Marcell Granát

Senior Data Scientist & Quantitative Researcher

Budapest, Hungary · marcell.granat@icloud.com · LinkedIn · GitHub

SUMMARY

Quantitative researcher and data scientist with deep expertise in applied machine learning, natural language processing (NLP), and high-throughput data engineering. Proven track record across central banking and institutional environments (European Central Bank, MNB), architecting production-grade AI systems, building distributed pipelines (DuckDB, Spark, Iceberg), and delivering empirical econometrics research. Driven by rigorous methodology, algorithmic problem-solving, and building robust, scalable analytical systems.

PROFESSIONAL EXPERIENCE

European Central Bank (ECB) — Oct 2025 -- Oct 2026

Data Science Expert (F/G level), DG/S Financial Market Data — *Frankfurt, Germany*

- **RAG Architecture & NLP:** Designed and deployed production RAG systems for the MMSR chatbot project, structuring semantic retrieval pipelines and grounding complex documentation for external stakeholders.
- **Distributed Pipelines & Data Engineering:** Managed high-throughput data flows and Apache Iceberg workflows; processed granular financial datasets (TRAX, ABS, iBoxx, €STR, MMSR) with automated pipelines in Python and SQL.
- **High-Precision Classification:** Engineered automated classification pipelines linking massive, granular SFTDS and BrokerTec datasets, achieving 99.9% classification accuracy on high-frequency transactions.
- **Financial Analytics & Market Dynamics:** Analyzed granular repo market data (SFTDS) to deliver actionable empirical insights into market liquidity and the impact of Eurosystem asset purchase programmes.
- **Software Engineering Discipline:** Enforced rigorous software standards across production pipelines, including unit testing (pytest), peer code reviews, containerization (Docker), and git workflows.

Central Bank of Hungary (MNB) — Feb 2022 -- Sep 2025

Data Scientist, International Monetary Policy Directorate — *Budapest, Hungary*

Data Scientist — Sep 2024 -- Sep 2025

- **Applied AI & Vision LLMs:** Directed an archival processing and validation pipeline using vision-enabled LLMs and strict programmatic validation to extract structured balance sheets from complex historical financial yearbooks.
- **High-Performance Pipelines & DSL Design:** Built high-throughput ETL pipelines with DuckDB for multi-million transaction records; designed declarative domain-specific language (DSL) abstractions for complex spatial matching, improving data matching rates from 54% to >93% (presented methodology at the BIS-hosted Real Estate Statistics Conference 2026 in Tokyo).
- **Asynchronous Web Scraping & API Library:** Engineered a Python API library for automated data extraction from the Arcanum Digital Archives, utilizing Playwright for resilient browser automation, async request scheduling, and robust execution under strict quota limits.
- **Institutional AI Enablement:** Served on the MNB working group for institutional AI strategy, actively tracking frontier AI developments and conducting internal training courses to upskill researchers on LLM utilization and AI-assisted workflows.

Research Expert — Feb 2022 -- Aug 2024

- **NLP & Labor Market Analytics:** Built automated web scraping and NLP pipelines using LLM APIs to quantify task-level occupational exposure to generative AI across millions of job advertisements.
- **Empirical ML, Econometrics & Methodology:** Engineered empirical models to test market pricing of geopolitical risk, published Q1 research in financial methodology and asset pricing (*Expect the unexpected: Did the equity markets anticipate the Russo-Ukrainian war?*), and developed structural topic models to quantify environmental sentiment from unstructured media sources.

PROFESSIONAL EXPERIENCE (CONT.)

Research Flow Aug 2024 -- Apr 2025

Data Scientist (Freelance) *Budapest, Hungary*

- **Healthcare Microdata & Bayesian Modeling:** Processed large-scale administrative healthcare microdata from NEAK via ODBC connections, applying Bayesian structured models to examine drug utilization patterns.
- **Application Development & Testing:** Developed client-facing Shiny applications and custom internal R packages to streamline reporting workflows; built comprehensive unit and integration testing suites (testthat) ensuring production reliability.

ACADEMIC & MENTORING EXPERIENCE

John von Neumann University Jul 2022 -- Jul 2024

Part-time Lecturer & Thesis Supervisor in Big Data *Kecskemét / Budapest, Hungary*

- **Mentorship:** Supervised 10 Master's theses (9 in Natural Language Processing, 1 in Deep Learning), guiding students in designing and deploying end-to-end NLP pipelines, predictive models, and LLM evaluation setups.
- **Teaching & Public Engagement:** Instructed the "Big Data and Data Visualisation" master's course (covering SQL, DuckDB, R/Python, and financial microdata); delivered a TEDx talk on statistical thinking, data in society, and applied machine learning.

EDUCATION

Eötvös Loránd University (ELTE) Sep 2022 – Present

PhD in Economics & Finance (Coursework credits & publications completed; degree expected Summer 2027) Budapest, Hungary

- **Research Focus:** Designing automated validation frameworks and leveraging Large Language Models to extract and evaluate structured information from unstructured data sources.
- **Academic Representation (LERU):** Represented ELTE in economics at the League of European Research Universities (LERU) Doctoral Summer School (Univ. of Copenhagen) on *"AI across University Disciplines"*, collaborating with 54 doctoral researchers from 24 European institutions on foundational AI and Large Language Models.

Corvinus University of Budapest 2017 – 2022

MSc in Economic Analysis (2020–2022) · *BSc in Applied Economics* (2017–2020) Budapest, Hungary

- **Awards & Competitions:** 1st place in National Data Science Competition (Hungarian Central Statistical Office - KSH); 1st place in Whiteshield Data Science Competition utilizing BERT; 2nd place at Morgan Stanley Quant Challenge (2023).

TECHNICAL SKILLS

- **Programming & Development:** Python, R, SQL, Rust, FastAPI, Git, Docker, CI/CD, Unit Testing (pytest, testthat), API Design.
- **Data Engineering & Distributed Systems:** DuckDB, Apache Spark, Apache Iceberg, ODBC, ETL Pipeline Architecture, High-Throughput Microdata Processing.
- **Applied AI, NLP & Machine Learning:** Large Language Models (LLMs), Production RAG Systems, Vision-enabled LLMs, Semantic Search, Embeddings, PyTorch, Scikit-learn, Structural Topic Modeling, LLM Evals & Validation.
- **Financial Analytics & Data Providers:** Bloomberg, Refinitiv, CRSP, Compustat, SFTDS, Repo Market Analytics, Empirical Asset Pricing.
- **Mathematics & Quantitative Methods:** Econometrics, Bayesian Modeling, Statistical Learning, Time Series Analysis, Experimental Design.