

Few-shot Robustness Benchmark and Evaluation

- **Supervisor:** Abdullatif Köksal
- **Examiner:** Prof. Schütze
- **BSc, MSc, Open:** BSc/MSc
- **Title:** Few-shot Robustness Benchmark and Evaluation.
- **Summary:** In this project, we will investigate the current state of few-shot learning approaches that make use of prompts/instructions, such as in-context learning, multitask instruction tuning, and prompt-based tuning. Despite the success of these methods, it has been observed that they can exhibit high variability across different features, such as prompt design [1], verbalizer design [1], selected few-shot examples [2], and randomization during finetuning [2]. The proposed research aims to explore and expand upon these features and create a benchmark for evaluation using existing datasets. Additionally, a set of evaluation metrics will be introduced and existing approaches, such as GPT-3/BLOOM/OPT for in-context learning, Flan-T5/BloomZ for instruction tuning, and PET/OpenPrompt for prompt-based tuning, will be analyzed.
- **Prerequisites:** Experience in Pytorch/HuggingFace.

References

- [1] True Few-Shot Learning with Prompts—A Real-World Perspective, https://direct.mit.edu/tac/article/doi/10.1162/tac/a_00485/111728
- [2] MEAL: Stable and Active Learning for Few-Shot Prompting, <https://arxiv.org/abs/2211.08358>