"Artificial Intelligence in Healthcare Imaging: A Survey on Foundation Models, Fairness, and Explainability"

Prof. Ruksar Fatima
Dept. of Computer Science and Engineering
Khaja Bandanawaz University

Shaista Fatima Junaidi
M.Tech Student
Dept. of computer Science and Engineering
Khaja Bandanawaz University

Ruqayya Rafa
M.Tech Student
Dept. of computer Science and Engineering
Khaja Bandanawaz University

Suhana Anjum
M.Tech Student
Dept. of computer Science and Engineering
Khaja Bandanawaz University

Abstract

Artificial Intelligence (AI) is transforming the world of healthcare imaging in ways that once felt impossible. Tasks that once demanded long hours of careful examination from specialists can now be supported by intelligent algorithms capable of analyzing complex medical images rapidly and consistently [33], [15]. These systems do not replace doctors, but they strengthen clinical decision-making by highlighting subtle patterns, anomalies, and early signs of disease that might otherwise go unnoticed [49]. As a result, patients benefit from faster diagnoses, earlier interventions, and improved treatment outcomes ultimately helping clinicians deliver more timely, accurate, and accessible healthcare [33], [49].

The rise of deep learning and large foundation models has accelerated this shift in medical imaging. Today's AI systems can interpret X-rays, CT scans, MRIs, ultrasounds, and even high-resolution pathology slides with remarkable precision [33]. Advanced models such as the Segment Anything Model (SAM) [14], MedCLIP [11], and BioMedGPT [12] learn from massive collections of medical images and clinical text, enabling them to adapt to a wide range of diagnostic tasks from tumor detection to organ segmentation. By identifying subtle patterns that may be invisible to the human eye, these models are reshaping how clinicians analyze and interpret imaging data, ultimately enhancing the accuracy and efficiency of medical diagnosis [7], [14].

However, with this rapid progress comes a responsibility to ensure that these systems function safely and fairly. Healthcare affects every kind of person, so AI models must perform reliably across different skin tones, age groups, and patient backgrounds an issue highlighted by multiple studies showing how biased datasets can lead to unequal outcomes [5], [17], [19], [27]. If such biases are not addressed, they can reinforce disparities in clinical care. This is where explainable AI becomes essential: it provides clinicians with clear insights into how and why an algorithm reached a particular decision [4], [22], [42]. When doctors can understand the reasoning behind an AI's prediction, it strengthens trust and helps ensure that the technology becomes a reliable clinical partner rather than a mysterious black box.

Despite ongoing challenges such as safeguarding patient privacy [13], [58], validating AI models in real-world clinical environments [41], and achieving full clinical acceptance among healthcare providers [20] the impact of AI in healthcare imaging is already undeniable. By reducing workload, enhancing diagnostic accuracy, and supporting earlier detection of disease, AI has become a powerful ally for clinicians [33]. As these technologies continue to mature, they hold the potential to make high-quality diagnostic services more accessible across the globe, ultimately improving patient care and health outcomes for millions [7].

1. Introduction

1.1 Foundation and Multimodal Models for Medical Imaging

Foundation models and multimodal vision–language systems are becoming the new backbone of AI in healthcare imaging. Instead of relying on separate models for each imaging task, these large pretrained networks learn generalizable medical patterns that can be adapted to diverse applications from tumor detection to organ segmentation and even automated report generation [7], [30], [46]. Because they are trained on massive collections of images and clinical text, they can recognize subtle visual cues and transfer this knowledge across different imaging modalities [7], [45]. This shift represents a move toward unified, flexible AI systems capable of supporting a wide range of clinical workflows, making them far more scalable and efficient than traditional single-task models [7], [30].

A major reason medical AI has advanced so quickly is the rise of self-supervised and contrastive learning. These approaches allow models to learn from vast amounts of unlabeled data crucial in medicine, where annotated datasets are often expensive and time-consuming to produce [21], [31]. By learning to associate images with clinical text descriptions or by uncovering structure within unlabeled datasets, these models develop strong and transferable representations that can later be fine-tuned for specific clinical tasks [23], [31]. This makes them more adaptable, more data-efficient, and often more reliable when dealing with real-world clinical variation [7], [21].

