Longtian Bao

lobao@ucsd.edu | github.com/LongtianBao | Google Scholar

EDUCATION

University of California, San Diego — *La Jolla, CA*
M.S. in Computer Science (GPA: 4.0/4.0) — *2024 – 2026 (Expected)*
B.S. in Computer Engineering (GPA: 3.8/4.0) — *2020 – 2024*

PUBLICATIONS

[1] Weili Cao*, Jianyou Wang, Youze Zheng*, **Longtian Bao***, Qirui Zheng, Taylor Berg-Kirkpatrick, Ramamohan Paturi, Leon Bergen. "Single-Pass Document Scanning for Question Answering." *Conference on Language Modeling (COLM)*, 2025. (Oral Spotlight, Top 2%). [Paper] [Code]

[2] Jianyou Wang*, Weili Cao*, **Longtian Bao**, Youze Zheng, Gil Pasternak, Kaicheng Wang, Xiaoyue Wang, Ramamohan Paturi, Leon Bergen. "Measuring Risk of Bias in Biomedical Reports: The RoBBR Benchmark." *Empirical Methods in Natural Language Processing (EMNLP)*, 2025. [Paper] [Code]

RESEARCH EXPERIENCE

Research Assistant — Laboratory for Emerging Intelligence, UC San Diego — *La Jolla, CA*
Advisors: Prof. Leon Bergen and Prof. Ramamohan Paturi — *Apr. 2024 – Present*

- Re-architected a Mamba-2 State-Space Model (SSM) by replacing its language modeling head with a classification head, creating a subquadratic model that preserves global context for efficient sentence retrieval in documents as long as 256k tokens [1].
- Designed and implemented a novel link-based synthetic data generation pipeline where an LLM identifies and leverages natural, long-range dependencies within documents to create coherent training examples for long-context question answering [1].
- Developed an LLM-assisted annotation-validation pipeline to construct a new information retrieval benchmark from 41 existing QA datasets, enabling a fine-grained evaluation of model performance in retrieving relevant sentences from long documents [1].
- Developed the RoBBR benchmark, a new resource for assessing risk-of-bias in biomedical research, derived from expert judgments on over 500 scientific studies to evaluate LLM capabilities in analyzing methodological strength [2].
- Designed two novel subtasks—Support Sentence Retrieval (SSR) and Support Judgment Selection (SJS)—to disentangle and independently measure the information retrieval and reasoning abilities of language models in the complex domain of bias assessment [2].
- Engineered an automatic, human-validated annotation pipeline using GPT-4 to create the ground truth for the retrieval task by decomposing expert judgments into fine-grained "aspects" and mapping them precisely to source sentences in biomedical papers [2].

TEACHING EXPERIENCE

Instructional Apprentice (Tutor) — UC San Diego — *La Jolla, CA*
Selected Quarters: Spring 2021 – Winter 2024

- **CSE 12 - Data Structures:** Spring 2021, Winter 2023, Winter 2024
- **CSE 11 - Accelerated Introduction to Programming:** Winter 2022, Fall 2022, Spring 2023
- **CSE 30 - Computer Organization:** Fall 2023
- **CSE 15L - Software Tools & Techniques Lab:** Spring 2021

PROJECTS

UCSD Autograder Platform *Fall 2023 – Spring 2024*

- Engineered and maintained a full-stack educational platform used by over 1,000 students across multiple departments, managing a database with millions of rows of student data.
- Developed key features including a containerized autograder for homework submissions and a real-time, ticket-based queue system for teaching assistant office hours.

Autonomous AI Racing Car *Winter 2024*

- Built and configured a 1/10 scale autonomous vehicle, integrating a Jetson Nano single-board computer with a VESC, LIDAR, and OAK-D AI camera.
- Utilized the DonkeyCar framework and TensorFlow to train a behavioral cloning model, enabling the car to autonomously navigate a racetrack by collecting and learning from human driving data.

DaCe Framework Reproducibility — Student Cluster Competition *Fall 2022*

- Evaluated the reproducibility of the DaCe high-performance Python framework on NPBench benchmarks using a novel AMD-based hardware configuration.
- Ported and debugged several NumPy benchmarks to the DaCe framework, analyzing performance speedups and identifying portability limitations.

SKILLS

Programming: Python (advanced), C/C++, Java, JavaScript/TypeScript, Bash
ML & NLP/IR: PyTorch, TensorFlow, Hugging Face Transformers, State-Space Models (Mamba-2), long-context QA & sentence retrieval, OpenCV
Backend & Data: Flask/FastAPI, PostgreSQL, REST APIs, real-time queue systems, basic PySpark
Systems & DevOps: Docker, Linux/SSH, Git/GitHub, AWS (EC2/S3), Distributed Training, Weights & Biases
Web/Frontend: React
Robotics/Edge: NVIDIA Jetson, LIDAR, OAK-D, VESC, DonkeyCar framework
LLM Tooling: OpenAI/Anthropic/Google APIs