The IRB As a Guardrail for Healthcare AI

Artificial intelligence (AI) in healthcare has significant promise, potentially allowing earlier disease detection, expanded access to care, and reduced administrative burden on clinicians. For Catholic and other mission-driven healthcare systems, this seems providential: technology that advances access for the poor and vulnerable and improves care for all. Yet medical breakthroughs have often produced novel moral hazards. AI is no exception.

The moral questions raised by AI are extensive, and should be extensively examined. They touch on privacy, bias, responsibility, and the shifting role of clinicians. They also reach deeper, into questions of dignity, justice, and the moral character of institutions and societies. Catholic healthcare, with its long tradition of ethical reflection, is well positioned to examine these risks. However, lofty principles alone are not enough. Best practices for ethically implementing healthcare AI require incorporating ethics input throughout the lifecycle of development and implementation of AI in hospitals and clinics.[1] As AI tools move from design to deployment, they encounter a crucial checkpoint: the Institutional Review Board (IRB).[2] We outline our experiences with early review of healthcare AI at the IRB level to explore how IRBs can be positioned to respond to the novel challenge of implementing AI at the bedside.

IRBs protect human subjects in research, a checkpoint for ensuring that study design is appropriate. Today, they face a new challenge: evaluating how healthcare AI will be developed and tested. This means tracking risks to study subjects from opaque algorithms, weighing benefits, and reviewing unanticipated events. IRB evaluation of AI studies will represent a re-review of the ethical principles involved in the development of the goals of AI and a targeted examination of impacts of integrating AI into a healthcare setting. The IRB sits at a "mid-level" of oversight—between the high-level ethical principles and regulation of clinical realities. Here, commitments to justice and dignity become concrete questions: What risks does this tool pose? Who might be harmed? Whose voices are missing?

In this sense, an IRB effectively tuned to evaluate the novel moral issues raised by healthcare AI is not merely a bureaucratic hurdle but serves as a moral forum. An IRB well-prepared to address AI can be a space where hype and skepticism are balanced. When the future of healthcare is being so strongly impacted by the integration of AI, an IRB can be a place where this is debated in miniature.

CHALLENGES IN AI REVIEW

Our experience with our IRB's first encounters with healthcare AI highlights the challenges others may encounter. Panelists steeped in

research ethics and medicine felt ill-equipped to evaluate technical aspects of AI. They asked, how does one assess risk when errors are probabilistic rather than deterministic? How can we evaluate implementation risks without the ability to predict possible mistakes? How confident can we be in the accuracy of performance evaluations of evolving algorithms? Panelists researched AI error sources and outcome predictions, but reported research was difficult because of the hype and misinformation surrounding AI. In addition to questions about how AI works, panelists had questions about the ethical considerations relevant to data use, AI design, and AI implementation. Panelists asked ethicists to provide summaries of ethical frameworks for AI design and implementation.

Researchers, especially those steeped in the culture of AI development rather than healthcare, too, had questions about IRB review. Some questioned whether IRB oversight applied to their work. They viewed AI implementation as quality improvement rather than research. Others equated AI ethics with data governance, overlooking broader questions of risk, justice, and responsibility.[3] Researchers unfamiliar with IRB guidance did not initially provide complete responses to IRB questions about potential harm or privacy, and were surprised by IRB follow-up questions. These education gaps left AI-interested researchers uneasy about engaging with the IRB structure. This experience, if repeated, might lead to a "cooling" of researchers' ability to reach out to IRB staff with questions about study design, or a hesitance for AI researchers to develop healthcare AI and engage with an unfamiliar regulatory structure.

These potential tension-points can be proactively addressed with education, offering IRB panelists and researchers alike guides for the unfamiliar space of IRB reviewing healthcare AI. IRB members should have training in the operation, risks, and ethical pitfalls of AI. One-pagers for IRB members to review, supplemental specialized education courses, or presentations from experts on AI design and ethics with Q&A may all be helpful. Researchers in the AI space, in turn, need grounding in the history and expectations of medical research ethics. Orientations and onboarding to these as well as technical IRB requirements could similarly be provided. Ethicists and those engaged in IRB development can develop these materials and organize this education before their institutions begin to review healthcare AI research, to avoid miscommunications.[4]

OPPORTUNITIES IN AI REVIEW

Tension and novel moral hazards are inevitable with major change, but IRBs are uniquely suited to respond to these. Research ethics culture is familiar with hype: IRB panelists reference memories of earlier “miracle cures” to temper more outlandish claims surrounding AI. IRBs are designed to ask scientific and ethical questions in moments of innovation and uncertainty, and to integrate existing guidelines for the ethics of particular treatments. The IRB’s ability to question, identify gaps, integrate new ethical principles, and to resist over-optimism or cynicism are invaluable in the rapidly changing world of healthcare AI.[5]

Moreover, the diversity of perspectives and on-the-ground experience with integration of new technology in healthcare on IRBs

proved fruitful. Clinicians, ethicists, and researchers raised questions about clinician work environments, data disclosure, and the pragmatics of implementation. These were concerns that had been overlooked. Panelists shared experiences with past changes in healthcare delivery and also raised questions for researchers about feasibility in study design based on this history. IRB panelists can have the research ethics resources to hold healthcare AI to realistic and firm standards, and AI researchers can find partners with novel perspectives and experience in innovative research. Panelists also reported being more optimistic about healthcare AI in their own practice following the opportunity to research for an IRB review.

