Stakeholder Suite: A Unified AI Framework for Mapping Actors, Topics and Arguments in Public Debates

Mohamed Chenene[1] **Jeanne Rouhier**[1] **Jean Daniélou**[2] **Mihir Sarkar**[3] **Elena Cabrio**[4]
[1]ENGIE Lab CRIGEN, France
[2]Centre de Sociologie de L'Innovation, CNRS, UMR 9217, Mines ParisTech, PSL University
[3]ENGIE Research & Innovation, France
[4]Université Côte d'Azur, CNRS, INRIA, I3S, France
mohamed.chenene1@gmail.com, jeanne.rouhier@external.engie.com,
jeandanielou@live.fr, mihir.sarkar@engie.com, elena.cabrio@univ-cotedazur.fr

Abstract

Public debates surrounding infrastructure and energy projects involve complex networks of stakeholders, arguments, and evolving narratives. Understanding these dynamics is crucial for anticipating controversies and informing engagement strategies. This paper presents **Stakeholder Suite**, a framework deployed in operational contexts for mapping actors, topics, and arguments within public debates. The system combines actor detection, topic modeling, argument extraction and stance classification in a unified pipeline. Tested on multiple energy infrastructure projects as a case study, the approach delivers fine-grained, source-grounded insights while remaining adaptable to diverse domains. The framework achieves strong retrieval precision and stance accuracy, producing arguments judged relevant in 75% of pilot use cases. Beyond quantitative metrics, the tool has proven effective for operational use: helping project teams visualize networks of influence, identify emerging controversies, and support evidence-based decision-making.

1 Introduction

In the era of social media, public controversies can emerge rapidly and significantly affect the reputation and operations of organizations. Controversy analysis aims to systematically track and interpret such public debates by identifying the stakeholders involved, their relationships, their stances and the evolution of discourse over time. A structured understanding of these dynamics enables organizations to anticipate conflicts, adapt communication strategies, and mitigate potential delays or financial losses. In the energy sector, large-scale renewable infrastructure projects (e.g., wind farms, solar plants, district heating, hydrogen, or ammonia transport) frequently trigger public consultations and debates (LaPatin et al., 2023). Numerous stakeholders (citizens, associations, local officials and activists) express their views across diverse digital channels, making the monitoring and analysis of these discussions both complex and time-consuming for project developers.

This work addresses the question of how computational methods can support and enhance a transparent analysis of public debates. Our objective is to develop a *generic and adaptable framework* capable of mapping stakeholders, identifying debated topics, and extracting arguments and stances to provide an exhaustive and objective view of controversial debates. We introduce **Stakeholder Suite**, that combines Large Language Models and Retrieval-Augmented Generation (RAG) (Lewis et al., 2020) to build a structured representation of public debates. LLMs (Wei et al., 2021; Brown et al., 2020; Chen et al., 2024) now perform effectively for various NLP tasks, enabling fast deployment of complex analysis pipelines without large annotated datasets. RAG improves factual grounding and transparency by coupling generation with evidence selection, mitigating hallucinations, and improving traceability (Kalai et al., 2025).

This paper makes the following contributions: 1. We present a transparent and solid framework for the analysis of public debates, which jointly models stakeholders, topics, and arguments within a unified data structure; 2. We propose a fine-grained approach to topic modeling and argument analysis that leverages LLM-based reasoning, enabling richer and contextual insights. To date, the system has been deployed across more than ten renewable energy projects, offering project teams actionable insights into local dynamics and stakeholder perceptions. To the best of our knowledge, this is among the first operational frameworks that combine stakeholder mapping, fine-grained argument extraction, and topic modeling within a single and updatable pipeline spanning heterogeneous texts. A demo video is available.[1]

[1]https://youtu.be/c1b6QgyVkws

Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics
Volume 3: System Demonstrations, pages 1–20
March 24-29, 2026

2 Related Work

Mapping stakeholders and their stances across public debates intersects with several markets and research areas: media intelligence, social listening, public affairs, controversy analysis, policy analysis and argument mining (AM).

