Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Whose Body Is the Algorithm For? Diagnostic Artificial Intelligence, Indigenous Health, and the Politics of Representational Validity

Américo Mota

americomota@gmail.com

Graduate Program in Human Ecology and Socio-Environmental Management (PPGEcoH), State University of Bahia (UNEB)

Bruna Almeida

AMMO Institute / CDI

Isabella Bunge

AMMO Institute / CDI

Jandir Nicacio

University Hospital of the Federal University of the São Francisco Valley (HU-UNIVASF / HU-BRASIL)

Vanessa Pereira

Federal University of the São Francisco Valley (UNIVASF)

Thiago Carvalho

AMMO Institute / CDI

Paula Maduro

University Hospital of the Federal University of the São Francisco Valley (HU-UNIVASF / HU-BRASIL)

Paulo Schwingel

University of Pernambuco (UPE)

Carlos Souza

Federal University of the São Francisco Valley (UNIVASF)

Orlando Gomes

University Hospital of the Federal University of the São Francisco Valley (HU-UNIVASF / HU-BRASIL)

Ana Pinto

AMMO Institute / CDI

Caroline Chieco

AMMO Institute / CDI

Anderson Armstrong

Graduate Program in Human Ecology and Socio-Environmental Management (PPGEcoH), State University of Bahia (UNEB)

Systematic Review

Keywords: Artificial Intelligence, Diagnostic Imaging, Representational Validity, Algorithmic Bias, Indigenous Data Sovereignty, Diagnostic Justice

Posted Date: May 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9801630/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Artificial intelligence has moved from a research field organized around the formalization of reasoning to a clinical infrastructure built on data-dependent classification. In diagnostic imaging, deep learning and foundation model systems have achieved specialist-level performance in selected tasks across ophthalmology, pathology, radiology, dermatology, and endoscopy. However, benchmark performance does not transfer reliably to deployment: external validation cohorts remain demographically narrow, dataset bias produces measurable algorithmic harm, and populations most under-represented in training are most vulnerable to misdiagnosis. This systematic search-informed integrative review synthesizes evidence across seven thematic axes — the philosophical foundations, historical trajectory, and conceptual taxonomy of AI; AI in healthcare and diagnostic imaging; algorithmic bias and external validation; and Indigenous health and data sovereignty — to argue that diagnostic AI cannot be evaluated by technical accuracy alone. This review advances the concept of *representational validity in diagnostic AI*, which refers to the degree to which the populations, contexts, imaging protocols, and sociocultural conditions used to train and validate an AI system correspond to those in which it will be deployed. The corpus contained no peer-reviewed empirical study that developed or validated a diagnostic-imaging AI system in an Indigenous population in Brazil or Latin America; this scarcity is treated as a substantive finding consistent with structural inequity analyses across global AI research. The article connects representational validity to algorithmic fairness, external validation, CARE and OCAP data-sovereignty frameworks, and diagnostic justice, and proposes a research agenda that treats community governance and population-representative validation as constitutive of, rather than supplementary to, clinical validity.

1. Introduction: AI, Representation, and Diagnostic Justice

A diagnostic-imaging model that reads a chest radiograph in São Paulo with specialist-level accuracy does not, on that evidence, know how to read one in a Distrito Sanitário Especial Indígena in Northeastern Brazil; whether it can be expected to is not a technical question, but a sociotechnical, epistemic, and governance question that the field has yet to confront systematically.

Artificial intelligence has become an integral part of contemporary healthcare. AI systems are now embedded in the pathways that determine who is diagnosed, when, and with what confidence across clinical decision support, screening and triage, risk prediction, and the analysis of medical images (Bates et al., 2021; Sharma et al., 2022). In diagnostic imaging in particular, deep-learning architectures have reached, in selected tasks, performance comparable to specialist clinicians (De Fauw et al. 2018; Ran et al. 2019; Li et al. 2024). These advances have been read as evidence that medical imaging is an exemplary site for clinical AI implementation, and they have authorized a wave of regulatory, commercial, and academic interest in deployment.

Interest has outpaced careful consideration of the conditions under which such systems are reliable. Performance metrics are produced on datasets that are assembled by particular institutions, in particular settings, with particular imaging equipment, for particular populations. When the same

systems are tested outside the conditions of their development, their performance typically degrades (Chen Q et al. 2025; Qian et al. 2023). When they are tested in populations under-represented in training data, the degradation is rarely symmetrical: the populations least visible in the dataset are most likely to bear the cost of the model's miscalibration (Yogarajan et al. 2022; Joseph 2025; Ghanem et al. 2025). What is at stake in these patterns is not only the precision of the next benchmark but also the question of which bodies a diagnostic AI is constitutively for.

This review proposes that the answer requires a conceptual move beyond external validity, fairness, and dataset shifts taken in isolation. It introduces representational validity in diagnostic AI as a unifying lens: the degree to which the populations, clinical contexts, imaging protocols, disease patterns, demographic groups, geographic settings, and sociocultural conditions used to train and validate an AI system correspond to those in which the system is deployed. Representational validity is a necessary condition for clinically reliable and ethically defensible diagnostic AI. It is not a property of an algorithm in isolation; rather, it is a property of the relationship between an algorithm, its data ecology, and its deployment context.

Indigenous populations are a critical case for this argument. The corpus assembled for this review, drawn from PubMed/MEDLINE, Scopus, Web of Science, IEEE Xplore, SciELO, LILACS, and Google Scholar, in English, Portuguese, and Spanish, contains no peer-reviewed empirical study that develops, validates, or deploys a diagnostic-imaging AI system in an Indigenous-defined population in Brazil or Latin America. Adjacent evidence from Aotearoa New Zealand (Yogarajan et al. 2022), Canada (Khan et al. 2025; Mashford-Pringle and Pavagadhi 2020), and the Western Pacific (Watego et al. 2025; Hurst et al. 2026) is informative but not substitutive. The structured absence of Indigenous-specific imaging-AI literature is consistent with the structural-inequity claims of recent commentary (Lancet Digital Health 2023; Joseph 2025; Ghanem et al. 2025; Walter et al. 2020). This manuscript treats this absence as a substantive finding rather than a gap to be apologized for.

Framing carries philosophical weight. Diagnostic AI converts clinical judgement into computational classification; the conversion is enabled and limited by the formalizations that make data and labels into a model. Several recent contributions in the literature have argued that intelligence resists clean formalization (Svensson 2021; Nadin 2019), that the Turing-test debate remains unresolved on what counts as machine thinking (Gonçalves 2023), and that the architectural standardization of contemporary AI has its own epistemic and ethical costs (Fishman and Hancox-Li, 2022). Read together with the equity literature, these contributions suggest that representational validity is not only a methodological standard but a philosophical question: which bodies become the normal body in an imaging dataset, and what is the cost — to those bodies and to those outside them — of treating the normalization as natural?

The review is organized as an analytical funnel. Section 2 describes the review design and the analytical strategy. Section 3 sets out the philosophical and historical foundations relevant to the diagnostic-AI argument. Section 4 traces the conceptual trajectory from symbolic systems to foundation models.

Section 5 examines AI in healthcare and diagnostic imaging, focusing on validation limits. Section 6 develops the concept of representational validity. Section 7 — the analytical climax — addresses Indigenous health, data sovereignty, and diagnostic AI; its argument is anchored in the gap-map findings produced by this review. Section 8 outlines the research agenda. Section 9 concludes.

The thesis of this manuscript is that diagnostic AI cannot be evaluated by technical accuracy alone; it must be assessed through representational validity — the correspondence between training-and-validation populations and deployment contexts — a condition whose failure is starkly visible in the near-absence of Indigenous populations from diagnostic-imaging AI literature.

2. Review Design and Analytical Strategy

This review is a systematic search-informed integrative review with scoping and evidence-mapping components. The combined design was chosen because the research question crosses philosophical, historical, technical, clinical, ethical, and equity dimensions, and because preliminary scoping suggested that direct evidence on AI-based diagnostic imaging in Indigenous populations might be scarce, a possibility that itself required analytical treatment.

2.1 Aim and review question

This review examines the historical-conceptual evolution of AI, its applications in healthcare and diagnostic imaging, and the evidence gap regarding Indigenous populations, algorithmic bias, and representational validity. The primary review question is as follows: how has AI evolved conceptually and technically; how is it deployed in healthcare and medical diagnostic imaging; and what evidence and gaps characterize its validation, bias, and equity implications in Indigenous and underrepresented populations?

2.2 Sources and search strategy

Records were retrieved through structured searches of PubMed/MEDLINE, Scopus, Web of Science, IEEE Xplore, SciELO, LILACS, and Google Scholar. Search blocks combined AI terms (artificial intelligence, machine learning, deep learning, convolutional neural networks, transformers, foundation models, large language models, generative AI, computer vision), healthcare terms (clinical decision support, screening, risk prediction, precision medicine, clinical NLP, digital health, epidemiological surveillance), imaging terms (radiology, computed tomography, magnetic resonance imaging, ultrasound, X-ray, ophthalmology, fundus imaging, optical coherence tomography, dermoscopy, histopathology, endoscopy), equity terms (algorithmic bias, dataset bias, fairness, health equity, external validation, generalizability, representativeness, underrepresented populations), and Indigenous terms in English, Portuguese, and Spanish (Indigenous Peoples, Indigenous Health, Aboriginal, First Nations, Native American, Māori, Povos Indígenas, Pueblos Indígenas, Saúde Indígena, Salud Indígena, Indigenous data sovereignty, CARE Principles, OCAP, algorithmic sovereignty, cognitive sovereignty).

No date restrictions were applied to philosophical, historical, or conceptual AI sources. Records of applied AI in medical imaging were prioritized from 2012 onward, reflecting the deep-learning inflection point in computer vision. Records on algorithmic fairness, bias, and Indigenous data sovereignty were prioritized from 2015 onward, while foundational sources predating this window were retained and remained in active circulation.

