Xiaoyu Yuan

AI Researcher | Trustworthy AI Systems, Computer Vision, Software Engineering

xiaoyuyuan19@gmail.com | xiaoyuyuan19.github.io/portfolio | Github | Linkedin

Education

University of Helsinki, Helsinki, Finland ***Master in Computer Science***

Aug. 2023 – Jul. 2025 (GPA: 4) Track: Algorithms and Machine Learning **Full Scholarship**

Key Courses: Neural Networks & Deep Learning, Machine Learning in Molecular Biology, Advanced Course in Machine Learning, Statistics for Data Science, Information Retrieval, Big Data Platforms, Academic Writing (CEFR C1)

University of Oulu, Oulu, Finland ***Bachelor in Software Engineering (Double Degree)***

Aug. 2019 – Jun. 2023 (GPA: 4) Conferred jointly by Nanjing Institute of Technology & University of Oulu

Key Courses: Software Engineering, Requirements Engineering, Software Architectures, AI and Software Engineering, Information Systems in Organizations, Software Quality and Testing

Work Experience

Human-AI RLHF Evaluation Tasker | Outlier.ai *Freelance, Jan. 2025 – Present*

- Conducted large-scale preference **evaluation of LLM outputs** across tasks (e.g., summarization, Q&A).
- Specialized in **detecting hallucination**, **evaluating reward models**, and prompting refinement.
- Contribute to **RLHF** and comparison alignment pipelines, focusing on Reward Modeling and cue-response pairing.
- Gained firsthand insight into weak supervision signals such as noisy rankings and response edits.

Software Tester Intern | Tongdahai Information Technology Co., Ltd. *Nanjing, Jun. 2021 – Jul. 2021*

- Conducted **black-box testing** and analyzed 2000+ **UX feedbacks**
- Collaborated in **Agile sprints** and resolved bugs across platforms

Research Experience

Attention-based High-Precision Visual SLAM for GI Gastrointestinal Navigation Helsinki, Finland

Master Thesis Research, University of Helsinki *Dec. 2024 – Jun. 2025*

- Developed a localization framework in **GI environments**, analogous to anatomical tracking in medical imaging.
- **Enhanced ORB-SLAM3 with deep learning** for biomedical image-based localization in **endoscopy scenes**.
- Integrated Transformer depth estimation**,** leveraging **Foundation Models** for robust trajectory prediction.
- Developed a **attention U-Net**, refining **SLAM-based** depth perception and localization accuracy.
- Experience with remote **GPU and HPC services**.

Ancient Character Recognition Website with Transformer-based OCR Nanjing, China

Bachelor Thesis Research, Nanjing Institute of Technology & University of Oulu *Jun. 2021 – Jun. 2023*

- Developed a **Transformer-based OCR system** with a custom dataset to improve model accuracy.
- Built a **real-time OCR pipeline** at HoumaOCR [YouTube Demo Link], with **Flask & JS-based interfaces** for live uploads and visualization.
- Experience in dataset annotation pipelines and OCR text authenticity checks, relevant to text detection workflows.
- Contributed to **two international conference papers, one software copyright, and one patent**.
- Recognized with the **1st Prize** in the Chinese Collegiate Computing Competition and the **2nd Prize** in the Jiangsu Province Bachelor Thesis Award.

Projects Experience

Trustworthy AI & System Reliability

Preference Optimization in LLM: RLHF & DPO Exploration *Self-Initiated Research Project* *2025*

- Explored **Direct Preference Optimization (DPO)** as an efficient alternative to **RLHF** in **LLM alignment**.
- Conducted **mathematical derivations** of the DPO loss function, visualized **gradient flow and parameter updates**, and connected DPO with logistic regression.
- Compared the **DPO** pipeline with **RLHF + PPO** in terms of implementation complexity and stability.
- Created a **video presentation and demo slides** demonstrating model behavior and potential for deployment in human-in-the-loop settings [Video link].
- Insights gained in detecting preference consistency and hallucinations across hybrid LLM-human outputs.

Scalable Inference in Extreme Multi-Label Classification (XMC) *2025*

- Replications and benchmarks of the **SPARTEX** and **SOTA XMC models** were performed to analyse performance **under memory and latency constraints**.
- **Sparse modes** (block sparse softmax, fan-in sparse auxiliary header) and their impact on the representation capability were investigated.
- Adaptive semantic **bottleneck** designs are proposed to improve the efficiency and generality of sparse inference.