Commercial solutions. Social listening and sentiment analytics platforms as Talkwalker, Brandwatch or Meltwater, aggregate large-scale social and web data, offering dashboards for trends, topics, influencer identification, and sentiment. A second line of commercial tools centers on network views of entities, topics, and audiences. NetBase Quid and Pulsar Platform provide interactive graphs, topic clustering, community analysis, and audience segmentation across social and news sources. These capabilities help reveal ecosystems (e.g., co-mention networks and narrative clusters) and identify influential accounts. The Stakeholder Company (TSC.ai) maintains a large stakeholder repository with influence mapping and engagement workflows and comes closest to stakeholder-centric use cases in practice. Public affairs and government-relations platforms provide structured data and workflows for legislative and regulatory engagement. Quorum integrates bills, hearings, and institutional actors across multiple jurisdictions, supporting monitoring, alerting, and outreach. Some products also incorporate retrieval-augmented assistants for quick knowledge access. While effective within institutional policy contexts, these systems are not intended as general-purpose frameworks across heterogeneous public debates.[2]

Academic research. Foundational work in controversy analysis (Venturini and Munk, 2022) establishes sociological methods for systematically mapping controversies and constructing networks from debates. Automated policy analysis applies NLP pipelines to structure regulatory content, extracting measures, actors, and impacts from environmental and institutional documents (Firebanks-Quevedo et al., 2022; Singh et al., 2024). For instance, (Planas et al., 2022)'s framework is a knowledge graph-oriented approach that allows rapid review of policy documents through entity search, topic analysis, and policy search. While these approaches enable efficient document analysis, they typically operate on formal policy texts rather than public debate corpora. AM offers complementary methods for analyzing contested discourse, including claim detection, stance classification, and argument clustering (Daxenberger et al., 2020; Slonim et al., 2021) in different kinds of structured contexts: essays, online debate platforms, legal documents, and political debates. Political debate analysis, which examines clash points, strategies, and argumentative structures (South et al., 2020; Chen et al., 2025; Goffredo et al., 2023), shares similarities with energy infrastructure debates in involving political actors, citizen associations, and multichannel information flows. While prior work provides valuable components (e.g., large-scale data access, influencer and network analytics, institutional monitoring, and argument mining methods), an end-to-end production-oriented approach that jointly maps stakeholders, detects and organizes topics, and extracts source-grounded arguments with stance across heterogeneous debate corpora has not yet been proposed. This gap motivates the development of retrieval-augmented frameworks adaptable across domains where public debate and stakeholder engagement are central.

[2]Commercial platforms discussed: Talkwalker (https://www.talkwalker.com), Brandwatch (https://www.brandwatch.com), Meltwater (https://www.meltwater.com), NetBase Quid (https://www.quid.com), Pulsar Platform (https://www.pulsarplatform.com), TSC.ai (https://tsc.ai), Quorum (https://www.quorum.us).

3 Methodology

The goal of **Stakeholder Suite** (Figure 1) is to provide a comprehensive snapshot of a public debate. Its components are designed to address three main questions: (1) *Who is speaking?* → Actor detection; (2) *What are they talking about?* → Topic modeling and extraction; (3) *What is their position?* → Argument mining and stance classification.

3.1 Data Collection

To start, two main components are required:

(1) **Stakeholder list:** manually or automatically identified actors participating in the debate;

(2) **Document database:** a corpus integrating diverse textual sources on the debate, including: (a) **Debate-centric data:** press articles, policy papers and administrative documents, typically obtained through data brokers, web scraping or manually uploaded; (b) **Actor-centric data:** social media posts, websites and blogs associated with listed stakeholders.