2.3 Inclusion and exclusion criteria

Records were included if they were peer-reviewed articles, systematic or scoping reviews, integrative or narrative reviews, landmark empirical studies, conceptual or theoretical articles, historical analyses, Indigenous governance frameworks, or policy documents with substantive analytical content; addressed at least one of the seven thematic axes; offered an analytical or empirical contribution beyond superficial commentary; and contained sufficient bibliographic metadata (author, year, title, journal or source, and either a DOI or stable URL) to be cited verifiably. Gray literature was included if it addressed Indigenous data sovereignty directly or if peer-reviewed evidence on a relevant subtopic was scarce.

Records were excluded if published in venues identified as predatory, if they were superficial commentary without conceptual or empirical contribution, if they were technical AI papers without engagement with health, imaging, ethics, equity, or society, if they conflated AI, ML, deep learning, large language models, and generative AI without conceptual clarity (unless cited as examples of conceptual confusion), if bibliographic metadata were insufficient for verification, if they duplicated previously included records (deduplication by DOI then by normalized title), or if they had been retracted (unless the retraction itself was under discussion).

Peer-reviewed sources from venues with limited or pending indexing in Scopus or Web of Science were deliberately retained. These include regional journals, such as the Journal of Electronics, Electromedical Engineering, and Medical Informatics (JEEEMI, Indonesia), the International Journal of Healthcare Information Systems and Informatics (IJHISI, IGI Global, India), the Revista Brasileira de Preservação Digital (Brazil), and the Anais do Workshop de Informação, Dados e Tecnologia (WIDaT, Brazil), and newly launched journals from established academic publishers whose Scopus/Web of Science coverage is still pending (notably Machine Learning: Health, IOP Publishing), in each case where the source offered a substantive technical, methodological, or Indigenous governance contribution to the thematic axes. This decision recognizes the value of diversity in the technical-AI literature and is consistent with the review's central commitment to representational validity: a corpus that systematically excludes regional venues or treats time-bound indexing lag as evidence of low quality would itself reproduce the geographic, epistemic, and institutional narrowness the review aims to surface.

2.4 Screening and charting

The screening process was conducted in three stages. The inventory of bibliographic exports retained 2,403 unique records across 53 export files (3,824 rows; 1,421 duplicates merged). At the eligibility stage, the title, abstract, and available excerpt were assessed against the inclusion and exclusion criteria; a record was retained for full charting when its abstract or excerpt clearly placed it within at least

one thematic axis. At the charting stage, the analytically central sources were extracted into a structured chart (bibliographic metadata, study type, evidence type, thematic axis, AI technique, imaging modality where applicable, health domain, population studied, whether Indigenous or otherwise underrepresented populations were included, country or region, dataset size where applicable, main findings, author-reported limitations, equity and bias considerations, data sovereignty considerations, relevance to diagnostic AI, Indigenous health, and representational validity, evidence strength judgement, and free-text notes). The identification, screening, eligibility, and inclusion flow is summarised in Fig. 1.

2.5 Synthesis

The synthesis was thematic and analytical rather than meta-analytic, and pooled effect estimates were not produced. Sources were grouped along seven thematic axes: philosophical foundations (E1), historical trajectory (E2), conceptual taxonomy (E3), AI in healthcare (E4), AI in diagnostic imaging (E5), algorithmic bias, fairness, and representational validity (E6), and Indigenous health and data sovereignty (E7), and integrated through an argument map that fixes the manuscript's main claim and the eight supporting claims that organize it. The intersection of E5 and E7 (diagnostic-imaging AI in Indigenous populations) received particular analytical attention.

2.6 Treatment of scarcity

Where direct evidence on AI-based diagnostic imaging in Indigenous populations in Brazil and Latin America was limited or absent, the scarcity itself was treated as an analytical finding rather than a defect of the review. The reasoning is twofold: the absence of evidence in a high-priority area of global health, after a wide and structured search, is itself informative about the field's priorities and accessibility; and the manuscript's central conceptual argument — that representational validity is a necessary condition of clinically reliable diagnostic AI — is *supported* rather than weakened by the demonstration that Indigenous populations are structurally underrepresented in the literature.

2.7 Limitations of the review design

This review has several limitations. The heterogeneity of sources precluded meta-analytic synthesis. The scarcity of Indigenous-specific AI literature constrained the depth of synthesis of E5 × E7 and required reliance on adjacent and analogical evidence (notably from Māori, First Nations, and remote-community contexts). Reliance on English-indexed databases, despite explicit Portuguese and Spanish search terms, likely undercounted Latin American Indigenous health publications. The distinction between diagnostic imaging AI and broader health AI is not always clear in the literature; sources addressing AI in healthcare without isolating imaging are used to characterize context rather than to ground imaging-specific claims.

3. From Intelligence to Computation: Philosophical and Historical Foundations

The history of AI is also a history of the assumption that intelligence can be formalized. The Turing test, proposed in 1950 in a paper whose contemporary readers continue to find layered and contestable (Turing, 1950; Gonçalves, 2023), articulates the methodological intuition that intelligent behavior can be operationalized through observable performance. The intuition was institutionalised at the 1956 Dartmouth proposal, where a small group articulated the hypothesis that every aspect of learning or any other feature of intelligence can in principle be precisely described so that a machine can be made to simulate it (McCarthy et al. 1955; see also Haenlein and Kaplan 2019 for the historical reconstruction). The hypothesis founded the symbolic AI tradition, which dominated the field for several decades and encoded human expertise through hand-crafted rules, formal logic, and explicit knowledge representations (Haenlein and Kaplan 2019; Mundlamuri et al. 2025).

The symbolic project encountered limitations that were not merely technical. Early symbolic systems performed strongly in constrained domains (theorem proving, game playing, controlled natural-language parsing) but proved brittle in the face of the ambiguity of everyday and clinical situations. The reliance on explicit rule formulation makes scaling exponentially expensive and exposes the field to enduring philosophical critiques of computational intelligence (Svensson 2021; Nadin 2019; Gonçalves 2023). Svensson (2021) argues, drawing on Nicholas of Cusa, that intelligence is an organic capacity to handle the unknown as it unfolds in the present. This argument matters for a diagnostic AI review because it locates the limits of computational intelligence in the situations on which clinical practice is built: novel presentations, unclear histories, and comorbidities that exceed any prior categorization. Nadin (2019) reaches a similar conclusion from a different starting point: computation, as algorithmic embodiment of know-how, is closed to meaning and can mimic but not create the situated judgements that healthcare requires.

A second tradition, connectionism, proposes a different account of how computation can approximate cognition. Pollack (1989) documented the foundations and resurgence of connectionist models — networks of simple processing units, inspired by biological neurons, that learn distributed representations from data rather than relying on explicit programming. The early articulation of artificial neural units traces back to McCulloch and Pitts (1943), whose formalism became the basis of artificial neural network research. The practical training of multilayer networks awaited the backpropagation algorithm of the mid–1980s (Rumelhart, Hinton, and Williams 1986), which enabled gradient-based learning of complex nonlinear mappings and prepared the ground for the deep-learning era that would arrive two decades later (Deng 2018; Mundlamuri et al. 2025).

Philosophical history is important for a diagnostic-AI review because diagnostic AI transforms clinical judgement into computational classification. Classification presupposes a formalization of which features of a case are clinically relevant; it presupposes that the relevant variation has been anticipated by the dataset and labelling protocol. The strong-vs-weak AI debate — central to a longer philosophical conversation on machine cognition (Luger and Chakrabarti 2017; Gonçalves 2023; Nadin 2019; Svensson 2021), and reaching back to Searle's (1980) Chinese Room argument — turns precisely on what such formalisation can and cannot capture. In a clinical setting, what is not captured is sometimes

the most diagnostically significant feature: an atypical presentation that does not fit the training categories, an imaging artifact specific to a particular ultrasound probe used in field settings, or a comorbidity pattern more common in a specific population than in any benchmark cohort. The philosophical question of formalization, in this sense, is not a prelude to the empirical literature; it is the same question that returns when we ask whether a model trained on one population is for another population.

Recent critical contributions on cognitive imperialism (Ofosu-Asare 2024) and the epistemic implications of architectural unification in machine learning (Fishman and Hancox-Li 2022) extend this line of thought. Ofosu-Asare argues that standard datasets and design assumptions of contemporary AI encode Western epistemic assumptions and treat Indigenous and non-Western knowledge systems as either invisible or available for extraction. Fishman and Hancox-Li argue that the dominance of the transformer architecture across application domains, while productive, generalizes a particular epistemic stance about how problems should be represented. Both arguments converge on the conclusion that the formalization question of early AI has not been resolved by scale; it has been displaced into the politics of data set construction and architectural standardization.

Philosophical history is relevant to this manuscript in three ways. This explains why the question of what gets formalized matters for clinical AI, why the question of whose body becomes the normal body in a dataset is continuous with older debates about whose cognition the field treats as paradigmatic, and why the move from rule-based to data-driven AI did not dissolve these questions but relocated them. The next section traces this relocation.

4. From Symbolic Systems to Foundation Models: Conceptual Trajectories of AI

The mature symbolic period — roughly the 1970s and the 1980s — produced expert systems that captured commercial and academic attention. Programs such as MYCIN (for bacterial infection diagnosis) and DENDRAL (for chemical analysis) encoded domain knowledge as if-then rule bases and represented the peak of the symbolic project's clinical aspirations (Haenlein and Kaplan 2019; Deng 2018). They also exposed the project's limits: the inability to handle uncertainty gracefully, the high cost of knowledge acquisition from experts, and the failure to generalize beyond narrow domains. The disillusionment that followed produced the so-called *AI winters* of the mid-1970s and the late 1980s, which were periods of reduced funding and credibility (Haenlein and Kaplan 2019).

