

Emily Alsentzer

Curriculum Vitae

✉ ealsentzer@stanford.edu
🌐 emilyalsentzer.com
machine learning for healthcare

Education

- 2022 **Massachusetts Institute of Technology & Harvard Medical School, Boston, MA,** Ph.D.
Harvard-MIT Division of Health Sciences and Technology, Medical Engineering and Medical Physics
Advisors: Peter Szolovits (MIT CSAIL) and Isaac Kohane (HMS DBMI)
- 2017 **Stanford University, Stanford, CA,** M.S.
Biomedical Informatics
- 2016 **Stanford University, Stanford, CA,** B.S.
Computer Science, With Distinction

Honors

- 2023 Conference on Health Inference and Learning (CHIL) Best Paper Award
- 2023 ISMB/ECCB Best Oral Presentation Award
- 2023 Microsoft Accelerate Foundation Models Research Grant Recipient
- 2020-2022 Microsoft Research PhD Fellowship (1 of 10 awardees in the U.S.)
- 2022 ICLR Highlighted Reviewer
- 2020 NeurIPS Top Reviewer
- 2018-2019 National Library of Medicine BIRT Fellow, Harvard Medical School
- 2016-2017 National Library of Medicine Training Grant Recipient, Stanford University
- 2016 Tau Beta Pi, Stanford University
- 2012 Intel Science Talent Search Semifinalist
- 2011 Siemens "We Can Change the World" Challenge, 3rd Place Overall

Research Experience (Selected)

- 2022-2024 **Postdoctoral Fellowship with David Bates, Brigham and Women's Hospital.**
Postdoctoral Fellow. Address the challenges of integrating AI into clinical workflows through research and clinical deployment initiatives. Focus on the development and evaluation of fair, trustworthy, and clinically useful large language models (LLMs) in healthcare. Spearhead the evaluation of two Mass General Brigham LLM pilots. Advisor: Dr. David Bates

2017–2022 **Isaac Kohane Lab**, *Harvard Department of Biomedical Informatics*.

Clinical Decision Making Group, *MIT Computer Science & Artificial Intelligence Laboratory*.

PhD Student. Designed, implemented, and analyzed machine learning and natural language processing algorithms for use in biomedicine. Focused on development of trustworthy systems that integrate clinical domain knowledge, excel in settings with limited data, leverage simulated patient data and clinical pretraining, and promote equitable healthcare. Medical applications include rare disease diagnosis and electronic health record (EHR) summarization and information extraction. Passionate about the development and release of open-source clinical models. Advisors: Dr. Isaac Kohane and Dr. Peter Szolovits

2019 **Microsoft Research**.

Developed verifiable, human-in-the-loop approaches for automatically extracting information for breast cancer registries from electronic health record (EHR) notes. Advisors: Dr. Tristan Naumann and Dr. TJ Hazen

2017 **Verily Life Sciences**.

Developed trustworthy NLP approaches to automatically summarize longitudinal electronic health records. Gave several research talks, including talk on ontology-guided clinical machine learning. Advisor: Dr. Paul Varghese

Clinical Experience

2021-2022 **Introduction to Clinical Medicine**, *Mount Auburn Hospital*, Cambridge, MA.

Two six week clinical rotations at Mount Auburn Hospital. Learned how to perform a history and physical exam, construct a differential diagnosis, generate an assessment and plan, and write admission, progress, and discharge notes. Participated as a member of the care team in Internal Medicine for five weeks. Assigned 1-2 patients each day to interview and examine, present on during morning rounds, and document in the electronic health record.

Work Experience

2023 Consultant for Clinical NLP Startups

2019 Research Intern, Microsoft Research

2018-2019 Consultant, Generate Biomedicines

2017 Research Intern, Verily Life Sciences

2016 Research Intern, HealthMap, Boston Children's Hospital

2015 Research Intern, Giulio De Leo Lab, Stanford Hopkins Marine Station

2014 Research Intern, Military Division of Tropical Medicine

2013 Research Assistant, Vanderbilt Vaccine Research Program, Vanderbilt University

Papers

I have published papers at **NeurIPS**, **NAACL**, **Lancet Digital Health**, **Nature Communications**, **NPJ Digital Medicine**, **the Conference on Health, Inference, and Learning (CHIL)**, and **the Pacific Symposium on Biocomputing (PSB)**.

