

# Foundations of Human-AI Interaction: Theory, Algorithms, Practice

ICML 2026 Workshop Proposal

Website: <https://human-ai-interaction-workshop.github.io/>

## Abstract

Human–AI Interaction (HAI) is rapidly evolving as a central part for modern AI research. As systems such as Large Language Models become embedded in our daily workflows, they begin to take active roles as multi-turn, conversational, problem-solving and decision-making agents. Despite its growing relevance and importance, HAI as a research direction currently lacks unifying foundations to fundamental aspects such as efficiency, optimality, privacy, and interpretability. Hence, we propose our workshop, *Foundations of Human-AI Interaction: Theory, Algorithms, Practice*, to gather researchers from different domains of research. In our proposal, we present six confirmed speakers from diverse research backgrounds, such as interpretability, uncertainty quantification and decision-making. We also include interactive games to engage the audience and showcase existing challenges in HAI research, and a panel discussion in the form of debates to foster insightful discussions. Broadly speaking, we aim to foster cross-disciplinary HAI research as a principled field of AI research with long-term impact in the foreseeable future.

**Description and Motivation.** Collaboration is central to knowledge creation and decision making. From negotiation and collective problem-solving to scientific discovery, progress often emerges through interaction rather than isolated reasoning. As today’s Artificial Intelligence (AI) systems, including Large Language Models (LLMs) [6, 15, 25, 35, 37] and diffusion models [19, 31, 38, 39], become increasingly generative, capable, and versatile, they are evolving into active participants in collaborative tasks. Through multi-turn conversational exchanges, humans routinely use AI systems to explore ideas, verify reasoning, navigate uncertainty, and support decision-making across domains. Consequently, Human–AI Interaction (HAI) is taking the form of collaborative, multi-agent behavior, characterized by asymmetric capabilities, responsibilities, and incentives between humans and machines.

Despite its growing importance, HAI lacks unifying foundations. Fundamental questions such as, “When is HAI helpful in practice?” [16], “What constitutes optimal collaboration under information imbalance?” [14], “What are appropriate notions of interpretability in collaborative systems?” [21], or “How can we build interactive-by-design AI systems?” [36] are being raised across several research communities, yet clear, unifying answers remain elusive. Developing general principles to guide algorithm design, provide formal guarantees, and enable systematic evaluation has grown urgent. This is especially important for the ICML community: while traditional machine learning models humans as static sources of labels or feedback, modern AI systems demand novel frameworks that formally capture humans as adaptive collaborators within learning and inference loops. Addressing these questions requires diverse perspectives, from learning theory and decision theory to cognitive science and policymaking.

**Scope and Topics.** The goal of this workshop is to establish a foundational science of HAI for modern interactive AI systems, bridging perspectives from well-established areas of research (e.g., machine learning, statistics, decision-theory, and economics) and identifying principled theory, algorithms, and practice that govern effective collaboration between humans and AI systems. To bring together researchers working across theory, algorithms, and practice to develop a shared language and tools for studying HAI, we invite contributions along three tightly connected dimensions:

1. Theory: foundational principles and formal models of HAI, including statistical learning theory [7], uncertainty quantification [26, 27], value of information [17], Bayesian frameworks [9, 34], sequential decision-making and agreement protocols [12–14], and characterization of human behavior [16].
2. Algorithms: implementations of formal HAI principles, such as learning with human feedback [28], conversational systems [10, 24], interactive mechanisms of explanation [21], evaluation metrics and online auditing [32].
3. Practice: real-world cases where HAI is essential, including healthcare [11, 23], scientific discovery [33], recommendation systems [20], robotic agents [30], preference elicitation [22], and scenarios involving partial information [8, 18, 29].

