

# 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](https://yao-xiao-200073244)

## EDUCATION

**Harvard University** | Master of Science | Computational Science and Engineering 2024.09 – 2026.05 (expected)

- GPA: 3.92/4.00, including: Computer Networks, HPC, Data Systems, Distributed Systems (MIT), etc.

**New York University** | Bachelor of Science | Honors Mathematics | Computer Science Shanghai | New York | 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: Data Structures, Algorithms, Operating Systems, 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
- [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:10.1007/978-981-97-5569-1\_19

## TEACHING EXPERIENCE

- Computer Networks**, COMPSCI.1450, Harvard School of Engineering and Applied Sciences, Teaching Fellow, Spring 2025
- 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

## WORKING EXPERIENCE

**Scikit-learn** | Open Source (GitHub 60K Star) | Core Developer | [128 Merged Pull Requests](#) 2023.04 – present

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

- Managed maintenance tasks e.g., test suite coverage, code refactoring, developer API improvement, automated GitHub workflows, etc.
- Enhanced sparse array and polars dataframe support, estimator representation, metrics visualization, multilabel data cross-validator, etc.
- Optimized IncrementalPCA on sparse data (>10x speedup, <3% memory usage), SPD matrix generator (<10% memory usage), etc.
- Led the redesign the entire scikit-learn main website and coordinated efforts in documentation improvements and UI / UX enhancements.

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

2023.05 – 2023.08

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

- Led the construction of 403K legal knowledge instruction data, curated with legal syllogism prompting for higher expertise.
- Fine-tuned DISC-LawLLM, an LLM specialized for legal services based on Baichuan 13B Chat, outperforming GPT-3.5 Turbo.
- 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.

## 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.

## HONORS AND AWARDS

- [1] **Magma cum laude**, NYU Shanghai, 2024
- [2] **Level I Certification**, [CRLA's International Tutor Training Program](#), 2024
- [3] **Dean's List of Academic Year**, NYU Shanghai, 2020 – 2021, 2021 – 2022, 2022 – 2023

## SKILLS

- [1] **Programming**: Python, Rust, JavaScript/TypeScript, C, C++, Go; SQL, Java, MATLAB, Julia
- [2] **Frameworks and packages**: Tauri, React; Numpy, Pandas, Polars, Scikit-learn, PyTorch; CUDA; SIMD/AVX; OpenMP, MPI
- [3] **DevOps**: Docker; Git; AWS, Google Cloud; Ansible; Kubernetes; GitHub Actions, CI/CD; Linux