

Ken (Tongli) Su

+1 404-384-4141 · ken.su@emory.edu · linkedin.com/in/ken-su · Atlanta, GA

Research Summary

Computer Science PhD student at Emory University (advisor: Liang Zhao) researching **LLM agents, alignment, and AI interpretability and safety**. My goal is to help people accomplish complex tasks with AI while retaining meaningful control. I am particularly interested in **multi-agent collaboration and theory of mind**, personalized assistance, and when agents should involve humans. My work includes repository retrieval for coding agents (**LARGER**, co-first author), software world models, and cross-modal physiological signal generation (**EMBC 2026 oral**, first author). Before the PhD I spent 2+ years as an ML engineer shipping production LLM, retrieval, and forecasting systems across manufacturing, finance, consumer goods, and insurance. Reviewer, NeurIPS 2026.

Education

Emory University

Ph.D. in Computer Science — Advisor: Liang Zhao

Atlanta, GA

Aug 2025 – Dec 2028 (expected)

Columbia University

M.S. in Data Science, GPA: 3.95/4.00

New York, NY

Dec 2023

Emory University

B.A. in Computer Science & B.A. in Economics (Dual Major), GPA: 3.84/4.00

Atlanta, GA

May 2022

Research Experience & Publications

Research Intern | Causal Dynamics Lab

Jul 2026 – Aug 2026

Research on coding agents and software world models, including predicting the consequences of changes across software systems.

Graduate Research Assistant — LLM Agents & Retrieval | Emory University, Advisor: Liang Zhao

Aug 2025 – Present

LARGER: Lexically Anchored Repository Graph Exploration and Retrieval — **co-first author**

arXiv:2605.16352, 2026 (under review)

- Studied repository localization errors in coding agents: lexical search can miss imports, call chains, type hierarchies, and code–test links, motivating structural retrieval integrated into the agent’s existing search workflow.
- Formalized **Lexically Anchored Structural Localization**: a fixed agent policy turns its own lexical observations into structural graph anchors and progressively exposes compact, confidence-filtered local neighborhoods — delivering graph evidence *within* lexical observations rather than through new agent actions, so the method drops into existing CLI agents unchanged.
- Decomposed the joint objective into **graph quality** and **retrieval efficiency**, proposing a divide-and-conquer approximation that reduces global subgraph search to lexical anchoring plus confidence-scored local expansion, and engineered the full system: multi-language **AST-based graph construction**, edge weighting, and community detection for graph quality; lightweight **sidecar storage**, **dynamic lexical propagation**, and **budget-aware filtering** for retrieval efficiency.
- Evaluated across **4 benchmarks** (LocBench, MuLocBench, SWE-Atlas Test Writing, SWE-Atlas Codebase QA) against **7 baselines** (BM25, Agentless, LocAgent, SWE-agent, OpenHands, Claude Code, Codex): **+13.9 Acc@5** tuned, **+11.8 Acc@5** with fixed hyperparameters (robust without per-benchmark tuning), and **+12.2 Recall@5**, surpassing strong proprietary baselines on codebase QA and test writing.

Graduate Researcher — Cross-Modal Generative Modeling | Emory University, Katebi & Clifford Lab

2025 – Present

Cross-Modal Signal Translation from Fetal–Maternal ECG to Fetal Doppler Waveforms — **first author**

IEEE EMBC 2026, **Oral**

- Developed cross-modal generation from fetal and maternal ECG to Doppler waveforms, studying which aspects of fetal cardiovascular function are recoverable from electrical signals and where reconstruction remains limited.
- Designed an architecture combining **dilated convolutions** (large receptive field for long-range temporal dependencies), **cross-modal attention** (fetal features as queries, maternal ECG as keys/values, selectively attending to maternal segments temporally correlated with fetal blood flow), and **self-attention** over the fetal signal.
- Trained on **885 synchronized fetal/maternal ECG–Doppler envelope segments from 39 pregnancies**: **PSD MSE 49.9 ± 15.8 dB²** (**51% below** the two-channel baseline) and heart-rate error **4.71 ± 0.77 bpm** (**1.5% better**; negligible against the 110–160 bpm physiological range).
- Compared ways to incorporate maternal ECG: **cross-modal attention reduces PSD MSE by 39%** relative to naive dual-channel concatenation, supporting selective use of maternal information for waveform reconstruction.
- Introduced a **composite loss** balancing pointwise error, derivative error, and correlation to enforce amplitude accuracy and morphological fidelity simultaneously.

A Review of Multimodal Pretraining Strategies for Physiological Wearable Sensor Foundation Models — **first author**

Under review, 2026

- Screened arXiv, Google Scholar, Semantic Scholar, PubMed, Web of Science, and IEEE Xplore (Jan 2024 – early 2026), synthesizing **22 papers** into the first review organized around **multimodal pretraining strategy** as the primary lens rather than as one topic among many.
- Proposed a taxonomy of **five pretraining paradigms** — contrastive, generative/reconstruction, language-grounded alignment, missing-modality learning, and cross-modal knowledge distillation — comparing objective design and architectural patterns across PPG, ECG, EEG, and ACC.
- Analyzed the recent model cohort (SleepFM, LSM/LSM-2, MoCA, NormWear, SensorLM, PhysioOmni, CSFM, AnyPPG) to extract shared successful principles (shared cross-modal representation spaces, physiology-aware positive pairing, masked cross-modal reconstruction) and open challenges in personalization, robustness to missing sensors, and demographic bias.

