Filip Nový

Machine Learning Engineer, MSc in Data Science
Prague, Czechia
filip.novy@hotmail.com · +420 739 737 537 · 24 March 2001
LinkedIn · GitHub



Relevant Experience

Machine Learning Engineer August 2026 – present
TV Nova, Prague

- Designed and built an end-to-end evaluation suite for a production agentic recommender built with LangGraph at the largest streaming platform in Czechia, automating API-driven query execution, response capture and scoring.
- Created behavioural tests and a golden retrieval dataset from scratch; combined a custom LLM judge with deterministic information-retrieval metrics to assess agent behaviour and retrieval quality.
- Reduced token expenditure by approximately 23% through prompt and tool-call optimization, while strengthening safeguards against jailbreaks and unintended usage.
- Delivered unit-tested Python code, translated research methods into production implementations, and integrated Azure- and GCP-hosted LLM APIs.
- Contributed to production database schema design.

NLP Researcher (Thesis and Publication Project) January – July 2026
University of Southern Denmark (SDU), AlgaeProBANOS

- Built an end-to-end GraphRAG system with LangChain over 879 scientific articles, combining PDF extraction, dense retrieval, cross-encoder reranking and a knowledge graph of approximately 60k entities, with taxon annotations linked to the GBIF taxonomic backbone.
- Implemented schema-validated LLM extraction, entity resolution and two-stage community summarization to reduce indexing costs.
- Evaluated four retrieval configurations across 203 questions with RAGAS and statistical testing; first-author publication at KONVENS 2026. Research within EU Horizon Europe project AlgaeProBANOS.

Data Trainee (analytics and BI) April 2023 – August 2024
T-Mobile Czech Republic (Deutsche Telekom Group), Prague
SQL data pipelines and Power BI dashboards for financial forecasting and sales-lead prioritization, adopted by regional sales teams nationwide.

Education

MSc Data Science Sept 2024 – June 2026
University of Southern Denmark (SDU), Kolding; degree awarded 19 June 2026
Coursework graded 12 (A, highest): Natural Language Processing, Deep Learning, Applied Machine Learning, Statistics for Data Science, Visualization, Programming for Data Science

Thesis: GraphRAG for scientific question answering; **12 (highest mark)**, 30 ECTS.

BSc Applied Computer Science 2020 – 2024
Prague University of Economics and Business (VŠE)
Thesis: An Analysis of the Current State of Research on Artificial General Intelligence (grade A).
Erasmus exchange semester in Poland.

Technical Skills

Programming	Python, SQL, R; PyTorch, scikit-learn, FastAPI, Pydantic
ML & NLP	Deep learning, NLP, information retrieval, recommender systems, RAG / GraphRAG, agentic LLM applications; LangChain, LangGraph
Evaluation	Evaluation framework development, benchmark design, LLM-as-a-judge, information-retrieval evaluation, experimental design and statistical analysis
Databases	Database schema design; PostgreSQL, Microsoft SQL Server, Oracle; Neo4j (graph), ChromaDB (vector), Redis
Engineering	API integration, unit and integration testing, code review and collaborative development; Git, Docker, CI/CD with GitHub Actions; exposure to Kubernetes
Cloud & MLOps	Azure, AWS, Databricks, MLflow; GCP-hosted LLM APIs

Selected Projects

Deep Learning Portfolio
PyTorch, TensorFlow
From-scratch implementations of CNNs, RNNs/LSTMs, autoencoders, VAEs, GANs, and Transformers, including the underlying optimization machinery (SGD variants, backpropagation).

Vision Transformer for Plant Disease Classification
PyTorch
Fine-tuned a Vision Transformer with custom attention heads on 87,000 leaf images across 38 disease classes, achieving 99.77% validation accuracy. Oral exam grade: 12 (highest).

Publications

Filip Nový and Tariq Yousef (2026). **Not All Context Helps: Isolating Graph Expansion and Community Summaries in Scientific RAG**. In Proceedings of KONVENS 2026. First author; peer-reviewed conference paper. [PDF]

Teaching and Mentoring

Volunteer student tutor (2025), MSc Data Science, SDU: mentored incoming students across 30+ nationalities; recognized in a written recommendation by Programme Coordinator Søren Brolund Ulriksen.

Languages

Czech (native); English **C2** (Cambridge CPE); Danish **B1**; German **A2**