Opinion diversity through hybrid intelligence

Meer, M.T. van der

1

Introduction

Michiel van der Meer. Facilitating Online Opinion Diversity through Hybrid NLP Approaches. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 4: Student Research Workshop)*, pages 272–284, Mexico City, Mexico. Association for Computational Linguistics.

1

The essence of democracy is that citizens have a say in how they are governed. From the public fora of the Ancient Greeks to the European Parliament, reasoning and arguing form the core of discussions where diverse perspectives are debated. However, modern governments struggle with declining citizen engagement [362] and diminished trust in political institutions [128]. At the same time, our society faces a multitude of complex, interwoven issues—climate change [181], misinformation [431], vaccination hesitancy [258], and many others [280]—that require democratic resolution. These societal issues share characteristics: problems are multifaceted and interdependent, they have no clear definite solution, decisions need to be made under strict time constraints, and solutions require fleshing out deeply-rooted ethical disagreements. These characteristics are typical for *wicked problems* [320]: issues that seemingly have no solutions due to the diverse needs of those involved.

Addressing wicked problems in society requires reshaping citizen participation [325]. *Deliberative democracy* underpins a wave of democratic transformation, advocating for decisions to be made through fair and reasonable discussion [289]. Central to deliberative democracy is the process of **deliberation**, where citizens, not just experts or politicians, are deeply involved in shaping solutions to societal issues [106]. Deliberation is based on egalitarian and rational debate, with expert information freely accessible [155]. Solutions stemming from deliberation benefit from the wisdom of the crowd effect: the collective judgment of a diverse crowd of humans is more accurate than any individual member in that group. Humans are good collaborative problem solvers [219], and collective decision-making builds sustainable solutions [163]. However, deliberations need careful facilitation to sustain the conditions for productive discussions and safeguard democratic ideals.

The diversity of perspectives is a driving factor in determining the quality of outcomes in a deliberation [36, 53, 87]. When citizens express their desires and provide insights from different backgrounds, diversity leads to effective decisions [227]. Diverse perspectives can spark creative solutions by challenging assumptions and encouraging innovative thinking. This is echoed in cases of democratic transformation where encouragement of diverse perspectives is hailed as a means of stabilizing democracy [114].

Facilitating diversity requires actively steering the deliberation process. First, participation from a broad group of representatives requires more organizational overhead to ensure an inclusive recruitment procedure. Second, deliberating the complex needs of individuals requires active perspective-taking from those involved in the discussion, imparting a significant cognitive and emotional load [133, 213, 391]. Third, the deliberation process requires moderators that play a crucial role in setting ground rules for respectful communication, encouraging participation from all members, managing conflicts constructively, and summarizing discussions to highlight different viewpoints [91, 136].

Existing deliberative practices have inherent limitations, such as a reliance on physical gatherings and the frequent use of small, supposedly representative, citizen groups [26]. Even small-scale deliberations see issues surrounding organization, effective participation, and collective decision-making [123]. For instance, gathering people to come together physically at a specific time is resource-intensive [115]. Further, there is a maximum number of people that can be feasibly included, limiting the diversity of that group.

Alternatively, contemporary social media platforms enable large-scale communication and may facilitate large-scale *online* deliberations [132], fostering citizen engagement [159, 348]. These platforms can serve as a channel for the rational exchange of ideas and opin-

ions, provide access to a broad range of information sources, and host facilitated discussion through moderator involvement [118]. Large technical leaps, like recommender systems [14] and automatic translation [444], can provide opportunities for all citizens to contribute to the public debate. Lowering the barrier to accessing societal discussions allows global issues like climate change to be addressed not by a limited group of representatives, but through engagement across all layers of society. However, whether such platforms serve as an inclusive public space or not remains debated [297]. Online discussion is fundamentally different from the conversations in offline deliberation [25]. Online discussions offer wider and more free participation but are less regulated and harder to moderate than offline ones. It is therefore important to highlight the prerequisites for achieving the wisdom-of-the-crowd effect in online discussions: the egalitarian participation of a diverse crowd of citizens.

