THE FOUNDATIONAL DOCUMENTS SERIES · Vol. I ¦ No. III ¦ November 2024

FOUNDATIONAL PUBLICATION

Issued by the Founding Council of the International Society of Medical AI (ISMAI)

Charter of Ethical Principles in Medical AI

Safeguarding Human Dignity, Clinical Integrity, and Responsible Innovation

ISMAI

International Society of Medical AI

Machinam Medicorum

Published by ISMAI · www.ismai.org

Preamble

Medicine has long rested upon a sacred covenant: to safeguard human life, first, to *do no harm*, and to nurture an enduring bond of trust between healer and patient. Throughout centuries of evolution, guided by the Hippocratic ethos and shaped by the broader arc of ethical progress, physicians, nurses, and allied professionals have stewarded this covenant, placing patient welfare at the forefront of every decision. Now, we have arrived at a momentous juncture. The ascent of Artificial Intelligence in health care promises extraordinary advancements in diagnosis, treatment, and the equitable delivery of care, while also imposing upon us a profound moral responsibility.

No algorithm, however remarkable, can replicate the empathy, judgment, and accountability that define the human clinician. Rather, technology must serve as no more than a conscientious partner, enriching clinical insights and efficiency while preserving the soul of the physician-patient relationship. AI is a tool: potent, yes, but ultimately subordinate to human purpose and duty. It falls to us, then, as architects of the design of AI and practitioners of its use, to ensure that AI remains firmly rooted in the moral soil of medicine, enhancing rather than eroding the bedrock values of compassion, autonomy, and justice that have guided the practice of medicine for millennia.

This new era demands that we govern our tools rather than succumb to them. It is our collective task to align AI with the same moral commitments that have guided the healing arts, ensuring that safety, beneficence, and respect for persons remain non-negotiable standards. Clinicians must remain the ultimate authorities for patient care, exercising critical oversight of AI-generated recommendations and safeguarding the trust patients place in them. Regulators, institutions, and developers alike share in this responsibility, working together to put AI in practice in ways that elevate patient outcomes without diminishing human dignity or degrading clinical authority or responsibility.

Accordingly, the International Society for Medical Artificial Intelligence (ISMAI) proclaims this Charter of Ethical Principles. Grounded in the inherited wisdom of the healing professions and inspired by the global endeavour to uphold human rights, this Charter sets a course for ethical progress in medical AI. It reaffirms that, while technology may expand our horizons, the final measure of success remains our fidelity to the patient's well-being. By adopting these principles, we carry the tradition of medical ethics forward into the age of AI, resolute in our conviction that true innovation cannot merely accelerate the pace of change—it must preserve, and indeed enrich, the sanctity of care.

I Preserve Clinical Autonomy

"No AI system or algorithm shall supersede the independent professional judgment of licensed healthcare providers, who retain full authority to accept, override, or reject any AI-derived recommendation whenever doing so is in the best interest of the patient."

Clinical autonomy in healthcare embodies the long-standing ethic that physicians and allied professionals, rather than automated systems, bear ultimate responsibility for patient well-being. This principle is rooted in the Hippocratic tradition and reinforced by international standards, including the World Health Organization's guidance on protecting autonomy in AI deployment. Clinical autonomy recognises that health professionals integrate deep medical expertise, empathy, and situational awareness; qualities no algorithm can fully replicate.

Clinicians remain legally accountable for their decisions, even when they incorporate AI recommendations. Insurance frameworks and institutional policies that force providers to blindly follow AI outputs erode trust, shifting liability in unclear and unpredictable ways. Ethically, mandating algorithmic compliance undermines the centuries-old social contract under which individual practitioners remain personally responsible for the care they render.

In practice, this principle demands that any AI tool be framed as an advisory mechanism, not prescriptive authority. Clinicians must retain the right to deviate from algorithmic guidance, whether due to a patient's unique presentation, emergent circumstances, or personal values. Healthcare institutions and insurers should align their policies to support, rather than penalise, such professional judgment. For example, if AI suggests intervening aggressively, but the provider deems it clinically inappropriate or contrary to patient preferences, the provider must feel empowered to override AI's recommendation.

