Participatory Research* (CBPR) [Fields, 2025].

Historicamente, mulheres foram tratadas como versões menores dos homens, com dosagens medicamentosas ajustadas apenas pelo tamanho corporal, sem considerar diferenças biológicas de sexo, que podem ser substanciais na expressão gênica, na prevalência, na idade de início, na sintomatologia e na mortalidade de doenças como doença coronária, acidente vascular encefálico (AVE) e diferentes tipos de câncer. Indivíduos LGBTQ+ são particularmente afetados por desigualdades no acesso à saúde

decorrentes tanto de necessidades específicas de tratamento quanto de vieses nos sistemas de saúde [Norori, 2021].

Estratégias de coleta de dados precisam ser redesenhadas para abranger áreas rurais, línguas sub-representadas e grupos marginalizados. O uso de auditorias de equidade, de dados sintéticos para casos sub-representados e de modelos de linguagem multilíngue pode ajudar a mitigar os pontos cegos sistêmicos [Joseph, 2025].

Algumas iniciativas buscam garantir que populações historicamente excluídas não apenas sejam representadas nos dados, mas também exerçam controle sobre a coleta, o uso e a governança desses dados. Entre elas, destacamos os Princípios CARE para Governança de Dados Indígenas (*Collective Benefit, Authority to Control, Responsibility, Ethics*), que são orientados a pessoas e propósito, reconhecendo o papel crucial dos dados no avanço da inovação e da autodeterminação indígenas [Carroll, 2020]. Os princípios Māori do *Te Mana Raraunga* na Nova Zelândia [Whittaker, 2023] e os princípios OCAP® do *First Nations Information Governance Centre* no Canadá. Outra iniciativa relevante é o relatório da UNESCO que demanda a inclusão participativa de comunidades indígenas em todas as fases do desenvolvimento de IA na América Latina e no Caribe [Zepeda, 2023].

O STANDING Together (NHS AI Lab) publicou 29 recomendações para a transparência e a diversidade em conjuntos de dados de IA em saúde [Alderman, 2025]. Estas recomendações estruturam-se em dois pilares fundamentais para promover a equidade na saúde digital: 18 diretrizes voltadas à documentação de conjuntos de dados de saúde, destinadas a curadores de dados, visando garantir a transparência sobre a composição e as limitações desses dados; e 11 diretrizes focadas no uso desses dados, de modo a orientar os usuários a identificar e mitigar vieses algorítmicos.

Os modelos *Masakhane e Lacuna Fund* [Ariyo, 2025] representam um movimento coordenado de soberania de dados e de representação cultural, construindo modelos, *datasets* e políticas de IA orientados pelo princípio de que a tecnologia para africanos deve ser desenvolvida por africanos.

Recentemente, a Assembleia Geral adotou uma resolução exigindo o respeito aos direitos humanos em todo o ciclo de vida dos sistemas, enquanto o Fórum Permanente sobre Questões Indígenas alertou que, sem salvaguardas adequadas, a IA pode reforçar vieses, exclusão e apropriação cultural sem consentimento.

No Brasil, a situação é ainda mais grave do que a média global. As grandes coortes epidemiológicas brasileiras têm cobertura concentrada nas regiões Sul e Sudeste, e os sistemas de informação em saúde do SUS, como os Sistemas de Informação de Mortalidade (SIM), Nascidos Vivos (SINASC) e Agravos de Notificação (SINAN), apresentam subnotificação estrutural nas regiões Norte e Nordeste. Praticamente inexistem estudos sistemáticos sobre o desempenho de modelos de IA em populações indígenas ou quilombolas brasileiras, ou mesmo em populações rurais, e o país não dispõe de um equivalente aos Princípios CARE, OCAP® ou aos *frameworks STANDING Together*, adaptado à sua realidade jurídica e epidemiológica.

A revisão da literatura apresentada acima permite mapear lacunas críticas em três dimensões interdependentes.

No plano dos dados e da representação, o Brasil carece de *benchmarks* públicos de IA em saúde calibrados para populações indígenas, quilombolas, ribeirinhas e rurais, problema

agravado pela subnotificação estrutural do SUS no Norte, Nordeste e Centro-Oeste, que gera uma invisibilidade estatística que precede qualquer tentativa de treinar modelos representativos, à qual se somam a ausência de métricas de *fairness* validadas para grupos étnicos brasileiros e de protocolos padronizados de coleta de dados adaptados a essas realidades.

