LLM TOOL: A HYBRID PIPELINE FOR AUTOMATED LARGE-SCALE TEXT ANNOTATION USING LOCAL LANGUAGE MODELS AND BERT CLASSIFIERS

Authors

Lemor, Antoine, Jérémy Gilbert, Shannon Dinan, and Yannick Dufresne

Abstract

The annotation of large-scale text corpora represents a fundamental bottleneck in computational social science research, particularly when dealing with complex multi-label classification tasks in political science. We present LLM Tool, a novel hybrid pipeline that combines local Large Language Models (LLMs) with BERT-based classifiers to enable fully automated annotation at scale. Our approach leverages state-of-the-art open-source LLMs (Gemma3:27B, Llama3.3:42B, Nemotron:42B, DeepSeek-R1:70B, GPT-OSS:120B) running entirely on local infrastructure to generate initial annotations on stratified samples, which then serve as training data for specialized BERT models capable of efficient large-scale inference. The pipeline implements an extended version of the Comparative Agendas Project (CAP) coding scheme adapted for Canadian political discourse, generating structured annotations across 21 policy themes, 9 political parties, 2 specific themes, and 3 sentiment dimensions. Through rigorous empirical validation on 1,593 manually annotated Canadian parliamentary debates and media articles, we demonstrate that BERT models trained on LLM-generated annotations significantly outperform those trained on human annotations, achieving a Micro F1 score of 0.6673 compared to 0.4601 for human-annotation-trained models—a 45\% improvement. We validate this finding across three dataset configurations: Small (1,343 sentences), Large (5,753 sentences), and Extra-Large (12,000 sentences with complete test isolation, still pending). The Extra-Large configuration, which ensures zero contamination between training and test data, will soon aim to confirm the robustness of our approach.