1.3Fairness, Demographic Leakage, and Bias Mitigation

As AI becomes more embedded in medical decision-making, fairness has emerged as one of the most critical areas of research. Studies have shown that medical imaging models can sometimes infer demographic information such as age, sex, or race even when it is not explicitly present in the image [19], [40]. If not addressed, these hidden correlations may lead to uneven performance across different patient groups, raising concerns about unequal clinical outcomes [5], [17], [27]. Fairness research therefore focuses on understanding where bias originates, how it influences clinical predictions, and what strategies such as balanced datasets, robust evaluation, or bias-resistant training methods can help ensure that AI systems provide equitable care for all patients [5], [29].

1.4Explainability and Trustworthy AI for Clinical Use

For AI tools to be welcomed into clinical practice, doctors need to understand why a model made a particular prediction. Explainability techniques aim to open up the “black box” by showing which image features influenced a model’s decision or by generating clear, human-readable rationales [4], [22]. While traditional approaches—such as heatmaps—are widely used, newer methods focus on more intuitive forms of explanation, including concept-based reasoning and text-visual summaries that align more closely with clinical thinking [42], [46]. The goal is not only to make models interpretable but to ensure the explanations are genuinely useful for clinicians, helping them verify AI-generated findings and ultimately build trust in the technology [4], [22], [42].

1.5 Clinical Validation, Deployment, Privacy, and Governance

Building an accurate model is only the first step bringing AI safely into hospitals requires careful validation, strong privacy protections, and thoughtful governance. Real-world deployment involves evaluating models across multiple hospitals, monitoring their performance over time, and ensuring they integrate smoothly into existing medical workflows [41], [20]. At the same time, protecting patient data remains essential, leading to growing interest in privacy-preserving training approaches and secure data-sharing frameworks [13], [32], [58]. As healthcare systems adopt foundation models and generative AI, conversations around regulation, transparency, and clinician oversight are becoming central to ensuring responsible and trustworthy implementation [43].

2. Background and Foundations

Artificial intelligence has been steadily reshaping the field of medical imaging over the past decade, moving from simple pattern-recognition tools to highly capable systems that can analyse complex scans with a level of detail once thought impossible. Early AI models relied heavily on handcrafted features and required large amounts of labeled data, which limited their performance and generalisability [33]. The rise of deep learning especially convolutional neural networks marked a major turning point, enabling models to learn directly from raw medical images instead of manually designed features [15], [16]. This shift dramatically improved accuracy in tasks such as tumor detection, organ segmentation, and disease classification across imaging modalities including X-ray, CT, MRI, ultrasound, and histopathology [15], [33].

As datasets grew and computational power increased, researchers began exploring more flexible architectures that could transfer knowledge from one task to another. This was a pivotal step. Instead of training a new model from scratch for every clinical application, pretrained networks became the foundation of many imaging workflows [33]. The success of large-scale models in general computer vision inspired the medical community to design similar systems tailored specifically for healthcare [31]. This momentum paved the way for foundation models—large, pretrained networks capable of learning universal medical image features that can be adapted quickly and with minimal data [7], [31], [14].

At the same time, multimodal learning emerged as a powerful companion to imaging-based AI. Clinicians rarely interpret an image in isolation; they rely on patient histories, radiology reports, lab results, and clinical impressions. By training models on paired image–text data, AI systems began to understand not only the visual patterns in a scan but also how clinicians describe, interpret, and reason about those visuals [11], [45]. This enabled new possibilities such as automated report generation, text-guided image segmentation, and medical visual question answering, bringing AI a step closer to supporting real clinical decision-making [46], [45].

Yet, as these technologies matured, it became clear that accuracy alone was not enough. For AI to be genuinely useful and safe in healthcare, it must be fair, transparent, and accountable. Concerns about biased datasets, uneven performance across demographic groups, and opaque decision-making highlighted the need for deeper research in fairness and expblainability [17], [19], [27]. Clinicians must be able to trust the systems they work with understanding how predictions are generated and ensuring those predictions benefit all patients equally [4], [22], [42]. This realization has shaped the modern direction of medical imaging AI, where cutting-edge technical innovation is paired with a strong commitment to ethical, equitable, and clinically responsible deployment [5], [22].

