

# Derek Xu

Software Developer

## Contact

☎ (647) 636 5430

✉ [derek\\_yhx@hotmail.com](mailto:derek_yhx@hotmail.com)

Oakville, ON

## Portfolio

<https://dxiled.github.io/>

## Education

Bachelor's Degree

Computer Science

Queen's University

2018 – 2023

Kingston, Ontario

## Skills

Python (PyTorch,  
TensorFlow, NumPy,  
pandas, scikit-learn,  
matplotlib, Jupyter)

Java, C, R, SQL

Version Control (Git)

Linux shell scripting

JIRA and Confluence

Agile Development (Scrum)

MS Office (Word, Excel)

Strong communication skills

Team player

## Languages

Fluent in English

Fluent in French

## Work Experience

Applied Generative AI Specialist at WordJog Inc.

2023 – Present / Remote in Newmarket, Ontario

- Worked in a small team to deliver an end-to-end conversational retrieval system using large language models.
- Used LangChain to prepare, process, and filter user queries and context datasets.
- Implemented core application features such as logging, telemetry, and security
- Succeeded in a fast-paced startup environment.

Data Science Intern at Bell Canada

2021 – 2022 / Hybrid in Toronto, Ontario

- Worked in a small team on natural language processing for a chatbot in both English and French.
- Analyzed and classified production data to evaluate the performance of a chatbot and generate reports.
- Created intents and verified their performance using a Rasa model with QBox testing, reaching 92% training accuracy in English and 87% training accuracy in French.
- Leveraged transformer models for data augmentation and English French translation.

## Projects

Enron Email Spam Classifier

- Worked in a group in an academic setting to develop various ML models for spam classification using the Enron dataset.
- Implemented a custom sub-word-based tokenizer using the byte-pair compression algorithm.
- Achieved a testing accuracy of 85% using a recurrent neural network and the custom tokenizer, performing 10% better than the baseline recurrent model using a word-based tokenizer.
- Achieved a testing accuracy of 97% using a SVM with the custom tokenizer.

Twitter Vaccination Dataset Analysis

- Worked in a group in an academic setting to implement various analytics models to explore and analyse a Twitter dataset about COVID-19 vaccinations.
- Explored a state-of-the-art transformer model for sentiment analysis, achieving similar results to other papers regarding the dataset.
- Implemented K-means clustering, achieving a silhouette score of 0.504 and a Davies-Bouldin index of 0.6.
- Analysed the clusters, revealing a group of tweets that were likely bots made to promote paid vaccinations.