No plano tecnológico e de infraestrutura, o Brasil Profundo enfrenta a ausência de modelos de IA explicável treinados com dados regionais, a escassez de infraestrutura computacional local e a inexistência de soluções de *edge computing* e de operação *offline* validadas para contextos de saúde comunitária em regiões remotas.

Nesse contexto, sensores eletroquímicos permitem a aquisição de dados clínicos diretamente no território. A integração com modelos de IA pode permitir a triagem, o monitoramento e o diagnóstico precoce de doenças em comunidades remotas, contribuindo diretamente para a redução das desigualdades no acesso ao diagnóstico. A literatura internacional tem demonstrado o potencial desses dispositivos como ferramentas portáteis, sensíveis e de baixo custo para diagnóstico rápido em campo. Entretanto, esses dispositivos são frequentemente calibrados e validados com amostras de populações urbanas ou de países desenvolvidos, o que pode introduzir vieses na resposta eletroquímica quando aplicados a populações com perfis metabólicos, ambientais e nutricionais distintos, como as do Brasil Profundo. Além disso, a eficácia dos modelos de IA integrados a esses sensores depende criticamente da qualidade e da representatividade dos dados eletroquímicos utilizados no treinamento, reforçando a necessidade de construir bases de dados locais que capturem as variabilidades biológicas e ambientais dessas populações.

No plano da governança, da ética e da política, o país carece de um *framework* de soberania de dados de populações vulneráveis adaptado à sua realidade jurídica. Apesar de a LGPD e as Resoluções do CNS oferecerem bases normativas que poderiam sustentar um modelo robusto, elas permanecem desarticuladas especificamente no que diz respeito à IA em saúde. A lacuna de participação comunitária, indígena, quilombola, LGBTQ+, migrante e rural no design, na validação e na governança das ferramentas de IA é talvez a mais estrutural de todas. Sem a presença efetiva dessas comunidades como sujeitos, e não apenas como objetos do processo de desenvolvimento tecnológico, arrisca-se reproduzir, com recursos computacionais sofisticados, as mesmas exclusões que a IA prometia superar.

3.4 Métricas de Avaliação

O avanço em relação ao desafio aqui proposto só poderá ser aferido com rigor se forem estabelecidas métricas operacionalizáveis que traduzam objetivos amplos em indicadores mensuráveis. As métricas propostas organizam-se em quatro dimensões avaliativas: dados e representação; infraestrutura e tecnologia; adequação ao contexto do Brasil Profundo; e governança e produção de conhecimento.

A representatividade dos *datasets* pode ser medida pela proporção de indivíduos do Brasil Profundo em conjuntos de dados nacionais de IA em saúde e pelo gap de desempenho intergrupos, definido como a diferença de assertividade entre o subgrupo com melhor desempenho e as demais populações sub-representadas em modelos diagnósticos prioritários.

Também é possível avaliar a adoção de IA explicável em contexto clínico pelo percentual de modelos publicados que incorporam um módulo descritivo auditável, o que é complementado pela capacidade computacional soberana instalada em instituições do Norte, Nordeste e Centro-Oeste, em comparação à capacidade nacional.

Na dimensão de governança e produção de conhecimento, a métrica avalia o número de protocolos aprovados pela Plataforma Brasil que incluam cláusula explícita de soberania comunitária sobre os dados gerados, com meta de longo prazo de tornar esse requisito obrigatório em todos os editais federais de pesquisa em saúde.

Tomadas em conjunto, essas métricas formam um painel integrado de monitoramento que indica o caminho a seguir rumo a uma IA em saúde verdadeiramente representativa e soberana.

3.5 Conexão com Desafios e Problemas de Saúde

O desafio proposto articula-se com múltiplos eixos tecnológicos e com os principais problemas de saúde que o Brasil enfrentará na próxima década.

