Ethical Considerations for Artificial Intelligence in Medical Imaging: Deployment and Governance


1Department of Health Humanities and Bioethics and Department of Philosophy, University of Rochester, Rochester, New York; 2Department of Bioethics, Hospital for Sick Children, and Dana Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; 3Department of Philosophy and Religion and Khoury College of Computer Sciences, Northeastern University, Boston, Massachusetts; 4Department of Radiology and Nuclear Medicine, Cancer Centre Amsterdam, Amsterdam University Medical Centres, Amsterdam, The Netherlands; 5Department of Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, Bethesda, Maryland; 6Department of Biomedical Engineering and Mallinckrodt Institute of Radiology, Washington University in St. Louis, St. Louis, Missouri; 7Departments of Radiology and Physics, University of British Columbia, Vancouver, British Columbia, Canada; 8Department of Radiology, University of Michigan Medical School, Ann Arbor, Michigan; 9Departments of Radiology and Physics, University of Iowa, Iowa City, Iowa; 10Mallinckrodt Institute of Radiology, Washington University in St. Louis, St. Louis, Missouri; and 11Siemens Medical Solutions USA, Inc., Hoffman Estates, Illinois

The deployment of artificial intelligence (AI) has the potential to make nuclear medicine and medical imaging faster, cheaper, and more effective and more accessible. This is possible, however, only if clinicians and patients feel that these AI medical devices (AIMDs) are trustworthy. Highlighting the need to ensure health justice by fairly distributing benefits and burdens while respecting individual patients’ rights, the AI Task Force of the Society of Nuclear Medicine and Molecular Imaging has identified 4 major ethical risks that arise during the deployment of AIMD: autonomy of patients and clinicians, transparency of clinical performance and limitations, fairness toward marginalized populations, and accountability of physicians and developers. We provide preliminary recommendations for governing these ethical risks to realize the promise of AIMD for patients and populations.

Key Words: AI ethics; software as medical device; fairness; explainability; justice

J Nucl Med 2023; 64:1509–1515
DOI: 10.2967/jnumed.123.266110

Artificial intelligence (AI) and machine learning systems will likely soon be incorporated into various aspects of patient care in nuclear medicine. These AI medical devices (AIMDs) fuse traditional medical devices with continuously learning software systems to improve patient care and health-care worker practices. AI in nuclear medicine offers a tremendous opportunity for faster and more reliable diagnoses (1), and over 340 medical imaging AIMDs have been approved by the U.S. Food and Drug Administration at the time of writing this publication (2). The ethical benefits of AIMD may be particularly profound in nuclear medicine, where the use of radiation generates a strong imperative to use all feasible means to minimize exposure doses (e.g., the ALARA principle) and improve the accuracy of treatment. However, the deployment of AI without regard to potential ethical risks may result in unintended harm to patients and health-care systems (3). Many have raised concerns regarding patient privacy, the opacity of algorithms, deskilling of clinicians, and the robustness of systems in lower resource contexts (4–7). Moreover, there is mounting evidence that AIMDs may exacerbate existing health disparities based on race, ethnicity, sex, and socioeconomic status (8–10). Grappling with these ethical issues is essential before the widespread adoption of AIMDs in nuclear medicine.

The AI Task Force of the Society of Nuclear Medicine and Molecular Imaging has set out to make clear the assignments of responsibility between developers, physicians, and regulators by distinguishing between the development and deployment of AIMDs. In a companion paper (11), we will discuss the ethical duties of researchers in 3 phases of the AIMD production pipeline: during data collection, training and validation, and evaluation of the tool. In this paper, we focus on the obligations of clinicians and regulators during the deployment of AIMD.

The use of medical devices has historically been constrained by the traditional 4 principles of medical ethics—autonomy, nonmaleficence, beneficence, and justice—with the greatest emphasis placed on patient autonomy and nonmaleficence (12). AIMDs are sometimes thought to make compliance with these first 2 principles more challenging. By automating diagnostic and prognostic tasks within opaque AI models, they make informing patients and catching errors more difficult. Professional societies have thus sought to extend the core ethical principles of autonomy, beneficence, nonmaleficence, and justice to include further principles such as explicability and transparency to buttress patient autonomy and prevent harmful errors (6).

