H2AI: A Framework for Experiential Learning and De-Risking Generative AI in Healthcare

Rohan Rani [1*], Jacqueline Sandling[1*], Francis Arellano[1], Nika Shroff[1], Yumin Gao[1], Raj Ratwani[1,2]

**Co-first authors.*
[1] Georgetown University School of Medicine
[2] MedStar Health National Center for Human Factors in Healthcare
`js5155@georgetown.edu, rmr137@georgetown.edu`

Abstract

The rapid evolution of generative AI presents immense opportunities for healthcare, yet a significant gap exists between technological potential and clinical application. This gap is exacerbated by a lack of supported environments where emerging leaders can apply new technologies to complex health problems. Traditional innovation pipelines are often too slow and siloed, while unstructured events like conventional hackathons frequently fail to produce lasting, clinically relevant solutions. This paper demonstrates the H2AI (Health and AI) hackathon as a working system and replicable framework designed to bridge this gap. Over two years, the H2AI model has evolved a structured, four-pillar approach: 1) curated interdisciplinarity in a university setting, 2) a technology-agnostic platform for agility, 3) patient-centricity by design, and 4) an integrated ethical and safety scaffolding. A longitudinal analysis of outcomes from 2024 to 2025 reveals significant growth in participation and a marked maturation in the technical sophistication of prototypes, evolving from general workflow tools to specialized, multimodal platforms targeting complex diseases like Parkinson's. By embedding a formal curriculum on trustworthy AI and creating a structured pipeline to clinical translation, H2AI functions as a de-risking mechanism for early-stage innovation. This paper presents H2AI as a scalable model for experiential learning that accelerates the development of responsible, patient-centered generative AI solutions.

1 Introduction

A fundamental challenge in translating artificial intelligence from code to clinic is the scarcity of opportunities for future leaders to apply novel technologies to healthcare problems within supported, interdisciplinary environments. Aspiring innovators face significant barriers: building a well-rounded team, accessing domain expertise from clinicians, and keeping pace with the rapid evolution of AI are all difficult undertakings for individuals or small groups.

In response to this challenge, healthcare-focused hackathons have emerged as an increasingly popular venue for fostering innovation. Since 2010, the number of such events has grown substantially, with organizations like MIT Hacking Medicine facilitating over 175 events across 29 countries [1], and even government agencies such as the Centers for Medicare & Medicaid Services (CMS) adopting the format to address systemic issues in rural health [2].

However, a debate surrounds the efficacy of these events. While they generate enthusiasm, many argue that healthcare hackathons "fall short of developing lasting innovations because healthcare

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: The Second Workshop on GenAI for Health: Potential, Trust, and Policy Compliance.

challenges are too complex to be addressed in a 48-hour event" [3]. This suggests that while the demand for rapid and collaborative innovation is high, a more robust framework is needed to produce solutions that are not only technologically novel but also clinically relevant and viable.

The H2AI hackathon is a demonstration of such a framework. It is designed as a structured, mission-driven collaborative engineering marathon to advance innovation, research, and education for better patient care. The model is explicitly architected as a platform for experiential learning; by providing a supported sandbox, H2AI offers a solution to the core problems of access and guidance, creating a second-generation hackathon model engineered to overcome the documented flaws of its predecessors.

2 The H2AI Framework: Four Pillars of Responsible Innovation

The H2AI model's success is the result of a deliberate design built on four integrated pillars. This framework creates a system of structured serendipity, actively steering the innovation process toward solutions that are clinically relevant, ethically sound, and commercially viable. The outputs are therefore not a random occurrence but a predictable result of the system's design.

2.1 Pillar 1: Curated Interdisciplinarity in a University Setting

The fusion of diverse talent is the foundational element of the H2AI model. Recognizing that healthcare problems are too complex for any single discipline to solve, the program leverages its university setting and institutional partnerships to convene a carefully curated cohort of participants. In its inaugural 2024 event, the participant demographics were balanced, with backgrounds in Computer Science & Engineering (33%), Medicine & Health (30%), Data Science & Analytics (14%), and Business & Management (10%). This composition ensured that technical development was grounded in clinical insight and business acumen from the outset.