Transitioning to online deliberation adds a new dimension to the challenge of facilitating diversity: that of **scale**. Considering the massive user bases online platforms can support, manual moderation becomes infeasible. Online opinions spread and evolve differently from guided offline deliberations [441, 447]. In offline deliberation, diverse participation is attained by representative sampling according to demographics. However, ubiquitous participation from online users leads to open questions on how to foster the development of diverse perspectives when such a strategy is infeasible. Since poorly designed online discussions can lead to polarized outcomes [437], this challenge needs to be considered carefully.

To effectively facilitate online discussions at scale, it is essential to have tools that can analyze these discussions. In this dissertation, we consider these interactions to be text-based exchanges of opinions. On social media platforms, humans engage with one another by communicating their viewpoints through written text. We turn to Natural Language Processing (NLP) and create new methods for harvesting insights from opinions. While investigating human behavior has long been the domain of social sciences, combining social science methodologies with NLP models has barely passed its infancy [454]. This emerging interdisciplinary approach offers new avenues for understanding large-scale human interactions. To uphold democratic ideals, it is essential to develop responsible tools [455], which requires a thorough understanding of the shortcomings of existing NLP techniques. We create an overview of these limitations and propose a strategy to overcome them in the form of Hybrid Intelligence (HI). HI refers to integrating human and machine intelligence, enhancing human capabilities instead of replacing them [5]. We dive into how we can create HI that combines citizens and NLP methods to facilitate diversity in online societal discussions.

Improving citizen engagement through deliberation requires effective collaboration between citizens and stakeholders, such as politicians or industry parties. The institutional uptake and implementation of deliberation efforts have thus far remained unfocused and scattered [140, 360]. One reason for the hesitant uptake of online deliberation is that legitimate deliberative processes need to account for non-included individuals to be considered representative [298]. Enhancing citizen participation by designing and implementing technical solutions for addressing societal issues at scale can help in achieving legitimacy [148]. This dissertation contributes to this goal by proposing to engage with a diverse public directly through NLP-supported facilitation. Focusing on finding wide-ranging perspectives in society-wide conversations leads to inclusive and informed decision-making. An integrated view of the humans involved in online discussions should limit adverse effects such as echo chambers [77], polarization [392], and other negative external and internal effects

1

[251, 263, 370]. In the long run, the positive effects of promoting diversity in online discussions can lead to the empowerment of citizens.

Structure The rest of this chapter is structured as follows: We provide an overview of the problem of facilitating online discussions with NLP based on the ideals of deliberation and introduce our Research Questions (RQs) in Section 1.1. We continue with a description of the relevance of each RQ in Section 1.2. We define the scope of this dissertation in Section 1.3, and finally provide an outlook on the findings of our work in Section 1.4.

1.1 Research Questions

Online discussions generate vast amounts of content, which is challenging to manage and navigate [88] because content is scattered across time and threads, and contains frequently repeating or unconnected arguments. This makes it difficult for users to know where to add new contributions, resulting in low-quality content [204]. These issues can be addressed by employing moderators, e.g., to structure the content of a discussion or to steer user interactions [390]. However, given the amount of data, manual moderation is not feasible.

Instead, we turn to NLP for interpreting text-based opinions at scale [374], powered by the recent surge of Large Language Models (LLMs) [20, 266]. LLMs have shown a remarkable ability to code novel texts with limited adaptation requirements [385]. Central to our approach to facilitation is extracting structured *perspectives* from users in a discussion. Perspectives provide high-level insights into the arguments employed by citizens [414] or the motivations underlying the opinions in a community [429]. These representations influence the facilitation strategies [121] and shape policies following the discussion [274].