No plano tecnológico, a invisibilidade das populações do Brasil Profundo nos repositórios globais é, em sua essência, um problema de IA e de ciência de dados: construir modelos equitativos sem dados representativos exige avanços em *fairness*, adaptação de domínio e ajuste fino com dados escassos. A isso se somam desafios de soberania tecnológica, dados de populações vulneráveis processados em nuvens estrangeiras, sem governança adequada, configuram vulnerabilidade estratégica e de computação em contextos adversos, já que, em vastas porções da Amazônia e do semiárido, a operação *offline* não é opcional, mas tecnicamente incontornável. Há ainda conexão com os aspectos éticos da computação aplicada à saúde, particularmente com o colonialismo algorítmico e a responsabilidade institucional sobre vieses que afetam sistematicamente populações vulneráveis, e com o risco de grandes modelos de linguagem, sem ajuste cultural, tornarem-se vetores de desinformação clínica em contextos de baixa densidade assistencial.

No plano sanitário, seis grandes problemas estruturam a urgência do desafio. A eliminação das Doenças Tropicais Negligenciadas, cuja meta da OMS para 2030 permanecerá inacessível sem ferramentas de diagnóstico capazes de operar em campo com dados representativos. A redução da mortalidade materna e neonatal, ainda concentrada nas populações com menor acesso a cuidados, onde modelos preditivos treinados sem dados do Brasil Profundo falharão sistematicamente, inclusive por ignorar a relação entre saúde bucal materna e desfechos gestacionais adversos: alterações no microbioma oral durante a gestação estão associadas a prematuridade, baixo peso ao nascer e sepse neonatal, e os dados que permitiriam treinar modelos preditivos para essa associação em populações do Brasil Profundo simplesmente não existem. O controle das doenças crônicas não transmissíveis em populações indígenas, cujos padrões de risco diferem dos grupos em que os modelos foram treinados, resulta em subestimação sistemática, com relevância também para as complicações oftalmológicas do diabetes não diagnosticado, como a retinopatia diabética e o edema macular, que levam à cegueira irreversível em contextos em que não há oftalmologistas. Também há o desafio da saúde bucal como problema de saúde pública. No Brasil Profundo, populações de baixa renda e comunidades tradicionais apresentam alta prevalência de cárie e periodontite, que, submetidas a métodos convencionais de diagnóstico bucal subjetivos e tardios, permitem

a detecção de lesões periapicais apenas em estágios avançados de desmineralização óssea. Modelos de IA capazes de identificar precocemente cárie, perda óssea periodontal e patologias pulpares a partir de imagens radiográficas, termográficas e clínicas poderiam transformar a atenção odontológica na atenção primária regional, mas dependem de datasets representativos das populações locais que ainda não existem. A equidade e a universalidade do SUS podem ser ampliadas por meio da disponibilização de ferramentas tecnológicas calibradas para a população a ser atendida. E, transversalmente, a soberania científica e tecnológica nacional, ameaçada pela dependência crescente de modelos desenvolvidos com dados não representativos da população brasileira.

3.6 Caminhos de Solução

O enfrentamento do desafio aqui proposto requer avanços simultâneos em cinco frentes interdependentes.

A primeira é o desenvolvimento de infraestrutura computacional soberana, instalada nas ICTs, com área de atuação no Brasil Profundo. *Clusters* com capacidade de treinamento de modelos em grande escala, operando sob governança institucional de dados sensíveis de populações tradicionais, são a pré-condição técnica para o restante. Questão de pesquisa em aberto: como definir arquiteturas de referência de baixo custo e alto desempenho adaptadas às condições de infraestrutura física e de recursos humanos disponíveis na região.

A segunda frente de atuação passa pela construção de metodologias de coleta de dados clínica e culturalmente adequados, cujas respostas podem ser fortemente dependentes da matriz amostral, das condições ambientais e da estabilidade dos materiais ativos empregados. A coleta em comunidades subrepresentadas precisa de protocolos compatíveis com a autonomia comunitária e em conformidade com os termos das Resoluções do CNS, bem como com as especificidades logísticas e relacionais do campo. Questão de pesquisa em aberto: como projetar esses protocolos para maximizar a qualidade e a representatividade dos dados, sem sobrecarregar as comunidades participantes nem reproduzir relações extrativistas de pesquisa.

