

Mohamed Amine Ketata

Seeking ML research internships, available anytime in 2027 for up to six months

a.ketata@tum.de

 Website ·  Google Scholar ·  GitHub ·  LinkedIn

Education

- Oct 2024–Present **PhD**, *Machine Learning*, Technical University of Munich (TUM), Germany
- Research focus: *Generative modeling for structured data (e.g., tables, relational databases, and graphs)*.
 - Supervisor: Prof. Stephan Günnemann; **expected graduation: 09/28**.
- 2021–2024 **MSc**, *Robotics, Cognition, Intelligence*, TUM, Germany
- Passed with high distinction; final grade: 1.0 (best possible grade); **top 0.8%** of 364 graduates.
- Aug–Dec 2021 **Exchange Semester**, University of Illinois Urbana-Champaign, USA
- Selected coursework: Deep Learning Theory, Statistical RL; GPA: 4.0/4.0.
- 2018–2021 **BSc**, *Electrical Engineering and Information Technology*, TUM, Germany
- Passed with high distinction; final grade: 1.0 (best possible grade); **top 0.3%** of 666 graduates.

Publications *Equal contribution

Conference Publications

- C1. **Ketata, M. A.**, Lüdke, D., Schwinn, L., & Günnemann, S. *Joint Relational Database Generation via Graph-Conditional Diffusion Models*. **NeurIPS 2025**. [Paper](#) ·  [Code](#)
- C2. **Ketata, M. A.***, Gao, N.*, Sommer, J.*, Wollschläger, T., & Günnemann, S. *Lift Your Molecules: Molecular Graph Generation in Latent Euclidean Space*. **ICLR 2025**. [Paper](#) · **Oral presentation** and **Best Paper Runner-up** at [GRaM Workshop](#) (ICML 2024).
- C3. Wollschläger, T., Gao, N., Charpentier, B., **Ketata, M. A.**, & Günnemann, S. *Uncertainty Estimation for Molecules: Desiderata and Methods*. **ICML 2023**. [Paper](#) ·  [Code](#)

Preprints and Workshop Papers

- W1. **Ketata, M. A.**, Schambach, M., & Günnemann, S. *CDMD: A Cross-Dataset Mixed-Type Diffusion Model for Tabular Data*. **Preprint, 2026** (under review). [Paper](#) · Accepted at [BeNTo Workshop](#) (NeurIPS 2026).
- W2. **Ketata, M. A.***, Laue, C.*, Mammadov, R.*, Stärk, H., Wu, M., Corso, G., Marquet, C., Barzilay, R., & Jaakkola, T. S. *DiffDock-PP: Rigid Protein-Protein Docking with Diffusion Models*. [Machine Learning for Drug Discovery Workshop](#) (ICLR 2023). [Paper](#) ·  [Code](#)

Research Experience

- Oct 2024–Present **Doctoral Researcher**, [Data Analytics and Machine Learning Group](#), TUM
- Led two research projects in collaboration with [SAP Business AI Research](#), from conception through model design, implementation, experimental evaluation, and manuscript writing.
 - Developed a diffusion model for relational databases that uses a graph neural network to jointly denoise attributes of interconnected rows; outperformed existing baselines in capturing multi-hop correlations across six real-world databases. [\[C1\]](#)
 - Developed a mixed continuous-discrete diffusion model for joint training across heterogeneous tables; outperformed baselines in average generation quality on seven datasets using one shared model; improved generation in limited-data settings via pretraining on 337 datasets. [\[W1\]](#)
 - Currently developing a diffusion-based tabular foundation model, targeting in-context synthetic data generation and a unified framework for generative and predictive tasks.
- May 2023–Jan 2024 **MSc Thesis**, Data Analytics and Machine Learning Group, TUM
- Developed a molecular graph generative model based on E(3)-equivariant diffusion in a synthetic 3D latent space; achieved over 26% higher normalized FCD score on ZINC250K than the strongest evaluated non-autoregressive baseline. [\[C2\]](#)

Oct 2022–Apr 2023 **Research Project**, TUM Data Innovation Lab

- Led a team of three students collaborating with MIT researchers to develop DiffDock-PP, a diffusion model for rigid protein-protein docking combining diffusion over relative translations and rotations with an equivariant graph neural network. [W2]

Industry Experience

Feb–Jul 2024 **Deep Learning Engineering Intern**, QuantCo

- Fine-tuned language models for intent classification and dense retrieval for a German insurer; improved top-1 classification accuracy by approximately 11% relative to the existing system on an internal test set and enabled adding new response candidates without model retraining.

Apr 2022–Dec 2023 **AI Research Engineer** (part-time), Helsing

- Developed a method for selecting data augmentation hyperparameters using embedding distances, avoiding repeated model training; co-inventor on a [published patent application](#).
- Implemented CLIP-based filtering to curate a LAION-400M subset for training self-supervised object detection models.

Mar 2020–Jul 2021 **SWE Intern** (Mar–May 2020) and **AI Junior Consultant** (part-time), TNG

- Trained a ResNet model and developed a real-time computer vision system on a Raspberry Pi to detect deviations from planned item trajectories for a pharmaceutical client.
- Contributed [three merged pull requests](#) to ROS 2 in C++, adding playback controls to rosbag2 and support for compressed image streams in RViz.

Teaching & Supervision

2025–Present **Instructor** (Exercises & Seminars), TUM

- [Machine Learning](#) (Winter 2025/26), [Advanced ML: Deep Generative Models](#) (Summer 2025, 2026): Led exercise and Q&A sessions; prepared exam questions.
- Seminar: Selected Topics in Machine Learning Research (Winter 2026/27): Co-organizing the seminar, selecting papers, and supervising two MSc students on tabular foundation models.

2024–Present **Research Supervisor**, TUM

- Supervised one BSc thesis and one guided research project; co-supervised two MSc theses. Topics: relational deep learning and generative models for tabular, relational, and graph data.

2019–2022 **Teaching Assistant** (part-time), TUM

- Led tutorials in Linear Algebra and Analysis; graded exercises and exams for ML.

Awards & Honors

2026 **Top Reviewer Award**, NeurIPS 2026.

2022–Present **relAI Funding (DAAD)**: MSc scholarship and full doctoral funding.

2017–2024 **Tunisian Government Scholarship** supporting studies in Germany.

2017 **Presidential Award for Academic Excellence**, Tunisia: Ranked **1st nationally** in the Baccalaureate mathematics track (11,826 registered candidates).

Academic Service

- **Conference Reviewer**: NeurIPS (2025, 2026); ICLR (2026, 2027); LoG (2026).
- **Workshop Reviewer**: DeLTa (ICLR 2025); BeNTTo (NeurIPS 2026).

Skills & Languages

- **Programming**: Python (primary), C++ (prior experience).
- **ML & scientific computing**: PyTorch, PyTorch Geometric, scikit-learn, NumPy, pandas.
- **Research tools**: Linux, Git, Slurm, Weights & Biases (wandb).
- **Languages**: Arabic (native); English, German, French (fluent).