# Omar Chenine

## AI Engineer and Data Scientist

Remote | Email: pending | GitHub: pending | LinkedIn: pending

Fourth year Data Science Engineering student. I build AI systems end to end: data preparation, modeling, retrieval, evaluation, and the interface. I care about baselines, leakage, and what happens when the model is wrong.

### EDUCATION

**Data Science Engineering**

Abdelhamid Ibn Badis University | Fourth year of a five year program | Remote

### SELECTED PROJECTS

**RAG with retrieval evaluation** — Python, retrieval, evaluation

- Built a staged retrieval pipeline with a frozen evaluation set and cited answers.
- Compared retrieval against a baseline before adding generation.
- Tracked corpus size, chunking, top k, and reranker settings for each run.

**ECG image analysis** — Python, NumPy, signal processing

- Digitized ECG images into a signal and extracted rhythm, axis, PR, and QRS.
- Generated a clinical format report with a review flag for low confidence cases.
- Documented calibration assumptions and medical limits.

**Ames housing stacked ensemble** — Python, scikit-learn, XGBoost, LightGBM, CatBoost

- Built a stacked regression pipeline with log target, outlier policy, and ordinal encoding.
- Fit preprocessing inside the fold and used out of fold predictions for the stack.
- Compared the stack against single model and baseline scores.

**BigQuery RFM and Power BI** — SQL, BigQuery, Power BI

- Wrote an RFM segmentation pipeline in BigQuery with explicit segment rules.
- Built a Power BI dashboard for segment size and customer behavior.
- Documented a warehouse and dashboard region conflict and the resolution.

**Voice agents and AI agent work** — Python, agents, voice

- Designed tool using agents with typed inputs, fallback paths, and trace logs.
- Set a per stage latency budget for voice turns.
- Defined a task set and failure taxonomy for evaluation.

**Sports outcome predictors** — Python, pandas, scikit-learn, REST APIs

- Built World Cup, F1, and boxing predictors with time based evaluation.
- Used Elo as a baseline and cached API snapshots.
- Reported calibration and sample size limits.

### SKILLS

**Programming**
Python / SQL / TypeScript / JavaScript

**ML and DS**
Pandas / NumPy / Scikit-learn / XGBoost / LightGBM / CatBoost / Feature engineering / Ensembling / Model evaluation

**Data and BI**
BigQuery / Power BI / RFM segmentation

**Frontend and Web**
React / Next.js / Tailwind / Framer Motion / React Three Fiber / REST API integration

**AI systems**
RAG / AI agents / Voice agents / NLP / Image analysis / ECG signal digitization

**Security**
OpenSSL / GnuPG / Nmap / Snort