

Kenji Hilasaca

✉ lhilasacas@uni.pe 🔗 [LinkedIn/in/KenjiHilasak](https://www.linkedin.com/in/KenjiHilasak) ☎ (+44) 7887762961 | Leeds, UK

Data Scientist and Engineer with 6+ years' experience designing end-to-end data solutions and advanced ML architectures across finance, consulting, and academia. Expert in building scalable data pipelines, deploying NLP/Generative AI models (RAG, LLMs, LangGraph), and translating complex technical insights into commercial and strategic value for senior stakeholders.

RELEVANT EXPERIENCE

Research Assistant | University of Leeds - *NLP & Data Science* January 2026 - present

- Designed and optimized distributed NLP extraction pipelines on enterprise Linux cluster equipped with NVIDIA GPUs, optimizing computing efficiency to slash data processing runtime by 87% (from 48 hours down to 6 hours).
- Co-authored and presented peer-reviewed research at the LREC-BUCC workshop, introducing novel text-alignment methodologies and parallel corpora extraction for automated text simplification tasks.
- Engineered robust text-processing workflows for the iDem project (EU-funded, €1M+ budget), ensuring data integrity, lineage, and strict adherence to UKRI quality frameworks.

Data Engineer | BBVA - *Corporate & Investment Banking* March 2024 - September 2024

- Designed and deployed distributed data pipelines (PySpark, Hadoop, AWS Step Functions) reducing data processing latency by 50%.
- Delivered a client classification ML model (Python, XGBoost) to personalise corporate lending strategies and increase retention.
- Created RESTful APIs for model integration into core banking apps, enhancing data-driven risk decisions.

Data Base Analyst | INDRA - *Telecoms Consulting* October 2022 - February 2024

- Developed predictive models for late payments and fraud (Scikit-learn, XGBoost), driving \$20M annual cash flow improvement.
- Designed feature engineering pipelines and automated ETL (Python, SQL) ensuring traceable, high-quality data flows.
- Communicated model insights and KPIs to senior finance and operations teams through interactive dashboards.

Software Engineer | PLAYTEC - *Software Development* January 2016 - October 2020

- Created end-to-end MLOps pipelines (Airflow, TensorFlow, Docker), reducing model retraining time by 60%.
- Delivered web and mobile apps integrating real-time ML predictions for fraud and transaction risk scoring.

EDUCATION

University of Leeds | United Kingdom September 2024 - September 2025

MSc in Data Science and Analytics | Distinction

- Advanced coursework in ML, Big Data, Statistical Modelling, and Cloud Computing.

Massachusetts Institute of Technology (MIT IDSS) | Remote May 2022 - October 2023

Micromaster in Statistic and Data Science

National University of Engineering (UNI) | Peru March 2010 - July 2016

BSc Mechatronic Engineering | GPA 3.5/4.0 (2:1 UK equivalent)

TECHNICAL SKILLS

- **Programming:** Python (pandas, NumPy, Scikit-learn, TensorFlow, XGBoost), R, SQL, React-JS.
- **AI & Natural Language Processing (NLP):** Generative AI, RAG, Semantic Search, LLMs (BERT), Text Simplification, Transformers, LangGraph.
- **Data Engineering:** Feature engineering, ETL pipelines, Hadoop, PySpark, Airflow, Git, Docker.
- **Cloud Platforms:** AWS (Step Functions, Glue, S3, EC2), Azure.
- **Data Visualization & Tools:** Power BI, Tableau, Jupyter Notebook, MLflow.