Connectionism, which had been pursued in parallel since at least the 1940s, re-emerged as the deep learning era began in earnest around 2012. The trajectory has been documented as the convergence of three factors: large labelled datasets (notably ImageNet for computer vision), advances in GPU computing, and architectural innovations that enabled the training of increasingly deep networks (Deng 2018; Mundlamuri et al. 2025). Convolutional neural networks (CNNs), originally developed for spatially structured inputs, have become dominant in computer vision and have produced breakthroughs that will

soon transform medical imaging. Recurrent architectures and Long Short-Term Memory (LSTM) variants excel at sequential data, and deep reinforcement learning demonstrates complex decision-making in game and simulation environments. The cumulative effect was a shift in the locus of validity: where symbolic AI had located validity in the design of rules and the curation of knowledge bases, deep learning located validity in the composition of training data and the conditions of model fitting.

The introduction of the Transformer architecture (Vaswani et al. 2017) reorganized sequence modelling around self-attention rather than recurrence, enabling parallelized training on extremely large text corpora. The architecture underpins the current generation of Large Language Models (LLMs) — from BERT and the GPT series to more recent multimodal systems — and has been documented across several recent surveys (Raiaan et al. 2024; Wang Z et al. 2024b; Zhou et al. 2025). The same architecture has been extended beyond text: vision transformers, multimodal foundation models, and increasingly retrieval-augmented systems are now active research directions in medical imaging (Morano et al., 2025). Fishman and Hancox-Li (2022) read this architectural unification critically, arguing that the apparent domain-agnostic success of transformers carries epistemic and ethical implications that the field has not yet adequately confronted: a single architectural family, trained on a particular composition of public data, becomes the lens through which a widening range of problems is approached.

Foundation models — large-scale models pre-trained on broad data and adaptable to many downstream tasks — represent a recent paradigm whose sociotechnical stakes are visible (Schneider 2022; Zhou et al. 2025). Schneider (2022) explicitly situates the development of foundation models in their historical and sociotechnical contexts, noting that the choices about pre-training data, the populations whose language is over- or under-represented, and the institutions that can afford training are not incidental to model behavior but constitutive of it. Wang et al. (2024b) survey the history and ethical principles of LLM development with a similar emphasis, while Sheth et al. (2023) propose a neuro-symbolic line of work in which the pattern-recognition capacities of neural networks are integrated with the reasoning capacities of symbolic systems, an integration whose appeal for clinical decision-making lies in the demand for explainability and transparent reasoning.

Taxonomic clarification is important for the rest of the manuscript and is often missing from the literature. Artificial intelligence is a broad project for building systems that perform tasks associated with human intelligence. Machine learning (ML) is a subfield in which systems learn patterns from data rather than explicit rules. Deep learning (DL) is a class of ML method built on artificial neural networks with multiple layers. Convolutional neural networks (CNNs) are deep architectures designed for spatially structured inputs and have dominated medical imaging. Transformers are deep architectures built on self-attention that dominate language and vision. Foundation models are large-scale models that are pre-trained on broad data and adapted to many downstream tasks. Large language models (LLMs) are foundation models that are specialized for language. Generative AI refers to models capable of producing text, images, code, or synthetic data, and overlaps but is not identical to LLMs. The conflation of these terms in clinical and policy writing is itself a source of confusion (Moulaei et al. 2024) and an obstacle to careful argument about which validity claims a particular system can and cannot support.

The transition from symbolic to data-driven AI carries an analytical lesson for diagnostic imaging. The locus of validity has shifted from rule design to data composition; the question of whether a model is reliable has become inseparable from the question of whose data the model was built on. This is the conceptual move that the next section makes.

5. AI in Healthcare and Diagnostic Imaging: Technical Promise and Validation Limits

AI in healthcare has matured along two related but distinguishable trajectories: the application of ML and DL methods to clinical decision-making across electronic health records, risk prediction, and surveillance, and the application of DL methods, particularly CNNs and increasingly transformer-based models, to medical images. Both trajectories show consistent patterns: strong performance under controlled validation conditions and a recurring gap between research-stage metrics and clinical deployment.

5.1 AI in healthcare

Clinical decision support systems built using ML have moved from prototypes to early implementation in multiple domains. A scoping review in *npj Digital Medicine* found that AI has been applied to adverse event detection, clinical deterioration prediction, and medication error prevention, with retrospective performance typically exceeding prospective evaluation (Bates et al., 2021). A systematic review of AI for clinical decision support in cardiovascular intensive care reported that ML-based monitoring systems can detect deterioration earlier than conventional approaches; however, external validation across clinical settings remains limited (Moazemi et al., 2023). A scoping review of ML tools implemented in hospital settings found that while many systems show promising research-stage performance, few have been rigorously validated in routine workflows (Tricco et al. 2023). Across these reviews, recurring challenges included data quality, algorithmic bias, limited explainability, and difficulty in integrating AI tools with existing health-system infrastructure.

Risk stratification is an early target. A scoping review in EClinicalMedicine documented how ML models integrating genetic, clinical, and lifestyle data can improve cardiovascular risk prediction beyond traditional calculators (Singh et al., 2024). Clinical natural language processing has expanded the scope of structured information extraction from unstructured clinical text, supporting phenotyping, prognosis estimation, and patient surveillance. A systematic review noted a persistent annotation bottleneck, which was partially addressed by active learning and distant supervision (Spasic and Nenadic 2020). The emergence of generative AI has further expanded clinical NLP capabilities; a scoping review documented applications from clinical-documentation automation to patient-communication support, alongside risks of hallucination, privacy violation, and the difficulty of validation in clinical contexts (Moulaei et al. 2024).

The pandemic accelerated the integration of AI into surveillance. Naseem et al. (2020) reviewed AI and ML applications for COVID–19 in low- and middle-income countries, identifying both opportunities (workforce shortages and diagnostic-access gaps) and barriers (digital infrastructure, locally representative training data, and regulatory gaps). A systematic review of AI in early warning systems for infectious disease surveillance found that ML models can detect outbreak signals earlier than traditional methods, while data quality and timeliness in resource-limited settings remain constraints (Villanueva-Miranda et al., 2025). A broad scoping review mapped the integration of AI in telehealth, mobile health, and remote monitoring, modalities particularly relevant for reaching underserved populations (Sharma et al., 2022).

5.2 AI in diagnostic imaging

Diagnostic imaging AI has produced the most visible clinical-grade demonstrations in the field. In ophthalmology, a deep-learning system was shown to diagnose retinal disease from optical coherence tomography (OCT) scans with performance matching or exceeding expert clinicians, organized as a two-stage architecture (segmentation of the OCT volume into tissue maps, followed by classification of pathology) that provided both accuracy and a degree of interpretability (De Fauw et al., 2018). A three-dimensional deep-learning system for glaucomatous optic neuropathy detection from spectral-domain OCT achieved an area under the receiver-operating-characteristic curve of 0.969 in primary validation and demonstrated performance across three independent external validation datasets, highlighting the importance of multicenter validation for establishing generalizability (Ran et al. 2019). A real-world multicenter external validation of a DL detection system for glaucomatous optic neuropathy reinforced the same lesson, and methodological guidance for AI workflow and external validation in eye-disease diagnosis has more recently been articulated (Qian et al., 2023; Chen Q et al., 2025). More recent work has extended the trajectory: an AI cloud platform for OCT-based retinal anomaly screening deployed in real clinical environments (Chen X et al. 2025); a multimodal foundation model and benchmark for comprehensive retinal OCT analysis (Morano et al. 2025); and a prospective multicentre clinical trial showing that DL assistance improved the diagnostic accuracy of junior ophthalmologists across thirteen major fundus diseases (Li et al. 2024).

Radiological imaging has been transformed by deep learning in CT and MRI modalities. A comprehensive survey documented progress in U-Net variants, attention mechanisms, and multi-scale architectures for glioma segmentation on MRI, with top models reaching Dice scores above 0.90 on benchmark datasets such as BraTS (Liu et al. 2022). A systematic review and meta-analysis of CNN models for meningioma segmentation synthesized evidence across architectures and training strategies (Wang et al., 2024a). A multimodality framework for segmenting 40 anatomical classes across MRI and CT illustrated the trend toward general-purpose segmentation models (Häntze et al. 2025), and body-region classification across MRI and CT has supported workflow automation (Raffy et al. 2023).

Computational pathology has advanced rapidly, with deep learning approaches applied to whole-slide images. Weakly supervised survival CNNs equipped with visual attention have been shown to predict overall survival from H&E-stained images using only patient-level survival labels, with attention maps

providing interpretable insights into prognostic tumor microenvironment regions (Qaiser et al. 2022). Deep learning has been applied to quantify NAFLD/NASH progression in liver biopsies (Heinemann et al. 2022) and to predict clinicopathological features, multi-omics events, and prognosis from digital pathology in HR+/HER2- breast cancer (Hu et al. 2023). In gastrointestinal endoscopy, deep learning models have been developed and validated for gastric mucosal lesion diagnosis (Nam et al., 2022), automated prediction of the revised Vienna Classification in colonoscopy with initial external validation (Yamada et al., 2022), and endoscopic severity grading in ulcerative colitis using multicenter clinical trial data (Gutierrez Becker et al., 2021). Notably, a CNN for neoplasia diagnosis on digital cholangioscopy was developed and validated in a Latin American multicenter context, one of the few studies in the present corpus from the region (Robles-Medranda et al., 2023).

