James Chua Shao Hwee

+ 65 9040 4373 | chuajamessh@gmail.com | **Github** https://github.com/thejaminator

Open source contributions

Conditionme (Creator)

- Conditionme is a library for easily retraining existing language models to work in a conditional /decision transformer offline reinforcement learning (RL) fashion. Primarily as a means to investigate safety benefits of conditional RL, it helps researchers easily specify a scalar target reward for finetuning.

Strawberry (Maintainer)

- Strawberry is a popular python package to create graphql services. Users can enhance their type safety by building their graphql schema from pydantic data models. I define the interface for pydantic conversions, making sure that changes are backward and forward compatible. I maintain the mypy plugin to catch potential errors. I review pull requests and answer developers' questions in using the library. See contributions here

Eleuther Elk (Contributor)

- Elk is a package for researchers to search for features in the hidden states of a language model. It searches for potential features that can elicit what the model truly knows. I contribute primarily by refactoring the codebase and making it testable, so that the codebase may progress rapidly in a maintainable manner. See contributions here

Slist (Creator)

- A drop in replacement for the built-in mutable python list But with more post-fixed methods in a typesafe manner. This allows developers to chain list operations confidently, avoiding runtime errors.

Typesafe Parmap (Creator)

- Run functions in parallel safely with python type checkers. This helps developers easily run functions in parallel.

LeadiQ

Senior Machine Learning engineer Jan 2020 - Current

Worked in the LeadIQ scribe team - We help salespeople generate great emails easily

- Labeling is expensive and time consuming. I created an active learning pipeline that sorted the feedback from users by entropy. This allowed us to prioritize labeling examples that are the most informative to our generative language model.
- Diversity in model generations is important to our product. I tackled entropy collapse in our machine learning feedback loop with active learning, various dataset sampling strategies, and entropy penalizers.
- Our product receives a large amount of preference feedback from users and labellers. I created streamlit apps for labellers to easier assign rewards. I added "Upside-down" / Decision Transformer style reinforcement learning to our training pipeline. This increased the quality of our generations and user satisfaction.

- Collecting data is very important for our machine learning pipeline. Designed APIs for our product such that user session information is easily saved in the database for future use.
- Python codebases often have runtime issues due to dynamic typing. I advocated and trained other engineers to write typechecked code with mypy, leading to less production bugs. I created mypy plugins to prevent production errors. An example of a plugin I created for strawberry, a popular graphql API.
- Led and coached the team in analytics of product success. Shared about statistical paradoxes, and cognitive biases for better decision making. An example of a lesson I gave.
- Conducted sharing sessions with the latest papers on Natural Language Processing.
- Trained transformer models for NLP generative tasks, taking into account user experience research.
- Built and maintained infrastructure for machine learning model serving in python. Fixed several issues regarding the python GIL and blocking I/O.
- Optimized redshift SQL queries to take advantage of columnar database optimisations for analytics.
- Worked using Scala and functional programming practices to build backend data pipelines and graphql API services.
- Conducted and designed the hiring of Machine Learning engineers, onboarded them with best python and graphql api design practices. An example of a style guide I created.
- Python stack: Pytorch, HuggingFace transformers, flask, mypy, strawberry, pydantic, streamlit, OpenAI api
- Scala stack: Monix, ZIO, Caliban
- Databases: Mongo, Redshift, Postgres

Patents

Neural network-facilitated linguistically complex message generation systems and method

- One of the inventors for LeadIQ's email generation system.

Education

National University of Singapore August 2017 - 2021
BBA (Business Analytics, First Class Honours) , Minor in French language studies
Honours Thesis: A Natural Language Explanation Framework for Machine Learning Decisions.
Implemented a framework to translate the output of explanation methods (e.g. Counterfactuals, LIME, Shapely values) into Natural Language Explanations.