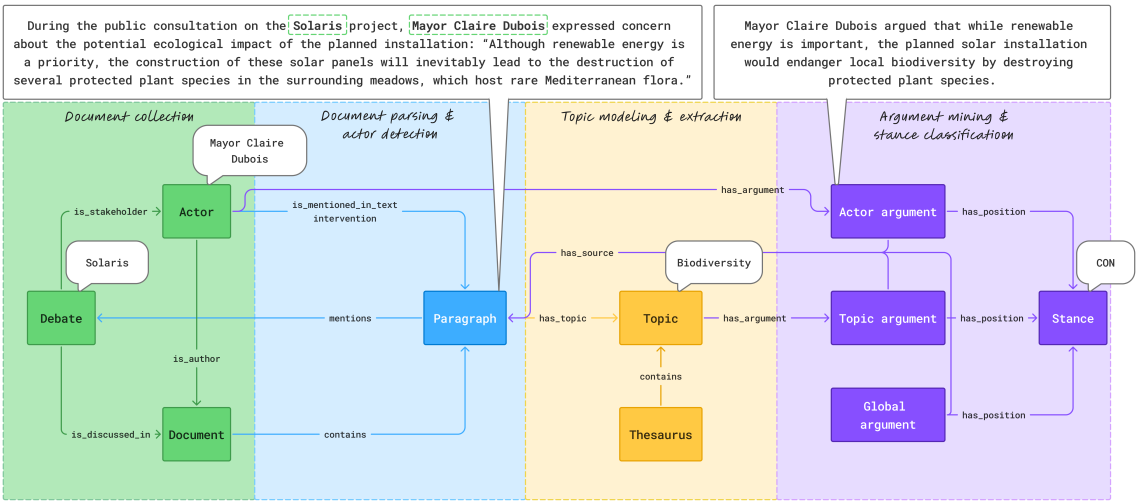

Figure 1: Blueprint of the **Stakeholder Suite** framework with an actor argument example: (1) Construction of a database containing stakeholders and documents related to the public debate; (2) Parsing of documents into paragraphs and detection of actor mentions and interventions within the text; (3) Linking of paragraphs to topics from the thesaurus using semantic similarity; (4) Generation of arguments through a RAG pipeline along three analytical axes (global, topic-specific, actor-specific) together with automatic stance classification for each argument.

Data quality. Both data types are essential for ensuring that stakeholders are properly represented within the corpus. The most informative argumentative content often originates from stakeholder-generated material. Based on our experience in the energy sector, high-yield argumentative sources include public debate transcripts, stakeholder blog posts, and social media statements. Conversely, press and institutional documents tend to be less argumentative but provide valuable contextual information for topic modeling.

3.2 Data parsing and actor detection

Data parsing. After constructing the database, the next step is to structure it and establish semantic links between entities. Given that sentences are often too short to capture full meaning, while documents can vary greatly in length, we adopt the paragraph as the atomic analysis unit, as it provides a balance between contextual coherence and processing granularity. In our implementation, paragraph segmentation, coreference resolution, and actor speech extraction are handled jointly through a single LLM prompt (see Appendix A.1). All paragraphs are embedded and indexed in OpenSearch vector database for full-text and semantic search.

Actor detection. Once paragraphs are extracted, we link them to the corresponding actors from the stakeholder list. We define three relation types:
(1) *is_author:* the actor of the document (e.g., blog post, tweet), known from the actor-centric corpus.
(2) *is_mentioned_in_text:* explicit textual mentions of an actor in a paragraph. To improve the precision for common or ambiguous names, we search for direct matches using the known actor list and verify them with FLAIR NER (Akbik et al., 2019).
(3) *intervention:* actor speech segments in debate transcripts. Interventions are more valuable content than mentions because they report the actor's speech. We use the LLM parser mentioned above.

Mentions of the debate are also detected at the paragraph level using a predefined list of aliases. This step is necessary because, even if a document refers to the debate as a whole, not all paragraphs within it necessarily discuss the debate directly.

