CourtPressGER: Generating German Court Press Releases with LLMs

Sebastian Nagl[1][0000−0000−0000−0000], Mohamed Elganainy[1][0000−0000−0000−0000], and Matthias Grabmair[1][0000−0000−0000−0000]

Technical University of Munich, Munich, Germany
{sebastian.nagl,mohamed.elganainy,matthias.grabmair}@tum.de

Abstract. This paper presents CourtPressGER, a system for automatically generating German court press releases using Large Language Models (LLMs). We present a curated dataset with 6.4k entries of court decisions with corresponding press releases from Germany's highest courts. The dataset is enhanced with synthetic prompts that enable automated generation of press releases from court decisions. We describe a pipeline for generating press releases with various state-of-the-art models and evaluate the results using human and automated metrics. Our approach combines specialized legal language models with domain-specific techniques to produce accurate and informative press releases that adhere to journalistic and legal standards.

Keywords: Legal NLP · Legal AI · Court Press Releases · German Legal Text Generation · Large Language Models

1 Introduction

The German legal system consists of a complex network of courts that regularly publish extensive decisions. To make these decisions accessible to the public, the highest courts create press releases that summarize the essential aspects and implications of the decisions in an understandable form. These press releases serve as an important interface between the judicial system and the general public by explaining complex legal matters in an accessible way.

However, the manual creation of such press releases requires significant resources. At the same time, recent advances in Large Language Models (LLMs) offer new possibilities for automated text generation in specialized domains. Our project CourtPressGER aims to leverage these capabilities for the automatic generation of court press releases.

1.1 Objectives

The main objectives of our project are:

- The creation of a curated dataset with 6.4k entries of court decisions with corresponding press releases from Germany's highest courts.

- The development of synthetic prompts for each decision-press release pair that can be used to automatically generate press releases.
- The implementation of a pipeline that queries various LLMs with the synthetic prompts and stores the results alongside the actual press releases.
- The evaluation of the generated press releases using human and automated metrics.

2 Related Work

2.1 Juristic Text Generation

In recent years, the automatic generation of juristic texts has made significant progress with the emergence of powerful language models. Various studies have focused on summarizing juristic documents, generating juristic arguments, and simplifying complex legal texts.

2.2 Large Language Models in the Juristic Context

LLMs have increasingly been used for juristic applications, including legal advice, document analysis, and decision predictions. Current research explores the ability of LLMs to understand and generate legal language, as well as their reliability and ethical implications in the legal context.

2.3 Press Releases in the Judicial System

Studies on the role and impact of court press releases have highlighted their importance for public perception and understanding of legal decisions. However, there has been little research on the automated generation of such releases, especially in the German legal system.

3 Dataset

3.1 Data Sources

Our dataset includes court decisions and corresponding press releases from Germany's highest courts, including the Federal Constitutional Court, the Federal Court of Justice, and other federal courts. The data was collected from publicly accessible sources.

3.2 Data Cleaning

The data cleaning process involved a combination of rule-based methods and semantic similarity analysis. The process included:

- Filtering and normalizing raw data
- Validating the mappings between decisions and press releases
- Removing duplicates and inconsistent entries

3.3 Dataset Statistics

The cleaned dataset contains 6.4k pairs of court decisions and press releases. The average length of decisions is X Tokens, while press releases average Y Tokens in length. The dataset covers various legal fields, including civil, criminal, administrative, and constitutional law.

4 Methodology

4.1 Synthetic Prompts

For each decision-press release pair, we developed synthetic prompts that serve as input for LLMs to generate press releases. These prompts were carefully designed to highlight the key aspects of the decision and to train the models to create relevant and precise press releases.

4.2 Pipeline for Generating Press Releases

Our pipeline includes various LLMs, which can be categorized into three groups:

- Large Models: GPT-4o, Llama-3-70B
- Small Models: Teuken-7B, Llama-3-8B, EuroLLM-9B

The pipeline is designed to send the synthetic prompts to the models, collect the generated press releases, and store them alongside the actual press releases. A checkpoint system allows for the continuation of interrupted generation processes.

4.3 Evaluation Metrics

We use various metrics to evaluate the generated press releases:

- ROUGE (Rouge-1, Rouge-2, Rouge-L)
- BLEU (BLEU-1 to BLEU-4)
- METEOR
- BERTScore (with EuroBERT model)

In addition to these automated metrics, we conduct a human evaluation where experts assess the quality, accuracy, and understandability of the generated press releases.

5 Results

5.1 Comparison of Models

[Here follow detailed results of the various models based on automated metrics.]

5.2 Human Evaluation

[Here follow the results of the human evaluation, including qualitative observations and quantitative assessments.]

5.3 Analysis of Case Studies

[Here, selected examples are analyzed that highlight specific strengths or weaknesses of the generated press releases.]

6 Discussion

6.1 Interpretation of Results

[Here follows a discussion of the main findings and their implications.]

6.2 Strengths and Weaknesses

[Here, we discuss the strengths and weaknesses of our approach, including specific challenges in generating legal texts in German.]

6.3 Ethical Considerations

[Here, we discuss ethical issues in the context of automated generation of court press releases, including transparency, accountability, and potential biases.]

7 Conclusion and Future Directions

Our project CourtPressGER demonstrates the potential of LLMs for automated generation of court press releases in the German legal system. The results show that modern language models are capable of producing understandable and informative press releases that accurately represent complex legal matters.

Future research directions include the integration of domain-specific knowledge bases, the improvement of accuracy and reliability through fine-tuning of the models, and the extension of the approach to other languages and legal systems.

Acknowledgments. This research was supported by [Funder/Project Number, if applicable]. This paper will be presented at the ICAIL Workshop 2025.

Disclosure of Interests. The authors declare that they have no conflicts of interest that are relevant to the content of this paper.

References

1. Author, A.: Article Title. Journal **2**(5), 99–110 (2016)
2. Author, B., Author, C.: Title of a Conference Paper. In: Editor, D., Editor, E. (eds.) CONFERENCE 2016, LNCS, vol. 9999, pp. 1–13. Springer, Heidelberg (2016). https://doi.org/10.10007/1234567890
3. Author, F., Author, G., Author, H.: Book Title. 2nd edn. Publisher, Location (1999)
4. Author, I.: Contribution Title. In: 9th International Conference, pp. 1–2. Publisher, Location (2010)
5. German Federal Court Website, http://www.example.de, last accessed 2024/06/10