FIND
Diagnosis for all

IMPLEMENTATION TOOLKIT V.1.0 FOR

AI-based Diagnostics

Acronyms/abbreviations

AI	Artificial intelligence
CAD	Computer-aided detection
CXR	Chest X-ray
TB	Tuberculosis
WHO	World Health Organization

Table of contents

Executive summary

Countries are aiming to adopt artificial intelligence (AI)-based tools to help address diagnostic gaps. We offer approaches to common challenges and draw attention to some key considerations that will help streamline the implementation and maximize the benefits of these tools.

Adopting AI-based diagnostic tools demands substantial resources and careful planning to avoid burdening healthcare systems without delivering value. Key considerations in planning include clear use case identification, infrastructure development, and capacity building as foundational steps for successful implementation and scaling.

Data are foundational for training, calibrating, and evaluating AI solutions in healthcare. Data requirements for AI in health are unique in their scale, nature, quality, diversity and confidentiality, necessitating specialized approaches for data collection and management. Key considerations include data attributes (e.g. metadata, annotation, versioning), ethical aspects, and ensuring security and privacy.

Independent evaluations are essential for building trust, ensuring compliance, and validating the performance, safety, and usability of AI solutions before and after deployment. These assessments identify gaps overlooked in vendor-led evaluations, foster transparency, and support successful adoption of the tools. Key evaluation metrics include performance measures, software quality, interpretability, and post-market surveillance.

AI tools require post-deployment optimization to calibrate for specific deployment settings. Frequent updates, driven by technological advancements and model improvements, necessitate ongoing monitoring and evaluation, which are less critical in typical software deployments.

Given the complexity of AI tools, investment in and procurement of the solutions require careful attention to factors such as regulatory approval, health system integration, and economic feasibility to ensure their long-term sustainability.

Introduction

Many countries are dealing with a shortage of trained professionals to deal with diagnostic gaps. As such, they are looking to adopt artificial intelligence (AI)-based diagnostic tools to help address these shortfalls. However, there are challenges in the adoption of AI tools due to lack of trust in the AI outputs, as well as inadequate awareness of how to select and evaluate these tools.

Additional challenges include the need for datasets to develop and validate context-appropriate tools, as well as potential pitfalls during and after deployment of the tools. In this document, we highlight key factors for implementers and innovators to consider when targeting AI-based diagnostic innovations and approaches to mitigate challenges in implementing AI-based diagnostics tools.

We draw on discussions with stakeholders, in-country workshops and experience with evidence generation and validation of chest X-ray computer-aided detection (CXR-CAD) tools, as well as 20+ years of evidence generation on diagnostics for screening and triage. Recognizing the context-dependent complexities in global health, we offer guidance on the adoption of AI-based diagnostics in this setting, without overly prescriptive solutions.

Scope

This toolkit provides guidance for the deployment and scaling of AI-based screening and diagnostic interventions in low-resource settings. The document focuses on the following key areas:

- Planning for the selection, implementation and scale-up of AI-based diagnostics across low-resource healthcare settings;
- High-level best practices for the creation and management of locally representative datasets;
- High-level guidance on the evaluation of AI-based diagnostics;
- Frameworks for post-deployment optimization, monitoring and evaluation;
- Criteria to inform procurement of and investment in AI-based diagnostic solutions.

Purpose

This document aims to bridge the gap between AI-based technology solutions and healthcare outcomes, by offering actionable insights and guidelines.

We draw on lessons distilled from shared experiences validating CXR-CAD tools globally, as well as discussions with in-country governments, non-government agencies, providers, regulators, developers, and partners implementing and regulating AI for diagnostics. This document offers considerations on the key workstreams essential for successful and sustainable deployments of AI-based diagnostic tools. These include planning, dataset creation, evaluation, procurement and deployment, and post-deployment activities. These workstreams are interconnected and often proceed in parallel. Where appropriate, we provide specific examples from experiences with CXR-CAD tools for screening and triage.

