

András Novoszáth

Budapest, Hungary | +36 30 889 4244 | nocibambi@gmail.com | [LinkedIn](#)

Over a decade of experience across data engineering, machine learning, and knowledge systems — applied to clinical research, and regulated data & analytics environments. Focusing on Real-World Data in clinical research.

EXPERIENCE

Data Platform & Operations Engineer | Diligent | Budapest, Hungary | June 2023 – Present

- Designed and operated **serverless data-platform infrastructure** for governance and voting-data pipelines.
- **Migrated data collection workflows** between languages (VBScript to Python) and infrastructures (CDK to Terraform). Resolved data-quality, performance and rate-limit issues.
- **Built out data operations and QA monitoring** including structured logging, dashboards, and alerting (CloudWatch, SNS, Slack/Teams).
- Built **PySpark / EMR pipelines and Apache Airflow DAGs** for ETL workflows within a medallion architecture.
- Orchestrated **multi-month backfills** with live run logs and per-stage data-coverage summaries.
- **Cut CI/CD build time by 75%** and expanded reliability and developer-experience tooling. Managed shared ECR / CodeArtifact images.
- **Integrated AI tools** into team workflow. Delivered internal **AWS training** sessions on databases and microservices.

Data Insights Squad | Aragon | Remote | Aug 2022 – Feb 2023

- Built community and governance analytics for a distributed organisation, integrating multiple primary data sources for reporting and oversight.
- Designed and shipped a financial-oversight dashboard tracking treasury funds across distributed wallets.

Data Scientist & Engineer | Freelance | Remote | Sep 2018 – Jun 2023

- **Renovia/Bold Type Consulting**: built a clinical research reporting pipeline from **clinical trial data** assessing a pelvic-floor therapeutic device for stress urinary incontinence; data was critical in securing funding.
- **Bioepic**: built a time-series glucose forecasting model from clinical-trial CGM data, **matching the prediction accuracy of market-leading commercial medical devices**.
- Built an ML pipeline for evaluating feature-engineering techniques on multi-label classification datasets.
- Delivered analytics work across health, energy and distributed-organisation contexts.

Junior Consultant & Technical Writer | Dorsum | Budapest, Hungary | Jan 2016 – May 2018

- Supported a **B2B wealth-management SaaS** through regulation analysis, technical documentation, business proposals and marketing content.
- Translated compliance requirements into developer-facing guidelines and introduced single-source documentation practices.
- Wrote white papers on wealth management, mobile banking and financial innovation.
- Produced the business and functional sections of five RFPs and related RFIs.

SKILLS

- **Clinical & research data:** clinical-trial reporting, biosignal time-series modelling.
- **Data engineering:** Python, pandas, NumPy, pyspark, SQL (MS SQL, BigQuery, Snowflake, InfluxDB), Pydantic, Apache Airflow, requests, BeautifulSoup, Selenium, Playwright.
- **Cloud & DevOps:** AWS (Lambda, EMR, SNS, SQS, CloudWatch, CodeBuild, CodePipeline, CodeArtifact, ECR, Step Functions, EventBridge), Terraform, AWS CDK, Docker, GitHub Actions, Linux, bash.
- **Machine learning:** scikit-learn, Keras, time-series forecasting, classification, clustering, cross-validation, feature-engineering evaluation.
- **Data visualisation & dashboards:** matplotlib, seaborn, altair, plotly, Streamlit, Dash, Anvil.
- **AI in development:** Cursor and VS Code agents, project-specific agent configurations, prompt engineering.
- **Methodology & research:** research design, ethnographic and evidence-practice analysis, technical writing, regulation analysis.
- **Languages:** English (fluent), Hungarian (native).

EDUCATION

- **Ph.D., Science & Technology Studies** – The Open University, 2010 – 2016. Ethnographic study of knowledge production and technologies. Fieldwork on the Bristol Pound local currency.
- **Diploma (BA + MA), Economics** – Budapest University of Technology and Economics, 2002 – 2007. Specialisation in economic analysis: statistics, econometrics, optimisation. Dissertation on economic growth models.

CERTIFICATIONS

- Mathematics for Machine Learning (Coursera, 2019).
- IBM Cognitive Class – Data Science Foundations & Business levels (2017).
- Arc Certified Remote Data Scientist (2022).
- Using Databases with Python (Coursera, 2018).

SIDE-PROJECTS

- **Aave Liquidity Provider TVL Point Tracker** (Oct – Nov 2024) – tracker computing token-balance points from primary on-chain data, with end-to-end source-to-metric provenance and verification; results exposed via FastAPI.
- **Token-Swap Pool / Market Comparison ETL** (Aug – Sep 2024) – pipeline integrating multiple distributed-market data sources with schema validation in Pydantic.
- **CTR outlier detection** (Mar 2022, Adverity challenge) – outlier detection on click-through-rate time-series data.
- Technical write-ups and essays: [Machines of Trust](#), [Medium](#), [Dev.to](#), [Data Science learning notes](#).

CURRENTLY STUDYING

OHDSI / OMOP data model (EHDEN academy) and clinical-evidence methodology — target-trial emulation, propensity-score matching, and external-control study design.