Complementarmente, vislumbra-se o desenvolvimento de sensores e biossensores eletroquímicos, integrados a sistemas de IA embarcada, com robustez analítica, reprodutibilidade e adaptação às variabilidades ambientais típicas do Brasil Profundo, mantendo acurácia diagnóstica comparável à de métodos laboratoriais convencionais. Com o estabelecimento de protocolos padronizados para aquisição, armazenamento e anotação de dados eletroquímicos, permitindo sua integração com outras modalidades (imagem, texto clínico, dados genômicos) em abordagens multimodais de IA em saúde.

A terceira frente de atuação é o desenvolvimento e a adaptação de técnicas de aprendizado de máquina orientadas para a equidade em contextos de escassez de dados representativos (e.g., transferência de aprendizado, *few-shot learning*, adaptação e generalização de domínio), capazes de generalizar a partir de poucos exemplos em contextos de alta heterogeneidade étnica e ambiental. Incorporando métricas de fairness adequadas para subgrupos étnicos pequenos e historicamente sub-representados.

Outra frente de trabalho a ser desenvolvida está ligada ao desenvolvimento de soluções de *edge computing* e de IA embarcada em dispositivos de baixo custo, capazes de operar sem conectividade estável e de sincronizar dados de forma assíncrona quando a rede estiver disponível [Zhou 2024]. Gerando modelos para operação eficiente em dispositivos

de saúde comunitária, com recursos de processamento limitados, com compromisso entre tamanho e acurácia diagnóstica.

A quinta frente é a especialização de grandes modelos de linguagem para o contexto da saúde indígena e regional, incorporando terminologia clínica em português regional, saberes tradicionais em saúde e vocabulário das etnias do Brasil Profundo. Questão de pesquisa em aberto: como avaliar a qualidade cultural de um modelo de linguagem voltado à saúde das populações sub-representadas, indo além das métricas linguísticas convencionais e incorporando critérios definidos pelas próprias comunidades.

Esses cinco caminhos não são independentes: o avanço em cada um deles cria condições para o avanço nos demais, e a ausência de progresso em qualquer um compromete o conjunto. Essa interdependência é precisamente o que faz do desafio proposto um problema complexo no sentido técnico do termo, ele não permite solução parcial.

3.7 Conclusão

O Desafio aqui proposto não é um problema técnico que aguarda uma solução algorítmica. É um problema estrutural que reflete séculos de exclusão das populações do Brasil Profundo da produção de conhecimento sobre si mesmas. A Computação Aplicada à Saúde tem hoje os instrumentos para reverter esse quadro, e a responsabilidade de fazê-lo antes que os modelos de IA da próxima geração se consolidem sem a participação desses “Brasis”.

Construir conjuntos de dados soberanos, desenvolver modelos equitativos, garantir que a IA em saúde funcione onde não há internet e seja compreendida por quem a usa: esses são desafios simultaneamente técnicos, éticos e políticos. Enfrentá-los é uma questão de justiça social, e para que o princípio constitucional da universalidade do SUS deixe de ser uma promessa e se torne uma realidade para os povos de todos os “Brasis”.

Referências

ALDERMAN, J. E. et al. Tackling algorithmic bias and promoting transparency in health datasets: the STANDING Together consensus recommendations. *Lancet Digital Health*, v. 7, n. 1, p. e64–e88, jan. 2025.

ARIYO, O. Bridging the AI divide: Building Africa's future in the age of intelligent machines. *Medium*, out. 2025. Disponível em: <https://medium.com/@roariyo/bridging-the-ai-divide-building-africas-future-in-the-age-of-intelligent-machines-b8d5d50696e8>. Acesso em: 27 mar. 2026.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília: Presidência da República, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm.

CARROLL, S. R. et al. The CARE Principles for Indigenous Data Governance. *Data Science Journal*, v. 19, n. 1, p. 43, 2020. doi: 10.5334/dsj-2020-043.

CELI, L. A. et al. Sources of bias in artificial intelligence that perpetuate healthcare disparities. A global review. *PLOS Digital Health*, v. 1, n. 3, e0000022, 2022. doi: 10.1371/journal.pdig.0000022.

COIMBRA JR., C. E. A.; SANTOS, R. V. Saúde e povos indígenas no Brasil: reflexões a partir do I Inquérito Nacional de Saúde e Nutrição Indígena. *Cadernos de Saúde Pública*, v. 28, n. 4, p. 641–644, 2012.

