MICHAEL S YAO michaelsyao.com michael.yao@pennmedicine.upenn.edu Google Scholar

RESEARCH STATEMENT

My research focuses on how we can build trustworthy, evidence-based ML systems for healthcare. I specifically work on leveraging prior knowledge and statistical methods to design algorithms for clinical decision making, diagnostic reasoning, and precision medicine.

EDUCATION

University of Pennsylvania, MD-PhD Bioengineering 2021 - Present
NIH F30 NRSA Fellow | HHMI-NIBIB T32 Fellow | Medical Education Concentration
Advised by Osbert Bastani and James Gee | PhD Thesis Link

University of Pennsylvania, MS Computer Science 2023 - 2025
Penn Engineering Teaching Award Recipient

California Institute of Technology, BS Applied Physics 2017 - 2021
Salutatorian | Full-Ride Merit Scholarship Awardee | Advised by Mikhail Shapiro

EXPERIENCE

ML Scientist Intern, **Genentech** San Francisco, CA (Hybrid) | 2025
Designed algorithms for optimizing personalized treatment strategies using LLMs.

Health VC Fellow, **25madison** New York City, NY | 2024
Bridged clinical and engineering teams to help drive clinical operations for stealth healthtech incubation. Led exploratory investment research into digital health sectors.

AI Clinical Fellow, **Glass Health** Remote | 2023-2024
Released and assessed clinical guideline articles as knowledge sources for large language models (LLM). Investigated applications of LLMs for medical education.

PhD Research Intern, **Microsoft Research** Redmond, WA | 2022
Developed ML methods for accelerated MRI imaging. Proposed novel techniques for better generalization of MRI image reconstruction models.

Software Engineer, **Hyperfine** Guilford, CT | 2021
Implemented and validated algorithms for more robust MRI signal acquisition and image post-processing in MR software across 25+ hospital sites.

SELECTED WORKS

[1] **Yao MS**, Bastani O, Andersson A, Biancalani T, Bentaieb A, Iriondo C. Knowledgeable language models as black-box optimizers for personalized medicine. **ICLR**. (2025). Link
[2] **Yao MS**, Chae A, Saraiya P, Kahn CE, Witschey WR, Gee JC, Sagreiya H, Bastani O. Evaluating acute image ordering for real-world patient cases via language alignment with radiological guidelines. **Nature Commun Med**. (2025). Link
[3] **Yao MS**, Huang L, Leventhal E, Sun C, Stephen SJ, Liou L. Leveraging datathons to teach AI in undergraduate medical education. **JMIR Med Educ**. (2025). Link
[4] **Yao MS**, Gee JC, Bastani O. Diversity by design: Leveraging distribution matching for offline model-based optimization. **ICML**. (2025). Link
[5] **Yao MS**, Zeng Y, Bastani H, Gardner J, Gee JC, Bastani O. Generative adversarial

model-based optimization via source critic regularization. **NeurIPS**. (2024). Link

[6] Wu Y, Liu Y, Yang Y, **Yao MS**, Yang W, Shi X, Yang L, Li D, Liu Y, Gee JC, Yang X, Wei W, Gu S. A concept-based interpretable model for the diagnosis of choroid neoplasias using multimodal data. **Nature Communications**. (2025). Link

[7] Yang Y, Gandhi M, Wang Y, Wu Y, **Yao MS**, Callison-Burch C, Gee JC, Yatskar M. A textbook remedy for domain shifts: Knowledge priors for medical image analysis. **NeurIPS (Spotlight)**. (2024). Link

[8] **Yao MS**†, Chae A†, MacLean MT, Verma A, Duda J, Gee JC, Torigian DA, Rader D, Kahn CE, Witschey WR, Sagreiya H. SynthA1c: Towards clinically interpretable patient representations for diabetes risk stratification. **Prime MICCAI**. (2023). Link