[1] **Emily Alsentzer***, Michelle M Li*, Shilpa N Kobren, Undiagnosed Diseases Network, Isaac S Kohane, and Marinka Zitnik. Deep learning for diagnosing patients with rare genetic diseases. *Under Review*. Also presented at ISMB/ECCB and the Symposium on Artificial Intelligence for Learning Health Systems (SAIL).

ISMB/ECCB 2023 Best Oral Presentation Award.

[2] Jorge A. Rodriguez, **Emily Alsentzer**, and David W. Bates. Leveraging large language models to foster equity in healthcare. *Journal of the American Medical Informatics Association*, In Press, 2024.

[3] Travis Zack, Eric Lehman, Mirac Suzgun, Jorge A Rodriguez, Leo Anthony Celi, Judy Gichoya, Dan Jurafsky, Peter Szolovits, David W Bates, Raja-Elie E Abdunour, Atul J. Butte, and **Emily Alsentzer**. Assessing the potential of gpt-4 to perpetuate racial and gender biases in health care: a model evaluation study. *The Lancet Digital Health*, 6(1):e12–e22, 2024. Also presented at the Symposium on Artificial Intelligence for Learning Health Systems (SAIL).

Covered by STAT News, BWH News, and MIT Jameel Clinic for AI & Health.

[4] **Emily Alsentzer**, Sam G Finlayson, Michelle M Li, Undiagnosed Diseases Network, Shilpa N Kobren, and Isaac S Kohane. Simulation of undiagnosed patients with novel genetic conditions. *Nature Communications*, 14(1):6403, 2023.

[5] **Emily Alsentzer**, Matthew J. Rasmussen, Romy Fontoura, Alexis L. Cull, Brett Beaulieu-Jones, Kathryn J. Gray, David W. Bates, and Vesela P. Kovacheva. Zero-shot interpretable phenotyping of postpartum hemorrhage using large language models. *npj Digital Medicine*, 6(1):1–10, November 2023.

[6] Eric Lehman, Evan Hernandez, Diwakar Mahajan, Jonas Wulff, Micah J Smith, Zachary Ziegler, Daniel Nadler, Peter Szolovits, Alistair Johnson, and **Emily Alsentzer**. Do we still need clinical language models? In *Conference on Health, Inference, and Learning*. PMLR, 2023.

Best Paper Award (Top 3 Submissions), Oral Presentation.

[7] Brett K Beaulieu-Jones, Mauricio F Villamar, Phil Scordis, Ana Paula Bartmann, Waqar Ali, Benjamin D Wissel, **Emily Alsentzer**, Johann de Jong, Arijit Patra, and Isaac Kohane. Predicting seizure recurrence after an initial seizure-like episode from routine clinical notes using large language models: a retrospective cohort study. *The Lancet Digital Health*, 5(12):e882–e894, 2023.

[8] Hammaad Adam, Aparna Balagopalan, **Emily Alsentzer**, Fotini Christia, and Marzyeh Ghassemi. Mitigating the impact of biased artificial intelligence in emergency decision-making. *Communications Medicine*, 2(1):149, 2022.

Covered by MIT News.

[9] Hammaad Adam, Aparna Balagopalan, **Emily Alsentzer**, Fotini Christia, and Marzyeh Ghassemi. Just following ai orders: When unbiased people are influenced by biased ai. In *Workshop on Trustworthy and Socially Responsible Machine Learning, NeurIPS*, 2022.

[10] Griffin Adams, **Emily Alsentzer**, Mert Ketenci, Jason Zucker, and Noémie Elhadad. What’s in a summary? laying the groundwork for advances in hospital-course summarization. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, June 2021.

[11] Subhrajit Roy, Stephen Pfohl, Girmaw Abebe Tadesse, Luis Oala, Fabian Falck, Yuyin Zhou, Liyue Shen, Ghada Zamzmi, Purity Mugambi, Ayah Zirikly, Matthew BA McDermott, and **Emily Alsentzer**. Machine learning for health (ml4h) 2021. In *Machine Learning for Health Symposium*. PMLR, 2021.