Computer Vision & Autonomous Systems

AI+BII: Generative AI for Architectural Image Inpainting *2024*

- Designed a **Stable Diffusion-based generative inpainting system** to reconstruct missing structures in historical architecture images.
- Integrated **VAE and U-Net** backbones in a latent compression pipeline to **boost fidelity and resolution**.
- Enabled multimodal generation via **CLIPText-guided conditioning** and **parameterized generation logic**.
- **Deployed** an **interactive demo on Hugging Face Spaces**, supporting custom mask drawing, prompt input, and guided sampling control for real-time generation.
- [YouTube Demo Link] & [GitHub Link]

Software Engineering & Data Systems

DSPaperUniverse Platform, Academic Literature Exploration and Visualization *2023*

- Designed an **Interactive Graph platform** to visualize citation networks and model keyword co-occurrence.
- Applied network analysis techniques to **explore structural properties** (e.g., edge clustering) and topic diffusion.
- The approach is extensible to **biological networks** (e.g., **PPI, GO DAG**) in bioinformatics applications.
- **Web-based service** design for scalable data visualization, extensible to data collection and management for NLP.
- [YouTube Demo Link] & [GitHub Link]

Rule-Based Conversational Agent with RASA *Course Project – Software Engineering* *2021*

- Led the design and implementation of a **RASA-based dialogue agent** for travel planning, with multi-intent support and **slot-based memory modeling**.
- Developed custom **NLU pipelines**, dialogue management stories, fallback and **edge-case handling policies**.
- Coordinated full-cycle documentation, from **stakeholder requirements** and slot mappings to evaluation reports.
- Delivered a 76-page engineering report outlining the **complete agent development lifecycle**.

Compiler Construction for Real-World Machine Code *2023*

- Built an e**nd-to-end compiler** from scratch as part of a project-based master's course.
- Implemented **lexical analysis, parsing, type-checking, intermediate representation**, and **code generation** for a real machine architecture.
- Increase proficiency in programming language theory and system abstraction through practical development.

Biomedical AI & Healthcare Applications

Enhancement Proposal for PANNZER via Protein Language Model *Self-Initiated Literature-Based Project* *2025*

- Conducted an independent literature-driven dissection of the PANNZER pipeline for Protein Function Prediction.
- Analyzed the integration of **Suffix-Array-based** retrieval with **GO enrichment mechanisms**.
- Proposed a future research agenda focusing on representation learning (e.g., **ProtBERT, ESM**) and **IR techniques** (e.g., **Transformer IR, Burrows-Wheeler Transform optimizations**).
- Mapped workflow logic with visual **annotations**, raising questions for potential modular enhancements.

CancerTypeNet: Tumor Type Classification using Mutational Signatures *2024*

Course project, Machine Learning in Molecular Biology

- Developed a classification pipeline to predict tumor types directly from trinucleotide-based **SBS mutational signatures** using **DNN, CNN**, **ensemble methods (RF, GB)**, and traditional classifiers.
- Performed dimensionality reduction and motif-level feature selection to extract high-impact sequence features.
- Achieved 79.46% test accuracy across 22 tumor classes in the **PCAWG** dataset.
- Integrated **explainable AI tools** to uncover context-dependent mutational hotspots, highlighting **sequence-level determinants** of cancer specificity.

Publications & Awards

- **X. Yuan**, Z. Zhang, Y. Sun, Z. Xue, X. Shao, & X. Huang. (2023). A new database of Houma Alliance Book ancient handwritten characters and its baseline algorithm. In Proc. of the 8th Int. Conf. on Multimedia Systems and Signal Processing (ICMSSP '23). ACM. DOI: 10.1145/3613917.3613923
- Z. Zhang, X. Huang, **X. Yuan**, & Y. Sun. (2023). HABFD: Houma Alliance Book facsimiles database. In Proc. of the IEEE Int. Conf. on Image, Vision and Computing (ICIVC '23). IEEE. DOI: 10.1109/icivc58118.2023.10269984
- 1st Prize, Chinese Collegiate Computing Competition (National-Level, China), 2023
- 2nd Prize, Jiangsu Province Bachelor Thesis Award (Provincial-Level, China), 2023
- 3rd Prize, Jiangsu Province University Mathematical Modeling Competition, 2021
- 1st Prize in the 1st "Prospective Cup" Meta-Intelligent Data Challenge (Track: Entity Recognition in Sedimentology Knowledge Map), held at the CENet2022 Conference, Haikou, China (Nov. 2022).

Technical Skills

Trustworthy AI & System Reliability: RLHF, DPO, AI Safety, Model Evaluation, Human-AI Interaction
Computer Vision & SLAM: ORB-SLAM3, Transformer Vision Models, Depth Estimation
Software Engineering: Agile Development, System Design, Production Deployment, Testing Frameworks
Machine Learning & Deep Learning: PyTorch, TensorFlow, Transformer Architectures, Optimization
Development & Deployment: Python, C++, Git, Docker, Linux, Cloud Services, HuggingFace, Web Frameworks