

# YAO XIAO

(+86) 186-2182-3612 | ✉ [yaoxiao@g.harvard.edu](mailto:yaoxiao@g.harvard.edu) | 🏠 [charlie-xiao.github.io](https://charlie-xiao.github.io) | 🌐 [Charlie-XIAO](#) | 📄 [yao-xiao-200073244](#)

## EDUCATION

**Harvard University** | Master of Science | Computational Science and Engineering 2024.09 – present

- **GPA: NA/NA**, ongoing...

**New York University Shanghai** | Bachelor of Science | Honors Mathematics | Computer Science 2020.09 – 2024.05

- **Honors Mathematics GPA: 4.00/4.00**, including: Linear Algebra, Math Modeling, Probability Theory, Numerical Analysis, etc.
- **Computer Science GPA: 3.97/4.00**, including: Algorithms, Operating Systems, Computer Networking, Software Engineering, etc.

## PUBLICATIONS

AUTHORS WITH <sup>†</sup> ARE SORTED BY  $\alpha$ - $\beta$  ORDER, OTHERS ARE SORTED BY CONTRIBUTION

- [1] Yuejie Wang, Qitong Men, **Yao Xiao**, Yongting Chen, and Guyue Liu. 2024. ConfMask: Enabling Privacy-Preserving Configuration Sharing via Anonymization. In *Proceedings of the ACM SIGCOMM 2024 Conference (ACM SIGCOMM'24)*. Association for Computing Machinery, New York, NY, USA, 465–483. [doi:10.1145/3651890.3672217](https://doi.org/10.1145/3651890.3672217)
- [2] Shengbin Yue, Shujun Liu, Yuxuan Zhou, Chenchen Shen, Siyuan Wang, **Yao Xiao**, Bingxuan Li, Yun Song, Xiaoyu Shen, Wei Chen, Xuanjing Huang, and Zhongyu Wei. 2024. LawLLM: Intelligent Legal System with Legal Reasoning and Verifiable Retrieval. In *29th International Conference on Database Systems for Advanced Applications (DASFAA'24)*. [doi:10.48550/arXiv.2309.11325](https://doi.org/10.48550/arXiv.2309.11325)
- [3] Xinyu Li<sup>†</sup>, **Yao Xiao**<sup>†</sup>, and Yuchen Zhou<sup>†</sup>. 2023. Efficiently Visualizing Large Graphs. [doi:10.48550/arXiv.2310.11186](https://doi.org/10.48550/arXiv.2310.11186)

## RESEARCH EXPERIENCE

**Privacy-Preserving Network Configuration Sharing via Anonymization** | [SIGCOMM'24](#) | [GitHub](#) 2022.10 – 2024.08

ADVISOR: Professor Guyue Liu, [guyue.liu@gmail.com](mailto:guyue.liu@gmail.com)

- Proposed the ConfMask framework to systematically anonymize topology and routing information in network configurations.
- Designed the anonymization algorithm for different protocols that mitigated deanonymization risks yet preserved important utilities.
- Managed to rigorously prove the route equivalence and routing utility preservation properties of the anonymization framework.
- Led the implementation of the end-to-end network configuration anonymization system and the artifact evaluation.

**Efficient Distributed Serving System for Large Language Model Inference** | Capstone 2023.09 – 2024.01

ADVISOR: Professor Guyue Liu, [guyue.liu@gmail.com](mailto:guyue.liu@gmail.com)

- Enabled larger batch sizes beyond KV cache limit for layers except self-attention, observing that only self-attention relies on KV cache.
- Batched prefills and decodes dynamically in self-attention to mitigate pipeline bubbles caused by varying transformer input lengths.
- Packed multiple short attention computations with the longest one, while concurrently swapping KV cache to minimize overhead.

**Efficiently Visualizing Large Graphs** | Dean's Undergraduate Research Fund | [ArXiv](#) 2022.05 – 2022.08

ADVISOR: Professor Jie Xue, [jiexue@nyu.edu](mailto:jiexue@nyu.edu)

- Designed t-SGNE specialized for graphs, leveraging the neighboring relations between nodes and achieving 6.7x computation efficiency.
- Proposed SPLEE, a graph embedding method based on Laplacian eigenmaps and shortest paths, intended to suit t-SGNE.
- Combined SPLEE and t-SGNE for visualization of graphs with 300K nodes and 1M edges, achieving 10% improvement in visual effect.

