



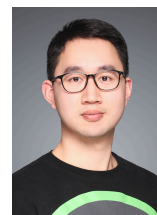
# Biyang Guo

Self-Motivated Researcher | Creating & Sharing | Have Fun with Tech

☎ +86 185-727-56723

✉ [guo\\_biyang@163.com](mailto:guo_biyang@163.com)

📍 Shanghai, China



## 🎓 Education

### AI Lab, Shanghai University of Finance and Economics

2020 - 2024

- **Ph.D.** NLP and Data-Centric AI, supervised by Prof. [Hailiang Huang](#) and Prof. [Songqiao Han](#)
- My research is published at [AAAI](#) (the top-tier conference of AI) and several papers are under review
- National Scholarship awardee

### Shanghai University of Finance and Economics

2018 - 2020

- **MA.** NLP applications, supervised by Prof. [Hailiang Huang](#) and Prof. [Songqiao Han](#)
- Served as a student team tech lead in a school-enterprise cooperated project to provide NLP-based AI analysis services for a large financial enterprise

### Shanghai University of Finance and Economics

2014 - 2018

- **B.S.** Electronic commerce and information management, supervised by Prof. [Wenting Tu](#)
- A college student entrepreneur, founder of an online notes sharing/selling platform
- Shanghai Outstanding Graduates awardee, Innovation and Entrepreneurship Fund

## 👤 Working Experience

### Microsoft Research Asia (MSRA)

2022.3 - 2022.11

- **Research Intern** at the NLC group. Supervised by Dr. [Yeyun Gong](#) and Dr. [Nan Duan](#).
- During internship, I proposed the **GENIUS model**, a sketch-based generative language model pre-trained on large corpus. GENIUS can generate coherent paragraphs given only some keywords/phrases, making it a powerful tool for AI-aided writing and NLP data augmentation. GENIUS is open-sourced and sponsored by Hugging Face, with **500+** downloads every month.
- Click to see the demo: [GENIUS Demo](#)

## 💻 Project

### ChatGPT Comparison and Detection

2022.12 - Now

**Project Initiator and Tech Lead.** Just 10 days after the launch of **ChatGPT**, I organized a team of **8** PhD students and engineers to conduct a project on the comparison between human and ChatGPT. After 40 days of data collecting, statistical analysis and deep-learning/machine learning models developing, we put forward the [world's first Human-ChatGPT comparison corpus](#) and the first open-sourced bilingual **ChatGPT content detectors**, which have raised wide interests across academic and industrial communities. Our online detectors have attracted **20K+** visits worldwide, with models downloads **3K+**/month, datasets downloads **1K+**/month and research paper receives **7** citations in a month.

Click to see the demo: [ChatGPT detectors](#)

### NLP Analysis Platform for Ping An Group

2020.6 - 2021.1

**Student Tech Lead.** In this school-enterprise project, I served as a student tech lead to develop several important NLP-based models, including classification, key-information extraction and semantic similarity/searching based on modern deep learning frameworks (PyTorch, Keras) and build RESTful services based on Flask, Unicorn and Docker.

## Research

---

1. **Label Confusion Learning to Enhance Text Classification Models**  
(AAAI-21, citations 26, Github Stars 100+, keywords: text classification, robust training)  
**Biyang Guo**, Songqiao Han, Xiao Han, Hailiang Huang, Ting Lu
2. **How Close is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection**  
(Arxiv-2023, citations 7, Github Stars 512+, keywords: ChatGPT, detection, comparison corpus)  
**Biyang Guo**, Xin Zhang, Ziyuan Wang, Minqi Jiang, Jinran Nie, Yuxuan Ding, Jianwei Yue, Yupeng Wu
3. **GENIUS: Sketch-based Language Model Pre-training via Extreme and Selective Masking for Text Generation and Augmentation**  
(Arxiv-2022, citations 1, Github Stars 120+, keywords: natural language generation, data augmentation)  
**Biyang Guo**, Yeyun Gong, Yelong Shen, Songqiao Han, Hailiang Huang, Nan Duan, Weizhu Chen
4. **Selective Text Augmentation with Word Roles for Low-Resource Text Classification**  
(Arxiv-2022, citations 3, keywords: data augmentation, text classification)  
**Biyang Guo**, Songqiao Han, Hailiang Huang

## Community

---

♦ **Beloved Technology Writer.** I enjoy writing **technical blogs** in Chinese knowledge sharing community **Zhihu**, pursuing to explain difficult and complex theoretical knowledge in easy-to-understand language. Representative works (in Chinese) are as follows:

- How does OpenAI train the GPT? Paper reading for InstructGPT, 2022 (160+ Likes)
- A friendly guide to Bayesian optimization and Gaussian process regression, 2022 (220+ Likes)
- The right way to understand L1/L2 normalization, 2021 (260+ Likes)
- 12 hours hard work to get you to figure out Seq2Seq in 20 mins, 2021 (900+ Likes)
- When will I know you - Graph Convolutional Networks (GCN), 2019 (3500+ Likes)
- Understanding Convolutional Neural Networks (CNN) from here, 2018 (2600+ Likes)

Including the above works, my series on **Zhihu** platform have received **20K+ likes** and **35K+ favourites**. Many of my articles have been included in **Zhihu Round Table Selections**.

♦ **Research Organization Founder.** Based on these activities, I founded the **SimpleAI** community with more than **10K** followers on WeChat. At the end of 2022, after the launch of ChatGPT, I gathered PhD students and engineers from 6 universities/companies in China and abroad in the SimpleAI community to develop and open source the world's first bilingual ChatGPT content detector and related datasets, which achieved a great response worldwide. The details are listed in the "Project" part above. This has been a very memorable experience for me in community research and team management.

♦ **Open-sourcing Enthusiast.** I have open-sourced several projects on GitHub, which have accumulated nearly **900 Stars**. I and the organization I founded have open-sourced **8 deep learning models and 3 datasets** on the Hugging Face platform, with several thousands downloads every month by the community. I also joined the Hugging Face Chinese Translation Founding Group, volunteering to translate technical courses for the Chinese community.

♦ **Academic Reviewer.** In the academic community, I served as a volunteer **reviewer** for conferences or journals including ACL, EMNLP, Applied Intelligence and Information & Management.