**Workshop Contributions.** We summarize the key contributions of our workshop as follows:

1. Interchange Ideas from Diverse Areas of Research: We aim to bring together research across theory and practice. Through such exchange, we aim to foster and build a community that develops methods with a stronger emphasis on HAI. Each of our six speakers represents an established field (i.e. uncertainty quantification, interpretability, optimization, decision-theory etc.), ensuring broad relevance for attendees and fostering cross-disciplinary dialogue.
2. Meaningfully Engage the Larger Community: Through dual poster sessions, spotlight talks, our proposed HAI game and panel through a Public Forum debate format, we facilitate fruitful idea exchanges, discussions and collaborations among researchers. Dedicated Q&A sessions and networking breaks further provide opportunities to connect with speakers and researchers from both academia and industry, bringing HAI’s relevance to every aspect of AI research.
3. Establish HAI as a Research Field: Our workshop takes a deliberate step toward building the research community and discovering open problems that can anchor HAI as a field in its own right. By convening researchers across these adjacent communities, we catalyze long-term collaborations and impact that extend well beyond the workshop itself.

**History of Workshop and Previous Workshops.** This is the first edition of the proposed workshop. While there exist related workshops [1–5], we are the first to emphasize the mathematical foundation of HAI. To highlight the major difference, we mention two of the most recent workshops: The *NeurIPS Workshop on Multi-Turn Interactions in Large Language Models* [3] primarily focuses on the methodology, evaluation, and multi-turn usage of LLMs. In contrast, our workshop broadens the lens to formally investigate HAI without restricting attention to a specific class of AI models. Similarly, the *MICCAI Workshop on Human-AI Collaboration* [2] focuses on the benefits of HAI in healthcare, for example, with AI-assisted radiology report generation. Our workshop, instead, seeks to build a research community to establish principled, mathematical foundations of HAI.

### Invited Speakers (All Confirmed, All In-Person).

- Dr. Been Kim, Senior Staff Research Scientist, Google DeepMind  
*Research areas: Interpretability*
- Dr. Jessica Hullman, Ginni Rometty Professor of Computer Science, Northwestern University  
*Research areas: Bayesian decision theory, uncertainty quantification, AI for decision-making*
- Dr. Hamed Hassani, Associate Professor of Electrical and Systems Engineering, University of Pennsylvania  
*Research areas: Trustworthy machine learning, information theory*
- Dr. Amine Bennouna, Assistant Professor of Operations at the Kellogg School of Management, Northwestern University  
*Research areas: Data-driven decision-making, optimization*
- Dr. Eric Zelikman, CEO and Co-Founder, Humans&  
*Research areas: LLMs, AI for science*
- Dr. Sin Yu Bonnie Ho, Senior Research Scientist, Microsoft Health and Life Sciences  
*Research areas: AI for decision-making, patient care*

**Interactive Engagement.** To provide more interaction and engagements with our audience, we introduce two major innovations to our workshop. 1) HAI Game: a live Human-AI collaboration game where participants experience AI-assisted decision-making firsthand. Participants will work with an AI assistant on a task where each party has access to different, incomplete information, making collaboration essential. Participants submit an initial answer, observe the AI’s suggestion, and decide whether to revise. We collect responses throughout and conclude by presenting live statistics: e.g. *How often did collaboration help? When did it hurt?* This transforms abstract concepts into tangible experience. 2) Debate Panel: Rather than a traditional Q&A panel session, we plan to divide the speakers into two teams and host a debate in “Public Forum” format about trending topics and popular talking points. The debate will have multiple rounds. In each round, each team will be given 5 minutes to answer questions such as: *“If an AI makes a mistake in a collaborative setting (e.g., medical diagnosis, legal advice), who is responsible?”*, *“Are certain elements of human judgment indispensable, or can smart machines eventually handle all decision-making?”*. Overall, we aim to create an engaging environment between the speakers and audience and find holistic answers to open-ended questions in HAI.