Using NLP for analyzing perspectives sourced from online discussions is challenging. For instance, social media platforms have been centered on managing large volumes of information, e.g., through personalized recommendations [3] or argument structuring [178] but have neglected inclusive design aspects [352]. This can cause majority opinions to be heard while suppressing dissent voices [282], or lead to filter bubbles [392]. Similarly, we see that LLMs capture majority opinions well, but do not distill all voices equally [e.g., 278, 405]. Further, LLMs lack deep social reasoning [232], may be biased [162, 333], and make mistakes in ways humans cannot anticipate [175]. LLMs can be readily applied in new contexts, but they remain fickle and inconsistent depending on the exact prompts used [254]. Straightforward automated discussion analysis runs the danger of ignoring diverse opinions, which undermines the wisdom-of-the-crowd effect [250]. To find out the nature of these challenges and whether they can be resolved, we ask our first research question:

Q1 *What are the fundamental issues in using NLP to analyze perspectives?*

Next, our goal is to obtain structured perspectives from online societal discussions that provide insights into the opinions involved. In particular, we aim to improve the degree to which **diverse** perspectives can be obtained. This requires us to combat the limitations of NLP by adopting a "hybrid" mindset, i.e., incorporating humans-in-the-loop to address diversity directly. We leverage LLMs and humans jointly, with their complementary capacities for interpreting opinions from text. This leads to our second research question:

Q2 *How to combine human intelligence and NLP to effectively capture diverse perspectives?*

Finally, in practice, analyzing opinions is modeled by different task formulations, all aimed at extracting various types of information based on language input. We propose a **perspective hierarchy** that incorporates *stance*, *arguments*, and *personal values* to represent perspectives at different levels of abstraction. We base our model on the complementary skills of humans and NLP methods, in which we mix higher-order abstractions with surface-level extraction tasks. Each task has been investigated separately, but little is known about their interaction in online discussions. We, therefore, ask our third research question:

Q3 *How to construct a perspective hierarchy based on diverse opinions in a discussion?*

1.2 Research Methodology

We introduce the methods for answering the research questions step by step.

1.2.1 Fundamental Issues (Q1)

There is an increasing interest [e.g., 84, 183, 440] in using NLP to facilitate online societal discussions. Existing work is focused on (1) using NLP tools, in particular few-shot prompted LLMs, to analyze the discussions [e.g., 377, 440], and (2) using discussion data to benchmark the capabilities of NLP tools [e.g., 124]. In the next two sections, we provide related work to the research methodologies adopted in this dissertation, highlighting fundamental techniques and applications.

Discussion Analysis

Using NLP to analyze large amounts of text in online interaction is studied under the broad umbrella of opinion mining [244]. Discussions happen in various contexts, such as climate change [249], pandemics [160], and others [49] . The scale of these discussions, combined with their pertinence, makes analyzing them interesting. Analyzing what humans express through text is the core task in many NLP areas, e.g., Opinion Summarization [244], Argument Mining [224], Sentiment Analysis [424], and Value Classification [237]. These tasks lie at the heart of creating insights into online (political) discourse. They can be used e.g., for estimating the quality of discussions [368], extracting the arguments involved [220], or reasoning over inconsistencies between choices and their justifications [243]. In the age of LLMs, these tasks have seen considerable performance improvements [186], although new challenges such as dealing with shortcut learning [138] or mitigating social biases [232] arise.

Extracting diverse views from online discussions is challenging for three reasons. First, data from social media platforms inherits biases present on these platforms, including fake news, trolling, and polarization [77]. This impacts how opinions are shaped [167] and the distribution of opinions [441]. Second, when analyzing the opinions about societal issues, not all citizens have equal access due to the digital divide [86] or differences in tech-literacy [206]. This makes the users in online discussions biased and less diverse. Third, since users are free to join in discussions of their choosing, there are undesired echo chambers or self-selection effects among the messages seen by users [363].

Despite these challenges, we can use NLP to investigate questions about human behavior at scale [225]. Analyses about behavior may lead to insights at both individual and group levels. This can be useful for improving democratic processes [80], but also applies in other areas, such as faithfully interpreting product feedback [34], service improvement [358], or course management for education purposes [233].

