Nils Fleischmann

ML Research Engineer

LinkedIn | GitHub | Munich, Germany

Experience

Pruna AI — Munich, Germany

ML Research Engineer, Applied Team — *Jan 2026 – Present*

- Founding member of the applied team, which builds and ships Pruna's own fast models.
- Core engineer on the **P-Models** family (P-Image-Edit, P-Video-2-Pro, ...), with **over 78M runs** on Replicate.
- Built **P-H3-Image** from MiniMax-H3: **over 600K downloads** on Hugging Face, now in diffusers and ComfyUI.
- Trained 5- and 8-step distillation LoRAs for Qwen-Image-2.1 (**over 500K downloads** on Hugging Face).

ML Research Engineer, Research Team — *Jan 2025 – Dec 2025*

- Built Pruna's proprietary diffusion **caching methods**, which power official BRIA and FLUX.2 endpoints.
- Added caching and quantization algorithms to the open-source `pruna` library; made its tests 2.4× faster.

Articly — Munich, Germany

Backend Developer, Working Student — *Jul 2024 – Nov 2024*

- Designed and built the backend of an AI-based audio-briefing product.
- Developed methods to automatically evaluate and improve LLM outputs.

Ehrenmüller — Kempten, Germany

Data Scientist, Working Student — *Jan 2022 – Dec 2022*

- Built an ML pipeline that finds relevant scientific papers for a medical device manufacturer's research team.
- Optimized the post-processing pipeline of a model that detects walls in building floor plans.

Publications & Talks

Predictive Feature Caching for Training-free Acceleration of Molecular Geometry Generation — TMLR 2026

*Sommer, **Fleischmann**, Rachwan, Günnemann, Charpentier*

TreeGen: A Bayesian Generative Model for Hierarchies — NeurIPS 2025

*Kollovieh, **Fleischmann**, Guerranti, Charpentier, Günnemann; M.Sc. thesis*

Talks on Inference Optimization for Diffusion Models

Hugging Face Computer Vision Hangout; Munich NLP

Education

Technical University of Munich — Munich, Germany

M.Sc. in Mathematics in Data Science; MDSI Master Scholarship; **Grade Average: 1.0** — *Apr 2022 – Jan 2025*

École Polytechnique Fédérale de Lausanne — Lausanne, Switzerland

Exchange semester (Swiss-European Mobility Programme); **Grade Average: 5.8/6** — *Sep 2023 – Feb 2024*

Ulm University — Ulm, Germany

B.Sc. in Mathematics and Management; Erasmus semester in Paris; **Grade Average: 1.3** — *Oct 2017 – Oct 2021*

Projects

Amazon Review Miner — LauzHack 2023, Finalist

Clustered complaints in thousands of product reviews with LLMs and embeddings

Target Search Platform for Infineon — TUM Data Innovation Lab 2023

Fine-tuned a transformer to find M&A targets from company descriptions