

GURUNG GAURAB

Data & AI Engineer | Computer Science Graduate | HK Permanent Resident

Website: hiregarry.com

Email: gurunggarry18@gmail.com

Contact: 5114 2437 Address: Yuen Long, HK

GitHub: github.com/GurungGaurab

LinkedIn: <https://www.linkedin.com/in/gaurab-gurung-24590b163/>

PROFESSIONAL SUMMARY

Recent Computer Science graduate (Distinction, 78% WAM) and HK Permanent Resident targeting data and AI roles. Skilled in SQL and Python (pandas) for cleaning and analysis, and Power BI with DAX for modelling and dashboards. Built a version-aware RAG pipeline and an open-source PDF ingestion tool (Docling, OCR, pgvector). Native English speaker; five years in hospitality, three as a bar manager. Working toward Microsoft PL-300.

TECHNICAL SKILLS

Data & Analytics: SQL (PostgreSQL), Python (pandas), Power BI, DAX, Power Query, star-schema modelling, ETL

AI & Retrieval: RAG, Docling, LlamaIndex, OCR, Gemini API, embeddings, pgvector, Ollama (local LLMs), n8n

Programming Languages: Python, Java, C++, SQL, JavaScript, HTML5/CSS3

Tools & Cloud: Git, pytest, AWS (Lambda, S3, DynamoDB, CloudFront), Express.js, Spring Boot, REST APIs

Languages: Native English, Native Nepali, Hindi (Fluent)

TECHNICAL PROJECTS

pdfingest – PDF Ingestion for RAG

Python · Docling · OCR · pytest

github.com/GurungGaurab/pdfingest

- Open-source CLI turning PDFs into Markdown, JSON and embedding-ready chunks, including text inside images
- Designed an OCR quality gate (confidence + vocabulary checks) with a Gemini vision fallback and per-image reports
- Settings-aware SHA-256 caching and automatic retry of incomplete extractions; 26 automated tests (pytest)

revision-rag – Version-Aware RAG Pipeline

Python · pgvector · LlamaIndex

github.com/GurungGaurab/revision-rag

- Built a RAG pipeline: Docling parsing → embeddings → pgvector search → local LLM (Qwen3 4B via Ollama)
- Showed that vector search returns outdated document versions; built per-question version filtering to fix it
- Incremental ingestion with SHA-256 change detection, so only new or changed documents are re-processed

Academic Performance Dashboard

Power BI · DAX · Power Query

github.com/GurungGaurab/UOW_Grade

- Modelled university results as a star schema (fact table + three dimensions) for clean, scalable analysis
- Built DAX measures including a credit-weighted WAM, prepared and shaped data with Power Query

Is Hong Kong Getting Hotter? – End-to-End Analysis

Python · pandas · matplotlib

github.com/GurungGaurab/hk-weather

- Cleaned 20 years (2006–2025) of Hong Kong Observatory data, handling malformed rows and type coercion
- Found the annual mean temperature rose about 1°C, faster than the global average, visualized the trend with matplotlib

EXPERIENCE

Research Assistant – Version-Managed RAG

Sep 2026 – Present

University of Wollongong College Hong Kong (UOWCHK)

- Contributor to a paper on keeping an AI assistant’s knowledge base accurate as documents change
- Showed that with default settings, text inside images embedded in PDFs never reaches Docling’s RAG chunks
- Compared four OCR approaches (incl. Apple Vision, Gemini) and designed OCR with an automatic Gemini fallback
- Documenting the data pipeline: PDF parsing and chunking (Docling), PostgreSQL/pgvector, n8n orchestration

Bar Manager – Amalfitana / Fratelli

Jun 2019 – Dec 2024

Repulse Bay, HK

- Promoted from Bar Supervisor to Bar Manager (3 years); led the team in a high-volume, fast-paced venue

Intern – HKEVPOWER, Science Park

Jan – Apr 2019

- Supported EV charging station installation and field operations; received a strong recommendation letter

EDUCATION

University of Wollongong (UOWCHK)

2026

Bachelor of Science in Computer Science – Distinction average (78% WAM)

IVE

2019

Higher Diploma in Computer Engineering – GPA 3.4 · ISTS Innovation Award