

Abdu Sallouh

Dresden, Germany
+49 152 59001424 | abdusalouh@gmail.com
abdu.codes | github.com/abduABS | linkedin.com/in/abdusallouh

PROFILE

MSc student in Computational Modelling and Simulation at TU Dresden (graduating around December 2026), focused on large language models, with computer vision as a secondary interest. First author of an award-nominated EMNLP 2026 paper on Arabic dialects in LLMs and of an ICICT 2026 paper on synthetic data for rare-species detection. Writing a thesis on how reliably LLMs materialize knowledge.

SKILLS

AI Focus: Large Language Models, Natural Language Processing, Computer Vision, Machine Learning

ML/LLM: Python, PyTorch, HuggingFace, Scikit-learn, pandas, NumPy, OpenCV

Programming / Tools (coursework level): C/C++, Java, C#, JavaScript, SQL, Git, Docker

Languages: English (C1), German (B2+), Arabic (native)

RESEARCH & PUBLICATIONS

Does Modern Standard Arabic (MSA) Dominate Arabic Dialects in LLMs? A Representation-Level Analysis | EMNLP 2026 Main Conference (ACL), short paper (award nominated)

10/2025 – 10/2026 | First author, co-authored with Nicholas Popovič and Michael Färber | Presented as a poster at EMNLP 2026

Paper: aclanthology.org/2026.emnlp-main.2178 | Code: github.com/abduABS/MSA_Dominance_LLM

- Adapted the Gaussian mixture model (GMM)-based language-dominance framework of Shani and Basirat (2025) to a cross-dialectal setting, measuring representational separability (NMI) and directional dominance (token-level likelihood ratio) across 26 Arabic varieties (MADAR corpus).
- Evaluated five LLMs of up to 8B parameters (Llama 3.1-8B, Qwen2.5-7B, BLOOM, mGPT, ArabianGPT-0.8B) on TU Dresden's Capella HPC cluster; found low representational separability (mean NMI 0.1 to 0.2) and, under this framework, no evidence that MSA or any dialect is a dominant internal representation, supported by replication and robustness checks.

Master's Thesis: Probing Knowledge Materialization in Large Language Models | ScaDS.AI

05/2026 – 12/2026 (expected completion)

- Building an evaluation harness to measure whether LLM-generated outputs reflect genuine, reliable model knowledge, combining deterministic and model-based scoring across multiple dimensions.
- Testing consistency of extracted assertions across six task variants, several using paraphrased prompts, and applying the framework to knowledge materialization systems including GPTKB.

Closing the Rare-Species Detection Gap in UAE Camera Traps with Diffusion-Generated Data | ICICT 2026, Springer, pp. 342–351

09/2023 – 02/2026 | First author, with 7 co-authors from American University of Sharjah and Emirates Nature-WWF

Paper: doi.org/10.1007/978-3-032-35982-7_30 | Project page: abdu.codes/CameraTraps

- Fine-tuned DreamBooth Stable Diffusion to generate ~700 synthetic images per class for five rare species (together under 6% of ~67,000 camera-trap images across 15 classes); detection F1 for rare classes improved for some species, with fluctuating results across classes.

WORK EXPERIENCE

Research Assistant (WHK) | DaTNet (DatenTreuhand Kompetenznetzwerk) *Dresden*
03/2025 – 06/2026

- Collected, organized, and presented research outputs of a BMBF-funded network on data trust and governance models, primarily by redesigning and maintaining the organization's website (datnet.eu).
- Moderated live workshops where network researchers presented their work on data sovereignty and regulatory frameworks to practitioners.

Digital Energy Intern | Schneider Electric *Dubai, UAE*
06/2023 – 07/2023

- Built a GPT-based customer support chatbot as an end-to-end GenAI prototype, iterating on answer quality through prompt engineering and systematic evaluation.
- Won an internal innovation and security competition with the solution and presented the final system directly to the regional VP.

IT Intern (Hybrid) | Dastech IT Solutions *Dubai, UAE*
07/2022 – 09/2022

- Analyzed a small business's infrastructure needs and implemented a cloud-based solution to support its setup and deployment.

PROJECTS

Cicero: AI Public Speaking Coach | Final Year Project, American University of Sharjah
09/2023 – 07/2024

- Built and deployed Cicero, an end-to-end multimodal GenAI web application (GPT-4, OpenCV, speech processing) delivering automated public speaking feedback.
- Designed pipelines for speech, facial expression, and transcript analysis to generate personalized recommendations.

Reinforcement Learning: 2D Racing Game | Technical University of Dresden
04/2025 – 08/2025 | Code: github.com/abduABS/CMSCarProject

- Developed an autonomous 2D racing agent using Proximal Policy Optimization (PPO) and Stable-Baselines3 for efficient navigation and obstacle avoidance.

EDUCATION

Technical University of Dresden *Dresden, Germany*
10/2024 – 12/2026 (expected completion, all coursework completed)

M.Sc. Computational Modelling and Simulation (Visual Computing), German grade: approx. 2.1 (excl. thesis)

American University of Sharjah *Sharjah, UAE*
08/2020 – 06/2024

B.Sc. Computer Science, Minor in Data Science, GPA: 3.61/4.0

Honors: Cum Laude; Tau Beta Pi and IEEE-Eta Kappa Nu (engineering honor societies)