Tom Hartvigsen www.tomhartvigsen.com
University of Virginia hartvigsen@virginia.edu

Interests

Machine Learning, Time Series, Language Models, Healthcare, Robustness, Safety, Interpretability

Employment

University of Virginia Aug 2023 –
Assistant Professor of Data Science

MIT, CSAIL Jan 2022 – Aug 2023
Postdoctoral Associate

Microsoft May 2021 – Aug 2021
Data Science Intern

UMass Medical School Aug 2018 – Sep 2019
Research Intern

University of Arizona May 2015 – Aug 2015
NSF REU Intern

Education

Worcester Polytechnic Institute Dec 2021
Ph.D. in Data Science
M.S. in Data Science

SUNY Geneseo May 2016
B.A. in Applied Mathematics

Publications

2023 Elizabeth Bondi-Kelly, **Thomas Hartvigsen**, Lindsay Sanneman, Swami Sankaranarayanan, Lauren Oakden-Rayder, Leo Celi, Julie Shah, Marzyeh Ghassemi. *Taking Off with AI: Lessons from Aviation for Healthcare*. *ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAMMO), 2023.*

2023 Jidapa Thadajarassiri, **Thomas Hartvigsen**, Walter Gerych, Xiangnan Kong, Elke Rundensteiner. *Knowledge Amalgamation for Multi-Label Classification via Label Dependency Transfer*. AAAI Conference on Artificial Intelligence (AAAI), 2023.

2023 Haoran Zhang, **Thomas Hartvigsen**, Marzyeh Ghassemi. *Algorithmic Fairness in Chest X-ray Diagnosis.* Case Studies in Social and Ethical Responsibilities of Computing, MIT, 2023.

2023 **Thomas Hartvigsen**, Swami Sankaranarayanan, Hamid Palangi, Yoon Kim, Marzyeh Ghassemi. *Aging with GRACE: Lifelong Model Editing with Discrete Key-Value Adapters*. *Under review.*

2023 **Thomas Hartvigsen**, Jidapa Thadajarassiri, Xiangnan Kong, Elke Rundensteiner. *Finding Short Signals in Long Irregular Time Series with Continuous-Time Attention Policy Networks*. *Under review.*

2023 Owen Queen, **Thomas Hartvigsen**, Huan He, Teddy Koker, Theodoros Tsiligkaridis, Marinka Zitnik. *Encoding Time-Series Explanations through Self-Supervised Model Behavior Consistency*. *Under review.*

2023 Anurag Vaidya, Yuzhe Yang, **Thomas Hartvigsen**, Faisal Mahmood. *Performance of Computational Pathology Oncology Models Varies by Demographics*. *Under review.*

2023 Haoran Zhang, Taylor Killian, Marzyeh Ghassemi, **Thomas Hartvigsen**. *Active Feature Acquisition for Early Classification of Clinical Time Series*. *Under review.*

2023 Tianhua Zhang, Hongyin Luo, Yung-Sung Chuang, Wei Fang, Luc Gaitskell, **Thomas Hartvigsen**, Xixin Wu, Danny Fox, Helen Meng, James Glass. *Interpretable Unified Language Assessment with Large Language Models*. *Under review.*

2023 Taylor Killian, Haoran Zhang, **Thomas Hartvigsen**, Ava Amini. *Continuous Time Evidential Distributions for Irregular Time Series*. *Under review. Working paper appeared at the ICML 2023 Workshop on Interpretable Machine Learning in Healthcare.*

2023 Ramesh Doddaiah, Prathyush Parvatharaju, Elke Rundensteiner, **Thomas Hartvigsen**. *Explaining Deep Multi-Class Time Series Classifiers*. *Under review.*

2023 Kimia Hamidieh, Haoran Zhang, **Thomas Hartvigsen**, Marzyeh Ghassemi. *Identifying Implicit Biases in Vision-Language Models*. *Under review. Working paper appeared at the ICML 2023 Workshop on Data-centric Machine Learning Research.*

2023 Bret Nestor, **Thomas Hartvigsen**, Jennifer Yu, Kathleen Sheehan, Fahad Razak, Amol Verma, Anna Goldenberg, Marzyeh Ghassemi. *Learning Limited: How limited data sharing impacts model performance in health*. *Under review.*

2023 Sujay Nagaraj, Sarah Goodday, **Thomas Hartvigsen**, Adrien Boch, Luca Foschini, Marzyeh Ghassemi, Stephen Friend, and Anna Goldenberg. *Dissecting "In-the-Wild" Stress from Multimodal Sensor Data*. *Under review. Working paper appeared at the 2022 Machine Learning for Health Symposium.*

2023 Hang Yin, Xiangnan Kong, Liping Liu, **Thomas Hartvigsen**, Xin Dai, Yao Su. *SkipSNN: Efficiently Classifying Noisy Spike Trains with Event Attention*. *Under review*.

