emilya@mit.edu

# Emily Alsentzer

MIT/Harvard Medical School PhD Candidate
Machine Learning + Medicine

## EDUCATION

**Massachusetts Institute of Technology / Harvard Medical School** – Cambridge, MA (9/17-Current)
- PhD, Medical Engineering and Medical Physics (MEMP), Harvard-MIT Health Science and Technology Program
- Computer science classes at MIT and medical school classes + 2 six-week clinical rotations at Harvard Medical School

**Stanford University** – Stanford, CA (9/12-5/17)
- MS, Biomedical Informatics
- BS, Computer Science, With Distinction, Tau Beta Pi

**Hume Fogg Academic High School** – Nashville, TN (8/08-5/12)
- Valedictorian and Distinguished Scholars Diploma

## RESEARCH & INTERNSHIP EXPERIENCE

**PhD Student, co-advised by Isaac Kohane (Harvard Medical School) & Pete Szolovits (MIT), Boston, MA (9/17-current)**
- Developing actionable machine learning methods for healthcare that leverage clinical domain knowledge
- Current focus is developing graph neural network algorithms to diagnose patients with rare genetic conditions in the Undiagnosed Diseases Network

**Intern, Microsoft Research, Cambridge, MA (Summer 2019)**
- Developed NLP models to extract information for breast cancer registries from EHR notes using deep learning
- Collaborated with clinicians to leverage the models to accelerate EHR annotation for cancer registries

**Consultant, Flagship Pioneering VL56, Cambridge, MA (11/18-6/19)**
- Quantified uncertainty in deep learning models for drug development

**Intern, Verily (Google Life Sciences), Cambridge, MA (Summer 2017)**
- Developed NLP methods to automatically summarize patients' electronic health records
- Presented work at a Verily-wide poster session and gave an Analysis Review talk, final internship presentation, and an ontology guided feature engineering review talk

**Intern, HealthMap, Boston Children's Hospital, Harvard University, Boston, MA (6/16-1/17)**
- Applied machine learning and NLP for topic classification of tweets related to patient experience at hospitals
- Developed library for Twitter topic classification that can be applied across projects at HealthMap
- Assessed LGBT healthcare quality using patient experience data from Twitter

**Intern, Giulio De Leo Lab, Stanford Hopkins Marine Station, Monterey, CA (Summer 2015)**
- Applied network theory to model the transmission of schistosomiasis, a water-borne infectious disease, using ODEs
- Demonstrated the influence of human mobility and network topology in disease transmission, which challenged the traditional assumption of homogenous mixing in populations
- Developed novel, data driven policy recommendations for mass drug administration and snail control

**Intern, African Federation for Emergency Medicine, Cape Town, South Africa (4/15-6/15)**
- Analyzed website traffic and use of AFEM emergency medicine teaching materials using Google Analytics
- Developed graphics and content for AFEM website and created SQL database to maintain membership records

**Intern, Military Division of Tropical Medicine, Bethesda, MD (Summer 2014)**
- Developed algorithms for detection of acute diarrheal disease outbreaks in an electronic disease surveillance system in the Peruvian Navy at NAMRU-6 in Lima, Peru
- Took a month long course in tropical medicine and travelled to Lima and Iquitos to learn about treatment of dengue, malaria, and other infectious diseases

**Research Assistant, Vijay Pande Lab, Stanford University, Stanford, CA (10/13 – 12/14)**
- Identified potential drugs effective against West Nile Virus by screening a cheminformatics database for promising hits using ROCS computational software and validated via experimental collaboration

emilya@mit.edu

**Research Assistant, Vanderbilt Vaccine Research Program, Nashville, TN (Summer 2013)**

- Developed case presentations, interpreted VAERS reports, and conducted background research for case discussions of the Clinical Immunization Safety Assessment (CISA), a program run by the CDC
- Helped develop a more user friendly and concise algorithm for assessing diagnostic certainty of adverse events following immunization and evaluated the efficacy and accuracy of the algorithm compared to the original Brighton case definition.

**Research Assistant, Richard Peek Lab, Vanderbilt University, Nashville, TN (Summer 2011)**

- Identified novel H. pylori adhesins responsible for binding the host receptor Decay Accelerating Factor (DAF)
- Used the following molecular techniques: immobilized metal affinity chromatography, silver staining and far western blotting, tandem mass spectrometry, and gene Splicing by overlap extension (gene SOEing)

## ACADEMIC SERVICE

**Conference General Chair:** ML4H Symposium (2021)
**Conference Organizer**: Conference on Health, Inference, and Learning (2020; 2021); ML4H Workshop at NeurIPS (2019; 2020); Symposium on Artificial Intelligence for Learning Health Systems (2020; 2021)
**Conference Reviewer:** ICLR (2021); ACL (2021;2020); ISMB (2021); ICML (2021); NeurIPS (2021; 2020); Pacific Symposium for Biocomputing (2020); Machine Learning For Healthcare Conference (2020;2019); American Medical Informatics Association Conference (2019); ML4H Workshop at NeurIPS (2018); ACL Medical Conversations Workshop (2020); EMNLP ClinicalNLP Workshop (2020)
**Journal Reviewer:** Nature Digital Medicine (2021); ACM Transactions on Computing for Healthcare (2020); Bioinformatics (2021); JAMIA Open (2021)