Target audience

This document is intended for stakeholders involved in the implementation, regulation, and scaling of AI-based diagnostic solutions, including:

GOVERNMENT MINISTRIES AND REGULATORS

To support the development of policies, standards, and regulatory frameworks for AI in diagnostics.

HEALTHCARE PROVIDERS AND IMPLEMENTERS

To provide guidance on data collection and the procurement and practical deployment and integration of AI tools within healthcare systems.

AI DEVELOPERS AND TECHNOLOGY PARTNERS

To align innovation efforts with regulatory requirements and real-world implementation needs.

GLOBAL AND LOCAL DEVELOPMENT PARTNERS

To coordinate and enhance efforts in supporting countries’ adoption and scaling of AI solutions.

INTERNATIONAL ORGANIZATIONS AND DONORS

To coordinate and enhance efforts in supporting countries’ adoption and scaling of AI solutions.

Guiding principles

Our work is grounded in the following principles:

INNOVATION IN AI-BASED DIAGNOSTICS SHOULD BE EQUITABLE, with accompanying measures to safeguard societal wellbeing, including incentives to prioritize transparency and incorporate safety and quality fail-safes.

REGULATIONS FOR AI-BASED DIAGNOSTICS MUST BE FLEXIBLE, allowing adaptation to the rapidly evolving nature of the landscape.

AI-BASED DIAGNOSTICS SHOULD BE INDEPENDENTLY ASSESSED to evaluate their performance and suitability for use in intended settings and populations.

DEPLOYMENT OF AI-BASED DIAGNOSTICS MUST TAKE A SYSTEMIC APPROACH, considering the holistic systems into which these tools will be integrated, and with careful consideration of potential downstream effects.

BUILDING TRUST IN AI-POWERED DECISION-MAKING IS ESSENTIAL for their acceptability and adoption, and is underpinned by principles of transparency and privacy.

Workstream considerations

The development and deployment of AI-based tools for diagnostics requires unique considerations, distinct from those for in vitro diagnostics or other software solutions. These must be accounted for by policy-makers, implementers, regulators and other stakeholders to enhance the likelihood of successful implementations. The following sections highlight key considerations and best practices, many of which may be missed during the planning and implementation of solutions. Where appropriate, we provide specific examples from experience, particularly with evaluating CXR-CAD tools.

Planning for the selection, implementation and scale-up of AI diagnostics

Adopting AI-based diagnostic tools requires significant resource investment. Without adequate planning, the deployment and scaling of these tools can add strain to already stretched healthcare systems without contributing significant value. Aligned with the Principles of Digital Development that FIND endorses **(1)**, the following sections outline key areas for consideration and foundational steps for selecting, implementing and scaling AI solutions.

Use case identification

Defining the tool's use case is a crucial step in the development of AI-based diagnostic solutions. This involves identifying the target users, such as healthcare providers, and understanding the specific needs of the population of interest, for example, whether they are asymptomatic or symptomatic. It also requires careful consideration of the medical settings in which the tools will be deployed to ensure alignment with broader healthcare goals and effective integration into existing workflows.

It is important to consider the following points:

- Before identifying potential AI-based tools, it is crucial to first define the use case to determine whether AI can effectively address the issue or if other, simpler solutions might suffice.
- Define the intended use of the tool by engaging with end users, healthcare providers, policy-makers, and funders early in the process to ensure the solution aligns with broader healthcare goals and receives the necessary support for its effective implementation.
- Outline measurable outcomes for the use case (e.g. reduction in diagnostic delays, improved screening sensitivity, or cost savings) to guide tool evaluation and track impact. The use case and expected outcomes will define subsequent steps like tool selection and evaluation criteria, the type of data that need to be collected and upskilling requirements.
- The use case may need to be refined based on feedback from pilot implementations or stakeholder consultations. An iterative process is key to aligning with real-world needs.

