Check for updates

Summarizing Online Discussions with Prototype Relation Networks Using Large Language Models

William Power(✉), Shelly Gupta, and Zoran Obradovic

Center for Data Analytics and Biomedical Informatics, Temple University, Philadelphia, PA, USA
{tug00038,shelly.gupta,zoran.obradovic}@temple.edu

Abstract. The internet contains an ever-increasing body of human communication. The emerging class of pre-trained generative language models represents a new opportunity for ingesting, understanding, and summarizing these communications. This work outlines a method of generating Prototype Relation networks using a pipeline of large language model prompt-completions. These networks are constructed with nodes of prototypical authors that have their views represented within the corpus, and edges containing possible argumentative relations between them. Methods of evaluating these networks are described, and are used to show that the content within the generated prototype descriptions and relation descriptions are in line with observations of hold-out sets and synthetically constructed comparisons. The pipeline and evaluation metrics make use of a prompt-based approach to argument role labeling, which is also tested, showing current generations of models can reach argument labeling accuracies on par with baselines.

Keywords: Data Mining and Information Retrieval · Natural Language Processing · Social Impact of AI

1 Introduction

For better [20,30,34] or worse [2,15,26], the internet has become the dominant hub for human communication [27,33]. User comments, like those posted in response to news articles hosted online, offer one window into the wealth of content and information hosted online. However, ingesting and understanding these comments can be time consuming and costly. Large language models, specifically auto-regressive models fine-tuned to align with human objectives [22,37], have become foundation models [5] for a wide range of tasks [9,13,35]. The flexibility and capabilities of these models suggest that they can be used to build a pipeline capable of understanding the arguments present within a large body of natural language content, as one might find in the comments sections of large online news networks.

Published by Springer Nature Switzerland AG 2025
I. Maglogiannis et al. (Eds.): AIAI 2025, IFIP AICT 758, pp. 389–402, 2025.
https://doi.org/10.1007/978-3-031-96235-6_28

Following inspiration from the domains of argument and discourse modeling, one such approach is to consider the body of content as representing a conversation between a small set of representative, or prototypical authors. A large corpus of comments or other social content could be summarized as a network where nodes represent prototypical authors containing descriptions of their points of view, and edges represent possible arguments being made between them. This contribution describes such a pipeline. In it, the following research questions are answered:

1. Can we extract descriptions of representative, prototypical authors from a corpus of content using a Large Language Model?
2. Can we extract descriptions of relations between these prototypical authors using a Large Language Model?
3. How can we evaluate the quality of the descriptions of generated prototypical authors and the relations between them?

