Shansan Gong

hisansas@gmail.com | Google Scholar | GitHub | Homepage   *The University of Hong Kong*

Summary

PhD candidate at HKU, supervised by **Lingpeng Kong**. I work on **text diffusion models** for conditional generation, reasoning, and scaling, and develop their next generation abilities as diffusion language models (**DLMs**). I'm also interested in **long-context modeling** and building personalized, tool-augmented systems.

Education

PhD in Computer Science   2023.09 – Present (estimated: 2027.01) *The University of Hong Kong (HKU)*   *Hong Kong*
M.Sc. in Computer Science (GPA 3.78)   2019.06 – 2022.03 *Shanghai Jiao Tong University (SJTU)*   *Shanghai*
B.Eng. in Information Engineering (GPA 3.84) 2015.09 – 2019.06 *Shanghai Jiao Tong University (SJTU)* *Shanghai*

Research Experience

Research Intern — Code Generation with DLMs; RL Post-training   2025.01 – 2025.07 *Apple*   *Seattle, US*

- **Masked diffusion models for code**: analyzed generation order vs. accuracy/diversity in DLMs for code; trained 7B-scale diffusion coder with competitive performance to AR coders.
- **RL post-training for DLMs**: introduced a coupled-sampling GRPO variant tailored to masked diffusion models.

Research Intern — Reasoning & Scaling of Diffusion LMs   2023.11 – 2024.10 *Tencent AI Lab*   *Shenzhen*

- **CoT for diffusion LMs**: proposed/implemented diffusion-native CoT (DoT: reasoning steps diffusing over timesteps); evaluated on arithmetic/logic/math reasoning with self-consistency variants.
- **Scaling via AR → Diffusion adaptation**: developed continual-pretraining pipelines to convert AR models (GPT/LLaMA scales) into diffusion LMs; ran ablations on training tokens/inference speed vs. quality.

Researcher — Diffusion Text Generation; Long-Context Modeling 2022.05 – 2023.08 *Shanghai AI Lab* *Shanghai*

- **Initial exploration of text diffusion for Seq2Seq**: designed classifier-free training and decoding recipes for conditional generation; established connections among AR/NAR/diffusion paradigms in practice.
- **Acceleration & bridging discrete/continuous spaces**: integrated ODE-based samplers and a soft absorbing state to speed up training and sampling for diffusion text generation.
- **Long-context modeling**: contributed to scaling/evaluation for long-context LMs (data processing, chunking strategies, evaluation harnesses), with takeaways later reused by collaborators.

Selected Publications * equal contribution

Shansan Gong, Ruixiang Zhang, Huangjie Zheng, Jiatao Gu, Navdeep Jaitly, Lingpeng Kong, Yizhe Zhang, "DiffuCoder: Understanding and Improving Masked Diffusion Models for Code Generation". arXiv, 2025.   [paper] [code] (*700+* ★)

Shansan Gong*, Shivam Agarwal*, Yizhe Zhang, Jiacheng Ye, Lin Zheng *et al.*, "Scaling Diffusion Language Models via Adaptation from Autoregressive Models". ICLR, 2025.   [paper] [code] (*300+* ★)

Jiacheng Ye, Jiahui Gao, **Shansan Gong**, Lin Zheng *et al.*, "Beyond Autoregression: Discrete Diffusion for Complex Reasoning and Planning". ICLR, 2025.   [paper] [code] (*70+* ★)

Jiacheng Ye*, **Shansan Gong***, Liheng Chen*, Lin Zheng *et al.*, "Diffusion of Thoughts: Chain-of-Thought Reasoning in Diffusion Language Models". NeurIPS, 2024.   [paper] [code] (*170+* ★)

Chenxin An, **Shansan Gong**, Ming Zhong *et al.*, "L-Eval: Instituting Standardized Evaluation for Long Context Language Models". ACL, **Outstanding Paper**, 2024.   [paper] [code] (*380+* ★)

Shansan Gong, Mukai Li, Jiangtao Feng, Zhiyong Wu, Lingpeng Kong, "DiffuSeq-v2: Bridging Discrete and Continuous Text Spaces for Accelerated Seq2Seq Diffusion Models". EMNLP Findings, 2023.   [paper] [code]

Shansan Gong, Mukai Li, Jiangtao Feng, Zhiyong Wu, Lingpeng Kong, "DiffuSeq: Sequence to Sequence Text Generation With Diffusion Models". ICLR, 2023.   [paper] [code] (*800+* ★)

Honors & Awards

HKU Presidential PhD Scholarship (HKU-PS) | ACL 2024 Outstanding Paper | SIGIR 2022 Student Travel Award

Service

Conference Reviewer: ACL; EMNLP; NeurIPS; ICLR; ICML; ARR; COLM
Open Source: Primary maintainer/contributor on DiffuSeq, DiffuLLaMA, DiffuCoder repositories