By 2025, the model demonstrated its adaptability, attracting a more technical cohort where developers and data scientists constituted a 73% majority. This shift did not diminish clinical relevance but rather increased the need for expert guidance. The program compensated by building a robust support network of over 20 clinical and business mentors (including MDs, PhDs, and MBAs) as well as patient advocates who provided continuous orientation toward real-world workflows and patient safety imperatives. This strategy is validated by external research, which identifies diverse interdisciplinary participants as one of the facilitators of a successful hackathon [3]. The institutional partnerships with Georgetown University's School of Medicine, McDonough School of Business, and School of Health, alongside the clinical partnership with MedStar Health, provide the mechanism that enables this curated diversity.

2.2 Pillar 2: A Technology-Agnostic Approach to Foster Agility and Creativity

The H2AI framework is designed for maximum adaptability in a rapidly changing technological landscape. This is embodied in its strategic decision to move from a prescribed tool (ChatGPT in its first year) to a technology-agnostic approach in its second year. This choice is justified by two key principles: speed and creativity. First, AI is evolving at an unprecedented pace. A new model, dataset, or paradigm may be released days before the competition begins. Being technology-agnostic grants participants the agility to leverage the most current and powerful tools available, ensuring the solutions are state-of-the-art. Second, empowering participants to use any tool unleashes creativity and accommodates the heterogeneous skill sets of an interdisciplinary group. This is particularly critical in an environment where not all participants have prior programming experience, allowing them to contribute using methods with which they are most familiar.

2.3 Pillar 3: Patient-Centricity by Design

A core principle of the H2AI model is the integration of the end-user voice directly into the innovation process. This is a critical factor often missing in technology development. H2AI achieved a potential solution to this problem by treating patients as domain experts within the competition. Their lived experience provides invaluable context that cannot be assumed or generalized, especially by development teams primarily composed of non-clinicians. This practice aligns with emerging trends in collaborative design, where patient-users are directly involved in the hackathon process [4]. The commitment to patient-centricity is further embedded programmatically. A representative from a patient safety nonprofit serves as a judge, and a dedicated Patient Safety Prize is awarded, signaling the premium placed on this dimension of innovation. In the second year, a second award was granted for the solution that exhibited true "care for the whole person." This pillar ensures that solutions are not just technically sound but are also designed in service of the patient, a stakeholder who is uniquely vulnerable in the healthcare context.

2.4 Pillar 4: Integrated Ethical and Safety Scaffolding

H2AI distinguishes itself by functioning as a de-risking model for innovation. Rather than treating ethics and safety as afterthoughts, the program provides a formal educational curriculum designed to equip participants with the knowledge to build trustworthy AI from the ground up. The 2024 curriculum included a series of expert-led workshops on critical topics, as detailed in Table 2. These principles are translated from theory into practice through programmatic incentives. The Patient Safety Prize ensures that safety-conscious design is actively rewarded. The power of this incentive was demonstrated in 2025, when the Grand Prize winner, Team Rabbit, was also the recipient of the Patient Safety Award. This alignment demonstrates that within the H2AI framework, safety is not an ancillary feature but a core, indispensable component of exemplary innovation.