EXAMPLES

The following sections offer some practical examples and considerations for defining use cases in the context of CXR-CAD tools, highlighting the nuances between screening and triaging applications. These insights are derived from extensive experience in evaluating CXR-CAD tools, with an emphasis on how they can be tailored to meet the distinct needs of various medical settings:

- CXR-CAD tools can be used for triaging or screening of tuberculosis (TB), COVID-19 and other diseases. For screening, we are trying to identify individuals with a particular condition, often from a population that is asymptomatic or at risk of the condition, to refer them for further diagnostic testing. The focus for this use case is on early detection to improve outcomes, so tools with high sensitivity are preferred to minimize false negative results (for example, people with COVID-19 who receive a negative test result), ensuring fewer missed cases. Lower specificity can be tolerated, i.e. some false positives (for example, people without COVID-19 who receive a positive test result) to ensure thorough coverage.
- For the triaging use case, we are trying to prioritize symptomatic individuals who need immediate attention based on likelihood of a condition. This is relevant for cases where individuals already present with symptoms, often in settings with constrained resources. For this use case, we need tools with a balance of sensitivity and specificity to ensure critical cases are prioritized, to best manage workflow and reduce the burden on the healthcare system.
- As screening focuses on identifying potential cases in asymptomatic or at-risk populations, validation data are required from diverse, asymptomatic populations to effectively assess sensitivity. Triaging, on the other hand, requires evaluation data from symptomatic populations that represent emergency settings. In both cases, similar tools might be used, but the choice of tool and configuration is determined by the specific use case.

Infrastructure

AI tools require foundational digital infrastructure such as computers, robust networks and power. The planning process should ensure that existing and planned investments in foundational infrastructure, such as digital public infrastructure[1] and national digital health strategies (2, 3), are taken into account when selecting tools to function effectively in the given settings.

These include the following key components.

COMPUTING CAPACITIES

Computationally intensive AI models might require advanced hardware such as graphics processing units and scalable storage. Some vendors offer solutions that can be run in the cloud and some offer solutions that can be run locally. In the latter setting, the resource utilization of the software needs to be determined, and adequate compatible local hardware resources might need to be procured accordingly.

CONNECTIVITY

While many AI vendors offer cloud-based platforms that can be accessed from anywhere, a seamless workflow demands reliable and secure network connectivity with low latency and high throughput so large files such as images or videos can be efficiently uploaded. It is crucial to assess requirements for each tool's use case and assess network performance metrics thoroughly before the tools are deployed to ensure they meet the necessary requirements.

ELECTRICITY

The stability of power is a foundational requirement for both computing and networking infrastructure. Solutions such as battery-powered systems, possibly supplemented by solar energy, can offer resilience against power interruptions.

INTEROPERABILITY

It is vital to ensure the compatibility of diagnostic machines that produce outputs with the AI solution under consideration. Efforts should be made to standardize data and their format, ideally following FAIR principles **(4)** and standards such as Fast Healthcare Interoperability Resources.

DIGITAL DIVIDE

AI-based solutions should be designed and deployed so they are accessible and usable by the whole target population. It is important to ensure that the deployment of such solutions doesn't create or exacerbate existing digital divides between people (e.g. by excluding those who don't have access to certain types of technology).

[1] Digital public infrastructure can be defined as basic digital capabilities that are essential to deliver economic opportunities and social services to the public.

EXAMPLES

- To integrate AI tools in diagnostics, specifically CXR-CAD and ultrasound, it is crucial to consider digital infrastructure, computing capacities, connectivity, power stability, and interoperability.

- For CXR-CAD, some vendors provide portable X-ray machines with computer hardware and software packaged in a backpack as a single solution. Some CXR-CAD vendors offer smartphone applications that can use photographs from analogue X-rays against a lightbox as input to the system. These applications offer feedback on the quality of the images prior to uploading to the AI engine.

- Handheld ultrasounds that can be used for full-body scans are commercially available. However, portability sometimes comes at the cost of performance or speed, and selection should be made based on available resources and acceptable trade-offs for the deployment setting.

Capacity building

Training and upskilling are especially critical for AI in diagnostics compared with traditional software or non-software diagnostics solutions. This is because AI introduces unique complexities and challenges at every stage of its lifecycle, requiring specialized skills and understanding.