Approach
We can employ discussion analysis to benchmark how well NLP approaches understand opinionated text. In benchmarking, we test the analysis procedure, and models used, for possible mistakes and biases. Representing subjectivity is difficult since LLMs do not faithfully capture the full range of opinions [108, 166, 405]. Whether LLMs can learn to represent them in the future remains unclear [337, 427], but research suggests that they cannot [20, 124]. Therefore, we work with the assumption that this is a fundamental limitation of LLMs, and we have to find other approaches for improving diversity.[1]

Creating diversity-enhancing techniques is gaining traction in NLP, but there are several aspects of diversity. For instance, creating more diverse news recommender systems is a common goal [216, 438] for shaping an individual's perspective [29]. Others strive to make LLMs better represent a diverse group of annotators based on their labeling behavior and demographics [28, 217]. In such approaches, models rely on annotated data. Labels are obtained from a few human annotators per instance and are often aggregated by majority voting, painting an incomplete picture of the true range of interpretations of opinionated text [302]. The role of subjectivity in these tasks remains unclear [21, 61]. This holds for traditional supervised learning, but also for the latest trends in instruction-tuning [393, 422].

Contributions
In Part I of this dissertation, we dive into the application of LLMs to analyze the opinions in online discussions. Our work centers on argumentation: the rationales behind human opinions. In Chapter 2, we begin by examining the diversity of the opinions in LLM-generated summaries of argumentative content. We find that automated methods for summarizing arguments struggle to represent arguments shared by few people, and such error cases usually go unnoticed using standard NLP evaluation practices. By examining how LLMs fare on complex argument quality assessment tasks under strong data constraints in Chapter 3, we aim to further investigate how we can best deal with low-resource settings. Here, we observe that zero-shot models can drive the state-of-the-art, but come with significant cost and data requirements to work well out of context. Overall, significant challenges remain when applying LLMs to tasks of analyzing opinionated data at scale.

1.2.2 Hybrid Intelligence (Q2)

In Part II, we argue that the aforementioned challenges can be overcome by using LLMs to **assist humans** in mining opinionated text, rather than replacing humans. This notion of *Hybrid Intelligence* [5, 97, 98] is central to our approach to uncovering diverse perspectives in online discussions. In Hybrid Intelligent Systems (HISs), Artificial Intelligence (AI) agents are collaborators that enhance human abilities such as reasoning, decision-making,

[1] Although linguistic diversity generally refers to the diversity of language proficiencies [103, 190], we are specifically interested in diversity in arguments, communication styles, and values in online discussions.

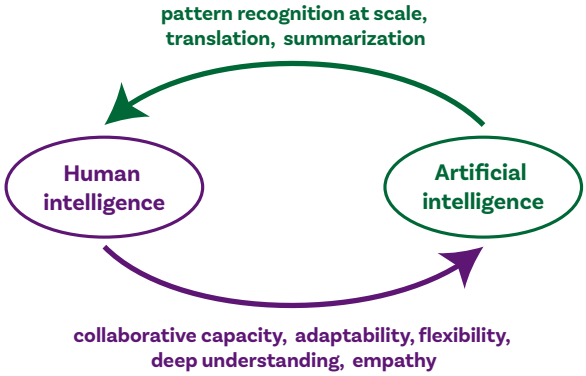

Figure 1.1: Feedback loops in Hybrid Intelligence.

and problem-solving [383]. Hybrid intelligence aims to augment intellect, creating a synergy between humans and NLP. For supporting online discussions, we combine the strengths of human intelligence with AI, highlighting bidirectional gains, as shown in Figure 1.1.