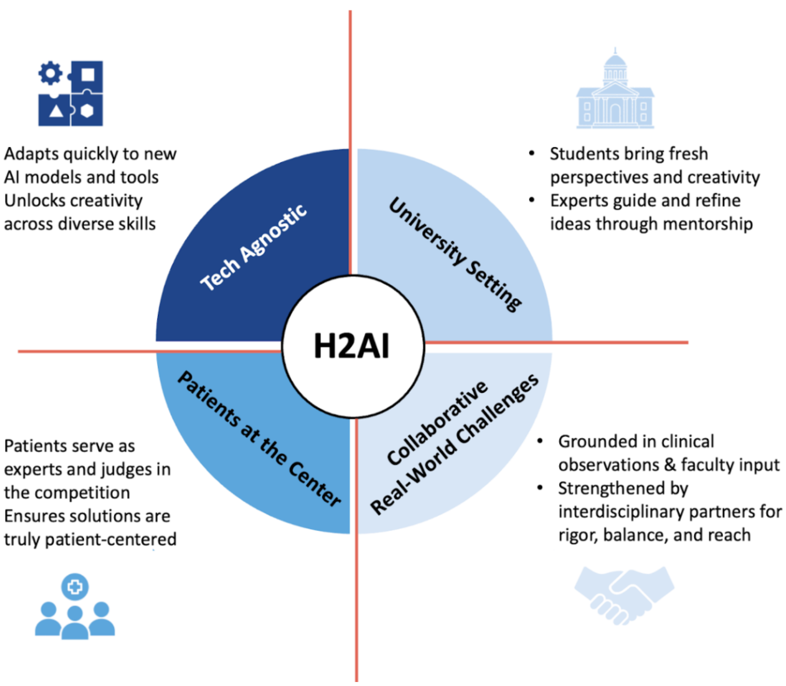

Figure 1: H2AI Main Features.

3 Outcomes and Impact: A Two-Year Longitudinal Analysis

The efficacy of the H2AI framework is best demonstrated through a longitudinal analysis of its outcomes. Over its first two years, the program has shown not only significant quantitative growth but also a clear maturation in the quality and sophistication of its outputs. Data from 2024 to 2025 reveals a rapidly expanding scale and influence, as detailed in Table 3. The number of participants doubled, the number of teams grew by 78%, and social media reach exploded by over 1000%, indicating a dramatic rise in the program's visibility and appeal.

While most metrics show growth, the 53% decrease in participating institutions reveals a deliberate strategic refinement. The inaugural event cast a wide geographic net to build brand awareness and encourage interdisciplinary collaboration. In contrast, the 2025 event featured an increase in participation from specific institutions that offer robust artificial intelligence and engineering programs, such as MIT, Johns Hopkins University, and Carnegie Mellon University. This pivot from maximizing geographic breadth to cultivating higher-impact participation likely enhanced the overall technical sophistication of teams' prototypes, while also allowing them to spend more time investigating the end user experience, go to market strategy, and research and development roadmaps.

3.1 Demonstration of Maturation: From Workflow Tools to Multimodal Platforms

The most compelling evidence of the model's effectiveness lies in the evolution of its outputs, summarized in Table 1. The winning projects show a rapid maturation of the program's innovative capacity, transitioning from foundational software applications to highly specialized, multimodal platforms.

In 2024, winning projects focused on leveraging generative AI to address systemic healthcare challenges identified by the event's organizing team. The grand prize winner, Aster, was a Software as a Service (SaaS) solution for clinical pre-charting and ambient documentation, designed with FHIR-based interoperability for any EHR system. Other winners like Filed.AI and NexusCare followed a similar theme of streamlining workflows and synthesizing information. These challenges were appropriate given the recent widespread adoption of LLMs like ChatGPT, the sophistication of readily available tools, and recent incorporation of both into universities' curricula.

The 2025 hackathon marked a significant technical leap forward. This was catalyzed by a programmatic decision to incentivize solving problems related to Parkinson's Disease. This choice acted as a forcing function, channeling the technical energy of the participants toward a specific set of patients' problems. The result was a portfolio of complementary solutions. The grand prize winner, Team Rabbit, developed a VR-based platform that simulated real-world mobility tasks while capturing detailed neuromuscular data. Other winners developed sophisticated tools using augmented reality, specialized Convolutional Neural Networks (CNNs) for handwriting analysis, and even custom ESP32 hardware with tremor-tracking sensors. This demonstrates that the design of the clinical challenge is a critical control variable for tuning the innovative output of the entire system, transforming the hackathon into a strategic instrument for targeted R&D.

Table 1: Longitudinal Comparison of Winning Prototypes (2024 vs. 2025), highlighting the maturation from software tools to specialized, multimodal platforms.