Dermatological imaging has produced extensive work on skin lesion classification (Han et al. 2020; Tiwari 2021). Critical equity concerns are now well documented: most benchmark dermoscopy datasets predominantly contain images from lighter Fitzpatrick skin types, raising serious questions about generalizability to diverse populations, including Indigenous peoples in Brazil and Latin America, whose skin phototypes span a wide range — a concern the manuscript returns to in Sections 6 and 7. Ultrasound has expanded across cardiac, obstetric, abdominal, and musculoskeletal applications (Wang et al., 2023; Putra et al., 2025; Milletari et al., 2017); ultrasound's portability, low cost, and real-time capability make AI-augmented point-of-care ultrasound a promising modality for remote and Indigenous-served settings.

A persistent challenge across all modalities is the gap between research-stage performance and clinical deployment of these technologies. Models trained on single-center data frequently degrade when deployed in new clinical environments with different patient demographics, imaging equipment and disease prevalence (Chen Q et al. 2025; Qian et al. 2023). Validation cohorts represented in the imaging-AI literature remain concentrated in high-income settings in North America, Europe, and East Asia; the populations most likely to encounter performance degradation in clinical use are also the least represented in the studies that produce the performance numbers (Yogarajan et al. 2022; Ghanem et al. 2025).

Therefore, the empirical picture is mixed in a particular way. Performance under controlled validation conditions is strong in selected tasks; external validation across diverse populations is uneven; the cohorts in which validation occurs are demographically and geographically narrow. This gap is discussed in Section 6.

6. Representational Validity, Algorithmic Bias, and the Limits of Generalization

The literature on algorithmic bias in healthcare has accumulated a vocabulary that is precise about some parts of the problem and silent about others. *Dataset bias* refers to systematic patterns of inclusion and exclusion in the data used to train and validate models. *Algorithmic bias* refers to

systematic differences in model behavior across groups, which may emerge from dataset bias, the design of objective functions, labelling practices, or interactions between these. *Fairness* refers to a family of formal criteria — equalized odds, equal opportunity, demographic parity, and disparate-impact ratios — by which group-level model behavior can be measured. *External validity* refers to whether model performance estimated on one dataset transfers to another; *dataset shift* and *population shift* describe specific failure modes (Yogarajan et al. 2022; Joseph 2025; Ghanem et al. 2025).

Vocabulary is necessary but incomplete. It treats the measurement of bias as a downstream property of an existing model. It does not name the upstream condition, without which bias measurement and external validation can only chase failures rather than prevent them. This manuscript proposes representational validity as an upstream condition. The concept is defined as follows.

Representational validity in diagnostic AI is the degree to which the populations, clinical contexts, imaging protocols, disease patterns, demographic groups, geographic settings, and sociocultural conditions used to train and validate an AI system correspond to the populations and settings in which the system will be deployed.

Representational validity is not equivalent to representativeness of a sample; it is not equivalent to external validity; it is not equivalent to fairness. This is a prior condition that determines whether external validity tests and fairness audits can be meaningfully conducted. A model whose training and validation populations are demographically narrow has low representational validity for any deployment context that lies outside that narrow envelope, regardless of the model benchmark accuracy. A model with low representational validity may still produce stable performance numbers in cohorts that resemble those used to develop it. Instability appears at deployment when the model is asked to act for populations for which it was never built.

Three categories of evidence converge on the importance of this concept: The first is the recurrent observation that external validation in mature imaging AI domains is narrow. Multicentre validation in ophthalmology has been important (Ran et al. 2019; Qian et al. 2023; Chen Q et al. 2025) but has typically extended across institutions within high-income settings, not across the full range of populations to which the systems might be deployed. The same pattern is documented across radiology, pathology, and endoscopy: models that perform strongly on benchmark cohorts have been tested on additional cohorts that resemble the original ones (Bates et al. 2021; Tricco et al. 2023; Wang T et al. 2024a).

Second, it demonstrates that algorithmic bias is measurable and present even in well-designed systems. Yogarajan et al. (2022) analyzed data and model bias in AI for healthcare in Aotearoa, New Zealand, using Māori health data, operationalizing disparate-impact, equal-opportunity, and equalized-odds metrics, and documenting evidence of both intentional algorithmic-design bias and unintentional bias inherited from pre-trained representation models. The study is methodologically significant for the present argument because it shows that tools for measuring Indigenous-specific algorithmic bias exist, have been applied successfully, and yield interpretable findings. Joseph (2025) characterises algorithmic

bias in public-health AI as a silent threat to equity in low-resource settings, arguing that models trained predominantly on high-income data systematically underperform for marginalised groups; Ghanem and colleagues (2025) reach a parallel conclusion for the Canadian public-health context.

Third, this pattern is not an accident but a structural feature of the field's current data ecology. Several recent commentaries argue that AI research systematically under-includes Indigenous and other historically marginalised populations (Lancet Digital Health 2023; Islam et al. 2024; Walter et al. 2020), that the result is a form of data colonialism in which the populations whose data are most extracted are also the populations least represented in the systems built on those data, and that the consequences for global health equity are non-trivial. The architectural standardization Fishman and Hancox-Li (2022) read critically — one architecture, one set of pre-training corpora, many downstream tasks — has the corollary that a single representational baseline becomes the universal lens through which an increasing range of clinical problems is approached. A clinically valuable model is not one that maximizes a single performance number on a single benchmark; it is one whose representational scope matches its deployment scope.

The concept of representational validity reorganizes the existing vocabulary of the field. Dataset bias is a specific manifestation of low representational validity at the data composition level. Algorithmic bias is a deployment-stage signature of low representational validity. External validity becomes a partial test of representational validity, which is useful but insufficient. Dataset shift becomes a predictable failure mode for models whose representational scope is narrower than their deployment scope. Fairness becomes a property that can only be meaningfully audited once representational validity is established; in the absence of representational validity, fairness audits chase the symptoms of a more fundamental problem.

Representational validity is a normative concept. This connects to what the manuscript calls diagnostic justice: the claim that the right to a reliable diagnosis cannot be conditioned on having been adequately represented in someone else's training data. Diagnostic justice does not require that every population be present in every model; it requires that the populations to whom a model is offered have been part of the conditions under which the model was made reliable for them or that they have been informed when they have not been. The technical reading of representational validity (a property of a model's relationship to its deployment context) and the normative reading (a condition for the legitimate clinical use of a model) are continuous rather than separable.

The next section examines the case in which the absence of representational validity is most visible:

7. Indigenous Health, Data Sovereignty, and Diagnostic AI

The corpus assembled for this review, drawn from PubMed/MEDLINE, Scopus, Web of Science, IEEE Xplore, SciELO, LILACS, and Google Scholar in English, Portuguese, and Spanish, with thematic blocks explicitly including Indigenous population terms in three languages, contains no peer-reviewed empirical study that develops, validates, or deploys a diagnostic-imaging AI system in an Indigenous-defined

population in Brazil or Latin America. The absence of these studies is structural and not a sampling artifact of the present review. This is the central finding of the gap-mapping that organizes this section.

7.1 What the absence is, and what it is not

The absence of evidence on AI in contexts adjacent to Indigenous health is not absolute. Yogarajan et al. (2022) is a rigorous empirical analysis of AI bias for Māori populations in Aotearoa New Zealand using electronic-health-record data and formal fairness metrics; it is, however, not an imaging study. Khan and colleagues (2025) propose a five-step deployment framework for AI-driven healthcare for remote and Indigenous populations in Canada, articulating infrastructure requirements, community engagement, cultural safety, and governance; it is a framework, not a single imaging study. Robles-Medranda et al. (2023) reported a multicenter CNN validation for cholangioscopy in Latin America; it is regionally significant but not Indigenous-defined. Motie-Shirazi et al. (2025) demonstrated edge-AI quality feedback on Doppler ultrasound for community midwives in remote settings; the deployment context is structurally analogous to many Indigenous-served settings, and the modality (portable ultrasound) is exactly the kind that the manuscript argues for in Section 8.

The absence of Brazilian and Latin American work on Indigenous data sovereignty is not an absence. Souza et al. (2023) analyse the CARE Principles in research-data management for Povos Indígenas in Brazil; Torino et al. (2023) propose a data-management plan specifically designed for research with Povos Indígenas; Carrasco et al. (2024) extend the CARE-Principles analysis to Indigenous archives and the right to collective memory, and the broader Indigenous Data Sovereignty literature (Walter et al. 2020) has begun to examine the specific challenges posed by generative AI (Banh and Strobel 2023). These contributions establish that the governance vocabulary for Indigenous data-governed AI is being articulated in Brazil. However, the diagnostic imaging AI literature is absent, which would be the empirical counterpart.

The absence is most precisely characterized as follows: while the methodological capacity to measure Indigenous-specific algorithmic bias exists (Yogarajan et al. 2022); while regional capacity for AI imaging validation has been demonstrated in Latin America (Robles-Medranda et al. 2023); while remote-community AI deployment using portable imaging modalities has been demonstrated (Motie-Shirazi et al. 2025; Khan et al. 2025); and while Brazilian and Latin American governance frameworks for Indigenous-data-governed research have been articulated (Souza et al. 2023; Torino et al. 2023; Carrasco et al. 2024) — the intersection of these capacities, in the form of diagnostic-imaging AI studies with Indigenous-defined populations in Brazil or Latin America, has not been realized in the published and indexed literature surveyed.

This is consistent with the broader structural inequality argument articulated in several recent studies. The Lancet Digital Health (2023) editorial on decolonizing health data argues that current AI-development pipelines reproduce extractive relationships with marginalized communities; Joseph (2025) and Ghanem et al. (2025) articulate the same pattern from public health perspectives in low-resource

and Canadian contexts; Kakraba and colleagues (2026) extend the analysis through the lens of cognitive sovereignty and decolonial public health, articulating how epistemic authority must be reclaimed by communities in the global AI era. The absence of Brazilian and Latin American Indigenous-specific diagnostic imaging AI is exactly what these analyses would predict.

