Opinion Identification using a Conversational Large Language Model

Chaya Liebeskind
Department of Computer Science
Jerusalem College of Technology
21 Havaad Haleumi St., P.O.B. 16031
9116001 Jerusalem, Israel

Barbara Lewandowska-Tomaszczyk
Department of Language and Communication
University of Applied Sciences in Konin
1, Przyjazni str.
62 510 Konin, Poland

Abstract

The paper focuses on testing the use of conversational Large Language Model (LLM), in particular chatGPT and Google models, instructed to assume the role of linguistics experts to produce opinions. In contrast to knowledge/evidence-based objective factual statements, opinions are defined as subjective statements about animates, things, events or properties in the context of an Opinion (Speech) Event in a social-cultural context. Taxonomy distinguishes explicit (direct/indirect) and implicit opinions (positive, negative, ambiguous, or balanced). Contextually richer prompts at the LLMs training phase are shown to be needed to deal with variants of implicit opinion scenario types.

Opinions

Opinions are produced in the context of an Opinion (Speech) Event (Lewandowska-Tomaszczyk et al. 2023), considered a semiotic act, embedded in a social-cultural context, and expresses an opinion holder's judgement on a person, animal, property or event. Examples of opinionated texts and instances of explicit opinion-marking discourse markers (words and phrases) we identified, as well as instances of opinion-marking mental verbs, evaluative and emotion phraseology, and expressive lexis, were provided in a series of prompts to test the use of conversational LLMs.

Taxonomy of Opinions

1. Explicit – introduced by semantically transparent structural/semantic opinionated markers: Syntactic framing imposes the order of linguistic elements used an opinion and together with Semantic framing identifies degrees of certainly and conviction by particular Agents: e.g., My/Our opinion is..../According to me... Lexical framing is marked by relevant lexical items, as e.g., Cognitive verbs (e.g., I think, I believe), Modifiers (adjectives slow, adverbs slowly) that express evaluation or judgement (e.g., good/bad, worthy, valuable; careless - carelessly), in the comparison degrees: positive pretty, comparative prettier (than), and superlative the prettiest (of....), as well as emotional/evaluative expressions (e.g., they love skiing).

2. Explicit indirect opinion markers: Opinions may be reinforced with persuasive language, such as rhetorical questions, appeals to authority, or emotional appeals (Roberts 1996), (i.a., offensive and vulgar language). Indirectly conveyed opinions: he said/I've heard/it seems to her... Those opinionated texts which are introduced by means of unambiguous opinion markers such as "I think/I don't think/I do not think", "in my opinion" or "according to me" or else by indirect Explicit Opinion markers heard/was told, repeated from outside sources or via intermediaries.

3. Implicit opinions are typically unaccompanied by any explicit opinion markers. They can also include reference to vague or imprecise targets. Therefore, in some contexts it is not easy to distinguish them from factual statements. However, some opinions may be produced in the contexts conducive to such differentiation and leaving no doubt as to their status. In recent papers context-focused considerations have been applied e.g., Lian et al. (2023) propose the F_vague detector to automatically detect vagueness in the text. The authors also show that a large part of individual vague sentences have at least one clarifying sentence in the documents. Their experiments showed good performance of high recall and precision. However, reference to relevant clarifying information is not present in all cases.

Our approach

Introduction of LLM generative tools has led to attempts to identify implicit toxicity in texts. Wen et al. (2023) demonstrate that LLMs produce outputs that are difficult to detect using zero-shot prompting.

We implemented the chain-of-thought prompting (CoT) methodology (Wei et al. 2022). CoT enhances the reasoning capacity of LLMs by incorporating systematic step-by-step reasoning procedures into the demonstration. CoT prompting enhances the model's comprehension of the question's complexities and the process of reasoning. In addition, the model produces a series of logical stages, providing us with a clear understanding of the model's cognitive process, hence improving its interpretability.

Prompts

White et al. (2023) presented a pattern-based collection of efficient engineering methods to address common LLM challenges. We experimented with four of these patterns:

1. The Persona Pattern - we asked the LLM to produce persona-like linguistic expert outputs.
2. We achieved the Reflection Pattern's goal of prompting the model to automatically explain user responses. The persona pattern was combined with a request for a variety of linguistic phenomena samples.
3. The Cognitive Verifier Pattern shows that LLMs enhance reasoning by dividing questions into sub-questions and combining answers (Zhou et al. 2022). We also tried this pattern. The description of explicit and indirect opinionated texts includes examples. For instance, texts with explicit opinions use cognitive verbs and modifiers for lexical framing. We queried twice. Lexical framing with cognitive verbs was first asked, then with modifiers. In the end, the LLM failed to separate the two queries and produced a mix of both types. LLM performed better by defining direct opinionated texts in detail and using many examples. It generated more effective examples without repetition and categorized them by language phenomena.
4. We set context for LLM conversations using the Context Manager Pattern. Instead of simply requesting examples based on a category name, we added context by describing opinionated text and integrating relevant examples. Examples that followed the category's rule were requested from the LLM.

It is observed that the utilization of the template pattern, which enables the user to specify a template for the output, was unnecessary in this case, as the bulleted list was already obtained in response to the examples request. In addition, we incorporated an emotional stimulus into our prompt based on prior research (Li et al. 2023) indicating that LLMs possess emotional intelligence and that their performance can be enhanced by the use of emotional prompts. We advised LLMs to avoid chatty behavior, such as inventing questions, and engaging in a 'question and answer' format (Pearce et al. 2022).