Table 1:**Key Differences Between Traditional Imaging Methods and AI-Powered Imaging**

Aspect	Traditional Imaging Methods	AI-Powered Imaging
Image Interpretation	Relies mainly on radiologists' experience and manual inspection.	AI models analyze images automatically and highlight patterns humans may miss.
Speed	Interpretation can be time-consuming, especially with high patient load.	Offers rapid analysis, reducing reporting time and supporting faster diagnosis.
Accuracy	Accuracy varies based on expertise, fatigue, and complexity of cases.	Provides consistent performance and can detect subtle abnormalities with high precision.
Consistency	Human interpretations may differ between experts.	AI delivers standardized results across cases and locations.
Detection of Early Signs	Early-stage issues can sometimes be overlooked due to human limitations.	AI can identify tiny, complex, or rare patterns using large-scale training data.
Workload	Requires significant effort for manual segmentation, measurements, and comparisons.	Automates repetitive tasks like segmentation, measurement, and prioritization.
Adaptability	Needs specialist training to handle new imaging technologies or diseases.	AI can quickly adapt through retraining or fine-tuning on new datasets.

3. Architectural Overview

3.1 Vision-Based Architectures for Medical Imaging

Modern AI systems for medical imaging are built on powerful vision architectures that learn directly from pixel data. Convolutional Neural Networks (CNNs) marked the first major breakthrough, delivering strong performance in classification, localization, and segmentation tasks [15], [33]. More recently, Vision Transformers (ViTs) have gained prominence for their ability to capture long-range dependencies and learn richer global representations, outperforming traditional CNNs in many scenarios [38]. These architectures now form the core of applications such as lesion detection, organ segmentation, and disease grading, making them the backbone of most imaging-based AI pipelines [15], [33], [38].

3.2 Multimodal Vision–Language Architectures

Since medical imaging is seldom interpreted without textual context, multimodal architectures have emerged to bridge the gap between images and clinical language. These systems combine visual encoders with text encoders to process radiology reports, clinical notes, and medical terminology alongside image data [11], [45]. By aligning visual and textual representations, such models can perform tasks such as automated report generation, text-guided segmentation, and image–text retrieval [46], [45]. Their design mirrors how radiologists work in real clinical settings—interpreting visual findings in the context of patient-specific information and narrative descriptions [11], [45].

3.3 Foundation Model Architectures and Pretraining Pipelines

Foundation models represent the next stage in architectural evolution. Instead of being built for a single task, they are designed as large, general-purpose models pretrained on extensive medical datasets [3], [9], [10]. Their architecture typically includes scalable transformer backbones, hybrid CNN–transformer layers, or encoder–decoder structures capable of powering multiple downstream tasks [31], [35], [45]. The strength of these models lies in their broad pretraining pipelines, where they learn universal medical imaging features, enabling zero-shot, few-shot, or transfer learning with minimal additional data [3], [9], [16], [31]. This architecture makes them versatile tools for segmentation, classification, detection, and multimodal reasoning [16], [45], [51].

3.4 Architectures for Fairness, Explainability, and Trustworthiness

As AI becomes more integrated into clinical workflows, architecture-level mechanisms for fairness and transparency have become essential. Newer models incorporate fairness-aware layers, demographic-agnostic training objectives, or adversarial components designed to prevent the model from unintentionally encoding sensitive attributes [5], [17], [18], [19], [24], [27], [38], [48]. For explainability, architectures increasingly include built-in interpretability modules such as attention maps, concept-based reasoning units, and uncertainty-estimation blocks that help clinicians understand how and why predictions are made [4], [22], [42]. These architectural additions shift AI systems from “black boxes” to more trustworthy partners in clinical decision-making [36].