Industry Experience

Machine Learning Engineer, GenAI | DuPont Specialty Products USA, LLC New York, NY | Jul 2024 – Aug 2025

- Collaborated cross-functionally to build a scalable lead identification and prioritization pipeline on **Azure Databricks** using **Spark**, **Delta Lake**, and generative models, powering an internal co-pilot for sales managers that became the company’s primary method for identifying potential customers and contributed **over \$60 million** in estimated business value.
- Fine-tuned a **Sentence-BERT** model with **triplet loss** on a retrieval-and-ranking task matching leads to relevant internal applications: **Recall@5 of 0.86** and **MRR of 0.64**, a **27% improvement** over the existing GPT-based system, with hallucinations eliminated. Enabled marketing efforts that converted leads into **over \$3 million** in revenue.
- Engineered a data enrichment module on the **OpenAI API** with round-robin load balancing, asynchronous processing, and batch handling: pipeline runtime **10 h → 6 h (40% reduction)**, token costs down **\$12,000/year**, and plug-and-play deployment saving **60 hours** of development time on related projects.
- Translated business feedback into model improvements and new features over a six-week sprint, **increasing DAU 48%** while raising average user satisfaction from 4.0 to 4.6/5. Presented results to company leadership, securing additional resources for continued development.

Machine Learning Engineer | ICBC Financial Service, LLC New York, NY | Feb 2024 – May 2024

- Collaborated with research scientists to develop an in-house **foundation model** based on **transformers**, establishing a robust training loop with the Bing Search API for periodic continuous re-pretraining. **Few-shot learning** for financial report translation reached **BLEU 0.59** (cumulative to 4-grams), reducing external translation costs **\$30,000/year** and turnaround time **50%**.
- Created a **Flask** application with Python-embedded **T-SQL** and **parameterized queries**, reducing data retrieval time **40%** (saving the team 20 hours/week on compliance requests) and eliminating SQL injection risk; **85% user adoption** among 50 team members in month one.
- Optimized **MySQL** scripts by rewriting queries and implementing indexing, cutting job runtime **50%** (4 h → 2 h) and server load **30%**, enabling timely investment decisions that contributed to a **3% increase** in portfolio returns.

Machine Learning Scientist | Unilever United States, Inc New York, NY | Sep 2023 – Dec 2023

- Developed a **multi-modal neural network** with **PyTorch** and **Keras** combining tabular product features with image embeddings extracted from a fine-tuned **VGG16** model, improving **R² by 54%** over the baseline **XGBoost** model in sales volume prediction and optimizing **\$3 million** in supply chain and inventory management cost.
- Engineered an **ETL pipeline** for e-commerce merchant datasets (~100 GB) in heterogeneous formats with **PySpark** and **Spark SQL**, enhancing data quality through advanced cross-joining techniques and Levenshtein distance matching.
- Deployed the model as an end-to-end pipeline on **GCP** enabling scalable real-time sales forecasting with scheduled and event-trigger workflows and managed task dependencies, enhancing operational efficiency **20%** per business team feedback.

Data Scientist, Personal Insurance R&D | The Travelers Companies, Inc Hartford, CT | Jun 2023 – Aug 2023

- Crafted and deployed a risk assessment model using **LightGBM**, augmented with rule-based triggers and fine-tuned via **Optuna** with customized objective functions, achieving estimated annual savings of **\$15 million** in claim resolving cost in backtesting.
- Developed a **Cox Proportional Hazard** model incorporating **SMOTE** to predict occurrences of roof discoloration and damage, achieving **Harrell’s c-index of 0.73** and enabling an inspection schedule that cut average customer life-time on-site inspection cost **12%** per policy.
- Conducted extensive **exploratory data analysis** and presented findings to leadership, demonstrating model-update benefits that boosted the project’s prioritization and lifted the referral action capture rate **33%** (from 32.4% to 43.1%).

Technical Skills

LLM & Generative AI	Transformers (Hugging Face), LLM & CLI coding agents, retrieval and ranking (RAG, BM25, graph retrieval), fine-tuning (LoRA, triplet loss), Sentence-BERT, few-shot learning, prompt design, OpenAI API, evaluation (Acc@k, Recall@k, MRR, BLEU)
ML & Deep Learning	PyTorch, TensorFlow, Keras, Scikit-learn, LightGBM, XGBoost, Optuna; attention mechanisms, cross-modal & multimodal architectures, generative models, self-supervised & contrastive learning, foundation model pretraining, time-series & biosignal modeling, survival analysis
Programming Languages	Python, SQL, R, Java, JavaScript, C++ , MATLAB, Bash, HTML, STATA
Big Data & Engineering	Spark, PySpark, Dask, Pandas, NumPy, Statsmodels, Hadoop, Airflow, Databricks, Delta Lake, MySQL, PostgreSQL, T-SQL, MongoDB, Flask, BeautifulSoup
Cloud & Visualization	AWS (EC2, S3, Redshift), Azure (VM, DevOps), GCP (VM, Cloud Storage, BigQuery); Tableau, Power BI, Seaborn, Matplotlib

Academic Service

- **Reviewer**, Conference on Neural Information Processing Systems (**NeurIPS**) **2026**.
- **Presenter (Oral)**, IEEE Engineering in Medicine & Biology Conference (EMBC) 2026.