FIELDS, C. T. et al. Governance for anti-racist AI in healthcare: integrating racism-related stress in psychiatric algorithms for Black Americans. *Frontiers in Digital Health*, v. 7, art. 1492736, 2025. doi: 10.3389/fdgth.2025.1492736.

FIGUEIREDO, B. S. et al. Mortalidade materna no Brasil (2020–2024): distribuição macrorregional, causas diretas e indiretas. *Revista JRG de Estudos Acadêmicos*, 2025. doi: 10.55892/jrg.v8i18.2423.

IBGE INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Censo Demográfico 2022: Quilombolas: Primeiros Resultados*. Rio de Janeiro: IBGE, 27 jul. 2023. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv102016.pdf>

IBGE INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Censo Demográfico 2022: Quilombolas e Indígenas, por sexo e idade, segundo recortes territoriais específicos: Resultados do universo*. Rio de Janeiro: IBGE, 3 maio 2024. Disponível em: <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=73107>

IEPS INSTITUTO DE ESTUDOS PARA POLÍTICAS DE SAÚDE. *Saúde dos Povos Indígenas e Quilombolas no Brasil: Relatório Técnico nº 1/2023*. São Paulo: IEPS, jun. 2023. Disponível em: <https://agendamais.org.br/wp-content/uploads/2023/06/ieps-boletim01-saude-povos-indigenas.pdf>

JOSEPH, J. Algorithmic bias in public health AI: a silent threat to equity in low-resource settings. *Frontiers in Public Health*, v. 13, art. 1643180, 23 jul. 2025. doi: 10.3389/fpubh.2025.1643180.

NORORI, N. et al. Addressing bias in big data and AI for health care: A call for open science. *Patterns*, v. 2, n. 10, art. 100347, 2021. doi: 10.1016/j.patter.2021.100347.

OBERMEYER, Z. et al. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, v. 366, n. 6464, p. 447–453, 2019. DOI: 10.1126/science.aax234

PASIPAMIRE, N.; MUROYIWA, A. Navigating algorithm bias in AI: ensuring fairness and trust in Africa. *Frontiers in Research Metrics and Analysis*, v. 9, art. 1486600, 2024. doi: 10.3389/frma.2024.1486600.

SASSEVILLE, M. et al. Bias mitigation in primary health care artificial intelligence models: scoping review. *Journal of Medical Internet Research*, v. 27, e60269, 7 jan. 2025. doi: 10.2196/60269.

SCHEFFER, M. et al. *Demografia Médica no Brasil 2023*. São Paulo: FMUSP/AMB, 2023. Disponível em: https://amb.org.br/wp-content/uploads/2023/02/DemografiaMedica2023_8fev-1.pdf

SEYYED-KALANTARI, L. et al. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nature Medicine*, v. 27, n. 12, p. 2176–2182, 2021. doi: 10.1038/s41591-021-01595-0.

TOPOL, E. J. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. Updated Edition. New York: Basic Books, 2024.

VICTOR, A. Artificial intelligence in global health: An unfair future for health in Sub-Saharan Africa? *Health Affairs Scholar*, v. 3, n. 2, art. qxaf023, 2025. doi: 10.1093/haschl/qxaf023.

WHO| WORLD HEALTH ORGANIZATION. *Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models*. Geneva: WHO, 18 jan. 2024.

WHITTAKER, R. et al. An example of governance for AI in health services from Aotearoa New Zealand. *npj Digital Medicine*, v. 6, n. 1, art. 164, 2023. doi: 10.1038/s41746-023-00882-z.

WU, Jiageng et al. Clinical text datasets for medical artificial intelligence and large language models: a systematic review. *NEJM AI*, v. 1, n. 6, 2024. Disponível em: <https://doi.org/10.1056/AIra2400012>

ZEPEDA, L. E. G.; MARTÍNEZ PINTO, C. E. *Inteligencia artificial centrada en los pueblos indígenas: perspectivas desde América Latina y el Caribe*. Paris: UNESCO, 2023. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000387814>

ZHOU, Z., Chen, X., Li, E., Zeng, L., Luo, K. and Zhang, J. (2019) Edge Intelligence: Paving the Last Mile of Artificial Intelligence with Edge Computing. Proceedings of the IEEE, 107, 1738-1762. <https://doi.org/10.1109/jproc.2019.2918951>

