Report on the outcomes of a Short-Term Scientific Mission[1]

Action number: CA21129

Grantee name: Ljubisa Bojic

Details of the STSM

Title: LLMs as a Tool for Textual Analysis

Start and end date: 07/10/2024 to 31/10/2024

Description of the work carried out during the STSM

During the Short-Term Scientific Mission (STSM) hosted by the University of Montenegro, we embarked on an intensive research endeavor aimed at deepening our understanding of large language models (LLMs) and their efficacy as tools for textual analysis. This mission was conceived as part of an ongoing research initiative that began at the University of Salamanca, structured under the COST Action CA21129 - OPINION, which focuses on integrating theory and methods for automatically analyzing opinionated communication.

Our primary objective during this mission was to complete and extend the foundational research initiated in Spain. We began by inquiring into the dataset previously curated at the University of Salamanca, which included a comprehensive collection of text analyses. These analyses were performed by two distinct groups—human participants and LLMs—allowing us to conduct a comparative evaluation of their outputs. Working closely with my host, Prof. Vuk Vukovic, we employed statistical techniques to scrutinize this dataset further. Utilizing advanced tools such as SPSS and Python libraries, we confirmed results of the previously conducted analysis such as inter-rater reliability and consistency scores. These methods were pivotal in validating our initial findings, which noted a remarkable alignment between the qualitative analyses of humans and the automated outputs generated by LLMs. However, except running analysis again and doing additional ones, the main work effort was related to writing multiple versions of the manuscripts. Main aspects of our work included aspects of the manuscript such as literature review, methodology, results and discussion. Our final output is draft that can be further improved towards submission to the journal. We sent the draft to the core team of authors for further refinement.

A critical aspect of our research involved not only validating our findings but also disseminating them effectively to the academic community. Hence, a considerable portion of our time was devoted to refining the manuscript that encapsulates our research insights. This entailed a rigorous process of integrating constructive feedback that we had received from various workshops previously conducted. We focused

[1] This report is submitted by the grantee to the Action MC for approval and for claiming payment of the awarded grant. The Grant Awarding Coordinator coordinates the evaluation of this report on behalf of the Action MC and instructs the GH for payment of the Grant.

on incorporating additional data analyses to bolster our manuscript's robustness, aiming to achieve the quality necessary for publication in leading, high-impact academic journals.

Beyond the confines of our direct research efforts, the STSM was characterized by a strong commitment to knowledge transfer. I led a series of informal workshops at the University of Montenegro, designed to equip colleagues and students with the latest methodologies and analytical techniques emerging from our research. This educational outreach not only facilitated a rich exchange of ideas but was also critically enhanced by the insightful feedback and perspectives offered by Prof. Vukovic and his team. Their contributions were invaluable in refining our analytical processes, ultimately strengthening the collaborative bond between our institutions and paving the way for future research synergies.

The activities and efforts carried out during this STSM were firmly aligned with the initially proposed work plan. Our adherence to this roadmap ensured that we achieved our objectives without significant deviations. The mission not only furthered our specific research goals but also contributed to the broader aims of the COST Action CA21129, encapsulating the essence of collaborative and interdisciplinary research that transcends geographical and institutional boundaries.

Description of the STSM main achievements and planned follow-up activities

The Short-Term Scientific Mission achieved significant milestones in evaluating the potential of Large Language Models for latent content analysis across various dimensions such as sentiment, political leaning, emotional intensity, and sarcasm detection. Our research involved comparisons between human annotators and LLMs, specifically employing 33 human participants and eight different LLMs, including advanced versions of GPT-4, Gemini, Llama, and Mixtral. Each participant analyzed 25 texts for each category, resulting in a comprehensive dataset comprising 3,300 human annotations and 19,200 LLM annotations.

One of the critical achievements of this mission was the validation of LLMs' capability to match and, in some cases, surpass human performance in specific analytical tasks. Using Krippendorff's alpha to measure inter-rater reliability, we found that both humans and LLMs exhibited high consistency in sentiment analysis ($\alpha \approx 0.95$). Interestingly, LLMs outperformed human annotators in assessing political leaning (LLMs: $\alpha \approx 0.80$ compared to humans: $\alpha \approx 0.55$) and emotional intensity (LLMs: $\alpha \approx 0.85$ compared to humans: $\alpha \approx 0.65$), underscoring their potential utility in these areas. However, both groups struggled with sarcasm detection, highlighting an area for future research and model enhancement.

Despite these findings, variations were evident, particularly in emotional intensity ratings, where humans provided higher and more varied scores than LLMs ($p < 0.001$, Cohen's $d = 0.880$). Notably, GPT-4 was the LLM that most closely aligned with human performance across categories, although disparities in emotional intensity persisted, emphasizing the continued need for human expertise in nuanced emotional interpretation.

The outcomes of this STSM underscore the growing role of LLMs in achieving consistent and high-quality content analysis. These insights contribute directly to the objectives of COST Action CA21129 - OPINION, pushing the boundaries of automated opinion analysis methods. Our findings indicate promising pathways for further refinement and optimization of LLMs, particularly in areas like sarcasm detection and emotional nuance.

We plan to build on this work by engaging in further collaborations with the University of Montenegro to explore advanced model training techniques and integrating insights drawn from human psychology to improve LLM interpretative capabilities. This collaboration will also focus on mentoring early-career researchers, equipping them with the skills necessary to leverage LLMs for comprehensive textual analysis. Through continued cooperation and shared expertise, we aim to enhance the applicability and accuracy of LLMs, reinforcing the interdisciplinary and collaborative ethos of the OPINION network.