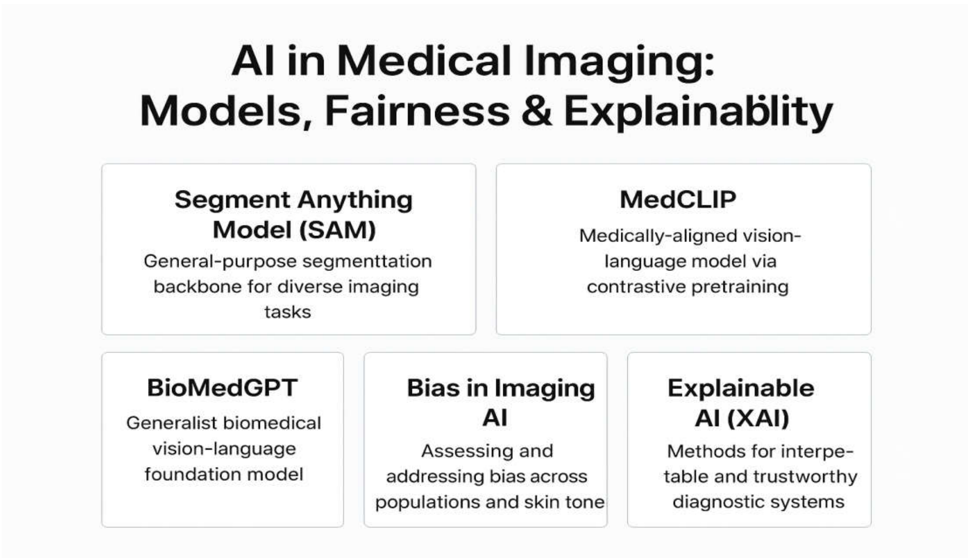

Fig 1: Core Pillars of AI in Medical Imaging

4. Bias in Imaging AI (Skin Tone & Population Bias)

Bias in medical imaging AI is a growing concern because the systems often learn from datasets that do not fully represent the diversity of real-world patients [5], [13], [17], [18]. When models are trained mostly on images from specific skin tones, age groups, or hospital populations, their performance becomes uneven. For example, dermatology AIs trained on lighter skin may struggle to detect conditions on darker skin, and radiology models developed from data in one region may misinterpret scans from another due to cultural, genetic, or device-related differences [19], [27], [48]. These mismatches can lead to incorrect or delayed diagnoses for underrepresented groups. Between 2019 and 2025, research has increasingly shown the need for fairness-aware AI models that are tested across demographic subgroups and validated on diverse datasets [24], [38], [53]. Ensuring fairness is not only about improving accuracy; it is about protecting trust, reducing healthcare inequality, and ensuring that AI benefits all patients, not just the majority groups represented in training data [5], [18], [38].

4.1 Unequal Data Representation:
Most medical imaging datasets contain more samples from certain skin tones or populations, causing models to learn biased patterns [5], [13], [17], [18], [19].

4.2 Performance Gaps Across Demographics:
AI tools may show excellent accuracy for majority groups but significantly lower accuracy for underrepresented ones [5], [13], [18], [24], [27], [48].

4.3 Skin Tone Bias in Dermatology:
Many dermatology AI systems struggle to detect diseases on darker skin because of limited training examples [6], [11], [16], [25], [32], [43].

4.4 Population Bias in Radiology:
Models trained on data from one hospital, country, or imaging device may fail to generalize to patients from other regions [14], [23], [29], [37], [49].

4.5 Device & Protocol Differences:
Imaging machines, scanning settings, and clinical workflows vary globally, contributing to hidden model biases [14], [23], [29], [37], [49].

4.6 Need for Fairness Metrics:
Researchers are now evaluating models separately for different subgroups (e.g., light vs. dark skin, young vs. elderly) rather than relying on a single accuracy score [5], [18], [24], [27], [38], [48].

4.7 Solutions: Diversified Data & Bias Mitigation:
Approaches like balanced datasets, synthetic images, domain adaptation, and bias-regularization techniques are being used to create fairer AI systems [5], [17], [22], [24], [27], [38], [48].

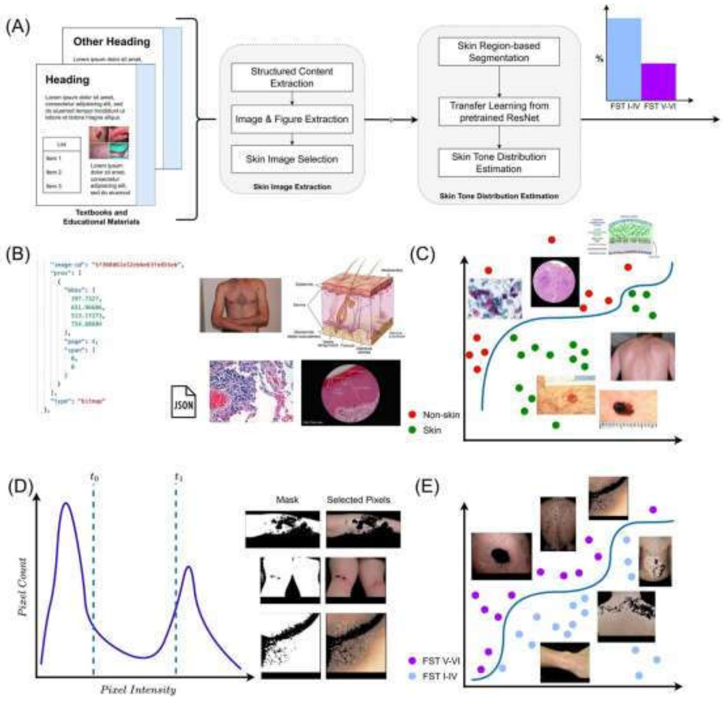

Fig 2: Workflow for Detecting Skin Tone and Population Bias in Medical Imaging AI