The application of AI to understanding human written language has had a profound impact on how researchers analyze human behavior at scale. To do so responsibly, we must ensure that our methods uphold democratic values, especially considering the pressing need to represent diverse perspectives. Previous work on hybrid approaches for NLP includes user adaptation [256], human-in-the-loop computing [423], human-AI interaction [164] and others [e.g., 82, 102]. Recent interest in explainable AI has focused on human understanding of NLP models [230]. Specifically for NLP, much focus is on approaches that mix crowd, expert, and automated decision-making, which have been applied to analyzing discussion content [208, 295]. However, these approaches have a one-way interaction between the NLP model and humans, as we will describe in the next section.

Approach

We observe that LLMs have many challenges to overcome in representing diverse perspectives (Section 1.2.1). Discussions are deeply human, who can adapt to incomplete and informal argumentation, behave flexibly, and provide empathic responses to foster collaboration. Thus, humans and NLP can benefit from each other. In the next paragraphs, we examine each benefit in either direction (humans aiding NLP or NLP aiding humans) separately, and lastly illustrate how both can be incorporated into an overall hybrid method.

Humans aiding NLP Humans provide the data that the NLP tools perform their analysis on, as gathered from interactions between different stakeholders, including casual and advanced users, moderators, or even site admins [336]. They provide text and behavioral data, such as likes or post-votes, which we, in turn, can use to analyze their attitude. Furthermore, NLP approaches learn from labeled data, obtained from annotators who observe a given text and draw labels from a predefined set of classes. Humans can be flexibly employed in such procedures, dealing with expanding label sets [396], free-form text response [294], asking a crowd of annotators rather than individuals [286], and more [e.g., 302, 334]. Humans contribute their opinions, either through text or by labeling, based on

lived experiences or professional expertise, and are capable of empathizing with others. While crowd annotators are usually uninformed lay users, they are assumed to adapt to tasks quickly given a set of instructions and examples. Since annotators adapt differently, addressing the problem of diverse opinion understanding requires selecting an appropriate set of annotators, to capture the human label variation [302].

NLP aiding humans NLP aids humans in online discussions in multiple ways. While we have mostly discussed the analysis of large-scale discussion data, there is a broader potential impact of NLP technologies in online deliberations [384]. First, NLP may enable, rather than restrict, access to certain services, for example by summarizing large amounts of content through summarization or using automatic translation to account for different language proficiencies. Second, since humans suffer from cognitive biases, NLP models may offer an alternative interpretation of the content. Machines do not get bored and treat each sample with equal consideration. Third, NLP models mirror biases captured in the data, which allows for obtaining synthetic opinion data or exposing biases in discussions. Lastly, since their scale, speed, and accessibility to researchers are advancing quickly, we can experiment with them rapidly.

Combination Existing work mostly offers one-directional benefits, either machine- or human-oriented. By constructing hybrid approaches, we aim to improve both humans and AI through an iterative process. We see that NLP methods are biased, leading to questions about the soundness of the analysis. Humans can repair biases and provide deeper interpretations, contexts, and explanations. Furthermore, we see that there are many opportunities for NLP to aid humans. Completing the loop allows bootstrapping: traversing the two feedback loops shown in Figure 1.1, iteratively refining the analysis procedure while performing discussion analyses. In this procedure, a human interprets opinions shown from the output from a model and possibly corrects it in a human-in-the-loop fashion [273]. However, to guide the human through a large amount of data, the NLP models will steer it through what content to observe. Through this collaborative approach, we hope to synthesize **bidirectional gains**. Bidirectional gains in hybrid intelligence refer to the mutual increase in capabilities achieved when human and artificial intelligence work together. We emphasize the synergistic nature of human-AI collaboration, where each side strengthens the other, leading to more powerful, efficient, and reliable intelligence than either could achieve alone. Hybrid approaches combine the strengths of humans and machines, offering immediate and long-term benefits. By keeping humans in the loop, their task proficiency improves, and additional data is generated to develop the hybrid collaboration. Further, advancements in NLP models can be integrated into the framework. However, doing so effectively requires broad contextual understanding.