Sobre os Autores

Ana Clécia Santos de Alcântara

Doutora em Química. Professora da Coordenação do Curso de Química Bacharelado da UFMA. Líder do Grupo de Pesquisa em Materiais Híbridos e Bionanocompósitos (Bionanos/UFMA). Professora dos Programas de Pós-Graduação em Química (PPGQuim/UFMA), Ciência e Tecnologia de Materiais (PPGCTM/UFMA) e do Doutorado Associativo em Química (DQUIM/UFMA). Atua no desenvolvimento de materiais híbridos nanoestruturados e bionanocompósitos voltados à remediação ambiental, aplicações biomédicas e energéticas. Bolsista de Produtividade em Pesquisa do CNPq Nível 2.

Anselmo Cardoso de Paiva

Doutor em Informática. Professor Titular da Coordenação de Ciência da Computação da UFMA. Professor do Programa de Pós-Graduação em Engenharia Elétrica, do Programa de Pós-Graduação em Ciência da Computação da UFMA, e do Doutorado em Ciência da Computação em Associação UFMA-UFPI. Coordenador do Núcleo de Computação Aplicada (NCA-UFMA) e do Doutorado em Ciência da Computação em Associação UFMA/UFPI. Atua em processamento de imagens médicas, processamento de linguagem natural, aprendizado profundo e visão computacional. Bolsista de Produtividade em Pesquisa do CNPq Nível 2.

Aristófares Correa Silva

Doutor em Informática. Professor Titular da Coordenação de Engenharia Elétrica da UFMA. Coordenador do Núcleo de Computação Aplicada (NCA-UFMA). Professor do Programa de Pós-Graduação em Engenharia Elétrica, do Programa de Pós-Graduação em Ciência da Computação da UFMA, e do Doutorado em Ciência da Computação em Associação UFMA-UFPI. Atua em Ciência da Computação com ênfase em processamento de imagens médicas, reconhecimento de padrões, inteligência artificial e visão computacional. Bolsista de Produtividade em Pesquisa do CNPq Nível 2.

Cecilia Claudia Costa Ribeiro

Doutora em Cariologia. Professora Titular da Coordenação de Odontologia da UFMA. Professora do Programa de Pós-Graduação em Odontologia e do Programa de Pós-Graduação em Saúde Coletiva da UFMA. Atua nas áreas de Odontologia e Epidemiologia, com ênfase em doenças crônicas não transmissíveis, integralidade em saúde, saúde da população indígena e primeiros mil dias de vida (abordagem DoHAD). Bolsista de Produtividade em Pesquisa do CNPq Nível C.

Elaine de Paula Fiod Costa

Médica com doutorado em Oftalmologia. Professora Adjunta IV. Atua em retina cirúrgica (buraco macular e descolamento de retina), trauma ocular, síndrome congênita

pelo Zika Vírus, propriedades biofísicas e biomecânicas de tecidos oculares, e rastreamento e tratamento de retinopatia diabética em povos originários do Maranhão.

Erika Bárbara Abreu Fonseca Thomaz

Doutora em Saúde Pública (Epidemiologia). Professora Associada da Coordenação Especial de Saúde Pública da UFMA e Professora dos Programas de Pós-Graduação em Saúde Coletiva e em Odontologia da UFMA. Atua em epidemiologia, análise de dados em saúde, avaliação de serviços de saúde, atenção primária à saúde, saúde materno-infantil e saúde bucal coletiva. Bolsista de Produtividade em Pesquisa do CNPQ Nível C.

Flavio Santos Damos

Doutor em Química Analítica. Professor da Coordenação de Química Bacharelado da UFMA. Docente do Programa de Pós-Graduação em Química (PPGQuim/UFMA) e do Doutorado Associativo em Química UFMA-IFMA. Atua em Química Analítica e Eletroanalítica, com ênfase em sensores e biosensores, investigação de processos interfaciais, filmes finos, sistemas fotoeletroquímicos, microscopia eletroquímica de varredura, ressonância de plásmem de superfície acoplada a técnicas eletroquímicas e microbalança eletroquímica de cristal de quartzo. Bolsista de Produtividade em Pesquisa do CNPq Nível 2.