Capacity building should focus on the following points.

INTERPRETATION

Unlike traditional software that produce deterministic outputs, AI models often work probabilistically, generating results with inherent uncertainty. End-users need training to interpret these outputs correctly and on how to properly set thresholds.

LIMITATIONS

AI models can have biases or limitations based on the data they are trained on. Training stakeholders helps them recognize and mitigate these potential pitfalls.

LOCALIZATION

AI solutions often need to be adapted to local contexts, such as specific disease prevalence, demographic variations, or resource constraints. Building capacity ensures that local teams can fine-tune, validate, and deploy solutions effectively.

TRUST

Training end-users of the system will engender trust and empower the teams to take ownership of the solutions, enabling them to make adjustments as needed.

EVALUATIONS

The evaluators and regulators should be sensitized to the characteristics of the tools such as intended use, performance metrics, data requirements, integration needs, as well as nuances of evaluation and global best practices in assessment.

Best practices for the creation and management of locally representative datasets

Data are foundational to the training, calibration and evaluation of AI solutions. Data requirements for AI in health are unique in their scale, nature, quality, diversity and confidentiality, necessitating specialized approaches for data collection and management. These requirements and best practices are described in this section.

Data requirements

Establishing representative and high-quality datasets involves following several essential requirements, as detailed here.

DATA COLLECTION

Data collection should ideally be guided by the use case ensuring it represents a wide range of scenarios, demographics, and conditions to avoid bias and ensure generalizability across different contexts. This applies to data used for training, calibration and evaluation.

METADATA

Capturing metadata, such as the device used, time of sample collection and patient demographic details, enables data organization, analysis and the creation of meaningful dataset cohorts.

ANNOTATION

Data used for training or validation require high-quality annotations. This should be factored in upfront, as annotation can be costly and time-consuming. Disagreement among annotators can introduce variability, necessitating processes like consensus annotation to resolve issues.

DATA QUALITY

Ensuring high data quality is essential for AI development. Key considerations include completeness, accuracy, consistency, timeliness, validity, and integrity. High-quality data ensures robust, reliable AI models that perform well for the intended use.

DATA VERSIONING

In real-world settings, AI models must be regularly updated and evaluated to remain relevant as input data characteristics change. Proper versioning of datasets is crucial to track changes and ensure reproducibility.

STANDARDIZATION

Data collection should ensure that data are easily reusable, following FAIR principles and standards such Fast Healthcare Interoperability Resources.

EXAMPLE

- For CXR-CAD, portable CXRs might produce lower-quality images such as when there is poor lighting, patient position is suboptimal, or due to lower resolution of the machines. Datasets should include data from these settings to determine how a tool might perform under these conditions. To train and evaluate CXR-CAD tools used in TB screening, corresponding reference standard data for both symptomatic and asymptomatic cases are needed. However, in using retrospective data, reference data may not have been collected for asymptomatic cases by default, so data might need to be collected afresh to train and evaluate these tools. This example underscores the importance of planning data collection based on the specific target use case for the AI based tool.

Ethical considerations

As AI systems become more prevalent and influential in decision-making processes, addressing the ethical dimensions of data collection, processing, and deployment is crucial. While the World Health Organization (WHO) provides a comprehensive overview of considerations for AI (5), this section outlines some key issues specific to dataset management that must be navigated to ensure that AI technologies are developed and utilized in a manner that is fair, transparent, and respectful of individual rights.

DIVERSITY AND INCLUSIVITY

It is essential for datasets to be diverse and inclusive to represent all relevant populations independent of race, gender, age or socioeconomic status, in order to prevent biases that might lead to unfair or discriminatory outcomes. For example, a tool trained on data from one demographic subgroup might inadvertently discriminate against another demographic not adequately represented in the data.

EXAMPLE

- When evaluating CXR-CAD products at FIND, we have found there can be considerable variation in the performance characteristics of tools based on the geographic region and gender of the cohort used for evaluation. This is likely because of biases in the training sets used for training.