Although these extended frameworks are well suited to interactions between clinicians and individual patients, they are less well suited to the governance of AIMDs within complex health systems. Governance of AIMDs requires distributing benefits and burdens between multiple stakeholders. For instance, reasonable people may disagree about the appropriate tradeoff between false-positive and false-negative rates for the detection of malignancies, but an AIMD may be able to encode only a single tradeoff for all
patients. The traditional principles of clinical ethics, which focus on the obligations of caregivers or researchers in direct contact with patients, cannot be straightforwardly applied to these multilateral decisions (13,14). In this respect, the governance of AIMDs ought to take seriously the problem of navigating the circumstances of justice, where multiple stakeholders must work together to produce a shared good (i.e., AIMDs) while respecting one another’s rights and fairly distributing benefits and burdens (14). Although some of the traditional principles of clinical ethics may be useful as starting points, their application to AIMDs requires a greater emphasis on the principle of justice. Throughout this paper, we consider the deployment and governance of AIMDs in nuclear medicine through the lens of 3 domains of value: patient welfare, patient autonomy, and health justice (Table 1).

<table>
<thead>
<tr>
<th>Ethical dimension</th>
<th>Clinicians during deployment</th>
<th>Governance by administrators and societies</th>
<th>Governance by regulators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Welfare</td>
<td>Ensure familiarity with performance and limitations of AIMDs before use</td>
<td>Ensure clinicians are trained in AIMD performance and limitations</td>
<td>Ensure that clinical utility, not just in silico performance, is evaluated and declared</td>
</tr>
<tr>
<td></td>
<td>Avoid off-label or unevaluated use of AIMD</td>
<td>Avoid automation bias through technical and policy interventions</td>
<td>Ensure that legal liability for harm is clearly demarcated</td>
</tr>
<tr>
<td>Autonomy</td>
<td>Ensure consent to direct data collection and that risks, benefits, and limitations of AIMD are declared to patient</td>
<td>Ensure clear protocols for declaring risks and benefits of AIMD to patients</td>
<td>Ensure that performance and limitations are declared in ways that enable patient and clinician decision-making</td>
</tr>
<tr>
<td></td>
<td>Consider notification for secondary reuse of data in learning AIMDs</td>
<td>Promote legitimacy by involving patients in policy priority setting for AIMD use</td>
<td>Promote techniques that explain salience of input features at population level</td>
</tr>
<tr>
<td>Justice</td>
<td>Carefully justify use of sensitive attributes (race, sex, etc.) as input features</td>
<td>Ensure regular auditing of performance disparities by race, sex, etc.</td>
<td>Require evaluation of performance disparities by race, sex, etc.</td>
</tr>
<tr>
<td></td>
<td>Avoid conflicts of interest as clinician/researcher</td>
<td>Ensure clear mechanisms for managing clinician/researcher conflicts of interest</td>
<td>Promote broad access to AIMDs through subsidy and regulation</td>
</tr>
</tbody>
</table>

Use of term ensure or required denotes legally or ethically required conduct; use of term promote or avoid denotes ethically ideal conduct.

### CLINICAL USE OF AIMD IN NUCLEAR MEDICINE

Clinicians possess the primary ethical responsibility for the use of AIMDs in patient care. This places burden on clinicians to understand the capacities and limits of algorithms but also reinforces the case for developers to clearly specify the performance of the algorithm and its intended-use cases. In this section, we review some ethical considerations for clinicians as they deploy AIMDs to improve patient well-being, respect patient autonomy, and promote health justice.

**Patient Well-Being**

One of a clinician’s primary responsibilities is to act in the best interests of the patient, avoiding harm and benefiting well-being when possible. This requires that clinicians be attentive to automation bias, knowledgeable about the task-specific performance and limitations of an AIMD, and appropriately cautious about the implementation of AIMDs in their practices. Moreover, identifying whether the use of an AIMD is in a patient’s best interest requires consideration of the specific values of individual patients.