In our Catholic healthcare context, the shared language of mission and dignity further enriched these discussions. Panelists were impressed by resources like the Rome Call for AI Ethics, giving language to the societal and spiritual hazards of AI use.[6] The Catholic philosophical and moral tradition on AI and the more secular tradition of AI ethics are evolving quickly, which will provide healthcare institutions with the resources to think through broad questions about AI.[7]

CONCLUSION

AI will transform healthcare, but this is not healthcare's first transformation. Healthcare innovation has long brought ethicists, clinicians, and institutions new challenges, revealing the enduring need for moral guidance. Properly equipped, IRBs can be pillars of this guidance. For Catholic healthcare, this is not merely a technical task but a moral vocation: shaping medicine to honor dignity and serve the vulnerable. At its best, an IRB is not a hoop for healthcare AI research to jump through or an outdated committee whose importance is waning in the world of AI, but a place where the ethical, clinical, and cultural dimensions of AI are tested, and where the story of AI in healthcare can be co-written as one of innovation as well as justice. ✠

SARA KOLMES, PH.D., HEC-C
Senior Clinical Ethicist
Providence Center for Healthcare Ethics
Eastern Washington/Montana

KAYLA TABARI, RN, MBE, HEC-C
Clinical Ethicist and Fellowship Supervisor
Providence Center for Healthcare Ethics
Oregon

ENDNOTES

1. Barina, Ryan, et al. "Data Ethics in Catholic Health Systems." National Catholic Bioethics Quarterly 22, no. 4 (2022)., Samuel, G., Derrick, G. E., & van Leeuwen, T. (2019). The ethics ecosystem: Personal ethics, network governance and regulating actors governing the use of social media research data. Minerva, 57(3), 317–343. https://doi.org/10.1007/s11024-019-09368-, Sankar, Baradwaj Simha, et al. "Building an ethical and trustworthy biomedical AI ecosystem for the translational and clinical integration of foundation models." Bioengineering 11.10 (2024): 984., Herzog, Christian, Sabrina Blank, and Bernd Carsten Stahl. "Towards trustworthy medical AI ecosystems–a proposal for supporting responsible innovation practices in AI-based medical innovation." AI & SOCIETY 40.4 (2025): 2119-2139., Waeiss, Quinn, and Mildred K. Cho. "An ecosystem approach to governing commercial actors in healthcare AI." Policy Studies (2025): 1-14., Eisenhauer, Lisa, "CommonSpirit

reviews clinical, ethical aspects in data and AI decision-making" *Catholic Health World* (2025).
2. Knight, Simon, Antonette Shibani, and Nicole Vincent. "Ethical AI governance: mapping a research ecosystem." AI and Ethics 5.2 (2025): 841-862., Nebeker, Camille, John Torous, and Rebecca J. Bartlett Ellis. "Building the case for actionable ethics in digital health research supported by artificial intelligence." BMC medicine 17.1 (2019): 137.
3. Bouhouita-Guermech, S., et al. "Specific Challenges Posed by Artificial Intelligence in Research Ethics." *Frontiers in Artificial Intelligence* 6 (2023).
4. Samuel, Gabrielle, et al. "Boundaries Between Research Ethics and Ethical Research Use in Artificial Intelligence Health Research." Journal of Empirical Research on Human Research Ethics 16, no. 3 (2021)., Li, Fan, Nick Ruijs, and Yuan Lu. "Ethics & AI: A systematic review on ethical concerns and related strategies for designing with AI in healthcare." Ai 4.1 (2022): 28-53.
5. Obermeyer, Ziad, et al. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." Science 366, no. 6464 (2019): 447–453., Bouhouita-Guermech, S., et al. "Specific Challenges Posed by Artificial Intelligence in Research Ethics." Frontiers in Artificial Intelligence 6 (2023), Samuel, Gabrielle, et al. "Boundaries Between Research Ethics and Ethical Research Use in Artificial Intelligence Health Research." Journal of Empirical Research on Human Research Ethics 16, no. 3 (2021).
6. Pontifical Academy for Life. Rome Call for AI Ethics. Vatican City: Vatican Press, 2020.
7. Umbrello, Steven. "Pope Leo XIV and the New Social Questions of Artificial Intelligence." *The Linacre Quarterly* 92.3 (2025): 354-357., Kockler, Nick, "Generating Insights from Catholic Social Teaching: Ethical Guidelines for Artificial Intelligence in Healthcare Ministries", *Healthcare Ethics USA* 17.2 (2023), Minda, Julie, "Ethicists have critical role to play as artificial intelligence use increases, says webinar presenter", *Catholic Health World* (2025), Laracy, Joseph R., et al. "Human Dignity and the Ethics of Artificial Intelligence: A Framework for Responsible Design and Use from the Perspective of Catholic Social Teaching." 2025 *IEEE International Symposium on Ethics in Engineering, Science, and Technology (ETHICS)*. IEEE, 2025., Nasir, Sidra, Rizwan Ahmed Khan, and Samita Bai. "Ethical framework for harnessing the power of AI in healthcare and beyond." IEEE Access 12 (2024): 31014-31035., Jobin, Anna, Marcello Ienca, and Effy Vayena. "The global landscape of AI ethics guidelines." *Nature machine intelligence* 1.9 (2019): 389-399., Hagendorff, Thilo. "The ethics of AI ethics: An evaluation of guidelines." *Minds and machines* 30.1 (2020): 99-120.