Developing hybrid approaches necessitates a new evaluation paradigm. We must assess the effectiveness of our method by comparing it against both human-only and machine-only baselines. In the field of NLP, test sets are typically compiled manually and with hidden data issues [99, 154, 329], which might introduce an unfair advantage to the upper bound of performance [56, 200]. Initial work shows that there are considerable performance gaps between hybrid and manual approaches [127, 443].

Contributions

We present our approach to incorporating humans and NLP methods for analyzing opinionated text data. First, we introduce a method for mining diverse arguments from citizen feedback in Chapter 4. Our method, HyEnA, finds more diverse arguments and improves the precision of the argument analysis by efficiently querying human annotators across three distinct phases. In Chapter 5, we further investigate how differences between annotators in subjective tasks, such as interpreting texts for extraction of arguments or personal values, can be modeled more efficiently. Our approach steers models to learn diverse label distributions by picking from a large pool of annotators. Central to our work, we create discussion analysis approaches that (1) select samples for human inspection that are interesting to annotate, (2) account for diversity (e.g., leveraging contextualized embeddings [314]), and (3) seek labels from multiple annotators. The hybrid nature of our methodology leads to bidirectional gains, serving the NLP system as well as the humans involved. For instance, we create approaches to capture more diverse interpretations of the arguments in discussions using a crowd of annotators. After the annotation, our method outputs a summary of the high-level argument involved, while annotators were able to develop their understanding of controversial discussions. We achieve a cost-effective crowd annotation, while actively engaging with the annotators, and developing their perspective. Moreover, we can also actively diversify which annotator we query an annotation from. We observe that an active selection of diverse annotators can inform a model more quickly of the label distribution underlying subjective tasks in cases where the annotator pool is large.

1.2.3 Perspective Hierarchy (Q3)

Given that NLP can process large amounts of discussion data, but is limited in its capabilities (Section 1.2.1) and that we may construct hybrid procedures to account for these limits (Section 1.2.2), we address the challenge on how to capture perspectives. Uncovering perspectives from online societal discussions requires a representation for identifying how people feel about potential decisions, how the considerations are communicated in the discussions, and the motivations underlying preferences held by individuals. There is a large amount of literature concerned with addressing these questions through separate NLP tasks. We attempt to integrate these tasks and find out how they model various aspects of perspectives. We propose a hierarchy to structure our approach to facilitating online discussions at scale.

Few attempts to comprehensively represent perspectives exist [71, 412]. These works focus on annotating utterances for low-level claim information [272], or investigating the reasoning behind the views held in discussions [104]. Stances and arguments are inherently linked in argumentation models [386, 408], and form the basis of frameworks for representing perspectives [72, 432]. Existing work on mapping deliberative discussions has focused extensively on capturing this reasoning and using it for facilitation [158, 205].

However, stances and arguments do not represent opinions at a deeper personal level. A fundamental concept for explaining the motivations underlying opinions is personal values [344]. There are various theories of personal values [e.g., 143, 321, 344]. Preferences among values describe the attitude of individuals and groups [304], and can be extracted from behavioral cues to investigate political affiliation [326], perform moral reasoning [271] or positively influence lifestyle [95]. Values are abstract and need to be interpreted inside their context, making it difficult for both humans and NLP methods to measure them reli-

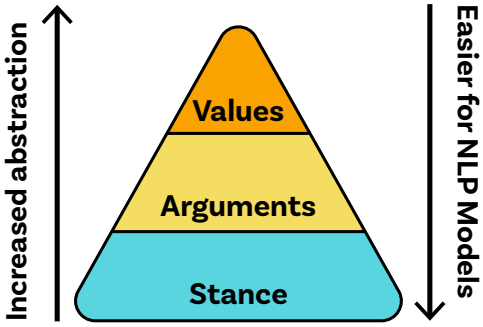

Figure 1.2: The perspective hierarchy. The higher the level of abstraction, the more human intelligence is required for interpreting the component.

ably [241]. One way to contextualize them is to connect values to argumentation, focusing on how choices are justified [201]. Using this insight, we incorporate personal values into our perspective representation and aim to obtain them using a hybrid approach.

