AAHRPP
Association for the Accreditation of Human Research Protection Programs, Inc.®

2025 AAHRPP Webinar Series

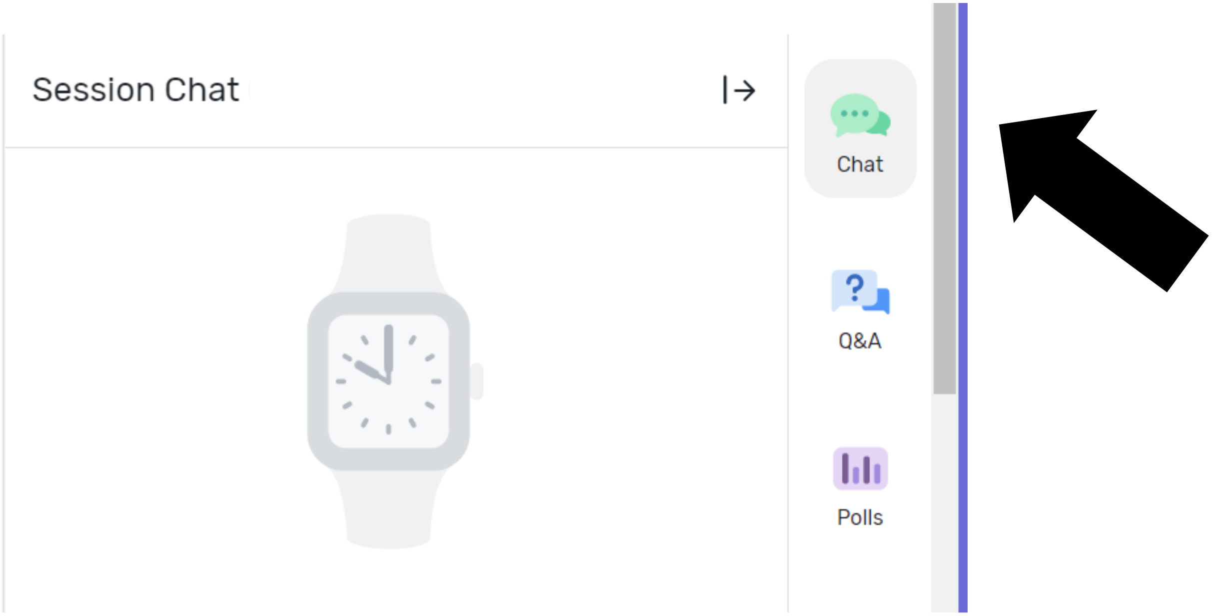
A 2025 “HRPP Innovations” Webinar:

Should We Use AI to Support the IRB Review Process? It Depends (of course!)

October 1, 2025; 1:00 pm – 2:30 pm ET

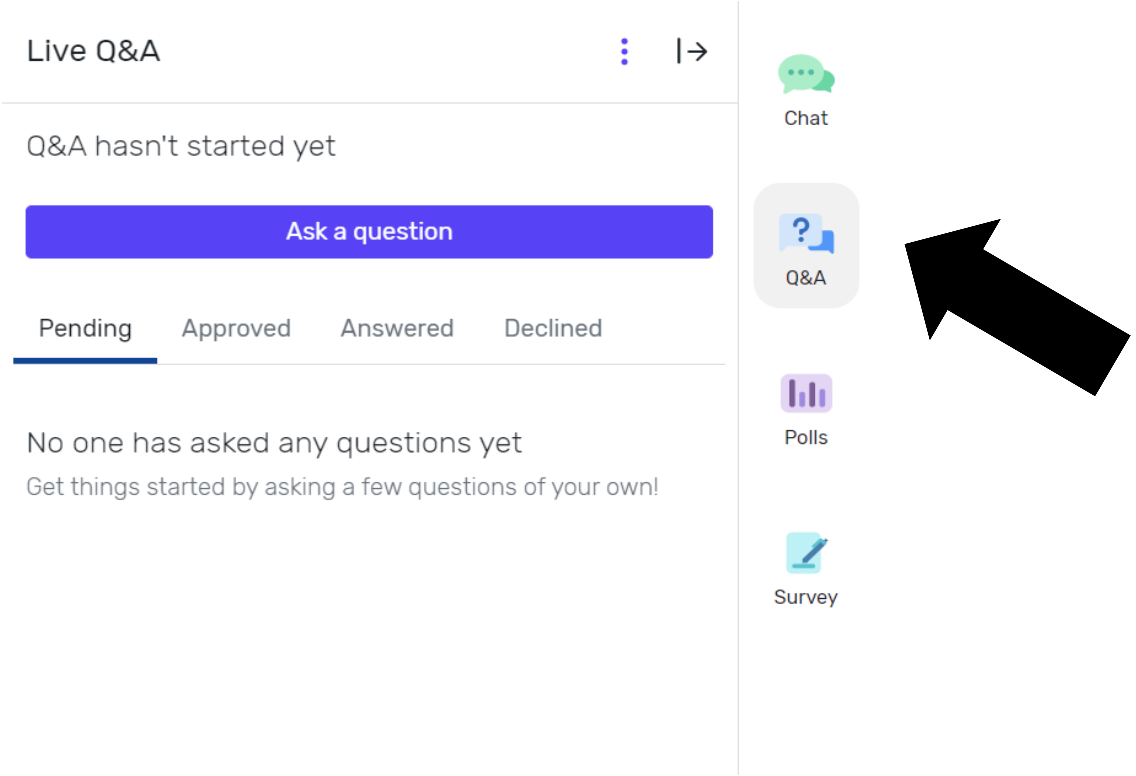
Chat Feature

To chat with your colleagues before and after the session, or if you have technical questions, use the "Chat" icon



Questions

To ask questions about the topic for the presenters, please use the “Q&A” icon:



AAHRPP Association for the Accreditation of Human Research Protection Programs, Inc.®

Upcoming Webinars

Ask... AAHRPP®

Save these dates for the remaining 2025 "Ask AAHRPP" webinars:

- October 14, 2025
- December 9, 2025

AAHRPP Association for the Accreditation of Human Research Protection Programs, Inc.®

HRPP Innovations

Webinar Series

Save these dates for the remaining 2025 “HRPP Innovations" webinars:

- November 2025 - TBD

Visit [Webinars (aahrpp.org)] for more information and registration links

Visit AAHRPP's Annual Conference page for more information

Presenter Introductions

AAHRPP Association for the Accreditation of Human Research Protection Programs, Inc.®

Mohammad Hosseini

Northwestern University

Yefrenia Henriquez Taveras

Teachers College, Columbia University

Swapnali Chaudhari
Rutgers University

AAHRPP
Association for the Accreditation of Human Research Protection Programs, Inc.®

Tonya Ferraro

AAHRPP

Northwestern Medicine®
Feinberg School of Medicine

Ethical Aspects of Using AI in Preparing or Reviewing IRB Applications

Mohammad Hosseini, Oct 1, 2025

Funding acknowledgments

- Office of Data Science Strategy (3OT2DB000013-01S1)
- National Center for Advancing Translational Sciences (5UL1TR001422-08)
- Network of the National Library of Medicine Evaluation Center (U24 LM013751)

Ideas and opinions expressed here do not represent the views of the ODSS, NCATS, NLM, NIH or the US government.