---

TEACHING

Instructor and Curriculum Lead, **Ethical Algorithms for the Modern Clinician** | Link
TA, **Distributed Computer Systems** (CIS 5050, Penn) | Spring 2025, Fall 2025
TA, **Principles of Deep Learning** (ESE 5460, Penn) | Fall 2024
TA, **Imaging Informatics** (EAS 5850, Penn) | Spring 2024, Summer 2024
Head TA, **Healthcare and Technology** (CIS 7000, Penn) | Fall 2023, Fall 2024
TA, **Diagnostic Ultrasound for Medical Students** (Penn) | 2023 - Present
TA, **Pre-Clinical Medicine** (Penn) | 2023 - Present
Head TA, **Applied Mathematics** (ACM 95ab, Caltech) | Winter 2021, Spring 2020
TA, **Graduate Classical Physics** (Ph 106a, Caltech) | Fall 2020
TA, **Quantum Physics** (Ph 12ab, Caltech) | Winter 2020, Fall 2019
TA, **Operating Systems** (CS 24, Caltech) | Fall 2019
TA, **Special Relativity, Electromagnetism** (Ph 1bc, Caltech) | Spring 2019, Winter 2019

SERVICE

Referee
AMIA Annual Symposium
npj Digital Medicine
Neural Information Processing Systems (NeurIPS)
International Conference on Learning Representations (ICLR)
International Conference on Machine Learning (ICML)
AAAI Conference on Artificial Intelligence
AHLI Conference on Health, Inference, and Learning (CHIL)

Organizations
Co-President, Penn HealthX Student Group
Technology Committee Vice-Chair, American Physician Scientists Association
Board Member, Medical Education Club, University of Pennsylvania SOM
AI Curriculum Task Force Member, University of Pennsylvania SOM
Admissions Committee, University of Pennsylvania SOM
Peer Tutor, University of Pennsylvania SOM
Director of Data Science & AI, MDplus

Editor-in-Chief, Caltech Undergraduate Research Journal
Volunteer Tutor, Caltech RISE Tutoring Program
Peer Tutor, Caltech Deans' Office
Student Body Representative, Caltech Academics and Research Committee

ADDITIONAL WORKS

[9] **Yao MS**. Large language model-based evaluation of the impact of gender in medical research. Preprint. (2026). Link

[10] Barinaga Z, Chae A, Hashimoto DA, Jiang J, Leventhal E, Meng S, Purcell M, **Yao MS**. Ethical algorithms for the modern clinician. **NeurIPS Education Program**. (2025). Link

[11] Meng S, Purcell M, **Yao MS**, Hashimoto D. Implementing a novel artificial intelligence in medicine elective curriculum. **AAMC Annual Meeting** (Poster). (2025).

[12] Gee JC, **Yao MS**. Effective structured information extraction from chest radiography reports using open-weights large language models. **Radiology** (Editorial). (2025). Link

[13] Chae A†, **Yao MS**†, Sagreiya H, Goldberg AD, Chatterjee N, MacLean MT, Duda J, Elahi A, Borthakur A, Ritchie MD, Rader D, Kahn CE, Witschey WR, Gee JC. Strategies for implementing machine learning algorithms in the clinical practice of radiology. **Radiology**. (2024). Link

[14] **Yao MS**, Hansen MS. A path towards clinical adaptation of accelerated MRI. **Proc ML4H**. (2022). Link

[15] Abedi MH†, **Yao MS**†, Mittelstein DR, Bar-Zion A, Swift MB, Lee-Gosselin A, Barturen-Larrea P, Buss MT, Shapiro MG. Ultrasound-controllable engineered bacteria for cancer immunotherapy. **Nature Communications**. (2022). Link | Patent

[16] **Yao MS**, Uhr L, Daghlian G, Amrute JM, Deshpande R, Mathews B, Patel SA, Henri R, Liu G, Reierson K, Johnson G. Demonstration of a longitudinal medical education model (LMEM) to teach point-of-care ultrasound in resource-limited settings. **POCUS Journal**. (2020). Link