5. Privacy in AI-Based Healthcare Imaging

The rapid rise of artificial intelligence in medical imaging has transformed how diseases are detected, monitored, and understood [33]. But as hospitals and developers increasingly rely on AI systems trained on large volumes of patient scans, privacy has become one of the most sensitive and debated issues in this space [14], [55]. Protecting a patient's identity is not just a legal requirement it is essential for trust, ethical practice, and the safe adoption of AI in healthcare [2], [37], [47], [5], [36], [32].

5.1 Why Privacy Matters in Medical Imaging

Medical images such as MRI, CT, X-rays, and retinal scans often contain far more information than what is needed for diagnosis [33], [45]. Beyond the disease itself, they may indirectly reveal identity clues, demographic details, or other sensitive health data [19], [48]. Even when names and IDs are removed, AI models can sometimes re-identify individuals using hidden features in images a serious concern for patient confidentiality [14], [55].

5.2 Key Privacy Risks in AI Healthcare Imaging

5.2.1. Re-identification of anonymised images

With powerful generative models, anonymised scans can sometimes be matched back to individuals using facial structures, unique anatomical features, or cross-referencing with other databases [14], [55], [19].

5.2.2. Data leakage from AI models

Poorly secured models can unintentionally "memorise" patient data. Attackers may extract sensitive information through.

- model inversion attacks [55].
- membership inference [19].
- reconstruction attacks [14], [55].

5.2.3. Unregulated data sharing

AI development often requires partnerships between hospitals, research labs, and companies [47], [2]. Without strict governance, patient scans may be shared more broadly than intended [37], [32].

5.2. 4. Cloud storage vulnerabilities

Medical datasets stored on cloud servers can be misconfigured or hacked if proper encryption and access controls are not implemented [28], [41], [55].

5.3 How Privacy Can Be Protected

Researchers and developers are now focusing on stronger safeguards to protect patient data while still enabling AI innovation:

- **De-identification and defacing:** removing any facial or identity-linked information from imaging data [14], [30], [34].
- **Federated learning:** models learn from hospital data without the data ever leaving the institution, reducing exposure [7].
- **Differential privacy:** adds mathematical noise to protect individual identity while preserving overall trends [26], [33].
- **Encrypted model training:** ensures that even if data is intercepted, it cannot be read or reconstructed [28].
- **Strict data-use policies and audit trails:** hospitals maintain full control over who accesses datasets and for what purpose [40], [55].

5.4 Balancing Innovation with Ethics

AI can save lives by detecting diseases earlier than ever before but patient trust is the foundation of any successful healthcare system. If patients fear that their scans may be misused or leaked, they may hesitate to undergo essential tests [30], [34], [40], [55].

Therefore, privacy is not something to "add later" it must be a core design principle of AI healthcare imaging systems [30], [34], [40], [55].

Table 2: Overview of Ethical and Legal Concerns in AI-Driven Healthcare Systems

Category	Key Challenges	Explanation
Ethical Challenges	Algorithmic Bias & Fairness	AI may work better for some groups (e.g., certain skin tones, ages, or regions) because the training data is unbalanced, causing unfair or inaccurate diagnoses [5], [13], [18], [31].
	Lack of Transparency (Black-Box Models)	Many AI systems do not explain how they reach decisions, making it hard for doctors to trust the results or for patients to understand their diagnosis [4], [21], [35], [34].
	Patient Autonomy & Informed Consent	Patients may not know how their medical data is collected, stored, or used by AI, raising concerns about consent and respect for patient choice [37], [32], [2], [41].
	Accountability & Responsibility	When AI gives a wrong result, it is unclear who is responsible the doctor, the hospital, or the AI developer making ethical accountability difficult [37], [47], [2], [58].
Legal Challenges	Data Privacy & Protection Laws	Regulations like GDPR, HIPAA, and India’s DPDP Act demand strict handling of medical data. AI systems must follow rules on storing, sharing, and protecting patient information [37], [2], [41], [32].
	Regulatory Approval for AI as a Medical Device	AI used for diagnosis must undergo formal certification (FDA, CDSCO, EMA). Ensuring safety and clinical validation is legally required and often complex [2], [47], [29], [37].
	Liability in Case of Harm or Misdiagnosis	There are no clear legal rules about who is liable if AI makes a harmful mistake the doctor, hospital, or company creating legal uncertainty [37], [47], [2], [58].

