

Conference on Health, Inference, and Learning (CHIL) 2026

Elizabeth Healey*

Boston Children's Hospital, Boston, MA, USA

ELIZABETH.HEALEY@CHILDRENS.HARVARD.EDU

Jason Fries*

Stanford University, Palo Alto, CA, USA

JFRIES@STANFORD.EDU

Tom Pollard*

Massachusetts Institute of Technology, Cambridge, USA

TPOLLARD@MIT.EDU

Shengpu Tang*

Emory University, Atlanta, GA, USA

SHENGPU.TANG@EMORY.EDU

Anna Zink*

Tufts University

ANNA.ZINK@TUFTS.EDU

Thomas Hartvigsen*

University of Virginia

HARTVIGSEN@VIRGINIA.EDU

Monica Agrawal*

Duke University

MONICA.AGRAWAL@DUKE.EDU

Vasiliki Bikia*

Stanford University

BIKIA@STANFORD.EDU

Samuel G. Finlayson*

University of Washington and Seattle Children's Hospital

SGFIN@UW.EDU

Benjamin Glicksberg*

Mount Sinai

BENJAMIN.GLICKSBERG@MSSM.EDU

Brett Beaulieu-Jones*

University of Chicago

BEAULIEUJONES@UCHICAGO.EDU

Kai Wang*

Georgia Institute of Technology

KWANG692@GATECH.EDU

Daseyra Fontalvo*

Association for Health Learning and Inference, New York, NY, USA

DASEYRA@AHLI.CC

Tasmie Sarker*

Association for Health Learning and Inference, New York, NY, USA

TASMIE.SARKER@AHLI.CC

Irene Y. Chen*

University of California, Berkeley, CA, USA

IYCHEN@BERKELEY.EDU

Emily Alsentzer*

Stanford University, Palo Alto, CA, USA

EALSENTZER@STANFORD.EDU

Abstract

The Conference on Health, Inference, and Learning (CHIL) focuses on advancing machine learning for health, bringing together clinicians and researchers, across both industry and academia. Since 2022, CHIL has been an official conference of the Association for Health Learning and Inference (AHLI). This volume contains

proceedings of the seventh annual CHIL conference, held at the Seattle Children's Research Institute in the US.

1. Introduction

The Conference on Health, Inference, and Learning (CHIL) convenes a cross-disciplinary community—clinicians and researchers from academia, health systems, and industry—around a shared mission to advance machine learning for health ([Conference on Health, Inference, and Learning \(CHIL\), 2026](#)). Each year, CHIL serves as a venue for this community to forge collaborations, exchange ideas, and grapple with the emerging directions shaping the field.

The questions facing machine learning for health have grown sharper over the past year. The early excitement around generative models and large-scale AI has given way to a more demanding set of problems—not only how to compose these models into agentic systems and reason over multimodal data, but how to evaluate the resulting systems rigorously, deploy them safely within real clinical workflows, and ensure their benefits reach patients and care providers equitably. Meeting these challenges is not a matter of methodological progress alone. It requires sustained collaboration across clinical practice, computer science, and the social sciences, and a community willing to hold innovation to a high standard of evidence.

The work gathered in this volume reflects that maturing agenda. A large share of this year’s papers turns to large language models for clinical language, spanning the generation and evaluation of clinical text, while a growing thread moves beyond single-model prompting toward agents and the simulation environments needed to evaluate them. Multimodal learning is a second current, with contributions that integrate and reason over diverse clinical data modalities, including foundation models whose generalization is scrutinized as carefully as their accuracy. Running through much of the volume is an insistence on trustworthiness, with work that revisits performance claims, examines calibration and selective prediction, and probes robustness under realistic shifts in data and deployment conditions. A final group turns toward clinical decision-making and care delivery, advancing causal, survival, and reinforcement-learning methods for treatment decisions and prognosis, and bringing machine learning to resource-constrained, real-world care settings.

Through CHIL, we aim to foster these collaborations and accelerate the responsible impact of machine learning for health. This volume presents the proceedings of the seventh annual CHIL conference,

held in Seattle, Washington at the Seattle Children’s Research Institute.

