

YUNWEI ZHAO

New York, NY | 206-351-1702 | yz5944@nyu.edu | [LinkedIn](#) | [GitHub](#)

EDUCATION

Ph.D. in Computer Science, New York University 2024 - 2029 (Expected)
Advisor: Lakshmi Subramanian, Research Area: NLP, ML

Master of Computer Science, Cornell University 2023 - 2024

Bachelor of Computer Science, University of Washington 2019 - 2023

RESEARCH PROJECTS

Privacy Logic Extraction and Reasoning with LLMs Based on Contextual Integrity Theory Ongoing
- Building a SOTA model for identifying contextual integrity components from text by fine-tuning open-source LLMs on the token classification task. Utilizing techniques such as FSQP and 8-bit QLoRA to increase training and generation efficiency.
- Formalizing logic flow extraction based on contextual integrity components.
- Implementing a new reasoning engine that step-by-step extracts clean logic flow in the form of first-order logic from natural language privacy policies.

Single-Cell Gene Network Discovery via Contrastive Learning and Bayesian ML Ongoing
- Designing an effective, robust, and scalable method for single-cell gene regulatory network (SC-GRN) prediction with contrastive learning methods to internalize regulatory relationships in embeddings and Gaussian ML to approximate the probability of an edge given a known network.
- Conducting comparison and ablation studies that show the method maintains consistently high AUROC and AUCPR scores across highly unbalanced SERGIO-simulated SC-GRN datasets with ratios of positive edges varying from 2.5% to 0.2%.

Designing and Scaling a Dual-Platform Agricultural Technology in Rural Tanzania
- 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
- 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 a 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. [\[Code\]](#)

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.