

Jonathan Ng

jonathan.n@u.nus.edu | @derpy
H1B1 Eligible | +65 83995537

EDUCATION

NUS (NATIONAL UNIVERSITY OF SINGAPORE) CADENZA LABS | RESEARCH ENGINEER

BCOMP IN INFORMATION SECURITY
December 2021 | Singapore, SG

Apr 2023 – Present

- Developed the EleutherAI/elk library, worked on productionizing, coordinating and conducting large sweeps across multiple GPU clusters, conducted original experiments into language model probing, developing reproducible research by incorporating WandB.

LINKS

Github:// [derpyplops](#)
LinkedIn:// [Jonathan Ng](#)

COURSEWORK

UNDERGRADUATE

Web Application
Database Management
Penetration Testing + Practicum
Systems & Web Security + Practicum
Computer Networks II + Practicum
Programming Methodology
(Senior Teaching Asst.)
Unix Tools and Scripting

SKILLS

PROGRAMMING

Proficient:
Python • Scala • VueJS
JS • SQL • React

Familiar:

Java • Shell
Golang • C++ • C

TOOLS & FRAMEWORKS

Apollo GQL • ZIO • Play Framework
AWS • GRPC • Jest
AsciiDoc • Redux
express.js • Kafka • Docker

INTERESTS

Effective Altruism
Arete Fellowship (SG) Organizer (Fall '20, Spring '21)
NUS Chapter Lead

PROFESSIONAL EXPERIENCE

SERI MATS | RESEARCH ENGINEER

Jan 2023 – March 2023 | Singapore

- Participated in seminars and workshops relating to AI Alignment, mentored by Dr Dan Hendrycks.
- Authored (with others) the *Machiavelli* paper (ICML23), created a data collection platform on Streamlit (repo) (paper).
- Worked with the trlx library to train LMs on an ethics dataset via RLHF.

LEADIQ | SOFTWARE ENGINEER (FULL-STACK)

Apr 2020 – June 2022 | Singapore

- Built Self-Serve Single Sign-On, a feature that automates SSO client setup, and cuts a 2-3 week long process down to a day. Won internal hackathon 1st place. After winning, we brought the code and made it production ready in a month.
- Built Slack Bot delivering Job Change Notifications.
> *Kafka, Scala, ZIO, Slack API*
- Built Assign Sequences with Outreach.io API.
> *Vue, CSS, HTML, ApolloGQL, Scala*

STRONGSALT | SOFTWARE ENGINEER INTERN (BACKEND)

Jan 2020 – Apr 2020 | California, San Jose

- StrongSalt is an encryption platform, featuring the world's first always-on searchable encryption infrastructure.
- Wrote and deployed AWS S3 Component.
- Built and deployed API documentation from scratch with *Docusaurus*.
- Implemented the Go and NodeJS API endpoints with GRPC.

COURSES

MLSS | ML SAFETY COURSE

July - September 2022

- Course teaching ML and DL concepts, and current topics in AI Safety with focus on Systems Safety.

ARENA | ML ENGINEERING

October - December 2022

- Course imparting deep understanding about fundamentals systems such as DL, RL.
- Implemented architectures PPO, BERT, GPT (and more) from scratch.
- Hands on experience with the latest in mechanistic interpretability research.
- Contributed to TransformerLens, an open source library for interpretability research.