1.1. Organizing Committee

The organizing committee consisted of the following individuals: Emily Alsentzer (General Chair), Irene Chen (General Chair), Sam Finlayson (Program Chair), Monica Agrawal (Program Chair), Vicky Bikia (Unconference Chair), Hong Truong (Unconference Chair), Suhana Bedi (Unconference Chair), Mjaye Mazwi (Research Roundtables Chair), Aishwarya Mandyam (Research Roundtables Chair), Emma Rocheteau (Doctoral Symposium Chair), Matt Engelhard (Doctoral Symposium Chair), Tom Pollard (Proceedings Chair), Elizabeth Healey (Proceedings Chair), Jason A. Fries (Proceedings Chair), Tom Hartvigsen (Track 1 Chair), Shengpu Tang (Track 2 Chair), Anna Zink (Track 3 Chair), Kai Wang (Fundraising Chair), Brett Beaulieu-Jones (Fundraising Chair), Ben Glicksberg (Fundraising Chair), Xinyang Han (Comms Chair), Feng Chen (Local Logistics Chair) Daserya Fontalvo (Conference Coordinator), and Tasmie Sarker (Executive Director).

2. Program

Our keynote speakers were selected to reflect a breadth of expertise in technology development, deployment, and evaluation:

Chris Longhurst, MD, MS: Chris Longhurst is CEO of Seattle Children’s, who until recently served in a dual role as chief medical officer (CMO) and chief digital officer (CDO) at UC San Diego Health, where he directed an AI portfolio across the UC San Diego System. He has published hundreds of articles in care quality, patient safety and health informatics.

Jessilyn Dunn, PhD: Jessilyn Dunn is Assistant Professor of Biomedical Engineering and Biostatistics and Bioinformatics at Duke University, and Director of the BIG IDEAs Laboratory whose goal is to detect, treat, and prevent chronic and acute diseases through digital health innovation.

Raj Manrai, PhD: Raj Manrai is an Assistant Professor in the Department of Biomedical Informatics at Harvard Medical School, where he leads a research lab that works broadly on applying machine learning and statistical modeling to improve medical decision-making. Raj is also a founding Deputy

Editor of NEJM AI, the new artificial intelligence-focused journal from the publishers of the New England Journal of Medicine, and co-host of the NEJM AI Grand Rounds podcast.

2.1. Panels

Our panel sessions explored topical and evolving challenges in the development and deployment of artificial intelligence:

AI Deployment and Operational Challenges: Clara Lin (Seattle Children’s Hospital), Nikesh Kotecha (Stanford Medicine), and Jean Feng (UCSF). Moderated by Samuel Finlayson (University of Washington, Seattle Children’s Hospital). The panel tackled key tradeoffs faced by healthcare institutions as they navigate the balance of safety and innovation amid an evolving technological landscape. The discussion was colored by their complementary experiences at hospitals with vastly different financial resources and patient populations.

AI and Mental Health: Tim Althoff (University of Washington), Saadia Gabriel (UCLA), Harini Suresh (Brown). Moderated by Lucy Wang (University of Washington). The panel discussed pressing challenges at the intersection of AI and mental health, where safety is paramount but present day realities include limited access to traditional resources.

2.2. Unconference

Building on the participant-driven format introduced in 2025, this year’s Unconference—Health AI Builders 2.0—sharpened its focus on the central challenge of the field: moving health AI beyond promising research and prototypes into ventures, products, and clinical practice that reach patients. The program was framed around the perspective of the builder—the founder, engineer, and operator—and convened entrepreneurs alongside clinicians, investors, and health-system and industry leaders for a full day of candid, practice-oriented exchange. The agenda moved deliberately from reflection to action: a keynote-style conversation and a practitioner panel in the morning, followed in the afternoon by two rounds of facilitated small-group discussion and an evening social.

The day opened with a fireside conversation on deploying health AI in the real world, featuring Jay Nanduri (Co-Founder and CTO, Truveta), which explored the journey from an initial idea to a functioning, deployed system. This was followed by a panel on

implementation stories from the field, in which Andy Chu (Chartis), Michael Han (Ambience Healthcare), Chris Carruthers (DexCare), and Arun Ravi (Microsoft) reflected on real-world experiences deploying AI in clinical settings and the lessons shaping future implementation efforts.

The afternoon’s first round, “Translating Health AI into Practice,” organized seven roundtables, each led by an expert practitioner who seeded discussion with questions drawn from their own experience and the perspectives that determine whether health AI succeeds in the field: health-system adoption (Nikesh Kotecha), the payer perspective (Corinne Stroum, SCAN), life sciences (Martin Kang, DanaHER), getting started (Ryan McRobert, Fenwick & West), products and partnerships (Arun Ravi, Microsoft), funding pathways (Hong Truong, Define Ventures), and technical infrastructure and deployment (Bradd Busick, Frazier Healthcare Partners). Participants moved between tables, allowing different groups to shape the conversation across topics.