House Keeping

- You have my consent to take photos, use quotes and tag me online if you find anything useful
  - LinkedIn: hosseinimohammad
  - Bluesky: @mhmdhsini
- Introduced apps or studies are not endorsed

Use of AI in Preparing IRB Applications

- Detailed and time-consuming process

ELSEVIER

SoftwareX

Volume 25, February 2024, 101601

Original Software Publication

IRB-draft-generator: A generative AI tool to streamline the creation of institutional review board applications

Ryan C. Godwin [a] [b], Ayesha S. Bryant [a], Brant M. Wagener [a], Timothy J. Ness [a], Jennifer J. DeBerry [a], LaShun L. Horn [a], Shanna H. Graves [a], Ashley C. Archer [c], Ryan L. Melvin [a]

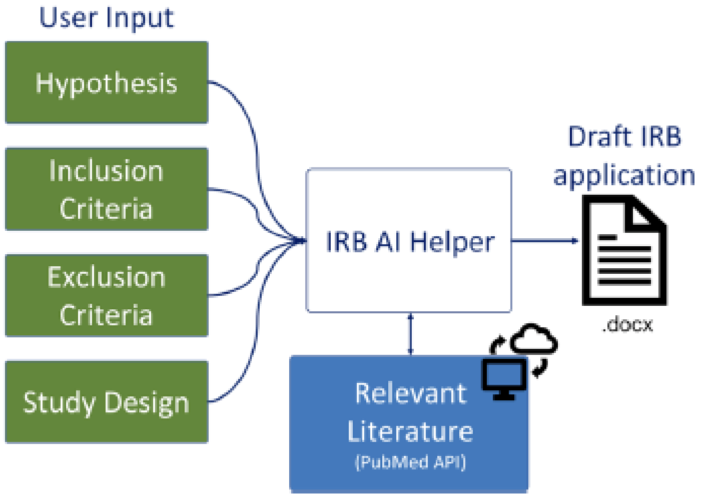
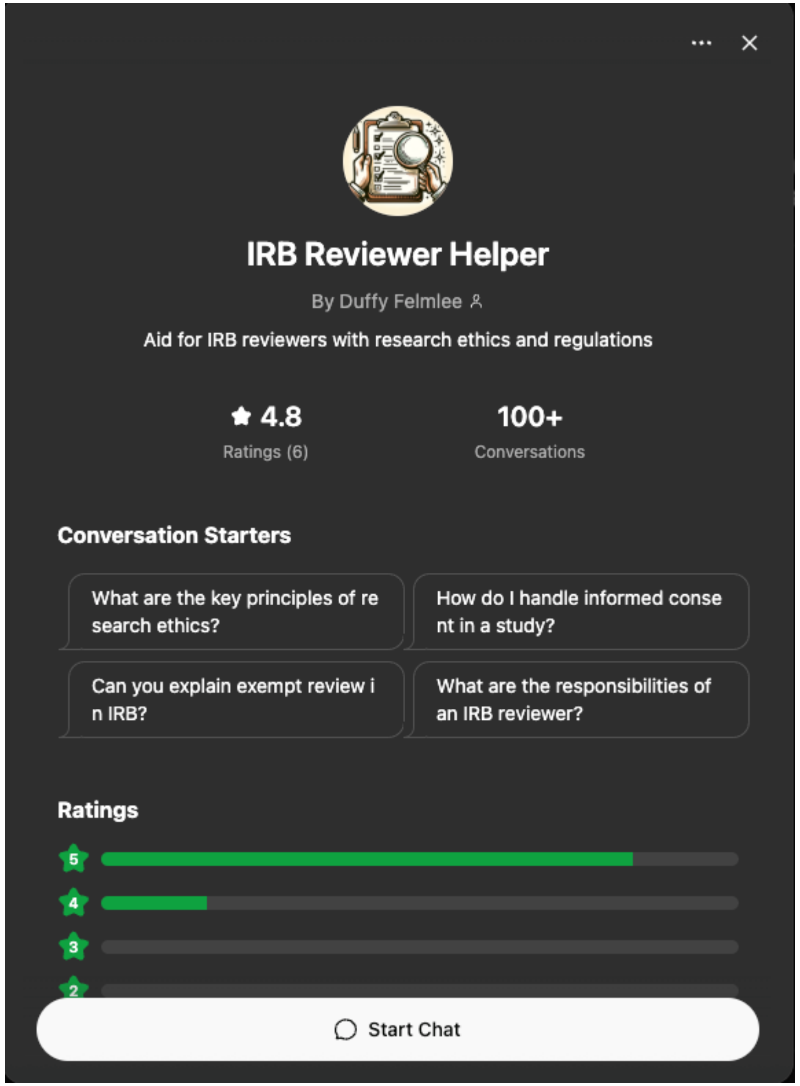
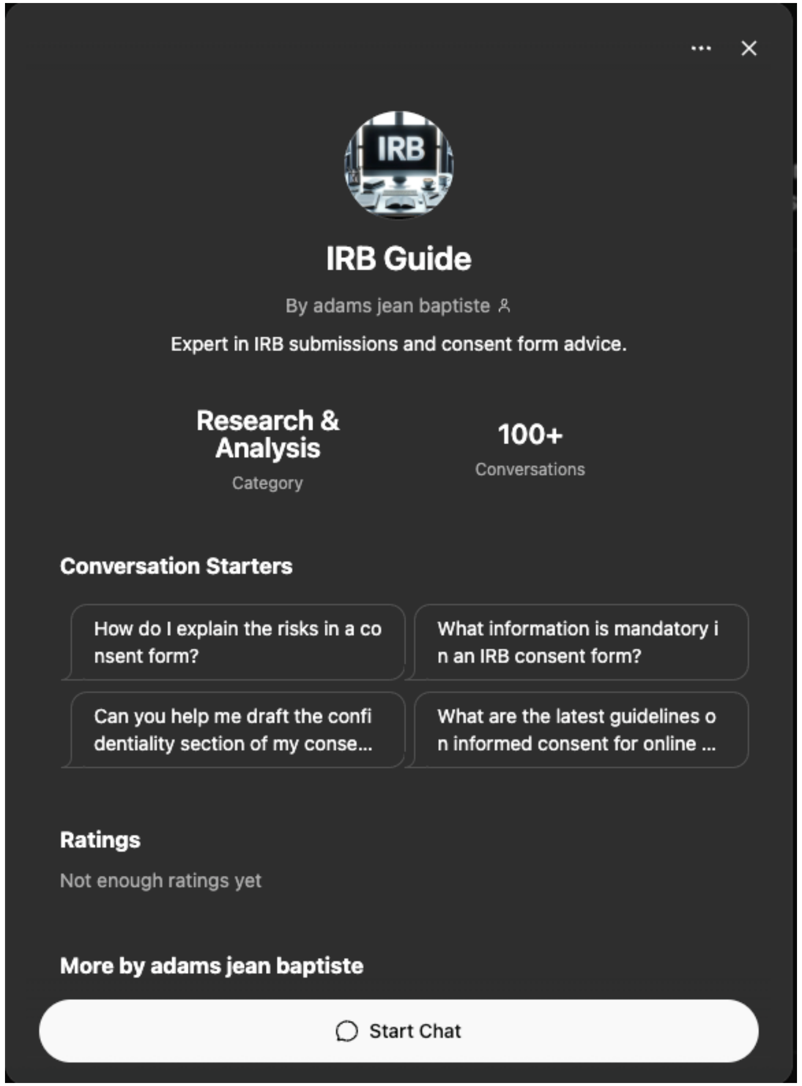
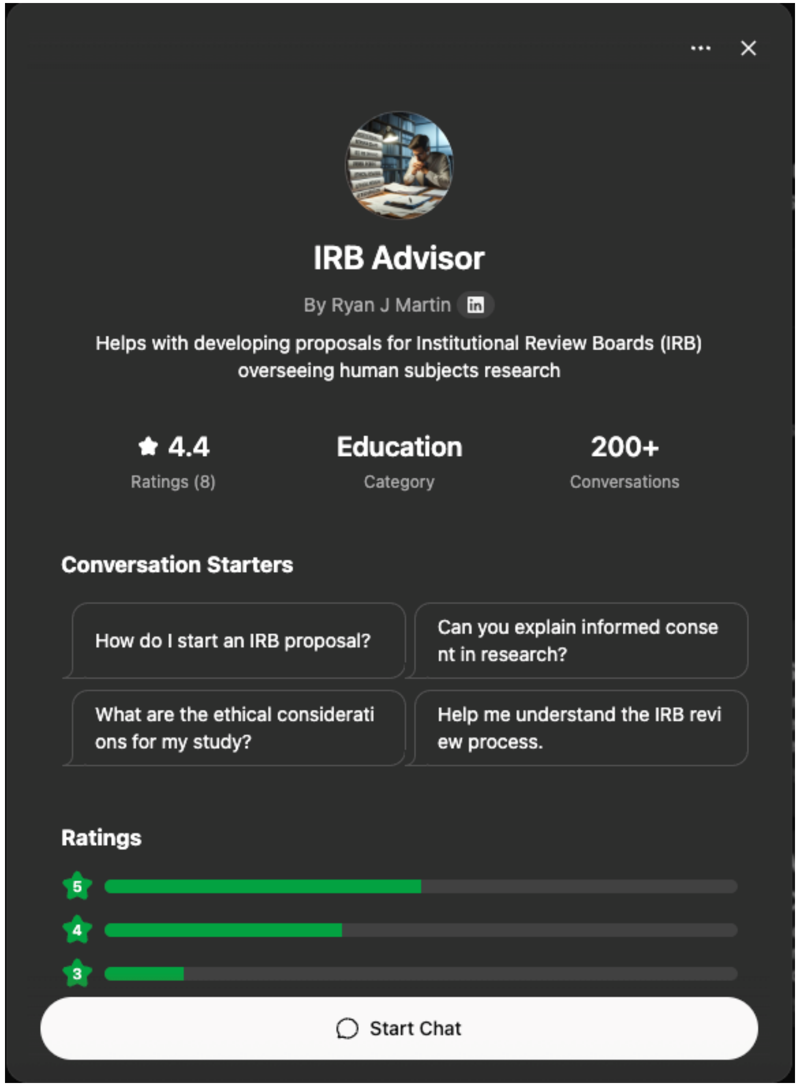