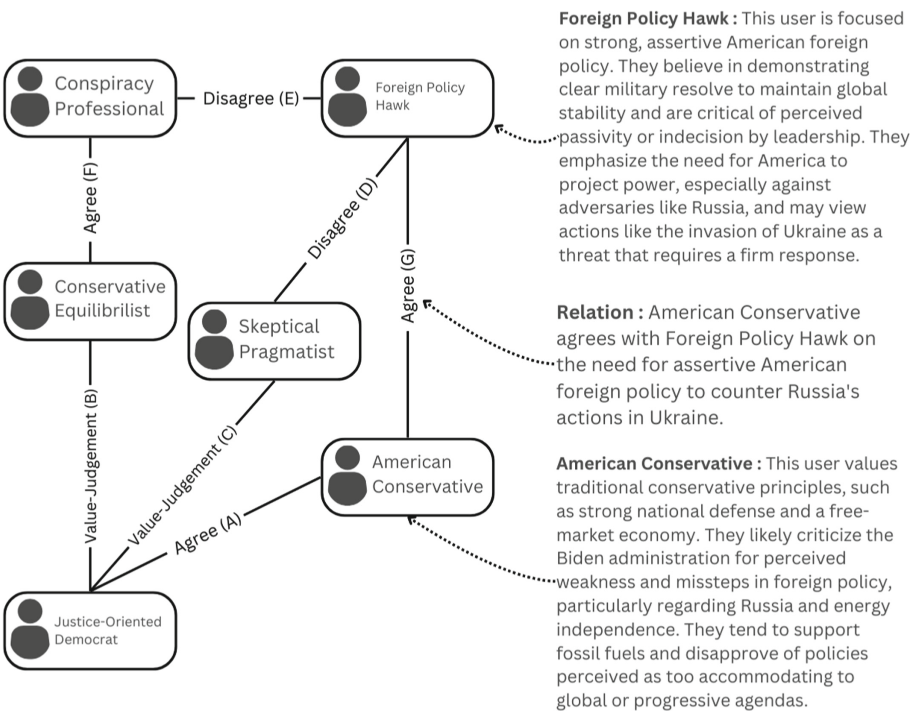

Fig. 1. Prototype Relation Network. This example highlights the possible content that can be obtained using the proposed pipeline. Two representative authors and a relation between them are depicted.

2 Related Work

The goal of the argument and discourse mining domains typically concerns themselves with the identification of argument and discourse elements [21,36], the identification of possible relations between these units [23,32], and the parsing of these relations into structures representing the overall conversation [7,12]. Work in this area often considers very specific, well labeled corpora, like manually parsed and annotated debate transcripts [19], the chat logs of complex social boardgames [29], or posts in online communities specific to persuasive discourse [8]. This is distinct from the much more general, and abstract setting of an arbitrary online discussion board, but the domain provides a rich body of underlying research and understanding to build from.

Large Language models have had a wide application, beyond the straightforward task of chatting [17], they can be used to help detect fake news [12], evaluate and build topic models [25], and even act as an agent within a complex social board game [16]. Effort has been made to apply the emerging class of large language models to the domain of argument mining, building from larger argument classification and labeling datasets [6]. By composing 'chain-of-thought' style representations of arguments and their clauses, the work of [18] leverages large models to predict the structure of a larger argument. Further, graph knowledge has been combined with contextual data to mine possible relations using large language models that can interact with knowledge graphs [28]. While these methods consider argument mining tasks, in a manner similar to the components of the described pipeline, they operate with full information of authorship and context, which is distinct from the conditions described here.

3 Methodology

This section describes the process of creating prototype relation networks intended to summarize the points of view and arguments contained within a large corpus of online comments. In this process, a network (an example can be seen in Fig. 1) is built where nodes represent prototypical authors of content, and edges represent argumentative relations between them. Prototype networks are generated by a pipeline, illustrated in Fig. 2, that interacts with generative, prompt-completion based large language models.

3.1 Generating Prototype Author Profiles

To generate the prototype profiles and their description, a prompt is constructed that contains the task instructions (seen in Listing 1.1), the text of a set of articles, and a set of user comments for these articles. The task description was crafted to encourage the model to find distinct sets of author descriptions, based on a formulation inspired by topic modeling approaches like those found in Latent Dirichlet Allocation [4] or Text-Based Ideal Points [31]. Specifically, the motivation for the task description is the hypothesis that the comments observed

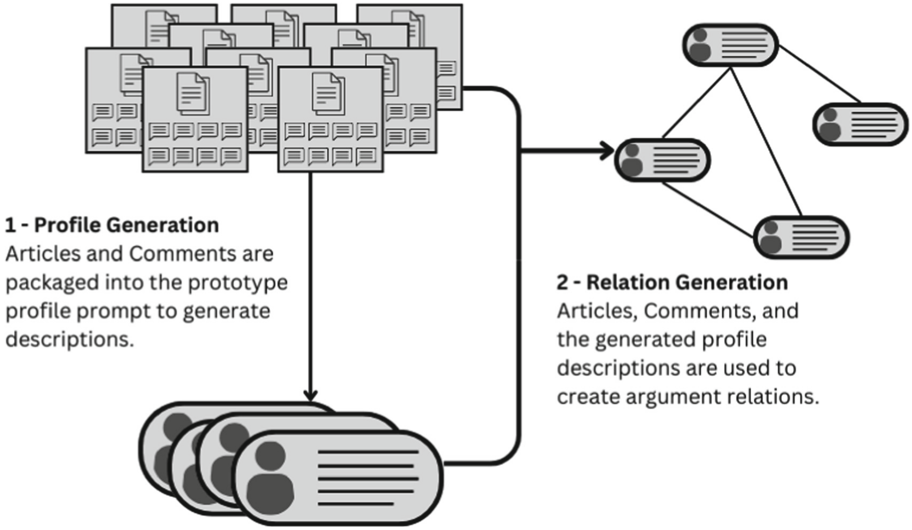

Fig. 2. Prototype Relation Network Pipeline. The corpus of content is first used to generate the prototype profile descriptions. Then, the corpus and generated profiles are used to generate argument relation edges between the representative prototype users. Not shown is the comment-labeling component of the pipeline, which is only used in the proposed evaluation method, not during typical generation of a network.

