Large Language Model-Based Speech Act Classification of Information Security Policy Statements

Leila Aro-Sati, Fredrik Karlsson, Shang Gao

Department of Informatics, School of Business, Örebro University, Sweden

Abstract

Information security policies (ISPs) are widely considered as the primary formal control by which organisations communicate to employees their expected information security behaviour. Despite the central role of ISPs within organisations, prior research has shown that ISPs' content frequently encounter ambiguity, vagueness, and lack of actionable advice. Large language models (LLMs) can offer new opportunities for assisting in the analysis of ISPs and, as a result, help in suggesting improvements of ISP statements (sentences from ISP). To do that, it is necessary to establish whether LLMs can reliably distinguish between different types of ISP statements. In this study, we draw on Speech Act Theory that explores how language is used to perform actions, not just to convey information. As such, we investigate how LLAMA3.3-70B classifies ISP statements into five speech act categories: assertive, directive, commissive, expresive, declarative. To our knowledge, this represents one of the first applications of Speech Act Theory in the context of ISP quality research. To evaluate this, we analysed a dataset of 600 ISP statements randomly sampled from ten British National Health Service (NHS) information security policies. We evaluated LLAMA3.3-70B under both zero-shot and few-shot learning conditions, running prompt configurations containing 1, 5, and 10 examples per speech act category. To assess the sensitivity of classification outcomes to example selection and quantity, three random seeds (12345, 1234, and 123) were applied to sample subsets from a pool of 10 examples per speech act category. The 10-example prompt used all available examples and therefore required no sampling. Under each seed, both a 1-example and a 5-example set were drawn. Each setting was executed over three runs to establish output consistency. Model performance was evaluated using balanced accuracy, precision, recall, and F1-score. Finally, statistical tests were applied to determine whether observed differences between configurations were statistically significant. The results indicate that LLAMA3.3-70B shows promising classification performance and strong output consistency across repeated runs. Few-shot configurations consistently outperform zero-shot prompting. Among the evaluated configurations, providing 5 examples per speech act category yields the best overall performance, while further increasing the number of examples does not lead to additional improvements. As such, the statistically significant gain associated with example count was observed in balanced accuracy when moving from 1 to 5 examples, suggesting that additional examples improve handling of less frequent speech act categories. Performance variability across seeds further indicates that classification outcomes are sensitive to the specific set of examples selected. At the category level, Directive and Assertive statements are classified with consistently high reliability, while Declarative and Expressive statements are more challenging due to their misrepresentation in the dataset. We conclude that LLAMA3.3-70B is a useful starting point for continued work on improving ISP content using LLMs.

Keywords: Few-shot Learning, Information Security Policy, Speech Act Theory