**Intended Use and Performance.** Although AIMDs are emerging as incredibly powerful new tools in health care, increasingly able to make diagnostic or treatment recommendations, the nature of the physician–patient relationship requires that those at the bedside retain responsibility and accountability for potential errors in AI-based medical diagnosis or risk stratification. Although developers and regulators carry an ethical burden to ensure that AIMD performance claims are warranted (15), the clinician who is credentialed by the appropriate professional body is responsible for the clinical action. This suggests 3 considerations.
Nonetheless, as CADx systems start to appear, and as AIMDs begin to treat a disease and may only be applied as a measurement tool, include a statement that the software is not intended to diagnose or treat a disease from the clinician. In all cases, clinicians should not use an AIMD outside its intended use. Most systems at the non-CADe level, such as a PET quantification tool, may pose lower risks since they simply provide additional information that the physician incorporates into decision-making. Although even at this level, AIMDs may inadvertently eliminate or de-emphasize malignant features in imagery (17). At the CADe level, the risk increases since physicians may deviate from their judgment on the basis of overreliance on AI-based detection and segmentation of tumors on, for example, 18F-FDG PET. The highest risk comes with CADx systems, since they provide binary (or categorical) diagnostic information and may obscure the underlying evidence or reasoning for the diagnosis from the clinician. In all cases, clinicians should not use an AIMD outside its intended use.

First, clinicians should be knowledgeable about the intended use of an AIMD system. The Food and Drug Administration has done substantial work to define a typology of AIMD (what they call Software as a Medical Device), based on the level of computer-aided detection (CADe) or computer-aided diagnosis (CADx) that the AIMD is intended to provide (Table 2) (16). Non-CADe systems provide measurement or annotation of imagery without interpretation. CADe refers to an AI device that intends to identify abnormalities but does not attempt diagnosis or treatment recommendations. CADx systems attempt to directly diagnose the presence (and severity) of a disease (4).

Table: U.S. Food and Drug Administration Grading System for Risk Evaluation of Software as Medical Device (15)

<table>
<thead>
<tr>
<th>Patient condition</th>
<th>Non-CADe (e.g., measure, denoise, annotate; no interpretation)</th>
<th>CADe (e.g., identify abnormalities; dosimetry calculation)</th>
<th>CADx (e.g., guide diagnosis or treatment)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not serious</td>
<td>I</td>
<td>I</td>
<td>II</td>
</tr>
<tr>
<td>Serious</td>
<td>I</td>
<td>II</td>
<td>III</td>
</tr>
<tr>
<td>Critical</td>
<td>II</td>
<td>III</td>
<td>IV</td>
</tr>
</tbody>
</table>

I–IV represent lowest risk through highest risk.

Grading AIMDs on this scale allows clinicians to identify the level of risk associated with using a specific AIMD in a clinical workflow. Consider AIMD systems involved in PET workflows. Most systems at the non-CADe level, such as a PET quantification tool, may pose lower risks since they simply provide additional information that the physician incorporates into decision-making. Although even at this level, AIMDs may inadvertently eliminate or de-emphasize malignant features in imagery (17). At the CADe level, the risk increases since physicians may deviate from their judgment on the basis of overreliance on AI-based detection and segmentation of tumors on, for example, 18F-FDG PET. The highest risk comes with CADx systems, since they provide binary (or categorical) diagnostic information and may obscure the underlying evidence or reasoning for the diagnosis from the clinician. In all cases, clinicians should not use an AIMD outside its intended use.

Second, responsible use of AIMDs requires that clinicians be familiar with the task-specific performance of an AIMD within the population that the clinician serves. In a previous paper (15), we noted that AIMDs should ultimately be evaluated by their performance on clinical tasks in representative clinical contexts and populations. To avoid inappropriate use, clinicians should familiarize themselves with these performance data, including differences in accuracy for race or sex subpopulations.

Finally, clinicians must consider the risks of automation bias (18). Automation bias occurs when users come to unquestioningly accept the output of AIMD, without appropriate regard for predictive errors or uncertainties. In general, clinicians should act with an appropriate level of skepticism with respect to the outputs of AIMDs, until such time as they are well integrated into routine clinical practice. Indeed, most Food and Drug Administration–approved AIMDs specifically include a statement that the software is not intended to diagnose or treat a disease and may only be applied as a measurement tool. Nonetheless, as CADx systems start to appear, and as AIMDs begin to demonstrate better accuracy than physicians at a specific task, automation bias may become difficult to resist (19). It is critical to remain attentive to the fact that ethical clinical decision-making requires sustaining the shared decision-making paradigm, where AI is but one source of information in a set of considerations that, together, contribute to a decision (20).