**Tentative Schedule.** The schedule of our workshop is designed to foster a welcoming, stimulating, and inter-disciplinary environment for sharing ideas and forming connections. Beyond keynotes from invited researchers and interactive sessions outlined above, we will host 2 poster sessions and 2 oral spotlight sessions to highlight outstanding submissions (3 talks per session, 10 min. each). Our tentative schedule is:

Time	Event	Time	Event
8:30–8:50am	Breakfast + <b>HAI Game #1</b>	12:30–1:30pm	Lunch break + <b>HAI Game #2</b>
8:50–9:00am	Opening remarks	1:30–2:00pm	<b>Invited talk #5</b> (25 min. + 5 min. Q&A)
9:00–9:30am	<b>Invited talk #1</b> (25 min. + 5 min. Q&A)	2:00–2:30pm	<b>Invited talk #6</b> (25 min. + 5 min. Q&A)
9:30–10:00am	<b>Invited talk #2</b> (25 min. + 5 min. Q&A)	2:30–3:30pm	<b>Poster session #2</b> + Coffee break (discussion)
10:00–11:00am	<b>Poster session #1</b> + Coffee break (discussion)	3:30–4:00pm	<b>Spotlight orals #2</b> (7 min. + 3 min. Q&A)
11:00–11:30am	<b>Invited talk #3</b> (25 min. + 5 min. Q&A)	4:00–4:55pm	<b>Discussion Panel</b>
11:30–12:00pm	<b>Invited talk #4</b> (25 min. + 5 min. Q&A)	4:55–5:00pm	Closing remarks
12:00–12:30pm	<b>Spotlight orals #1</b> (7 min. + 3 min. Q&A)		

**Workshop Submissions and Timelines.** Paper Submission: We will invite two types of submissions: 1) We will invite extended abstracts of up to 4 pages (excluding references). Accepted papers will be presented as in-person posters. We will make sure to recruit reviewers from a wide variety of institutions and training levels to ensure the review process is unbiased; 2) We will invite submissions from accepted conference/journal papers up to a year ago (ICML 2025). Acceptance decisions will be made by organizers based on relevance to the workshop. Timeline: *Call for papers*: March 22 | *Submission deadline*: April 24 | *Reviewing period*: April 26 - May 10 | *Notification of acceptance*: May 15 | *Workshop Date*: July 10/11. Sponsorships: We are actively seeking sponsorship from both academia and industry. The obtained funds will be used to give awards to students who submitted to our workshop and achieved spotlight (Top 1-2% submission) or to provide travel grants to students with accepted papers from underrepresented groups.

## Organizers

Our core organizing team covers a variety of workshop-organizing experience, from junior researchers establishing their academic networks to well-experienced members:

Name	E-mail	Designated contact	Website	Scholar	Other ICML 2026 workshop proposals
Kwan Ho Ryan Chan	<a href="mailto:ryanckh@seas.upenn.edu">ryanckh@seas.upenn.edu</a>	✓	<a href="#">link</a>	<a href="#">link</a>	✗
Sima Noorani	<a href="mailto:nooranis@seas.upenn.edu">nooranis@seas.upenn.edu</a>	✓	<a href="#">link</a>	<a href="#">link</a>	✗
Jacopo Teneggi	<a href="mailto:jtenegg1@jhu.edu">jtenegg1@jhu.edu</a>	✗	<a href="#">link</a>	<a href="#">link</a>	✗
Christopher Chiu	<a href="mailto:chyc3@cam.ac.uk">chyc3@cam.ac.uk</a>	✗	<a href="#">link</a>	<a href="#">link</a>	Yes, 1
Hyewon Jeong	<a href="mailto:hyewonj@mit.edu">hyewonj@mit.edu</a>	✗	<a href="#">link</a>	<a href="#">link</a>	Yes, 1
Dr. Aditya Chattopadhyay	<a href="mailto:achatto@amazon.com">achatto@amazon.com</a>	✗	<a href="#">link</a>	<a href="#">link</a>	✗
Dr. René Vidal	<a href="mailto:vidalr@seas.upenn.edu">vidalr@seas.upenn.edu</a>	✗	<a href="#">link</a>	<a href="#">link</a>	✗
Dr. George Pappas	<a href="mailto:pappasg@seas.upenn.edu">pappasg@seas.upenn.edu</a>	✗	<a href="#">link</a>	<a href="#">link</a>	✗

