Eric Zhu

0412-137-895 | linkedin.com/in/ericfzhu | ericfzhu909@gmail.com | github.com/ericfzhu | ericfzhu.com | Randwick, Sydney

ABOUT

Result-driven software engineer and artist with expertise in using Python, Java, and TypeScript to build large distributed systems for B2B applications. Currently pursuing a Master of IT at UNSW, specialising in Artificial Intelligence, and actively developing NotesCast, a platform utilising LLMs and STT models to enhance knowledge discovery and curation through podcasts. Also expanding knowledge through books and podcasts on technology, philosophy, design, and investment strategies.

Check out https://ericfzhu.com/resume for the complete resume. Able to defer or reduce workload on Master's program for a full-time position.

EDUCATION

University of New South Wales **Expected Nov 2025**
Master of Information Technology (Artificial Intelligence)
WAM: 81
Relevant courses: Computer Graphics, Computer Networks and Applications, Information Retrieval and Web Search

University of New South Wales **Nov 2022**
Bachelor of Science, Computer Science (Artificial Intelligence)
Relevant courses: Database Systems, Artificial Intelligence, Business Finance, Portfolio Management, Neural Networks & Deep Learning, Cryptocurrency and DLT, Computer Vision, Recommender Systems, Knowledge Representation

PROFESSIONAL EXPERIENCE

Amazon Web Services **Mar 2025 - Present**
Cloud Support Engineer I (SVO) [Fulltime]

- Providing support for customers building on SageMaker and Bedrock

NotesCast **Jul 2023 - Present**
Software Engineer [Founder]

- Automatic Speech Recognition STT models and LLMs are used alongside various prompting techniques, such as Chain-of-Density, to transform podcast audio files into summaries
- Generated data is embedded in sliding windows to power Alexandria, a Retrieval Augmented Generation agent that allows users to query for industry-specific knowledge, as well as a recommendation engine for users to discover related content
- User authentication is handled via Firebase, while the other server features including RAG is deployed to AWS EC2
- v0.1 launched in Dec 2023 with roughly 300 MAUs, and currently working on improving summary text quality and response accuracy for Alexandria for a v1 release
- Building a pipeline for speaker diarization and speaker-specific retrieval

National Australia Bank **Feb 2022 - Feb 2023**
Software Engineer Intern [Fulltime]

- Delivered robust features for the New Payments Platform using Java, Spring, and Jenkins on AWS in an Agile environment
- Led the implementation of PayID transaction processing feature after analyzing customer payment pain points, resulting in faster payment processing and projected revenue impact of 3-5M/year
- Explored and mastered new monitoring tools to better understand system performance, leading to the identification and resolution of transaction dropping issues during high-traffic periods
- Investigated and resolved a critical Kafka backpressure performance issue by diving deep into system metrics and logs, ensuring reliable payment processing for customers during peak loads of 5,000 transactions/second
- Enhanced CI/CD pipelines using Python, Jenkins, and Terraform, increasing QA efficiency by 8%
- Proactively learned and implemented multiple AWS services and Spring framework features to enhance the New Payments Platform, demonstrating adaptability in a complex financial technology stack

The University of Sydney **Nov 2021 - Mar 2022**
Research Engineer [Project-based]

- Developed a Fast Low-cost Online Semantic Segmentation machine learning model for detecting abnormal contextual changes in streaming data from sensors placed around a building
- Designed and implemented a research project website using Gatsby and Contentful CMS

MAPFRE Insurance **Nov 2019 - Feb 2020**
Software Engineer Intern [Fulltime]

- Collaborated with a senior engineer to develop a new mobile app with Vue.js and Node using REST APIs
- Engineered a new routing algorithm, which led to an 18% reduction in incident response time

TECHNICAL SKILLS

Languages**:**	Python, TypeScript, Java, JavaScript, HTML, CSS, C, SQL, Solidity, C++
Libraries**:**	React, Next.js, Gatsby, Vue.js, Framer Motion, Three.js, Spring, Selenium, React Native
AI/ML:	PyTorch, TensorFlow, LangChain, SciPy, CNN, GAN, LLM, TorToiSe, FLOSS
Databases**:**	PostgreSQL, PineconeDB, MySQL, MongoDB, SQLite, DynamoDB, Redis, OpenSearch
Cloud:	AWS, Firebase, Cloudflare, Vercel, Terraform, SAM, Docker, CloudFormation, CloudFront

RELEVANT WORKS

CODEX | https://github.com/ericfzhu/codex | https://codex.ericfzhu.com/ **2024**

- Convert text from books and articles into vector embeddings
- Exploring how to interact and visualise embedding maps

TypeScript, Next.js, Three.js, PineconeDB

TOOLBOX | github.com/ericfzhu/toolbox | https://toolbox.ericfzhu.com/ **2024**

- Create dot patterns, ASCII art, and Gaussian blurs from images
- Pick a primary colour and generate a palette from an image

TypeScript, Next.js

REPLICA | github.com/ericfzhu/replica | https://replica.ericfzhu.com/ **2023**

- Training ML models from scratch using original papers
- Completed AlexNet and VGG-16, currently working on ResNet50 and a ResNet-based GAN

Python, TypeScript, Next.js, PyTorch, GAN, CNN

"WEBSITE" | http://ericfzhu.com/ **2022**

- Personal website designed to mimic a macOS desktop
- Each app is a window into the different perspectives of website design
- Smooth animations powered by Framer Motion

TypeScript, Next.js, Framer Motion, Figma, Notion

CERTIFICATIONS AND AWARDS

AWS Certified Developer Associate	2024
AWS Certified Cloud Practitioner	2022
Microsoft Certified: Azure AI Fundamentals	2021

3rd place at Build Together: AI Hackathon 2024

- Guided: Sensor-based guided meditation and exercise

INTERESTS

Maintain an active learning mindset through continuous self-development, completing 33 books in 2024 primarily focused on technical skills, leadership, and business strategy.