Patient Best Interest. The outputs of AIMDs will inform clinician decision-making about a host of tradeoffs in medical imaging: for example, between false-positive and false-negative diagnoses, or acceptable dosage of radioisotopes relative to investigational value. Two considerations are relevant.

First, minimizing harm and maximizing benefit require that we recognize imbalances in the harm of false positives and false negatives for a specific task. For example, in a cancer diagnosis task, false negatives will often have higher costs for patients than false positives (21). This suggests that common performance measures for AIMDs may not provide sufficient information to clinicians and patients involved in shared decision-making. For instance, the area under the receiver-operating-characteristic curve is a threshold agnostic performance metric that treats false positives and negatives as equivalently weighted (which is rarely true in the clinical context) (22). Nor does task-specific selection of simple metrics (e.g., minimizing false negatives) solve the problem, since for almost all tasks both false negatives and false positives harm patients (i.e., through under- and overtreatment). Instead, clear communication of confusion matrices may be a necessary component of clinical evaluation, to ensure that doctors and patients can navigate the complex assessment of costs and benefits themselves.

Second, minimizing harm and maximizing benefit require careful consideration of the different ways patients make tradeoffs between the risks and benefits of interventions. Many interventions in nuclear medicine carry grave tradeoffs between longevity and quality of life. In this respect, AIMDs—and especially CADx or CADe systems (23)—should avoid unnecessarily hard coding judgments about the appropriate risk and benefit tradeoffs (24). For instance, during radiation therapy planning, an AIMD that segments tumors in PET/CT could provide an estimate of how much diseased tissue is present in each voxel of the image (25), allowing caregivers and physicians to discuss risk tolerances with patients. Of course, not all value judgments can be avoided in the development of an AIMD. Tasks such as image denoising or instrument calibration (26)—although they affect the error rates of downstream diagnosis or intervention—are too abstracted from patient outcomes for meaningful dialogue with each individual patient to occur. In these cases, reasonable effort should be made to ensure
that embedded value judgments—that is, aggressiveness of denoising, or sensitivity to patient motion—reflect broadly held standards. If well-established standards do not exist to guide the selection of critical thresholds, developers should seek to involve stakeholders, including patients and providers, in the selection process (27).

**Patient Autonomy**

Clinicians must respect their patients’ autonomy, and this requires that they provide patients with sufficient information to consent to interventions (12). At a minimum, patients must be notified of the use of an AIMD during diagnostic or therapeutic interventions, the safety and efficacy of the AIMD for patients such as them, and any known risks or limitations associated with the AIMD. Furthermore, whereas explainability techniques may facilitate patient autonomy in the future, physicians should be cautious in overly relying on them given current limitations.

*Notification and Risk Declaration.* Clinicians have an obligation to notify patients regarding the use of AIMDs in a clinical workflow when the clinicians have reason to believe that this information would be material to the patients’ decision-making. First, performance information should be clinically relevant (e.g., false-positive and -negative rates in clinical contexts) and not simply abstract performance metrics (e.g., area under the receiver-operating-characteristic curve). This will enable informed discussions with patients about the relative risks and benefits of AIMD use and the relevance of an AIMD’s findings to the overall prognosis or treatment plan. Second, performance limitations for racial or sex subpopulations should be declared to patients who are members of the disadvantaged class. This requires that the performance of the AIMD be evaluated in subpopulations that are likely to be encountered. Moreover, alternatives to the use of the AIMD, and the relative performance of these alternatives, should also be provided to patients.

*Explainability.* The black box nature of deep learning means that detailed information about the decision procedure of AIMDs is not always accessible or interpretable, arguably undermining clinician and patient understanding (28,29) (the standard practice for preparing informed consent forms is to use an eighth-grade reading level). Explainability refers to a cluster of techniques that aim to help physicians understand and explain the AI’s internal decision-making process. Although explainability techniques may sometimes be useful, we argue that understanding the performance and limits of the AI system is likely more important. We do so for 2 reasons.