[12] **Emily Alsentzer***, Samuel Finlayson*, Michelle M Li, and Marinka Zitnik. Sub-graph neural networks. In *Advances in Neural Information Processing Systems*, volume 33, pages 8017–8029, 2020.

[13] Irene Y Chen, **Emily Alsentzer**, Hyesun Park, Richard Thomas, Babina Gosangi, Rahul Gujrathi, and Bharti Khurana. Intimate partner violence and injury prediction from radiology reports. In *Proceedings of the Pacific Symposium on Biocomputing*, pages 55–66, 2020.

Spotlight Presentation.

[14] William Boag, Tzu-Ming Harry Hsu, Matthew McDermott, Gabriela Berner, **Emily Alsentzer**, and Peter Szolovits. Baselines for chest x-ray report generation. In *Proceedings of the Machine Learning for Health Workshop at NeurIPS*. PMLR, 2020.

[15] **Emily Alsentzer**, Sarah-Blythe Ballard, Joan Neyra, Delphis M Vera, Victor B Osorio, Jose Quispe, David L Blazes, and Luis Loayza. Assessing 3 outbreak detection algorithms in an electronic syndromic surveillance system in a resource-limited setting. *Emerging infectious diseases*, 2020.

[16] Suproteem K Sarkar, Subhrajit Roy, **Emily Alsentzer**, Matthew BA McDermott, Fabian Falck, Ioana Bica, Griffin Adams, Stephen Pfohl, and Stephanie L Hyland. Machine learning for health (ml4h) 2020: Advancing healthcare for all. In *Machine Learning for Health Workshop at NeurIPS*. PMLR, 2020.

[17] **Emily Alsentzer**, John Murphy, William Boag, Wei-Hung Weng, Di Jin, Tristan Naumann, and Matthew McDermott. Publicly available clinical BERT embeddings. In *Proceedings of the 2nd Clinical Natural Language Processing Workshop*. Association for Computational Linguistics, 2019. https://huggingface.co/emilyalsentzer/Bio_ClinicalBERT

Oral Presentation, 61,195,735 total downloads from Huggingface as of March 2023.

[18] Adrian V Dalca, Matthew BA McDermott, **Emily Alsentzer**, Samuel G Finlayson, Michael Oberst, Fabian Falck, Corey Chivers, Andrew Beam, Tristan Naumann, and Brett Beaulieu-Jones. Machine learning for health (ml4h) 2019: What makes machine learning in medicine different? In *Machine Learning for Health Workshop at NeurIPS*. PMLR, 2019.

[19] Albee Y Ling, **Emily Alsentzer**, Josephine Chen, Juan M Banda, Suzanne Tamang, and Evan Minty. Scalable electronic phenotyping for studying patient comorbidities. In *AMIA Annual Symposium Proceedings*. American Medical Informatics Association, 2018.

[20] Yulin Hswen, Kara C Sewalk, **Emily Alsentzer**, Gaurav Tuli, John S Brownstein, and Jared B Hawkins. Investigating inequities in hospital care among lesbian, gay, bisexual, and transgender (lgbt) individuals using social media. *Social Science Medicine*, 2018.

[21] **Emily Alsentzer** and Anne Kim. Extractive summarization of ehr discharge notes. *arXiv preprint arXiv:1810.12085*, 2018.

[22] Joshua S Lichtman, **Emily Alsentzer**, Mia Jaffe, Daniel Sprockett, Evan Masutani, Elvis Ikwa, Gabriela K Fragiadakis, David Clifford, Bevan Emma Huang, Justin L Sonnenburg, Kerwyn Casey Huang, and Joshua E Elias. The effect of microbial colonization on the host proteome varies by gastrointestinal location. *The ISME journal*, 2016.

[23] **Emily Alsentzer**, Delphis M Vera, Joan Neyra, Luis Loayza, Ricardo A Hora, Victor B Osorio, Jose Quispe, Sarah-Blythe Ballard, and David Blazes. Monitoring acute diarrhea via an electronic surveillance system in the peruvian navy. *Online journal of public health informatics*, 2015.