Year	Project Name	Clinical Problem	Key Technologies	Output Type
2024	Aster	Provider Burnout	Generative AI, FHIR	SaaS Platform
2024	Filed_AI	Administrative Burden	OCR, GPT, Python/Flask	Automation Tool
2024	NexusCare	Geriatric Decision Support	Gen AI, EHR Integration	Chatbot
2025	Team Rabbit	Parkinson's Mobility	VR, ML, Computer Vision	Diag. & Therap. Platform
2025	ParkSuite	Parkinson's Remote Monitoring	AR, NLP, Computer Vision	Monitoring Suite
2025	Tremor Trace	Parkinson's Motor Skills	1D CNN, Attention	Diagnostic Tool

Beyond producing prototypes, the H2AI program serves as an intense, immersive educational experience. Post-event survey data from 2024 provides quantitative evidence of its impact on participants. On average, participants rated the likelihood of recommending H2AI to a peer at 8.5 out of 10 and reported that the event had a strong positive impact on their motivation to make a difference with their careers (7.8 out of 10). Furthermore, approximately 70% of participants planned to learn more about the concepts they were exposed to, and 50% expressed a desire to continue working on their projects. This aligns with external findings that hackathons can significantly increase participants' confidence and provide a positive change in attitude toward professional development [5].

4 Discussion: Limitations and Future Directions

4.1 Lessons Learned and Iterative Improvement

The H2AI model is self-reflective and continuously evolving. The organizing team, itself learning by doing, has transparently acknowledged limitations and used them as catalysts for improvement. Challenges such as participant burnout from the intense schedule and logistical inefficiencies were identified in the 2024 impact report. In response, the organizers implemented concrete solutions, such as hosting some workshops prior to the main hackathon weekend to alleviate time pressure.

A key challenge is ensuring that the participating teams have sufficient clinical context. In 2024, participants from medicine and health backgrounds constituted a strong 30% of the cohort. However, by 2025, this figure dropped to 9%. Without sufficient clinical input during the intensive building phase, there is a risk of developing solutions that are technically impressive but disconnected from real-world workflows and patient needs. Other limitations, such as inconsistent funding and natural team turnover, are being addressed through strategic planning for long-term partnerships.

4.2 A Scalable Model for Overcoming the "Valley of Death"

A common criticism of hackathons is that innovative ideas often "die on Monday morning" as participants return to their regular routines [6]. The H2AI model is deliberately designed to counteract this phenomenon by building a structured pipeline to support projects post-event. This approach mirrors best practices from other successful models that utilize post-hackathon offerings to stimulate progress toward commercialization.

The H2AI program formalizes a pathway to both commercialization and clinical validation by connecting winning teams with regional entrepreneurship ecosystems and facilitating clinical research partnerships within MedStar Health. The involvement of institutional bodies like Georgetown's Office of Technology Commercialization provides expert guidance on intellectual property and commercial strategy. Following the 2025 event, the Tremor Trace team is reported to be "now pursuing clinical research," while Team Rabbit is planning to develop its interfaces "in close collaboration with care teams specializing in Parkinson's disease" [7]. This structure provides a tangible roadmap for the most promising teams, addressing the critical valley of death for early-stage health tech ideas. Future directions aim to scale this pipeline by partnering with a student-run healthcare technology accelerator program, building a robust alumni network, and connecting with other healthcare AI competitions to form a powerful meta-network for innovation.

Because the original event occurred within the last two years, assessments of long-term outcomes for this model of student-led innovation are limited. Time will ultimately determine whether these structures translate into sustained projects and the development of tangible patient-centered generative AI products that patients and their clinicians can use. What can be concluded at this stage, however, is that the creation of a pipeline for innovation at the student level is feasible and provides the necessary infrastructure for emerging leaders to develop patient-centered generative AI solutions.

5 Conclusion

The H2AI hackathon, as demonstrated through its 2024 and 2025 iterations, represents more than just a successful student-run event. It serves as a scalable, replicable, and highly effective model for addressing persistent challenges in the development and deployment of healthcare AI. Its success is the result of a deliberate design that establishes a cohesive framework. By strategically convening students and mentors from technical, clinical, and business backgrounds, H2AI creates a microcosm of the real-world teams needed to build and launch successful health technologies. Through a formal curriculum and embedded incentives, H2AI de-risks innovation at the point of conception, fostering a culture of responsibility essential for building trust in medical AI. By building formal pathways to clinical research and commercialization, the program provides a bridge for promising early-stage ideas to mature into clinically impactful tools.

