Yizirui Fang

(667) 391-4284 | yizirui.fang@gmail.com | GitHub: yizirui | yiziruifang.com | LinkedIn: easey-f |

Education

Johns Hopkins University May. 2024
MSE in Computer Science, GPA: 3.9 *Baltimore, MD, USA*

Imperial College London Jul. 2019 - Aug. 2019
Machine Learning and Applied Statistics Program, 7.5 ECTS, 3.7/4.0 *London, UK*

University of Nottingham Jul. 2022
BSc in Computer Science with Honors, First Class, Top 5% *Nottingham, UK*

Research Interests

Embodied AI, Human-centered Machine Learning, Bayesian Decision Making, Uncertainty Quantification

Publication

- Hongming Li, **Yizirui Fang**, Shan Zhang, Seiyon Lee, Yiming Wang, Mark Trexler, and Anthony F. Botelho. "*ARCHED: A human-centered framework for transparent, responsible, and collaborative AI-assisted instructional design.*" To appear in *iRAISE 2025: Innovation and Responsibility for AI-Supported Education at the 39th AAAI Annual Conference on Artificial Intelligence*, Philadelphia, PA, USA. (Accepted)

- **Yizirui, Fang**, and Eric Nalisnick. "Learning to Defer with an Uncertain Rejector via Conformal Prediction." (*AISTATS* Under Review), also in *NeurIPS 2024 Workshop on Bayesian Decision-making and Uncertainty*

- Ying, Lance, Jason Xinyu Liu, Shivam Aarya, **Yizirui Fang**, Stefanie Tellex, Joshua B. Tenenbaum, and Tianmin Shu. "SIFToM: Robust Spoken Instruction Following through Theory of Mind." *arXiv preprint arXiv:2409.10849* (2024)

- Ying, Lance*, **Yizirui Fang***, Shivam Aarya*, Joshua B. Tenenbaum, and Tianmin Shu. "UnclearInstruct: An Embodied Assistance Challenge for Multi-modal Speech Instruction Following with Diverse Speech Condition" 2024 AAAI Symposium on Unifying Representations for Robot Application Development

- **Yizirui, Fang**, and Anthony Bellotti. "Investigating Data Usage for Inductive Conformal Predictors." arXiv preprint arXiv:2406.12262 (2024).

Research Experience

Social Cognitive AI Lab, Johns Hopkins Jan. 2024 - Present

- Proposed a Spoken Instruction Following through Theory of Mind (SIFToM) model to interpreate acoustic wave and human auditory perception to infer robot goals via Bayesian inverse planning algorithm' on simulated and real-world data advised by Prof. Tianmin Shu at Johns Hopkins University
- Developed and scaled simulation of human indoor activities and human spoken instructions with various adversarial environment to evaluate human-robot collaboration with **Habitat**
- Developed a LLM enhanced representation learning for a fusion of speech and visual representation data with **PyTorch** and **litgpt** to estimate the goal of human under unclear and ambiguous instructions
- Surveyed multi-modal LLM with vision of human motion environemtn and audio of human speech for semantic learning and their representation fusion with **OpenAI API**, **LlaMA** and **Gemini**
- Created **SIFT** GPU optical flow tracking pipeline with **OpenCV** with object and motion path highlighting

Learning to Defer with an Uncertain Rejector via Conformal Prediction Jan. 2024 - Dec. 2024

- Proposed a **uncertainty-based distribution-free** post-train component for learning to defer in enhancing the collaborative performance of human and AI team and rendering safer decisions on tasks ranging from object to hate speech detection via **uncertainly quantification**, advised by Prof. Eric Nalisnick at Johns Hopkins University
- Developed surrogate loss functions in **Bayesion sub-optimal** approaches for learning to defer problem on wide ResNet, human expert simulators and data augmentation with **PyTorch**