## TEACHING EXPERIENCE

**Deep Learning for Biomedical Data TA, Harvard Medical School, Boston, MA (3/19-5/19)**

- Hold linear algebra and calculus review session and office hours, grade problem sets and final projects

**Algorithms for Computational Biology Grader, Stanford University, Stanford, CA (9/15-12/15)**

- Graded problem sets (genetic alignment, molecular dynamics, drug-ligand networks, etc.) and quizzes

**CS106A Section Leader, Stanford University, Stanford, CA (9/13-3/15)**

- Taught CS106A section, graded assignments, staffed the LAIR, a tutoring service for CS 106 students, and held weekly interactive grading sessions to help students improve the functionality and style of their code

## MENTORSHIP EXPERIENCE

**MIT-Harvard Women in Artificial Intelligence Mentor, Cambridge, MA (2021)**

- Mentor students who are interested in pursuing research in machine learning

**MEMP Application Assistance Program (MAAP) Mentor, Cambridge, MA (2020)**

- Mentor students from underrepresented backgrounds who are applying to the MEMP PhD program

**Undergraduate Research Mentor, Cambridge, MA (2018)**

- Mentored an undergraduate student in Pete Szolovits' research group on clinical NLP

## LEADERSHIP EXPERIENCE

**General Chair, Machine Learning for Healthcare Symposium (1/21-12/21)**

- Led a team of over 20 people to organize the ML4H symposium, an annual event that publishes top research in the machine learning for healthcare community
- Oversaw review process, submission and research mentorship initiatives, speaker selection, website and virtualization stack creation, and many other aspects of the conference organizing process

**Editor in Chief, Stanford Undergraduate Research Journal, Stanford, CA (9/15-6/16)**

- 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

**Team Coordinator, Stanford Partners in Health Engage, Stanford, CA (9/15-6/16)**

- Led Stanford's PIH Engage chapter to fundraise for PIH Engage (>$9,000), advocate for global health policies, and educate the Stanford community about global health issues

## AWARDS

- Microsoft Research PhD Fellowship (2020-2022)
- NeurIPS Best Reviewer Award (2020)

- Intel Science Talent Search Semifinalist (2012)
- Siemens "We Can Change the World" Challenge, 3rd Place Overall (2011)
- National AP Scholar (2011)
- TN Junior Academy of Sciences, 2nd Place Presentation (2011)

## SKILLS & CLASSWORK

**Programming Languages/Libraries:** Python, Pytorch, Keras, Tensorflow, R, Matlab, SQL, Java, C++, C, Javascript, HTML/CSS
**Computer Science Classes:** Machine Learning, Natural Language Processing, Databases, Biomedical Computation, Algorithms for Computational Biology, Data Driven Medicine, Social and Information Networks, Mining Massive Datasets, Design and Analysis of Algorithms, Statistical Inference, Nonparametric Statistics
**Harvard Medical School Classes:** Pathology, Cardiovascular Pathophysiology, Genetics, Immunology, Renal Pathophysiology

## CLASS PROJECTS

**Social and Information Network Analysis, CS224W (9/16-12/16)**

- Characterized and predicted dogmatic communities on Reddit using network features with an AUC of 0.8
- Performed logistic regression, SVM, Random Forest, and Gradient Boosting classifiers as well as k-means clustering

**Biomedical Informatics Research Methodology, BIOMEDIN 212 (4/16-6/16)**

- Worked with Nigam Shah's lab to develop semi-supervised ML algorithms with noisy labeling for phenotyping of patients with comorbidities
- Used Observational Medical Outcomes Partnership (OMOP) common data model to develop algorithms for electronic health records that are portable across institutions
- Assessed the tradeoff between increasing labeling accuracy, model complexity, and training sample size to maximize algorithm scalability

**Computational Modeling of Microbial Communities, BIOE115 (4/14-6/14)**

- Analyzed mouse proteome in the stomach, jejunum, ileum, cecum and proximal colon under three colonization states (germ-free, colonized with Bacteroides thetaiotaomicron, and conventionally raised) to study proteomic abundance profiles along the gastrointestinal (GI) tract
- Performed K-means clustering, PCA, hierarchical clustering, random forest classification, and comparison of GO term abundances to assess differences in protein abundances

