Megha Thukral

mthukral3@gatech.edu | meghathukral.github.io | linkedin.com/in/megha-thukral | Google Scholar | +1 470-815-1982

Research Summary

My research lies at the intersection of machine learning, ubiquitous sensing, and digital health, with a focus on robustness under distribution shift. I develop generalizable models for movement and physiological sensing data using large-scale unlabeled data, self-supervised learning, and cross-modal transfer. I also explore how foundation models from data-rich domains can be adapted to improve the modeling of sensing and waveform data in low-resource, distribution-shifted settings.

Education

Georgia Institute of Technology — Atlanta, USA
Ph.D. in Machine Learning, School of Interactive Computing — Aug 2023 – Present
GPA: 3.87/4.00
Selected coursework: Mathematical Foundations of ML; Probabilistic Graphical Models; Optimization, Stat. ML

Georgia Institute of Technology — Atlanta, USA
M.S. in Computer Science — Aug 2021 – May 2023
GPA: 4.00/4.00
Selected coursework: Deep Learning; Artificial Intelligence; Machine Learning; Computer Vision; Algorithms

Work Experience

Samsung Research America — Digital Health Lab — Mountain View, USA
AI Research Intern — May 2025 – Dec 2025

- Proposed a wavelet-driven novel time-frequency pretraining task and trained a large-scale photoplethysmography (PPG) foundation model on 18 million unlabeled PPG segments from 32K smartwatch users to learn robust representations.
- Implemented self-supervised objectives (masked reconstruction, contrastive) and data pipelines for wearable PPG in PyTorch.
- Designed evaluation and ablation studies to measure representation quality across 19 different downstream tasks, including cardiovascular health tasks (e.g., hypertension, arrhythmia).

Georgia Institute of Technology — Atlanta, USA
Graduate Research Assistant — Aug 2023 – Present

- Research on multimodal adaptation of clinical EHR foundation models with continuous sensing data.
- Developed a layout-agnostic, cross-modality transfer framework for smart-home human activity recognition (HAR), leveraging pre-trained text encoders to improve generalization across homes.
- Leveraged LLM agents to simulate synthetic sensor data in Virtual Home environments and improve HAR performance.

Bloomberg LP — AI Group — New York, USA
Software Engineer (ML) Intern — May 2022 – Jul 2022

- Built and deployed an FAQ retrieval system using sentence encoders, and automating ML workflows with Argo while containerizing models on Kubernetes/KServe for resilient serving.

Publications

- **Cross-Domain HAR: Few-Shot Transfer Learning for Human Activity Recognition**.
  Megha Thukral, Harish Haresamudram, Thomas Plötz.
  ACM TIST 2025 (published).
  Proposes a novel few-shot transfer learning method for cross-domain HAR (across body positions) and benchmarks various adaptation methods under label-scarce settings.
  Link: dl.acm.org/doi/10.1145/3704921

- **Layout Agnostic Human Activity Recognition via Textual Descriptions of Sensor Triggers (TDOST)**.
  Megha Thukral*, Sourish Dhekane*, Shruthi K. Hiremath, Harish Haresamudram, Thomas Plötz.
  Proc. ACM IMWUT / UbiComp 2025 (published).
  Introduces layout-agnostic HAR by leveraging textual descriptions of sensor triggers to generalize across diverse home layouts.
  Link: dl.acm.org/doi/pdf/10.1145/3712278

- **Wavelet-Driven Masked Multiscale Reconstruction for Photoplethysmography (PPG) Foundation Models**.
  Megha Thukral, Cyrus Tanade, Simon A. Lee, Juhyeon Lee, Hao Zhou, Keum San Chun, Migyeong Gwak, Viswam Nathan,Md Mahbubur Rahman, Li Zhu, Mehrab Bin Morshed, Subramaniam Venkatraman,Sharanya Arcot Desai.
  NeurIPS TS4H 2025 (accepted), under review at ICML'26
  Proposes a novel wavelet-based pretraining for smartwatch PPG, improving representations for downstream tasks such as hypertension and PVC detection.
  Link: https://openreview.net/pdf?id=kJbY2CsTPb