In this way, ensuring clinical autonomy will preserve the central human element in medicine and safeguards the trust patients place in their clinicians. This principle reminds us that AI is meant only to assist—not to replace—the well-honed science and the art that underpin ethical patient care.

2 Ensure Patient Safety

"All AI tools in healthcare must be rigorously validated, continuously monitored, and designed to minimise harm. No system can be introduced into clinical care absent compelling evidence of both its safety and its efficacy."

Non-maleficence, the injunction to "first, do no harm," has anchored medicine throughout history. This principle demands that emerging AI technologies uphold the core ethic of non-maleficence through robust testing before deployment and through ongoing surveillance. Medical AI is typically regulated under frameworks akin to medical devices or pharmaceuticals, requiring evidence of clinical benefit through trials or real-world performance assessments. When AI enters clinical workflows before it has proven itself through robust testing, especially in high-stakes domains like ICU triage or surgical navigation, patients can suffer from misdiagnoses and harmful treatments, and trust in healthcare also suffers.

From a legal standpoint, regulatory bodies such as the U.S. FDA and the European Commission require that high-risk AI undergo standardised evaluations. These might include multi-site validation studies, compliance

Published by ISMAI · www.ismai.org

audits, and risk assessments. Ethically, patient safety requires that any AI system have clear protocols for detecting failures. This includes designing transparent channels through which clinicians can report anomalies (e.g., patterns of overdiagnoses, missed pathologies) and can systematically refine the algorithm.

Ensuring AI's non-maleficence in practice means adopting "safety by design." For example, diagnostic AI might be configured with fail-safes so that it halts or flags uncertain predictions, rather than presenting them as conclusive. Post-market surveillance then becomes essential: if performance metrics drift or if patient populations change, immediate corrective action (such as software updates or product recalls) can be taken.

Ultimately, non-maleficence in AI is a commitment to continuous monitoring. Institutions must integrate these systems thoughtfully, mindful that innovation without safeguards can expose patients to undisclosed, even unknown, risks. Developers, clinicians, and regulators collectively are responsible for ensuring that technology complements—rather than compromises—the clinician's oath to "first, do no harm."

3 Protect Data Privacy

"Patient data used in AI systems shall be collected, stored, and analysed with the utmost care, ensuring confidentiality, security, and explicit patient-informed consent wherever feasible, thereby upholding trust and the rights of the individual patient."

Healthcare data are profoundly sensitive, capable of revealing both medical diagnoses and intimate personal details. In an era in which AI algorithms thrive on large data sets, the need for robust data governance grows exponentially. Privacy and data protection regulations, exemplified by the GDPR in the European Union and HIPAA in the United States, offer a legal baseline. Yet, ethically, healthcare institutions and AI developers must uphold an even higher standard of stewardship, reflecting their moral obligation to respect not only patient privacy, but patient autonomy and dignity.

Practically, this entails de-identifying data whenever possible, employing strong encryption, and adhering to "minimum necessary use" principles, so that superfluous personal identifiers are not exposed. Clear data-sharing agreements and rigorous technical safeguards (including secure servers, restricted access, and privacy-preserving machine learning techniques) help prevent breaches and unauthorised exploitation of patient data. Institutions should also integrate data protection officers, ethics boards, and review committees into the AI lifecycle to ensure continuous oversight.

Another core responsibility is transparency in how and why data is collected. Patients should be aware that their scans, lab results, or electronic records are feeding an algorithm. If individuals wish to opt out, systems should ideally accommodate this preference unless overriding public health needs exist (e.g., epidemic surveillance within legal frameworks). It is also vital to ensure that any secondary use of data—like algorithmic training beyond immediate care—is ethically and legally authorised.

