

Yash Jain

yash012@e.ntu.edu.sg | [LinkedIn](#) | [GitHub](#) | yashjain14.github.io | +65 87903463

EDUCATION

Nanyang Technological University, Singapore

Bachelor of Engineering in Computer Science (specialization in Artificial Intelligence)

Jul 2022 - Jul 2026

Bachelor of Business (specialization in Business Analytics)

(Double Bachelors)

WORK EXPERIENCE

Mastercard

Singapore

AI Research Intern (veNTUre x Mastercard)

March 2025 – Present

- Developed an Edge AI kiosk system to run local AI models on commodity hardware for customer segmentation and targeted ad serving.
- Utilized JavaScript, HTML/CSS, Python, and small vision models (Moondream2, Smol-VLM) for segmentation.
- Delivered a functional prototype featuring ad-serving, face recognition for user journey tracking, model benchmarking, and a research report on low-cost AI solutions.

Seatrium

Singapore

Full Stack Developer Intern (NTU Peak)

Jan 2025 – March 2025

- Engineered an online platform and PCF chatbot to help Seatrium's suppliers to report Scope 3 emissions data.
- Developed using React/Next.js (TypeScript), integrating SSO authentication and PostgreSQL for database
- Spearheaded the transition from price-based to weight-based accounting, enhancing emissions reporting granularity and overall data reliability.

MMLab (S-Lab, NTU)

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 250+ 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.
- Acquired 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](#)

Open Context

- Built an AI agent inspired by Open AI's Deep Research which searches web and local files to generate reports and PowerPoint presentations.
- Utilized React, Next.js, Python, pptx library for generating PPT, Tavily for web search and RAG for file search.
- Optimized performance through multithreading to execute parallel searches and faster data report generation.

SOPPU: Scalable One PEFT per User (Paper accepted to DAI 2024)

- Decentralized LLM inference server for personalized AI on scale via compressed PEFT (LoRA) adapters.
- Built with Transformers and PEFT library, and integrated LoRAX for dynamic serving of LoRA adapters at scale.
- Achieved multi-adapter joint compression via diagonalization, with 2x ratio and 49.7% memory savings.

StoryLine AI

- Long-form video generator with 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

- AI powered Textile designing tool for printing, weaving and mockups for textile manufacturers in India.
- Developed using Next.js, React, Typescript, MongoDB, Clerk, AWS S3, Stripe Payments and CI/CD with Docker.
- Secured a grant of \$10,000 from NTUitive. Reduced manufacturers' design process from days to minutes.

AWARDS & ACHIEVEMENTS

- Global Top 50:** Alibaba Global E-commerce Challenge, 2024
- NTUitive MDT Grant:** Multidisciplinary Team Grant of \$10,000 SGD, 2024
- NTU SIE Program:** Kwek Leng Joo Colab4Good Grant of \$10,000, 2023
- NTU SEP Program:** SEP Grant of \$3,000 SGD by NTU's I-Lab, 2023
- 1st Runner Up:** YouthxHack Total Defence Hackathon, 2023
- 600k+ 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, PyTorch, FastAPI, GraphQL
- Technologies: AWS, Google cloud Platform, Azure, Git/GitHub, Docker, MongoDB, PostgreSQL, Linux, Docker