3.3 Topic modeling and extraction

With a structured database of paragraphs linked to actors, the next step is to identify the topics discussed. Applying standard unsupervised clustering to highly domain-specific corpora often produce generic or irrelevant topics. Following (Tamkin et al., 2024), we leverage LLMs to construct a domain-specific thesaurus: a structured list of topics and their corresponding descriptions. The process involves the following steps:
(1) **Corpus chunking:** split the database into coherent subsets according to the debate characteris-

tics. For long-running cases, use temporal windows (e.g., monthly or quarterly); for political debates, segment by speaker or party.
(2) **Chunk description via RAG:** for each subset, run a RAG with a query such as: *"What are the emerging topics related to {debate_name}?"*, and extract the resulting topic candidates.
(3) **Topic clustering:** cluster the generated topics using methods such as HDBSCAN or Spectral Clustering based on a semantic affinity matrix (Campello et al., 2013; Shi and Malik, 2000).
(4) **Cluster summarization:** apply an LLM to name each cluster and provide representative examples, producing two descriptive levels (topic and subtopic) to preserve granularity.
(5) **Thesaurus aggregation:** run a final LLM pass to consolidate results into a unified thesaurus containing topics, descriptions, and subtopics.

Although this pipeline heavily relies on LLMs, it has proved most effective for generating high-resolution topic descriptions. Moreover, analyzing topics across temporal slices reveals trends and weak signals of business relevance. Empirically, naive clustering methods consistently failed to capture debate-specific nuances, whereas our LLM-augmented approach produced coherent, non-generic topics. Appendix A.2 provides an example of a thesaurus built with this approach and the prompts used for steps 4 and 5. Once the thesaurus is defined, topic occurrences are computed across the corpus. We compute cosine similarity between topic descriptions and paragraph embeddings, retaining matches above a predefined threshold ($t = 0.15$) to balance quality and quantity of paragraphs per topic. When multiple topics qualify, the top three by similarity are selected.

3.4 AM and stance classification

We adopt a RAG architecture for argument detection, because (i) it is easily implemented in production environments through LLM API calls (in particular, we use `gpt-4o-mini` (OpenAI et al., 2024) on the Azure platform, selected for compliance and cost-quality balance); (ii) it enables the generation of concise, one-sentence arguments that are easily interpretable by end users; (iii) it preserves traceability by maintaining explicit links between generated outputs and their source documents.

A known limitation of RAG, however, is its non-exhaustive coverage of the corpus, as it retrieves only the top-K relevant passages prior to generation. To mitigate this, we partition the database into coherent chunks and query each subset independently, allowing the extraction of more than 300 arguments per debate. This strategy achieves a practical balance between corpus coverage, computational cost, and analytical depth.

3.4.1 Temporal and Dimensional Splits

To maximize coverage of the public debate and reveal dynamics, we partition the corpus along two axes: (i) **time**, to trace how arguments evolve; and (ii) **dimension**, to focus on *stance*, *actors*, or *topics* derived from previous steps. We operationalize this with three uniform query families.

(1) Global argument. **Goal:** enumerate debate arguments *in favor of* or *against* for each year.

Database filter (pseudo-Cypher):
```
MATCH (p:PARAGRAPH) WHERE p.date.year = $year
RETURN p
```

Prompt to LLM: *"What are all the arguments {in favor/against} {debate_name}?"*

Rationale: yearly stance snapshots surface trends and turning points.

(2) Actor argument. **Goal:** extract arguments attributable to a specific stakeholder, prioritizing paragraphs that are most on-topic for the target debate (no temporal split due to data sparsity).

Database filter (pseudo-Cypher):
```
(a:ACTOR)->(p:PARAGRAPH)-[:MENTIONS]->(d:DEBATE)
RETURN p AS p_debate
(a:ACTOR)->(p:PARAGRAPH)-[:HAS_TOPIC]->(:TOPIC)
RETURN p AS p_topic
(a:ACTOR)->(p:PARAGRAPH) RETURN p AS p_all
IF COUNT(p_debate) > 25 THEN RETURN p_debate
ELSE IF COUNT(p_topic) > 25 THEN RETURN p_topic
ELSE RETURN p_all
```

Prompt to LLM: *"What are the arguments of the stakeholder {actor_name} about {debate_name}?"*

Rationale: We prioritize *debate-specific* paragraphs for precision; if these are insufficient (≤ 25), we fall back to *topic-linked* paragraphs, and otherwise include *all* paragraphs for coverage. The threshold was determined empirically from a study of more than 150 actors and was optimized to maximize the number of stances extracted.