in the corpus could be generated by a probabilistic model. One where a comment is generated as a weighted combination of some influence from a small set of prototypical authors. These prototypical authors are described as needing to be representative of the view points being discussed within the corpus. The model prompt provides a short description of this assumed method of generation, and encourages descriptions that are similarly relevant and distinct. This complex task description is removed in the ablation studies, retaining only structural information relevant to proper execution and parsing.

Listing 1.1. Profile Generation Prompt

```
You will be sent sets of data that contain a summary of
an article , and a set of comments written by users who
have read the article . Once you have read each article ,
and all of the comments written about them, you will be
asked to create a set of descriptions of hypothetical
ideal users , that represent the topics and positions
being discussed in the comments.

You will assume that these ideal users represent a
distribution over terms and topics contained in the
article and comments. When creating the descriptions of
the ideal users , you will assume that each comment you
have read was written by an actual user , whose
```

```
viewpoints are some weighted sum of the viewpoints of
the ideal users.
```

3.2 Generating Argument Relations

To generate descriptions of the possible argumentative relations between the prototype profiles, a prompt is constructed that contains, as before, task instructions (seen in Listing 1.2), a description of the prototype profiles (the output of the first step of the pipeline), and a context (the articles and/or comments). The choice of role labels is motivated by the typical labels given in similar argument classification and argument role labeling tasks [19,29]. Argument classification concerns itself with parsing content, possibly dialogues, and extracting representations of possible arguments that are occurring with the text. That is, a body of text contains spans. Arguments can be represented as edges between certain groups of tokens within these spans. Argument mining is the process of extracting these groupings of tokens, and identifying possible edges. Argument classification is tasked with assigning labels to these edges. The prompt instructions contain examples of possible comment role labels, which are hand constructed and manually labeled from the corpus.

Listing 1.2. Relation Prompt

```
You will create a list of possible relations between comment
authors that will be described to you. These relations should
be based on the topics present in the articles provided. The
authors being described are authors of comments written in
response to these articles. Relationships should describe
agreements or disagreements about topics within the articles.
These relationships should be coherent with the content of
the provided comment author descriptions

Use relations of the following form:
Relation: Agreement
"Author A agrees with Author B on topic C"
"Author A agrees with Author B that the statements made
 by Russian sources are harder to trust than those made
 by their own government"

Relation: Disagreement
"Author A disagrees with Author B on topic C"
"Author A disagrees with Author B on the importance of
 the effects of climate change. They believe resources
 being spent on that issue are better spent on healthcare
 or the military."

Relation: Value-Judgment
"Author A thinks the belief B of Author C is X"
"Author A thinks that the belief of Author B, that the
 issue of climate change is overblown, is wrong. They
 think that people who hold this belief are ignorant."
```

3.3 Proposed Metrics

A set of metrics must be defined to evaluate the efficacy of the generated profiles and the relations between them. In this work, we propose methods of evaluating the quality of the profile content based on topic models and perplexity on various corpus', as well as distribution-distance based approaches to measuring the quality of the relations and their descriptions. In addition to these measures, a word-cloud based method of evaluating the content of the relations is used. The following sections will outline these metrics in more detail.

Topic-Based Prototype Description Metric. To evaluate the content of the prototype profiles, a measure based on the perplexity of a topic model is used. Building a topic model (specifically, the LDA model [4]) fit to the corpus of comments and articles used to create the prototype networks provides a way to investigate the topic-match between the content of the generated profiles, and the content of the underlying articles and comments. This work assumes that calculating the perplexity of a topic model with respect to a provided set of documents is a good proxy for quality of 'fit' between the model and the documents. By looking at the relative quality of the model against various document sets, we can understand how well the generated profiles fit with the underlying topic model of the corpus used. The topic model will have its perplexity evaluated against corpora composed of; (1) the generated profiles, (2) a set of hold-out comments, and (3) a random corpus with similar lengths to the hold-out comments. The random corpus is generated by creating a copy of the hold-out corpus, and randomly selecting new token-ids for each entry in the bag-of-words representation of the document.

