



Universiteit
Leiden
The Netherlands

Opinion diversity through hybrid intelligence

Meer, M.T. van der

Citation

Meer, M. T. van der. (2025, March 26). *Opinion diversity through hybrid intelligence*. SIKS Dissertation Series. Retrieved from <https://hdl.handle.net/1887/4209024>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4209024>

Note: To cite this publication please use the final published version (if applicable).

Summary

Problem In a representative democracy, citizens elect representatives to act in their interest for shaping public policy. Modern democracies face a critical issue of declining citizen participation, leading to a disconnect between citizens and their elected officials. Deliberative democracy, which emphasizes open dialogue and encourages wider participation, is one way to address this issue. However, traditional in-person deliberation methods face challenges, such as including participants with a diverse set of perspectives and backgrounds, and ensuring all voices in a discussion are considered equally. Online social media platforms offer an alternative venue for large-scale deliberation, allowing for discussions with a wider audience and rapid access to information. However, concerns exist about whether these platforms can foster truly inclusive and diverse discussions. For instance, locating contributions of relevant perspectives can be difficult due to the large amounts of scattered content. Further, interactions on platforms lead to echo chambers that drive polarization, threatening the egalitarian basis of the discussions. Artificial Intelligence (AI), and Natural Language Processing (NLP) in particular, for facilitating text-based online discussions have become attractive as a solution. However, its impact on the diversity of perspectives in online deliberation is unexplored.

Methods This dissertation (1) identifies the challenges involved in facilitating large-scale online discussions with NLP, (2) suggests solutions to these challenges by incorporating hybrid human–AI technologies, and (3) investigates what these technologies can reveal about individual perspectives in online discussions. We propose a three-layered hierarchy for representing perspectives that can be obtained by a mixture of human intelligence and Large Language Models (LLMs). This combination is known as Hybrid Intelligence (HI). We illustrate how these representations can draw insights into the diversity of perspectives and allow us to investigate interactions in online discussions.

- In Part I of this dissertation, we show that existing opinion analysis methods, particularly those involving LLMs, are limited in understanding the perspectives expressed by minorities. Nonetheless, the models' capabilities of processing text-based data at scale make them attractive in analyzing online discussions. Despite the complexity of understanding free-form opinionated texts, we can effectively use models in low-resource settings. In particular, LLMs can address abstract tasks fluently and interpolate missing information. However, the sensitivity of, e.g., zero-shot prompting procedures for LLMs underscores how they still need human oversight to perform well across contexts. Further, LLMs behave differently from humans, failing to align with human disagreement and making errors different from us, thus necessitating careful supervision.
- In Part II, we harness the potential of HI systems for opinion analysis. By strategically incorporating human input with LLMs and fostering a back-and-forth process between

humans and AI, HI systems can be designed to capture diverse opinions precisely and efficiently. HI requires careful task allocation and balancing. Our methods use human annotators to provide a nuanced understanding of e.g. arguments while using NLP techniques for sampling interesting opinions from a large dataset. Through repeated interaction, both sides continuously adjust and learn from each other. Our hybrid setup paves the way for a future where humans and NLP technology can join forces to cultivate a deeper understanding of the multifaceted nature of online discussions.

- In Part III, we show that HI systems can extract an individual's *Perspective Hierarchy* based on the tasks used in Part I. To extract the perspective hierarchy, we leverage the complementary abilities of humans and NLP models. We show how arguments, next to stances and personal values, are a core component of the hierarchy by experimenting with extracting a direct relation between values and stance.

Findings We find that there are fundamental issues to fostering diversity when analyzing online discussions on social media platforms. These issues include: (1) ensuring that minority and marginalized voices are participating on the platforms, (2) an emphasis on frequently repeated opinions that fail to bridge political divides, and (3) an aggravation of this problem by the straightforward application of LLMs for opinion analysis. HI can help alleviate these problems. In our approach to HI, we encourage more explicit communication between humans through repeated interaction with LLMs. All of this feeds improvements on two ends: humans benefit from explicit communication and closely considering each other's point of view, while the rationales provided by them are useful resources for AI to learn from.

In this dissertation, we provide one of the first demonstrations of how HI can be used to integrate humans and AI, mixing human collaborative capacity with LLMs. Nonetheless, showing that HI leads to improved and diverse discussions remains difficult. Existing evaluation paradigms are insufficient for measuring how HI leads to improvements over AI-only or manual approaches, indicating the need for more dynamic and context-sensitive evaluation approaches.