(3) Topic argument. **Goal:** collect arguments tied to a specific topic, per year, decoupled from the debate stance if necessary.

Database filter (pseudo-Cypher):
```
MATCH (p:PARAGRAPH)-[:HAS_TOPIC]→(t:MACRO_TOPIC)
WHERE p.date.year = $year AND t.name = $topic_name
RETURN p
```

Prompt to LLM: *"What are the arguments related to the topic {topic_name}?"*

Rationale: topics may drift from debate framing; isolating them improves thematic resolution. It can also give a broader vision of the subject.

For a setup with ∼50 actors, 8 topics, and a 5-year span, and assuming ∼3 arguments are generated per query we get approximately 300 arguments per debate. Combining temporal and dimensional partitions with standardized prompts yields a scalable, interpretable, and trend-aware argument set while keeping retrieval focused and noise low.

3.4.2 RAG Pipeline

The RAG pipeline structures the final argument extraction process. For each query identifier, it operates in four sequential stages:
(1) **Retrieval:** the database is pre-filtered according to the previously mentioned split. We retrieve the top-K paragraphs ($K = 25$) using Maximum Marginal Relevance (Carbonell and Goldstein, 1998) with $\lambda = 0.8$ to balance relevance and diversity.
(2) **Argument generation:** the query and retrieved paragraphs are passed to the LLM, which returns a structured list of arguments, each explicitly linked to one or more source paragraphs. The output is then parsed using regex to extract the argument text and the corresponding source identifiers.
(3) **LLM-as-a-Judge** (Zheng et al., 2023): a second LLM performs an automatic quality check, verifying that each generated argument is supported by its sources and flagging potential hallucinations.
(4) **Stance classification:** each validated argument is assigned a stance label (`PRO`, `CON`, or `NEUTRAL`).
Appendix A.3 shows the prompts for steps 2 to 4, together with examples of extracted arguments.

3.5 Visualization: Stakeholder Mapping

To visualize stakeholder relationships, we construct a connected network where each node represents an actor. Node color encodes stance (`green` = pro, `red` = con, `grey` = neutral), and node size reflects the frequency of debate mentions during actor interventions. Edges are drawn when one actor explicitly references another within an intervention, capturing direct interactions and influence links. An example can be found in Appendix A.4.

4 Evaluation

As previously described, our framework addresses three main tasks, i.e., NER for actor detection, semantic text similarity for topic extraction and RAG for argument generation. Since for the first two we rely on standard, validated approaches (Akbik et al., 2018; Muennighoff et al., 2023), in the following we focus on the RAG pipeline, with emphasis on retrieval, generation, and stance classification.

4.1 Retrieval Evaluation

Retrieval performance was assessed across eight energy projects for both `PRO` and `CON` stances (16 queries in total). Each evaluation mixed relevant and randomly sampled paragraphs to approximate realistic corpus conditions, producing datasets of 200 candidates per query, with an average of about 35 relevant documents. We report Precision@K as the primary metric, focusing on diversity–relevance trade-offs using Maximum Marginal Relevance (MMR). As shown in Table 1, optimal performance occurs for $\lambda \in [0.7, 0.8]$, confirming that moderate diversity does not deteriorate retrieval quality. The mean Precision@20 reaches 0.59, with variability driven by project-specific content density, sufficient for reliable argument generation downstream.

MMR λ	Precision@10	Precision@20
0.5	0.41±0.13	0.40±0.09
0.6	0.59±0.17	0.55±0.14
0.7	**0.64±0.20**	0.57±0.18
0.8	0.64±0.25	**0.59±0.17**
0.9	0.63±0.26	0.58±0.19
1.0	0.63±0.26	0.58±0.19

Table 1: Retrieval performance across different λ values. Metrics are reported as mean ± standard deviation.