Distribution-Distance Relation Metric. To evaluate the identified relations we consider the n-gram relation-label distribution for the generated relation descriptions. We consider the universe of possible events consisting of all pairs of possible n-grams and relation labels. This distribution can be calculated over the relation descriptions and for a hold-out set of validation comments. For the labeled comments, a large language model is used to label the corpus of comments with argument role labels, from the set of target role labels: Agrees, Disagrees, and Value-Judgment.

In these experiments, we consider the case where $n = 2$, or bigrams. We create a distribution that is defined as the set of possible pairs of bigrams and relation labels. That is, for a set of relation labels R containing 'Agrees', 'Disagrees', and 'Value-Judgment', and a set of possible bigrams from the corpus, $B = \{(i, j) | i, j \in C\}$, we consider the space of events defined by the cross product $R \times B$. By observing the distribution of these events from the text of the relations (pairing bigrams from the descriptions with their LLM-provided role label) and the text of the comments (by pairing bigrams from the comments with their LLM-provided role label) we define instances of these distributions. By considering the distances (via KL-divergence [14]) between these observed distributions, and other meaningfully-crafted distributions (Uniform and Random), we learn about the quality of the argument relations being generated by the pipeline.

Relation Word Cloud. We also consider word clouds to qualitatively evaluate the argument relations based on TF-IDF weighted term values. To do so, we consider each type of relation as a single 'document', and calculate TF-IDF values for each term as it appears within each of these single relation documents versus the others. By doing so, the terms most distinct to each relation, with respect to the entire set of relations, will be given higher weight than others. By doing this, and then plotting word clouds over the terms, we can get visual feedback for the quality of the overlap in 'distinct-to-the-relation' words for each relation for the generated content and the labeled-comments.

4 Experiments

To investigate the efficacy of the prototype relation network pipeline, the following experiments are conducted: (1) an overall evaluation of the pipeline using the metrics described in §3.1 and §3.2, (2) an evaluation of the comment-role labeling component using labeled argument classification datasets, and (3) an ablation study that considers the impact of the semantic task information contained in the pipeline prompts.

For each experiment, two different language models will be used, with slight differences in how the prompting occurs for each. These experiments interact with the GPT4-Omni model [1] and the DeepSeek-R1 model [10]. The main difference between these two (for the sake of this publication) is how structured output is handled. GPT4-Omni presents an API that allows for a specific definition of the schema for returned data, allowing for easy parsing of returned data into objects amenable to programmatic manipulation. This contrasts with the DeepSeek-R1 model, which requires explicit instructions within the prompt guiding the format of the returned data. This requires additional processing to parse the results and handle spurious, non-data content. Overall, the generation and collection of prompt returns, the GPT4-Omni model had zero parsing or formatting errors over the roughly 400 prompt requests made during pipeline execution. In contrast, the hand-crafted parsing approach used for the DeepSeek-R1 model generated four total errors in the same number of requests. Additionally, the relation descriptions returned by the DeepSeek-R1 model required manual correction to parse properly. This did not impact the quality of the resulting data, as the process for the DeepSeek-R1 pipeline was made error-tolerant, retrying to ensure each data batch obtained fully structured returns from the models. However, this human cost is not required by the Omni model results.

A dataset consisting of 20 batches of content, each containing five articles, with a minimum of 20 comments each from four news outlets (FOX, WSJ, ABC, HP) [11]. For each of the two large language models the batches of articles and comments are used to generate and evaluate prototype relation networks.

Networks are built using the full, and ablated prompts, and have the quality of their profile descriptions and relations evaluated based on the metrics described in §3.

4.1 Evaluating News Corpus Prototype Relation Network

In the first experiment, a set of prototype relation networks is generated from a corpus of online news articles and comments [11]. A set of 20 article and comment batches are randomly sampled from a set of online news outlets. Each of these batches is run through the pipeline, generating a set of prototype profile descriptions, a set of relation descriptions, and a set of role-labeled comments for use in evaluation.