- **AgentSense: Virtual Sensor Data Generation Using LLM Agents in Simulated Home Environments**.
  Zikang Leng*, **Megha Thukral***, YaQi Liu*, Hrudhai Rajasekhar, Shruthi K. Hiremath, Thomas Plötz.
  AAAI 2026 (published).
  Uses LLM-based behavior/ routine generation in simulated homes to generate synthetic ambient sensor data, boosting performance in low-data smart home settings.
  Link: arxiv.org/abs/2506.11773

- **Himae: Hierarchical masked autoencoders discover resolution-specific structure in wearable time series**.
  Simon Lee, Cyrus Tanade, Hao Zhao, Juhyeon Lee, **Megha Thukral**, Minji Han, and others.
  ICLR 2026 (published).
  Develops a lightweight SSL objective that competes with much larger transformer Foundation models that also serve as an interpretability tool
  Link: https://openreview.net/pdf?id=iPAy5VpGQa

- **Towards On-device Foundation Models for Raw Wearable Signals**.
  Simon A. Lee, Cyrus Tanade, Hao Zhou, Juhyeon Lee, **Megha Thukral**, Baiying Lu, Sharanya Arcot Desai.
  NeurIPS 2025 TS4H Workshop (accepted);
  Develops edge-efficient foundation models for raw wearable signals, enabling scalable pretraining.
  Link: openreview.net/forum?id=i6DUa3k63o

- **Multimodal Self-Supervised Learning for Wearable Sleep Staging Using Photoplethysmography and Accelerometer Signals**.

Juhyeon Lee, Simon A. Lee, Cyrus Tanade, Viswam Nathan, **Megha Thukral**, Hao Zhou, Keum San Chun, Sharanya Arcot Desai.
ICASSP 2026 (accepted).
Pretrain joint wearable PPG+ACC representations on 23K hours of data and use a causal Mamba sequence model for scalable, real-time sleep staging

- **A Personalized Real-Time Proactive Voice Memory Assistant**.
Hao Zhou, Md Mahbubur Rahman, Simon A. Lee, Baiying Lu, Juhyeon Lee, **Megha Thukral**, Cyrus Tanade, Md. Sazzad Hissain Khan, Samsad Ul Islam, Subramaniam Venkatraman, Sharanya Arcot Desai.
ICASSP 2026 (accepted).
Introduces, MemoryAids, a privacy-aware, proactive real-time voice memory assistant to spot missing details and summarize conversations

- **How Much Unlabeled Data is Really Needed for Effective Self-Supervised HAR?**
Sourish Dhekane, Harish Haresamudram, **Megha Thukral**, Thomas Plötz.
ACM ISWC 2023 (published).
Empirically studies the impact of unlabeled data scale on self-supervised HAR
Link: dl.acm.org/doi/10.1145/3594738.3611366

Services

Organising

- GenAI4HS @ UbiComp 2025
Co-organized Generative AI and Foundation Models for Human Sensing Workshop – a workshop bringing researchers and practitioners to discuss the recent trends and challenges in building foundational models for human sensing, and integrating advances from Generative AI into sensor pipelines.

Reviewing

- PACM IMWUT 2025
- PACM ISWC 2025
- NeurIPS 2025 Time Series for Health
- GenerativeAi4HS @ Ubicomp 2025

Mentoring

- Tyler Stennett – now CS PhD student at Georgia Tech
- Abdullah Altaweel – now Software Engineer at Civic-DX
- Brandon Lo – now Software Engineer at Amazon
- Juan Macías Romero – undergrad student at Universidad Carlos III de Madrid
- Aarushi Biswas – MS CS at Georgia Tech
- Kevin A Daniel – MS CS at Georgia Tech
- Khush Attarde – Research Intern at Ogmen Robotics