INFORMED CONSENT

Patients should be fully informed about the potential use of their data in the development and evaluation of AI tools and provided with the option to opt out of having their data used for this purpose.

AUDIT TRAIL

It is essential to maintain clear documentation of data sources, collection methods, and preprocessing steps to ensure transparency. Audit mechanisms should be put in place to track how data are used to address potential misuse or ethical breaches.

Security and privacy

In the context of AI, safeguarding data security and privacy is of utmost importance. As AI systems increasingly rely on frequent access to vast amounts of sensitive personal information, such as health records and biometric data, the potential risks associated with data breaches and misuse escalate.

This section addresses the critical measures and best practices that must be implemented to protect data integrity and maintain user trust.

DATA PROTECTION

AI relies on large datasets, often containing sensitive personal information, like health records or biometric data. As such, robust data protection using techniques such as encryption and de-identification are essential to minimize the risk of and potential damage from data breaches.

DATA INTEGRITY

AI systems are susceptible to unique threats, such as adversarial inputs (i.e. those designed to trick the system into performing an unintended action) and data poisoning (i.e. deliberate manipulation of training data to compromise performance), which can manipulate model outputs or compromise training data. These risks require specialized security measures, such as training with adversarial data, data integrity checks and post-deployment monitoring.

DATA OWNERSHIP

Laws like the General Data Protection Regulation and California Consumer Privacy Act allow for individuals' right to withdraw consent to have their data used. Revocation of individual patient data from a dataset can be disruptive to ongoing training (e.g. models might need to be retrained) and evaluation unless processes are created to facilitate this upfront.

DATA PUBLICATION

The integrity of evaluation data needs to be maintained, so these data are not used for training or fine-tuning, as it may bias validation of the tools.

DATA STORAGE

Depending on local regulations, health-related data might have to be stored on servers located within the country. When setting up the data infrastructure for an AI-based health solution, it is important to check for local regulations and align the storage and processing infrastructure accordingly. This might mean relying on local cloud providers, instead of the typically used large international providers.

EXAMPLE

- During evaluation of CXR-CAD software at FIND, vendors often seek to push updates to their models. This requires granting vendors access to the processing environment that has been exposed to validation data. To ensure that the validation data remain secure and inaccessible to vendors, FIND has had to implement techniques to snapshot our infrastructure, allowing us to revert it to a state that predates any exposure to validation data.

Evaluation criteria and framework

The evaluation of solutions by independent organizations, both prior to deployment and on an ongoing basis, plays a critical role in building trust among providers and policy makers, ensuring the software meets established criteria, and increasing the likelihood of successful adoption. Independent evaluations provide an unbiased assessment of performance, safety, and usability, helping identify gaps that may not be apparent in vendor-led evaluations. They also serve as a mechanism for building trust, ensuring regulatory compliance, validating real-world utility, and fostering transparency. In addition to independent evaluation, it is also necessary to perform local evaluations prior to deployment and on an ongoing basis, to customize, validate and fine-tune models for local contexts **(6)**. Some quantitative and qualitative metrics for evaluation are outlined in this section.

Performance measures

Performance measures are critical in assessing the effectiveness and reliability of AI diagnostic tools within their defined use cases. Key factors that play a pivotal role in properly measuring AI diagnostic tool performance are as follows.

- It is important to clearly identify primary and secondary performance measures for evaluating the AI diagnostic tool. This is usually based on the defined use case(s).
- For each measure, it is necessary to establish and benchmark against a reference standard to determine if the tool meets the quality level for deployment in a given setting. It is important to spend time on properly defining the reference standard.
- It is essential to define the sample size and data collection strategy for evaluation in a particular setting.

EXAMPLE

- For CXR-CAD, as per WHO guidelines **(7)**, the sensitivity and specificity of the tool are compared with that of a trained radiologist using reference standards for TB. This performance measure exists as it is expected that these tools might be used in the absence of radiologists.