2023 Haoran Zhang,* Qixuan Jin,* **Thomas Hartvigsen**,* Miriam Udler, Marzyeh Ghassemi. *A Pipeline for Interpretable Clinical Subtyping with Deep Metric Learning*. *ICML 2023 Workshop on Interpretable Machine Learning in Healthcare.* *Best Paper Award*.

2022 **Thomas Hartvigsen**, Saadia Gabriel, Hamid Palangi, Maarten Sap, Dipankar Ray, Ece Kamar. ToxiGen*: A Large-Scale Machine-Generated Dataset for Adversarial and Implicit Hate Speech Detection.* Annual Meeting of the Association for Computational Linguistics (ACL), 2022.

2022 **Thomas Hartvigsen**, Walter Gerych, Jidapa Thadajarassiri, Xiangnan Kong, Elke Rundensteiner. Stop&Hop*: Early Classification of Irregular Time Series*. ACM International Conference on Information and Knowledge Management (CIKM), 2022.

2022 Swami Sankaranarayanan, **Thomas Hartvigsen**, Lauren Oakden-Rayner, Marzyeh Ghassemi, Philip Isola. *Real-world Relevance of Generative Counterfactual Explanations*. *Working paper appeared at NeurIPS 2022 Workshop on Trustworthy and Socially Responsible Machine Learning*.

2022 Qixuan Jin, Haoran Zhang, **Thomas Hartvigsen**, Marzyeh Ghassemi. *Learning Fair Multimodal Checklists for Interpretable Disease Classification*. *Working paper appeared at NeurIPS 2022 Workshop on Learning from Time Series for Health.*

2022 Ramesh Doddaiah, Prathyush Parvatharaju, Elke Rundensteiner, **Thomas Hartvigsen**. *Class-Specific Explainability for Deep Time Series Classifiers*. International Conference on Data Mining (ICDM), 2022.

2022 Walter Gerych, **Thomas Hartvigsen**, Emmanuel Agu, Elke Rundensteiner. *Robust Recurrent Classifier Chains For Multi-Label Learning With Missing Labels*. ACM International Conference on Information and Knowledge Management (CIKM), 2022.

2022 Walter Gerych, **Thomas Hartvigsen**, Luke Buquicchio, Emmanuel Agu, Elke Rundensteiner. *Recovering the Propensity Score from Biased Positive Unlabeled Data*. AAAI Conference on Artificial Intelligence (AAAI), 2022 (Oral Spotlight).

2022 **Thomas Hartvigsen***, Walter Gerych*, Marzyeh Ghassemi. *On Detecting COVID-Risky Behavior from Smartphones*. epiDAMIK Workshop @ KDD 2022 (Spotlight).

2022 Aparna Balagopalan, Haoran Zhang, Kimia Hamidieh, **Thomas Hartvigsen**, Frank Rudzicz, Marzyeh Ghassemi. *The Road to Explainability is Paved with Bias: Measuring the Fairness of Explanations*. ACM Conference on Fairness, Accountability, and Transparency (FAccT), 2022.

2022 Walter Gerych, **Thomas Hartvigsen**, Luke Buquicchio, Kavin Chandrasekaran, Abdulaziz Alajaji, Hamid Mansoor, Elke Rundensteiner, Emmanuel Agu. *Positive Unlabeled Learning with a Sequential Selection Bias*. SIAM International Conference on Data Mining (SDM), 2022.

2022 Ruofan Hu, Dongyu Zhang, Dandan Tao, **Thomas Hartvigsen**, Hao Feng, Elke Rundensteiner. *TWEET-FID: An Annotated Dataset for Multiple Foodborne Illness Detection Tasks*. Language Resources and Evaluation Conference (LREC), 2022.

