

KATIE MATTON

kmatton@mit.edu $\diamond$ <https://kmatton.github.io>

EDUCATION

Massachusetts Institute of Technology

Ph.D. in Electrical Engineering and Computer Science

2020 - present

· Advisors: John Guttag and Rosalind Picard

· Topics: machine learning, self-supervised learning, interpretability, causality, healthcare applications

University of Michigan, Ann Arbor

M.S.E. in Computer Science & Engineering

2018 - 2019

B.S.E. in Computer Science & Engineering, *Summa Cum Laude*

2014 - 2018

· Minor in Mathematics

PUBLICATIONS & PREPRINTS

* denotes equal contribution

Katie Matton, Robert Ness, John Guttag, Emre Kiciman. “Walk the Talk? Measuring the Faithfulness of Large Language Model Explanations.” *ICLR 2025*. (spotlight, top 3%)

Hyewon Jeong, Sarah Jabbour, Yuzhe Yang, . . . , **Katie Matton**, *et al.* “Recent Advances, Applications and Open Challenges in Machine Learning for Health: Reflections from Research Roundtables at ML4H 2023 Symposium.” *ArXiv Preprint arXiv:2403.01628*, 2024.

Robert Ness, **Katie Matton**, Hayden Helm, Sheng Zhang, Junaid Bajwa, Carey E. Priebe, Eric Horvitz. “MedFuzz: Exploring the Robustness of Large Language Models in Medical Question Answering.” *ArXiv Preprint arXiv:2406.06573*, 2024.

Katie Matton^{*}, Robert Lewis^{*}, John Guttag, Rosalind Picard. “Contrastive Learning of Electrodermal Activity Representations for Stress Detection.” *ACM CHIL 2023*.

Katie Matton^{*}, Robert Lewis^{*}, Rosalind Picard, John Guttag. “Improving Domain Generalization in Contrastive Learning using Adaptive Temperature Control.” *Workshop on Self-Supervised Learning, NeurIPS 2023*.

Katie Matton, John Guttag, Rosalind Picard. “Invariance-Based Causal Estimation in the Presence of Concept Drift.” *Workshop on Causal Representation Learning, UAI 2022*.

Laura Biester, **Katie Matton**, Janarthanan Rajendran, Emily Mower Provost, Rada Mihalcea. “Understanding the Impact of COVID-19 on Online Mental Health Forums.” *ACM Transactions on Management Information Systems*. 2021.

Laura Biester^{*}, **Katie Matton**^{*}, Janarthanan Rajendran, Emily Mower Provost, Rada Mihalcea. “Quantifying the Effects of COVID-19 on Mental Health Support Forums.” *Workshop on NLP for COVID-19, EMNLP 2020*. (oral)

John Gideon, **Katie Matton**, Steve Anderau, Melvin G McInnis, Emily Mower Provost. “When to Intervene: Predicting Individual Outcomes for Rapid Intervention (PRIORI) Using Passively Collected Speech from Smartphone Conversations.” *Bipolar Disorders*, 2020.

Katie Matton, Melvin G McInnis, Emily Mower Provost. “Into the Wild: Transitioning from Recognizing Mood in Clinical Interactions to Personal Conversations for Individuals with Bipolar Disorder.” *Interspeech 2019*. (oral)

EXPERIENCE

Research Intern, Microsoft Research

Summer 2023

Causal Machine Learning Team, advised by Emre Kiciman and Robert Ness

- *Measuring the Faithfulness of LLM Explanations*: Developed novel method for quantifying the degree to which an LLM's explanations are consistent with its underlying decision process.
- *Evaluating the Clinical Generalizability of LLMs*: Contributed to new method for testing the ability of medical LLMs to generalize to real clinical settings and surfacing potential vulnerabilities.

PhD Candidate, Massachusetts Institute of Technology

2020 - Present

Clinical and Applied Machine Learning Group (CSAIL), advised by John Guttag

Affective Computing Group (Media Lab), advised by Rosalind Picard

- *Improving Domain Generalization in Contrastive Learning*: Developed novel contrastive learning objective that re-weights samples using domain information to learn more generalizable representations.
- *Contrastive Learning for EDA Data*: Designed novel, EDA-specific data augmentations to incorporate domain knowledge and learn improved representations within a contrastive learning framework.
- *Invariant Representation Learning (IRL) with Hidden Variables*: Developed new IRL method that relaxes assumptions of prior work, enabling out-of-distribution generalization in more realistic settings.

Research Assistant, University of Michigan

2017 - 2020

Computational Human AI (CHAI) Lab, advised by Emily Mower Provost

- *Detecting Bipolar Mood Severity from Speech Data*: Investigated machine learning methods for detecting symptom severity in individuals with Bipolar Disorder from “in-the-wild” speech and text.
- *Quantifying Language Patterns for Bipolar Mood Detection*: Examined methods of computationally modeling language patterns to detect bipolar mood states.
- *Measuring the Effects of COVID-19 on Mental Health Forums*: Used time series intervention analysis, topic modeling, and network analysis to measure pandemic-initiated changes in behavior on forums.

Research Assistant, University of New South Wales

01/2020 - 02/2020

Speech Processing Research Laboratory, advised by Vidhyasaharan Sethu

- *Measuring the Uncertainty of Depression Prediction Models*: Explored methods for estimating different types of uncertainty (e.g. aleatoric, epistemic) in ML models for depression severity prediction.

Software Engineering Intern, Microsoft

Summer 2017, Summer 2016

- Designed and implemented new algorithms for quantifying the quality of the Cortana Virtual Assistant's responses to user queries. Developed scraping engine to request and parse query results from Cortana.
- Collaborated to create Windows app to measure and display Secure Digital card quality.

