

Vincenzo Timmel

AI/ML Engineer | Scientific ML, LLMs, Agentic Systems

📍 Basel, Switzerland | 📞 +41 76 383 18 33 | ✉️ vincenzo.timmel@gmail.com

🌐 vincenzo-timmel | 🌐 kenfus.github.io

Experience

AI/ML Engineer — University of Applied Sciences and Arts Northwestern Switzerland (FHNW)

Brugg, Switzerland | Sept 2022 - Present

Build and deliver AI/ML solutions across the full model lifecycle, from research prototypes to production systems and publications.

NLP, Foundation Models & Agentic Systems

- Fine-tuned **OpenAI Whisper (PyTorch, custom training loop)** for Swiss-German → Standard-German, improving BLEU **62 → 78 (+16)** on STT4SG-350; published at **SwissText 2025**.
- Reduced training runtime by **50%** and scaled effective batch size **16×** via mixed precision, activation checkpointing, 8-bit optimizers, and distributed multi-GPU training.
- Built a **RAG-based verbatim correction system** using parliamentary protocols as structured context, reducing WER **15% → 10%** on the 650h **SPC_R** corpus, creating the biggest Swiss-German ASR dataset; published at **SwissText 2025**.
- Re-architected and benchmarked a commercial ASR stack for a Swiss startup, evaluating model families, preprocessing, and decoding strategies; achieved **+15 BLEU** average improvement in production.
- Fine-tuned and deployed foundation models (OpenAI, Hugging Face) for downstream tasks including classification and translation, implementing task-specific training, evaluation, and inference pipelines.

Space Science & Signal Processing

- Built an automated **solar radio-burst detector** (PyTorch + ResNet-18) for **e-Callisto**, achieving **+13%** higher recall at equal precision than domain experts.
- Developed **ecallisto_ng**, a reusable Python data pipeline converting domain-specific solar spectrometer (FITS) scientific data into structured, ML-ready Torch tensor datasets, enabling efficient and convenient preprocessing for deep learning.
- Scaled the **Karabo** (a radio-telescope simulation framework) to multi-node, multi-GPU clusters using Dask-based distributed orchestration and data-parallel execution, reducing end-to-end runtimes by **80%** for ESA workloads.

Additional Impact

- Engineered a multi-LLM content pipeline with automated post-editing and quality gates, reducing narrative content production time by **90%** while meeting style and factuality requirements.
- Established internal ML engineering standards and model lifecycle practices, introducing MLflow and Hugging Face-based checkpointing, experiment tracking, and reproducible training workflows across teams.
- Improved internal model accessibility by packaging and deploying previously siloed research models as reusable services, enabling cross-team adoption.

Data Analyst - Onedot

Zürich, Switzerland | May 2019 - Jul 2021

Applied machine learning and scalable data processing techniques to integrate and enhance large-scale product data for enterprise clients.

- Automated categorization, cleaning, and normalization of **4.1M** product records from **50k** semi-structured XML files using rule-based and ML-driven pipelines.
- Eliminated a 90-minute daily manual integration process by building a direct ingestion pipeline for **600k** articles into the data warehouse, saving **375+ hours/year**.
- Managed up to four concurrent data integration projects, coordinating product managers and engineers to deliver scoped solutions on schedule.

Selected Publications

1. **Beinlich F.R.M.**, et al. (2024).

Oxygen Imaging of Hypoxic Pockets in the Mouse Cerebral Cortex.
Science, 383(6690), 1471-1478.

Contribution: Developed semi-matrix factorization signal extraction methodology to isolate spatially localized transient oxygen dips (hypoxic pockets) from bioluminescence imaging data.

[DOI](#)

2. **Timmel V.**, et al. (2025).

Fine-Tuning Whisper on Low-Resource Languages for Real-World Applications.
SwissText 2025.

[DOI](#)

3. **Timmel V.**, et al. (2025).

SPC_R: RAG-Based Verbatim Transcription Correction for Swiss-German ASR.
SwissText 2025.

[DOI](#)

4. **Sharma R.**, et al. (2025).

Karabo: A Versatile SKA Observation Simulation Framework.
arXiv:2504.00303.

[DOI](#)

Technical Skills & ML Engineering

- **Programming:** Python, Bash, Git
- **Deep Learning:** PyTorch (custom training loops, mixed precision, distributed multi-GPU training)
- **Foundation Models & LLMs:** Hugging Face, FAISS, LangChain, OpenAI APIs, fine-tuning, RAG pipelines, agentic workflows
- **Model Development:** Benchmarking, robustness testing, error analysis, domain shift / OOD evaluation
- **MLOps & Deployment:** Docker, MLflow, CI/CD (GitHub), reproducible training and inference pipelines
- **Scientific Computing:** Signal processing, FITS-based data pipelines, large-scale dataset engineering

Education

M.Sc. Computer Science (OMSCS), Georgia Institute of Technology, USA

2025 - 2027 (part-time, alongside full-time AI/ML engineering role; expected)

B.Sc. Data Science — FHNW, Brugg, Switzerland

2021 - 2022

Mechanical Engineering — ETH Zürich

2016 - 2020

Interests

- Swimming, scuba diving (PADI Advanced Open Water), strength training, golf, building side projects with OpenClaw (self-hosted AI assistant/agent platform)