

Shasha Zhou

📍 Exeter, United Kingdom ✉ sz484@exeter.ac.uk ☎ +44 7827 092546

🎓 EDUCATION

University of Exeter

PhD in Computer Science (Supervised by Dr. Ke Li)

Sep 2023 – present (expected Sep 2027)

Exeter, UK

University of Electronic Science and Technology of China (UESTC)

M.Sc. in Computer Science

Sep 2020 – Jun 2023

Chengdu, China

Hangzhou Dianzi University

B.Eng. in Computer Science

Sep 2014 – Jul 2018

Hangzhou, China

💼 EMPLOYMENT

Hundsun Technologies

Test Engineer

May 2018 – Sep 2019

Hangzhou, China

🔬 RESEARCH INTEREST

- **Foundation Models for Biology and Medicine.** Building and rigorously evaluating large-scale models for genomic and biomedical data, with a focus on biology-grounded interpretability that goes beyond anecdotal evidence.
- **LLM Agents for Scientific Discovery.** Knowledge-graph-grounded agents combining structured retrieval with multi-hop reasoning, deployed in high-stakes domains including medicine, plant science, and scientific-literature exploration.
- **Trustworthy and Robust ML.** Stress-testing modern ML systems through adversarial evaluation and multi-objective optimization, exposing failure modes in code intelligence and language models.

📖 PUBLICATIONS

* indicates equal contribution.

Theme 1: Foundation Models for Genomics and Biology

- [1] **S. Zhou***, M. Huang*, K. Li. “Position: Genomic Model Research Must Move Beyond Anecdotal Evaluation of Interpretability Methods.” *International Conference on Machine Learning (ICML)*, 2026. **Spotlight (top 5%)**.
- [2] M. Huang, **S. Zhou***, K. Li. “Augmenting Biological Fitness Prediction Benchmarks with Landscape Features from GraphFLA.” *Neural Information Processing Systems (NeurIPS)*, 2025. **Spotlight (top 2.2%)**.

Theme 2: LLM Agents for Science and Medicine

- [3] **S. Zhou**, M. Huang, J. Cole, C. Britton, M. Yin, J. Wolber, K. Li. “Assessing Automated Fact-Checking for Medical LLM Responses with Knowledge Graphs.” *AAAI Conference on Artificial Intelligence (AAAI)*, 2026. **Oral (top 5%)**.
- [4] M. Huang*, **S. Zhou***, Y. Chen, K. Li. “Conversational Exploration of Literature Landscape with LitChat.” *International Joint Conference on Artificial Intelligence (IJCAI)*, 2025.
- [5] H. Yu*, **S. Zhou***, M. Huang*, L. Ding, Y. Chen, Y. Wang, Y. Ren, N. Cheng, X. Wang, J. Liang, H. Zhang, Y. Ding, K. Li. “PlantScience.ai: An LLM-Powered Virtual Scientist for Plant Science.” *Molecular Plant (IF=24.1)*, 2026.

Theme 3: Adversarial Robustness and Multi-objective Optimization

- [6] **S. Zhou**, M. Huang, Y. Sun, K. Li. “Evolutionary Multi-objective Optimization for Contextual Adversarial Example Generation.” *ACM International Conference on the Foundations of Software Engineering (FSE)*, 2024.
- [7] **S. Zhou**, K. Li. “Empirical Studies of Resampling Strategies in Noisy Evolutionary Multi-objective Optimization.” *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2023.

- [8] **S. Zhou**, K. Li, G. Min. “Attention-Based Genetic Algorithm for Adversarial Attack in Natural Language Processing.” *International Conference on Parallel Problem Solving from Nature (PPSN)*, 2022.
- [9] **S. Zhou**, K. Li, G. Min. “Adversarial Example Generation via Genetic Algorithm: A Preliminary Result.” *Genetic and Evolutionary Computation Conference Companion (GECCO)*, 2022.

Pre-prints

- [10] M. Huang, **S. Zhou**, K. Li. “Mapping Literature Landscapes with Data-Driven Discovery: A Case Study on MOEA/D.” Under review at *IEEE Computational Intelligence Magazine (IF=10.3)*, 2026.

🔗 OPEN-SOURCE SOFTWARE

- **OmniGenBench** [COLA-Laboratory/OmniGenBench]
A modular benchmarking platform for evaluating genomic foundation models (gFMs) across diverse tasks like RNA structure prediction, gene function classification, and multi-species generalization.
- **Genomic FM Interpretability Toolkit** [COLA-Laboratory/EvalXAI]
Implements attention visualization, feature-attribution methods, and biology-grounded evaluation protocols (motif overlap, counterfactual perturbations) to diagnose what genomic foundation models actually learn versus what they latch onto from dataset artifacts (companion to the ICML 2026 Position paper).
- **PlantScience.ai** [COLA-Laboratory/autoSKG]
An LLM-powered virtual scientist for plant biology, backed by a domain knowledge graph aggregated from 11 databases and large-scale literature (4M+ entities, 10M+ edges). Supports multi-hop, source-grounded reasoning for plant-science researchers (companion to the *Molecular Plant* paper).
- **EMOC** [COLA-Laboratory/EMOC]
A C++ framework for single- and multi-objective evolutionary optimization with 30+ algorithms and 80+ benchmark problems, packaged with a cross-platform GUI. Co-led development; adopted by the EC research community with substantial runtime speedups over Python/Java alternatives.

🏆 GRANTS AND HONORS

Grants & Scholarships

- Modal Research Grant (\$10,000) 2026-04
- Modal Research Grant (\$10,000) 2025-12
- OpenAI Researcher Access Program (\$5,000) 2025-06

Grant Proposal Contributions (Co-drafted proposal and ran proof-of-concept experiments)

- **BBSRC Transformative Research Technologies (£194,564)** (PI: Dr. Ke Li). 2025
Explainome: Open-Source Automated Explainable AI Agent Powered by Large Language Models and Knowledge Graphs for Computational Biology
- **EPSRC IAA Impact Visionary Award (£48,882)** (PI: Dr. Ke Li). 2025
Large Language Model Empowered Explainable Automated Machine Learning for Early Diagnosis of Parkinson’s Disease

🌐 SERVICES

Reviewing: IEEE Transactions on Evolutionary Computation (2023 - 2025), KDD (2025), ICML (2026).

Teaching Assistant: Software Development (2023)

🎤 TALKS AND PRESENTATIONS

- ICML’26 Spotlight Poster Seoul, Jul 2026
- AAAI’26 Oral Presentation Singapore, Jan 2026
- NeurIPS’25 Spotlight Poster San Diego, Dec 2025
- SMC’23 Oral Presentation Hawaii, Oct 2023