**Kwan Ho Ryan Chan** is a fifth-year Electrical and Systems Engineering Ph.D. Candidate at University of Pennsylvania, advised by Dr. René Vidal. His work intersects the theory and application of human-AI collaboration, such as building algorithms for uncertainty quantification and decision-making with Large Language Models. He is a recipient of the NSF Graduate Research Fellowship, Penn Engineering Dean’s Fellowship and UPenn AWS-ASSET Fellowship. Previously, he interned at Amazon AWS Agentic AI (2026) and AI/ML Health AI team at Apple (2024). He received his Bachelors of Art in Applied Mathematics from University of California, Berkeley. In the past, he helped with organization of conferences including Conference on Lifelong Learning Agents 2025, DeepMath Conference 2024, as well as volunteered at International Conference on Learning Representations 2022. *Workshop Contributions: initiating formulation, proposal writing, scheduling, inviting speakers.*

**Sima N. Noorani** is a Ph.D. student at the University of Pennsylvania, working at the intersection of machine learning, uncertainty quantification, and conformal prediction. Her research focuses on building reliable uncertainty estimation methods for modern ML systems—including generative models—and on extending these ideas to human–AI interaction settings where people and AI collaboratively reason and make decisions. She received her B.S. in Electrical Engineering from Drexel University with a minor in Computer Science. Previously, she held industry roles spanning applied ML, software engineering, and cybersecurity at Lockheed Martin, Comcast, and Bristol Myers Squibb. *Workshop Contributions: initiating formulation, proposal writing, organizing events, including HAI Game, inviting speakers.*

**Jacopo Teneggi** is a fourth-year Ph.D. student in Computer Science at Johns Hopkins University, working with Dr. Jeremias Sulam. His research centers on the development of statistically-valid methods for the responsible use of AI systems, with a focus on interpretability and uncertainty quantification in medical imaging. He has interned at Profluent and Polymathic AI (Flatiron Institute) to study AI systems for scientific application in biology. In the past, he has organized several local and national TEDx events with 100+ attendees, and lead a student-run non-profit organization. *Workshop Contributions: organizing and structuring of workshop logistics, scheduling, inviting speakers.*

**Christopher Chiu** is a Ph.D. candidate at the University of Cambridge, advised by Prof. Mihaela van der Schaar. His research focuses on developing multi-agent systems for scientific discovery and applications in healthcare. He is a medical doctor with an M.S. in Computer Science from Georgia Institute of Technology and B.Med/M.D. from University of New South Wales. Previously, he worked at Harrison.ai, where he led an international team of 150 radiologists in dataset curation, and contributed to developing foundational vision-language models for radiology. He has extensive experience in organizing medical conferences at both university and national levels, and served as Sponsorship Director for UNSW Medical Society, raising over AUD \$48,000 in funding. *Workshop Contributions: Managing workshop website, organizing and structuring of workshop logistics.*

**Hyewon Jeong** is a Ph.D. Candidate at MIT EECS and a medical doctor. Her research background is primarily in machine learning for healthcare. Recurring technical themes include building foundational models with large-scale real-world clinical data (tabularized time-series data such as EHRs, signals (e.g., electrocardiograms, electroencephalograms), wearables, and notes), fairness, and applying all-purpose foundational models (e.g., LLMs) for clinical purposes. Previously, she completed her medical degree at the Yonsei University College of Medicine, a master’s in Computer Science from KAIST, and a bachelor’s in Biological Sciences from KAIST. She has been an active member of machine learning for the healthcare community, serving for two years as organizer for ML4H, organized NeurIPS 2025 workshop on Learning from Time Series for Health (TS4H), Multimodal Large Language Models (MLLMs) in Clinical Practice