Software quality

AI-enabled software is likely to need context-specific adaptations to be effective. These issues are particularly important for AI software, given the complexity and opaqueness of solutions. Ensuring suitable software quality requires a comprehensive approach that encompasses all stages of the development and deployment lifecycle. Prioritizing software quality during evaluation significantly enhances the likelihood of successful deployments.

Key factors to consider are as follows.

USABILITY

User interfaces should ensure ease of use, with ongoing user feedback loops to identify and mitigate usability gaps.

ACCESSIBILITY

Systems should be designed to accommodate diverse user needs, including for people with disabilities, by leveraging features such as screen reader compatibility, customizable interfaces, and alternative interaction methods where appropriate. By aligning with international standards such as Web Content Accessibility Guidelines **(8)**, the system can be made accessible to a wide range of users and promote equitable access to AI-powered solutions.

SCALABILITY

Systems should be designed to efficiently handle increasing workloads. Performance should be regularly monitored under high demand to address scalability challenges.

EXTENSIBILITY

Systems should easily incorporate additional functionalities, using modular design principles to facilitate updates and allow for extensibility (e.g. the addition of new capabilities or functionalities).

LOCALIZATION

Software should be adapted to meet regional language, cultural, and regulatory requirements, with localization testing to identify and correct regional disparities.

RESOURCE UTILIZATION

Software should be optimized to use system resources efficiently, employing monitoring tools to detect and mitigate any resource bottlenecks.

MAINTAINABILITY AND SUPPORT

Clear documentation and support channels should be established to ensure easy maintenance, with regular training for technical teams to address maintainability challenges.

TESTING

Comprehensive testing strategies should be implemented to ensure that models remain performant and to quickly identify and mitigate quality gaps.

INTEROPERABILITY

AI-based tools should seamlessly integrate with existing healthcare system infrastructure.

Interpretability

AI tools are probabilistic in nature and results are often provided in the form of scores or binary output. It is therefore essential that AI tool outputs are interpretable, to provide clinicians with confidence in these tools and help them identify biases and errors. Ensuring the interpretability of tools can also help clinicians determine how the tools arrived at their conclusions.

EXAMPLE

- For CXR-CAD, some vendors provide bounding boxes or heatmaps to identify key artifacts that can help clinicians understand the basis for scores or outputs.

Post-market surveillance

- A framework should be put in place for ongoing evaluation of the tool's performance after deployment, to detect issues such as data drift or errors due to model updates.
- Evaluating software updates requires a different approach compared with fresh installations, as updates must ensure compatibility with existing systems, preserve data integrity, and address potential issues introduced by changes without disrupting ongoing workflows.

Post-deployment optimization, monitoring and evaluation

AI tools need to undergo calibration to optimize performance for the setting within which they are deployed. Additionally, these tools tend to undergo frequent updates due to the changing nature of the technology and constant improvement of models with new training data. These factors necessitate additional considerations, which are less critical for typical software deployments.

Post-deployment optimization and local calibration

- As it is difficult to train general models for very specific settings, after deployment it is prudent to run an initial pilot with test data during which phase the model can be fine-tuned with local data.
- AI tools generate numerical scores to indicate the likelihood of diseases or artifacts. A threshold is a predefined cut-off value on these scores that helps classify conditions (for example as high-risk or low-risk), guiding decisions and actions. These thresholds must be tailored to the use case, considering factors such as disease prevalence and healthcare system capacity. Due to variability in vendor scoring, thresholds must be configured on a case-by-case basis.
- A major challenge in CAD implementation is the considerable effort and potential for error involved in establishing an appropriate threshold score through local studies. The choice of threshold is critical and must be aligned with the tool's intended use. For instance, a higher sensitivity might be prioritized for screening purposes to ensure no cases are missed, even at the cost of reduced specificity, which differs from the balance sought in other applications.

