

Karl Jürgen Mollan Neyra

Senior Data Engineer · AI Engineer · Payment Systems & Fintech

Lima, Peru · +51 982 959 528 · karlmollan97@gmail.com

linkedin.com/in/karlmollanneyra · karlmollan.com · ORCID: 0000-0001-8361-4638

PROFESSIONAL SUMMARY

Senior Data Engineer with 5+ years building end-to-end data pipelines, generative AI systems, and fraud detection solutions in payment processing environments. Specialized in multi-cloud architectures (GCP, Azure) processing millions of daily transactions across BigQuery, Synapse, and AS400 legacy systems. Published researcher with four DOIs (IEEE and Zenodo) on hybrid RAG, GNN fraud detection, automated data governance, and mechatronics design. Production experience with RAG systems, conversational AI chatbots, autonomous agents, and real-time settlement analytics. Reduced pipeline latency by 85% and built ML models achieving 85% precision in production.

TECHNICAL SKILLS

Languages & Data: Python, SQL (Advanced), R, Power BI, ETL/ELT Pipelines, Data Modeling

AI & ML: RAG Systems (Hybrid, HyDE, GraphRAG), LLMs, Prompt Engineering, Machine Learning, Computer Vision, NL2SQL, LangChain, LangGraph, Autonomous Agents

Cloud & Infra: GCP (Cloud Run, BigQuery, Firestore, Vertex AI), Azure (Synapse, ADF, OpenAI, Logic Apps), PostgreSQL, Docker, Kafka CDC, Debezium

Domain: Payment Systems, Merchant Acquiring, Settlement (T+1/T+2), Fraud Detection, PCI Compliance, KYC/AML

Management: PMBOK, Agile/Scrum, MS Project, Team Leadership, Cross-functional Coordination

PROFESSIONAL EXPERIENCE

Data Engineering Specialist — Data & Analytics · Izipay (Intercorp)

Peru's largest payment acquirer, 2.5M+ merchants

Nov 2025 — Present · Lima, Peru

- Design and deploy stored procedures in Synapse and SQL Server for churn analytics with mobile-period logic, weighted averages, and volume segmentation across 2.5M+ merchant accounts.
- Build end-to-end Azure Data Factory pipelines with data existence validation, incremental copy, and quality controls ensuring daily SLA compliance for settlement processing.
- Execute cross-platform validations between BigQuery and Azure Synapse, identifying and resolving discrepancies in critical multi-cloud data migrations.
- Develop and deploy generative AI APIs and conversational chatbots on GCP Cloud Run (LangChain, Azure OpenAI, Firestore, vector stores) with 100% resolution of critical cases reported in production.
- Architect a Hybrid RAG system (vector search + FTS + HyDE + GraphRAG) with a formal evaluation framework over 52 ground-truth questions (Precision@5: 0.704, MRR: 0.853).
- Design a multi-agent architecture with LangGraph (specialized agents coordinated by a supervisor with controlled handoffs) to automate analytical workflows across heterogeneous domains.
- Implement a 6-tier NL2SQL system (metadata, semantic layer, few-shot, reranking) covering 63 tables and 2,810 columns, with operational logging and formal evaluation.
- Implement automated logging and alerting system with Azure Logic Apps + Blob Storage + multi-recipient PDF reports for 24/7 monitoring of AI APIs.

Analyst I — Planning & Commercial Intelligence · Izipay (Intercorp)

Apr 2025 — Nov 2025 · Lima, Peru

- Built retention KPI pipelines processing multi-million-row datasets in 20 min (vs. 2h prior) — 85% latency reduction eliminating legacy models.

- Refactored ecosystem of 8 legacy SQL/Python scripts, restoring 100% data reliability and cutting validation time to 5 minutes.
- Automated segmentation workflows from hours to 15 min with clustered indexes and dynamic pivoting, achieving >98% segmentation precision.
- Mapped >38 sensitive data flows per Peru's Data Protection Law (Ley 29733), minimizing PII exposure across settlement pipelines.

Innovation & AI Coordinator · LOLIMSA

Healthcare software company, 14 countries, 5,000+ clients

Jan 2025 — Mar 2025 · Lima, Peru

- Designed 2 internal WebSocket APIs for legacy-AI interoperability, connecting distributed SQL Server systems without public IPs; centralized monitoring for 5+ hospital clients.
- Built an end-to-end ML pipeline for anemia detection from images (85% precision, image → ML → PDF/email).
- Implemented advanced prompt engineering for pharmaceutical data extraction achieving 70% success rate, overcoming rate limiting across drug price comparison systems.