Component	Class	Precis.	Recall	F1
LLM-as a-Judge	BAD	0.42	0.75	0.54
	GOOD	0.77	0.45	0.57
	Macro avg	**0.59**	**0.60**	**0.55**
Stance Classification (ACTOR)	AGAINST	0.96	0.93	0.94
	NEUTRAL	0.69	0.76	0.73
	PRO	0.66	0.60	0.63
	Macro avg	**0.77**	**0.76**	**0.77**
Stance Classification (TOPIC)	AGAINST	1.00	0.61	0.76
	NEUTRAL	0.72	0.89	0.80
	PRO	0.78	0.70	0.74
	Macro avg	**0.83**	**0.73**	**0.76**

Table 2: Performance of the LLM-as-a-Judge and Stance Classification modules.

4.2 LLM-as-a-Judge: Generation Evaluation

To assess generation quality, a secondary LLM evaluated each argument-source pair to identify unsupported or hallucinated outputs. Our primary objective was to minimize false positives visible to end users, prioritizing a low error rate over exhaustive recall. In practice, this meant maximizing `BAD` argument recall while maintaining acceptable coverage

of GOOD ones, to preserve user trust and prevent misinterpretation of the system's outputs. Several prompt configurations were tested to balance strictness and recall. On a validation set of 739 samples, the best-performing validator achieved high BAD recall to protect users from unsupported claims, accepting some loss of GOOD recall (Table 2).

4.3 Stance Classification

Following the cross-topic argument classification approach proposed by (Stab et al., 2018), 3 annotators with expertise in Computational Linguistics annotated 350 arguments with their true stance toward one energy project to evaluate LLM classification performance. As shown in Table 2, we report the overall performance. For reference, (Reimers et al., 2019) estimates human performance on cross-topic stance classification at 0.81 (F1-score). Our results fall close to this upper bound, indicating reliable performance for our application.

5 Case Studies

5.1 A Case Study for Users Evaluation

User evaluations globally demonstrate that the system's analytical outputs are both accurate and operationally valuable (Appendix A.4 shows screenshots of the application). The application was used by 10 end-users across 9 different energy pilots. During onboarding, roughly 600 documents were processed per project, followed by about 20 new documents per month for continuous monitoring. Users particularly valued the argument summaries and mapping visualizations for improving internal communication and evidence-based engagement planning (*"Stakeholder Suite gives us a territorial radar. It saves us days of reading and helps us get straight to what really matters."*, from a pilot user's notes). In parallel, we directly collected their feedback through the application. On a sample of approximately 200 arguments, 75% were judged relevant or useful by project teams.

In the following, we present some insights gathered with the Stakeholders Suite on a solar project. Located in southern France, the *Montagne de Lure* area has experienced several solar development projects, some of which have triggered tensions among local associations and residents. Using Stakeholder Suite, project teams aggregated data from past and ongoing debates to identify active opponents, supportive officials, and recurring public concerns such as deforestation, landscape alteration, and threats to protected species.

Topic Extraction. Automatic clustering surfaced two dominant themes, i.e., land use conflicts and regulation and participation. The land-use cluster revealed strong attention to biodiversity and landscape protection, with repeated mentions of the ocellated lizard, a locally protected species. This insight led the project team to initiate additional environmental assessments before site validation. The regulation cluster showed that many objections targeted the lack of transparency and citizen participation rather than the solar technology itself, prompting stronger consultation efforts. A third cross-topic finding indicated a clear preference for solar installations on brownfields or rooftops instead of forests or farmland.

Argument Analysis. AM and stance classification exposed a well-structured opposition led by environmental groups. Their discourse gained visibility through local media and a published book that became a symbolic reference against solar projects in forested areas. Conversely, arguments from supportive actors (including other project owners) highlighted the site's strong solar potential and its strategic alignment with EU renewable goals, offering communication models for future initiatives.

Network Insights. It revealed tightly connected opposition clusters, suggesting coordinated activism rather than isolated critics. It also identified intermediary actors bridging detractors and supportive municipalities, indicating potential spaces for dialogue and balanced engagement.