Fig. 1. The Workflow of the IRB Draft Generator includes 4 user inputs, the literature search tool, the generative AI of the IRB Helper and the output document. The IRB Helper uses the user input to generate a relevant literature search query which is fed to a literature search tool that interfaces directly to PubMed.

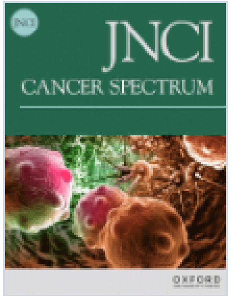
Use of AI in Preparing IRB Applications







Use of AI in Preparing/Enhancing Consent Forms



Volume 9, Issue 2
April 2025

JOURNAL ARTICLE

AI meets informed consent: a new era for clinical trial communication

Michael Waters, MD, PhD

JNCI Cancer Spectrum, Volume 9, Issue 2, April 2025, pkaf028,
https://doi.org/10.1093/jncics/pkaf028

Published: 18 March 2025 **Article history** ▾

Recruiting

Evaluating AI-Generated Plain Language Summaries on Patient Comprehension of Ophthalmology Notes Among English-Speaking Patients

ClinicalTrials.gov ID NCT06859216

Sponsor University of California, Los Angeles

Information provided by Prashant Tailor, University of California, Los Angeles (Responsible Party)

Last Update Posted 2025-03-05

Use of AI in Reviewing IRB Applications

Extended essay

PDF

Chat-IRB? How application-specific language models can enhance research ethics review

Sebastian Porsdam Mann [1, 2, 3], Jiehao Joel Seah [3], Stephen Latham [4], Julian Savulescu [3, 5], Mateo Aboy [6], Brian D Earp [3, 5]

Correspondence to Professor Julian Savulescu; julian.savulescu@uehiro.ox.ac.uk

Abstract

Institutional review boards (IRBs) play a crucial role in ensuring the ethical conduct of human subjects research, but face challenges including inconsistency, delays, and inefficiencies. We propose the development and implementation of application-specific large language models (LLMs) to facilitate IRB review processes. These IRB-specific LLMs would be fine-tuned on IRB-specific literature and institutional datasets, and equipped with retrieval capabilities to access up-to-date, context-relevant information. We outline potential applications, including pre-review screening, preliminary analysis, consistency checking, and decision support. While addressing concerns about accuracy, context sensitivity, and human oversight, we acknowledge remaining challenges such as over-reliance on artificial intelligence and the need for transparency. By enhancing the efficiency and quality of ethical review while maintaining human judgement in critical decisions, IRB-specific LLMs offer a promising tool to improve research oversight. We call for pilot studies to evaluate the feasibility and impact of this approach.

Journal of Empirical Research on Human Research Ethics

PRIM&R PUBLIC RESPONSIBILITY IN MEDICINE AND RESEARCH

Impact Factor: **1.1** / 5-Year Impact Factor: **1.7**

Journal Homepage

Restricted access | Research article | First published online June 17, 2024 | Request permissions

Assessing the Decision-Making Capabilities of Artificial Intelligence Platforms as Institutional Review Board Members

Kannan Sridharan and Gowri Sivaramakrishnan View all authors and affiliations

Volume 19, Issue 3 | https://doi.org/10.1177/15562646241263200

Contents | Get access | Cite | Share options | Information, rights and permissions | Metrics and citations

Abstract

Background: Institutional review boards (IRBs) face delays in reviewing research proposals, underscoring the need for optimized standard operating procedures (SOPs). This study assesses the abilities of three artificial intelligence (AI) platforms to address IRB challenges and draft essential SOPs. **Methods:** An observational study was conducted using three AI platforms in 10 case studies reflecting IRB functions, focusing on creating SOPs. The accuracy of the AI outputs was assessed against good clinical practice (GCP) guidelines. **Results:** The AI tools identified GCP issues, offered guidance on GCP violations, detected conflicts of interest and SOP deficiencies, recognized vulnerable populations, and suggested expedited review criteria. They also drafted SOPs with some differences. **Conclusion:** AI platforms could aid IRB decision-making and improve review efficiency. However, human oversight remains critical for ensuring the accuracy of AI-generated solutions.

Similar articl

Restrict

Underst
Enabler
Review:
Boards

Show D

Restrict

An Over

Ethical Issues of Using AI in Preparing and Reviewing IRB Applications

- Errors
- Biases
- Incomplete or inaccurate literature reviews
- Confidentiality
- Deskilling

IRB applications are not published
Attributions remain unclear

Conclusion

- AI cannot make value-based judgements
- Ethical planning, oversight and reasoning should rely on humans
- When is the right time to employ AI?
- Guidelines are needed

Association for the Accreditation of Human Research Protection Programs, Inc.®

2025 AAHRPP Webinar Series

AI in IRB Review: A Resource–Risk Spectrum

Yefrenia Henriquez Taveras, MPH, MHA, CCRP, CHES, CIP
QA & Education Specialist - Teachers College, Columbia University
IRB Compliance Coordinator – University of New England

Disclosure Statement

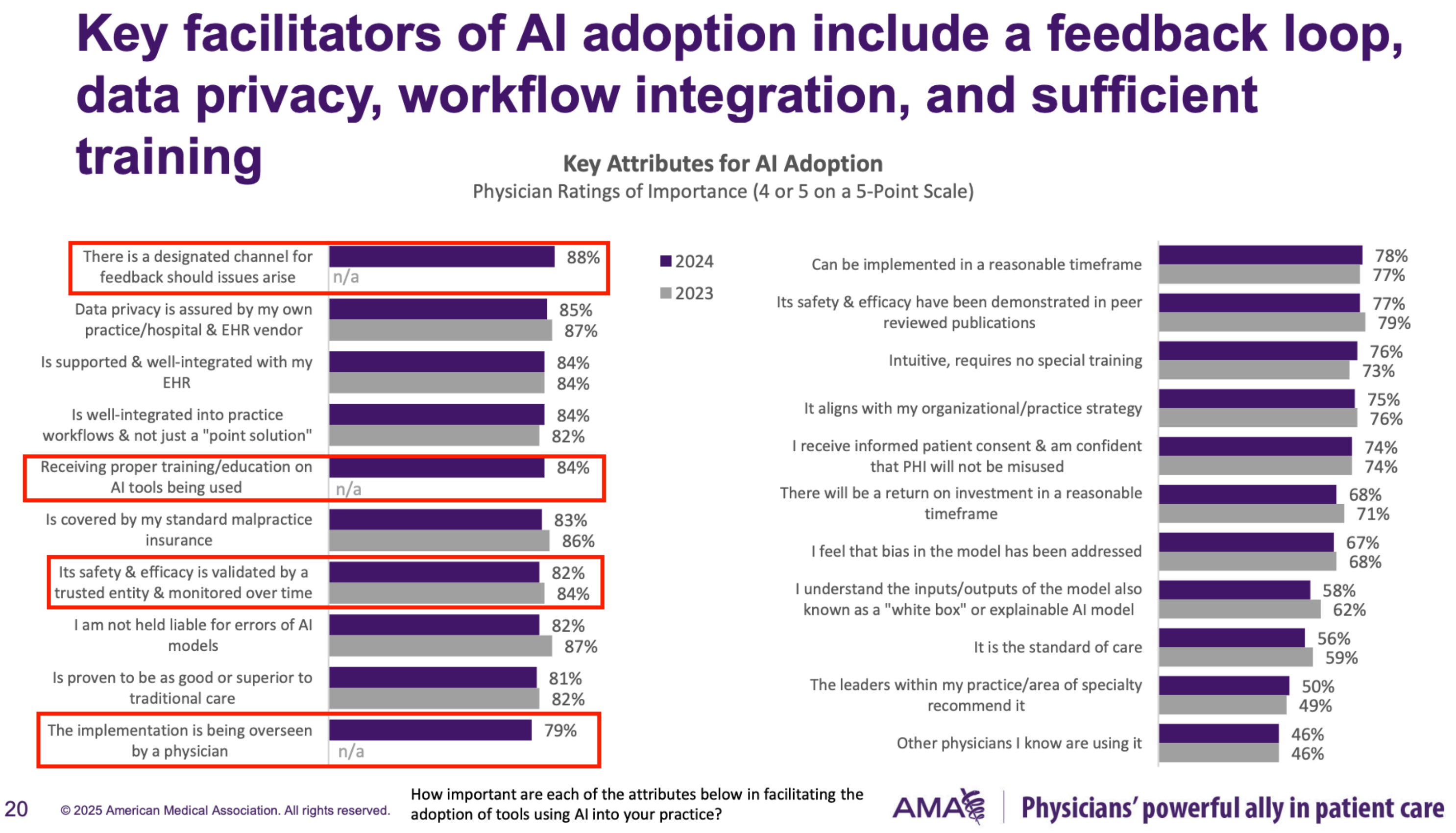
I have relevant professional relationship with respect to this educational activity with the following organization(s):