Steadfast adherence to privacy principles protects both individuals and the public interest. Data misuse erodes the public trust, undermines support for the adoption of beneficial AI, and harms vulnerable communities disproportionately. Thus, strong data stewardship forms the backbone of ethically sound healthcare innovation, enabling

AI to flourish without sacrificing the inviolable confidentiality owed to each patient.

4 Maintain Transparency and Explainability

"The involvement of any, and all, AI-driven processes in healthcare must be disclosed to both clinicians and their patients, who must also be provided with comprehensible rationales for key AI outputs, and which processes shall remain open to auditing, ensuring that decision-making about patient care is transparent."

Trust in AI-mediated care can flourish only when users understand both the presence of and rationale for an algorithm's suggestions. This principle aligns with universal calls for transparency, highlighted by bodies like the WHO and legislation like the EU AI Act, which call for clarity around design, validation, and performance. For the use of AI to be ethically sound, patients should know when, and how, AI is shaping their care, whether it is flagging abnormal radiology findings or suggesting a treatment plan. Similarly, only with sufficient information about AI's strengths, limitations, and error rates can clinicians integrate AI's outputs into their clinical decision-making responsibly.

Legal implications of transparency arise when patients or practitioners are misled or inadequately informed. For instance, a hospital that deploys a triage algorithm without disclosing its use to its patients or staff risks liability if harm arises from the algorithm's unrecognised biases or errors. Ethically, the "right to explanation" demands that critical medical decisions, such as a recommended chemotherapy regimen, cannot hinge on inscrutable statistical models. Even if fully revealing AI's source code is impractical, offering interpretable summaries or highlighting salient factors can enable clinicians to verify or contest questionable outputs from AI.

In practice, transparency and explainability entail a system design that brings relevant features to the surface for a given case. Diagnostic AI for mammograms, for instance, might visually highlight suspicious regions on the image and, at the same time, clarify the basis for its suspicion of malignancy. Hospitals and regulatory bodies should require logs or "audit trails" that capture which data were considered and how a recommendation arose. This fosters accountability when things go wrong and, importantly, supports continuous quality improvement.

Ultimately, transparency is not merely a legal formality but a cornerstone of shared decision-making. By disclosing AI usage and providing comprehensible justifications for following, or ignoring, its recommendations and conclusions, healthcare systems empower clinicians and patients alike to make informed choices, preserve autonomy, and safeguard the integrity of care decisions.

5 Define and Enforce Accountability

"Any and all use of AI systems in healthcare requires human oversight, with clinicians, developers, and institutions sharing responsibility for patient outcomes and retaining the authority to halt or revise any algorithmic action."

Medical AI often risks creating a "responsibility gap," in which blame for errors shifts between clinicians, vendors, and administrative frameworks. This principle of oversight ensures no such gap can take root: final accountability rests with human stakeholders, especially when patient well-being is at stake. It echoes global regulatory approaches, such as the EU's "high-risk AI" classification, that demand risk management, transparency, and human oversight of medical algorithms.

Legally, the chain of accountability can be complex, but must be delineated. A clinician employing AI must remain mindful that licensing and malpractice standards remain. Institutions deploying AI must remain responsible for confirming that the tool is safe, relevant for their patient population, and well-integrated into clinical workflows. Developers, in turn, are accountable for any latent flaws, biases, or misleading marketing claims that compromise patient care.

In day-to-day practice, "human-in-the-loop" models offer an operational safeguard. A radiologist reviewing AI-marked images does not automatically cede final interpretation to the system. Likewise, a hospital might have a specialised AI governance committee to periodically audit the tool's performance, examine near-miss events, and address reported errors. Shared accountability prevents overly simplistic assumptions that "the AI is always correct," reinforcing the need for ongoing professional vigilance.

From an ethical perspective, accountability honours the principle that patients deserve individualised care from identifiable, responsible individuals or entities. Without clear oversight, patients could be subjected to opaque, unreviewable decisions; an untenable scenario, especially in life-critical decisions like drug dosing or resource allocation. Traceability further strengthens accountability: all AI outputs should be recorded, so that when an adverse outcome occurs, stakeholders can analyse the data inputs and algorithmic reasoning. This fosters a continuous feedback loop, driving improvements and building trust.