5.2 Application to Political Debate

Context. To assess the adaptability of the framework, we applied it to a political debate in the energy domain. The Programmation Pluriannuelle de l'Énergie (PPE3) defines France's energy strategy for the period 2022-2035, covering the reduction of fossil fuel dependence, expansion of renewable energy sources (solar, wind, biogas, hydrogen), maintenance of the nuclear fleet, energy efficiency measures, grid modernization and transport decarbonization. The plan is shaped through a public consultation process involving parliamentary debates and stakeholder contributions.[3]

Data sources. The pipeline is applied to three corpora drawn from the PPE 3 consultation:
(1) **Sénat:** 2 transcripts of senatorial debates (1,224 interventions from 41 senators). (2) **As-**

[3]The code used for this experiment is available at https://github.com/stakeholder-suite/demo_eacl_2026

semblée Nationale (AN): 3 transcripts of debates in the lower chamber (722 interventions from 99 deputies). (3) **Actor Workbook:** 329 position papers from institutional stakeholders (industry federations, NGOs, local authorities, energy operators), each corresponding to one organization.

Thesaurus construction. The thesaurus was constructed using the actor workbooks, which provides a diverse set of viewpoints comparable to the press corpus used in previous experiments. Topic clustering was performed using HDBSCAN, although spectral clustering may offer finer control over the number of clusters.

Argument extraction. In this setting, we focus on actor and topic-level arguments. For each stakeholder, arguments are generated using our RAG approach. Given the broad scope of PPE3, global stance questions were not appropriate. Therefore, we retain only arguments that explicitly mention specific energy sources (solar, nuclear, wind) using keyword filtering. For each selected argument, stance is then computed relative to the identified energy source through a secondary query. This process yielded more than 2,000 arguments, including over 500 energy-specific arguments.

Insights. The analysis shows that a large proportion of arguments focus on nuclear energy, with most parliamentary actors expressing support for this source. In contrast, wind energy appears more contested, with a higher proportion of opposing arguments in parliamentary debates. These results demonstrate the framework's ability to extract source-grounded arguments and capture stakeholder positions from relatively small corpora.

6 Conclusion

We presented a production-oriented framework for analyzing public debates that unifies *stakeholders*, *topics*, and *arguments* in a single data model. The system combines paragraph-level retrieval, LLM-based argument extraction, and stance classification to produce source-grounded insights that scale across domains. Tested on multiple energy-infrastructure projects, and successfully piloted in legislative contexts, the Stakeholder Suite offers transparent, reusable workflows that complement existing media intelligence and public-affairs tools. Our evaluation indicates that diversity-aware retrieval provides adequate coverage for downstream generation, and that stance classification achieves high accuracy at the paragraph–argument level. LLM-as-a-judge effectively filters unsupported claims but can be conservative, discarding valid arguments in ambiguous contexts.

7 Ethical & broader considerations

Stakeholder management has become a key issue for multinational corporations across the world. The threat of controversies led by the stakeholders of an industrial project can result in project cancellation, especially in the field of major infrastructure projects (energy, transportation, telecommunication. . .). Public affairs and corporate social responsibility departments traditionally in charge of stakeholder management and controversy risk mitigation, are facing a new challenge, namely the growing digitization of controversies happening on a wide array of social media platforms. Being able to make sense of all the digital traces stakeholders leave on social media, in press releases, meeting minutes of public debates, etc., requires the elaboration of new management tools connected to heterogeneous data sources in different languages. AI-based solutions are bringing a wind of change in corporate daily practices for both managing stakeholders' data and designing communication strategies. It seems very likely that multinational corporations will equip themselves with AI-based solutions to refine the monitoring of potential controversies and better manage associated risks. The implementation of such solutions in the corporate toolbox implies actions of change management to ensure solution adoption and effective use. Research in sociology of innovation has shown that the disruptive impact of innovation on daily routines can lead to users' resistance and cause innovative projects to fail (Akrich et al., 2002). To become a success, AI-based stakeholder management solutions will have to integrate a change management strategy to ensure adoption in public affairs and corporate social responsibility departments. Regarding the European geographical area, one limitation of the development of such tools will be the respect of GDPR, which necessitates privacy-by-design solutions when manipulating social media data. We process publicly available texts only; personal data are handled under GDPR legitimate-interest.

