

CV

Name: Menghao Guo

Email: menghao.guo.319@gmail.com

GitHub: <https://github.com/JackKuo666>

Huggingface: <https://huggingface.co/jackkuo>



Profile

- **Five years of AI experience:** Currently a Research Specialist at [Zhejiang Lab](#).
- **Expert in NLP, deep learning, and LLM applications:** Led projects in text mining, knowledge graph construction, and multi-modal data integration, with a focus on leveraging LLM agents and AI for life sciences, including biocomputing, bioinformatics, and computational biology. Specialized in: Protein and genome foundation models, Multi-agent systems for life sciences data mining, Bioinformatics analysis and database construction, Drug repurposing, and Genomics-based personalized medicine.
- **Open-source contributor:** Actively contributed to open-source projects such as [AllenNLP](#) and [MindSpore](#) by enhancing model performance and functionality.
- **Biocomputation projects:** Involved in AMP-SEMiner, CRISPR-Cas Systems Discovery, Enzyme Kinetics Data Extraction (Enzyme Co-Scientist), and LLM Agent-based Chemical Literature Data Mining.
- **Diverse industry background:** Extensive experience across Huawei (2012 Lab), Ant Group, Ping An Technology, China Tobacco Zhejiang Industrial, and the Research Center for Life Sciences Computing of Zhejiang Lab.
- **PhD goal:** I am currently seeking a PhD position to expand my research on AI applications in life sciences, with a focus on developing data-driven solutions - integrating LLM agents with bioinformatics, causal inference, and reinforcement learning to advance precision medicine, enable drug repurposing, and optimize therapeutic strategies while delivering genotype-tailored recommendations.

Education

- **Master's Degree in Navigation, Guidance and Control**
Northeastern University, China (Sep 2017 – Jan 2020)
Focused on unsupervised deep learning, multitask learning, and visual odometry.

- **Bachelor's Degree in Electronic Information Science and Technology**
Nanyang Normal University, China (Sep 2013 – Jun 2017)
Coursework in **signal processing, software programming, and mathematical modeling.**

Work Experience

- **Zhejiang Lab | Research Specialist**
Mar 2022 – Present
Leading and collaborating on various AI and bioinformatics projects, with a focus on LLM applications and data-driven solutions.
- **China Tobacco Zhejiang Industrial Co., Ltd. | Algorithm Researcher**
Aug 2020 – Mar 2022
Applied machine learning and spectral analysis techniques to optimize tobacco quality, integrating heterogeneous data sources to develop predictive models.
- **Huawei Communication Technology Co., Ltd. | Algorithm Engineer**
May 2020 – Aug 2020
Optimized deep learning frameworks, including MindSpore, for NLP applications such as BERT, enhancing model training efficiency and scalability.
- **Zhejiang Ant Micro Financial Service Group | NLP Algorithm Engineer (Intern)**
Jun 2019 – Sep 2019
Developed multi-task learning enhancements within the AllenNLP framework, improving model robustness in dialogue and data extraction tasks.
- **Ping An Technology (Shenzhen) Co., Ltd. | NLP Algorithm Engineer (Intern)**
Oct 2018 – Jan 2019
Built an information extraction system for contract review using BERT and rule-based methods, advancing practical applications of NLP in structured data analysis.

Research/Project Experiences (Zhejiang Lab)

- **LLM Literature Extraction:** Led a project to develop prompt design and model fine-tuning techniques for automatic extraction of key scientific data from literature, enabling scalable biomedical data mining (A scientific paper published).
- **Knowledge Graph Construction:** Integrated LLM-Agents with knowledge graphs in the KOSA project to fuse multi-source data into complex relational networks (A scientific paper published).
- **Enzyme Kinetics Extraction (Enzyme Co-Scientist):** As co-first author, developed an LLM-based framework that extracted over 91,000 enzyme kinetics entries from ~3,500 papers with high accuracy (A manuscript has been preprinted and is under review).

- **AMP-SEMiner:** As co-first author, created a framework using Protein Language Models to discover 1.6 million antimicrobial peptide candidates by integrating genomics, molecular dynamics, and LLM technology (A manuscript has been preprinted and is under review).
- **SLMP Platform:** Proposed and implemented an LLM-powered platform for automated literature summarization, keyword extraction, and article analysis (A scientific paper published).
- **Chemical Literature Mining:** Contributed to an LLM Agent system for extracting reaction predictions and molecular properties from chemical literature (A manuscript has been preprinted).
- **CRISPR-Cas Discovery:** Participated in leveraging foundation models to identify novel Cas λ homologs for advancing gene editing research (A scientific paper published).

Skills & Technologies

- **Programming:** Python, Java, C/C++, JavaScript
- **NLP & AI:** Proficient with LLMs (e.g., GPT-4o, DeepSeek-R1, BioBERT), BERT variants, and classical NLP techniques (HMM/CRF, LDA, word2vec, FastText)
- **Frameworks & Tools:** HuggingFace, TensorFlow, PyTorch, Docker, Git
- **Data & Graph Analysis:** Pandas, NumPy, Matplotlib, Neo4j, RDF, SPARQL
- **Advanced Techniques:** RL, Multi-Agent Coordination, Causal Inference, KG Construction

Publications and Manuscripts

1. **Menghao Guo**, et al. "Investigations on scientific literature meta information extraction using large language models." *IEEE ICKG 2023*.
2. **Menghao Guo**, et al. SLMP: A Scientific Literature Management Platform Based on Large Language Models. *IEEE ICKG 2024*.
3. Fei Huang, et al., **Menghao Guo**, et al., "KOSA: KO enhanced salary analytics based on knowledge graph and LLM capabilities." *IEEE ICDMW 2023*.
4. Li, W., et al., **Menghao Guo**, et al. "Discovering CRISPR-Cas System with Self-processing Pre-crRNA Capability by Foundation Models." *Nature Communications (2024)*.
5. [Preprints] Wenhui Li[†], Baicheng Huang[†], **Menghao Guo**[†], et al. "Unveiling the Evolution of Antimicrobial Peptides in Gut Microbes via Foundation Model-Powered Framework". *preprint and under review at Cell Reports*.
6. [Preprints] Jinling Jiang[†], Jie Hu[†], Siwei Xie[†], **Menghao Guo**[†], et al. Enzyme Co-Scientist: Harnessing Large Language Models for Enzyme Kinetic Data Extraction from Literature.

*preprint and under review at **Nature Computational Science**.*

7. *[Preprints]* Chen, K., et al., **Menghao Guo**, et al. "An Autonomous Large Language Model Agent for Chemical Literature Data Mining." (*arXiv*)
8. *[Master Thesis]* **Menghao Guo**. Research on Visual Odometry Algorithm Based on Unsupervised Deep Learning [D]. 2020.

(†Co-first author)

Additional Qualifications

- **Certifications:** Advanced Information Systems Project Manager (2024)
- **Honors:** Zhejiang E-Class Talent (2024)
- **Conferences:** Oral presentations at IEEE ICKG, ICDM workshops
- **Patents:** 11 related to **tobacco quality optimization** and **biotech algorithms**. [[Search in Google Patents](#)]