AAHRPP

Member, Board of Directors

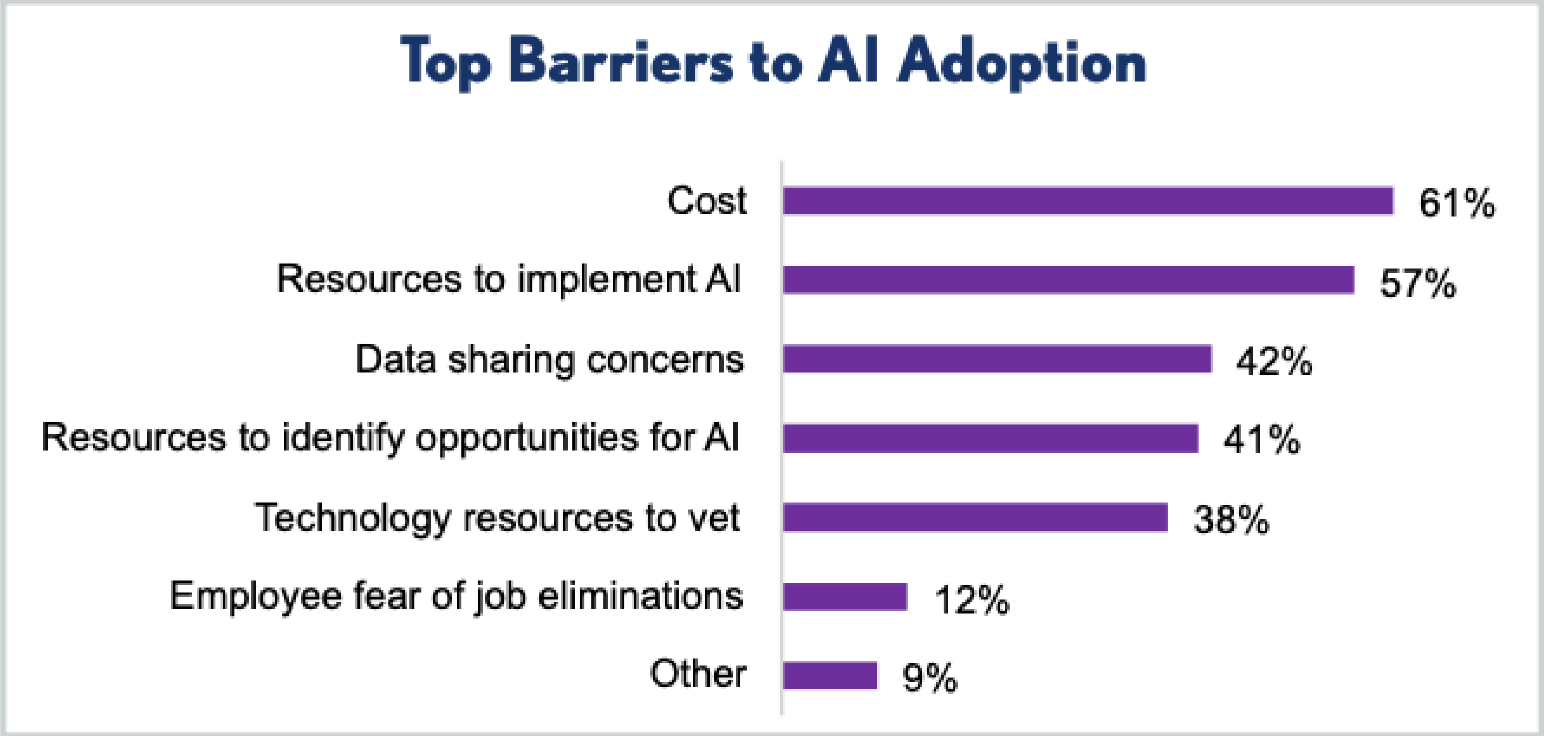
Responsible AI Requires Resources

- Feedback loops – issues reporting & monitoring (88%)
- Data privacy – secured integrations (85%)
- Training – ongoing education (84%)
- Validation & monitoring – proven safety & audits (82%)
- Human oversight – reviewer/physician time (79%)



➢ See AMA Physician AI Sentiment Report, 2025

System Readiness & Resource Divide



- See Health System Readiness for Artificial Intelligence

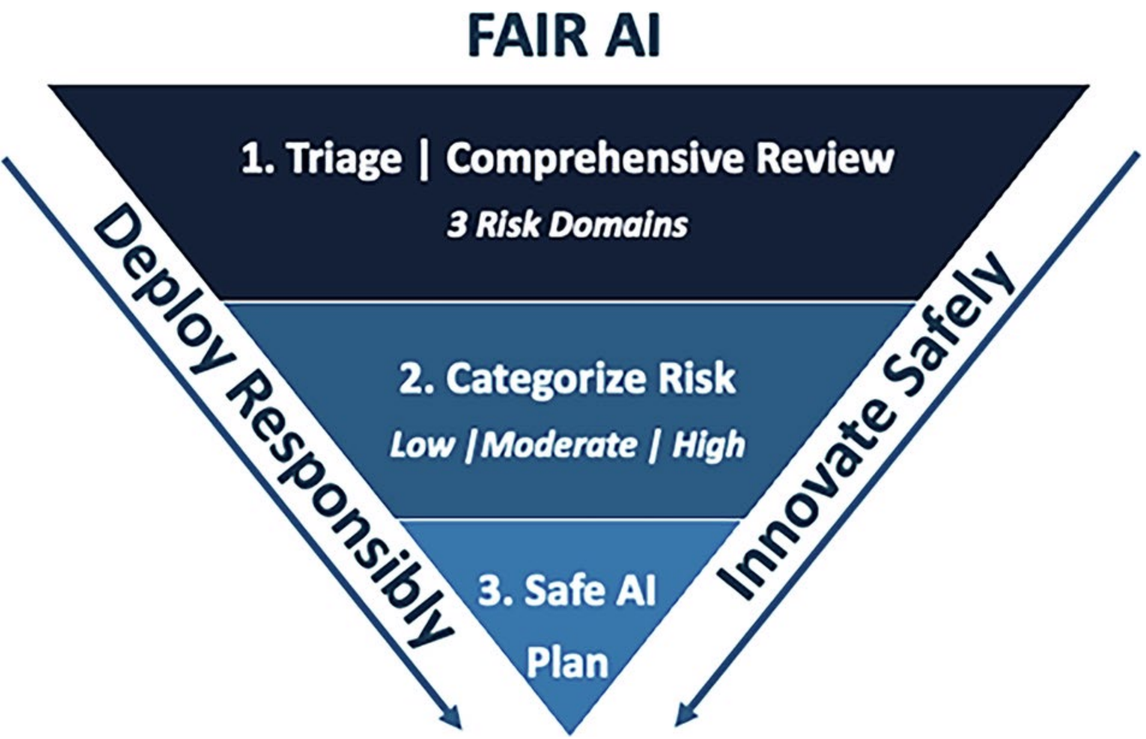
- See NTT Data: Global GenAI Report How organizations are mastering their GenAI destiny in 2025

What Does this Mean for IRBs?

