YUNWEI ZHAO

New York, NY | 206-351-1702 | yz5944@nyu.edu | LinkedIn | GitHub

EDUCATION

Ph.D. in Computer Science, New York University   2024 - 2028 (Expected)
Advisor: Lakshminarayanan Subramanian, Research Area: Large-scale Reasoning in Health and Security

Master of Computer Science, Cornell University   2023 - 2024

Bachelor of Computer Science, University of Washington   2019 - 2023

RESEARCH PROJECTS

Single-Cell Gene Network Discovery via Contrastive Learning and Bayesian ML (In Submission)   2024 - present
- Designed BEACON, a novel gene regulatory network inference framework that combines contrastive learning with a two-headed projection mechanism.
- Conducted large-scale experiments and ablation studies that show the method achieves state-of-the-art AUROC and AUPRC scores across highly unbalanced SERGIO-simulated and real-world single-cell GRN datasets with edge density varying from 2.6% to 0.03%.

Privacy Policy Document Verification for Code Compliance (Ongoing)   2024 - present
- Working on large-scale policy-compliant code reasoning.
- Designed an end-to-end reasoning system that extracts and verify logic flows from natural language privacy policies.
- Built a SOTA model based on Llama 3.1 for extracting essential privacy information from policy, while balancing between performance, training efficiency, and computational cost with FSQP and 8-bit QLoRA.

Designing and Scaling a Dual-Platform Agricultural Technology in Rural Tanzania   2023
- Built and optimized an accessible Tanzanian farming directory app, eKichabi v2, to enable farmers to search for local businesses. Developed user authentication and efficient logging systems for survey group data collection and analysis. Work published in CHI 2024.

PUBLICATIONS

[1] eKichabi v2: Designing and Scaling a Dual-Platform Agricultural Technology in Rural Tanzania
- Ananditha Raghunath, Alexander Metzger, Hans Easton*, XunMei Liu*, Fanchong Wang*, Yunqi Wang*, **Yunwei Zhao***, Hosea Mpogole, Richard Anderson
- Accepted by CHI 2024 [paper]

OTHER PROJECTS

Veloda: A query-to-fashion-design interface supported by LLMs. Collected and formatted fashion design data into JSON format with a designed CoT prompt for the dataset, defined the query-to-fashion-design task as a text-to-JSON translation task, fine-tuned CodeFuse-DeepSeek-33B and GPT-3.5 with the dataset, and built an interactive local demo. [model] [demo]

BioTranslator: An open-source framework based on PubMedBert for translating textual descriptions to non-textual biological data (e.g., gene expression). The method encodes cell type controlled vocabularies to cell type embeddings and cell expressions to cell embeddings, then performs cross-modal learning to map cell type embeddings to cell embeddings, thus enabling users to get the embedding of a cell by giving a description of it. [Code]

MiniTorch: A mini version of PyTorch that supports automatic differentiation and backpropagation, tensor broadcasting, mathematical operations, auto-gradient functions, and CV-related functions including conv1d, conv2d, pooling, softmax, and dropout.

Byte-SVDCNN: A text classifier that tokenizes input text with byte-level BPE and uses Squeezed Very Deep Convolutional Networks (SVDCNN) to verify whether given text is in real or fake English. [Code]

INDUSTRY EXPERIENCE

Software Engineering Intern @Ketogenic.com   2024
- Led the development of the Frontend system for Pro2col Health, a cross-platform mobile app built in Flutter and Dart that allows customers to choose or design fitness plans, track diet and exercise progress, chat with fitness advisors, compete with friends, and win prizes.

TEACHING

CSE473: Introduction to Artificial Intelligence   Spring'22, Fall'22
- Answered questions on the Ed Discussion Board, led sessions with 10-20 students, and held weekly and extra office hours. Designed and developed several assignments.