Towards A Unified Multi-Domain Framework for Dialogue Generation

- [ ] Supervisor: Yongkang Liu
- [ ] Examiner: Prof. Schütze
- [ ] BSc, MSc, Open: MSc/BSc
- [ ] Summary: Recently, dialogue systems of specific domains have received extensive attention in the research community. A large-scale corpus is indispensable for dialogue systems. Unfortunately, domain-specific dialogue materials are usually not available due to security and privacy concerns. One line is to improve the performance of the target domain (i.e., corpus insufficient side) through Domain Adaptation. Domain adaptation usually requires us to find a source domain with a rich corpus for cross-domain transfer, which limits the popularity of dialogue systems in specific domains. We observe that large-scale pre-trained language models have achieved excellent performance in open-domain dialogue generation, such as DialoGPT, BART, Meena, and ChatGPT. We believe that language models have great potential in specific domains. In this project, having a source domain with a rich corpus is not a requirement. We will explore the potential of large-scale language models in multiple low-resource domains, or even zero-resource domains.
- [ ] Prerequisites: enthusiasm, Good programming background (preferably python), basic knowledge of NLP, Generation task, and Pytorch