- **More targeted, more controlled** -> Small, less resourced IRBs
- **Deeper integrations, with multi-layered safeguards** -> Larger, more resourced IRBs

AAHRPP Association for the Accreditation of Human Research Protection Programs, Inc.®

Institutional Risk Tolerance Drives AI Scope



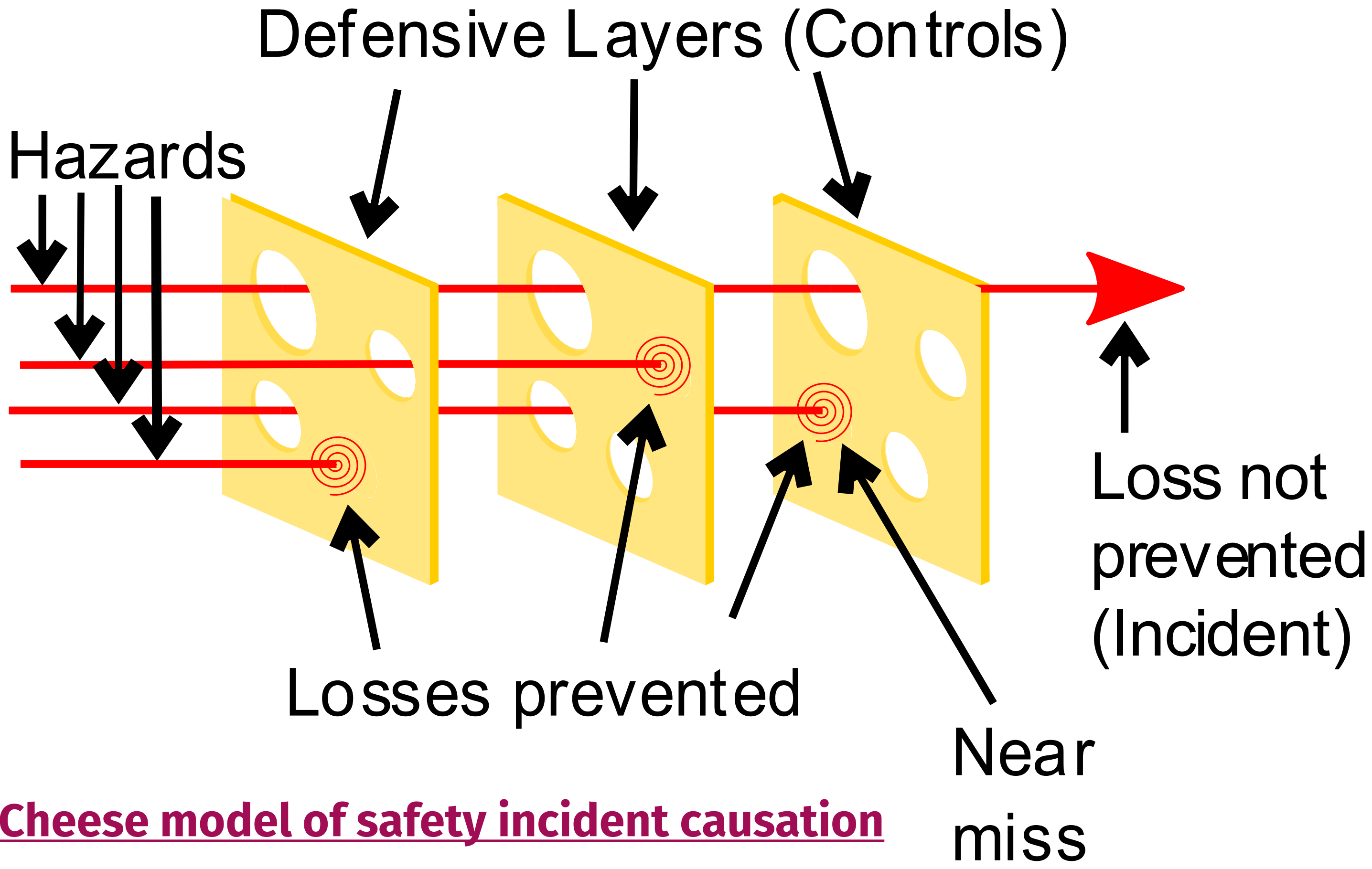
- **Conservative vs. permissive** interpretations of regulations determine whether AI stays in an advisory role or edges closer to decision support.
- Institutions with low appetite for reputational, legal, or accreditation risk will fence AI use tightly; **those with higher tolerance may pilot bolder applications**.
- **Appetite must be matched to capacity**—risk-tolerant adoption without monitoring resources creates disproportionate exposure.
- Leadership priorities, innovation incentives, and peer benchmarking often drive appetite as much as compliance obligations.

➢ **See A practical framework for appropriate implementation and review of artificial intelligence (FAIR-AI) in healthcare**

Association for the Accreditation of Human Research Protection Programs, Inc.®

Swiss Cheese Model: Borrowed from patient safety, Adapted to IRB Oversight

- **Layer 1: Researcher diligence**
  - Mandatory revision of AI drafts
  - Checklist confirmations
- **Layer 2: System safeguards**
  - AI flags inconsistencies, missing sections
  - Audit logs of AI use
- **Layer 3: IRB reviewer checks**
  - Human verification of risk/benefit
  - Focus on consent, vulnerable groups
- **Layer 4: Ongoing training & monitoring**
  - Training for researchers + IRBs
  - Policies on acceptable AI use
  - Ongoing QA/auditing



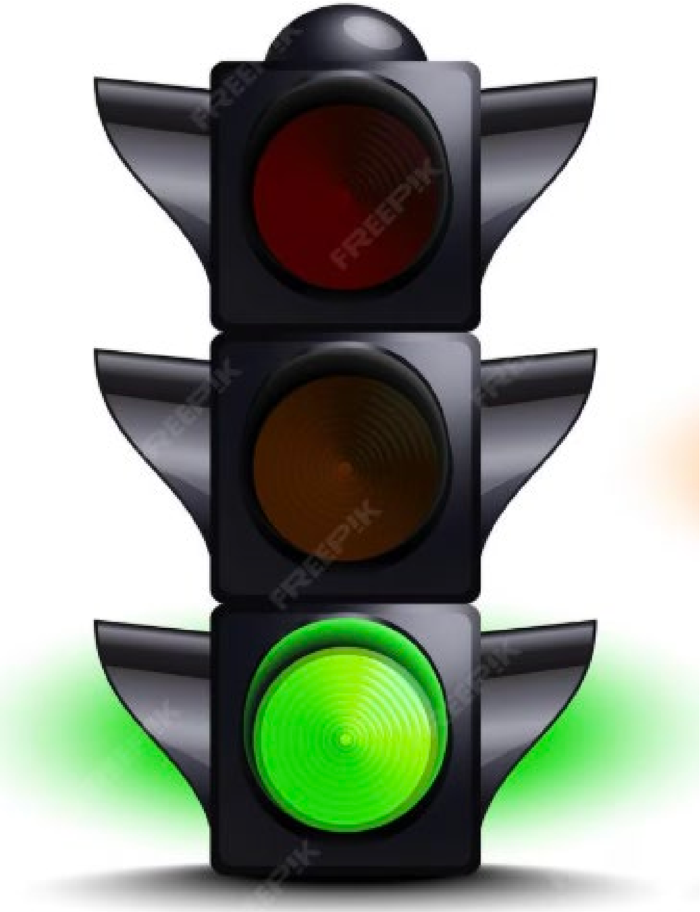
➢ **See The Swiss Cheese model of safety incident causation**

Scaling Oversight to Your Context

- AI use needs ongoing training, not one-time orientation.
- Define where AI assists vs. where human judgment is required.
- Treat AI like studies — continuous review and auditing.
- Oversight should match resources (dashboards vs. spot-checks).