First, currently existing explainability techniques are not able to reliably explain predictions at the individual level (30). Calls for greater explainability often assume that these techniques can describe the precise computational pathway between individual inputs and outputs (31). One purported strategy is to create a parallel regression model that identifies the statistical association of particular input features to an AIMD’s output. Another is to develop heat maps that purport to identify the areas of an image that were relevant to the prediction. It is unclear, however, whether these techniques deliver information about individual predictions or about the general parameters of the model (6,19). Moreover, emerging evidence indicates that explainability techniques may actually compromise clinicians’ ability to identify incorrect outputs and thus worsen automation bias (19).

Second, calls for explainability often conflate the different forms of explanation that are desirable in different contexts (32). Ferretti et al. (28) distinguish among 3 forms of opacity in medical AI: lack of disclosure, epistemic opacity, and explanatory opacity. Lack of disclosure refers to instances in which patients are unaware that diagnostic or interventional decisions are being made with the aid of an AIMD. These can be dealt with through simple notifications on the use and performance of AIMD. Second, epistemic opacity refers to the inability to inspect the precise computational pathway (i.e., feature weights and parameters) between inputs and predictions. Although this information may be useful to developers, this level of transparency about decision pathways is rarely demanded of other medical interventions (33). Finally, explanatory opacity refers to the inability to explain why the input data are causally connected to the prediction—a problem of particular importance in machine learning, which relies on identifying statistical regularities that may not have well-characterized causal explanations. It is unclear, however, whether principles of informed consent require clinicians to explain to patients the precise causal pathway between diseases and diagnostic tests (33,34). Moreover, a detailed explanation of the causal mechanisms underlying imaging results and disease diagnosis would require a greater understanding of statistics and nuclear medicine than most patients possess. These considerations suggest that explainability techniques—although of technical interest to developers—may not be necessary to satisfy existing informed consent practices.

**Justice and Algorithmic Fairness**

Basic principles of procedural and distributive justice require that AIMDs treat subgroups within a population fairly. It is well established, however, that machine learning models in medicine can exhibit race, sex, or socioeconomic biases (35,36). Many of these biases may be encoded before deployment (37), but even carefully trained models can create unfairness when they are inappropriately deployed (i.e., in out-of-sample populations or by ignoring shifts in the deployment population). In this context, users and administrators are obligated to ensure that the deployment of AIMDs is procedurally fair and promotes distributive fairness.

*Procedural Fairness.* Procedural fairness requires that patients be treated with equal consideration, regardless of their race, sex, religion, or other protected characteristics (12). Unfortunately, there is some evidence that clinicians (both in nuclear medicine and in other specialties) have often failed to live up to this requirement with respect to the provision of medical imaging (38). Although structural barriers mean that equal access does not ensure equal opportunity to benefit from AIMD, procedural fairness requires that the medically indicated use of AIMDs be offered to patients regardless of their protected characteristics.

Procedural fairness also discourages the use of a patient’s protected characteristics as an input feature. Historically, some medical decision-making tools have used features such as patient race or ethnicity as direct inputs (39,40), informed by ill-conceived genetic and biologic understandings of race (41). For instance, a common breast-cancer screening tool uses race alongside family history, age, and the Breast Imaging Reporting and Data System breast density score (42), with the result that it may underestimate risk in nonwhite patients (40). Although some now argue that race can act as a proxy for the influence of racist oppression on patient’s health, naive attempts to try to correct systemic bias may unintentionally introduce biases of their own (43). For instance, it can impose tradeoffs on marginalized groups: by using features such as race to improve the accuracy of risk predictions, we may reduce access to desired treatments or interventions for that
group (39). Moreover, the capture of demographic information (e.g., through self-identification or physician assessment) is imprecise and can contribute to the reification of stereotypes among clinicians about patient risk. As our understanding of the link between health, race, and other socially salient attributes improves, we believe that the use of these categories to make clinical decisions requires careful justification. These justifications should include robust knowledge of the data sources for demographic information, the causal structure of the association between the attribute and health, and the effect on marginalized people of using socially salient attributes to make clinical decisions.