AWARDS AND HONORS

ML4H Top Reviewer Award

2022

MIT Abdul Latif Jameel Fellowship for Machine Learning and Health Solutions

2020

NSF Graduate Research Fellowship Honorable Mention

2020

James B. Angell Scholar Award, University of Michigan

2018, 2019

Engineering Innovation Award, Michigan Engineering Graduate Research Symposium

2018

Best Research Poster, Michigan AI Symposium

2018

University of Michigan Engineering Honors Program

2016 - 2018

D.E. Shaw Discovery Fellowship

2016

TALKS AND POSTER SESSIONS

Contrastive Learning of Electrodermal Activity Representations for Stress Detection. Conference on Health, Inference, and Learning 2023. Cambridge, MA, USA.

Improving Domain Generalization in Contrastive Learning using Adaptive Temperature Control. Self Supervised Learning Workshop, NeurIPS 2023. New Orleans, USA.

Contrastive Learning of Electrodermal Activity Representations for Stress Detection. Department of Biomedical Informatics, Emory University & Georgia Institute of Technology, 2023.

Improving Domain Generalization in Contrastive Learning using Adaptive Temperature Control. Distribution Shift Workshop, NeurIPS 2023. New Orleans, LA, USA.

Machine Learning for Mental Health Monitoring. AstraZeneca Artificial Intelligence Journal Club, Virtual, 2023.

Contrastive Learning of Electrodermal Activity Representations for Stress Detection. Workshop on Learning from Time Series for Health, NeurIPS 2022. New Orleans, LA, USA.

Invariance-Based Causal Estimation in the Presence of Concept Drift. Causal Representation Learning Workshop, UAI 2022. Eindhoven, Netherlands.

Contrastive Learning for Wearable Sensor Data. MIAI Meeting, 2022. Grenoble Alpes University, Grenoble, France.

Contrastive Learning for Wearable Sensor Data. Digital Mental Health Workshop, 2022. MIT Media Lab, Cambridge, MA.

Machine Learning for Mental Health. Media Lab Focus Series, 2021. MIT, Cambridge, MA.

Quantifying the Effects of COVID-19 on Mental Health Support Forums. Workshop on NLP for COVID-19, EMNLP, 2020. Virtual.

Into the Wild: Transitioning from Recognizing Mood in Clinical Interactions to Personal Conversations for Individuals with Bipolar Disorder. Interspeech 2019. Graz, Austria.

Into the Wild: Transitioning from Recognizing Mood in Clinical Interactions to Personal Conversations for Individuals with Bipolar Disorder. University of Michigan AI Symposium 2019. Ann Arbor, MI.

Mood State Prediction for Individuals with Bipolar Disorder. Workshop for Young Female Researchers in Speech, Interspeech 2018. Hyderabad, India.

Mood State Prediction for Individuals with Bipolar Disorder. University of Michigan Engineering Graduate Research Symposium 2018. Ann Arbor, MI.

Mood State Prediction for Individuals with Bipolar Disorder. University of Michigan AI Symposium 2018. Ann Arbor, MI.

TEACHING & MENTORSHIP

Massachusetts Institute of Technology

- Master's Thesis Supervisor: *Puja Balaji, MEng* 2024-2025
- Instructor: *6.S091 Causality* January 2024
- Teaching Assistant: *6.S091 Causality* January 2023
- Teaching Assistant: *6.C51 Modeling with Machine Learning* Spring 2023

University of Michigan, Ann Arbor

- Teaching Assistant: *EECS 281 Data Structures and Algorithms* 2017 - 2018
- Proof Tutor: *Math 217 Linear Algebra* 2016 - 2017

REVIEWING

ICLR	2024
Machine Learning for Health (ML4H)	2022, 2024
Self Supervised Learning Workshop, NeurIPS	2023

SERVICE & OUTREACH

Graduate Student Panel for Undergraduate Researchers , Cambridge, MA	2025
Provided advice on research and navigating grad school to MIT undergraduate students.	
Graduate Application Assistance Program , Cambridge, MA	2020 - 2024
Mentored individuals from underrepresented groups in preparing graduate school applications.	
Research Roundtables at Machine Learning for Health , New Orleans, LA	2023
Junior chair of causality roundtable. Facilitated discussion and prepared report with key takeaways.	
Causal Representation Learning Reading Group , Cambridge, MA	2022 - 2023
Led bi-weekly reading group to discuss papers related to causal representation learning.	
Talk at Cohere for AI , Virtual	2023
Gave a talk on my research journey to a <i>Beginners in Machine Learning Research</i> interest group.	
Graduate Women in Course 6 , Cambridge, MA	2021 - 2022
President of student group dedicated to supporting women and gender minorities in EECS at MIT.	
AI Mentorship Program , Cambridge, MA	2020 - 2021
Mentored undergraduate MIT students from underrepresented groups interested in AI research.	
Girls Encoded Middle School Outreach , Ann Arbor, MI	2018 - 2019
Taught weekly CS lessons at local middle school and assisted with curriculum development.	
AI Family Festival , Detroit, MI	2018
Discussed experience as AI student and taught project-based AI lesson.	
Better Living Using Engineering (BLUE) Lab , Ann Arbor, MI	2014 - 2017
Led 30+ member student team in implementing engineering projects to improve health and education for international and local partner communities.	
Elementary Engineering Education Day , Ann Arbor, MI	2015
Served on committee to organize day of fun engineering activities for local 1st-5th grade students.	

TECHNICAL SKILLS

Programming Languages	Proficient: Python, C++, Matlab Some Experience: Julia, C#, SQL, Bash, JavaScript, HTML
Libraries & Tools	PyTorch, TensorFlow, Scikit-learn, Kaldi (familiar), Git