## WORKING EXPERIENCE

**Scikit-learn** | Open Source | Core Developer | [GitHub](#) 2023.04 – present

SKILLS: Python, Cython, JavaScript, Sphinx, scikit-learn, numpy, scipy, pandas, polars

- Participated in project maintenance, e.g., test suite coverage, code refactoring, robust validation, automated GitHub workflows, etc.
- Coordinated efforts in sparse array support, estimator representation, and website-related tasks.
- Led the redesign the entire scikit-learn main website with various documentation improvements and UI/UX enhancements.
- Contributed [118 merged pull requests](#) including bug fixes, performance improvements, new features, and documentation.
- Triaged issues, reviewed pull requests, and guided new contributors towards the project as a project maintainer.

**DISC Lab, Fudan University** | Lab Assistant | [DASFAA'24](#) | [GitHub](#) 2023.05 – 2023.08

SKILLS: Python, PyTorch, HuggingFace, LLM, instruction tuning

- Led the construction of 403K legal knowledge instruction data, curated with legal syllogism prompting for higher expertise.
- Fine-tuning DISC-LawLLM, a large language model specialized for intelligent legal services based on Baichuan 13B Chat.
- Participated in designing a verifiable knowledge retrieval module to inject external knowledge and enhance output actuality.
- Drove the implementation of a comprehensive benchmark for legal systems evaluation in both objective and subjective dimensions.

TEACHING EXPERIENCE

---

- **Linear Algebra**, MATH-SHU.0140, NYU Shanghai, Learning Assistant, Spring 2024
- **Calculus II**, MATH-SHU.0131, NYU Shanghai, Learning Assistant, Fall 2021, Fall 2023
- **Operating Systems**, CSCI-UA.0202, New York University, Teaching Assistant, Spring 2023

PROJECTS

---

**Deskulpt: A Cross-Platform Desktop Customization Tool** | [GitHub](#) 2024.03 – present

SKILLS: Rust, TypeScript, Tauri, React, Vite, SWC

- 
- 

**CampusHelper: WeChat / Alipay Miniprogram** 2023.12 – present

SKILLS: TypeScript, Alipay miniprogram cloud, WeChat cloud, miniprogram frameworks

- 
- 

**YouTube Interface Customizer** | Course Project | [GitHub](#) 2023.02

- Built a Firefox extension that supports changing color themes, rearranging, and customizing elements of the YouTube interface.
- Created the documentations of features and contribution guides, and released (self-distributed) v1.0 at Mozilla Add-ons.

**Inequality Process Simulation** | Course Project | [Paper](#) | [GitHub](#) 2022.12

- Simulated inequality process in economic systems via nuanced random transactions functions, reflecting on real-world economy.
- Discovered that the final distribution of wealth in a real-world economic system fits the shape of a gamma or beta prime distribution.

**Gyro-Tower Simulation** | Course Project | [Paper](#) | [GitHub](#) 2022.10

- Modeled gyroscopes as networks of springs, formulated the system with differential equations, and solved it via Euler’s method.
- Simulated vertical stacks of gyroscopes, and found that they obeyed gyroscopic precession assuming a flexible middle axle.

HONORS AND AWARDS

---

- [1] **Magma cum laude**, NYU Shanghai, 2024
- [2] **Level I Certification**, [CRLA’s International Tutor Training Program](#), 2024
- [3] **2nd Prize, 4th Place**, [Alipay Miniprogram Developers’ Competition](#), 2023
- [4] **Meritorious Winner**, [Mathematical Contest in Modeling](#), 2023
- [5] **Dean’s List of Academic Year**, NYU Shanghai, 2020 – 2021, 2021 – 2022, 2022 – 2023

SKILLS

---

- [1] **Programming:** Proficient in Python, Rust, C, TypeScript; Intermediate in Java, C++, Julia, MATLAB
- [2] **Frameworks and packages:** Tauri; React; Numpy, Scipy, Pandas, Polars, Scikit-learn, Matplotlib, Seaborn, PyTorch
- [3] **DevOps:** CircleCI, Github Actions; Docker; Git/GitHub; AWS, WeChat / Alipay Cloud; Computer networks; Linux
- [4] **Languages:** English (fluent); Mandarin Chinese (native)