Distributive Fairness. Another potential concern is the accuracy of AIMDs by ethnicity, socioeconomic status, or sex. Encouragingly, there has been much recent work on technical methods to remove bias from AIMDs (37). In a companion paper (11), we will note that a variety of techniques during the data collection and development phase can help ameliorate biases introduced during the training phase. Nonetheless, these techniques are not sufficient to ensure that AIMDs reduce health disparities, for at least 2 reasons.

First, the impact of an AIMD on disparities will be dependent on its precise pattern of accessibility. There are well-documented inequities in structural access to medical imaging for coronavirus disease 2019 diagnostics, mammography, and lung cancer screening (8). The cost of AI implementation is not only limited to the cost of developing or purchasing the AIMD but also includes supportive technology infrastructure, training staff, and patient education. Careful assessment of whether AIMDs should be adopted is especially critical for low-income or rural areas, where implementing expensive AI technology may divert resources from lower-tech interventions, with a greater impact on patient outcomes (44).

Second, even if an AIMD is equally accurate and accessible for all demographic subgroups, it may exacerbate existing structural inequalities (45). For instance, an AIMD designed to schedule future appointments based on a no-show predictor may schedule patients with a history of missed appointments to overbooked days. Of course, missed appointments are often related to a patient’s structural determinants of health—an inability to cover transportation costs or childcare or to take time off from work—and hence the very patients who may need additional care must now experience longer wait times or overbooked clinics. Thus, even if the AIMD predicts absenteeism correctly regardless of race or socioeconomic status, its deployment may widen health disparities due to background structural injustice.

GOVERNANCE OF AIMD

Alongside awareness of the ethical considerations of clinical deployment, a sustainable governance framework for AIMD is necessary to ensure its potential is realized. In this section, we will discuss the need for clear and effective performance claims and the appropriate way to navigate the assignment of ethical responsibility for the use of AIMDs.

Ensuring Safety and Efficacy

To discharge their ethical duties, physicians must have reliable information about the task-specific performance, safety, and functional limits of AIMDs. In a prior paper, we note that AIMDs should be evaluated by robust trials on clinical tasks, not just retrospective datasets, before deployment (15). Three additional considerations arise for regulators seeking to evaluate the safety and efficacy of an AIMD.

First, the safety and efficacy of an AIMD may be relative to the deployment environment. Consider that an algorithm trained in a high-resource context with a predominately high-income population may not have the same accuracy in other patient populations or clinical contexts. If the AIMD is deployed in a lower-resource context, it may expose marginalized patients to harm through misdiagnosis, inappropriate treatment recommendations, or misdirected therapy. At minimum, the deployment of AIMDs to environments with more limited resources should be done with caution to avoid situations in which patients receive inaccurate diagnoses or treatment recommendations. More expansively, if health equity is of overriding importance, regulators might require demonstrations of efficacy in both low-resource and high-resource contexts before deployment anywhere.

Second, even after initial clinical evaluation, performance testing is an unfinished project. Postdeployment evaluation is essential for maintaining the trustworthiness of a system and for ensuring that the AI continues to perform as expected not only when changes are made to the software but also when populations, diseases, and clinical ecosystems change over time (18). Tools such as algorithmic audits can help support continuous monitoring of AIMD performance after deployment (46).

Finally, regulators should be attentive to the fact that, once certified as efficacious and safe, AIMD solutions may be deployed as a replacement for physician-directed care. AIMDs can potentially assist when there may otherwise be no available resources. For instance, in a region without an on-call trained nuclear medicine physician, an AI assistant could notify the primary care team and generate a preliminary report for unexpected urgent findings, such as pneumothorax in an oncologic PET/CT examination, while waiting for evaluation by a qualified physician. Although this ability appears to improve access to high-quality care, reliance on AIMD decisions by untrained clinicians may lead to inappropriate diagnosis or treatment that may be worse than delayed care.

Supportive Infrastructure

The performance of an AIMD may have less to do with the model itself and more to do with the ecosystem around the AIMD. The successful deployment of an AIMD in nuclear medicine thus requires key supportive investments.

First, data and standards are needed to support the development, validation, and testing of AIMDs. Relevant and reliable training data, comprehensive test sets (26), standardized evaluation and validation methods (47), and comprehensive imaging archive repositories are crucial to fair, efficacious, and reliable development of AIMDs. Some of this can be achieved by professional self-governance. Cognate professional societies in nuclear medicine, radiology, and medical imaging should consider collaborative efforts to develop a central repository for developing standards and sharing reliable datasets for AIMD research and development.

