# Generation of German Text Controlled by Sentiment and Keywords

Paulina Aleksandra Żal

Master's Thesis
Master of Science in Applied Information and Data Science
School of Business
Lucerne University of Applied Sciences and Arts

Wettingen AG, 22nd of December 2023


| | |
|---|---|
| Author: | Paulina Aleksandra Żal<br>paulina.zal@stud.hslu.ch |
| Supervisor: | Dr. Guang Lu<br>guang.lu@hslu.ch |
| Co-Supervisor: | Dr. Nianlong Gu<br>nianlong@ini.ethz.ch |

# Management Summary

The introduction of transformer models into Natural Language Processing (NLP) led to Pre-trained Language Models (PLM), which are constantly becoming more attractive for a variety of today's use cases. Due to their power, these models can be employed in Controllable Text Generation (CTG) to exert precise control over attributes like sentiment or topic, advancing text customization in various applications. While proficient, the probabilistic nature of those models may lead to outputs not precisely aligned with user intent. Emerging methods such as Fine-Tuning and Plug-and-Play Models (PPLMs) empower users to steer text generation, aiming for more tailored outputs, particularly in marketing and content creation domains.

This thesis contributes by refining text generation techniques, ensuring outputs adhere to predefined guidelines related to both topic and sentiment. The aim is to modify a pre-trained GPT-2 model in German, enabling it to produce paragraphs derived from an input sequence, a sentiment token, and a specific keyword set.

The pipeline consists of Supervised Fine-Tuning (SFT), optimization with Reinforcement Learning (RL), and a Logit Modification mechanism. SFT is utilized to adapt the GPT-2 model to generate text concerning the specified sentiment token. RL optimizes the model and improves the text generation in terms of sentiment. The model is rewarded by scores from a Sentiment Discriminator that was previously created. In the last step a logit modification mechanism from Pascual et al. (2020) is adapted and tested with different decoding strategies.

SFT and RL show promising results in generating sentiment-based text. However, implementing keyword control mechanisms reduces the performance in terms of keyword utilization compared to the GPT-2. The human assessment indicates that texts generated with sentiment tokens exhibit moderate fluency and coherence. The introduction of keyword control negatively impacts fluency and coherence. Using the method may save time and effort during the creative process of content creation. However, the proposed pipeline is not yet suitable for practice, as it faces some limitations in generating coherent and fluent texts with a specific set of keywords. The author believes that keyword control could be improved within the SFT and RL processes to achieve a balanced emphasis on both sentiment and keywords in Fine-Tuning models.

***Keywords.*** Natural Language Processing, Controllable Text Generation, Pre-trained Language Models, Supervised Fine-Tuning, Reinforcement Learning

# Contents