7.2 Indigenous data sovereignty as a constitutive condition

The Indigenous Data Sovereignty (IDS) literature provides analytical resources to read the absence not only as a gap but also as a structural relationship. Carroll et al. (2020) articulate the CARE Principles for Indigenous Data Governance — *Collective Benefit, Authority to Control, Responsibility, Ethics* — as complementary to the FAIR principles (findable, accessible, interoperable, reusable) and as centering Indigenous rights and self-determination in data ecosystems. The CARE Principles have gained broad international endorsement and are increasingly recognized as essential for ethical AI development involving Indigenous data.

In the Canadian context, Mashford-Pringle and Pavagadhi (2020) describe the application of OCAP (Ownership, Control, Access, and Possession) and Inuit Qaujimajatuqangit (IQ) frameworks to address historical trauma in Indigenous health research, arguing in AMA Journal of Ethics that these principles must extend to AI and algorithmic systems that process Indigenous health data. In the Brazilian context, Souza et al. (2023) and Torino et al. (2023) extend the CARE-Principles framing to Povos Indígenas, and Carrasco et al. (2024) develop the analysis for Indigenous archives and the right to collective memory; the implications of generative AI for Indigenous data governance (Banh and Strobel 2023; Walter et al. 2020) remain an underexamined frontier in this literature. The Māori algorithmic sovereignty framework articulated by Brown et al. (2024) — a multi-author Māori-led contribution that includes Te Taka Keegan — operationalizes sovereignty principles for algorithmic systems, with potential adaptability to other Indigenous contexts. Ofosu-Asare (2024) proposes a participatory methodology that incorporates Indigenous epistemologies into AI development as a corrective to cognitive imperialism.

The collective claim of this literature is that Indigenous data sovereignty is not an additional ethical constraint on AI development but a different account of what data are to be considered. Data extracted from a community without that community's authority, governance, and participation are not, on this account, data about that community in any clinically meaningful sense; they are data about a particular extractive relationship. The implication for representational validity is direct: representational validity is not satisfied by the technical presence of data nominally about a population. Representational validity requires that the data be collected, governed, validated, and interpreted in relation to the population that produced them. The CARE Principles, OCAP, and algorithmic-sovereignty frameworks are the operational vocabulary for that relationship; the parallel debate on data sovereignty in genomics has reached similar conclusions for biospecimen and sequence data (Nature Machine Intelligence Editorial 2022).

7.3 The shape of an adequate Indigenous-health diagnostic-AI literature

If the present absence is a symptom, the shape of adequate literature can be specified by combining the available adjacent evidence. Such a literature would (i) develop and validate imaging-AI systems with population-representative cohorts that include Indigenous populations as defined and protagonist participants, not as silent additions to a residual category; (ii) report performance stratified by ethnicity, geography, socioeconomic condition, and imaging-equipment ecology, in keeping with the reporting standards the field has begun to articulate (Chen Q et al. 2025); (iii) embed community governance, contextual consent, and benefit-sharing into the data-production pipeline from the outset, as the CARE Principles and OCAP require (Carroll et al. 2020; Mashford-Pringle and Pavagadhi 2020; Souza et al. 2023; Torino et al. 2023); (iv) prioritise portable, low-resource imaging modalities — ultrasound, fundoscopy, dermoscopy — that are deployable in the geographies in which Indigenous communities live, building on the demonstrated feasibility of edge-AI deployment for remote-community settings (Motie-Shirazi et al. 2025; Khan et al. 2025); (v) integrate Indigenous knowledge systems as substantive components of the design process, following the participatory methodology articulated for AI more broadly (Ofosu-Asare 2024) and the community-engagement approach now being articulated for AI-driven prevention science (Haroz et al. 2026); and (vi) make the limits of any model's representational validity explicit, so that clinicians and communities can make informed decisions about whether and how to use it.

7.4 Why this matters: convergence of philosophical, political, and clinical questions

Indigenous health is not only a clinical question. It is the clearest instance in contemporary AI of the problem that recent critical AI scholarship has articulated: that AI systems are not neutral technical artifacts; that they encode the assumptions of the institutions that produced them; and that the populations rendered invisible by dataset construction become invisible to the systems built on those datasets (Pasquinelli and Joler 2020; Ofosu-Asare 2024; Fishman and Hancox-Li 2022; Svensson 2021). This is the case in which the philosophical question of formalization, the political question of data colonialism, and the clinical question of diagnostic reliability converge.

The argument can be stated as follows. Indigenous populations are statistically invisible in diagnostic imaging AI in Brazil and Latin America. This invisibility is consistent with the documented patterns of structural inequity in AI research. The available governance frameworks (CARE, OCAP, algorithmic sovereignty) provide the operational vocabulary for an alternative arrangement. The available technical adjacent evidence (Yogarajan et al. 2022; Robles-Medranda et al. 2023; Motie-Shirazi et al. 2025; Khan et al. 2025) demonstrates that this alternative is feasible. Therefore, the absence is not a research limitation; it is a research agenda whose components are present and whose integration is overdue.

The next section makes this agenda explicit.

8. Toward an Agenda for Equitable Diagnostic Intelligence

Three tensions organize the agenda. The first is between *technical capability* and *representational validity*. A model that performs well on a benchmark cohort is technically capable in a specific sense; it is representationally valid only for the cohorts for which it has been developed and validated. Treating capability as sufficient evidence of validity has been a recurring source of disappointment when systems have been deployed (Chen Q et al. 2025; Qian et al. 2023; Tricco et al. 2023). The first requirement of the agenda is that diagnostic AI evaluation become explicit about its representational scope: a model's claim to validity is bounded by the populations and settings used to develop it, and the bound should be reported as visibly as the headline accuracy.

The second tension is between automation and situated clinical judgement. Diagnostic AI is most clinically useful, and most ethically defensible, when it augments situated clinical judgement rather than displaces it (Li et al. 2024; Qaiser et al. 2022; Sheth et al. 2023). The augmentation model is harder to operationalize than the automation model; it requires careful work on workflow integration, communication of uncertainty, and how the model's outputs interact with the clinician's reasoning. However, it is the model that respects the limits of formalization set out in Section 3. The second requirement of the agenda is that the deployment design be augmentation-first, with automation reserved for tasks in which the representational validity is high and the cost of error is well understood.

The third tension is between data extraction and Indigenous Data Sovereignty. The default mode of dataset construction in machine learning treats data as a resource to be aggregated and reused; IDS frameworks treat data as a relational artifact whose collection, governance, and use are inseparable from the communities they concern (Carroll et al., 2020, 2021; Walter et al., 2020; Mashford-Pringle and Pavagadhi, 2020; Brown et al., 2024). The third requirement of the agenda is that diagnostic-AI development for Indigenous-served populations adopt the IDS vocabulary as a precondition rather than as a downstream constraint.

From these tensions, a research agenda emerges. Dataset transparency: Diagnostic-AI publications should report the demographic, geographic, institutional, and equipment ecology of their training and validation cohorts as a routine reporting requirement and should make the representational scope of the model an explicit claim (Chen Q et al. 2025; Yogarajan et al. 2022). External validation in under-represented populations: external validation should be designed in advance to include populations beyond the development context, with cohort composition justified against the deployment scope (Ran et al. 2019; Qian et al. 2023). Participatory and community-governed AI: the design, governance, and validation of imaging-AI systems intended for Indigenous-served populations should include community governance from the design stage, in accordance with CARE and OCAP (Carroll et al. 2020; Mashford-Pringle and Pavagadhi 2020; Souza et al. 2023; Torino et al. 2023). Reporting standards for population representativeness: existing reporting frameworks for clinical AI should be extended to require explicit characterisation of population representativeness; absence of diversity should be reported as a limitation rather than left implicit (Yogarajan et al. 2022; Joseph 2025; Ghanem et al. 2025). Capacity building in historically under-represented territories: meaningful equity requires that the institutional capacity to develop and validate AI exists in the territories where the populations live — a long-horizon

investment in computational infrastructure, training, and partnerships (Khan et al., 2025; Naseem et al., 2020). Prioritization of portable, low-resource imaging modalities: portable ultrasound, fundoscopy, and dermoscopy are the modalities most deployable in the geographies of Indigenous-served care; AI development for these modalities should be prioritized on equity grounds, building on the demonstrated feasibility of edge-AI in remote settings (Motie-Shirazi et al., 2025). Diagnostic-AI evaluation beyond accuracy: evaluation criteria should include representational validity, deployment-context concordance, community-governance compliance, and clinical-uncertainty communication characteristics that augmentation-mode use requires.

The agenda is intentionally modest in its individual components and ambitious in its integration into the curriculum. Each component has been previously discussed in the literature. Their absence as a coherent package — most visibly in the Indigenous-imaging-AI literature in Brazil and Latin America — is what gives the manuscript its analytical urgency.

9. Conclusion

The contribution of this review is both conceptual and evidentiary. Conceptually, it articulates representational validity in diagnostic AI as a necessary condition for clinically reliable and ethically defensible imaging. Representational validity is not equivalent to external validity, fairness, or data sovereignty taken in isolation; it is the upstream condition that determines whether any of those properties can be meaningfully obtained. It is a property of the relationship between a model, its data ecology, and its deployment context, and can be operationalized as a reporting requirement, evaluation criterion, and governance principle.

The review demonstrates that diagnostic AI has reached strong technical performance in selected tasks under controlled validation conditions; this performance does not transfer reliably to populations under-represented in training and validation; Indigenous populations in Brazil and Latin America are absent from the imaging-AI literature in a way that is structural, not accidental; and the adjacent evidence available from Aotearoa, New Zealand, Canada, the Western Pacific, and remote-community settings elsewhere both substantiates the equity concern and demonstrates that an alternative is feasible.