Low-Resource, Low-Risk AI Use Cases

- FAQ chatbots trained only on official policies
- Basic document formatting checks
- Intake completeness scans
- Proactive education tools that walk investigators through requirements before submission



Pre-submission

Intake
(pre-review)

Medium-Resource AI Use Cases

- Drafting structured review summaries
- Cross-checking protocols against IRB checklists
- Flagging omissions or inconsistencies
- Suggesting precedent from similar cases
- **Must include HUMAN in the loop (HITL)!**

- **See Human-in-the-Loop Testing for LLM-Integrated Software for QA frameworks to reduce hallucinations, bias, and prompt injection.**
- **See Formalising Human-in-the-Loop for typologies of oversight, failure modes, and legal–moral responsibility.**

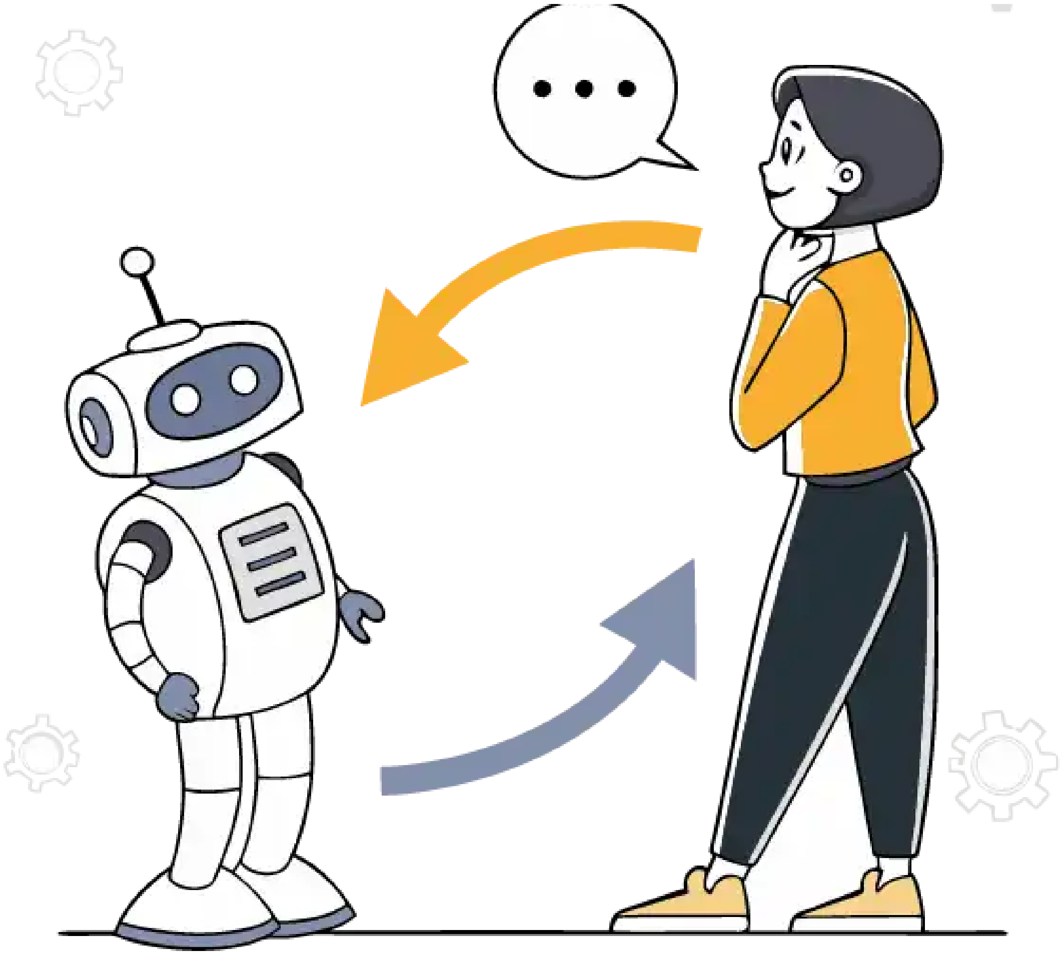
Review

High-Resource, High-Responsibility AI Use Cases

- Predictive post-approval monitoring
- Identifying systemic risks across protocols
- Compliance database integration
- Human in the loop: **constant oversight structure**

Post-determination

Human-in-the-Loop: Oversight in Practice



Defines use cases, data flows, failure modes; policy sign-off.

Verifies AI flags; document accept/reject with rationale.

Conducts QA sampling, incident/near-miss logging, periodic recalibration.

Any output affecting determinations is independently reviewed; never auto-approve.

Insights from Our IT Neighbors

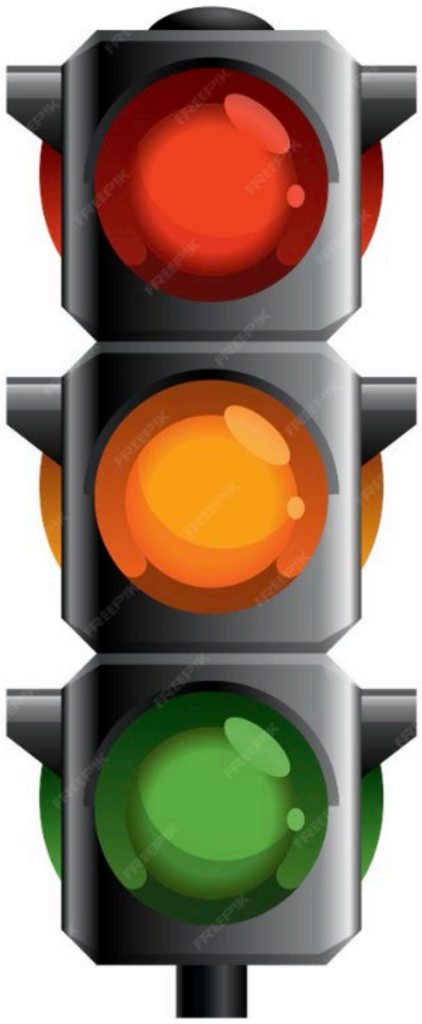
- **Audit logs**: Robust logs should capture:
  - Data lineage (origin and transformation of datasets)
  - Model versioning (which model was used)
  - Traceability (which human reviewer acted on AI output)
- **De-identification standards:**
  - Safeguards against re-identification and alignment to governance policies.
- **Vendor vetting:** Contracts with AI vendors must address:
  - Data ownership
  - Breach notification
  - Right to security audits
- **Risk evaluation:**
  - Adapting existing frameworks
  - Framework-guided evaluation

➢ **See ISO/IEC 27001 vendor controls and NIST CSF 2.0: Updated Third Party & Supply Chain Risk Management**

Adoption Test: Use Only When All Are True

- ✓The **IRB makes the determination,** not AI.
- ✓Reviewers **can interrogate outputs**; logs exist for QA.
- ✓**No identifiable data** enters models without explicit safeguards.
- ✓People and process for **training, incidents, updates**.
- ✓If efficiency reduces transparency/control, **do not use.**

Your resources, risk tolerance, and safeguards determine your AI adoption curve!



Adopt AI Responsibly

- Match adoption to monitoring capacity
- Keep humans in the loop
- **Ethics first**, efficiency second

➢ **See [Responsible artificial intelligence governance: A review and research framework]()**

Table 4
Responsible artificial intelligence (AI) principles.