[24] Deepa Joshi, **Emily Alsentzer**, Kathryn Edwards, Allison Norton, and Sarah Elizabeth Williams. An algorithm developed using the brighton collaboration case definitions is more efficient for determining diagnostic certainty. *Vaccine*, 2014.

Book Chapters

Peter Szolovits and **Emily Alsentzer**. "Knowledge-Based Systems in Medicine". *Intelligent Systems in Medicine and Health: The Role of AI*. Springer, 2022. pp. 75–108.

Invited Talks

- 2024 "Leveraging Large Language Models for Phenotyping and Clinical Decision Support", Mass General Brigham CAIBILS Invited Talk. April, 2024.
- 2024 "Evolving Landscape of Large Language Models in Healthcare", Mass General Brigham Clinical Informatics Fellowship Seminar. April, 2024.
- 2024 "Knowledge-Enhanced Machine Learning for Healthcare", Broad Institute of Harvard and MIT, Machine Learning for Healthcare Invited Talk. April, 2024.
- 2024 Columbia University, Department of Biomedical Informatics. March, 2024.
- 2024 Carnegie Mellon University, School of Computer Science. March, 2024.
- 2024 University of California Los Angeles, Department of Computational Medicine. March, 2024.
- 2024 Harvard Medical School, Department of Biomedical Informatics. March, 2024.
- 2024 University of California San Francisco, Division of Clinical Informatics and Digital Transformation. March, 2024.
- 2024 Stanford University, Department of Biomedical Data Science. February, 2024.
- 2024 Duke University, Department of Biostatistics and Bioinformatics. February, 2024.
- 2024 Vanderbilt University, Department of Biomedical Informatics. February, 2024.
- 2024 University of Washington, Department of Biomedical Informatics and Medical Education. January, 2024.
- 2023 Responsible AI for Social and Ethical Healthcare (RAISE) Symposium, Invited Participant. October, 2023.
- 2022 "Few Shot Learning for Rare Disease Diagnosis", HMS-Clalit Research Institute Symposium, Harvard Medical School. December, 2022.

- 2022 “Machine Learning for Rare Diseases”, Emory University, Computer Science Seminar Series. December, 2022.
- 2022 “Promise and Challenges of Language Models in the Clinical Domain”, Broad Institute of Harvard and MIT, Models Inference and Algorithms (MIA) Seminar. November, 2022.
- 2021 “Applying State of the Art Language Models to Enable Better Clinical Natural Language Processing”, AMIA 2021, Panel. November, 2021.
- 2021 “Promise and Challenges of BERT in the Clinical Domain”, Brigham and Women’s Hospital MTERMS Seminar. November, 2021.
- 2020 “Clinical BERT for rapid cohort creation.” MERCK-Harvard Medical School Steering Committee. December 2020.
- 2020 “Geometric Deep Learning for Healthcare”, University of Massachusetts Lowell COMP5800, Guest Lecture. November, 2020.
- 2020 “Clinical BERT Embeddings”, Verily Life Sciences, Guest Lecture. January, 2020.
- 2019 “EHR Annotation for Clinical Trial Recruitment”, Allen Institute for AI AllenNLP Summit. August, 2019.
- 2016 “Parasite Dynamics in a Connected World: The effect of network topology and human mobility on schistosomiasis transmission and control”, Stanford Global Health Research Convening. January, 2016.
- 2014 “Monitoring Acute Diarrhea via an Electronic Surveillance System in the Peruvian Navy”, International Society for Disease Surveillance. December, 2014.

Teaching

- 2019 **Teaching Assistant**, *Department of Biomedical Informatics*, Harvard University.
 Deep Learning for Biomedical Data (Prof. Kun Hsing Yu).
 Held linear algebra and calculus review sessions, held office hours, and graded problem sets and final projects. The course includes both foundational lectures on computer vision and natural language processing and applied lectures on applications in biomedicine.
- 2015 **Course Assistant**, *Department of Biomedical Informatics*, Stanford University.
 BIOMEDIN 214: Algorithms for Computational Biology (Prof. Russ Altman).
 Graded and provided feedback on problem sets and quizzes on genetic alignment, molecular dynamics, drug-ligand networks, and other computational biology fundamentals.
- 2013-2015 **Section Leader**, *Department of Computer Science*, Stanford University.
 CS 106A: Programming Methodology (Prof. Mehran Sahami).
 Taught CS106A section of 15-20 students, graded assignments, staffed the LAIR, which is a tutoring service for CS 106 students, and held weekly 1:1 interactive grading sessions to help students improve the functionality and style of their code