By affirming responsibility at every level (developer, institution, and clinician) this principle ensures that advanced algorithms remain firmly tethered to the human duty of care.

6 Adapt Informed Consent

"Patients and participants have the right to know when, and how, AI is informing their diagnosis or treatment, the right to know the nature of AI's role, and the right to be given the option to seek human-driven alternatives or second opinions whenever feasible."

Medical ethics have long treated informed consent as a bedrock principle: patients must comprehend the risks, benefits, and alternatives of any proposed intervention. AI complicates this dynamic when patients are unaware of algorithmic involvement in their care, and because standard consent forms typically do not address machine learning processes. Ethically, failing to disclose AI's role undermines the patient's autonomy and capacity for meaningful engagement in their care and for informed consent.

Legally, emerging frameworks increasingly view the use of AI as material information that must be disclosed. For instance, data protection laws in many jurisdictions oblige practitioners to inform patients if automated systems significantly shape their decisions. Ethically, being transparent about AI's presence and limitations respects the patient's right to question or refuse such technology. Although refusing algorithmic assistance may be impractical in some settings (like a fully digital radiology suite), giving patients at least a conceptual understanding anyway promotes trust.

In practice, updated consent processes or disclaimers can explain that "both an AI-assisted diagnostic system and human radiologists review your X-rays," or "a predictive model will help guide your medication dosing." Patients who are uncomfortable can inquire about alternative approaches or request human review. Especially for high-stakes decisions, clinicians should be prepared to articulate AI's known accuracy, scope of validation, and relevant biases or constraints.

Such honesty need not impede adoption; many patients welcome innovation if it is transparently and safely applied. By clarifying that the ultimate decision remains a clinician's responsibility (and that patients can decline AI-influenced recommendations), healthcare systems reinforce patient dignity and uphold the moral imperative of informed choice. This reciprocity fosters shared decision-making, in which AI complements the healing relationship, rather than obscuring it.

7 Safeguard Equity

"All healthcare AI initiatives must proactively prevent algorithmic bias, promote fair access, and uphold the ideal that no individual or community shall receive inferior care due to systemic inequities or discriminatory models."

Justice in healthcare demands equitable treatment and equitable distribution of resources, a demand only magnified by AI's ability to impact patient populations on large scales. When training data lack demographic diversity by omitting minority ethnicities, older patients, or particular socioeconomic backgrounds, algorithms risk perpetuating and even amplifying disparities. If, for instance, a risk stratification model systematically underestimates disease severity in disadvantaged groups, these patients may lose access to necessary interventions.

Legally, many jurisdictions have anti-discrimination laws that extend to automated processes, prohibiting AI systems from biased decision-making. From an ethical standpoint, the principle of equity resonates with global declarations that healthcare is a fundamental right. Institutions deploying AI must therefore conduct "bias audits," verifying that predictive tools maintain comparable accuracy across relevant subgroups. Developers can incorporate balanced datasets and re-weight model training to minimise skewed outcomes.

For practical implementation, "health equity by design" compels AI developers and healthcare administrators to consider how to extend benefits to underserved regions and populations, ensuring, for example, that telemedicine AI can run on low-bandwidth connections or that interfaces are localised into multiple languages. If certain communities historically lack robust data, making extra effort during data collection or taking a cautious approach to algorithmic recommendations may be warranted.

An ongoing challenge is cost: advanced AI can be expensive, potentially limiting it to wealthier health systems. Ethical practice insists that breakthroughs are also channelled to low-resource settings, possibly through tiered licensing or philanthropic support. Ensuring that bias corrections and that context-specific validations are part of an AI rollout are essential to fairness. Ultimately, achieving equity requires more than mere technical calibration: it requires acknowledging historical inequities and deliberately crafting solutions that benefit all, preventing the emergence of digital and diagnostic divides.