To evaluate the profiles, the LDA model is fit to the underlying article and comment corpus. Then, using the same term-token dictionary, the evaluation corpus' are crafted. This includes a corpus based on a hold out comment set, the profile descriptions provided by the pipeline, and a synthetic random corpus. Finally, the perplexity of the comment model is calculated against these three evaluation corpora. To evaluate the relations, the bigram-relation distribution is generated for the dataset and the generated relation descriptions. Then, a random bigram-relation distribution is generated by changing the token ids within the bag-of-words representation of a sampled document to random values. Finally, the distances between the relation-based distribution and the evaluation distributions are calculated.

4.2 Evaluating LLM Argument Role Labeling

To show that the pipeline is justified in using large language models to label the source comments with role labels, another experiment is performed that assesses the ability of a prompt-based approach to argument role labeling. To do this two argument role classification datasets are found, each containing examples of content, labeled with one of two classes: 'attacks' or 'supports'. We suggest that this task is similar in scope and difficulty to the comment labeling task used in the described evaluation metrics. Further, the hypothesis is that high accuracy on this task on the M-arg and BRAT dataset will support the decision to use a large language model in the evaluation pipeline. The prompt used to obtain the comment labels can be seen in Listing 1.3.

Listing 1.3. Comment Argument Role Labeling Prompt

```
You will be tasked with labeling comments from an online news
outlet about an article with labels related to the comments
role in an argument. You will read the article , and then the
comments. For each comment you will provide an argument label.
Use the following argument role labels for comments
Label: Agreement
Use when:
 * The comment author agrees with the article about a
```

```
  topic or statement.
* The comment author agrees with another comment about
  a topic or statement.

Label: Disagreement
Use when:
* The comment author disagrees with the article about a
  topic or statement.
* The comment author disagrees with another comment
  about a topic or statement.

Label: Value-Judgment
Use when:
* The comment author is criticising a statement from the
  article or another comment.
* The comment author is insulting the author of the
  article, or the author of another comment.
```

4.3 Evaluating Prompt Ablation

In the final experiment, the same news comment corpus is used to evaluate the importance of the semantic task information provided in the system prompts used within the pipeline. To do this, the same pipeline and evaluation process is used. However, the prototype prompt, relation prompt, and role-labeling prompt will be reduced to include only structural information about the returned values, and have the elements describing or suggesting the impetus behind the task removed. The output of the pipeline will then be evaluated using the topic model and distribution-based methods described, as well as compared against the baseline methods of labeling comments.

5 Results

5.1 News Corpus Prototype Relation Networks

Table 1 reports the results from the profile evaluation metrics found using the two models, when using the full and ablated prompts. We see that the generated profiles have the smaller perplexity with respect to the corpus-based topic model. Further, the perplexity value of the hold-out comment set with respect to this topic model sits between the randomly generated document corpus and the profiles, suggesting that the generated profiles are an efficient representation of the topics. Further, ablation study results show that the removal of the task-specific instruction language within the prompt pushes the perplexity of the generated profile descriptions closer to that of the hold-out document set. This suggests that the output of the large language model has skewed closer to sampling the provided comments, than it has to the task of generating the profiles.

Table 2 reports the results from the relation evaluation metrics found using the two models, when using the full and ablated prompts. We see that for each

Table 1. Evaluating Prototype Profiles - Perplexity of an LDA topic model fit on the corpus of comment data with respect to the content of the profile descriptions and synthetically constructed, informative term topic vectors.

	Comment Model Perplexity		
	Profiles	Hold Out	Random
GPT4-Omni	-17.81	-18.13	-29.05
DeepSeekR1	-17.20	-20.81	-22.23
GPT4-Omni, Ablated	-18.78	x	x
DeepSeekR1, Ablated	-20.54	x	x

model the distance between the profiles and comments is the minimum, with a smaller distance than that of the hold-out comment set. Further, these distances are closer to zero than those of the uniform, and randomly constructed synthetic datasets. This suggests that the relations contain related content with similar arguments. When we consider the ablation study, we see that, overall, only the ablated GPT4 Omni model has a distance that suggests the quality of the relations moves closer to the hold-out sample when the task-specific semantic information is removed. This trend also holds with the DeepSeek models if the change in distance between the comment and hold-out distance is considered. The ablated DeepSeek relations move slightly towards that of the hold-out set.