Mentorship and Outreach

- 2022-2023 Research Mentor for MIT PhD student and UCSF Research Fellow
- 2021-2022 MIT-Harvard Women in Artificial Intelligence Mentorship Program
 - 2021 Society for the Advancement of Chicanos/Hispanics and Native Americans in Science Conference PhD Outreach
- 2020-2021 MEMP PhD Application Assistance Mentorship Program

2018 Research Mentor for John Murphy (MIT undergrad)

Funding

- 2023 Microsoft Accelerate Foundation Models Research Grant Program. Awarded funding to study the use of large language models for verifiable summarization of EHRs.
- 2020-2022 Microsoft Research PhD Fellowship. Full tuition and stipend for two years of PhD research. One of ten awardees across the United States.
- 2018-2019 NLM Biomedical Informatics and Data Science Research Training (BIRT) Fellow, Harvard University.
- 2016-2017 NLM T15 Biomedical Informatics and Data Science Research Training Fellow, Stanford University.

Conference Organization

- 2020–2024 Founding Organizer and Program Committee Member, Symposium on Artificial Intelligence for Learning Health Systems
 - 2023 Program Chair, Conference on Health, Inference, and Learning (CHIL)
 - 2022 Communications Chair, Conference on Health, Inference, and Learning (CHIL)
 - 2021 General Chair, Machine Learning for Health (ML4H) Symposium
- 2020-2021 Proceedings Chair, Conference on Health, Inference, and Learning (CHIL)
- 2019-2020 Program Chair, Machine Learning for Health (ML4H) Workshop at the Conference on Neural Information Processing Systems (NeurIPS)

Journal Reviewer

- 2023-24 New England Journal of Medicine (NEJM) AI
 - 2023 Nature
 - 2023 Nature Communications
 - 2023 PLOS Digital Health
- 2022 Journal of the American Informatics Association (JAMIA)
- 2022 Journal of Patient Safety
- 2020 NPJ Digital Medicine
- 2020 Bioinformatics
- 2020 JAMIA Open
- 2020 ACM Transactions on Computing for Healthcare

Conference Reviewer

- 2024 COLM
- 2020-21,2023 NeurIPS
- 2019-20,2023 Machine Learning For Healthcare Conference
 - 2020, 2023 ClinicalNLP Workshop at EMNLP
 - 2022 EMNLP

- 2021 ICLR
- 2021 ICML
- 2020-2021 ACL
- 2021 ISMB
- 2020 Pacific Symposium for Biocomputing
- 2020 Medical Conversations Workshop at ACL
- 2019 American Medical Informatics Association (AMIA)
- 2018 Machine Learning for Health Workshop at NeurIPS

Leadership

- 2021-present **Founding Board Member**, *Association for Health Learning and Inference (AHLI)*.
AHLI (ahli.cc) is a non-profit dedicated to building the machine learning and health community. AHLI oversees the ML4H and CHIL conferences and hosts health data challenges. Served terms as Secretary and Treasurer for AHLI.
- 2021 **General Chair**, *Machine Learning for Health (ML4H) Symposium*.
Led a team of over 20 organizers, expanded ML4H from a NeurIPS workshop to a stand alone symposium, oversaw peer review process for over 150 submissions, speaker selection, website development, and submission and research mentorship initiatives
- 2015-2016 **Editor in Chief**, *Stanford Undergraduate Research Journal*.
Led a team of 50 staff members to produce a peer-reviewed research journal that publishes articles in engineering, natural sciences, social sciences, and humanities. Coordinated journal design, submissions, distribution, and outreach and cultivated relationships with Stanford University
- 2015-2016 **Team Coordinator**, *Stanford Partners in Health Engage*.
Led Stanford's PIH Engage chapter to fundraise for Partners in Health, advocate for global health policies, and educate the Stanford community about global health issues



Emily Alsentzer

Last update: August 27, 2024