

# XU SHUYAO

✉ shuyao@u.nus.edu · 🌐 <https://github.com/Tim-Siu> · 🔗 <https://tim-siu.github.io>

## 🎓 EDUCATION

---

**National University of Singapore (NUS), Singapore** Aug 2022 – Present

*Bachelor of Computing in Computer Science (with Honours)*

Second major in Statistics **GPA: 4.8/5.0**

**Stanford University, USA**

Jun 2023 – Aug 2023

*Summer Quarter*

International Honors Program **GPA: 4.15/4.0**

## 👥 PROJECT EXPERIENCE

---

**MapleCG Research Group @ NUS**

February 2024 – Present

*Undergraduate Researcher* Advisor: Professor Bryan Low

- Conducted research on data synthesis and model distillation for LLMs. Responsible for implementing the distillation and model performance evaluation pipeline.
- Proposed improved methods based on existing academic approaches. Deployed algorithms using PyTorch, designed and executed a series of experiments to validate the performance of distilled smaller models. Small LLMs fine-tuned with synthetic data showed significant accuracy improvements on the CQA dataset. Ablation studies are ongoing.

**MarkBind** [GitHub Project Link](#)

January 2024 – Present

*Mentor / Developer* Advisor: Professor Damith C. Rajapakse

- MarkBind is a static site generator based on Markdown syntax. In maintenance, it requires new features to enhance user experience. I am responsible for enhancing table filtering and sorting, mermaid diagram support, global search support, and automatic line wrapping for code blocks.
- Wrote over 8k lines of code using TypeScript, Node.js, and Vue, and utilized tools like Webpack, ESLint, Jest for development. Implemented continuous integration and deployment through GitHub, improving team collaboration and code review. Successfully added new features, significantly enhancing the user experience of MarkBind, and received positive feedback from users.

## ⚙️ PROFESSIONAL SKILLS

---

- **Programming:** Python, Java, C. Familiar with LaTeX and Shell Scripting.
- **Machine Learning Theory and Practice:** Proficient in statistical learning (linear models, kernel methods, convex optimization, PAC learning...). Familiar with recent deep learning models (VAEs, BERTs, GPTs, ViTs...) and training practices. Familiar with PyTorch, HuggingFace Transformers...
- **NLP/LLM:** Familiar with regex, tokenizers, architectures of GPTs, full pipeline of pretraining, finetuning, alignment, agent prompting, evaluation. Proficient in model compressing via pruning and distillation.

## 📖 COURSES

---

- Deep Learning Specialization (Coursera), 2023
- **NUS Courses** (incl. ongoing courses): **Graduate level:** CS4243 Computer Vision, ST4234 Bayesian Statistics, MA4270 Data Modelling and Computation, **Undergraduate level:** Probability and Statistics (A+), Discrete Structures (A+), Calculus for Computing (A+), Programming Methodology (A+), Database Systems (A+), Data Structure and Algorithms (A), Numerical Analysis (A)...
- **Stanford Courses:** CS148: Introduction to Computer Graphics and Imaging (A+), CS221: Artificial Intelligence: Principles and Techniques (A)