Yonghao Yang

2023302061214@whu.edu.cn | WeChat: yyh142857369 | GitHub: miracleyang-dev | Generative AI

EDUCATION

The Chinese University of Hong Kong, Department of Statistics *Aug 2027 entry*

PhD in Statistics — supervisor confirmed; formal admission in progress

Wuhan University, School of Mathematics and Statistics *Sep 2023 – Jun 2027 (expected)*

Major: Data Science and Big Data Technology (BSc)

Major GPA: 93.8/100 (3.92/4.0) **Major Rank:** 1/63

Coursework: Probability (99), Statistics (97), Numerical Analysis (100), Real Analysis (94), Neural Networks (98).

RESEARCH EXPERIENCE

Independent Study: Score, Flow & Drifting Models *Jul 2026 – Aug 2026*

- Read NCSN, DDPM, Score-SDE, Flow Matching and Drifting, tracing how one denoising objective is re-derived under different assumptions; derived the forward / reverse-time SDEs, the probability-flow ODE and the continuity equation, checking each step against the papers.
- Reproduced 1-D / 2-D toy samplers for Euler–Maruyama and probability-flow ODE, comparing discretisation error and sampling cost.

Sexism Detection in Chinese Social Media *Feb 2026 – Apr 2026*

- Built a unified SWSR pipeline (text normalisation, pinyin / wubi features, stratified splits, fixed seeds) and benchmarked N-gram, RNN, LSTM, Transformer and BERT on macro F1, with feature and classifier ablations.
- Traced the bottleneck to annotation noise: three LLMs re-annotated every sample with majority-vote arbitration, compared on the original and cleaned corpora; after cleaning, BERT macro F1 rose from 0.7824 to 0.9007 (+11.8 pp), and pinyin / wubi became the best configuration.
- Designed CMLG (CKBERT, BGE, pinyin and wubi embeddings → additive attention → Bi-GRU). Sole-author preprint, Zenodo 2026, DOI 10.5281/zenodo.21104764.

Deep Learning for Spatial Multi-omics *Oct 2025 – Jan 2026*

- Surveyed ResNet (image encoding), GNN (spatial adjacency), VAE (latent representation) and GAN (cross-modal translation) in spatial multi-omics, and how self-supervised, contrastive and distillation objectives are adapted to this setting.
- Reproduced selected papers under one PyTorch data and training pipeline, checking each reported metric and recording any deviation and its plausible cause.
- Benchmarked GNN / VAE / GAN on single-cell and bulk RNA-seq under shared splits and evaluation code, reporting performance together with failure modes in a comparison report.

AWARDS & COMPETITIONS

National Scholarship	Ministry of Education, 2025
Diangong Cup (Mathematical Modeling)	National First Prize, 2026
Lanqiao Cup (Algorithm Design)	Provincial First Prize, National Third Prize, 2025
First-Class Academic Scholarship	Wuhan University, 2025
Outstanding Student	Wuhan University, 2026

SKILLS & LANGUAGES

Programming	Python, C / C++, R, SQL
ML / DL	PyTorch, scikit-learn, NumPy, Pandas, Matplotlib
Tools	Git, LaTeX, Jupyter, Markdown
Languages	IELTS 6.5 (all bands ≥ 6.0)