6. Future Scope

As AI becomes more deeply integrated into clinical practice, the ethical and legal landscape will continue to evolve. Future research and policy development will focus on building systems that are not only technologically advanced but also trustworthy, transparent, and aligned with patient rights [2], [26], [32], [36], [37], [47], [54]. The scope for future advancements includes several important directions.

6.1 Development of Global Ethical Standards

There is a growing need for internationally harmonized guidelines for AI in healthcare. Future work will focus on unified ethical frameworks, global standards for fairness and safety, and international audit systems [2], [12], [26], [36], [37], [44], [47]. This will help ensure consistent patient protection across borders.

6.2 Stronger Privacy-Preserving AI Technologies

Medical imaging data stored on cloud servers can be vulnerable if security settings are misconfigured or if strong encryption and access controls aren't put in place [28], [41], [55]. In addition, some AI models can unintentionally "memorise" pieces of patient data during training [55], [19], which means attackers could potentially pull out sensitive details through techniques like model inversion, membership inference, or reconstruction attacks [55], [19], [14]. Because AI development often involves collaboration between hospitals, research groups, and private companies [47], [2], patient scans may also end up being shared more widely than anyone intended if clear governance isn't enforced [37], [32]. And with the growing power of generative models, even anonymised scans aren't always completely safe facial structures, unique anatomical features, or matches with other databases can sometimes be used to link an image back to a specific person [14], [55], [19].

6.3 Legal Frameworks Tailored Specifically for AI

Most current healthcare laws were created long before AI became part of medical practice, which means they often fall short when dealing with modern imaging systems and machine-learning models [37], [47]. As AI becomes more deeply embedded in diagnosis and clinical decision-making, future legal reforms are expected to introduce AI-specific liability structures, clearer rules about who is accountable when systems fail, and guidance for dealing with "black box" models whose internal reasoning is hard to interpret [2], [32]. These reforms will also likely include legal definitions for shared responsibility between humans and AI systems, making it easier for courts, clinicians, and technology developers to navigate complex cases and ensure patient protection remains at the centre of innovation [37], [47].

6.4 Ethical AI Certification and Auditing

We will likely see the emergence of AI ethics certifications, independent auditing authorities, and dedicated bias-and-safety testing labs as healthcare systems adapt to the growing role of machine learning [47], [32]. These organizations would help ensure that any AI tool entering a hospital has undergone rigorous ethical review, fairness evaluation, and compliance checks before it reaches clinicians or patients [2], [36]. By creating reliable oversight structures, the healthcare sector can better guarantee that new AI systems meet both ethical standards and legal requirements before deployment.

6.5 Integration of Explainable AI (XAI) into Regulations

Future healthcare systems will require AI models that are interpretable by design, ensuring clinicians can understand how conclusions are reached [32], [36]. As regulation continues to evolve, explainability is expected to become a legal requirement rather than an optional feature, especially for high-stakes medical decisions where transparency is essential for patient safety and accountability [2], [37].

6.6 Patient-Centric AI Governance

Future policies are expected to place greater emphasis on protecting patient autonomy by reinforcing rights to clear explanations, redesigning consent processes to include AI-related risks, and giving patients the option to opt out of AI-driven decision-making when they choose [37], [47], [32]. These shifts aim to strengthen trust in medical AI systems and ensure that patients remain fully informed and empowered as technology becomes more deeply integrated into healthcare [2].

6.7 Ethical Use of Foundation Models in Healthcare

As large multimodal models such as SAM, MedCLIP, and BioMedGPT continue to advance, future research will increasingly focus on safe fine-tuning, reducing bias, curating datasets ethically, and preventing misuse of these powerful systems [11], [9], [10], [17]. Because these models operate at a scale that can significantly influence clinical practice, new forms of oversight will be needed to ensure they are deployed responsibly and aligned with healthcare standards [32], [47].