New this year, the second round introduced the Research-to-Venture Forum, which paired early-stage entrepreneurs with experienced facilitators at mixed-discipline tables of founders, technical leads, and business strategists. With no judges, rankings, or winners, each table was a collaborative working session where founders shared early-stage ideas, received candid feedback from participants with diverse expertise, and explored opportunities, challenges, and next steps for bringing their ventures to practice. Five ventures took part:

- **PARK** (Md. Saiful Islam, University of Rochester; facilitated by Babak Parviz, New-Days.ai) — a web-based tool that uses brief webcam and microphone tasks to screen for signs of Parkinsonian disorders, addressing widespread underdiagnosis and limited access to movement-disorder specialists by offering a low-burden front-door screen that flags who should seek clinical follow-up.
- **Cellulae Labs** (Lorenzo Bini, University of Geneva; facilitated by Martin Kang, DanaHER) — a B2B platform that automates the analysis of high-dimensional biological data, targeting slow, manual, and inconsistent workflows that bottleneck drug development and translational research.
- **PediaMed** (Xu Cao and James Rehg, University of Illinois Urbana-Champaign; facilitated by Amrita Sehgal, SCAN) — an at-home, video-

based solution for early detection of childhood developmental conditions, addressing the shortage of pediatric specialists and long diagnostic waitlists with objective, scalable assessment enabling earlier intervention.

- **Equanimity Labs** (Evan Miller, Microsoft; facilitated by Michael Han, Ambience Healthcare) — an AI-driven platform to enhance medical education and training, aimed at making high-quality, personalized learning and assessment scalable for clinicians in training.
- **Baseline** (Favour Narisse, Stanford University; facilitated by Neha Patadia, SeaHealthTech and Hong Truong, Define Ventures) — a personal wellbeing “autopilot” that continuously monitors health signals and delivers proactive, individualized support, shifting emphasis from reactive care toward continuous prevention.

The day concluded with closing remarks and an evening social that carried the connections begun at the tables into the broader CHIL community.

2.3. Research Roundtables

This year, CHIL hosted 10 research roundtables, with topics covering a variety of interdisciplinary subjects that are of interest to both machine learning and clinical communities. Examples include “Frameworks for measuring value: Towards Reliable Health AI” in which participants are asked to discuss how current AI systems are evaluated for healthcare deployment and whether those practices reflect how institutions use AI. Another roundtable topic, “On the Opportunities and Risks of LLM Agents in Healthcare” presents an opportunity for practitioners and model developers to anticipate how LLM agents can improve or negatively affect existing workflows.

Each research roundtable was led by two experts in the field. These experts include individuals with clinical, scientific, business development, and machine learning backgrounds. The roundtable leaders were asked to draft a list of relevant topics prior to the conference, which were used to direct conversation at the tables. The research roundtable sessions were held on both days of the main conference to ensure that different groups of individuals could contribute to the discussion for an individual topic. Roundtable leads were asked to take notes during the discussion so that the CHIL organizing committee could compile a whitepaper summarizing the general perspectives on the selected topics.

2.4. Doctoral Symposium

The Doctoral Symposium gave PhD students an opportunity to present ongoing or proposed research and receive feedback from CHIL attendees and senior experts. The symposium included 31 posters, selected spotlight talks, and mentorship sessions focused on research development and career advice.

3. Papers: Guidelines and Selection

Papers for the proceedings were solicited across three tracks at the intersection of health and machine learning: 1) *Models and Methods*; 2) *Applications and Practice*; and 3) *Impact and Society*. The submission and review process was entirely managed through the OpenReview platform ([OpenReview, 2026](#)).

3.1. Submissions

We received a total of 145 submissions across the three tracks. All submissions were reviewed by a minimum of three reviewers, except where the submission was subsequently desk rejected (n=2) or withdrawn (n=11).

3.2. Peer review

CHIL follows a mutually anonymous (“double blind”) review process with five broad phases:

1. Reviewer bidding
2. Assignment of papers to reviewers
3. Completion of reviews
4. Discussion, and author rebuttals
5. Decision and notification

Reviewers were asked to consider relevance to healthcare, clarity, soundness, and significance when preparing their reviews. After reviews were completed, a discussion phase allowed open communication between authors, reviewers, and area chairs. Area Chairs then created meta-reviews and recommendations based on the reviews. Final decisions were made by Track Chairs in discussion with Area Chairs, Proceedings Chairs, General Chairs, and AHLI board, where appropriate.