8 Apply Equal Standards in Underserved Settings

"AI deployments in underserved communities must meet the same safety and ethical standards as those in well-resourced contexts, must be tailored to local realities, and must never involve exploitative practices. Experimental deployments must be subject to strict oversight, ethical review, and equitable benefit-sharing."

One of AI's most promising virtues is its potential to expand access to healthcare in remote or under-resourced regions. Automated analysis of X-rays or point-of-care ultrasounds can bring specialist-level insight to communities without radiologists or consultants. However, heedless deployment of these tools (often by foreign organisations or companies) risks "data colonialism," in which vulnerable populations become testing grounds for unproven technologies, reaping little benefit but bearing disproportionate risk.

Legally and ethically, any pilot or commercial AI endeavour in underserved settings must conform to international research ethics standards, ensuring informed consent, local ethical review, and an equitable distribution of benefits. Institutions or governments hosting AI should require robust safety data, local validation studies, and clear accountability frameworks before AI is integrated into patient care. If the technology fails or injures patients, these communities must not be left without legal and clinical recourse.

Tailoring AI to local epidemiology is also critical. A model trained predominantly on urban hospital data in high-income countries might not detect diseases prevalent in rural Africa or Asia, thus undermining its clinical utility. Ethically, providing a "low-fidelity" version of advanced AI (one that is poorly tested or lacks essential safeguards) constitutes an unjust double standard. Instead, developers could prioritise "lightweight" yet rigorously validated AI tools, adapted for the reality of intermittent power or spotty internet connections.

Ensuring sustainability matters, too. If local healthcare workers cannot maintain or troubleshoot the system, it risks abandonment once its external backers leave. Hence, training local staff, investing in technology transfer, and establishing supportive infrastructure remain vital. Transparent community engagement is equally important: local perspectives on data usage, technology acceptance, and cultural norms should shape AI's design and its deployment strategy.

Ultimately, responsibly harnessing AI in low-resource environments can reduce health disparities. This principle enshrines the obligation to do so ethically: safeguarding communities from exploitation, respecting cultural contexts, and ensuring AI's benefits flow to all who need them, not just the well-served or the well-resourced.

9 Prevent Misuse and Dual Use

"Medical AI shall not be weaponised, repurposed for harmful surveillance, or diverted to unethical ends. All stakeholders must implement safeguards to prevent the exploitation of healthcare data or algorithms for malicious purposes."

Technological advances often carry the latent threat of "dual use": a beneficial innovation that can be twisted to destructive ends. In medicine, the possibility of turning diagnostic algorithms into invasive surveillance tools or using AI research to design chemical or biological weapons is not merely speculative. Prior cases have shown how easily a drug-discovery engine can be rerouted to generate toxic compounds in silico.

Ethically, medical professionals and developers uphold a code of conduct that forbids harming patients or populations. This moral norm is embodied in the Hippocratic Oath, bioethics conventions, and humanitarian laws—none of which condone harnessing healthcare techniques for oppression or warfare. Consequently, AI firms, researchers, and institutions have a heightened responsibility to restrict access once they detect suspicious or unauthorised usage. For instance, licensing agreements should prohibit using medical imaging AI for facial recognition or mass profiling. Export controls may also be necessary when dealing with high-risk AI technologies.

In practice, safeguarding against misuse involves vigilant risk assessments. If an AI system is capable of identifying genetic predispositions, could it facilitate discriminatory insurance policies or unethical eugenics programs? Robust accountability structures and whistleblowing channels can deter such abuses. Developers must train staff to watch for anomalies indicating that an AI's API or dataset is being exploited for non-medical ends.

Ultimately, shielding medical AI from perversion is indispensable to preserving public trust and preventing grave societal harm. By affirming that medical tools exist solely for healing and well-being, we sustain the moral centre of healthcare in an era of unprecedented technological power.

10 Respect Clinical Judgment

"No matter how advanced AI becomes, the professional judgment, wisdom, and experiential knowledge of healthcare providers must remain integral to clinical decisions, shaping and tempering algorithmic suggestions."

