

Yash Jain

yash012@e.ntu.edu.sg | [LinkedIn](#) | yashjain14.github.io

EDUCATION

Nanyang Technological University, Singapore

Bachelor of Engineering in Computer Science

(Double Bachelors)

Bachelor of Business (specialization in Business Analytics)

Jul 2022 - Jul 2026

EXPERIENCE

MMLab (S-Lab), Nanyang Technological University

Singapore

AI Research Intern

May 2024 – Nov 2024

- Worked on Large Multimodal Models and evaluation visual generation, focusing on Image, Video, and Motion.
- Implemented Structured Generation Inference for LLaVa-NEXT using SGLang.
- Maintained Awesome-Evaluation-of-Visual-Generation repository and led to 180+ stars on GitHub.

TUV SUD

Singapore

AI Intern

May 2023 – Jul 2023

- Led market research of AI certification in APAC region, delivering strategic insights for business expansion.
- Developed proficiency in AI Governance and AI life cycle, and learned about AI regulations (including EU AI Act).
- Improved AI certification course curriculum and training materials, focusing on practical industry applications.

PROJECTS

[More on GitHub](#)

SOPPU – Scalable One PEFT per User (Paper Accepted to DAI 2024)

- Decentralized LLM inference for scalable personalized AI using compressed PEFT (LoRA) adapters.
- Built with Hugging Face Transformers and PEFT library, and integrated LoRAX for dynamic, scalable serving of personalized adapters at scale.
- Achieved multi-adapter joint compression using diagonalization, with 1.99x ratio and 49.7% memory savings.

StoryLine AI

- Long-form video generation platform that creates script-consistent videos using Stable Video Diffusion and Flux.
- Implemented backend with PyTorch, Diffusers, NLTK and Gradio for UI. Uses AWS CloudFront for CDN.
- Secured a \$3,000 grant from NTU's Innovation Lab. Achieved faster video generation through parallel processing.

Weaver

- AI powered Textile designing tool for printing, weaving, and knitting designs for textile manufacturers in India.
- Developed using NextJS, ReactJS, Typescript, Tailwind CSS, MongoDB, Clerk Auth, AWS S3, & Stripe Payments.
- Secured a grant of \$10,000 from NTUitive. Reduced manufacturers' design process from days to minutes.

GitAI

- CLI-native AI agent for executing Git related tasks using natural language, tailored for beginners to Git and GitHub.
- Finetuned Qwen2 and Llama 3.2 on Git documentation and Stack Overflow data using MLX LoRA implementation.
- Implemented to run locally on CPUs, achieving speeds up to 100 tokens/second (MacBook Air M1 16GB).

KamaGPT

- Full-Stack multilingual RAG chatbot for sex education targeting youth in Asia Pacific region.
- Built with Groq Inference, LangChain, ChromaDB, Sentence-Transformers and SQL Database. Frontend using Python Flask, HTML, JavaScript, and Tailwind CSS.
- Secured \$10,000 NTU SIE grant and scaled to 22,000 users with 100,000+ answered queries.

CO-CURRICULAR ACTIVITIES

Chairperson – Open Source Society

- Hosted workshops, managed external speakers, and increased students' workshop participation rate.

President (Student Lead) - Google Developer Student Club

- Led a team of 150 members and organized workshops, and hackathons, boosting student engagement.

AWARDS & ACHIEVEMENTS

- **Global Top 50 (Semi-finalist):** Alibaba Global E-commerce Challenge, 2024
- **NTUitive MDT Grant:** Awarded Multidisciplinary Team Grant of \$10,000 SGD, 2024
- **Top 10 Teams (Finalist):** Kumar Sustainability and Innovation Challenge, 2024
- **NTU SIE Program:** Awarded Kwek Leng Joo Colab4Good Grant of \$10,000, 2023
- **NTU SEP Program:** Awarded SEP Grant of \$3,000 SGD by NTU's I-Lab, 2023
- **1st Runner Up:** YouthxHack Total Defence Hackathon, 2023
- **500k+ Views on YouTube:** Videos on Tech, vlogs, and higher education in Singapore.

SKILLS

- Languages: Python, C/C++, Java, R, JavaScript, TypeScript, SQL, NoSQL, HTML/CSS
- Frameworks/Libraries: React, Node.js, NextJS, Flask, ShadCn UI, Tailwind CSS, Bootstrap, Hugging Face, Transformers, Diffusers, LangChain, PyTorch, TensorFlow, MLX, FastAPI, Pandas, NumPy, Scikit-learn, Matplotlib
- Technologies: AWS, GCP, Git/GitHub, Docker, MongoDB, MySQL, PostgreSQL, Unity, MS Office, G-Suite, Linux