A diagnostic model whose representational scope does not include the population it is offered to is not merely unfair to that population in some downstream ethical sense; it is a model that is not for that population in any technically meaningful sense, and its deployment risks reproducing algorithmic harm that has long operated as institutional invisibility. The CARE Principles, OCAP, and emerging algorithmic-sovereignty frameworks are not a constraint on diagnostic AI; they are the operational vocabulary through which diagnostic AI can become valid for the populations it serves. A diagnostic AI worthy of the populations it claims to serve is one whose representational scope is made visible, whose validity claims are bounded honestly, and whose data and governance arrangements are built with the communities whose health it proposes to support.

Declarations

Author Contribution

A.M. conceived the idea for the article, designed the integrative review, performed the systematic literature search across PubMed/MEDLINE, Scopus, Web of Science, IEEE Xplore, SciELO, LILACS, and Google Scholar in English, Portuguese, and Spanish, conducted the data charting and evidence mapping across the seven thematic axes, developed the conceptual framework of representational validity in diagnostic AI, prepared Figure 1, and wrote the first draft of the manuscript. A.C.A., as senior researcher and group leader of the underlying research programme, supervised the project, contributed to the conceptual framework and to the analytical interpretation of the findings, and provided continuous critical revision of the manuscript throughout its preparation. B.F.A., I.D.S.L.B., T.A.C.C., A.P.P.R.C.P., and C.G.F.C. contributed to the screening of records, the analytical discussion of the diagnostic-imaging and Indigenous-health axes, and the critical revision of the manuscript for important intellectual content. J.M.N., V.C.P., and C.D.F.S. contributed to the analytical discussion of the public-health, epidemiological, and health-equity axes and to the critical revision of the manuscript. P.A.M. and O.V.G. contributed to the methodological discussion of the review design, to the validation of the bibliographic charting, and to the critical revision of the manuscript. P.A.S. contributed to the methodological discussion of the systematic search strategy and reporting standards and to the critical revision of the manuscript. All authors read and approved the final version of the manuscript and accept responsibility for the integrity and accuracy of its content.Initials — quick referenceInitialsAuthorA.M.Américo Mota (first author; corresponding author)B.F.A.Bruna Felix de AlmeidaI.D.S.L.B.Isabella Diniz de Souza Lins BungeJ.M.N.Jandir Mendonça NicacioV.C.P.Vanessa Cardoso PereiraT.A.C.C.Thiago Augusto Cavalcante de CarvalhoP.A.M.Paula Andreatta MaduroP.A.S.Paulo Adriano SchwingelC.D.F.S.Carlos Dornels Freire de SouzaO.V.G.Orlando Vieira GomesA.P.P.R.C.P.Ana Paula Pereira Rolim Coimbra PintoC.G.F.C.Caroline Guimarães da Fonseca ChiecoA.C.A.Anderson da Costa Armstrong (senior author)

Acknowledgement

The authors gratefully acknowledge the Truká people of Cabrobó, Pernambuco, and the Fulni-ô people of ÁguasBelas, Pernambuco, whose long-standing partnership with the underlying research programme has shaped, overmore than a decade, the questions, the methods, and the ethical commitments that inform this work. The authors areparticularly indebted to the community lideranças, the Equipes Multidisciplinares de Saúde Indígena (EMSI), andthe Conselho Distrital de Saúde Indígena (CONDISI) of the Distrito Sanitário Especial Indígena (DSEI)Pernambuco, whose institutional coordination and continuous dialogue have made this partnership possible. Whilethe present manuscript is a conceptual and methodological review and does not itself report empirical data collectedwithin these communities, the analytical orientation of the work — its commitment to representational validity,Indigenous data sovereignty, and diagnostic justice — is owed in large part to years of dialogue, fieldwork, andreciprocal learning that this partnership has sustained.The authors also thank the institutional teams at the State

University of Bahia (UNEB), the Federal University of theSão Francisco Valley (UNIVASF), the University Hospital UNIVASF (HU-UNIVASF / EBSERH), the AMMOInstitute / CDI, and the University of Pernambuco (UPE) — Petrolina Campus — for institutional and administrativesupport during the preparation of this manuscript.

References

1. Banh L, Strobel G (2023) Generative artificial intelligence. Electron Markets 33:63. https://doi.org/10.1007/s12525-023-00680-1
2. Bates DW, Levine D, Syrowatka A, Kuznetsova M, Thomas Craig KJ, Rui A, Jackson GP, Rhee K (2021) The potential of artificial intelligence to improve patient safety: a scoping review. npj Digit Med 4:54. https://doi.org/10.1038/s41746-021-00423–6
3. Brown PT, Wilson D, West K, Escott K-R, Basabas K, Ritchie B, Lucas D, Taia I, Kusabs N, Keegan TT (2024) Māori algorithmic sovereignty: idea, principles, and use. Data Sci J 23(15):1–16. https://doi.org/10.5334/dsj–2024–015
4. Carrasco LB, Vidotti SABG, Torino E (2024) Arquivos, princípios CARE e a governança de dados indígenas. Anais do VII Workshop de Informação, Dados e Tecnologia (WIDaT), Porto Velho, Brazil, 25–27 June 2024. https://doi.org/10.22477/vii.widat.194
5. Carroll SR, Akee R, Chung P, Cormack D, Kukutai T, Lovett R, Suina M, Rowe RK (2021) Indigenous Peoples' data during COVID–19: from external to internal. Front Sociol 6:617895. https://doi.org/10.3389/fsoc.2021.617895
6. Carroll SR, Garba I, Figueroa-Rodríguez OL, Holbrook J, Lovett R, Materechera S, Parsons M, Raseroka K, Rodriguez-Lonebear D, Rowe R, Sara R, Walker JD, Anderson J, Hudson M (2020) The CARE Principles for Indigenous Data Governance. Data Sci J 19(1):43. https://doi.org/10.5334/dsj–2020–043
7. Chen Q, Keenan TDL, Agron E, Allot A, Guan E, Duong B, Elsawy A, Hou B, Xue C, Bhandari S et al (2025) AI workflow, external validation, and development in eye disease diagnosis. JAMA Netw Open 8(7):e2517204. https://doi.org/10.1001/jamanetworkopen.2025.17204
8. Chen X, Wang J, Qian T, Wang J, Ding Y, Zhang S, Liao J, Cheng Q, Yang T, Mateen M et al (2025) An artificial intelligence cloud platform for OCT-based retinal anomalies screening system in real clinical environments. npj Digit Med 8:559. https://doi.org/10.1038/s41746-025-01959–7
9. De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, Askham H, Glorot X, O'Donoghue B, Visentin D et al (2018) Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med 24(9):1342–1350. https://doi.org/10.1038/s41591-018-0107–6
10. Deng L (2018) Artificial intelligence in the rising wave of deep learning: the historical path and future outlook. IEEE Signal Process Mag 35(1):177–180. https://doi.org/10.1109/MSP.2017.2762725
11. Fishman N, Hancox-Li L (2022) Should attention be all we need? The epistemic and ethical implications of unification in machine learning. In: 2022 ACM Conference on Fairness,

Accountability, and Transparency (FAccT '22). ACM, pp 1246–1260. https://doi.org/10.1145/3531146.3533206

12. Ghanem S, Moraleja M, Gravesande D, Rooney J (2025) Integrating health equity in artificial intelligence for public health in Canada: a rapid narrative review. Front Public Health 13:1524616. https://doi.org/10.3389/fpubh.2025.1524616
13. Gonçalves B (2023) Can machines think? The controversy that led to the Turing Test. AI Soc 38(6):2499–2509. https://doi.org/10.1007/s00146-021-01318–6
14. Gutierrez Becker B, Arcadu F, Thalhammer A, Gamez Serna C, Feehan O, Drawnel F, Oh YS, Prunotto M (2021) Training and deploying a deep learning model for endoscopic severity grading in ulcerative colitis using multicenter clinical trial data. Therapeutic Adv Gastrointest Endoscopy 14:1–15. https://doi.org/10.1177/2631774521990623
15. Haenlein M, Kaplan A (2019) A brief history of artificial intelligence: on the past, present, and future of artificial intelligence. Calif Manag Rev 61(4):5–14. https://doi.org/10.1177/0008125619864925
16. Han SS, Moon IJ, Kim SH, Na J-I, Kim MS, Park GH, Park I, Kim K, Lim W, Lee JH, Chang SE (2020) Assessment of deep neural networks for the diagnosis of benign and malignant skin neoplasms in comparison with dermatologists: a retrospective validation study. PLoS Med 17(11):e1003381. https://doi.org/10.1371/journal.pmed.1003381
17. Häntze H, Xu L, Mertens CJ, Dorfner FJ, Donle L, Busch F, Kader A, Ziegelmayer S, Bayerl N, Navab N, Rueckert D, Schnabel J, Aerts HJWL, Truhn D, Bamberg F, Weiss J, Schlett CL, Ringhof S, Niendorf T, Pischon T, Kauczor H-U, Nonnenmacher T, Kröncke T, Völzke H, Schulz-Menger J, Maier-Hein K, Hering A, Prokop M, van Ginneken B, Makowski MR, Adams LC, Bressem KK (2025) Segmenting whole-body MRI and CT for multiorgan anatomic structure delineation. Radiology: Artif Intell 7(6):e240777. https://doi.org/10.1148/ryai.240777
18. Haroz EE, Goklish N, Dillard A, Adams R, Bull SS, Gonzalez-Fisher RF, Valenza P, Manson SM, Thorpe RJ Jr (2026) Addressing health disparities through community engagement in artificial intelligence-driven prevention science. Prev Sci. https://doi.org/10.1007/s11121-025-01863–2
19. Heinemann F, Gross P, Zeveleva S, Hu SQ, Hill J, Höfer A, Jonigk D, Diehl AM, Abdelmalek M, Lenter MC, Pullen SS, Guarnieri P, Stierstorfer B (2022) Deep learning-based quantification of NAFLD/NASH progression in human liver biopsies. Sci Rep 12:19236. https://doi.org/10.1038/s41598-022-23905–3
20. Hu J, Lv H, Zhao S, Lin CJ, Su GH, Shao ZM (2023) Prediction of clinicopathological features, multi-omics events and prognosis based on digital pathology and deep learning in HR+/HER2 – breast cancer. J Thorac Disease 15(5):2528–2543. https://doi.org/10.21037/jtd–23–445
21. Hurst S, Dhein K, Yracheta J, Mackey T (2026) Sovereignty in the digital age: Indigenous perspectives on health data and emerging technologies. SSM – Qualitative Res Health 9:100698. https://doi.org/10.1016/j.ssmqr.2025.100698
22. Islam M, Valiani AA, Datta R, Chowdhury M, Turin TC (2024) Ethical and equitable digital health research: ensuring self-determination in data governance for racialized communities. Camb Q