2021 Walter Gerych, **Thomas Hartvigsen**, Luke Buquicchio, Emmanuel Agu, Elke Rundensteiner. *Recurrent Bayesian Classifier Chains for Exact Multi-label Classification*. Advances in Neural Information Processing Systems (NeurIPS), 2021.

2021 Hang Yin, John Boaz Lee, Xiangnan Kong, **Thomas Hartvigsen**, Sihong Xie. *Energy-Efficient Models for High-Dimensional Spike Train Classification using Sparse Spiking Neural Networks*. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2021.

2021 Jidapa Thadajarassiri, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner. *Semi Supervised Knowledge Amalgamation for Sequence Classification*. AAAI Conference on Artificial Intelligence (AAAI), 2021.

2021 Prathyush Parvatharaju, Ramesh Doddaiah, **Thomas Hartvigsen***, Elke Rundensteiner*. *Learning Saliency Maps to Explain Deep Time Series Classifiers*. ACM International Conference on Information and Knowledge Management (CIKM), 2021.

2021 Luke Buquicchio, Walter Gerych, Kavin Chandrasekaran, Abdulaziz Alajaji, Hamid Mansoor, **Thomas Hartvigsen**, Elke Rundensteiner, Emmanuel Agu. *Variational Open-Set Recognition*. IEEE International Conference on Big Data, 2021.

2021 Dongyu Zhang, Cansu Sen, Jidapa Thadajarassiri, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner. *Human-like Explanation for Text Classification with Limited Attention Supervision*. IEEE International Conference on Big Data, 2021.

2020 **Thomas Hartvigsen**, Cansu Sen, Xiangnan Kong, Elke Rundensteiner. *Recurrent Halting Chain for Early Multi-label Classification*. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2020.

2020 Cansu Sen, **Thomas Hartvigsen**, Biao Yin, Xiangnan Kong, Elke Rundensteiner. *Human Attention Maps for Text Classification: Do Humans and Neural Networks Focus on the Same Words?* Annual Meeting of the Association for Computational Linguistics (ACL), 2020.

2020 **Thomas Hartvigsen**, Cansu Sen, Xiangnan Kong, Elke Rundensteiner. *Learning to Selectively Update State Neurons in Recurrent Networks*. ACM International Conference on Information and Knowledge Management (CIKM), 2020.

2020 Jidapa Thadajarassiri, Cansu Sen, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner. *Learning Similarity Preserving Word Meta-Embedding*. IEEE International Conference on Big Data, 2020.

2020 Erin Teeple, **Thomas Hartvigsen**, Cansu Sen, Kajal Claypool, Elke Rundensteiner. *Clinical Performance Evaluation of a Machine Learning System for Predicting Hospital-Acquired Clostridium Difficile Infection*. International Conference on Health Informatics, 2020. 🏆 Best Poster.

2019 **Thomas Hartvigsen**, Cansu Sen, Xiangnan Kong, Elke Rundensteiner. *Adaptive-Halting Policy Network for Early Classification*. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2019.

2019 Cansu Sen, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner. *Patient-Level Classification of Clinical Note Sequences Guided by Attributed Hierarchical Attention*. IEEE International Conference on Big Data, 2019.

2019 Cansu Sen, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner. *Learning Temporal Relevance in Longitudinal Medical Notes*. IEEE International Conference on Big Data, 2019.

2019 Daniel Johnston†, Liubuo Klindziuk†, Lolita Nazarov†, **Thomas Hartvigsen***, Elke Rundensteiner*. *Early Diagnosis Prediction with Recurrent Neural Networks*. IEEE URTC 2019. 🏆 Best Paper runner up.

2019 Jidapa Thadajarassiri, Cansu Sen, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner. *Comparing General and Locally-Learned Word Embeddings for Clinical Text Mining*. IEEE International Conference on Biomedical and Health Informatics, 2019.

2018 **Thomas Hartvigsen**, Cansu Sen, Elke Rundensteiner. *Detecting MRSA Infections by Fusing Structured and Unstructured Electronic Health Record Data*. International Joint Conference on Biomedical Engineering Systems and Technologies, 2018.

2018 **Thomas Hartvigsen**, Cansu Sen, Sarah Brownell†, Erin Teeple, Xiangnan Kong, Elke Rundensteiner. *Early Prediction of MRSA Infections using Electronic Health Records*. International Conference on Health Informatics, 2018. 🏆 Best Student Paper runner up.