4. Proceedings

Thirty-nine submissions (30%) were accepted for publication in the archived conference proceedings. Some prominent themes include:

- Clinical language models, documentation, and diagnostic reasoning.** Several papers study large language models for clinical text, documentation, diagnostic reasoning, social signal extraction, privacy, and text generation. These include work on reconstructing sepsis trajectories from clinical case reports (Noroozizadeh and Weiss), adapting new-generation LLMs using legacy clinical models (Ronaghi et al.), evaluating stepwise diagnostic reasoning (Choi et al.), assessing the robustness of ambient scribes for SOAP note generation (Latif et al.), deriving extubation-failure features from respiratory therapy notes (Chaiken et al.), identifying breast cancer radiation side effects (Seah et al.), benchmarking LLM-patient interactions for spina bifida care (Azhar et al.), extracting social signals from patient-provider communication (Bedmutha et al.), generating bilingual synthetic clinical notes (Lee et al., a), applying reinforcement learning to acute-care clinical text generation (Wang et al., b), and anonymizing masked language models (Boutet et al.).
- Agents, simulation, and workflow-centered health AI.** A second theme concerns agentic systems, simulation environments, queueing and workflow optimization, and operational deployment settings. Papers in this group evaluate LLM agents for multimodal clinical prediction (Al Jorf and Shamout), introduce a multi-agent simulator for hospital administrative workflows with FHIR integration (Lee et al., b), optimize triage and queues for resource-constrained free clinics (Grewal), and use agent-based synthetic EHR data to evaluate robustness under mass casualty incidents (Delos Reyes et al.).
- Multimodal learning, imaging, audio, and vision-language evaluation.** Several contributions focus on medical imaging, radiology, ultrasound, respiratory audio, video, and vision-language models. These include work on brachytherapy treatment planning and visualization (Chen et al.), multiple instance learning for 3D neuroimage classification (Harvey et al.), structured and sequential chest X-ray interpretation (Moon et al.), zero-shot respiratory audio classification (Wang et al., c), ultrasound video question answering (Wang et al., d), revisiting chest X-ray model performance using clinical context (Wang et al., a), improving paraphrase consistency in medical vision-language models (Sadanandan and Behzadan), studying quantization effects in medical imaging models (Ozdemir and Li), and simulating video-based disease progression (Cao et al.).
- Disease progression, prognosis, treatment effects, and policy evaluation.** A further set of papers develops methods for longitudinal disease modeling, treatment-effect estimation, survival analysis, reinforcement learning environments, off-policy evaluation, and diagnosis forecasting. These papers address treatment modeling in deep survival analysis (Hong et al.), medication repurposing and treatment optimization in Alzheimer’s disease (Brady and Yeh), temporal disease progression from cross-sectional data (Hao et al.), multi-state event history analysis for patient-centered disease trajectories (Cottin et al.), counterfactual annotated off-policy evaluation (Mandyam et al.), and dependency-aware learning under extreme label imbalance in EHRs (Ho et al.).
- Biomedical foundation models, omics, signals, and efficient inference.** Papers in this theme examine biomedical foundation models and architectures for signals, single-cell data, microscopy, and efficient EHR inference. Contributions include a framework for evaluating EEG foundation models (Kommineni et al.), multinomial attention masking for sparse single-cell RNA-seq data (Naziri et al.), diffusion modeling for robust single-cell manifold generation (Bini and Marchand-Maillet), interpretable cell counting through prototype density decomposition (Mohammed and Tavanapong), and cached summary embeddings for memory-efficient EHR inference (Al Attrach et al.).
- Population inference, robustness, calibration, and causal evaluation.** Finally, several papers focus on the validity and reliability of machine learning systems under realistic inferential and deployment conditions. This includes survey-aware machine learning for population health inference (Oh et al.), robustness to provenance shift (Tan et al.), calibration and selective prediction in multimodal clinical condition classification (Lopez et al.), and causal evidence on residential relocation and wellbe-

ing using an integrated econometric and machine learning framework (Chen and Wan).

4.1. Awards

Two papers were selected for a Best Paper Award across the conference tracks.

- **Video-Based Disease Progression Simulation** by Xu Cao, Kaizhao Liang and Kuei-Da Liao et al. (Cao et al.)
- **H-AdminSim: A Multi-Agent Simulator for Realistic Hospital Administrative Workflows with FHIR Integration** by Jun-Min Lee, Meong Hi Son, and Edward Choi (Lee et al., b).

5. CHIL 2027 in Chicago

CHIL 2027 will be held in Chicago, IL, USA in June 2027. If you would like to volunteer at the conference or serve on our committee, please contact us at info@chilconference.org.

Acknowledgments

We thank the Association for Health Learning and Inference (AHLI) for their continued support.

5.1. Sponsors

We are deeply grateful for the generous support of our sponsors, whose contributions made CHIL 2026 possible and continue to propel the community at the intersection of health, inference, and learning.