Approach
We propose a perspective hierarchy to represent a person's perspective at different levels of abstraction, shown in Figure 1.2. Our perspective hierarchy is composed of stances, arguments, and values. We adopt the following definitions:

Stance Whether, or how much, support or opposition is expressed to a claim. Stance detection has been studied extensively and remains a popular NLP task [214].

Arguments The reasons given for adopting a stance towards a claim. In real-world contexts, argumentation manifests in many forms and is predominantly informal [146]. Mining arguments from text works well within known contexts [112], but suffers from implicit reasoning [157]. Hence, we require more human guidance to correct for possible mistakes in automated methods.

Values The motivations underlying opinions and actions [344]. Values are communicated implicitly through actions or written motivations. Estimating values automatically remains difficult even within their context [202]. Only through iterative hybrid procedures can we accurately reason about preferences among human values.

We combine the three components into a layered hierarchy, to indicate a tradeoff with respect to (1) the capabilities of NLP models to capture information from text, and (2) the level of abstraction that the component captures. Higher-order abstraction requires "filling in" more implicit knowledge. For instance, for stance detection, one or a couple of sentences can be enough to determine the stance of an individual concerning a topic [31]. However, for estimating value preferences, we need continued interaction over time to infer how values are prioritized within their context [240].

We illustrate how we used data from large online social media platforms to investigate perspective hierarchies for individuals [400]. Our main objective is to investigate whether

we can connect stances and values directly, omitting arguments to challenge their inclusion in the hierarchy.

Given a societal discussion on an online platform [305], we first identify relevant controversial topics and apply our automated methods for obtaining stances and value preferences. Because of the aforementioned limitations, we utilize the human-in-the-loop approach to uncover possible mistakes from the extraction pipeline. In particular, we compare human-provided self-reported value preferences to those estimated from behavioral data. Using this data, we can (1) compare how well the automated approaches work versus manual ones, (2) mix information from self-reported and behavior-based value preferences, and (3) investigate the relationship between components of the perspective hierarchy.

Contributions
In Part III, we make use of our hybrid setup to investigate the perspectives of participants in online discussions at scale. In Chapter 6, we investigate connections between value conflicts and disagreements in online discussions on societally relevant topics. Our experiments show that only when values are diverse, automatically-identified conflicts in values can correlate to stance disagreement. No strong evidence points towards a consistent and context-independent link between disagreement and value conflicts. However, when we incorporate human-provided self-reports, the evidence becomes stronger, showing that the hybrid approach is crucial to performing a meaningful analysis. When strong value diversity is absent, we cannot correlate disagreement and value conflict directly at all. A lack of a direct link means we require a more complete picture, and thus we incorporate arguments to complete the perspective hierarchy.

1.3 Dissertation Scope

The topic of this dissertation lies in the intersection of computer science, natural language processing, social science, and political theory. It is, therefore, inherently interdisciplinary and therefore can be approached from multiple angles. We provide a scope of the research involved before we dive into the description of how we address each research question.

In our work, we consider *online discussions* as text-based user interactions that happen on contemporary online platforms such as Twitter/X[2] or Reddit. Furthermore, we include data from specific questionnaires that gather citizen responses on proposed policy. We focus on topics that are *societally relevant*, such as climate change, due to the difficulty of addressing them. Lastly, we concern ourselves with deliberation among a group of people, as opposed to individual deliberation for self-reflection purposes.

Core to our work is diversity of perspectives. Depending on the context, the definition of diversity encompasses differences in various attributes, including social categories (e.g., gender, age, race) and informational or functional attributes (e.g., functional background, educational background) [30, 168, 409]. Research on group deliberation and diversity has primarily focused on a limited set of dimensions within these categories, or on the interplay between these two dimensions. In this work, we adopt *diversity of perspectives* as the full range of beliefs, opinions, stances, and values held by a given group of people. For any two people, these components might be in conflict at arbitrary levels, requiring extensive deliberation to uncover common ground. Our definition is similar to those adopted in other work

[2]Starting from July 2023, Twitter was renamed to "X."