Geraldo Braz Júnior

Doutor em Engenharia de Eletricidade. Professor Associado III da Coordenação de Ciência da Computação da UFMA. Docente do Programa de Pós-Graduação em Ciência da Computação (PPGCC/UFMA) e do Doutorado em Ciência da Computação em associação UFMA-UFPI. Atua em visão computacional, aprendizado de máquina e aprendizado profundo, com foco em análise de imagens médicas, diagnóstico auxiliado por computador e inteligência computacional aplicada ao setor público. Bolsista de Produtividade em Pesquisa do CNPq Nível 2.

Harvey Alexander Villa Vélez

Doutor em Engenharia e Ciência de Alimentos. Professor Adjunto da Coordenação de Engenharia Química da UFMA. Atua nas áreas de biomateriais, energia renovável, aproveitamento de resíduos sólidos e biorremediação de efluentes. Desenvolve projetos de pesquisa voltados ao monitoramento e ao tratamento de potabilização da água em comunidades quilombolas e tradicionais do Maranhão.

João Dallyson Sousa de Almeida

Doutor em Engenharia Elétrica. Professor Associado da Coordenação de Ciência da Computação da UFMA. Coordenador do Laboratório de Visão e Processamento de Imagens (VipLab-UFMA) e Vice-Coordenador do Programa de Pós-Graduação em Ciência da Computação (PPGCC/UFMA). Atua em processamento de imagens

oftalmológicas, aprendizado profundo, visão computacional e séries temporais. Bolsista de Produtividade em Pesquisa do CNPq Nível C.

João Otávio Bandeira Diniz

Doutor em Engenharia Elétrica. Professor EBTT do Instituto Federal do Maranhão. Coordenador da Fábrica de Inovação do Campus Grajaú e do Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas. Professor dos Programas de Pós-Graduação em Ciência da Computação da UFMA e em Computação Aplicada do IFMA. Atua em inteligência artificial aplicada à saúde, com foco em ciência de dados, aprendizado profundo, visão computacional e modelos de linguagem ampla.

Luciana Salles Branco De Almeida

Doutora em Farmacologia, Anestesiologia e Terapêutica. Professora Associada da Coordenação de Odontologia. Docente no Programa de Pós-Graduação em Odontologia da UFMA. Atua em métodos diagnósticos e alternativas terapêuticas visando o controle de infecções, dor, inflamação e reabsorção óssea inflamatória

Rita de Cássia Silva Luz

Doutora em Química Analítica. Professora Titular da Coordenação de Química do Bacharelado da UFMA. Co-coordenadora do Laboratório de Sensores, Dispositivos e Métodos Analíticos da UFMA e Coordenadora do Doutorado Associativo em Química UFMA-IFMA. Professora do Programas de Pós-Graduação em Química e do Programas de Pós-Graduação em Biotecnologia da UFMA. Atua em Química Analítica, com ênfase em sensores e biossensores, sistemas eletroquímicos, fotoeletroquímicos e eletroquimioluminescentes, com aplicações analíticas e no desenvolvimento de nanomateriais. Bolsista de Produtividade em Pesquisa do CNPq Nível 1D.

Soraia de Fátima Carvalho Souza

Doutora em Odontologia. Professora Titular da Coordenação de Odontologia da UFMA. Docente do Programa de Pós-graduação em Odontologia da UFMA. Atua em Endodontia, principalmente explorando a relação entre as alterações bucais das doenças sistêmicas, com foco na anemia falciforme, doenças cardiovasculares, diabetes mellitus, doenças autoimunes reumáticas, e diagnóstico das patologias pulpares e periapicais, buscando elucidar o papel da infecção endodôntica na expressão de marcadores precoces das doenças cardiovasculares e resistência insulínica.

Tiago Bonini Borchart

Doutor em Computação pela Universidade Federal Fluminense (UFF). Professor do Curso de Computação da UFMA. Coordenador do Bacharelado em IA da UFMA. Coordenador do Laboratório de Inteligência Interativa (LINT-UFMA). Realizou Pós-Doutorado na UFF. Atua nas áreas de gamificação, inteligência artificial, visão computacional e tecnologias na educação. Bolsista de Produtividade em Pesquisa FAPEMA/CNPq Nível C.