Table 2. Evaluating Argument Relations - Reporting distance between the bigram-entity distribution observed in the prototype profile content vs the distributions observed from the comments, and generated synthetically.

	KL-Divergence Profiles vs:			
	Comments	Hold-Out Set	Uniform	Random
GPT4-Omni	0.5625	0.6039	0.7442	-1.0443
DeepSeekR1	1.1154	1.2714	1.4339	-1.3951
GPT4-Omni, Ablated	0.7210	1.2714	0.9625	-1.4406
DeepSeekR1, Ablated	0.9893	1.1299	1.2693	-1.1521

In Fig. 3, the qualitative evaluation, we can see that the content of the word clouds for each of relation-specific, weighted term sets, suggests a good overlap of content and intensity for the Agreement and Disagreement labels. However, this is less obvious in the content of the word clouds for the Value-Judgment relation. The output of the large language model descriptions for the relations contains significantly fewer negatively balanced words.

Fig. 3. Word clouds of TF-IDF weighted terms contained in the content for each type of relation.

5.2 Argument Role Labeling

In Table 3, we see the accuracy of prompt-based approach to the argument labeling task, broken down by label type. The results of the model approach for the M-arg dataset [19] align with the text-only baseline provided by the researchers (86%). Similarly, the accuracy of the pipeline approach on the persuasive essay dataset [29] sits between the 'human standard' (86%) and the baseline classifier (79%).

Table 3. Reporting the accuracy of prompt-based approaches to the argument role labeling task.

	Overall	Attacks	Supports
	M-Arg		
GPT4-Omni	0.853	0.961	0.745
DeepSeekR1	0.784	0.853	0.716
	BRAT		
GPT4-Omni	0.817	0.733	0.900
DeepSeekR1	0.711	0.601	0.908

6 Conclusions

The results of this study suggest that the proposed pipeline can generate profile descriptions that contain content that is in line with the content of the underlying comments. Further, the distance-based metrics suggest that the descriptions of the relations and the distribution of relations themselves, are in line with observed distributions of term-relation distributions from the underlying corpus. However, there is some discrepancy when considering the content of the Value-Judgment relation. The lack of negatively-balanced content within the word clouds suggests a bias occurring within the large language models. A possible hypothesis is that this is an artifact of the fine-tuning and alignment stage; models are encouraged to be helpful and harmless [3]. Such alignment might cause a bias against more 'confrontational' or adversarial content within the relation descriptions. Further work will consider the question of such bias as well as the application of prototype relation networks within downstream tasks, in which features derived from the networks can be used to aid in prediction tasks, like Social Media Popularity Prediction.

Acknowledgments. Research was sponsored by the DEVCOM Analysis Center and was accomplished under Cooperative Agreement Number W911NF-22-2-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the DEVCOM Analysis Center or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

1. Achiam, J., et al.: GPT-4 technical report. arXiv:2303.08774 (2023)
2. Alonzo, R., et al.: Interplay between social media use, sleep quality, and mental health in youth: a systematic review. Sleep Med. Rev. **56**, 101414 (2021)
3. Bai, Y., et al.: Training a helpful and harmless assistant with reinforcement learning from human feedback. arXiv preprint arXiv:2204.05862 (2022)
4. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. (2003)
5. Bommasani, R., et al.: On the opportunities and risks of foundation models. ArXiv (2021). https://crfm.stanford.edu/assets/report.pdf
6. Cheng, L., et al.: IAM: a comprehensive and large-scale dataset for integrated argument mining tasks. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (2022)
7. Choubey, P.K., et al.: Discourse as a function of event: profiling discourse structure in news articles around the main event. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (2020)
8. Dutta, S., et al.: Can Unsupervised knowledge transfer from social discussions help argument mining? In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (2022)