Workshop in MICCAI 2025, and served as a guest editor for Frontiers in Digital Health. *Workshop Contributions: Writing on motivation and scope, and proposal and workshop formatting, scheduling.*

**Dr. Aditya Chattopadhyay**<sup>1</sup> is an Applied Scientist at AWS Agentic AI, where his research focuses on developing efficient architectures for Large Language Models, with special emphasis on long-context training and inference. He received his Ph.D. in Computer Science from Johns Hopkins University in 2024, advised by Profs. René Vidal and Donald Geman. He has made seminal contributions to the field of Interpretable AI, most notably his work on Grad-CAM++, which has been cited more than 4,000 times. He has also organized tutorials at ICCV and CVPR on the Foundations of Interpretable AI. *Workshop Contributions: proposal writing, finding sponsorships, managing logistics.*

**Dr. René Vidal** is the Penn Integrates Knowledge and Rachleff University Professor of Electrical and Systems Engineering & Radiology and the Director of the Center for Innovation in Data Engineering and Science (IDEAS) at the University of Pennsylvania. He is also an Amazon Scholar, an Affiliated Chief Scientist at NORCE, and a former Associate Editor in Chief of TPAMI. His current research focuses on the foundations of deep learning and trustworthy AI and its applications in computer vision and biomedical data science. His lab has made seminal contributions to motion segmentation, action recognition, subspace clustering, matrix factorization, deep learning theory, interpretable AI, and biomedical image analysis. He is an ACM Fellow, AIMBE Fellow, IEEE Fellow, IAPR Fellow and Sloan Fellow, and has received numerous awards for his work, including the IEEE Edward J. McCluskey Technical Achievement Award, D'Alembert Faculty Award, J.K. Aggarwal Prize, ONR Young Investigator Award, NSF CAREER Award as well as best paper awards in machine learning, computer vision, signal processing, controls, and medical robotics. *Workshop Contributions: initiating formulation, sending invites to our senior speakers, advising and writing on motivation and scope, finding sponsorships.*

**Dr. George Pappas** is the UPS Foundation Professor at the Department of Electrical and Systems Engineering at the University of Pennsylvania. He also holds a secondary appointment in the Departments of Computer and Information Sciences, as well as Mechanical Engineering and Applied Mechanics. He currently serves as the Associate Dean for Research and Innovation in the School of Engineering and Applied Science and as the Director of the Raj and Neera Singh program in Artificial Intelligence. Pappas's research focuses on control systems, robotics, autonomous systems, formal methods, and machine learning for safe and secure cyber-physical systems. He has received numerous awards, including the NSF PECASE, the Antonio Ruberti Young Researcher Prize, the George S. Axelby Award, the O. Hugo Schuck Best Paper Award, and the George H. Heilmeier Faculty Excellence Award. Pappas has mentored more than fifty students and postdocs, now faculty in leading universities worldwide. He is a Fellow of IEEE, IFAC, and was elected to the National Academy of Engineering in 2024. *Workshop Contributions: initiating formulation. Sending invites to our senior speakers, advising and writing on motivation and scope, finding sponsorships.*

## Program Committee

This is the list of Program Committee members who have all agreed to review the workshop papers: Yuyan Ge (University of Pennsylvania), Tianjiao Ding (University of Pennsylvania), Darshan Thaker (University of Pennsylvania), Beepul Bharti (Johns Hopkins University), Andrea Wynn (Johns Hopkins University), Jie Gao (Johns Hopkins University), Zheng Zhang (Notre Dame University), Drew Prinster (Johns Hopkins University), Shayan Kiyani (University of Pennsylvania), Natalie Collina (University of Pennsylvania), Luca Muscarena (University of Cambridge), Silas Ruhrberg (University of Cambridge). We will continue to recruit reviewers depending on the number of submissions and needs of the workshop.

