Furrer **Stanislas**

Sr. AI Engineer | Software Engineer | ETHZ • EPFL

in stanfurrer 🗘 github 🏾 www.stanfurrer.ch @ stanislas.furrer@gmail.com 💡 Zürich, Switzerland

Experienced Sr. Al Engineer (+3.5Years) specializing in deploying and scaling production-grade Deep Learning solutions. Proven track record from roles at IBM, Credit Suisse, Logitech, and Start-ups with research contributions at ETHZ and EPFL universities. Skilled in transforming cutting-edge AI research into revenue-driving products, focusing on Asynchronous Multi-Agent framework with continuous efficient fine-tuning and scalable serving of transformer-based models.

I strive to collaborate and translate research findings to make better product with stronger **business impact**.

SKILLS

Languages	Proficient : Python, Pyspark	Good : C/C++, SQL, HTML, CSS
Machine Learning Framework	Proficient : PyTorch (Lightning)	Good : TensorFlow, Keras, Scikit-Learn
GPU/CPU Computing	SLURM, Kubernetes, DistributedDataParallel (DDP), DeepSpeed.	
Cloud / Framework	AWS, Azure, Palantir, IBM Cloud.	
Developer Tools	Docker, Git, bash/UNIX, Jira, Databricks, Weight & Biases.	

PERSONAL WEBSITE ~ WORK EXPERIENCE

SR. AI ENGINEER • IBM • ZÜRICH SWITZERLAND

- Advanced IBM products by integrating cutting-edge AI research and scalable technologies.
- Developed a distributed asynchronous agentic framework with WebSocket communication.
- Built a hardware-optimized fine-tuning framework (SFT, Instruct FT, DPO) using LoRaX on multi-GPU systems.
- Leveraged multi-threading, async queues, memory management, multi-modalities, multi-LLM, tools, and RAG.
- Co-led InstructLab in Europe, IBM's synthetic data generation and fine-tuning framework.
- Delivered agentic and fine-tuning products that drove 15+ client opportunities, generating multimillion-dollar revenue.
- Offered part-time expertise in GenAI and DL/ML for IBM projects, including SQL and Python generation.

DATA SCIENTIST • CREDIT SUISSE • ZÜRICH SWITZERLAND

- Developed/Optimized terabyte-scale ETL and ML pipelines in Python/PySpark, boosting Hedge Fund strategies.
- Developed production LLM (NER, Information Retrieval, Search,...), guiding over 100+ collaborators' daily investments.
- Optimized ETL pipelines, cutting compute & memory by 30%, saving millions in operational costs per year.
- Served as Deputy Tech Lead, driving CI/CD processes and mentoring a team of 10+ developers.

MACHINE LEARNING RESEARCH ASSISTANT • EPFL-LASA • LAUSANNE SWITZERLAND

- Designed an LSTM/CNN algorithm for real-time robotic manipulation using auditory and tactile inputs.
- Developed a Bayesian control framework to enhance grip stability via predicted inertial properties.
- Implemented in Python and transitioned to a **robust C++** solution on ROS.
- Published at AI-HRI. Work carried out concurrently with my Master Thesis. C ArXiv Paper

MACHINE LEARNING INTERN • LOGITECH • LAUSANNE SWITZERLAND

- Led a deep learning project for early tremor detection (F1-score : 0.91) using LSTM-CNN models.
- Predicted human diseases from keyboard and mouse data, focusing on SHAP-based interpretability.
- Built scalable ETL pipelines with in-house apps, AWS S3, PySpark, and Docker.
- Improved MLOps with feedback monitoring, A/B testing, and iterative enhancements.
- Supported additional Logitech research projects in computer vision and time series.

DATA SCIENTIST (PART-TIME) • SMATCH SA • LAUSANNE SWITZERLAND

- Contributed to Develop a sports connectivity platform at a dynamic startup
 Website
- Conducted user interaction analysis to drive strategic decisions and platform improvements.
- Contributed to scaled user base to 50,000+ and secured \$50,000 funding in Lausanne.