- **Diamond Sponsors:** *Sage Bionetworks* and *Seattle Children's Research Institute*. We extend our special thanks to our Diamond-level sponsors for their exceptional generosity, leadership, and partnership. Their support has been foundational to the success of this year's conference and helps create a vibrant forum for researchers, clinicians, industry leaders, and advance rigorous, impactful work in machine learning for health.
- **Silver Sponsors:** *Oracle Health*, *Apple*, *Vega Health*, *Google*, and *Fenwick*. We sincerely thank our Silver-level sponsors for their strong commitment to advancing research and innovation in health and machine learning.

- **Startup Sponsors:** *Mount Sinai*, *Stanford Health Care*, *Washington Research Foundation*, and *Define Ventures*. We thank our Startup-level sponsors for their enthusiastic engagement and support of the CHIL community.

The collective generosity and support of all our sponsors has been instrumental in making CHIL 2026 a success.

5.2. Reviewers

The proceedings would not have been possible without the help of our area chairs and reviewers.

5.2.1. LIST OF SENIOR AREA CHAIRS

Shengpu Tang, Anna Zink, Tom Hartvigsen.

5.2.2. LIST OF AREA CHAIRS

Jidapa Thadajarassiri, Walter Gerych, Kimia Ghobadi, Haoran Zhang, Zepeng Frazier Huo, Michael Oberst, Jean Feng, Samantha Kleinberg, Jiayu Yao, Sarah Tan, Kirk Roberts, Katayoun Farahi, Intae Moon, Marika Cusick, Aya Zirikly, I. Elizabeth Kumar, Weimin Zhou.

5.2.3. LIST OF REVIEWERS

Shashank Yadav, Iris Szu-Szu Ho, Cheng Fei, Laven-der Yao Jiang, Naoki Nonaka, Ryan King, Jennifer Yu, Vedant Thakkar, Chufan Gao, Zhengjian Kang, Arinbjörn Kolbeinsson, Kei Sen Fong, Jessica Maria Echterhoff, Sheshananda Reddy Kandula, Yuhan Li, Miao Zhang, Lorenzo Bini, Jeffrey Feng, Chandrasekar Ramachandran, Frank Rudzicz, Kuk Jin Jang, Wallapak Tavanapong, Vishal Shah, Shiva Kaul, Zhuolin Hao, Madiha Shakil Mirza, Farzana Islam Adiba, Kevin Wu, Asad Aali, Aditya Kommineni, Frances Dean, Shahriar Noroozizadeh, Ziyue Yu, Sunny Dhamnani, Keyu Li, Jiacheng Lin, Fanyang Yu, Trisha Das, Mirza Farhan Bin Tarek, Leopoldo Julian Lechuga Lopez, Eduardo Davalos, Kai Zhang, Pratham Yashwante, Yurui Cao, David Restrepo, Alissa Andrea Valentine, Elena Sizikova, Pranav Kulkarni, Mingxuan Liu, Hyunjung Gloria Kwak, Van Tuan NGUYEN, Lucia Morris, Zhengyu Fang, Ariel Guerra-Adames, Benjamin C Warner, Aishwarya Mandyam, Chirag Patel, Yuxin Xiao, Michael Brudno, Yunyang Li, Kevin S. Xu, Chang Liu, Yubin Kim, Antoine Boutet, Xinyu Xiong, Jong-Hwan Jang, Tomer Meir, Jiaying Qiu, Simon A.