The art of medicine resides both in data analysis and in the clinician's capacity to synthesise medical knowledge, patient history, and empathic understanding. This principle reaffirms that AI insights, whether diagnosing tumours on CT scans or recommending medication dosages, cannot replace the nuanced evaluation that only human providers can offer. This principle builds on clinical autonomy while highlighting the complementary nature of professional judgment.

Legally and ethically, ignoring a qualified physician's or nurse's discretion contradicts the spirit of malpractice statutes that hold practitioners, rather than machines, responsible for adverse outcomes. Moreover, healthcare professionals, through their patient-facing role, can incorporate intangible factors like psychosocial context, cultural preferences, and patient-specific values, which no algorithm can fully capture. A borderline decision might be swayed by intangible cues (such as patient anxiety, home situation, or co-morbidities) that AI cannot observe.

In clinical practice, respecting professional judgment requires that, when AI suggests an aggressive intervention based on population-level data, a physician can adapt that recommendation to the individual patient's case. Similarly, nursing staff might sense that a patient is deteriorating in a way not captured by the algorithm's numerical readouts. By encouraging providers to question, confirm, and override AI guidance when appropriate, we preserve the synergy between computational efficiency and human insight.

This synergy fosters mutual reinforcement: AI systems can alert busy clinicians to overlooked findings, while human oversight addresses contextual subtleties that the machine cannot interpret. The result is a higher standard of care. In short, the best outcomes arise where AI's systematic strengths merge with the seasoned acumen of health professionals, never reducing them to mere overseers or reporters of automated processes.

11 Align with Evidence-Based Practice

"All AI interventions must align with established clinical guidelines, be subject to rigorous evaluation for safety and efficacy, and evolve alongside new medical evidence without contradicting recognised standards of care."

Evidence-based medicine (EBM) has transformed modern healthcare by demanding robust clinical validation. AI must similarly demonstrate that it improves (or at least does not degrade) standard patient outcomes. Regulators, such as the FDA and EMA, often classify medical AI under device regulations that require clinical trials, retrospective validation, and post-marketing surveillance.

Ethically, adherence to EBM ensures that AI is integrated prudently, preserving the trust that patients and professionals place in established guidelines. Technology should not disrupt proven protocols without compelling evidence; for instance, newly minted AI that suggests unconventional dosing for hypertension cannot be permitted to bypass the recognised standard of care unless peer-reviewed data and expert consensus support doing so. If AI evolves through adaptive learning, it should do so only under monitored conditions, to prevent "algorithmic drift" from undermining safe practice.

In routine use, healthcare providers should appraise AI outputs as they do drug recommendations or clinical trial findings—through critical assessments of reliability, applicability, and relevance. The principles of EBM also mandate ongoing updates: if new research lowers the recommended blood pressure target, AI managing hypertension must be updated accordingly. Failing to keep pace with clinical knowledge can lead to patient harm and liability for developers or institutions that remain unaware of the state of the art.

By anchoring AI in EBM, we pair innovation with clinical prudence, enabling the system to serve as a dynamic tool that advances along with emerging science. This safeguards patients, supports clinician confidence, and keeps technology from outpacing the fundamental requirement to deliver only proven effective care.

12 Require Developer Responsibility

"AI developers have a duty to proactively design, test, and continuously refine their systems for real-world clinical use, to anticipate socio-technical effects, to ensure user training, and to assume liability when design flaws compromise patient safety."

