Dr. Benjamin Matthias **Ruppik**

Deep Learning Researcher · Using Topological Data Analysis to Evaluate LLMs and Augment NLP Systems

ruppik.net | ben300694 | benjamin-matthias-ruppik | @ben300694 | Benjamin's Google Scholar

Summary

Machine Learning Researcher, working on **large language models**, **representation learning**, and **unsupervised evaluation**. As a Postdoctoral researcher in the *Dialog Systems and Machine Learning* Lab at Heinrich Heine University Düsseldorf, I am involved in the technical work of research projects covering natural language understanding, dialogue state tracking, RL policy learning, knowledge & ontology extraction, and natural language generation. My main expertise is in **Topological Deep Learning** and I use **Topological Data Analysis (TDA)** to study the latent spaces of large language models, with applications to task-oriented dialogue systems. I completed a **Ph.D.** in **low-dimensional topology** and knot theory at the *Max Planck Institute for Mathematics* in Bonn, resulting in multiple peer-reviewed publications in mathematics research journals.

Work Experience

Heinrich Heine University Düsseldorf, Dialog Systems and Machine Learning Lab

Düsseldorf, Germany

Postdoctoral Researcher (Lab Chair: Prof. Milica Gašić) *Jan. 2022 - Present*

- First author and technical lead on **NeurIPS 2025 publication** introducing novel framework using local intrinsic dimension estimates to analyze LLM representations, linking **dimension shifts** to **generalization, grokking, and training dynamics**.
- Lead **international research collaboration** with TUM, Helmholtz AI Munich and University of Fribourg on topological deep learning for NLP.
- Developed and fine-tuned **transformer-based language models** using PyTorch and Hugging Face (transformers, PEFT for LoRA fine-tuning); implemented **topological data analysis pipelines** with gudhi, giotto-tda, and custom Python modules for studying contextual embedding spaces.
- Managed large-scale **ML experiment grids** on **HPC clusters** (Hilbert at HHU) and GCP **cloud infrastructure**; orchestrated parallel training jobs using **PBS and SLURM schedulers**, Skypilot, and `hydra` launchers; implemented vector databases with `faiss` for **nearest neighbor search**.
- First author on SIGDIAL 2024 paper **(Nominated for Best Paper Award)** on dialogue term extraction using local topology measures in embedding spaces of transformer-based language models.
- Co-author on **10+ publications** in top-tier venues (TACL, AAAI, ACL, SIGDIAL, SIGIR, IEEE/ACM TASLP) covering LLM evaluation, emotion integration in dialogue systems, uncertainty estimation, noisy label learning, automatic knowledge extraction, and ontology construction.
- **Presented research** at top-tier **international conferences** (NeurIPS 2025 San Diego, SIGDIAL 2024 Kyoto, SIGDIAL 2022 Edinburgh) and delivered invited talks at ETH Zürich, MIT CSAIL, Columbia University NLP Seminar, University of Pittsburgh, and KTH Stockholm.
- **Supervised** master's students on **AI/ML thesis projects**; Example projects: "Hallucination detection in LLMs via topological data analysis"; "Topological autoencoders for dialogue data".
- **Designed and taught** graduate-level courses on word embedding spaces, representation learning and topological data analysis (planned for next term) for Master CS and AI & Data Science programs.

Education

Max Planck Institute for Mathematics, University of Bonn

Bonn, Germany

Ph.D. in Mathematics (magna cum laude) *Oct. 2018 - Jun. 2022*

- Funded by International Max Planck Research School on Moduli Spaces; member of Bonn International Graduate School of Mathematics.
- **Published 7 peer-reviewed papers** in mathematics research journals (Algebr. Geom. Topol., Pacific J. Math., Proc. Amer. Math. Soc.) on low-dimensional topology, knot theory and knotted surfaces in 4-dimensional manifolds.
- **Teaching assistant** for undergraduate and graduate-level courses in Analysis, Linear Algebra, Algebraic Topology.

University of Bonn

Bonn, Germany

M.Sc. & B.Sc. in Mathematics, Minor in Computer Science *Oct. 2013 - Aug. 2018*

- Relevant coursework: Linear and Integer Optimization (C/C++), Chip Design (C++), Functional Programming (Haskell), Systems-Level Programming (Assembly, Network programming with sockets, Go), Computer Vision (OpenCV, Python).
- Student associate in **Computer Vision** lab (Project: Semantic segmentation of RGB images and point clouds from LiDAR sensors, using ResNets in TensorFlow and additional tooling in MATLAB).

Key Publications

Less is More: Local Intrinsic Dimensions of Contextual Language Models

NeurIPS 2025

First Author

- **Unsupervised** evaluation framework using **local intrinsic dimension** estimates to analyze **LLM representations**, linking dimension shifts to **generalization, grokking, and training dynamics**. **Code:** aidos-lab/Topo_LLM_public; **Paper:** *Neural Information Processing Systems.*

Local Topology Measures of Contextual Language Model Latent Spaces With Applications to Dialogue Term Extraction

SIGDIAL 2024 (Best Paper Award Nominee)

First Author

- **Augments** language model based sequence tagging models with **local topology measures** in contextual embedding spaces for dialogue **term extraction**. **Code:** tda4contextualembeddings-public; **Paper:** *ACL Anthology.*