

Igor Kołodziej

Warsaw, Poland | igor.kolodziej@proton.me | +48 728 983 612
igorkolodziej.com | github.com/IgorKolodziej | linkedin.com/in/igorkolodziej

EDUCATION

MSc Data Science (02/2026 -- Present)

Warsaw University of Technology, Faculty of Mathematics & Information Science

BSc Data Science (Engineering) (10/2022 -- 02/2026)

Warsaw University of Technology, Faculty of Mathematics & Information Science

Grade: 5.0 (Very Good) • Thesis: *NMAR: An R Package for Estimation of Population Parameters Under Data Not Missing at Random.*

EXPERIENCE

Research Software Engineer (03/2025 -- 03/2026)

Poznań University of Economics and Business (NCN OPUS 20 project)

Project: Census-like statistics for foreign-born populations – quality, data integration and estimation.

- Co-develop **NMAR** (CRAN R package): implement estimators from the literature behind a stable `nmr()` API.
- Run reproducible simulation studies and real-data experiments to compare methods and validate behavior.
- Presented the work at **uRos 2025 (Romanian NSI)** and **ElementsX 2025 (AGH)**.

Product & Software Development (Student Venture) (01/2023 -- 10/2024)

Capitalize, Warsaw

- Shipped a demo mobile app to **Google Play (testing track)**; built backend features and APIs (**FastAPI**).
- Wrote Python scripts for basic analysis and visualizations of usage telemetry from **Amplitude exports**.
- Award: **2nd Place**, Enactus Poland National Competition 2023.

Data Warehousing Intern (07/2023 -- 09/2023)

Accenture, Warsaw

- Maintained and extended an enterprise **SAP BW** data warehouse within production change processes. Used **ABAP** for fixes and enhancements.

SELECTED PROJECTS (DATA/ML SYSTEMS)

Real-Time Finance -- Big Data Pipeline

[GitHub](#)

Docker Compose reproducible batch+stream pipeline (**NiFi, Kafka, HDFS, Spark, Hive, HBase**) integrating NBP daily FX rates with a live Binance trade stream. *Two-person team project; my scope: Spark ETL/analytics, data quality checks, Hive schemas, HBase serving.*

- Implemented Spark ETL for curated datasets and analytics (OHLC/volume, returns, rolling BTC--USD/PLN correlation); added run scripts for reproducible execution.
- Designed Hive external tables/views and added sanity checks for analytics outputs to catch data issues early.
- Built the serving layer: loaded analytical facts into HBase and integrated them with analytical views for fast key-based reads.

Mamut -- AutoML toolkit for tabular classification

[GitHub](#) | [Docs](#) | [PyPI](#)

Python package for tabular classification: preprocessing + Optuna-driven HPO, model comparison, and dynamic ensemble search, with HTML reports/plots. Packaged on PyPI with docs and CI/tests.

NMAR -- Estimation under nonignorable nonresponse

[GitHub](#) | [Docs](#) | [CRAN](#)

Open-source package providing a unified `nmr()` interface for NMAR estimation on i.i.d. data and complex survey designs. Includes multiple engines, vignettes, CI, and test coverage.

QuantumRAG -- Grover-inspired top-k selection for RAG

[GitHub](#)

Research RAG prototype: dense retrieval (FAISS + transformer embeddings) + GroverTopK selector (Qiskit Aer) with evaluation on **SQuAD**

1.1. Benchmarked `llama-3-8b / mixtral-8x7b / phi-3.5`. Grover matched classic selection in **99.11%** of cases with \approx **30 ms** overhead, and retrieval average of **0.29 s**.

LEADERSHIP & OUTREACH

President, Data Science Club (05/2024 -- 05/2025)

Warsaw University of Technology

- Led a student team organizing talks/workshops; hosted guests from Google, ING, and Allegro.
- Co-organizer, ensembleAI hackathon** (2024, 2025): Managed sponsors, logistics, and on-site operations.

SKILLS

Data/Systems: SQL, Spark, Hive, HDFS, HBase, NiFi

Infra/Workflow: Docker, Linux, Git, GitHub Actions (CI)

Programming: Python, Java, Scala, R

Backend: FastAPI, Spring Boot

ML: PyTorch, scikit-learn, Transformers, Optuna

Languages: Polish (Native), English (C2, CAE Certificate), German (Basic)