9. Fan, W., et al.: A survey on rag meeting LLMs: towards retrieval- augmented large language models. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (2024)
10. Guo, D., et al.: Deepseek-r1: incentivizing reasoning capability in LLMs via reinforcement learning. arXiv preprint arXiv:2501.12948 (2025)
11. He, L., et al.: On the dynamics of user engagement in news comment media. Wiley Interdisc. Rev. Data Min. Knowl. Disc. **10**(1) (2020)
12. He, X., et al.: Harnessing Explanations: LLM-to-LM interpreter for enhanced text-attributed graph representation learning. In: The 12th International Conference on Learning Representations (2023)
13. Hou, X., et al.: Large language models for software engineering: a systematic literature review. ACM Trans. Software Eng. Methodol. **33**(8) (2024)
14. Joyce, J.M.: Kullback-leibler divergence. In: International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg (2011). https://doi.org/10.1007/978-3-642-04898-2_327
15. Klikauer, T., Simms, N.: What is Digital Fascism? In: CounterPunch (2021)
16. Lan, Y., et al.: LLM-based agent society investigation: collaboration and confrontation in Avalon gameplay. In: Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (2024)
17. Lin, Y.-T., Chen, Y.-N.: LLM-Eval: unified multi-dimensional automatic evaluation for open-domain conversations with large language models. In: Proceedings of the 5th Workshop on NLP for Conversational AI (NLP4ConvAI 2023), pp. 47–58 (2023)
18. Liu, B., et al.: Argument mining as a multi-hop generative machine reading comprehension task. In: Findings of the Association for Computational Linguistics: EMNLP 2023 (2023)
19. Mestre, R., et al.: M-Arg: multimodal argument mining dataset for political debates with audio and transcripts. In: Proceedings of the 8th Workshop on Argument Mining (2021)
20. Mora-Rivera, J., Garcia-Mora, F.: Internet access and poverty reduction: evidence from rural and urban Mexico. Telecommun. Policy **45**(2) (2021)
21. Murray, G., et al.: Incorporating speaker and discourse features into speech summarization. In: Proceedings of the Human Language Technology Conference of the NAACL, Main Conference. Association for Computational Linguistics (2006)
22. Ouyang, L., et al.: Training language models to follow instructions with human feedback. Adv. Neural. Inf. Process. Syst. **35**, 27730–27744 (2022)
23. Paul, D., et al.: Argumentative relation classification with background knowledge. In: Computational Models of Argument. IOS Press (2020)
24. Peldszus, A.: Towards segment-based recognition of argumentation structure in short texts. In: Proceedings of the First Workshop on Argumentation Mining, pp. 88–97 (2014)
25. Pham, C., et al.: TopicGPT: a prompt-based topic modeling framework. In: 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (2024)
26. Richards, I.: A philosophical and historical analysis of "Generation Identity": fascism, online media, and the European new right. Terrorism Polit. Violence **34**(1) (2022)
27. Robertson, C.E., et al.: Negativity drives online news consumption. Nat. Hum. Behav. **7**(5) (2023)

28. Saadat-Yazdi, A., Pan, J.Z., Kokciyan, N.: Uncovering implicit inferences for improved relational argument mining. In: The 17th Conference of the European Chapter of the Association for Computational Linguistics (2023)
29. Stab, C., Gurevych, I.: Parsing argumentation structures in persuasive essays. Comput. Linguist. **43**(3) (2017)
30. Office of the Surgeon General et al.: Our epidemic of loneliness and isolation: the US Surgeon General's Advisory on the healing effects of social connection and community [Internet] (2023)
31. Vafa, K., Naidu, S., Blei, D.: Text-based ideal points. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (2020)
32. Varia, S., Hidey, C., Chakrabarty, T.: Discourse relation prediction: revisiting word pairs with convolutional networks. In: Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue (2019)
33. Walker, M., Matsa, K.E.: News Consumption Across Social Media in. In: Pew Centre Research (2021)
34. Wang, J., Xu, Y.: Internet usage, human capital and CO2 emissions: a global perspective. Sustainability **13**(15), 8268 (2021)
35. Wang, L., et al.: A survey on large language model based autonomous agents. Front. Comput. Sci. **18**(6) (2024)
36. Zhang, A., Culbertson, B., Paritosh., P.: Characterizing online discussion using coarse discourse sequences. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 11, issue 1 (2017)
37. Ziegler, D.M., et al.: Fine-tuning language models from human preferences. arXiv preprint arXiv:1909.08593 (2019)