- Lee, Hongtao Hao, Ali AhmadiTeshnizi, Edward Choi, Srinivasarao Daruna, Po-Chih Kuo, Kamile Stankevičiūtė, Thomas Demeester, Scott Ye, Emma Charlotte Rocheteau, Karanbir Singh, YongKyung Oh, Vasanth Rajendran, William Lotter, Chongmin Lee, Baraa Al Jorf, Elise Jortberg, Rafi Al Attrach, Vasiliki Bikia, Hyewon Jeong, Subasree Venkatsubhramaniyen, Wanying Dou, Yuwen Chen, Manas Bedmutha, Junyi Gao, Harsh Sharma, Siwei Wang, Ethan Harvey, William La Cava, Botao Zhang, Michael Hellstern, Avinash Patil, Jiho Kim, Yingchuan Sun, Divya D. Kulkarni, Marcus Theodor Moen, Donald Xin Lin, Shashank Kapadia, Zichen Wang, Sirui Ding, Ya Gao, Hariharan Ragothaman, Sayantan Kumar, Hangqi Ding, Arnav Saxena, Ajitesh Jain, Wai Tak Lau, Baheti-hazi Maidu, Alban Zammit, Mashrin Srivastava, Pallavi Gudipati, Yinghao Zhu, Meysam Ghaffari, Mudit Mangal, Brighton Nuwagira, Guanghao Mei, Fabian Gröger, William Han, Keshu Wu, Jong Hak Moon, Brett Beaulieu-Jones, Winston Chen, Jiaqing Zhang, Itamar Efrati, Fahmida Liza Piya, Ashim Dhor, Jahnavi Anilkumar Kachhia, Feng Wu, Elizabeth W. Miller, Dhruv Naik, Jonas Knecht, Justin Xu, Quanchao Lu, Saj Maru, Yeonsu Kwon, Bala Siva Sai Akhil Malepati, Yunwon Tae, Haotian Hu, Xinyang Han, Li Zhang, Yining Liu, Bhawesh Kumar, Aaron Fanous, Zhongyuan Liang, Megan Coffee, Huy Nghiem, Zhen Huang, Hongjia Yang, Christopher Yau, Anugya Srivastava, Xiaozhou Shi, Kai Shu, Yue Guo, Aaqib Saeed, Rui Yang, Yifan Shen, Ihsan Ayyub Qazi, Haochen Sui, Yuyang Jiang, Alireza Irani Helabad, Ahmad Rezaie Mianroodi, Junu Kim, Arash Asgari, Max Homilius, Tiantian Feng, Eike Petersen, Asfandyar Azhar, Charles B. Delahunt, Ayush Noori, Yingxin Lai, Yuwei Zhang, Aryan Jadon, Michael Gensheimer, Yike Zhang, Xu Cao, Xiaoman Zhang, Nina Fatehi, Hao Zhang, Warren Woodrich Pettine, Gaurav Mishra, Hongzhou Luan, Edward Chen, Jing Wang, Md Shahrar Fatemi, M Saifur Rahman, Shaurjya Mandal, Jun-Min Lee, Jung Min Lee, Minjia Wang, Karen Sachs, Jinghan Cao, Behdokht Kiafar, Aziliz Cottin, Sara Ketabi, Maryam Khalid, Xiaoli Yang, Rebecca Hwa, Meera Krishnamoorthy, Mahak Shah, Emily Zhou
- Lee, Hongtao Hao, Ali AhmadiTeshnizi, Edward Choi, Srinivasarao Daruna, Po-Chih Kuo, Kamile Stankevičiūtė, Thomas Demeester, Scott Ye, Emma Charlotte Rocheteau, Karanbir Singh, YongKyung Oh, Vasanth Rajendran, William Lotter, Chongmin Lee, Baraa Al Jorf, Elise Jortberg, Rafi Al Attrach, Vasiliki Bikia, Hyewon Jeong, Subasree Venkatsubhramaniyen, Wanying Dou, Yuwen Chen, Manas Bedmutha, Junyi Gao, Harsh Sharma, Siwei Wang, Ethan Harvey, William La Cava, Botao Zhang, Michael Hellstern, Avinash Patil, Jiho Kim, Yingchuan Sun, Divya D. Kulkarni, Marcus Theodor Moen, Donald Xin Lin, Shashank Kapadia, Zichen Wang, Sirui Ding, Ya Gao, Hariharan Ragothaman, Sayantan Kumar, Hangqi Ding, Arnav Saxena, Ajitesh Jain, Wai Tak Lau, Baheti-hazi Maidu, Alban Zammit, Mashrin Srivastava, Pallavi Gudipati, Yinghao Zhu, Meysam Ghaffari, Mudit Mangal, Brighton Nuwagira, Guanghao Mei, Fabian Gröger, William Han, Keshu Wu, Jong Hak Moon, Brett