Principle	Sub-dimensions	References
Accountability	**Auditability:** ability to assess AI applications concerning the algorithms, data, and design processes. **Responsibility:** oversight of the various stages and activities involved in AI deployment and how it should be allocated to people, roles, or departments	(de Almeida et al., 2021; European Commission, 2019; Mikalef et al., 2022)
Diversity, non-discrimination and fairness	**Accessibility:** design of systems in a manner that makes them accessible and usable for everyone, regardless of age, gender, abilities, and characteristics **No unfair bias:** inclination of prejudice toward or against people, objects, or positions, as well as inherent biases in datasets, which can precipitate undesirable outcomes	(Fjeld et al., 2020; Singapore Government, 2020)
Human agency and oversight	**Human review:** right of a person to challenge a decision made by an AI **Human well-being:** the notion that AI must include human well-being as a primary success factor for development	(European Commission, 2019; Singapore Government, 2020)
Privacy and data governance	**Data quality:** accuracy of values in a dataset, matching the true characteristics of the entities described by the dataset **Data privacy:** AI systems' development and operation in a manner that considers data privacy throughout the data lifecycle **Data Access:** national and international rights laws during the design of an AI for data access permissions	(Matthews, 2020; Singapore Government, 2020)
Technical robustness and safety	**Accuracy:** AI system's ability to make correct judgments, such as correctly classifying information into the appropriate categories or being able to predict, recommend, or make intelligent decisions based on data or data models **Reliability:** AI system's ability to work properly when subjected to a range of inputs or situational contexts **General Safety:** safety rules and fallback plans that should be established for AI systems in the event of problems **Resilience:** AI systems that should be protected against vulnerabilities that adversaries can exploit, e.g., hacking	(European Commission, 2019; Singapore Government, 2020)
Transparency	**Explainability:** ability to explain the technical processes of an AI system and related human decisions (e.g., application areas of a system) **Communication:** human right to be informed in advance when interacting with an AI agent **Traceability:** ability to track data and processes that yield the AI system's decision, including data gathering, labeling, and algorithms.	(Fjeld et al., 2020; Mikalef et al., 2022; Singapore Government, 2020)
Social and environmental well-being	**Social well-being:** ubiquitous exposure to social AI systems in all areas of society, such as work and education. **Environmental well-being:** most pressing environmental and climate concerns facing the planet	(European Commission, 2019; Singapore Government, 2020)

How AI tools might be incorporated into the IRB review process or to support other IRB and HRPP functions.

Swapnali Chaudhari, MBBS, MS, CRC/CRA
Director, IRB/HRPP
Rutgers, The State University of New Jersey

Disclosure Statement

I have relevant personal/professional/financial relationship(s) with respect to this educational activity with the following organization(s):

AAHRPP

Site Visitor

AI and IRB Partnership

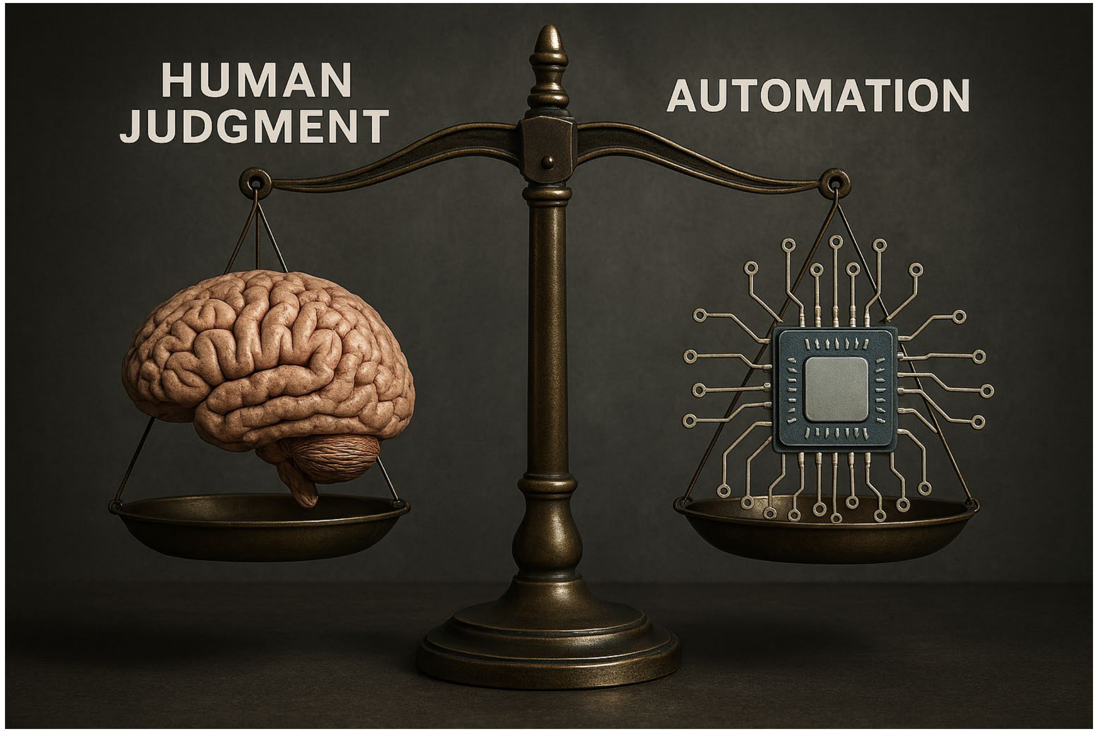
AI and IRB: A Symbiotic Relationship, Not a Takeover

🌀 Streamline IRB operations

✅ Support application checks

👤 Human expertise remains essential

🤝 AI as partner, not substitute



Association for the Accreditation of Human Research Protection Programs, Inc.®

Enhancing IRB Operations Through Improved Communication and AI

- Clear PI communication drives timely review
- Six-week AI vs. human study
- AI helps, experts remain essential
- Next: hybrid models & eIRB+ integration

Quality Assurance and AI Analysis of IRB Memorandums

Judy Kwak MA, CIP, Swapnali Chaudhari MBBS, MS, Naveena Yanamala MS, PhD

RUTGERS UNIVERSITY Office for Research

Goal

Enhancing IRB memorandum review efficiency by comparing human expert assessments with AI-based analysis over six weeks.

Background

Institutional Review Boards (IRBs) utilize Turnaround Time (TAT) as a key metric to assess the efficiency of research protocol reviews

The Rutgers University IRB enhanced turnaround time by:

- Identifying modifications and continuations for expedited review.
- Conducting regular assessments of queues (weekly, then biweekly).
- Establishing a clear timeline for processing and action steps.

Another area for improvement in TAT was the delays in Principal Investigators' response to requested changes which continues to impact turnaround time.

We questioned whether the use of AI tools in administrative workflows may improve TAT in 2 ways:

- Enhancing the quality of IRB memoranda,
- Faster responses from the PIs.

However, before formally implementing AI-assisted memorandums, it is critical to assess how AI interpretation compares to manually crafted memorandums and whether these differences impact overall efficiency in research administration.

Methodology

- **Study Duration**: Conducted over six weeks.
- **Sample Size**: Twelve IRB memorandums were randomly selected each week, covering *Full Board, Expedited, and Exempt* categories. Total sample size: 12 memorandums.
- **Review Process**: Each memorandum was independently assessed by two human expert reviewers, as well as AI-based tools, including Copilot and Gemini.
- **Evaluation Metrics**: The memorandums were reviewed for clarity, tone, completeness, readability and conciseness. Reviews were analyzed for discrepancies, deviations, and clarity in presenting deliberations.
- **Scoring Method**: A 1-to-5 rating scale was used to assess review quality. Microsoft tools and Grammarly supported scoring assessments. Specific Voting Metrics analyzed included:
  - Preference count for Copilot, Gemini, or Human.
  - Cross expert votes (e.g., Expert 1 voting for Expert 2)
  - Self-voting tendencies.
- **Third-Party Evaluation**: An anonymized review was conducted by a third evaluator, measuring gap percentages, discrepancies, and overall quality scores.
- **Comparative Analysis**: Performance was compared across two sets to assess differences in clarity, readability, tone, completeness, and conciseness.
  - Human vs. Copilot
  - Human vs. Gemini

Results

Summary of Voting Distributions:

	Expert 1	Expert 2
Gemini	30% (3/10)	50% (5/10)
Human Expert	70% (7/10)	50% (5/10)

Average Gemini Acceptance :(3+5)/20 = 40%

	Expert 1	Expert 2	Expert 3
Copilot	70% (7/10)	10% (1/10)	50% (5/10)
Human Expert	30% (3/10)	90% (9/10)	50% (5/10)

Average Copilot Acceptance: (7+1+5)/30=13/30 ≈ 43.3%

AI tools are accepted in less than half the cases overall and with high variance (SD ≈ 25%), suggesting subjectivity or bias in voting.

