Aashna P. Shah

aashnapshah | aashna-p-shah | aashnashah@g.harvard.edu | +1.516.633.4355

Education

Harvard University June 2021 – May 2026
Ph.D. in Systems, Synthetic, and Quantitative Biology
Dissertation: Redefining Normal for Clinical Medicine
Advisors: Arjun Manrai, Ph.D. and Chirag Patel, Ph.D.

Northeastern University June 2021
B.S. in Mathematics (GPA: 3.94/4.0, Summa Cum Laude)

Academic & Professional Experience

Harvard Medical School, Department of Biomedical Informatics Sep 2022 – Present
Graduate Student Researcher *(Arjun Manrai, Chirag Patel)*

Simons Foundation, Center for Computational Biology Jun 2020 – Jan 2021
Machine Learning Researcher *(Richard Bonneau)*

Harvard University, Department of Systems Biology Jun 2018 – Dec 2021
Research Assistant *(Debora Marks)*

Merck Pharmaceuticals, Research & Discovery Jun 2019 – Dec 2019
Computational Chemist Co-op

Wyss Institute for Biologically Inspired Engineering Jun 2016 – June 2019
Research Assistant *(Pamela Silver)*

Selected Research Projects

- **Personalized Interpretation of Routine Blood Tests:** Developed **NORMA**, a transformer-based conditional model trained on 100k MIMIC and EHRSHot longitudinal sequences, evaluated on 1.9 billion laboratory measurements from 4.9 million adults in the Clalit Health Services database. NORMA predicts personalized laboratory reference ranges, capturing within-person physiological dynamics and outperforming static population thresholds in predicting clinical outcomes such as mortality, type 2 diabetes, and chronic kidney disease.

- **Generalist LLMs for Enhancing Patient-Doctor Communication:** Developed HealthCoach, a chatbot that guides patients through structured visit preparation and generates clinical pre-notes using patient input. Fine-tuned the Llama model on 7,000 annotated reports based on the Patient–Clinician Assessment Tool (PCAT) to improve alignment between patient narratives and clinical documentation. Designed prompting and meta-prompting workflows to elicit visit priorities, recent history, and concerns, and built a backend that produces clinician-facing summaries and updates EHR records with patient feedback using RAG retrieval and real-time data integration

- **Proxy Identification of Demographics in Clinical Algorithms:** Developed **ARC**, a framework to identify explanatory factors underlying group-level differences in clinical equations. Applied ARC to pulmonary function testing across 159,893 participants to evaluate alternatives to race-based adjustments. Introduced ARC_{PFT}, a new reference equation incorporating sitting height and waist circumference, which outperformed the race-neutral GLI-Global standard with up to 24% lower mean absolute error and improved robustness across diverse populations.

Publications

1. **Shah, A.P.**, Diao, J. Pierson, E., Patel, C., Manrai, A.K. Disentangling Proxies of Demographic Adjustments in Clinical Equations *In review.*

2. Diao, J.A., Movva, R., Cheng, L., Kadoma, K., **Shah, A.P.**, Ferryman, K., Manrai, A.K., Pierson, E. A national survey of patient preferences regarding the use of race in clinical algorithms. *In review.*

3. Buckley, T.A., Conci, R., Brodeur, P.G., Gusdorf, J., Beltrán, S., Behrouzi, B., Crowe, B., Dockterman, J., Muham-

mad, M., Ohnigian, S., Sanchez, A., Diao, J.A., **Shah, A.P.**, et al. Advancing medical artificial intelligence using a century of cases. *arXiv*, 2025.

4. Sagers, L.†, **Shah, A.P.†** et al. Directing generalist vision–language models to interpret medical images across populations. *NeurIPS 2024 Workshop on GenAI4Health.*

5. Buckley, T.A., Kanjee, Z., Crowe, B., Pettinato, A.M., **Shah, A.P.†** et al. Automated assessment of large language models in open-ended medical prompts. *Manuscript in preparation.* Demo: perceptron-md.streamlit.app.

6. **Shah, A.P.**, Diao, J., Manrai, A.K., Patel, C. TRACE: A data-driven framework for explaining subgroup detection from medical imaging. *Manuscript in preparation.*

7. Gibbs, C., Jackson, C., Saldi, G., **Shah, A.P.** et al. Single-cell gene regulatory network inference at scale: The Inferelator 3.0. *Bioinformatics*, 2022.

8. Ziesack, M., Rollins, N., **Shah, A.P.** et al. Chimeric fatty acyl–acyl carrier protein thioesterases provide mechanistic insight into enzyme specificity and expression. *Applied and Environmental Microbiology*, 2018.

Skills

Programming Languages & Tools: Python (PyTorch, TensorFlow, Keras, W&B, NumPy, Pandas, scikit-learn, Matplotlib, Flask), R (tidyverse, dplyr, caret, ggplot2), Git, SQL, Docker, Bash, HTML

Relevant Coursework: Machine Learning, Advanced Data Science, Probability and Statistics, Stochastic Processes, Systems Development, Computer Science I–III, Real Analysis, Linear Algebra

Community & Outreach

Ethics Consultant for Conduct and Communication of Science *Sep 2024 – Nov 2024*
Department of Biomedical Informatics, Harvard Medical School

Teaching Fellow, Data Science for Medical Decision Making *Jan 2024 – May 2024*
Harvard Medical School

Peer Reviewer *Jun 2023 – Present*
NEJM AI, *Nature Medicine*

Preliminary Qualifying Exam Support Group Fellow *Jun 2022 – Present*
Department of Systems Biology, Harvard Medical School

Seminars & Symposiums

Speaker, Biomedical Informatics Science Day, Harvard Medical School, September 2025

Attendee, Women in Machine Learning (WiML) @ NeurIPS, Vancouver, December 2024

Poster, NeurIPS Demo Track, GenAI for Health Workshop, Vancouver, December 2024

Live Demo, NeurIPS Demo Track, GenAI for Health Workshop, Vancouver, December 2024

Poster, Biomedical Informatics Science Day, Harvard Medical School, September 2024 – **Awarded Best Poster**

Poster, Doctoral Symposium, Conference for Health and Inference Learning (CHIL), June 2024

Speaker, Pranav Rajpurkar Lab Meeting, Harvard Medical School, December 2023

Speaker, Computational Health Informatics Program (CHIP), Boston Children's Hospital, January 2024

Speaker & Poster, Systems Biology Symposium, Harvard University, October 2023

Speaker & Poster, Spring Systems Biology Symposium, Harvard University, October 2023

Volunteer, Precision Medicine Conference, September 2023