LANGUAGES



CERTIFICATIONS

- AWS Machine Learning Specialty
- AWS Solutions Architect Associate
- Databricks Machine Learning Associate

FEB 2021 - JUNE 2021

FEB 2022 - FEB 2024

FEB 2024 - TODAY

AUG 2020 - FEB 2021

JUL 2017 - JUL 2019

4 JANVIER 2025

PROJECTS

EMOTION FINETUNING ON OPENSSUBTITLE (BERT) | 🗘 github 🛛 🖸 Paper FEB-JUL 2020 • Created a model-driven method to align movie subtitles with speakers, generating a unique dialogue resource. • Annotated dialogues with 32 emotions and 9 empathetic responses, fine-tuning BERT for emotion classification. META-LEARNER LSTM FOR FEW-SHOT LEARNING | 🖓 github 🛛 📿 Paper FEB-JUL 2020 • Optimized performance and explored alternative architectures for an LSTM-based meta-learner in few-shot learning. COMPARISON BETWEEN TWO DIMENSIONALITY REDUCTION TECHNIQUES | 🗘 github 🛛 🖸 Paper FEB-JUL 2019 • Reviewed and compared LLE and its variant in terms of stability with diverse data and hyperparameters. • Evaluated topology preservation and classification performance, benchmarking against t-SNE and UMAP. LEARNING TO PLAY PONG WITH DEEP REINFORCEMENT LEARNING | 🗘 github MAY-JUN 2019 Built a Pong agent using Actor-Critic and A2C policy gradients. HACKATHON MICROSOFT X IBM | 1ST /250 TEAMS **MAY 2023** • Integrated a multi-agent framework into Teams calls to enhance efficiency and enable data-driven decisions. • Leveraged Function Call, RAG, tools like Email and Jira, and Whisper for audio-to-text processing. • Built using AutoGen and Azure services, including WebApp, Functions, Power Automate, and Copilot Studio. HACKATHON UBS | FINALIST TOP 10/230 TEAMS Nov 2023 • Build a Multi-Agent (summarize, Q&A, RAG) application for Legal Document. Used across the Bank. • Integrated OCR, tree-based chunking, agent orchestration, chatbot memory, and evaluation. HACKATHON DATABRICKS | 2ND/130 TEAMS **SEP 2023** • Built a LangChain **RAG application** in Python, scaled with PySpark, and deployed using MLflow. • Processed data, generated embeddings, indexed vectors, and implemented prompt augmentation for LLM. NOMINATED PROJECT | 50 YEARS EVENT @ EPFL | 🖓 github 🛛 🖓 Paper **JUL 2019** • Engineered **biocompatible** thin-film sensors for vein temperature and blood flow in **cleanroom**. • Develop the C++ interface for ESP32 to collect sensor data. HIGH FREQUENCY TRADING (SHS) - BEST POSTER 2015 EPFL **DEC 2015** • 1rst place among 1,700 EPFL students for the best poster analyzing high-frequency trading.

EDUCATION

Sep 2018 - Jul 2021	MSc. Robotics & Data Science EPFL • ETHZ
Sep 2015 - Jul 2018	BSc. Microengineering EPFL

- Author of a robust multimodal contrastive framework for attention-based Vision & Language models.
- Leveraged self-supervised pretraining to strengthen cross-modal relationships using adversarial samples.
- Managed large-scale multi-GPU training in collaboration with ETHZ, EPFL, and NYC universities.

MASTER THESIS : ROBUST MULTIMODAL CONTRASTIVE LEARNING | 🖸 github 🛛 🖸 Paper

• Utilized PyTorch, SLURM, DeepSpeed, and Weight & Biases for scalable, mixed-precision multi-modal training.

Awards

FEB-AUG 2021