Healthc Ethics 33:1–11. https://doi.org/10.1017/S096318012400015X

23. Joseph J (2025) Algorithmic bias in public health AI: a silent threat to equity in low-resource settings. Front Public Health 13:1643180. https://doi.org/10.3389/fpubh.2025.1643180
24. Kakraba S, Agyemang EF, Srivastav SK (2026) Cognitive sovereignty and decolonial public health: reclaiming epistemic authority in the global AI era. Front Public Health 14:1785170. https://doi.org/10.3389/fpubh.2026.1785170
25. Khan A, Galea S, Mendez I (2025) Five steps for the deployment of artificial intelligence-driven healthcare delivery for remote and Indigenous populations in Canada. Digit Health 11:1–6. https://doi.org/10.1177/20552076251334422
26. Li B, Chen H, Yu W, Zhang M, Lu F, Ma J, Hao Y, Li X, Hu B, Shen L, Mao J, He X, Wang H, Ding D, Li X, Chen Y (2024) The performance of a deep learning system in assisting junior ophthalmologists in diagnosing 13 major fundus diseases: a prospective multicenter clinical trial. npj Digit Med 7:8. https://doi.org/10.1038/s41746-023-00991–9
27. Liu Z, Tong L, Chen L, Jiang Z, Zhou F, Zhang Q, Zhang X, Jin Y, Zhou H (2022) Deep learning based brain tumor segmentation: a survey. Complex Intell Syst 9:1001–1026. https://doi.org/10.1007/s40747-022-00815–5
28. Luger GF, Chakrabarti C (2017) From Alan Turing to modern AI: practical solutions and an implicit epistemic stance. AI Soc 32:321–338. https://doi.org/10.1007/s00146-016-0646–7
29. Mashford-Pringle A, Pavagadhi K (2020) Using OCAP and IQ as frameworks to address a history of trauma in Indigenous health research. AMA J Ethics 22(10):E868–E873. https://doi.org/10.1001/amajethics.2020.868
30. McCarthy J, Minsky ML, Rochester N, Shannon CE (1955) A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. AI Magazine 27(4):12–14 (reprint, 2006. https://doi.org/10.1609/aimag.v27i4.1904
31. McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 5:115–133. https://doi.org/10.1007/BF02478259
32. Milletari F, Ahmadi S-A, Kroll C, Plate A, Rozanski V, Maiostre J, Levin J, Dietrich O, Ertl-Wagner B, Bötzel K, Navab N (2017) Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound. Comput Vis Image Underst 164:92–102. https://doi.org/10.1016/j.cviu.2017.04.002
33. Moazemi S, Vahdati S, Li J, Kalkhoff S, Castano LJV, Dewitz B, Bibo R, Sabouniaghdam P, Tootooni MS, Bundschuh RA, Lichtenberg A, Aubin H, Schmid F (2023) Artificial intelligence for clinical decision support for monitoring patients in cardiovascular ICUs: a systematic review. Front Med 10:1109411. https://doi.org/10.3389/fmed.2023.1109411
34. Morano J, Fazekas B, Sükei E, Fecso R, Emre T, Gumpinger M, Faustmann G, Oghbaie M, Schmidt-Erfurth U, Bogunović H (2025) Multimodal foundation model and benchmark for comprehensive retinal OCT image analysis. npj Digit Med 8:576. https://doi.org/10.1038/s41746-025-01852–3
35. Motie-Shirazi M, Nikookar S, Ahmad M, Rafiei A, Sameni R, Rohloff P, Clifford GD, Katebi N (2025) Real-time quality feedback on Doppler data for community midwives using edge-AI. Mach Learning:

Health 1:015014. https://doi.org/10.1088/3049–477X/ae1bad

36. Moulaei K, Yadegari A, Baharestani M, Farzanbakhsh S, Sabet B, Afrash MR (2024) Generative artificial intelligence in healthcare: a scoping review on benefits, challenges and applications. Int J Med Informatics 188:105474. https://doi.org/10.1016/j.ijmedinf.2024.105474
37. Mundlamuri R, Gunnam GR, Mysari NK, Pujuri J (2025) The evolution of AI: from classical machine learning to modern large language models. IEEE Access 13. https://doi.org/10.1109/access.2025.3621344
38. Nadin M (2019) Machine intelligence: a chimera. AI & Society. https://doi.org/10.1007/s00146-018-0842–8
39. Nam JY, Chung HJ, Choi KS, Lee H, Kim TJ, Soh H, Kang EA, Cho S-J, Ye JC, Im JP, Kim SG, Kim JS, Chung H, Lee J-H (2022) Deep learning model for diagnosing gastric mucosal lesions using endoscopic images: development, validation, and method comparison. Gastrointest Endosc 95(2):258–268e10. https://doi.org/10.1016/j.gie.2021.08.022
40. Naseem M, Akhund R, Arshad H, Ibrahim MT (2020) Exploring the potential of artificial intelligence and machine learning to combat COVID–19 and existing opportunities for LMIC: a scoping review. J Prim Care Community Health 11:1–11. https://doi.org/10.1177/2150132720963634
41. Nature Machine Intelligence Editorial (2022) Data sovereignty in genomics and medical research. Nat Mach Intell 4:905–906. https://doi.org/10.1038/s42256-022-00578–1
42. Ofosu-Asare Y (2024) Cognitive imperialism in artificial intelligence: counteracting bias with Indigenous epistemologies. AI Soc 40:3045–3061. https://doi.org/10.1007/s00146-024-02065–0
43. Pasquinelli M, Joler V (2020) The Nooscope manifested: AI as instrument of knowledge extractivism. AI Soc 36:1263–1280. https://doi.org/10.1007/s00146-020-01097–6
44. Pollack JB (1989) Connectionism: past, present, and future. Artif Intell Rev 3(1):3–20. https://doi.org/10.1007/BF00139193
45. Putra RA, Pradipta GA, Wulaning Ayu PD (2025) Gallbladder disease classification from ultrasound images using CNN feature extraction and machine learning optimization. J Electron Electromedical Eng Med Inf 7(4):1089–1111. https://doi.org/10.35882/jeeemi.v7i4.1030
46. Qaiser T, Lee C-Y, Vandenberghe M, Yeh J, Gavrielides MA, Hipp J, Scott M, Reischl J (2022) Usability of deep learning and H&E images predict disease outcome-emerging tool to optimize clinical trials. npj Precision Oncol 6:37. https://doi.org/10.1038/s41698-022-00275–7
47. Qian X, Xian S, Su Y, Guo W, Liu H, Xi X, Chu C, Yin Y, Yu S, Ma K, Cheng M, Qu Y (2023) External validation of a deep learning detection system for glaucomatous optic neuropathy: a real-world multicentre study. Eye 37:3813–3818. https://doi.org/10.1038/s41433-023-02622–9
48. Raffy P, Pambrun J-F, Kumar A, Dubois D, Patti JW, Cairns RA, Young R (2023) Deep learning body region classification of MRI and CT examinations. J Digit Imaging 36:1291–1301. https://doi.org/10.1007/s10278-022-00767–9
49. Raiaan MAK, Mukta MSH, Fatema K, Fahad NM, Sakib S, Mim MMJ, Ahmad J, Ali ME, Azam S (2024) A review on large language models: architectures, applications, taxonomies, open issues and

