Kevin Pek Yue Ting

Singapore Citizen | kevin_pyt@u.nus.edu | github.com/kevin-pek | kevinpek.com | Graduating Nov 2024

EXPERIENCE

GovTech – Data Scientist Intern | *Singapore* May 2024 – Oct 2024
- Applied few shot learning techniques to CLIP models, and trained CV/NLP models for document classification task.
- Setup Docker image for training on AWS Sagemaker, profiled and optimized GPU utilization for PyTorch training.
- Wrote various data processing and automation scripts using Bash and Python to automate existing workflows and tests.
- Made performance optimisations and bug fixes to speed up inference on deployed machine learning API.

NUS IORA/NCS – Research Assistant | *Singapore* Jun 2023 – Dec 2023
- Exploratory data analysis on time series data in Python (pandas, plotly) in Azure ML Linux server environment.
- Used the diffusion equation and finite volume method to visualise temperature distribution in data centers.
- Trained PyTorch forecasting model and setup batch inference endpoints on Azure ML.

NUS IORA/SIA – Research Assistant (Full Stack Dev) | *Singapore* Aug 2022 – May 2023
- Designed and developed a machine learning powered quiz webapp. Features include dashboards, user management features and password-based login using JWTs. Stack: React (Material UI), Express (Typescript), MySQL, AWS S3.
- Implemented multi-arm bandit algorithm (epsilon greedy) and developed a question recommender API using FastAPI.
- Handled deployment of entire application stack with Nginx reverse proxy on AWS and on-premises using Docker.
- Produced OpenAPI specification for various API services and produced technical documentation for handover.
- Converted Python scripts implementing multi-arm bandit algorithms into Flask webapp and deployed it on Heroku.

Idemia – Full Stack Developer Intern | *Singapore* May 2022 – Jul 2022
- Developed proof-of-concept for an orchestrator application in Java using the Apache Camel framework and Spring Boot.
- Wrote integration tests using JUnit5 and Testcontainers, and fixed bugs for various applications.

NUS Department of Biomedical Engineering – Game Developer | *Singapore* Feb 2022 – May 2022
- Developed a quiz game in Unity to assess the cognitive capabilities of people suffering from Alzheimer's Disease.

DSTA – Modelling and Simulations Intern | *Singapore* Mar 2021 – Jun 2021
- Built high fidelity drone flight simulation environment in Unity (C#) for training reinforcement learning models.
- Trained reinforcement learning agents using Unity ML Agents for multi-stage AI system.

EDUCATION & AWARDS

National University of Singapore, B.Sc Data Science, 2nd Major Computer Science Aug 2021 – Nov 2024
- Expected First Class Honors. Awarded Dean's List for AY2022/23 Semester 1.
- **Courses:** DS&A (A+), OOP, Machine Learning, Operating Systems (A+), NLP (A+), Info. Retrieval, Parallel Computing.

Stanford University, International Honors Program Jun 2023 – Aug 2023
- Awarded the SG-US Summer Exchange Scholarship.
- **Courses:** CS161 Design and Analysis of Algorithms, CS221 Artificial Intelligence: Principles and Techniques.

SKILLS

- **Programming Languages:** Python, Typescript, Javascript, C, Java, Swift, R, C#
- **Web Development:** NodeJS, React, Express, NextJs, FastAPI, Flask, Spring Boot, HTML, CSS
- **Data Science:** pandas, numpy, pytorch, scikit-learn, huggingface transformers, langchain
- **Cloud/Tools:** Git, Docker, Linux, Bash, AWS, Google Cloud Platforms, Azure ML
- **Databases:** PostgreSQL, MySQL, MongoDB, SQLite, Neo4j

PROJECTS

bitnet.c – *Zero dependency, C implementation of the BitNet paper*
- Implementation of neural network with single bit weights based on the BitNet research paper. Github repo.
- Implemented inference for Byte-Pair Encoding tokenisers trained with tiktoken, and neural network training on MNIST.
- Parallelisation of bit matrix multiplication and neural networks layers using OpenMP to speed up training time.

PDFSearch – *Open-source extension for the MacOS app Raycast*
- Author of app extension allowing users to search collections of PDF and text files, more than 400 users. Github repo.
- Written in Typescript (React, NodeJS) and native MacOS Swift packages for document parsing and similarity search.
- Used multithreading, NodeJS stream API, and file-based IPC to improve search latency for large collections of documents.

http-rs Aug 2024
- Multithreaded HTTP server written while learning the Rust programming language. [Github repo](#).

Shiftly Staff Scheduler – *Automated staff scheduling for f&b restaurants* Mar 2024 – Apr 2024
- Done as coursework for DSA3101 Data Science in Practice. [Github repo](#).
- Designed and developed webapp dashboard and interactive staff schedule planner for restaurant managers. Features include demand forecasting, staff and booking management system, and an automated staff schedule generator.

AcademIQ – *Knowledge graph creation from documents* Mar 2024 – Apr 2024
- Done as coursework for DSA4213 Natural Language Processing. [Github repo](#).
- H2OWave app that uses RAG to extract and build topic graphs from uploaded PDF documents and stores it within a Neo4j database. User can interact with the graph to view summaries, and generate questions based on selected topics.

Toxic Speech Detection on Twitter Dataset Mar 2024 – Apr 2024
- Done as coursework for CS3244 Machine Learning.
- Finetuned LLaMA2-7B using QLoRA and RoBERTA models on text classification tasks Google Cloud GPU.

Pix2Latex – *MacOS menubar app to run both normal and LaTeX OCR* Nov 2023
- Built in Python. Either runs MacOS OCR engine or pretrained OCR models on selected region of screen. [Github repo](#).

Document Semantic Search – *Hybrid search on PDF documents using embedding models* Jan 2023
- Gradio interface to search documents using FAISS index for semantic similarity and BM25 keyword search. [Github repo](#).

EXTRACURRICULARS

NUS Google Developer Student Club – Tech Lead Jul 2023 – Jan 2024
- Led a team of students to build a peer-to-peer sharing platform. Stack: Typescript, Express, NextJs, React, shadcn/ui.
- Setup CICD for testing and deployment. Troubleshooting for deployed services on Vercel and fly.io cloud environment.

NUS Google Developer Student Club – Software Engineer Aug 2022 – May 2023
- Full stack development, CICD setup using GCP Cloud Run. Stack: Typescript, Express, MongoDB, NextJs (React, MUI).

NUS Student Union IT Committee, CharITeach – Volunteer Tutor Mar 2022
- Taught underprivileged kids basic programming by teaching them to make games in the Scratch programming language.

42nd NUS Rovers Adventure Club – Honorary General Secretary Oct 2021 – Oct 2022
- Managed database of 2000+ members, planned and led 8 local and overseas trekking/cycling events with 50+ participants.

42nd NUS Science Club Alumni Relations Committee – Marketing Executive Oct 2021 – Oct 2022
- Prepared prospectus and reached out to organisations for sponsorships for the committee's events.