In synthesizing these elements, the H2AI model demonstrates that the gap between AI's potential and its clinical application can be bridged through structured, collaborative, and ethically-grounded innovation. As a working system, H2AI provides a powerful framework for de-risking, accelerating, and ultimately realizing the promise of trustworthy, patient-centered generative AI.

Acknowledgments and Disclosure of Funding

The authors thank the mentors, judges, sponsors, participants, and institutional partners who supported the H2AI initiative and contributed to its development and implementation.

References

[1] Gubin, Tatyana A., et al. "A Systems Approach to Healthcare Innovation Using the MIT Hacking Medicine Model." *Cell Systems*, vol. 5, no. 1, Jul. 2017, pp. 6–10. PubMed, `https://doi.org/10.1016/j.cels.2017.02.012`.

[2] 2024 Rural Health Hackathon | CMS. `https://www.cms.gov/priorities/innovation/2024-rural-health-hackathon`. Accessed 2 Sep. 2025.

[3] Poncette, A-S., et al. (2020). Hackathons as stepping stones in health care innovation: Case study with systematic recommendations. *Journal of Medical Internet Research*, 22(3), e17004. `https://doi.org/10.2196/17004`.

[4] Leung, K., et al. (2019). Using patient and family engagement strategies to improve outcomes of health information technology initiatives: Scoping review. *Journal of Medical Internet Research*, 21(10), e14683. `https://doi.org/10.2196/14683`.

[5] Nandi, A., & Mandernach, M. (2016). Hackathons as an informal learning platform. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education* (pp. 346-351). `https://doi.org/10.1145/2839509.2844590`.

[6] Pe-Than, E. P. P., et al. (2022). Corporate hackathons, how and why? A multiple case study of motivation, projects proposal and selection, goal setting, coordination, and outcomes. *Human-Computer Interaction*, 37(4), 281-313. `https://doi.org/10.1080/07370024.2020.1760869`.

[7] *This citation has been anonymized as it contains the authors' names within it.*

A Appendix

This appendix contains supplementary tables referenced in the main text.

Table 2: H2AI Educational Curriculum Workshops (2024)

Workshop Title	**Description**
AI Safety and Ethics	Led by a legal and ethics expert, this session covered foundational principles of consent, data privacy, and transparency.
Guidelines to Address Racial and Ethnic Bias	Presented by a director from the Agency for Health Research and Quality (AHRQ), this workshop provided a practical framework for bias mitigation in algorithms.
Redteaming Large Language Models in Medicine	This session introduced participants to adversarial testing, teaching them to identify and mitigate critical failure modes like factual "hallucinations" and privacy breaches.
Patentability of Generative AI Innovations	A workshop led by a law partner grounded the teams' work in the realities of intellectual property and the requirements for human contribution in AI-assisted inventions.

Table 3: H2AI Program Evolution (2024-2025): Key Metrics.

Metric	2024	2025	Year-over-Year Change
Registrants	210	240	+14%
Participants	80	160	+100%
Teams	18	32	+78%
Prizes Awarded	$15,000	$15,000	-
Social Media Impressions	2,425	30,000	+1136%
Participating Institutions	36	20	-53%

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The paper has been carefully reviewed by the authors for accuracy.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Common criticisms of hackathons are explored in juxtaposition with H2AI's success.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Justification: No theoretical results are provided.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: No code is provided, however the basic layout of the hackathon is provided and easily reproducible.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [NA]

   Justification: Non-experimental

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [NA]

   Justification: Non-experimental

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

   Justification: No statistics were provided

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [NA]

   Justification: Not an experiment

9. **Code of ethics**

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: All ethical guidelines were followed.

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: Both positives and negatives of the hackathon were discussed.

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: Not misused

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [NA]

    Justification: No assets were described, though the H2AI organization's members are in the authorship.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [No]

    Justification: No new assets introduced.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: No participants

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: No experimental subjects and thus no IRB approval was required.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: LLMs were not used in the core methods of this paper.