- Proposed active learning pipeline with **uncertainty quantification** methods including batch ensemble, SNGP, MC-Dropout, and BNN for Wide ResNet on CIFAR-10/100 dataset using **TensorFlow** and **Python**
- Developed and automated experiments on CIFAR10 w/ corruption, human, Hate Speech, and Street View dataset
- Surveyed distribution shift on wide ResNet using **OpenCV** and visualized using **matplotlib** and **seaborn**
- Automated and distributed experiments over GPUs with **Slurm**, **Shell**, and **Docker** at **Linux HPC**

Investigation of Data Usage for Inductive Conformal Predictors May. 2021 - Jun. 2022

- Proposed and demonstrated a data inducting algorithm to refine the machine learning uncertainty quantification with 54% upgrade in accuracy and 16% in efficiency with **Scikit Learn**, **seaborn**, and **TensorFlow** advised by Prof. Anthony Bellotti at University of Nottingham Ningbo China
- Proposed and proved with theoretical and empirical analysis the relationship among an uncertainty quantification algorithm, conformal predictor, three data sets and verified hypotheses with **Python**, **Scipy** and **Statsmodel**
- Surveyed neural networks architecture for cover type classification and assumption-free uncertainty quantification

Augmentation for Distribution Drift in Credit Scoring May. 2020 - Aug. 2021

- Proposed data augmentation algorithms with **kernel density estimation** against distribution drift of credit scoring models, and improve the AUC of ML models from 0.73 to 0.85 with **LightGBM** and **PyTorch** under various economic factors
- Surveyed credit risk models in gradient boosting, neural network algorithm with **Python**, **PyTorch**, **NumPy**
- Analysis and visualize experimental data statistically with **Pandas** and **matplotlib**
- Created large-scale databases for ∼2 bn financial time-series data points with **Spark** and **SQLAlchemy**

Development Experience

Software Dev Engineer Dec 2024 - Present

Amazon.com *Seattle, WA, US*

- Developed RESTful API for Amazon's global distributed warehouse inventory management, tracking system and cost component with **Java** and **AWS** in collaboration with **2** cross-functional teams

Full-stack Software Engineer Intern May. 2023 - Aug. 2023

League of Southeastern Credit Unions, Digital Operation Team *Remote, US*

- Delivered 2 web systems for 100+ credit unions using **TypeScript** and **Python**, attracting 10k users from 0 to 1
- Designed and developed for **RESTful API**s, middleware and data mutation including search, customizable dashboard and form with **Next.js**, **node.js** and **NextAuth**, customizable event triggers with **webhook**
- Engineered robust cache and real-time notification systems with **Redis** and **RabbitMQ** infra and designed TTL, cache eviction and polling policies, reduced fetch and search time by 32%, scaled to handle 50k tps
- DevOpsed with **Nginx**, **Docker**, **Jenkins**, **Shell** at **AWS** EC2, ECS and EKS, Tested with **Jest** in CI/CD

Software and Technology Developer Sept. 2020 - Aug. 2022

VR-Based Classroom Lab, University of Nottingham *Remote*

- **One pending patent**, One Innovational Software for Education with Immersive Technology (XR) for education proposes, Award Vice-Chancellor's Medal 2022
- Designed and implemented multi-player features: scene customization, player and object sync for PC and VR, by **RPC** and improved the network waiting time from 142s to 122s in **C#**, **OpenXR**, and **Photon Engine**
- Developed data storage for game status restoration and player inventory with **serialization** in **Unity**
- Developed UX of gaming, including player movement, player object interaction, avatar interaction, player dialogue, and, embedded browser with **Steam VR** and **Google VR** in MVC pattern for multi-devices with **C#**
- Coached STEM Programming Summer School and faculty training workshops each engaging over 50 people

Awards

- 2022 Dean scholarship at University of Nottingham
- 2021 Department Head Scholarship at University of Nottingham
- 2021 Mathematical Contest in Modeling, Honorable Mention

Technical Skills

Languages: Python, C/C++, C#, TypeScript, JavaScript, Java, SQL, Shell, OCaml, Haskell
Frameworks & Tools: PyTorch, TensorFlow, LightGBM, OpenCV, Slurm, Unity, PostgreSQL, Django, Docker