Mika Senghaas

+45 60740477 | mail@mikasenghaas.de | mikasenghaas.de | github.com/mikasenghaas | linkedin.com/in/mikasenghaas

Data Science Master Student and Excellence Fellow at EPFL. Seeking an industry internship to work along bright people in the broader field of data science and machine learning on projects with impact.

Education

École Polytechnique Fédérale de Lausanne, *MS Data Science* | Lausanne — GPA: 5.8 / 6.0 — 2023-2025
IT University of Copenhagen, *BS Data Science* | Copenhagen — GPA: 11.4 / 12 (Top 1%) — 2020-2023
Gymnasium Rissen, *High School* | Hamburg — GPA: 1.0 (Top 1%) — 2011-2019

Fellowships: EPFL Master Excellence Fellowship
Volunteering: Founder & President AITU (AI Student Group) | ITU Ambassador | Volunteering Barista in Analog
Recent Courses: Modern NLP (CS552) | ML (CS433) | Applied Data Analysis (CS401) | DL in Biomedicine (CS502)

Experience

Data Science Lab (DLAB), EPFL, *Research Assistant* | Lausanne — 2024

- Spearheaded data integration for Project MegaBites under the guidance of Prof. Robert West, targeting a 33% reduction in EPFL's food system carbon footprint in alignment with EPFL's Sustainability Strategy 2030
- Conducted a comprehensive carbon footprint analysis at the ingredient level by scraping and merging diverse data sources, setting a benchmark for sustainability measures
- Led a series of observational studies to decode the dynamics of individual and collective food choices, laying the groundwork for evidence-based intervention designs

Birds At Five/ WDP, *Freelance Data Scientist* | Copenhagen — 2023

- Led the development of copywriting tool to enable brand-aligned content creation through advanced LLM agents
- Leveraged OpenAI APIs and Langchain (chain-of-thought reasoning, semantic search, prompt injection) for building LLM agents capable of generating copy customised for the brand, type of content, and target audience
- Deployed a robust architecture, incorporating FastAPI for backend operations and PostgreSQL database

IT University of Copenhagen, *Teaching Assistant* | Copenhagen — 2021-2023

- Taught in Introduction to Data Science and Programming (Fall 2021), Applied Statistics (Spring 2022/2023), Algorithms and Data Structures (Spring 2022), Linear Algebra and Optimisation (Fall 2022), Machine Learning (Fall 2022), Data Visualisation (Spring 2023), StudyLab (Spring 2022/2023, Fall 2022)

FC St. Pauli, *Youth Football Player* | Hamburg — 2014-2020

- Competed in Germany's premier youth football leagues, advancing from U12 to U19 at FC St. Pauli
- Selected for the extended squad of the German national youth between the U15 and U17 levels

Skills & Extra-Curriculars

Languages	Python, R, SQL, Typescript, C(++), F#, Java, LaTeX
Technical Skills	ML, Statistics, Data Analytics, Data Visualisation, Backend & Database, DevOps
ML	Deep Learning, NLP, CV, GNNs, Few-Shot Learning
Stack	PyTorch, Lightning, Hydra, HF, W&B, Sci-Kit Learn, NumPy, Git, Docker, HPC (SLURM), GCP
Soft Skills	Clear communicator, Analytical thinker, Organised collaborator
Languages	German (Native), English (Fluent), Danish (Basic), French (Basic)
Other Projects	Websites (Portfolio, ITU Coffee Talks, AITU), EDUML, AITU Blog, LauzHack 2023

Selected Projects

Advancing Homepage2Vec with LLM-Generated Datasets for Website Classification [] — 2023

- Led a collaborative initiative between the Machine Learning and Optimization (MLO) and Data Science Lab (DLAB) at EPFL, enhancing the capabilities of Homepage2Vec, an open-source tool for website embedding
- Engineered an innovative approach for the creation of expansive, high-quality datasets tailored for website topic classification, leveraging the power of Large Language Models (LLMs)
- Achieved a notable 4% increase in the Macro F1-score by fine-tuning the model on the LLM-derived dataset, demonstrating a significant improvement in reliability in real-world applications.

Benchmarking SOT Feature Transform for Biomedical Few-Shot Learning Tasks [GitHub, Report] — 2023

- Extended biomedical few-shot learning benchmark by incorporating the SOT (Self-Optimal Transport) feature transform module, evaluated on the task of cell type (Tabula Muris) and protein function prediction (SwissProt)
- Implemented and evaluated common few-shot learning algorithms, like Baseline, MAML, ProtoNet, and MatchingNet, and integrated the SOT module in a 6000 line codebase

Navigating Indoors with Computer Vision - DL Approaches for Indoor Localisation [GitHub, Report] — 2023

- Developed an end-to-end deep learning system for room level localisation for indoor spaces
- Implemented and evaluated various image and video classification models trained on a self-curated video dataset
- Successfully deployed the system on mobile devices via PlayTorch, allowing for real-time indoor navigation and localisation through a user-friendly application