R&D Coordinator · LOLIMSA

Jan 2023 — Feb 2025 (2 years) · Lima, Peru

- Led cross-functional team (analysts, designers, developers) delivering innovation projects with data science and generative AI.
- Designed Power BI dashboards improving executive decision speed by 30%; implemented predictive analytics models reducing response times by 20%.
- Led desktop-to-web migration increasing operational efficiency by 25% through automation of key clinical workflows.

IT Coordinator · LOLIMSA

Jul 2022 — Dec 2022 · Lima, Peru

- Managed 200+ monthly support tickets, reducing downtime by 30% and recurrence to 7%; built PMBOK/BPM hybrid framework with USD 15K annual savings.
- Led cloud migration initiatives and billing audits with <5% deviation in timelines and budgets.

Junior Engineer / Intern · LOLIMSA

Mar 2021 — Jul 2022 · Lima, Peru

- Developed a mechatronics prototype (Autodesk Inventor) reducing fabrication costs by 23%; migrated 5 hospital servers (Huawei → OVH) with 40% less downtime.

FEATURED PROJECT

Distributed Semantic Inference Engine for BI Democratization

Mar 2025 — Oct 2025 · Private Project

Distributed architecture for natural language query (NLQ) processing translating business questions into SQL. Asynchronous WebSocket orchestration with adaptive timeouts (5s → 40s), semantic engine using open-source LLMs (Ollama/DeepSeek) with ontological schema mappings (50+ business terms), hierarchical fallbacks, and conversational interface on WhatsApp. Democratized BI access for non-technical enterprise users.

PUBLICATIONS

Data Quality Over Algorithmic Complexity: Empirical Evidence from a Production Hybrid Search System

Zenodo, 2026 · DOI: 10.5281/zenodo.19394029

Demonstrated 27% improvement in Precision@5 through data quality alone — zero code changes, zero additional cost on a production RAG system.

GRAFID: GNN-Based Anomaly Detection for Fraud Identification in Payment Systems

Zenodo, 2026 · DOI: 10.5281/zenodo.19493674

Graph Neural Networks (GraphSAGE vs XGBoost) for fraud ring detection in instant payment networks. Contribution: dataset-dependent cost-benefit guide for GNN vs XGBoost, validated on IEEE-CIS and EU Credit Card with 6.3M+ transactions.

DataGov-Factory: Framework for Enterprise Data Governance Automation

Zenodo, 2026 · DOI: 10.5281/zenodo.19520901

Framework for automated data governance with metadata capture, automatic PII classification, and lineage traceability applied to payment pipelines.

Conceptual Mechatronics Design: Autonomous Inverted Pendulum on Two-Wheeled Mobile Robot

IEEE ICAECT 2023 · DOI: 10.1109/ICAECT57570.2023.10118071

Conceptual design and prototyping with stress simulations and material validation. Published in IEEE Xplore.

EDUCATION

MSc Data Science — Universidad Nacional de Ingeniería (UNI)	2024 — 2026
MSc Project Management — Escuela de Postgrado UTP	2021 — 2023
BSc Mechatronics Engineering — Universidad San Ignacio de Loyola (USIL)	2020 — 2023
BSc Environmental Engineering — Universidad San Ignacio de Loyola (USIL)	2015 — 2019
Diploma in Agile Project Management — CENTRUM PUCP Graduate Business School	2022
Diploma in Biomedical Equipment Management — TECH SENATI	2021 — 2022

CERTIFICATIONS

Autodesk Inventor Advanced (UNI) · AutoCAD 2D & 3D (UNI) · Professional Graphic Engineering (UNI) · English Competency (USIL)

LEADERSHIP & VOLUNTEERING

- **Founder & President** — **Mechatronics Engineers Society (M.E.S.), USIL** (2020–2021). 4 divisions: Events, Marketing, R&D, Academic.
- **Teaching Assistant** — **Digital Circuits, USIL** (2021). Undergraduate course.
- **Scientific Research Director** — **Environmental Engineering Club, USIL** (2016–2017).
- **Physics Tutor** — **General Physics, USIL** (2016).