PhD in Linguistics, Universities of Bozen-Bolzano and Verona. Project-specific PhD scholarship funded by Eurac Research (http://www.eurac.edu)

Title

AI for Legal Text Quality Evaluation: Automatic Evaluation of Terminology and Overall Accuracy in a Lesser-used Standard Variety of German with Assessment of Generated and Translated Text

Description

Generative Artificial Intelligence (GenAI) has experienced a surge of fields of applications, including specialized communication and translation. Language models (LM) are at the core of GenAI and are trained on massive amounts of data. The quality and pertinence of data are central to aligning applications to their intended use. Achievements for under-resourced languages and language varieties have been comparatively modest, as their language data are scarce, easily outnumbered by major language varieties, and hard to mine.

In the South Tyrolean multilingual context, the use of GenAI applications for text drafting and translation poses notable challenges concerning the specific standard German variety used by all public authorities, which shows remarkable conceptual and terminological incongruence with respect to other German-speaking legal systems. This is due to the system-boundedness of legal language. In addition, law and administration are high-stakes domains where errors may have serious financial and personal consequences. It is therefore essential that GenAI tools are able to align with the official standardized terminology (DPR no. 574/1988) and customary use of the local variety.

The core research question is to what extent recent advances in AI can support legal drafting and translation evaluation across language varieties in non-English languages and language combinations. Currently, there is a research gap in the design of automated evaluation frameworks in line with the peculiarities of specific language varieties in monolingual and multilingual settings. Existing solutions consist of string-based metrics and neural metrics. However, their performances on non-English language combinations and minor language varieties are yet to be demonstrated. An approach to filling the gap for German language varieties could leverage existing data from the Information System for Legal Terminology *bistro* (https://bistro.eurac.edu/) with the purpose of fine-tuning an AI model as a custom evaluation metric.

Preferred requirements

- Experience with using and training machine learning models, including Transformers and LLMs.
- Knowledge of theory, practices and most used metrics of Natural Language Generation Evaluation, especially in the field of Machine Translation.
- Statistical testing.
- Basic theoretical knowledge of terminology and termbase management.
- Familiarity with the processes of Machine Translation adaptation pipeline and the latest research in Machine Translation.