We tested a conversational LLM's ability to understand different types of opinionated text. First, we explained a category with examples, and the LLM provided its own examples. We then introduced new categories and repeated the process. Finally, we challenged the LLM to distinguish opinion from fact by turning its generated examples into factual statements. By applying the specified prompting method, we effectively provided examples for each of the preset categories of opinionated text, which proved challenging for corpus linguistic techniques.

Results and conclusions

We report the results of two popular conversational LLM: OpenAI's ChatGPT-4[1] and Google's Gemini[2]. Both models were given identical prompts. Next, we detail the examples extracted for each category in the taxonomy of opinions.

Explicit opinions Both LLMs extracted 50 accurate examples as requested. However, Gemini independently classified them into linguistic phenomena (general, semantic framing, lexical framing, expressions of personal feelings or experiences, and other). All semantic framing examples were first person singular/plural. A peripheral opinion type. The semantic framework revolves around people (excluding the first person singular/plural). Expect: You/He/She/It/They...strongly believe(s). Ten examples of personal feelings (I got panicked) or experiences (We've seen this accident) are not opinions. As mentioned, they lack a nested target, making them peripheral opinions. All other linguistic phenomena examples are correct, but most of them supplement existing categories rather than create new ones.

Explicit indirect opinions Both LLMs extracted 50 accurate examples as requested. Although ChatGPT used rhetorical questions to emphasize opinions in all the extracted examples, such as ”Can't you see the blatant injustice in our legal system?”, Gemini extracted a variety of persuasive language examples that indirectly conveyed opinions. Gemini additionally explained each example.

Implicit opinions Tasked with 50 examples, the LLM gave ChatGPT and Gemini 25 and 30. All ChatGPT examples were explicit opinionated texts. Facts replaced implicit opinions. It became ”The movie received positive reviews” from ”I think that the movie was fantastic.” Geminis wrote 28 examples of 30 well to express implicit opinions. Among 30 examples, Implicit Opinions to be converted to Factual statements, 3 are implicit to implicit opinions and 2 are explicit indirect opinions converted to synonymous indirect ones. Implicit opinions converted to Explicitly Indirect ones (25) were obtained in the majority of instances, advancing opinion typology content clarification.

Separately, the missing reference to specific contextually-anchored Opinion Event context can make some examples taxonomically ambiguous. According to the definition, opinion as an event typically identifies its holder, sources, target, effects, relation to evidence data, etc. Identifying opinions from factual statements is harder without such a reference. The recipient of a context-free message ”It is raining” cannot verify its ambiguity between a factual statement made by the sender in a heavy rain or in a cozy room when the s/he sees water falling from the roof. It is access to relevant contextual information that would make it possible to disambiguate the sense. To resolve the implicitness ambiguity LLM systems would need to use relevant Opinion Event contextual clues to expand their taxonomic options. Our future research will refine prompts to make opinion event contextual clues more transparent and use computational methods to incorporate such data into a contextual information transfer system.

Acknowledgement

The present study was carried out within COST Action CA 21129 What are Opinions? Integrating Theory and Methods for Automatically Analyzing Opinionated Communication (OPINION).

[1]`https://chat.openai.com`

[2]`https://deepmind.google/technologies/gemini/`

References

Lewandowska-Tomaszczyk, B.; Liebeskind, C.; Baczkowska, A.; Ruzaite, J.; Dylgjeri, A.; Kazazi, L.; and Lombart, E. 2023. Opinion events: Types and opinion markers in english social media discourse. *Lodz Papers in Pragmatics* 19(2):447–481.

Li, C.; Wang, J.; Zhang, Y.; Zhu, K.; Hou, W.; Lian, J.; Luo, F.; Yang, Q.; and Xie, X. 2023. Large language models understand and can be enhanced by emotional stimuli. *arXiv preprint arXiv:2307.11760*.

Lian, X.; Huang, D.; Li, X.; Zhao, Z.; Fan, Z.; and Li, M. 2023. Really vague? automatically identify the potential false vagueness within the context of documents. *Mathematics* 11(10):2334.

Pearce, H.; Tan, B.; Krishnamurthy, P.; Khorrami, F.; Karri, R.; and Dolan-Gavitt, B. 2022. Pop quiz! can a large language model help with reverse engineering? *arXiv preprint arXiv:2202.01142*.

Roberts, W. R. 1996. Rhetoric by aristotle. *HyperText Presentation © 1996 Procyon Publishing*.

Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Xia, F.; Chi, E.; Le, Q. V.; Zhou, D.; et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems* 35:24824–24837.

Wen, J.; Ke, P.; Sun, H.; Zhang, Z.; Li, C.; Bai, J.; and Huang, M. 2023. Unveiling the implicit toxicity in large language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 1322–1338.

White, J.; Fu, Q.; Hays, S.; Sandborn, M.; Olea, C.; Gilbert, H.; Elnashar, A.; Spencer-Smith, J.; and Schmidt, D. C. 2023. A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint arXiv:2302.11382*.

Zhou, D.; Schärli, N.; Hou, L.; Wei, J.; Scales, N.; Wang, X.; Schuurmans, D.; Cui, C.; Bousquet, O.; Le, Q.; et al. 2022. Least-to-most prompting enables complex reasoning in large language models. *arXiv preprint arXiv:2205.10625*.