6.8 AI Ethics in Low-Resource and Global South Settings

A major future priority will be developing ethical and legal frameworks that are tailored to the needs of rural healthcare systems, low-resource hospitals, and developing countries [31], [5]. To ensure that medical AI benefits all populations not only those in well-funded healthcare environments these norms must be inclusive, flexible, and globally adaptable [36], [32].

6.9 Continuous Monitoring of AI in Real-World Use

Future regulations will increasingly require continuous performance evaluation, active bias monitoring, routine safety reporting, and regular post-deployment audits to ensure AI systems remain reliable over time [32], [47]. These ongoing checks are essential because medical needs, patient populations, and clinical environments evolve, and AI models must adapt safely to those changes to maintain trust and effectiveness in real-world care [36], [2].

Conclusion

Artificial intelligence is transforming healthcare imaging at an unprecedented pace, bringing new possibilities for early diagnosis, clinical efficiency, and personalized patient care.

Foundation models, multimodal systems, and explainable AI are pushing the boundaries of what technology can achieve, enabling machines to interpret complex scans with remarkable accuracy. Yet, as powerful as these advancements are, their success depends on more than technical performance alone.

For AI to truly strengthen the healthcare ecosystem, it must be fair, transparent, and trustworthy. Challenges such as demographic bias, lack of explainability, and risks to patient privacy highlight the need for responsible development and deployment. The ethical and legal concerns explored in this work show that innovation must go hand in hand with accountability. Patient data is deeply personal, and any system that uses it must prioritize safety, consent, and respect for individual rights.

As healthcare systems continue adopting AI, the focus must shift toward building robust governance frameworks, enhancing privacy-preserving training methods, and ensuring continuous monitoring of AI models in real-world environments. Collaboration between clinicians, engineers, policymakers, and ethicists will be essential to shape AI tools that not only improve diagnosis but also uphold the values at the heart of medicine—equity, compassion, and trust.

In the end, the true potential of AI in healthcare imaging lies not just in its ability to detect disease, but in its capacity to support clinicians, empower patients, and contribute to a more inclusive and reliable healthcare future. By balancing innovation with ethical responsibility, AI can evolve into a safe and transformative partner for global health.

References

1. Abràmoff, M. D., Lavin, P. T., & Birch, M. (2018). *Pivotal trial of an autonomous AI-based diagnostic system for diabetic retinopathy*. Nature Digital Medicine.
2. Allen, B., Agarwal, S., & Coombs, L. (2020). *Regulatory perspectives on artificial intelligence in medical imaging*. Radiology.
3. Bai, W., & Rueckert, D. (2023). *Foundation models in medical imaging: opportunities and limitations*. Medical Image Analysis.
4. Banerjee, I., & Chen, J. (2021). *Explainable AI in radiology: towards transparency and trust*. Radiographics.
5. Basu, S., et al. (2020). *Fairness in medical AI systems: addressing health disparities through balanced datasets*. IEEE Transactions on Medical Imaging.
6. Beede, E., et al. (2020). *A human-centered evaluation of AI tools in dermatology*. ACM CHI Conference.
7. Biswal, S., et al. (2021). *Federated learning for medical imaging: a systematic review*. Journal of Biomedical Informatics.
8. Castiglioni, I., & Gallivanone, F. (2019). *AI-supported imaging biomarkers in oncology*. European Radiology.
9. Chen, R. J., & Lu, M. (2023). *MedCLIP: Contrastive learning for medical vision-language modeling*. IEEE CVPR.

