Siwen Zhang

ssswzh.github.io | LinkedIn | GitHub

Location: Amsterdam, The Netherlands

Email: sylviezh9080@hotmail.com | Mobile: (31)0626049080

Research Interests

Large Language Models, Agentic AI, Trustworthy AI, Bioinformatics, Biomedical Data Science, and AI for Healthcare.

Education

Leiden University, Leiden, The Netherlands *Master of Science in Computer Science* Sept 2025 – Aug 2027

The University of Edinburgh, Edinburgh, United Kingdom *Master of Science in Bioinformatics* Sept 2015 – Aug 2016

Related Courses

Agentic Large Language Models | Advances in Data Mining | Automatic Machine Learning | Evolutionary Algorithms | Introduction to Deep Learning | Reinforcement Learning | Text Mining | Information Retrieval | Recommender Systems

First year transcript in Leiden University: **average 8.5/10 across 12 MSc courses**

Technical Skills

Languages and Dev Tools	:	Python, R, Shell, C (basic), SQL (basic), Docker, Git
Machine Learning	:	TensorFlow, PyTorch, Keras, Sci-kit Learn, HuggingFace, LLMs
Visualization	:	Matplotlib, Seaborn, ggplot2, Plotly, Tableau (basic)

Projects

Planning and Reflection for Agentic LLMs *Agentic LLMs, ReAct, Reflexion*

- Built and evaluated ReAct- and Reflexion-inspired LLM agents with planning, loop detection, and verbal memory across three interactive benchmarks through BALROG framework.

End-to-End Small LLM Development with Nanochat *LLM Training Pipeline, Instruction Fine-Tuning*

- Built an end-to-end LLM pipeline with nanochat, covering BPE tokenizer training, pre-training a 13M-parameter Transformer, mid-training on MMLU/GSM8K, instruction fine-tuning, benchmark evaluation, and inference deployment.

Recommendation Challenge for Amazon Products *Kaggle, PyTorch, SASRec*

- Built an Amazon Top-10 product recommendation system using SASRec-based sequential modeling and metadata-aware candidate filtering, achieving 8th/60 place on the Kaggle leaderboard.

Reinforcement Learning Algorithms for CartPole Problem *TensorFlow, Keras*

- Implemented and compared Naive DQN, REINFORCE, AC, A2C, and PPO in CartPole environment. Several engineering tricks (mini-batch, GAE, etc.) were applied to improve performance and stability.

Experience Snapshot

Middle Bioinformatics Engineer *Feb 2017 – Sept 2024* Beijing, China

- Delivered MRD/HRD analytics for oncology studies, built platelet RNA-seq ML classifiers, supported 5 cell and gene therapy products, and productionized cloud workflows that reduced runtime and cost by 50%.
- Led the first long-read targeted sequencing development for rare disease diagnosis in China, built COVID-19 pathogen surveillance workflows, and benchmarked Nanopore, PacBio, and Illumina platforms.
- Developed oncology biomarker assays and reproducible genomic workflows for CNV, MSI, TMB, survival analysis, and multi-omics interpretation.

Certifications

- 5 published articles and 3 patents
- IBM Data Science Professional Certificate