---

<sup>1</sup>The content of this proposal does not relate to Aditya Chattopadhyay's position at Amazon.

## References

- [1] Ai for human-centered interaction (aihci) workshop. <https://sites.google.com/view/aihci/home>. Accessed: 13 February 2026.
- [2] Miccai workshop on human-ai collaboration. <https://haic-miccai.github.io/>. Accessed: 2026-02-13.
- [3] Neurips workshop on multi-turn interactions in large language models. <https://workshop-multi-turn-interaction.github.io/>. Accessed: 2026-02-13.
- [4] Human-ai interaction for augmented reasoning: Improving human reflective and critical thinking with artificial intelligence. <https://aireasoning.media.mit.edu/>, 2025. CHI 2025 Workshop, Accessed: 13 February 2026.
- [5] Iclr 2025 workshop on human-ai coevolution (haic). <https://sites.google.com/stanford.edu/haic2025/home>, 2025. Accessed: 13 February 2026.
- [6] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.
- [7] Nikhil Agarwal, Alex Moehring, and Alexander Wolitzky. Designing human-ai collaboration: A sufficient-statistic approach. Technical report, National Bureau of Economic Research, 2025.
- [8] Chinmaya Andukuri, Jan-Philipp Fränken, Tobias Gerstenberg, and Noah D Goodman. Star-gate: Teaching language models to ask clarifying questions. *arXiv preprint arXiv:2403.19154*, 2024.
- [9] Gagan Bansal, Besmira Nushi, Ece Kamar, Eric Horvitz, and Daniel S. Weld. Is the most accurate ai the best teammate? optimizing ai for teamwork, 2021. URL <https://arxiv.org/abs/2004.13102>.
- [10] Kwan Ho Ryan Chan, Yuyan Ge, Edgar Dobriban, Hamed Hassani, and René Vidal. Conformal information pursuit for interactively guiding large language models. *arXiv preprint arXiv:2507.03279*, 2025.
- [11] Christopher Chiu, Silviu Pitis, and Mihaela van der Schaar. Simulating viva voce examinations to evaluate clinical reasoning in large language models. *arXiv preprint arXiv:2510.10278*, 2025.
- [12] Natalie Collina, Surbhi Goel, Varun Gupta, and Aaron Roth. Tractable agreement protocols. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 1532–1543, 2025.
- [13] Natalie Collina, Surbhi Goel, Aaron Roth, Emily Ryu, and Mirah Shi. Emergent alignment via competition. *arXiv preprint arXiv:2509.15090*, 2025.
- [14] Natalie Collina, Ira Globus-Harris, Surbhi Goel, Varun Gupta, Aaron Roth, and Mirah Shi. Collaborative prediction: Tractable information aggregation via agreement. In *Proceedings of the 2026 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 4712–4798. SIAM, 2026.
- [15] Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.
- [16] Ziyang Guo, Yifan Wu, Jason D Hartline, and Jessica Hullman. A decision theoretic framework for measuring ai reliance. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, pages 221–236, 2024.
- [17] Ziyang Guo, Yifan Wu, Jason Hartline, and Jessica Hullman. The value of information in human-ai decision-making. *arXiv preprint arXiv:2502.06152*, 2025.
- [18] Kunal Handa, Yarin Gal, Ellie Pavlick, Noah Goodman, Jacob Andreas, Alex Tamkin, and Belinda Z Li. Bayesian preference elicitation with language models. *arXiv preprint arXiv:2403.05534*, 2024.
- [19] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