2018 Julia Friend†, Alec Hauck†, Sruthi Kurada†, Cansu Sen, **Thomas Hartvigsen***, Elke Rundensteiner*. *Handling Missing Values in Multivariate Time Series Classification*. IEEE URTC 2018.

2017 Sarah Brownell†, **Thomas Hartvigsen**, Elke Rundensteiner. *MRSA Infection Prediction System*. IEEE URTC 2017.

2017 Cansu Sen, **Thomas Hartvigsen**, Kajal Claypool, Elke Rundensteiner. *CREST - Risk Prediction for Clostridium Difficile Infection Using Multimodal Data Mining*. European Conference on Machine Learning (ECML), 2017.

*equal contribution †undergraduate mentee

Invited Talks

Stanford, *MedAI seminar series*	May 2023
IIT Delhi, *LC2S seminar series*	Apr 2023
Harvard Medical School	Mar 2023
University of Virginia	Feb 2023
Rochester Institute of Technology	Feb 2023
AJCAI Workshop on Toxic Langauge Detection	Dec 2022
Worcester Polytechnic Institute	Nov 2022
Northeastern University	Oct 2022
Tufts University	Jul 2022
University of Rochester	Jul 2022
Google Research, *Responsible AI seminar series*	Jun 2022
Massachusetts Institute of Technology, *MIT Horizons Spotlight series*	Apr 2022
Massachusetts Institute of Technology	Nov 2021
Harvard University	Sep 2020
Computational Sustainability Doctoral Consortium	Sep 2020
Florida State University	Jun 2020
MITRE	Mar 2020
IBM Research	Jan 2020
University of Minnesota, *IMA Data Science Workshop*	Sep 2019
Worcester Polytechnic Institute, *Arts & Sciences Showcase*	May 2019

Service

Organizing:

- **General Chair**, Machine Learning for Health (ML4H) Symposium 2023
- Conference on Health, Informatics, and Learning (CHIL) 2023
- ICML Workshop on Deployment Challenges in Generative AI 2023
- NeurIPS Workshop on Learning from Time Series for Health (TS4H) 2022

Reviewing:

- **Program Committees.** AAAI, ARR, WSDM, ACL, EMNLP, NeurIPS, NeurIPS Datasets & Benchmarks, CVPR, ICCV, KDD, CHIL, TSRL4H @ ICLR, epiDAMIK @ KDD, RobustSeq @ NeurIPS.
- NeurIPS 2023 Workshop Proposals

Teaching and Mentoring

MIT 6.S898 (Graduate-level Deep Learning), guest lecture on *Responsible Deep Learning* 2022

***MIT EECS GAAP* mentor**, Helping underrepresented students apply to graduate school 2022

"*Is graduate school right for me?*" panelist, SUNY Geneseo 2021

Led tutorial on Deep Learning with PyTorch for Undergrads, WPI 2019

TA for six courses, SUNY Geneseo 2014 – 2016

Selected Honors & Awards

🏆 **Best Paper**, IMLH Workshop, ICML 2023	2023
🏆 **AAAI Oral Spotlight**	2022
🏆 **Best Poster**, International Conference on Health Informatics	2020
🏆 **Outstanding Graduate Research Award**, WPI	2019
🏆 **Best Poster**, Graduate Research Innovation and Exchange, WPI	2019
IMA Travel Award, University of Minnesota	2019
🏆 **Best Student Paper Runner Up**, HEALTHINF	2018
🏆 **People's Choice Poster Award**, Graduate Research Showcase, WPI	2017
GAANN PhD Fellowship (5 years Tuition/Stipend), U.S. Dept. of Ed.	2016 – 2021

Media Coverage

`MSR`: (De)ToxiGen: Leveraging large language models to build more robust hate speech detection tools.

`TechCrunch`: Microsoft claims its new tools make language models safer to use.

`MIT News`: In bias we trust?

Patents

Dipankar Ray, Thomas Hartvigsen, and Misha Bilenko. *A novel data mining technique for utilizing large language models to generate/augment datasets for use in training Responsible AI systems*.

Hamid Palangi, Saadia Gabriel, Thomas Hartvigsen, Dipankar Ray, Maarten Sap, and Ece Kamar. *Adversarial language imitation with controlled exemplars*.

References

Available upon request