

Tay Zhi Wen Jeremiah

HP: +65 86125565

Email: tayjeremiah2000@gmail.com/ e0726076@u.nus.edu

GitHub: <https://github.com/jeremiah-tay>

EDUCATION

National University of Singapore

Aug 2021 - Dec 2025

Bachelor of Science in Data Science and Analytics (Hons)

- **Grades:** Honors with Distinction
- **Coursework:** Convex Optimization, Data Science in Practice, Data Structures and Algorithms, Data Visualization, Deep Learning, Efficient Algorithms and Data Structures 2, Machine Learning, Modelling and Simulation, Numerical Computation, Regression Analysis, Statistical Computing in R, Statistical Learning, Database Technology and management

WORK EXPERIENCE

AgileAlgo

Sep 2024 – Jan 2025

Data Science Intern

Python, LangChain, RAG, LLM, Power BI

- **Designed and implemented** a back end large language model (LLM) pipeline using Python, delivering a scalable architecture for the company's Virtual System Implementer Platform.
- **Enhanced** LLM performance by integrating responses with domain-specific knowledge graphs, leveraging graph-based context to improve output relevance and accuracy.
- **Conducted** exploratory data analysis (EDA) and feature engineering using Pandas and Power BI, resulting in the identification of key insights that improved decision-making and predictive accuracy for client's datasets.
- **Applied** advanced AI techniques, including prompt engineering and RAG, to fine-tune model responses, aligning them with business objectives.
- **Utilized** NLP similarity search techniques to optimize information retrieval and enhance response accuracy within the company's platform.

Almazing

July 2023 – Dec 2023

Data Analyst Intern

Regex, JavaScript, PostgreSQL

- **Developed and implemented** automated data generation processes using JavaScript on Google Apps Script, resulting in a significant enhancement in operational efficiency and improved data handling capabilities within the company.
- **Conducted** comprehensive research on a Named-Entity Recognition (NER) model, significantly optimizing the company's workflow.
- **Collaborated** seamlessly with international colleagues, fostering a culture of knowledge sharing and leveraging diverse experiences and expertise to drive collective success in projects.
- **Designed and implemented** parsing rule groups utilizing regular expressions, effectively converting unstructured text into structured data, contributing to streamlined information processing and data analysis.
- **Proficient** in using SQL queries to retrieve, analyze, and manipulate large datasets.

PROJECTS

Natural Language Processing (NLP)-Based ESG Report Extraction Tool | Web-sourced ESG Reports

Mar 2025

RAG, LLM, Git, GitHub, Docker, ChromaDB, PostgreSQL, OCRmyPDF, PyPDF2, Power BI

- **Automated** the extraction and preprocessing of ESG reports from web-sourced PDFs using OCRmyPDF and PyPDF2, enabling scalable document handling and accurate text parsing across diverse report formats.
- **Implemented** Retrieval-Augmented Generation (RAG) by storing parsed text in a ChromaDB vector database, enhancing contextual relevance for Large Language Model (LLM)-powered ESG data extraction.
- **Engineered** a robust PostgreSQL schema hosted on Supabase for structured ESG data storage, supporting seamless remote access, query optimization, and future scalability.
- **Built** interactive Power BI dashboards to visualize ESG metrics across companies, industries, and continents, improving interpretability for stakeholders and supporting data driven ESG evaluations

Neo4js Knowledge Graph Data Science | Knowledge Graphs

Aug 2024

- **Developed** a python script to automate data extraction and loading into Neo4j, streamlining the data preparation process for knowledge graph analysis.
- **Leveraged** Cypher, Neo4j's query language, to perform graph analytics, extracting and analyzing specific relationships within the graph structure.
- **Applied** predictive link methods using the Graph Data Science package in python to forecast potential relationships that are not explicitly present in the existing graph, enhancing the knowledge graph's predictive capabilities.

SKILL SETS & PROFICIENCY

- **Programming Languages:** Python, R, Java, JavaScript, MATLAB, MySQL, PostgreSQL, SparkSQL, Scala, Cypher, Git, Docker
- **Data Science Tools & Frameworks:** Power BI, Tableau, Jupyter Notebook
- **Deep Learning & Machine Learning Framework:** Keras, TensorFlow, Scikit-Learn, LangChain
- **Core Competencies:** Data Cleaning, Data Modelling, Exploratory Data Analysis, Data Visualization, Predictive Analysis
- **Optimization & Operation Research:** Linear Programming, Convex Optimization
- **Other Tools:** Microsoft Word, Excel, and PowerPoint
- **Interests:** Machine Learning, Artificial Intelligence, Mathematics, Statistics