EXAMPLE

- **Many CXR-CAD software do not come with pre-set manufacturer recommended threshold settings that define abnormal CXRs. Though some manufacturers may recommend threshold settings, there can be considerable variations in recommendations. Local calibration of CAD thresholds by operators of the tools is essential to maximize their performance (9).**

Ongoing monitoring

Data drift is the change in model input data over time, causing discrepancies between the training data and new data, which can degrade model performance. Factors like changing user behaviour, seasonal effects, or technological advancements might contribute to this drift. Addressing data drift necessitates ongoing monitoring and, when necessary, model updates or retraining with current data to ensure sustained accuracy and applicability. Implementing such monitoring mechanisms prior to deployment is essential for maintaining the long-term reliability of AI-based CAD tools. It is important to monitor both input and output data to identify any distribution changes and address them as soon as possible.

Criteria for investment and procurement

The above considerations focus on technical and operational dimensions of the solution. However, broader considerations are essential for investment and procurement to ensure the solution's practicality, sustainability, and real-world impact.

Selection

CRITERIA

After deciding on a valid use case, selection criteria should be defined for tools based on technical and non-technical considerations. These considerations include performance metrics, software quality, budget availability, infrastructure and resource needs for deployment, scalability and regulatory compliance.

EVALUATION

How these selection criteria might be evaluated should be defined, particularly if in-country regulatory bodies do not have capacity or the maturity to effectively evaluate the tools. For example, certifications and independent evaluations can be used to validate advertised performance characteristics, or it may be necessary to run exhaustive pilots prior to deployment.

EXAMPLES

- Obtaining United States Food and Drug Administration certification for TB-specific CAD tools can be challenging due to the low prevalence of TB in the United States. However, low-resource countries with a higher burden of TB may lack the robust regulatory frameworks necessary to certify these tools. In such cases, certifications from international organizations, like WHO, could be considered as a replacement, particularly if they can be supplemented with local pilot studies.

- WHO is now inviting manufacturers to submit CAD products for an assessment of performance by an external expert group **(10)**. This will involve an independent evaluation of the tools, using a digital X-ray library hosted by FIND, as well as the submission of regulatory and marketing documentation. Following the assessment, WHO will issue an updated list of products that demonstrate adequate accuracy for use in TB screening and triage by Member States and implementing partners. The list of products will be determined by the technical expert group evaluation and based on the benchmarks set by the initial 2021 recommendation **(10)**.

Integration with the healthcare system

To truly drive the adoption of AI-powered solutions in healthcare, these solutions must integrate with existing workflows and systems, to avoid placing additional burden on the healthcare system. In line with the aforementioned Principles for Digital Development **(1)**, endorsed by FIND, digital interventions should be designed to enhance frontline health worker delivery without increasing the workload of providers.

Many digital tools unintentionally create inefficiencies by duplicating tasks or disregarding existing infrastructure and processes. To ensure meaningful adoption, AI solutions should align with clinicians' routine practices, minimizing disruptions and enabling smooth integration into daily operations. This will facilitate better acceptance and adoption of tools by healthcare professionals and enable full utilization of the technology's potential. In addition, AI tools must be interoperable with existing data systems and support bidirectional flow of data, allowing input data. For example, the tool should be able to process imaging data and patient history from the electronic health record system or picture archiving and communication system. Results from the tool should also be fed back into the electronic health record system to streamline decision making and avoid data silos.

Regulatory framework

In many countries, regulatory processes for AI-enabled diagnostics are still evolving **(11, 12)**. To aid responsible adoption of these tools, international bodies like WHO have released an expression of interest to CXR-CAD manufacturers to get their tools evaluated (10). Additionally, independent evaluators, such as FIND and some universities, also conduct performance analyses of these tools. In countries where regulations are still evolving, it might be necessary to leverage external certifications and evaluations with appropriate caveats, for example:

- Where appropriate, regulatory authorities in low-resource countries can leverage certifications from Stringent Regulatory Authorities or international bodies like the WHO to fast-track approvals. This approach can be supplemented with localized testing and pilot programmes to validate the tool's performance under local conditions and specific use cases.
- AI technologies evolve rapidly, with frequent updates to models and software. Regulatory frameworks must address this dynamic nature by establishing mechanisms for continuous oversight, including streamlined processes for approving updates, monitoring post-market performance, and ensuring tools remain safe, effective, and aligned with their intended use over time.