Second, the deployment of AIMDs may require substantial investment in supportive infrastructure by hospitals and public agencies. The digital divide prevails in lower-resource contexts and results in uneven access to, or efficacious use of, medical imaging technologies due to a lack of basic information technology, bioinformatics, and database support (48). Efficacious AIMD deployment may require vastly increased access to remote and telemedicine services, as well as access to reliable power and Internet service in many parts of the globe (49).

Third, the deployment of AIMD requires appropriate training and policies. Too often, especially in lower-resource health
systems, new technologies are not used or are used inappropriately, thereby wasting scarce resources that could have been invested in simpler, proven interventions (50). Moreover, the use of AIMDs without up-to-date patient privacy or data protection policies (or appropriate regulators to enforce these policies) can compound, rather than remediate, the harm of the digital divide in medicine.

**Legal and Regulatory Oversight**

Establishing responsibility and accountability pathways is crucial to maintain trust, respect legal limits, and protect human rights (49). We identify 3 problems for regulators seeking to build community trust in the use of AIMDs within diagnostic radiology and nuclear medicine.

First, assigning legal liability for the harm generated by AIMD may be difficult because of the so-called responsibility gap generated by systems that automate some components of tasks previously supervised by humans. Clinicians are ethically responsible for the use of both CADe and CADx systems, but this does not absolve developers and administrators of responsibility. If a systematic error occurs because of inaccurate performance claims, obscured limitations, or failures of the system in intended-use cases, responsibility should be placed on those who trained, tested, and validated the AI device. The assignment of legal liability for harm is made more complex by the evolving legal status of AIMDs as regulated medical devices in the United States and Europe (51,52) and the unsettled requirements for premarket and postmarket disclosure. As we build a regulatory structure around AIMDs, the locus of liability for harm must be proactively addressed by lawmakers and regulators.

Second, AI software is usually proprietary, and external researchers typically have little or no access to training data or performance evaluations. Ensuring that developers meet their obligations with respect to performance transparency and fairness may thus require independent auditors, researchers, or government agencies to have access to underlying models and performance info (53). This may build trust by ensuring there is a public mechanism for holding AIMDs accountable, reliable, and consistent (54).

Third, the task of oversight and regulation of AI in health care may not be feasible through a single regulatory agency. The responsibilities of most current agencies in the Department of Health and Human Services are narrow. For instance, the Federal Trade Commission aims to prevent anticompetitive harm, the Food and Drug Administration ensures the safety and efficacy of devices, and the Centers for Medicare and Medicaid Services administer and regulate payment for health-care services (55). AIMD implementation within nuclear medicine involves all of these regulatory areas. As a result, there may be either a need to create a formal method for these agencies to collaborate or a need to form a new agency to regulate the safety, competition, and ethical aspects of AI-based devices.

**CONCLUSION**

There is undoubtedly enormous potential in the use of AI tools and software in medical imaging. Appropriate implementation of these technologies can not only increase efficiency and accuracy but also reduce the burden on clinicians and narrow health inequality gaps. This paper aimed to anticipate potential ethical considerations for widespread rapidly evolving AIMDs in medical imaging. Viewing these considerations through the lens of health disparities permits us to conceive of potential harm to certain groups in the population and protect against them.

**DISCLOSURE**

Melissa McCradden acknowledges funding from the SickKids Foundation pertaining to her role as the John and Melinda Thompson Director of AI in Medicine at the Hospital for Sick Children. Abhinav Jha acknowledges support from NIH R01EB031051-02S1. Peter Scott also acknowledges support from the National Institutes of Health (R01EB0211555). Sven Zuehlsdorff is a full-time employee of Siemens Medical Solutions USA, Inc. No other potential conflict of interest relevant to this article was reported.

**ACKNOWLEDGMENT**

The members of the task force acknowledge Bonnie Clarke for her support.

**REFERENCES**

16. IMDRF Software as a Medical Device (SaMD) Working Group. "Software as a Medical Device": Possible Framework for Risk Categorization and Corresponding Considerations. International Medical Device Regulators Forum; 2014;14.