KE LIN

✉ leonard.keilin@gmail.com   ⚙ leonardodalinky   in LinkedIn   ᴿᴳ Google Scholar   ⌂ Homepage

Education

College of William & Mary	**Sep. 2025 – Jun. 2030 (Expected)**
Ph.D. in Data Science	*Williamsburg, Virginia, US*
Advisor: Prof. Qingyun Wang	

Tsinghua University	**Sep. 2022 – Jun. 2025**
M.Eng. in Software Engineering (with Honors)	*Beijing, China*

Tsinghua University	**Sep. 2018 – Jun. 2022**
Bachelor in Software Engineering	*Beijing, China*

Experience

Tencent	**Apr. 2024 – Sep. 2024**
AI Research Intern	*Shenzhen, China*

- Developed a VLLM-based pipeline for evaluating the realism, reliability, and aesthetics of generated 3D scenes.
- Built a 2D rendering framework using Blender for generating views of 3D objects as inputs to VLLMs.
- Optimized the LLM agents to generate 3D layouts for the compositions of pre-defined 3D assets.
- Investigated the spatial reasoning capability of LLMs to produce precise 3D layouts utilizing arithmetic coordinates.

Momenta	**Jan. 2021 – Apr. 2021**
Backend Intern	*Beijing, China*

- Developed an automated service for scheduling autonomous driving model training tasks based on Kubernetes.
- Utilized Golang to reduce redundant resource consumption and estimated the approximate cost of training sessions.
- Processed the runtime logs of training tasks and stored them in AWS Cloud Storage for visualization.
- Automated the deployment of driving models on AWS and the updating of the K8s image from the upstream repository.

Publications

1. **Ke Lin**, Yilin Lu, Shreyas Bhat, Xuehang Guo, Junier Oliva, Qingyun Wang. "SciDER: Scientific Data-centric End-to-end Researcher." *Preprint.* arXiV:2603.01421.
   - Developed a data-centric multi-agent system that automates the full research lifecycle from ideation to experiments.
   - Designed a self-evolving memory mechanism using RAG to enable continuous test-time learning and memorization.
   - Achieved SOTA performance on MLE-Bench, AI-Idea-Bench, and SciCode benchmarks for research tasks.

2. **Ke Lin**, Yiyang Luo, Zijian Zhang and Ping Luo. "Zero-shot Generative Linguistic Steganography." In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics* (**NAACL 2024**).
   - Presented a zero-shot approach for linguistic steganography based on in-context learning using samples of covertexts.
   - Improved both the binary coding process and the embedding process by differential coding and annealing penalty.
   - Designed several metrics and language evaluations to evaluate both the perceptual and statistical imperceptibility.

3. **Ke Lin**, Yasir Glani, and Ping Luo. "Low-Latency Privacy-Preserving Deep Learning Design via Secure MPC." In *Proceedings of the IJCAI-24 Workshop on Artificial Intelligence Safety* (**IJCAI-AISafety 2024**).

4. Yiyang Luo*, **Ke Lin***, Chao Gu*, Jiahui Hou, Lijie Wen and Ping Luo. "Lost in Overlap: Exploring Logit-based Watermark Collision in LLMs." In *Findings of the Association for Computational Linguistics*: **NAACL 2025**.
   - Proposed the concept of watermark collisions, where multiple watermarks are present simultaneously in the same text.
   - Analyzed the potential risks and the vulnerability of existing watermarking techniques.

5. Yiyang Luo*, **Ke Lin*** and Chao Gu*. "Context-Aware Indoor Point Cloud Object Generation through User Instructions." In *Proceedings of the 32nd ACM International Conference on Multimedia* (**ACM MM'24**).
   - Designed a GPT-aided data pipeline for paraphrasing the descriptive texts in the ReferIt3D dataset to generative ones.
   - Proposed an end-to-end multi-modal diffusion model for generating indoor 3D objects into specific scenes.
   - Introduced the visual grounding task to assess the quality of an augmented scene, along with other metrics.

6. Chao Gu, **Ke Lin**, Yiyang Luo, Jiahui Hou, and Xiang-Yang Li. "ViRED: Prediction of Visual Relations in Engineering Drawings." In *Proceedings of the 2024 19th Int. Conf. on Mobility, Sensing and Networking* (**IEEE MSN 2024**).
   - Developed a relation prediction pipeline to differentiate between the circuit and tabular parts in engineering drawings.
   - Achieved a 10% enhancement in detection accuracy compared to prior studies on the electrical diagram dataset.

7. Glani Yasir, Ping Luo and **Ke Lin**. "MDASC: Advanced Dual-Layered Code Cloning Technique for Identifying Reused Malicious Code." *International Conference on Software Engineering and Knowledge Engineering* (2024).

8. <u>**Ke Lin**</u> and Ping Luo. "Skipping Scheme for Gate-hiding Garbled Circuits." arXiv preprint arXiv:2312.02514 (2023).

9. Glani Yasir, Ping Luo, <u>**Ke Lin**</u>, et al. "AyatDroid: A Lightweight Code Cloning Technique Using Different Static Features." *2023 IEEE 3rd International Conference on Software Engineering and Artificial Intelligence* (SEAI). IEEE, 2023.

10. Yue-Xun Huang, Ming Li, <u>**Ke Lin**</u>, et al. "Classical-to-quantum transition in multimode nonlinear systems with strong photon-photon coupling." **Physical Review A** 105.4 (2022): 043707.

*Equal contribution

Projects

Farthest Point Sampling Library | *Python, Rust, C++* **Sep. 2023**
- Developed a high-performance farthest point sampling library `fpsample` and `pytorch_fpsample` for Numpy arrays.
- Achieved 100× faster than vanilla implementation in pure Numpy for simplified preprocessing of 3D point clouds.
- Published PyPI packages and earned over **200 GitHub stars** for this open-source project.

Technical Skills

Languages: Mandarin (native), Cantonese (native), English (fluent) | Python, Golang, Rust, C++, Java, ReactJS, SQL
Tools/Frameworks: PyTorch, Transformers, LangGraph | Blender, Ubuntu, ArchLinux, PostgreSQL

Awards & Grants

- 2025 Google Cloud Research Grant
- 2025 Lambda Research Grant
- 2025 THSS Outstanding Graduate with Honors
- 2024 Tsinghua-Ubiquant Scholarship