- [20] Xiong Junwu, Xiaoyun Feng, YunZhou Shi, James Zhang, Zhongzhou Zhao, and Wei Zhou. Digital human interactive recommendation decision-making based on reinforcement learning. *arXiv preprint arXiv:2210.10638*, 2022.
- [21] Been Kim, John Hewitt, Neel Nanda, Noah Fiedel, and Oyvind Tafjord. Because we have llms, we can and should pursue agentic interpretability. *arXiv preprint arXiv:2506.12152*, 2025.
- [22] Belinda Z Li, Alex Tamkin, Noah Goodman, and Jacob Andreas. Eliciting human preferences with language models. *arXiv preprint arXiv:2310.11589*, 2023.
- [23] Stella Li, Vidhisha Balachandran, Shangbin Feng, Jonathan Ilgen, Emma Pierson, Pang Wei W Koh, and Yulia Tsvetkov. Mediq: Question-asking llms and a benchmark for reliable interactive clinical reasoning. *Advances in Neural Information Processing Systems*, 37:28858–28888, 2024.
- [24] Jessy Lin, Nicholas Tomlin, Jacob Andreas, and Jason Eisner. Decision-oriented dialogue for human-ai collaboration. *Transactions of the Association for Computational Linguistics*, 12:892–911, 2024.
- [25] Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, et al. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*, 2024.
- [26] Laura R Marusich, Jonathan Z Bakdash, Yan Zhou, and Murat Kantacioglu. Using ai uncertainty quantification to improve human decision-making. *arXiv preprint arXiv:2309.10852*, 2023.
- [27] Sima Noorani, Shayan Kiyani, George Pappas, and Hamed Hassani. Human-ai collaborative uncertainty quantification. *arXiv preprint arXiv:2510.23476*, 2025.
- [28] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744, 2022.
- [29] Wasu Top Piriyakulkij, Volodymyr Kuleshov, and Kevin Ellis. Active preference inference using language models and probabilistic reasoning. *arXiv preprint arXiv:2312.12009*, 2023.
- [30] Allen Z Ren, Jaden Clark, Anushri Dixit, Masha Itkina, Anirudha Majumdar, and Dorsa Sadigh. Explore until confident: Efficient exploration for embodied question answering. *arXiv preprint arXiv:2403.15941*, 2024.
- [31] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.
- [32] Shuvom Sadhuka, Drew Prinster, Clara Fannjiang, Gabriele Scalia, Aviv Regev, and Hanchen Wang. E-evaluator: Reliable agent verifiers with sequential hypothesis testing. *arXiv preprint arXiv:2512.03109*, 2025.
- [33] Erzhuo Shao, Yifang Wang, Yifan Qian, Zhenyu Pan, Han Liu, and Dashun Wang. Sciscigpt: advancing human-ai collaboration in the science of science. *Nature Computational Science*, pages 1–15, 2025.
- [34] Mark Steyvers, Heliodoro Tejeda, Gavin Kerrigan, and Padhraic Smyth. Bayesian modeling of human ai complementarity. *Proceedings of the National Academy of Sciences*, 119(11):e2111547119, 2022. doi: 10.1073/pnas.2111547119. URL <https://www.pnas.org/doi/abs/10.1073/pnas.2111547119>.
- [35] Kimi Team, Yifan Bai, Yiping Bao, Guanduo Chen, Jiahao Chen, Ningxin Chen, Ruijue Chen, Yanru Chen, Yuankun Chen, Yutian Chen, et al. Kimi k2: Open agentic intelligence. *arXiv preprint arXiv:2507.20534*, 2025.
- [36] Shirley Wu, Michel Galley, Baolin Peng, Hao Cheng, Gavin Li, Yao Dou, Weixin Cai, James Zou, Jure Leskovec, and Jianfeng Gao. Collabllm: From passive responders to active collaborators. *arXiv preprint arXiv:2502.00640*, 2025.
- [37] An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.
- [38] Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Wentao Zhang, Bin Cui, and Ming-Hsuan Yang. Diffusion models: A comprehensive survey of methods and applications. *ACM computing surveys*, 56(4):1–39, 2023.
- [39] Sherry Yang, KwangHwan Cho, Amil Merchant, Pieter Abbeel, Dale Schuurmans, Igor Mordatch, and Ekin Dogus Cubuk. Scalable diffusion for materials generation. *arXiv preprint arXiv:2311.09235*, 2023.