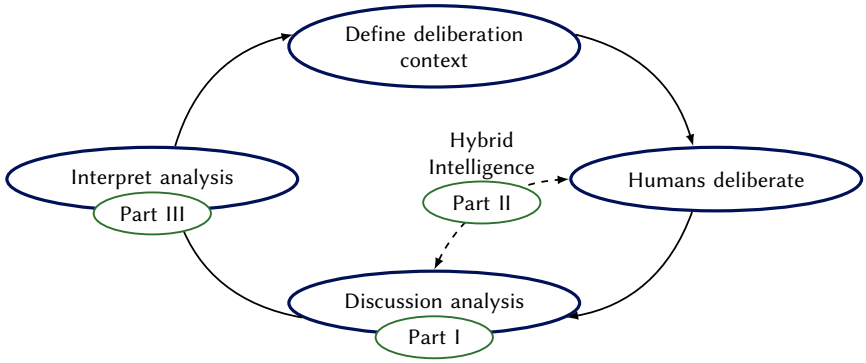

Figure 1.3: The deliberation cycle, annotated with the three parts addressed in this dissertation.

in group deliberation, often referred to as cognitive diversity [194], or viewpoint diversity [105]. It is distinct from demographic diversity [221] since we target the opinions and not the opinion holders, or linguistic diversity [190], which focuses on language proficiency.

This dissertation is focused on developing hybrid approaches to analyzing discussions from a technological perspective, with *hybrid* indicating human–AI cooperation [5, 97, 312]. We make modifications to computational artifacts (such as NLP models and datasets) and design processes for discussion analysis. Other strategies for improving discussion analysis, such as teaching humans analytical skills or implementing interventions for behavioral change, are left as future work but are compatible with our setup.

Lastly, our work is concerned with creating AI methods that focus on understanding human opinions based on digital text. Neural approaches from Natural Language Processing, in particular Transformer-based models, are the workhorse in the experiments performed in this dissertation. Other behavioral information, such as direct polling, referenda, post-voting, and others may provide different and possibly conflicting information for interpreting an individual's perspective. Consolidating such information with text-based opinions is nontrivial and requires careful prioritization of signals [354].

1.4 Outlook

Our goal is to augment the diversity of the opinions present in online societal discussions. These discussions are rooted in deliberative ideals, aiming to foster inclusive, informed, and respectful exchanges that lead to collective decision-making and problem-solving. We enhance the discussion analysis process by considering discussion analysis a hybrid undertaking, bringing HI to aid the deliberative cycle as shown in Figure 1.3. We separate our work into three parts as follows. First, we identify the strengths and weaknesses of using NLP to analyze discussions with diverse perspectives in Part I. Second, we see how HI can improve the capture of diverse perspectives in societal discussions in Part II. Our work proposes hybrid methods to sustain a high degree of diversity in discussions with a large crowd. Third, in Part III, we propose a perspective hierarchy to guide the investigation of human opinions in online societal discussions at scale.

The outlook of using HI, where we augment human intellect with AI, particularly sup-

ports deliberative discussions and decision-making processes. Our approach can democratize access to information and enhance the quality of public discourse by providing structured data analysis, fact-checking, and summarization pipelines, enabling more informed and evidence-based conversations. HI also facilitates inclusivity by assisting individuals with different abilities and backgrounds, ensuring a broader range of voices are heard. It aids in navigating complex societal challenges, such as climate change or public health crises, by integrating diverse data sources and perspectives. However, it is crucial to ensure that these technologies are developed and deployed ethically, mitigating biases and maintaining transparency to foster trust and acceptance in society. Ultimately, HI has the potential to empower communities, strengthen democratic processes, and drive more effective problem-solving for societal issues.