

Fabian Jetzinger

(he/him) | Nationality: Austrian

Valencia, Spain (Open to local & remote worldwide)

Email: jetzinger.fabian@gmail.com | Mobile: (on request; removed from public version)

Web: <https://fabianjetzinger.github.io/>

Education

Industrial PhD in Bioinformatics | BioBam Bioinformatics S.L. & University of Valencia

2023 – Present (Expected Defense: Q4 2026 – Q1 2027)

Fellow of the Marie Skłodowska-Curie Actions (MSCA) Doctoral Network LongTREC.

Industrial research in collaboration with Ana Conesa (I2SysBio).

Selected PhD Projects:

- Developed a LangGraph-based LLM agent (including tool calling, RAG, web search, etc.) for bioinformatic analysis within OmicsBox, helping users explore their data, identify suitable analyses, configure and launch workflows, and interpret results.
- Independently designed and developed a deep learning method for protein function prediction, PLM2GO (manuscript in preparation). This involved data engineering of public protein data, use of protein language models, model training and evaluation, and orchestration in AWS infrastructure (including S3 and EC2 spot instances).
- Developed a Nextflow pipeline for the benchmarking of transcript reconstruction tools for long-read RNA-sequencing data sets with multiple biological replicates.
- Development, maintenance, and user support of the OmicsBox long-read transcriptomics module, making long-read RNA-seq analysis accessible to biologists.

MSc in Data Science and Engineering | University of Applied Sciences Upper Austria

2021 – 2023

Thesis: *Named Entity Recognition on German traffic-related texts*. Built a hybrid NER pipeline combining a custom rule engine with fine-tuned transformer language models and a disambiguation step, benchmarked against GPT-3.

BSc in Medical and Bioinformatics | University of Applied Sciences Upper Austria

2018 – 2021

Thesis: *Dimension Reduction on Time Series Data*. Applied PCA, t-SNE, UMAP, and an autoencoder (TensorFlow/Keras) to anomaly detection in clinical time-series data.

Matura Diploma (Higher Technical College for Informatics) | HTBLA Grieskirchen, Austria

2012 – 2017

Thesis: *Towards Predicting the Popularity of Music Artists*. Presented at the 10th International Workshop on Machine Learning and Music at Universitat Pompeu Fabra, Barcelona.

Publications

- Jetzinger, F. et al. "To join or not to join: handling biological replicates in long-read RNA sequencing data." *bioRxiv* (2025). (Accepted for publication in *Nature Communications*.) First author. Performed pre-processing, transcript reconstruction and benchmarking analysis across PacBio, ONT, and Illumina sequencing platforms.
- Liu, T., et al. "TUSCO: benchmarking transcriptome reconstruction with endogenous single-isoform controls." *Nature Communications* (2026). Contributed to pre-processing of long-read RNA-seq data and transcript reconstruction.

Professional Experience

Industrial PhD Student | BioBam Bioinformatics S.L.

2023 – Present

Developing methods for protein function prediction, and responsible for development, maintenance, and support of the OmicsBox long-read transcriptomics submodule.

Junior Data Scientist | RISC Software GmbH

2021 – 2023

Contributed to development of a hybrid named-entity recognition pipeline (rule-based engine + fine-tuned transformer language models, with a disambiguation step) to automate German-language traffic-incident reporting from speech-to-text data, as part of a multi-partner research project; benchmarked against GPT-3.

Data Science Intern | RISC Software GmbH

2019 – 2021

Applied dimension reduction techniques (PCA, t-SNE, UMAP) and deep convolutional autoencoders to real-world clinical time-series data for visualization and anomaly detection.

Skills

Programming: Python, R, Bash, Java, C#, C++, MATLAB

Data Science: LLM Agents (LangGraph, RAG, ...), Deep Learning (PyTorch), Natural Language Processing, Quantitative Statistics, SQL / Relational Databases, Visual Analytics, Big Data

Bioinformatics: Analysis of short- and long-read RNA-seq data, protein function prediction, fusion transcript analysis

Tools & Infrastructure: Linux / WSL, Nextflow, HPC (SLURM), Cloud Computing (AWS), Containerization (Docker, Singularity), CI/CD, FAIR principles

Soft Skills: Teamwork, Leadership, Science Communication, Customer Support, Public Presentations

Languages:

- German: Native
- English: Full Professional Proficiency (Near-Native)
- Spanish: Conversational (B2; planning C1 certification)