

Hyeontae Kim

 hyeontae.me |  github.com/h383kim |  h383kim@uwaterloo.ca |  [LinkedIn](#)

EDUCATION

University of Waterloo

Waterloo, ON, Canada

Data Science Co-op, Bachelor of Honours Mathematics

EXPERIENCE

Data Scientist @ Canadian Intellectual Property Office

Sep 2024 - Present

- Designed and deployed **advanced entity matching (record linkage)** methods to cleanse and standardize Canadian (**13M+** records) and worldwide (**330M+** records) patent databases stored on Amazon S3.
- Experimented with **SBERT-based Siamese networks (LLM fine-tuning)** and a **gradient-boosting** approach on fuzzy similarities, ultimately reducing data redundancy by **35%** and improving search accuracy by **40%**.
- Incorporated **advanced indexing and data chunking** techniques to optimize record linkage workflows.

Data Engineer @ Canadian Intellectual Property Office

May 2024 - Aug 2024

- Optimized a **batch ETL pipeline for classification tasks**, reducing end-to-end processing time by around **82%** handling large-scale data (**350M+** records) in PostgreSQL.
- Created a dynamic **PowerBI prototype** for senior management, utilizing **advanced data modeling** and **optimization in DAX and SQL** to address data challenges and enable interactive KPI dashboards.

Deep Learning Research & Implementation Community Member @ MetaCode

March 2024 - Present

- Led a **research-driven community** presenting bi-weekly on cutting-edge deep learning topics, including classical models and state-of-the-art (SOTA) architectures.
- Implemented deep learning models from scratch.** Low-level coding practices were applied to build neural networks, manage data pipelines, and optimize training.

Cloud Engineer @ ShinwooTNS - Cato Networks

Aug 2023 - Jan 2024

- Developed manuals and scripts for the installation of **sockets at various sites** and **vSockets in cloud environments (AWS, Azure)**, and actively participated in a **Proof of Concept (POC) project for the deployment**.
- Played a key role in developing **configuration diagrams** and **creating network and security rules**, showcasing strategic planning and technical proficiency in network security.
- Managed over **10,000 global VPN users (SDP Users)** for LXPantos using the Administrator Console(CMA).

PROJECTS

Deep Learning from Scratch ([GitHub](#)) *Python, PyTorch, OpenCV, HuggingFace, Wandb*

- Implemented **many deep learning architectures** from scratch (CNNs, RNNs, and Transformers) pipelining diverse downstream tasks: **Translation, Next-Token Prediction, NER, Classification, Detection, Segmentation.**

RAG-Based Tech Article QA System ([GitHub](#))

Python, LangChain, OpenAI

- Built a simple question-answering system using **RAG**, utilizing **LangChain, ChromaDB, and OpenAI's GPT** for document retrieval and response generation.

AI arXiv Daily ([GitHub](#))

Python, Airflow, OpenAI

- Developed a **fully automated pipeline** that fetches, filters, and translates AI-related research papers from arXiv into Korean on a daily basis.
- Translates papers using OpenAI's GPT API with **prompt engineering**, compiling them into high-quality PDFs.

CERTIFICATIONS

- AWS Cloud Practitioner (CLF-C02)
- Certified ADsP (Advanced Data analytics Semi-Professional), by Korea Data Agency (K-DATA)
- Certified SQLD (Structured Query Language Developer), by Korea Data Agency (K-DATA)

TECHNICAL SKILLS

Programming Languages: Python, C++/C, SQL(PostgreSQL), R, Go, Bash, HTML, CSS

Frameworks & Tools: PyTorch, TensorFlow, LangChain, MLflow, Apache Spark(PySpark), AirFlow, AWS, Docker, Databricks, Snowflake, PowerBI, Wandb

Libraries: Scikit-Learn, Pandas/Polars, Numpy, OpenCV, Matplotlib, NLTK, SpaCy, Selenium, BeautifulSoup