challenges. IEEE Access 12:26839. https://doi.org/10.1109/access.2024.3365742

50. Ran AR, Cheung CY, Wang X, Chen H, Luo L-y, Chan PP, Wong MOM, Chang RT, Mannil SS, Young AL, Yung H-w, Pang CP, Heng P-A, Tham CC (2019) Detection of glaucomatous optic neuropathy with spectral-domain optical coherence tomography: a retrospective training and validation deep-learning analysis. Lancet Digit Health 1(4):e172–e182. https://doi.org/10.1016/S2589–7500(19)30085–8
51. Robles-Medranda C, Baquerizo-Burgos J, Alcívar-Vásquez J, Kahaleh M, Raijman I, Kunda R, Puga-Tejada M, Egas-Izquierdo M, Arévalo-Mora M, Méndez JC, Tyberg A, Sarkar A, Shahid H et al (2023) Artificial intelligence for diagnosing neoplasia on digital cholangioscopy: development and multicenter validation of a convolutional neural network model. Endoscopy 55:719–727. https://doi.org/10.1055/a–2034–3803
52. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323(6088):533–536. https://doi.org/10.1038/323533a0
53. Schneider J (2022) Foundation models in brief: a historical, socio-technical focus. https://doi.org/10.48550/arXiv.2212.08967. arXiv:2212.08967 [cs.AI]
54. Searle JR (1980) Minds, brains, and programs. Behav Brain Sci 3(3):417–424. https://doi.org/10.1017/S0140525X00005756
55. Sharma M, Savage C, Nair M, Larsson I, Svedberg P, Nygren JM (2022) Artificial intelligence applications in health care practice: scoping review. J Med Internet Res 24(10):e40238. https://doi.org/10.2196/40238
56. Sheth A, Roy K, Gaur M (2023) Neurosymbolic artificial intelligence (why, what, and how). IEEE Intell Syst 38(3):56–62. https://doi.org/10.1109/MIS.2023.3268724
57. Singh M, Kumar A, Khanna NN, Laird JR, Nicolaides A, Faa G, Johri AM, Mantella LE, Fernandes JFE, Teji JS, Singh N, Fouda MM, Singh R, Sharma A, Kitas G, Rathore V, Singh IM, Tadepalli K, Al-Maini M, Isenovic ER, Chaturvedi S, Garg D, Paraskevas KI, Mikhailidis DP, Viswanathan V, Kalra MK, Ruzsa Z, Saba L, Laine AF, Bhatt DL, Suri JS (2024) Artificial intelligence for cardiovascular disease risk assessment in personalised framework: a scoping review. EClinicalMedicine 73:102660. https://doi.org/10.1016/j.eclinm.2024.102660
58. Souza LP, Sousa RSC, Löw MM, Barros THB (2023) Promovendo a justiça epistêmica: uma análise dos princípios CARE na gestão de dados de pesquisa em relação aos povos indígenas. Anais do VI Workshop de Informação, Dados e Tecnologia (WIDaT). 1–11. https://doi.org/10.22477/vi.widat.73
59. Spasic I, Nenadic G (2020) Clinical text data in machine learning: systematic review. JMIR Med Inf 8(3):e17984. https://doi.org/10.2196/17984
60. Svensson J (2021) Artificial intelligence is an oxymoron: the importance of an organic body when facing unknown situations as they unfold in the present moment. AI Soc 38:363–372. https://doi.org/10.1007/s00146-021-01311-z
61. The Lancet Digital Health (2023) Decolonising health data. Lancet Digit Health 5(8):e477. https://doi.org/10.1016/S2589–7500(23)00132–2

62. Tiwari S (2021) Dermatoscopy using multilayer perceptron, convolution neural network, and capsule network to differentiate malignant melanoma from benign nevus. Int J Healthc Inform Syst Inf 16(3):58–73. https://doi.org/10.4018/IJHISI.20210701.oa4
63. Torino E, Monteiro ECSA, Vidotti SABG (2023) Plano de gestão de dados de pesquisa de povos indígenas: considerações acerca dos princípios FAIR e CARE. Revista Brasileira de Preservação Digit 4:1–23. https://doi.org/10.20396/rebpred.v4i00.17936
64. Tricco AC, Hezam A, Parker A, Nincic V, Harris C, Fennelly O, Thomas SM, Ghassemi M, McGowan J, Paprica PA, Straus SE (2023) Implemented machine learning tools to inform decision-making for patient care in hospital settings: a scoping review. BMJ Open 13(2):e065845. https://doi.org/10.1136/bmjopen–2022–065845
65. Turing AM (1950) Computing machinery and intelligence. Mind 59(236):433–460. https://doi.org/10.1093/mind/LIX.236.433
66. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. In: Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA. arXiv:1706.03762v7 [cs.CL]. https://doi.org/10.48550/arXiv.1706.03762. NeurIPS proceedings: https://papers.nips.cc/paper/7181-attention-is-all-you-need
67. Villanueva-Miranda I, Xiao G, Xie Y (2025) Artificial intelligence in early warning systems for infectious disease surveillance: a systematic review. Front Public Health 13:1609615. https://doi.org/10.3389/fpubh.2025.1609615
68. Walter M, Kukutai T, Carroll SR, Rodriguez-Lonebear D (eds) (2020) Indigenous data sovereignty and policy. Routledge, Abingdon. https://doi.org/10.4324/9780429273957
69. Wang C, Zhang J, Liu S (2023) Medical ultrasound image segmentation with deep learning models. IEEE Access 11:10158. https://doi.org/10.1109/access.2022.3225101
70. Wang T-W, Hong J-S, Lee W-K, Lin Y-H, Yang H-C, Lee C-C, Chen H-C, Wu H-M, You WC, Wu Y-T (2024a) Performance of convolutional neural network models in meningioma segmentation in magnetic resonance imaging: a systematic review and meta-analysis. Neuroinformatics 23:14. https://doi.org/10.1007/s12021-024-09704–3
71. Wang Z, Chu Z, Doan TV, Ni S, Yang M, Zhang W (2024b) History, development, and principles of large language models: an introductory survey. AI Ethics 5(3):1955–1971. https://doi.org/10.1007/s43681-024-00583–7
72. Watego K, Brodhurst-Hill M, Murray C, Mangoyana C, Emery A, Woon J, Williams K, Reid N, Engstrom T, Nelson C, Sullivan C, Setchell J (2025) Indigenous people's perspectives on sharing health data for service delivery purposes: an inquiry using Indigenous methodologies. Lancet Reg Health – Western Pac 65:101753. https://doi.org/10.1016/j.lanwpc.2025.101753
73. Yamada M, Shino R, Kondo H, Yamada S, Takamaru H, Sakamoto T, Bhandari P, Imaoka H, Kuchiba A, Shibata T, Saito R (

Figures

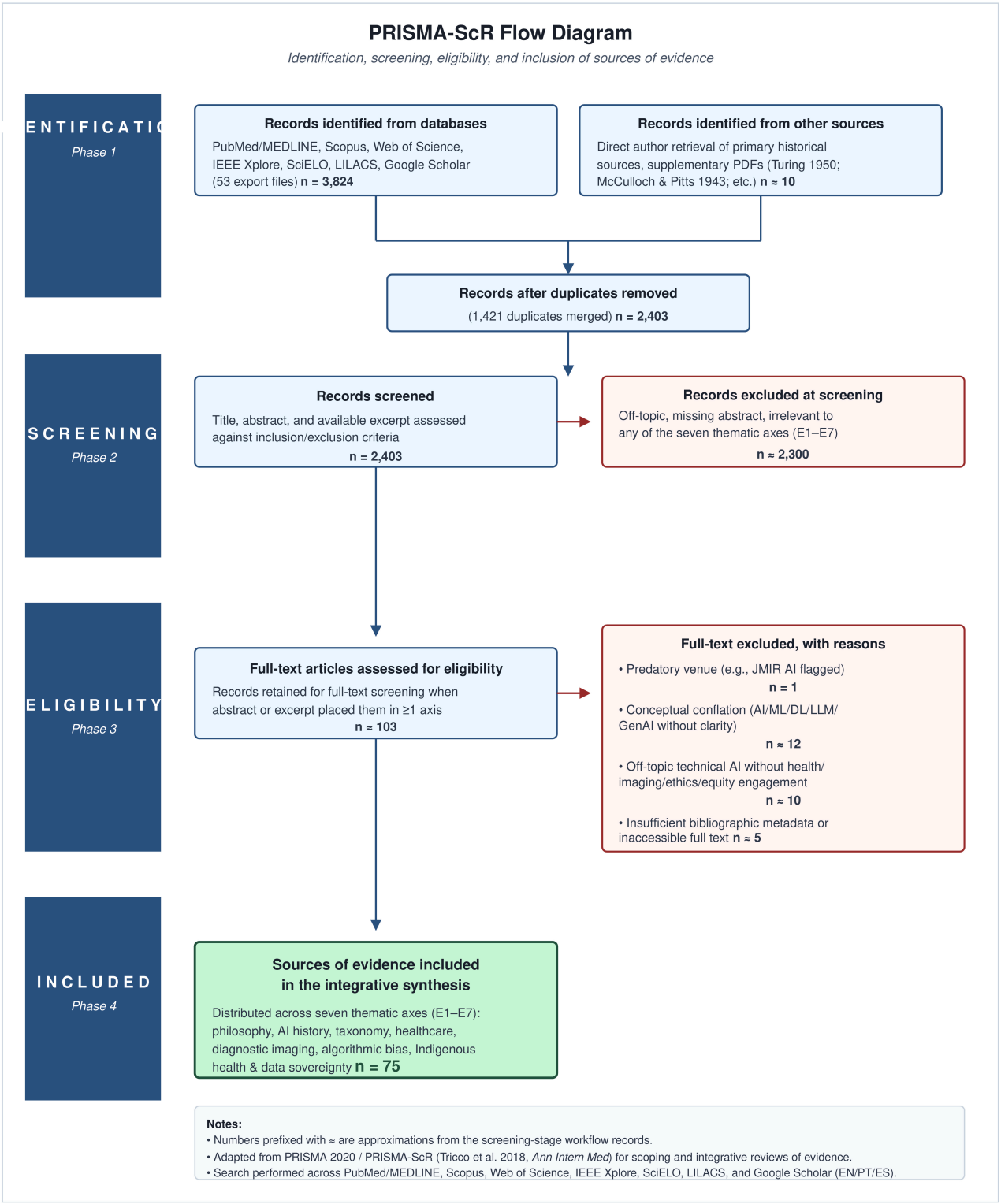


Figure 1

PRISMA-ScR flow diagram summarising the identification, screening, eligibility, and inclusion stages of the integrative review. Records were identified through structured searches of PubMed/MEDLINE, Scopus, Web of Science, IEEE Xplore, SciELO, LILACS, and Google Scholar (n = 3,824), supplemented by a small number of records retrieved directly by the authors for primary historical sources (n ≈ 10). After deduplication, 2,403 unique records were screened; approximately 103 were assessed for full-text

eligibility; 75 sources were included in the final integrative synthesis across the seven thematic axes (E1–E7). The diagram is adapted from the PRISMA 2020 statement and its PRISMA Extension for Scoping Reviews (PRISMA-ScR).