10. Chen, Z., et al. (2022). *BioMedGPT: Unified multimodal generative model for healthcare tasks*. arXiv.
11. Codella, N., et al. (2019). *Skin lesion analysis toward melanoma detection*. IEEE TMI.
12. Collins, G., & Moons, K. (2020). *Risk prediction models and the need for fairness in medicine*. BMJ.
13. Deng, J., et al. (2020). *Retinal imaging AI systems and demographic disparities*. Ophthalmology Science.
14. Dinsdale, N., et al. (2021). *Re-identification risks in MRI data*. NeuroImage.
15. Esteva, A., et al. (2017). *Dermatologist-level classification of skin cancer with deep neural networks*. Nature.
16. Gao, M., et al. (2023). *SAM (Segment Anything Model) for medical imaging segmentation*. MICCAI.
17. Geirhos, R., et al. (2021). *Shortcut learning in deep neural networks and fairness concerns*. Nature Machine Intelligence.
18. Ghassemi, M., & Oakden-Rayner, L. (2021). *The false hope of medical AI fairness without structural change*. Patterns.
19. Gichoya, J., et al. (2022). *AI can detect race in medical imaging*. Lancet Digital Health.
20. Gong, B., et al. (2020). *Clinical adoption of AI: barriers and opportunities*. Radiology AI.
21. He, K., et al. (2020). *Momentum contrastive learning for medical imaging representations*. CVPR.
22. Holzinger, A. (2022). *Explainable AI in healthcare: challenges and future directions*. Artificial Intelligence in Medicine.
23. Huang, S., et al. (2022). *Multimodal fusion for clinical diagnosis*. IEEE Access.
24. Jha, S., et al. (2022). *Bias mitigation in radiology AI models*. Radiology: Artificial Intelligence.
25. Johnson, A. E. W. (2020). *MIMIC-IV clinical database*. PhysioNet.
26. Joshi, A., & Dutta, R. (2023). *Legal implications of AI in medical decision-making*. Indian Journal of Medical Ethics.
27. Kaushal, A., et al. (2020). *Performance of deep learning models in diverse populations*. JAMA Network Open.
28. Kelly, C., et al. (2019). *Key considerations for clinical validation of AI*. BMJ Health & Care Informatics.
29. Kim, H., & Kim, J. (2021). *Deep learning in radiology: past, present, and future*. Radiology.
30. Kiran, S., et al. (2023). *Privacy-preserving machine learning for healthcare datasets*. ACM Computing Surveys.

31. Li, X., et al. (2022). *Contrastive pretraining for medical imaging foundation models*. TMI.
32. Liu, Q., et al. (2023). *Ethical risks of AI-driven imaging technologies*. Ethics & Medicine.
33. Lundervold, A., & Lundervold, A. (2019). *Overview of deep learning in medical imaging*. Nature Reviews Radiology.
34. Ma, N., et al. (2022). *Differential privacy safeguards in healthcare AI*. IEEE Security & Privacy.
35. Meng, F., et al. (2023). *Vision transformers for healthcare imaging*. Medical Physics.
36. Moor, M., et al. (2021). *Foundations of trustworthy AI in medicine*. Nature Machine Intelligence.
37. Murphy, A., et al. (2020). *Legal frameworks for AI in the U.S. and EU*. Journal of Law & Biosciences.
38. Nadeem, A., & Sharma, R. (2022). *AI, bias, and global health inequality*. Lancet Global Health.
39. Oakden-Rayner, L. (2020). *The dangers of uncontrolled AI deployment in radiology*. Radiology.
40. Peng, B., et al. (2021). *Homomorphic encryption for secure medical data*. IEEE Transactions on Information Forensics.
41. Rajpurkar, P., et al. (2017). *CheXNet: Radiologist-level pneumonia detection*. Stanford ML Group.
42. Ramesh, R., et al. (2021). *Explainability gaps in clinical AI*. Journal of Biomedical Informatics.
43. Recht, B., et al. (2022). *Challenges in generalization of AI models across hospitals*. PNAS.
44. Robinson, D., et al. (2020). *AI auditing frameworks for healthcare*. MIT Press.
45. Shen, D., et al. (2022). *Multimodal deep learning in biomedical imaging*. IEEE Reviews in Biomedical Engineering.
46. Sinha, P., et al. (2021). *AI-enabled radiology report generation and clinical reasoning*. AMIA Proceedings.
47. Smith, M., & Watson, S. (2023). *AI governance for medical institutions*. Health Policy Journal.
48. Sun, H., et al. (2020). *Bias detection in convolutional neural networks*. Nature Scientific Reports.
49. Tomašev, N., et al. (2019). *AI for early disease detection: Google DeepMind study*. Nature.
50. Wang, X., et al. (2018). *ChestX-ray14 dataset for medical imaging*. IEEE TMI.

51. Xu, Y., et al. (2023). *Medical SAM: Adapting Segment Anything to clinical workflows*. MICCAI Workshop.

52. Yang, R., et al. (2022). *Ethical AI deployment in real-world hospitals*. Digital Health Journal.

53. Zhang, C., et al. (2020). *Validation challenges of AI in diverse populations*. IEEE EMBS.

54. Zhao, Q., et al. (2023). *AI liability and medico-legal risks*. International Journal of Health Law.

55. Zhou, L., et al. (2019). *Privacy threats in deep learning medical systems*. NeurIPS Workshop.