Yufu (Yvette) Li

yufuli2@illinois.edu | LinkedIn | GitHub

Research Interests

Core Focus: Explainable AI (XAI), Graph Representation Learning, Intelligent Tutoring Systems.

Themes: Self-Regulated Learning, Ethical AI in Education, Trajectory Modelling, Human-AI Alignment.

Education

University of Illinois Urbana-Champaign Urbana, IL, USA
Master of Computer Science (MCS) *Jan. 2022 – Aug. 2023*

- **Relevant Coursework:** Reproducibility in Machine Learning, Mining Cognitive and Metacognitive Models, Natural Language Processing, Practical Statistical Learning.
- **Key Project:** Conducted a comprehensive reproducibility study on Graph Attention Networks (GAT), analyzing scalability and interpretability trade-offs in session-based recommendation.

University of Waterloo Waterloo, ON, Canada
Bachelor of Mathematics, Honours Statistics & Computational Mathematics *Sep. 2015 – Aug. 2020*

- **Minors:** Computer Science, Economics
- **Undergraduate Research:** Developed stochastic SIR/SEIR models for COVID-19 progression, focusing on latent trajectory estimation using R-package development.

Research Experience

ICALM (Interactive-Constructive-Active Learning Machine) | *Principal Investigator* May 2023 – Present

- **Graph-Based Knowledge Representation:** Designed and implemented a pipeline to convert unstructured educational text into a **Concept Knowledge Graph** using Neo4j. Utilized LLMs for entity-relation extraction to map pedagogical dependencies (e.g., "Prerequisite-Of").
- **Pedagogical Alignment (ICAP):** Structured learner interactions around the **ICAP framework** (Interactive, Constructive, Active, Passive). Implemented diverse learning pathways, including "MetaNode" (structural review) and "Interest-Based" exploration, to support different cognitive states.
- **GraphRAG & Explainability:** Developed a Retrieval-Augmented Generation (RAG) system that grounds responses in specific graph nodes. This reduces hallucinations and enables the system to explain recommendations by tracing edges and surrounding nodes in the knowledge graph.
- **Funding & Impact:** Secured competitive funding ($7,000 USD total) from Microsoft for Startups and Google for Startups to support computational infrastructure.

Reproducibility Study: Graph Attention Networks (GATs) | *UIUC CS598* Feb. 2023 – Apr. 2023

- **Rigorous Validation:** Reproduced the original GAT paper (Velickovic et al.), achieving a 15% boost in micro-F1 accuracy on the PPI dataset by refining the training loop and hyperparameters.
- **Automated Optimization:** Built an automated hyperparameter tuning pipeline using **Ray Tune** and ASHAScheduler to efficiently explore the attention head and hidden layer search space.
- **Explainability Analysis:** Investigated the interpretability of attention weights. Conducted perturbation analysis to identify which subgraphs were driving predictions, highlighting specific scalability limitations in transductive learning settings.

Stochastic SIR/SEIR Modeling of COVID-19 | *Undergraduate Research* Jan. 2020 – Apr. 2020

- **Latent Trajectory Modeling:** Addressed the limitations of deterministic ODE models by developing a stochastic SEIR framework that accounts for unobserved incubation and exposure periods.
- **Statistical Inference:** Implemented **Markov Chain Monte Carlo (MCMC)** and **Maximum Likelihood Estimation (MLE)** in R (utilizing Rcpp for performance) to estimate parameters for latent infection trajectories.
- **Software Development:** Led a team of 4 students to package the model into a custom R-library, focusing on variance-based diagnostics to quantify uncertainty in pandemic progression predictions.

Mining Cognitive Models from Text | *UIUC CS447* Aug. 2022 – Dec. 2022

- **Metacognitive Analysis:** Conducted a literature review and synthesis on Natural Language Processing methods for extracting cognitive states (e.g., confusion, planning) from student text.
- Synthesized findings on how linguistic patterns correlate with self-regulated learning strategies, informing the design of the chatbot interaction module in ICALM.

Industry Experience

Bell Canada Enterprises Canada
Data Scientist II *Sep. 2020 – Present*

- **Recommender Systems at Scale:** Designed and deployed high-dimensional personalization models serving millions of users. Architectures included **Two-Tower Retrieval** for candidate generation, **Wide&Deep** for ranking, and Tabular Transformers for feature interaction.

- **Ethical AI & Metric Alignment:** While achieving a 24% uplift in nDCG, identified significant gaps between "engagement metrics" and user welfare (transparency/reflection). This disconnect directly motivated my pivot to research on explainable, human-centric objectives.
- **ML Engineering & MLOps:** Built robust inference pipelines using **FastAPI** and **Redis** for low-latency serving. Containerized applications with **Docker** and managed deployment on OpenShift/GCP, ensuring high availability for real-time marketing decisions.
- **Business Translation:** Acted as a technical lead bridging the gap between engineering teams and business stakeholders, translating complex model outputs into actionable insights for marketing strategy.

GeoMate Canada (Remote)
Computer Vision AI Engineer *Jan. 2020 – Mar. 2023*

- **AI for Accessibility:** Developed object detection models (EfficientDet) to identify sidewalk obstacles, directly supporting the creation of safe navigation maps for wheelchair users.
- **Bias Mitigation:** Addressed severe class imbalance in geospatial data through **Active Learning** strategies and targeted data re-balancing, ensuring the model performed equitably across different urban environments.
- **Pipeline Automation:** Established Continuous Integration (CI) pipelines for image labeling and model retraining, significantly accelerating the iteration cycle for accessibility features.

Ministry of Advanced Education and Skills Development Ontario, Canada
Research Assistant (Co-op) *Jan. 2018 – Apr. 2018*

- **Labor Market Analysis:** Conducted quantitative research on socioeconomic trends to support vocational training policy. Developed interactive **RShiny** dashboards to visualize workforce dynamics, aiding evidence-based decision-making.

Technical Skills & Grants

Programming: Python, R, SQL, C++, Scala, Bash, MATLAB
Deep Learning: PyTorch, TensorFlow (Keras), Scikit-learn, Horovod, Ray Tune
Graph & LLM: Neo4j, LangChain, GraphRAG, NetworkX
Data Engineering: Hadoop, Spark, BigQuery, Redis, PostgreSQL, Tableau
DevOps: Docker, Kubernetes, Google Cloud Platform (Vertex AI, Dataflow), Git
Certifications: Google Cloud Professional Machine Learning Engineer, PRINCE2 Project Management
Grants: Microsoft for Startups Founders Hub ($5,000), Google for Startups Cloud Program ($2,000)