While frontline care providers remain accountable for patient outcomes, the creators of healthcare AI cannot relinquish

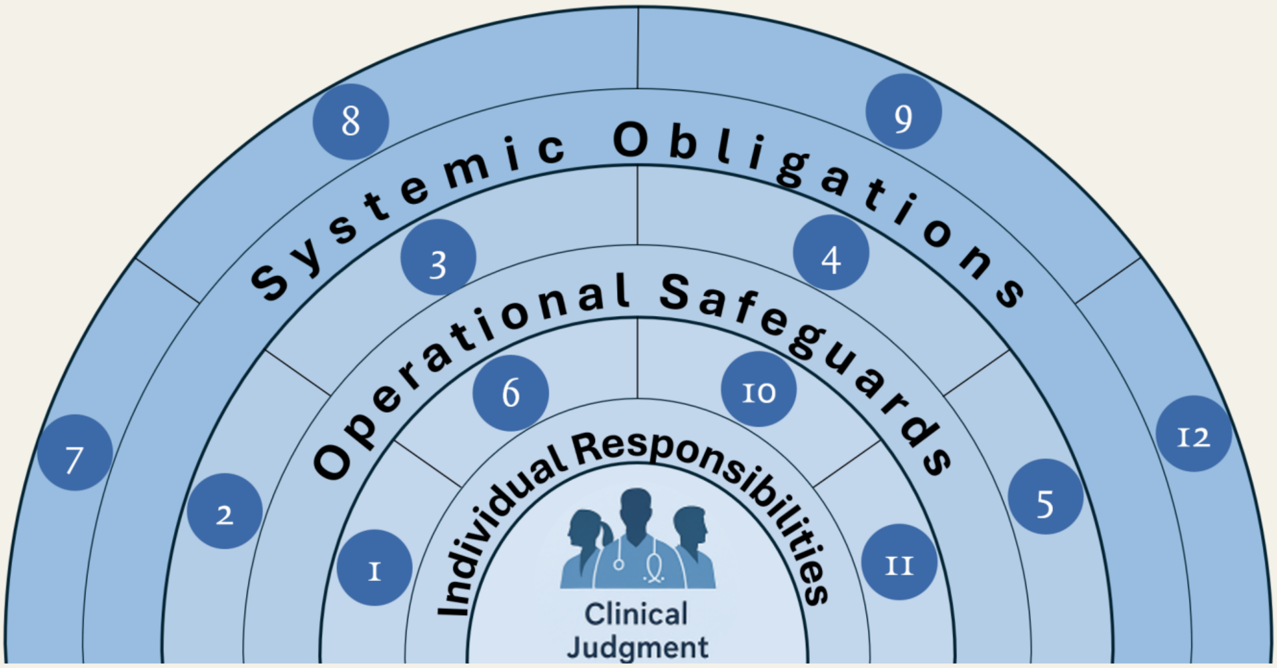
Published by ISMAI · www.ismai.org

responsibility once their code is written. This principle reflects a growing consensus that ethical AI development involves more than mere programming: it requires robust quality assurance, user-centred design, and vigilance for unintended consequences. From legal and ethical standpoints, software vendors and AI companies may bear product liability similar to that of traditional device manufacturers. If an algorithmic flaw or misleading claim precipitates patient harm, the developer's negligence or inadequate testing can be scrutinised and can create legal liability.

In practice, "developer responsibility" entails structured validation protocols—conducting multi-site trials, collecting feedback from clinicians, and iterating real-world data. Ethical guidelines from the WHO and professional societies encourage an ongoing dialogue: if clinicians detect anomalies or biases, the developer should swiftly investigate, correct, and update the product. Similarly, thorough documentation of AI functionality, recognised limitations, and population scope fosters transparency and reduces misuse.

From a human factors' perspective, developers must consider how clinicians and nurses interact with the tool, providing training materials and supportive user interfaces that minimise errors. Overselling the system's accuracy or failing to disclose known blind spots can undermine clinical judgment and lead to suboptimal decisions. Hence, developers are urged to adopt "responsible innovation" frameworks, acknowledging that healthcare AI is never purely technical, because it impacts lives in myriad, often unpredictable ways.

Ultimately, this principle shifts the paradigm from "ship it and forget it" to a partnership model. By embracing accountability for design, performance, and ongoing improvement, AI developers uphold the same standard of care that clinicians bring to their own professional duties, completing the chain of ethical stewardship in medical AI.


Systemic Obligations
Operational Safeguards
Individual Responsibilities
Clinical Judgment
8
9
3
4
7
2
6
10
12
5
1
11