Beaulieu-Jones, Winston Chen, Jiaqing Zhang, Itamar Efrati, Fahmida Liza Piya, Ashim Dhor, Jahnavi Anilkumar Kachhia, Feng Wu, Elizabeth W. Miller, Dhruv Naik, Jonas Knecht, Justin Xu, Quanchao Lu, Saj Maru, Yeonsu Kwon, Bala Siva Sai Akhil Malepati, Yunwon Tae, Haotian Hu, Xinyang Han, Li Zhang, Yining Liu, Bhawesh Kumar, Aaron Fanous, Zhongyuan Liang, Megan Coffee, Huy Nghiem, Zhen Huang, Hongjia Yang, Christopher Yau, Anugya Srivastava, Xiaozhou Shi, Kai Shu, Yue Guo, Aaqib Saeed, Rui Yang, Yifan Shen, Ihsan Ayyub Qazi, Haochen Sui, Yuyang Jiang, Alireza Irani Helabad, Ahmad Rezaie Mianroodi, Junu Kim, Arash Asgari, Max Homilius, Tiantian Feng, Eike Petersen, Asfandyar Azhar, Charles B. Delahunt, Ayush Noori, Yingxin Lai, Yuwei Zhang, Aryan Jadon, Michael Gensheimer, Yike Zhang, Xu Cao, Xiaoman Zhang, Nina Fatehi, Hao Zhang, Warren Woodrich Pettine, Gaurav Mishra, Hongzhou Luan, Edward Chen, Jing Wang, Md Shahrar Fatemi, M Saifur Rahman, Shaurjya Mandal, Jun-Min Lee, Jung Min Lee, Minjia Wang, Karen Sachs, Jinghan Cao, Behdokht Kiafar, Aziliz Cottin, Sara Ketabi, Maryam Khalid, Xiaoli Yang, Rebecca Hwa, Meera Krishnamoorthy, Mahak Shah, Emily Zhou
- Cached summary embeddings for memory-efficient ehr inference. pages N–N.
- Baraa Al Jorf and Farah Shamout. Agentrx: A benchmark study of llm agents for multimodal clinical prediction tasks. pages N–N.
- Asfandyar Azhar, Shaurjya Mandal, Zaid Khan, Nidhish Shah, Curtis Langlotz, and Brad Dicianno. Toward improving diagnostic reasoning for spina bifida care: Benchmarking llm–patient interactions. pages N–N.
- Manas Satish Bedmutha, Feng Chen, Andrea L Hartzler, Trevor Cohen, and Nadir Weibel. Sociallm: Social signal processing of patient-provider communication using llms and contextual aggregation. pages N–N.
- Lorenzo Bini and Stephane Marchand-Maillet. Lapddpm: Spectral perturbation diffusion for robust single-cell manifold generation. pages N–N.
- Antoine Boutet, Lucas Magnana, and Juliette Senechal. Towards the anonymization of masked language modeling. pages N–N.
- Nolan Brady and Tom Yeh. Alpaca: A reinforcement learning environment for medication repurposing and treatment optimization in alzheimer’s disease. pages N–N.
- Xu Cao, Kaizhao Liang, Kuei-Da Liao, Tianren Gao, Zhiguang Ding, Jianguo Cao, Zheng Chen, Jintai Chen, James M Rehg, and Jimeng Sun. Video-based disease progression simulation. pages N–N.
- Izzy Chaiken, Aditya Khowal, Neha A Sathe, Mark M Wurfel, and Lucy Lu Wang. Enhancing extubation failure prediction with llm-derived features from respiratory therapy clinical notes. pages N–N.
- Edward Chen, Natalie Dullerud, Pang Wei Koh, Thomas Niedermayr, Elizabeth Kidd, Sanmi Koyejo, and Carlos Guestrin. Almo: Interactive aim-limit-defined, multi-objective system for personalized high-dose-rate brachytherapy treatment planning and visualization for cervical cancer. pages N–N.
- Jerry Chen and Li Wan. Does where you live affect how you feel? causal evidence from an integrated econometric and machine learning framework. pages N–N.