Cost-effectiveness and economic viability

The upfront costs of adopting AI-based diagnostic tools need to be fully accounted for. Examples of upfront costs are those associated with infrastructure setup, the customization of AI based tools to local needs, and seamless integration into existing healthcare systems. Other key costs include the purchase of necessary hardware, software licenses, and professional services for implementation. In addition, ongoing costs related to the operation and maintenance of the AI-based diagnostic tool need to be evaluated. This encompasses regular software updates, system support, personnel training, and any necessary hardware upgrades to maintain optimal performance and security standards.

In addition to evaluating the costs of AI-based diagnostic tools, it is essential to compare them with non-AI-based solutions to determine if AI-based tools are truly the most cost-effective option. AI is not a one-size-fits-all solution, and for example, training and hiring more radiologists may prove to be more cost-effective than implementing a CXR-CAD solution in some settings.

The AI-based diagnostic tool should be also compatible with local healthcare reimbursement models, funding opportunities, and grant mechanisms. Understanding the financial landscape, including insurance coverage and government support for innovative healthcare technologies, can significantly influence the tool's economic viability and adoption rate.

References

1. Principles for Digital Development. 2024 [Available from: https://digitalprinciples.org/.

2. Foundation BMG. What is digital public infrastructure? [Available from: https://www.gatesfoundation.org/ideas/digital-public-infrastructure.

3. United Nations Development Programme. Digital Public Infrastructure [Available from: https://www.undp.org/digital/digital-public-infrastructure.

4. GO FAIR. FAIR Principles [Available from: https://www.go-fair.org/fair-principles/.

5. World Health Organization. Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models 2024 [Available from: https://www.who.int/publications/i/item/9789240084759.

6. Youssef A, Pencina M, Thakur A, Zhu T, Clifton D, Shah NH. External validation of AI models in health should be replaced with recurring local validation. Nat Med. 2023;29(11):2686-7.

7. World Health Organization. WHO consolidated guidelines on tuberculosis: module 2: screening: systematic screening for tuberculosis disease 2021 [Available from: https://www.who.int/publications/i/item/9789240022676.

8. Web Accessibility Initiative. WCAG 2 Overview 2024 [Available from: https://www.w3.org/WAI/standards-guidelines/wcag/.

9. TDR. Determining the local calibration of computer-assisted detection (CAD) thresholds and other parameters 2021 [Available from: https://tdr.who.int/publications/i/item/determining-the-local-calibration-of-computer-assisted-detection-(cad)-thresholds-and-other-parameters.

10. World Health Organization. Call for expression of interest by manufacturers of software for computer-aided detection of tuberculosis (CAD) to submit products for WHO expert assessment 2024 [Available from: https://www.who.int/news-room/articles-detail/call-for-expression-of-interest-by-manufacturers-of-software-for-computer-aided-detection-of-tuberculosis-(cad)-to-submit-products-for-who-expert-assessment.

11. U.S. Food and Drug Administration. Good Machine Learning Practice for Medical Device Development: Guiding Principles 2021 [Available from: https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles.

12. Forum IMDR. Machine Learning-enabled Medical Devices—A subset of Artificial Intelligence-enabled Medical Devices: Key Terms and Definitions 2021 [Available from: https://www.imdrf.org/sites/default/files/2021-10/Machine%20Learning-enabled%20Medical%20Devices%20-%20A%20subset%20of%20Artificial%20Intelligence-enabled%20Medical%20Devices%20-%20Key%20Terms%20and%20Definitions.pdf.

Acknowledgements

The development of this toolkit was made possible through our partnership with the Patrick J. McGovern Foundation.

FIND
Diagnosis for all

GLOBAL HEALTH CAMPUS
CHEMIN DU POMMIER 40
1218 GRAND-SACONNEX
SWITZERLAND

www.finddx.org