Self-Voting & Peer Expert Preference:

Expert	Self-Voting vs Copilot	Self-Voting vs Gemini	Other Expert vs Copilot	Other Expert vs Gemini
Expert 1	33%	67%	25%	75%
Expert 2	83%	33%	100%	75%

The perception of peer expertise shifted between tools, with Gemini making Expert 2 more trusting of Expert 1, while Expert 1 self-rated higher when comparing against Gemini.

IRB Category & Expert 3 Voting Preference:

IRB Type	Expert 1	Expert 2	Copilot
Exempt (n=4)	1	1	2
Expedited (n=4)	1	1	2
Full Board (n=4)	0	2	2

Copilot (50% in all cases) followed by Expert 2 (33.3% overall and 50% in Full board) was favored by Expert 3 (completely blinded).

Conclusions & Future Directions

Conclusions:

- Overall, neither AI tool tested as part of this study outperformed human experts across all IRB categories.
- Expert perception was observed to be tool dependent (Copilot vs Gemini) and was highly variable – recommending the need for broader testing with more diverse experts and blinded trials
- The preference for AI versus Human did not differ in IRB memorandums for Exempt or Expedited studies.
- Expert 2/3's judgments aligned most with perceived quality, suggesting domain expertise remains critical in complex reviews.
- AI could be a viable co-reviewer or assistant to provide consistency where human skill diverges, or reviewer variability is substantial – especially for complex cases (Full Board protocols).

Next Steps:

- Expand to larger, balanced expert cohorts with overlapping case reviews.
- Explore hybrid review models: AI + Human expert collaboration, especially for Full Board complexity.
- Explore AI-generated memorandums' impact on readability, structure, and PI response time.
- Identify key areas to improve communication strategies for faster study resubmission & maintain professional standards while enhancing institutional expectations.
- Assess how AI-driven clarity influences workflow efficiency and whether optimized memoranda lead to more timely and effective PI responses

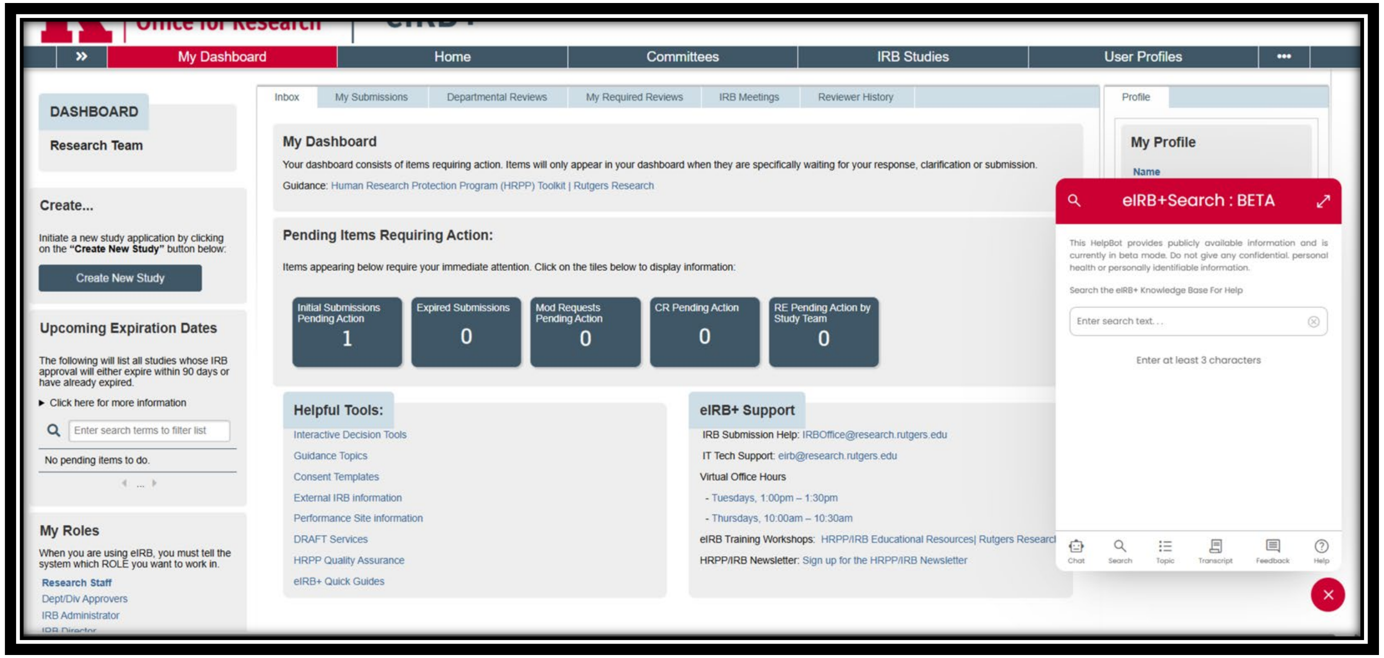
Limitations:

- **Small Sample Set & Style Leakage:** Considering only 6 cases per expert and they reviewed/voted their own previously written work (or AI-generated work prompted by their own input), there is a possibility for remembering/recognizing the content, phrasing or structure.
- **Training Bias:** Individual preferences influenced outcomes, as reviewers' backgrounds affected their assessments.

The IRB's 24/7 "Answer Man" The Research Bot

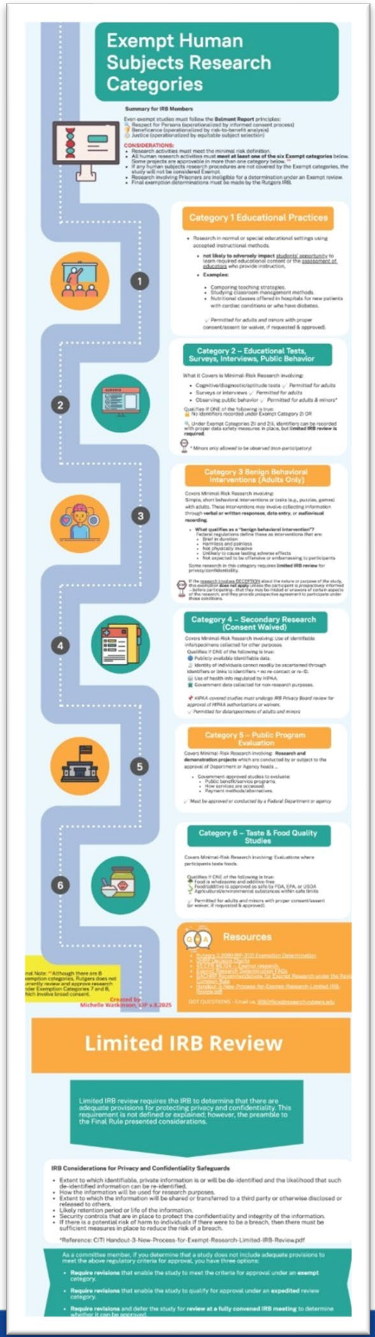
The Research Bot

- Three Models of Use
- Easy Access to Info
- Downloadable Transcripts
- Feedback Driven



Education with Pictograms

- Meeting Compliance Demands with Strategy
- Transitioning From Lectures To "Bite-Sized" Learning
- Infographics as a 24/7 Resource
- Boosting Engagement through Two-Minute Briefs



AI and Office for Research Partnership

- AI-Powered Contract Pre-Review
- Strategic Insight from Sponsored Data
- Enhanced Research Portfolio Mapping

Next Innovations

ADDING AI TOOLS FOR THE IRB PRE-REVIEW PROCESS.

AUTOMATING REVIEW OF EXEMPT SUBMISSIONS.

DEVELOPING A PREDICTIVE AI MODEL.

Reference List:

- The ethics of using artificial intelligence in scientific research: new guidance needed for a new tool
- Beyond principlism: Practical strategies for ethical AI use in research practices
- When combinations of humans and AI are useful: A systematic review and meta-analysis
- Augmented Intelligence Framework for Human–Artificial Intelligence Teaming in Cybersecurity.
- NIST Researchers Suggest Historical Precedent for Ethical AI Research.
- "Does Black Box AI In Medicine Compromise Informed Consent?"

Thank You!