References

Rafi Al Attrach, Rajna Fani, David Restrepo, Yungang Jia, Leo Anthony Celi, and Peter Schuffler.

- Yera Choi, Yeong Hwa Kim, JaeDeok Lee, Taekang Kim, Sangdoon Yun, and Seong-Eun Moon. Ascent: A benchmark for evaluating and advancing step-wise diagnostic reasoning in large language models on common clinical scenarios. pages N–N.
- Conference on Health, Inference, and Learning (CHIL). CHIL Conference, 2026. URL <https://chilconference.org/>. Accessed: 2026-06-16.
- Aziliz Cottin, Marine Zulian, Sandrine Katsahian, and Agathe Guilloux. Msnet: A deep neural network based on piecewise-constant proposals within multi-state event history analysis. pages N–N.
- Roben Delos Reyes, Daniel Capurro, and Nicholas Geard. Generating synthetic electronic health record data using agent-based models to evaluate machine learning robustness under mass casualty incidents. pages N–N.
- Armaan Grewal. ML-powered triage and queue optimization for resource-constrained free clinics. pages N–N.
- Hongtao Hao, Joseph L Austerweil, and Alzheimer’s Disease Neuroimaging Initiative. Tempo: Transformers for temporal disease progression from cross-sectional data. pages N–N.
- Ethan Harvey, Dennis Johan Loevlie, Amir Ali Satani, Wansu Chen, and Michael C Hughes. A multi-dataset benchmark of multiple instance learning for 3d neuroimage classification. pages N–N.
- Iris Szu-Szu Ho, Lars Werne, Konrad Rawlik, Bruce Guthrie, and Sohan Seth. Learning under extreme label imbalance in ehrs: A dependency-aware loss for multi-label classification. pages N–N.
- Natalia Hong, Krishnarajah Nirantharakumar, and Christopher Yau. Structured treatment modeling in deep survival analysis via hazard factorization. pages N–N.
- Aditya Kommineni, Emily Zhou, Kleantlis Avramidis, Tiantian Feng, and Shrikanth Narayanan. A multi-dimensional framework for evaluating generalization in eeg foundation models. pages N–N.
- Ehsan Latif, Aleema Faisal, Shaheer Hammad, Dayyan Ali Akhtar, Agha Ali Raza, and Ihsan Ayyub Qazi. Evaluating robustness of llm-based ambient scribes for soap note generation. pages N–N.
- David Seung U Lee, Seeun Park, Seoyoon Jang, Suny-oung Lee, Chaeyoung Chang, Sungwook Choi, and Howard Lee. Generation of bilingual synthetic clinical notes for realistic data augmentation. pages N–N, a.
- Jun-Min Lee, Meong Hi Son, and Edward Choi. H-adminsim: A multi-agent simulator for realistic hospital administrative workflows with fhir integration. pages N–N, b.
- L Julian Lechuga Lopez, Farah E Shamout, and Tim G J Rudner. An empirical analysis of calibration and selective prediction in multimodal clinical condition classification. pages N–N.
- Aishwarya Mandyam, Shengpu Tang, Jiayu Yao, Jenna Wiens, and Barbara E Engelhardt. Candor: Counterfactual annotated doubly robust off-policy evaluation. pages N–N.
- Abdurahman Ali Mohammed and Wallapak Tavapanpong. Proto4dme: Interpretable cell counting via additive prototype density decomposition and optimal-transport coverage. pages N–N.
- Jong Hak Moon, Geon Choi, Paloma Rabaey, Min Gwan Kim, Jung-Oh Lee, Hyuk Gi Hong, Eun Woo Doe, Hangyul Yoon, Jiyoun Kim, Harshita Sharma, Daniel C Castro, Javier Alvarez-Valle, and Edward Choi. Llanguage: A benchmark for structured and sequential chest x-ray interpretation. pages N–N.
- Amirreza Naziri, Arash Asgari, Aijun An, Eleftherios Sachlos, and Laleh Seyyed-Kalantari. Mam: Multinomial attention masking for foundation models on sparse single-cell rna-seq data. pages N–N.
- Shahriar Noroozizadeh and Jeremy Weiss. Reconstructing sepsis trajectories from clinical case reports using llms: the textual time series corpus for sepsis. pages N–N.
- YongKyung Oh, Henry W Zheng, Jeffrey Feng, and Alex A T Bui. Survey-aware machine learning: A guideline for valid population health inference based on scoping review. pages N–N.
- OpenReview. Openreview, 2026. URL <https://openreview.net/>. Accessed: 2026-06-10.
- Okan Bilge Ozdemir and Ruowang Li. Bridging the reliability gap: Int8 quantization effects on discrimination and calibration in medical imaging. pages N–N.

Sasha Ronaghi, Chloe Stanwyck, Asad Aali, Amir Ronaghi, Miguel Angel Fuentes Hernandez, Tina Hernandez-Boussard, and Emily Alsentzer. Training-free adaptation of new-generation llms using legacy clinical models. pages N–N.

Binesh Sadanandan and Vahid Behzadan. Mechanistically guided lora improves paraphrase consistency in medical vision-language models. pages N–N.

Natalie Seah, Danielle S Bitterman, Daphna Spiegel, and Thomas Hartvigsen. Can language models identify side effects of breast cancer radiation treatments? pages N–N.

Yongsen Tan, Zhecheng Sheng, Xiruo Ding, Serguei V S Pakhomov, and Trevor Cohen. Decondtn-toolkit: A library for evaluation and enhancement of robustness to provenance shift. pages N–N.

Andrew Wang, Jiashou Zhang, and Michael Oberst. Revisiting performance claims for chest x-ray models using clinical context. pages N–N, a.

Minjia Wang, Luyang Luo, Sung Eun Kim, Fang Cao, David A Kim, and Pranav Rajpurkar. Open-ended clinical text generation for acute care: Applying reinforcement learning with clinically grounded rewards. pages N–N, b.

Tsai-Ning Wang, Herman Teun den Dekker, Lin-Lin Chen, Neil Zeghidour, and Aaqib Saeed. Adaptive test-time scaling for zero-shot respiratory audio classification. pages N–N, c.

Xucheng Wang, Xiaoman Zhang, Ankit Pal, and Pranav Rajpurkar. Rexsonovqa: A video qa benchmark for procedure-centric ultrasound understanding. pages N–N, d.