Song Ouyang

Algorithm Engineer Intern | +86-158-2750-5700
Beijing, China | ouyangsong@whu.edu.cn
AI for Science, Multimodal Learning | https://ouyang-song.com

Education

- **Wuhan University (WHU)** *Sep. 2023 – Jun. 2026*
  M.S., Artificial Intelligence (Recommended Admission) GPA: 3.64/4.0
  - **Scholarships & Honors:** First-Class Scholarship for Outstanding Freshmen, Second-Class Scholarship, Outstanding Graduate Student
  - **Key Courses:** Machine Learning (A), Advanced Computer Graphics (A), Natural Language Processing Technology (A), Machine Learning in Computer Vision (A), Research of Methodology of Computer Science (A+)

- **Northeastern University (NEU)** *Sep. 2019 – Jun. 2023*
  B.E., Software Engineering (Big Data & AI Track) GPA: 4.23/5.0 (**Top 1%**)
  - **Scholarships & Honors:** National Scholarship (2 times, **Top 1%**), First-Class Scholarship (4 times), *Xianggangyucai* First-Class Scholarship (Top 1/400+), Outstanding Graduate of Liaoning Province, Outstanding University Student of Shenyang City, Outstanding Student (Cadre) Model (3 times)
  - **Key Courses:** Advanced Mathematics (A+), Discrete Mathematics (A+), Python for Data Analysis (A+), Data Mining: Principles and Algorithms (A+), Basic of Program Design (A+), Object-Oriented Programming and Design (A+), Introduction to Artificial Intelligence (A+), Introduction to Deep Learning (A+)

Publications

- **Song Ouyang**, Huiyu Cai, Yong Luo, Kehua Su, Lefei Zhang, Bo Du. *MMSite: A Multi-modal Framework for the Identification of Active Sites in Proteins.* **NeurIPS**, 2024.
- **Song Ouyang**, Yong Luo, Huiyu Cai, Dongjing Shan, Kehua Su, Fei Liao, Na Zhan, Huangxuan Zhao, Tailang Yin, Lin Zhao. *M^3Site: Leveraging Multi-Class Multi-Modal Learning for Accurate Protein Active Site Identification and Classification.* **Briefings in Bioinformatics**. *(Under review)*
- Ke Shi†, **Song Ouyang**†, Kehua Su, Yong Luo, Bo Du, Gang Liu, Zhiwen Liang. *Anatomy-Conserving Unpaired CBCT-to-CT Translation via Schrödinger Bridge.* **MICCAI**, 2025. *(Under review)*

Projects

- **Multimodal Protein Active Site Identification & Classification (Paper Under Review)** *Jul. 2024 – Jan. 2025*
  Independent Project
  - Collected and preprocessed 53.5k high-quality protein samples from Swiss-Prot and AlphaFold DB, including functional annotations, active site labels, and 3D structures. Clustered dozens of active site labels with similar meanings into six categories using Large Language Models (LLMs).
  - Proposed M^3Site, a multimodal learning framework leveraging ESM-3, PubMedBERT, and EGNN (with physicochemical features). Designed FunICross and AWF modules for multimodal fusion and introduced inter-loss and center-loss to tackle class imbalance. Achieved a 31.7% improvement in F1 score over baselines.
  - Deployed M^3Site using Gradio and Docker, enabling users to predict, analyze, and visualize active sites.

- **Protein Inverse Folding Using Discrete Diffusion & Autoregression** *Aug. 2024 – Sep. 2024*
  Independent Project
  - Reviewed research on inverse protein folding and sequence generation with diffusion models.
  - Addressed exposure bias in Teacher Forcing models, which struggle to model dependencies between generated amino acids. Proposed a diffusion strategy in noise space, where each amino acid was assigned a distinct noise during training.
  - Utilized an RNN-based generative architecture with EGNN-extracted structural features and D3PM to generate discrete amino acid types. Preliminary results achieved 40% sequence recovery on CATH 4.2.

- **Renewable Energy Power & Generation Forecasting Using AI** *Oct. 2023 – Dec. 2023*
  Core Algorithm Developer
  - Used multiple imputation and random forests to fill missing photovoltaic (PV) power generation data and engineered domain-specific features.
  - Designed a hybrid regression-time series (HGBR + Informer) model for distributed scenarios, utilizing attention-based adaptive weighting and cross-domain transfer learning to mitigate data scarcity issues. Achieved 93.7% power prediction accuracy and 82.6% energy generation accuracy.
  - Implemented stacking ensemble learning for centralized PV farms, combining random forests, HGBR, and DNNs, with Lasso as a meta-learner, reaching 93.7% accuracy.

Experience

- **BioGeometry Biotech (Beijing)** *Feb. 2023 – May. 2023*

  Backend Engineer Intern

  – Developed the backend for an intelligent protein computing platform using Flask, integrating Slurm for deep learning model scheduling, Nginx & uWSGI as web servers, MySQL & Redis for database management, and Celery & WebSockets for task execution.
  – Independently completed Version 1.0 backend, which was showcased at the 2023 Hangzhou Yunqi Conference and adopted by multiple leading pharmaceutical companies.

- **Huawei Ascend AI Innovation Center (Shenyang)** *May. 2022 – Aug. 2022*

  Algorithm Engineer Intern

  – Migrated and optimized the MAE model for Huawei's MindSpore framework, achieving comparable performance to the original paper on ImageNet.
  – Researched cross-modal retrieval (text-to-image search) and contributed to the development of an AI + Security cross-modal information processing platform.

Competition Awards

- **National Second Prize** | China Post-Graduate Mathematical Contest in Modeling*, 2024
- **National Bronze Award** | National College Student Algorithm Design and Programming Challenge*, 2022
- **National Second Prize** | Lanqiao Cup National Software and Information Technology Professional Talent Competition, 2022
- **National Third Prize** | China College Student Computer Design Competition*, 2021
- **Honorable Mention** | Mathematical Contest in Modeling*, 2021
- **Provincial First Prize** | National Mathematical Modeling Competition*, 2021
- **Provincial Silver Medal** | China International College Students' "Internet+" Innovation and Entrepreneurship Competition, 2021
- **Second Prize in the Northeast Region** | WeChat Mini Program Application Development Competition, China Collegiate Computing Design Competition*, 2021
- **Provincial Second Prize** | National College Mathematics Competition, 2020

* I am the team leader in this competition.

Positions of Responsibility

- **Class President**, Administration Class in WHU *2023 – President*
- **Assistant Minister**, Student Union of Software College in NEU *2020 – 2022*
- **Class Monitor**, Administration Class in NEU *2020 – 2023*

Skills & Competencies

- **Languages**: CET-6
- **Programming**: Proficient in Python, Java, C/C++, familiar with Matlab, Shell
- **Framework & Tools**: Expert in PyTorch, NumPy, Scikit-learn, Pandas, Matplotlib, familiar with MindSpore, BioPython, TorchDrug