## PUBLICATIONS

**MANUSCRIPTS**:

- G Adams, E Alsentzer, M Ketenci, J Zucker, N Elhadad. What's in a Summary? Laying the Groundwork for Advances in Hospital-Course Summarization. NAACL. 2021.
- E Alsentzer, SG Finlayson, M Li, M Zitnik. Subgraph Neural Networks. NeurIPS. 2020.
- IY Chen, E Alsentzer, H Park, R Thomas, B Gosangi, R Gujrathi, B Khurana. Intimate Partner Violence and Injury Prediction From Radiology Reports. Pacific Symposium on Biocomputing. 2020.
- E Alsentzer, S Ballard, J Neyra, D Vera, V Osorio, J Quispe, D Blazes. Assessing 3 Outbreak Detection Algorithms in an Electronic Syndromic Surveillance System in a Resource-Limited Setting. Emerging Infectious Diseases. 2020 Sep;26(9):2196.
- W Boag, TH Hsu, M McDermott, E Alsentzer, G Berner, and Pete Szolovits. Baselines for Chest X-Ray Report Generation. Proceedings of the Machine Learning for Health Workshop at NeurIPS. 2019.
- E Alsentzer, J Murphy, W Boag, WH Weng, D Ji, T Naumann, M McDermott. Publicly Available Clinical BERT Embeddings. Proceedings of the Clinical Natural Language Processing Workshop at NAACL 2019. 2019; 72-78.
- WH Weng, E Alsentzer, D Jin, SY Min, P Raghavan and P Szolovits. Logical Form Information for Clinical Question Answering. DSHealth2019: KDD Workshop on Applied Data Science for Healthcare. 2019
- A Ling, E Alsentzer, J Chen, J Banda, S Tamang, E Minty. Scalable Electronic Phenotyping for Studying Patient Comorbidities. AMIA Annu Symp Proc. 2018; 2018:740-749.
- Y Hswen, K Sewalk, E Alsentzer, G Tuli, J Brownstein, J Hawkins. Investigating inequities in hospital care among lesbian, gay, bisexual, and transgender (LGBT) individuals using social media. Social Science & Medicine. 2018 Oct; 215:92-97.
- J Lichtman, E Alsentzer, M Jaffe, D Sprockett, E Masutani, E Ikwa, G Fragiadakis, D Clifford, B Huang, J Sonnenburg, K Huang, J Elias. The effect of microbial colonization on the host proteome varies by gastrointestinal location. ISMEJ. 2016 May; 10(5):1170-81.
- E Alsentzer, D Vera, J Neyra, L Loayza, R Hora, V Osorio, J Quispe, S Ballard, D Blazes. Monitoring Acute Diarrhea via an Electronic Surveillance System in the Peruvian Navy. Online Journal of Public Health Informatics. 2015.
- D, Joshi, E Alsentzer, K Edwards, A Norton, SE Williams. An Algorithm developed using the Brighton Collaboration case definitions is more efficient for determining diagnostic certainty. Vaccine. 2014 Jun; 32(28):3469-72.

- E Alsentzer, HD Bitner, and LK Moribe. "Recolonization of Algal Assemblages After Flooding in Nashville Creeks." Tennessee Junior Academy of Sciences Proceedings. 2011.
- E Alsentzer, HD Bitner, CA Caffey, LC Lu, LK Moribe, and SM Rucker. "The Effect of Fertilizer Pollution on the Algal Profiles of Richland and Henry Creek." Water Professionals Conference Proceedings. 2010.
- E Alsentzer, HD Bitner, and LK Moribe. "The Use of Algae as Bioindicators of Water Quality in Nashville Creeks." Water Professionals Conference Proceedings. 2010.

**SEQUENCES:**

- AM Eeds, EM Alsentzer, H Jin, T Khan, LC Lu, LK Moribe, SM Rucker, VL Shepherd, JI Creamer, CP Vanags, and KA McCue. GAPDH sequence of Minuartia cumberlandensis, GenBank, Accession HM017509, 2010.

## PRESENTATIONS

- E Alsentzer. "Clinical BERT Embeddings." MERCK- Harvard Medical School Steering Committee. (Boston, MA, December 2020)
- E Alsentzer. "Machine Learning for Healthcare." University of Massachusetts Lowell COMP5800. (Boston, MA, November 2020)
- E Alsentzer. "Clinical BERT Embeddings." Verily Life Sciences. (Boston, MA, January 2020)
- E Alsentzer. "Publicly Available Clinical BERT Embeddings." NAACL Clinical NLP Workshop. (Minneapolis, MN, June 2019)
- E Alsentzer. "Parasite Dynamics in a Connected World: The effect of network topology and human mobility on schistosomiasis transmission and control." Talk: Stanford Global Health Research Convening. (Stanford, CA, January 2016).
- E Alsentzer. "Parasite Dynamics in a Connected World: The effect of network topology and human mobility on schistosomiasis transmission and control." Poster: Symposia of Undergraduate Research and Public Service. (Stanford, CA, October 2015).
- E Alsentzer. "Monitoring Acute Diarrhea via an Electronic Surveillance System in the Peruvian Navy." Poster: International Society for Disease Surveillance Conference. (Philadelphia, PA, December 2014).
- E Alsentzer. "Monitoring Acute Diarrhea via an Electronic Surveillance System in the Peruvian Navy." Poster: Symposia of Undergraduate Research and Public Service. (Stanford, CA, October 2014).
- E Alsentzer, HD Bitner, and LK Moribe. "Recolonization of Algal Assemblages After Flooding in Nashville Creeks." Talk: Tennessee Junior Academy of Sciences Symposium. (Nashville, TN, April 2011).
- E Alsentzer. "The Use of Algae as Bioindicators of Water Quality in Nashville Creeks." Poster: Water Professionals Conference. (Nashville, TN, July 2010).