NILS HÖHING

nils.hoehing@gmail.com Dublin nilshoehing.github.io nilshoehing Scholar

PhD Candidate at University College Dublin (ML-Labs) working on benchmarking and spatial reasoning for Vision-Language Models. Experienced in NLP and Computer Vision with both academic and industry exposure. Passionate about building benchmarks that measure real ability rather than superficial cues. Seeking a 3-6 month research or engineering internship in 2026 focussing on benchmarking and evaluation of LLMs/VLMs. Willing to relocate.

EDUCATION

PhD in Machine Learning, *ML-Labs, University College Dublin* 09/22 – now
Research focus on benchmarking and spatial reasoning for Vision-Language Models (VLMs).

Master of Computational Linguistics, *University of Heidelberg* (moved to PhD) 03/22 – 07/22

Bachelor of Computer Science, *Munich University of Applied Sciences* 09/18 – 02/22
GPA 3.6/1.4 US/GER, Thesis 4.0/1.0 US/GER

EXPERIENCE

Teaching Assistant and Tutor, *University College Dublin & University of Applied Sciences Munich* 10/19 – now
Big Data Programming, Introduction to AI, Human Language Technologies, Software Development I & II, Databases, Intelligent autonomous systems.

NLP & Software Engineering Intern, *Alexandrya AI*, Munich 09/20 – 03/21
Built a browser extension for extractive summarisation, deployed via Google Cloud & Kubernetes. Worked with customers to refine semantic search and topic classification pipelines, improving F1-score.

Internship Business Intelligence, *Amadeus Data Processing*, Munich 07/18 – 10/18
Planned and programmed a script that automated the compilation and evaluation of controlling data (excel sheets) thus reducing human workload and speeding up the process.

PUBLICATIONS

Chain-of-Thought (CoT) Reasoning for Solving Puzzles currently in progress
Benchmarking reasoning capabilities of CoT models on puzzles and fine-tuning open source models. Responsible for the evaluation code.

Understanding Space Is Rocket Science - Only Top Reasoning Models Can Solve Spatial Understanding Tasks arxiv
Designed and collected benchmark for spatial understanding in VLMs revealing strong improvements of CoT models over non-CoT models. Conducted human baseline incl. ethics approval.

What's Left Can't Be Right - The Remaining Positional Incompetence of VLMs ICML KLR Workshop 2023 / arxiv
Analysed shortcomings of existing spatial understanding benchmarks and fine-tuned CLIP models with synthetic data on SLURM cluster. Outperformed existing CLIP-like models on the "left" vs. "right" distinction.

PROJECTS

Benchmark Blogpost Blog
Wrote detailed introduction to benchmarks covering main tradeoffs and decisions for benchmark design.

Open source contribution Github
Implemented RocketScience benchmark in openbench following the inspect.ai evaluation framework. Sped up preprocessing by over 90% by replacing unnecessary write operations.

Bachelor's Thesis Thesis
Trained an image captioning model using only text (food reviews) and a pretrained image-text encoder. Demonstrated accurate and creative captions on images of foods.

Autonomous Toy Car Project Team project
Collaborated with a team to collect benchmark data, trained an object detection model, and deployed it using ROS and Docker, improving mAP score and speed. Improved docker layering reducing build time by more than 60%.

Alexa Skill for Dementia Patients Course project
Developed an Alexa Skill that helps elders remember sequences of tasks on AWS, using DynamoDB and S3.

SKILLS, TOOLS & LANGUAGES

Programming Languages Experienced: Python, Java - Familiar: C, C++, Go, Javascript, Matlab, Haskell
Data Science & ML PyTorch, Huggingface, SpaCy, NumPy, Pandas, Inspect AI, OpenBench, SQL, Plotly, OpenCV
Tools & Environments Git, Docker, SLURM, Unit tests, GCP, AWS
Relevant Coursework Semantic Technologies and Knowledge Graphs, Applied Mathematics, Machine Learning, Deep Learning, AWS "Developing on AWS" seminar
Languages German (native), English (C1/TOEFL 110 points), French (B1/DELF)