Acknowledgments

We thank Stakeholder Suite project team for their support and contributions throughout this work, including Alexis Courtin, Elena Hinnekens, Juliette

Lagrange, Bilel Loussaief, Gilles Olivié, Ousmane Traoré. This work has been funded by ENGIE Research & Innovation. The work of E. Cabrio has been supported by the French government, through the 3IA Cote d'Azur investments in the project managed by the National Research Agency (ANR) with the reference number ANR-23-IACL-0001.

References

Alan Akbik, Tanja Bergmann, Duncan Blythe, Kashif Rasul, Stefan Schweter, and Roland Vollgraf. 2019. FLAIR: An easy-to-use framework for state-of-the-art NLP. In *NAACL 2019, 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pages 54–59.

Alan Akbik, Duncan Blythe, and Roland Vollgraf. 2018. Contextual string embeddings for sequence labeling. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1638–1649, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Madeleine Akrich, Michel Callon, and Bruno Latour. 2002. The Key to Success in Innovation Part I: The Art of Interessement. *International Journal of Innovation Management*, 6(2):187 – 206.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, and 12 others. 2020. Language models are few-shot learners. *CoRR*, abs/2005.14165.

Ricardo J. G. B. Campello, Davoud Moulavi, and Joerg Sander. 2013. Density-based clustering based on hierarchical density estimates. In *Advances in Knowledge Discovery and Data Mining*, pages 160–172, Berlin, Heidelberg. Springer Berlin Heidelberg.

Jaime Carbonell and Jade Goldstein. 1998. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '98, page 335–336, New York, NY, USA. Association for Computing Machinery.

Guizhen Chen, Liying Cheng, Anh Tuan Luu, and Lidong Bing. 2024. Exploring the potential of large language models in computational argumentation. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2309–2330, Bangkok, Thailand. Association for Computational Linguistics.

Qianhe Chen, Yong Wang, Yixin Yu, Xiyuan Zhu, Xuerou Yu, and Ran Wang. 2025. Conch: Competitive debate analysis via visualizing clash points and hierarchical strategies. *Preprint*, arXiv:2507.14482.

Johannes Daxenberger, Benjamin Schiller, Chris Stahlhut, Erik Kaiser, and Iryna Gurevych. 2020. Argumentext: Argument classification and clustering in a generalized search scenario. *Datenbank-Spektrum*, 20(2):115–121.

Daniel Firebanks-Quevedo, Jordi Planas, Kathleen Buckingham, Cristina Taylor, David Silva, Galina Naydenova, and René Zamora-Cristales. 2022. Using machine learning to identify incentives in forestry policy: Towards a new paradigm in policy analysis. *Forest Policy and Economics*, 134:102624.

Pierpaolo Goffredo, Elena Cabrio, Serena Villata, Shohreh Haddadan, and Jhonatan Torres Sanchez. 2023. DISPUTool 2.0: A Modular Architecture for Multi-Layer Argumentative Analysis of Political Debates. In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI-23 Special Programs, IAAI-23, EAAI-23, Student Papers and Demonstrations, pages 16431–16433, Washinghton, DC, United States. AAAI.

Adam Tauman Kalai, Ofir Nachum, Santosh S Vempala, and Edwin Zhang. 2025. Why language models hallucinate. *arXiv preprint arXiv:2509.04664*.

Michaela LaPatin, Lauryn A. Spearing, Helena R. Tiedmann, Miriam Hacker, Olga Kavvada, Jean Daniélou, and Kasey M. Faust. 2023. Controversy in wind energy construction projects: How social systems impact project performance. *Energy Policy*, 176:113507.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks. *CoRR*, abs/2005.11401.

Niklas Muennighoff, Nouamane Tazi, Loic Magne, and Nils Reimers. 2023. MTEB: Massive text embedding benchmark. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 2014–2037, Dubrovnik, Croatia. Association for Computational Linguistics.

OpenAI, :, Aaron Hurst, Adam Lerer, Adam P. Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, Aleksander Mądry, Alex Baker-Whitcomb, Alex Beutel, Alex Borzunov, Alex Carney, Alex Chow, Alex Kirillov, and 401 others. 2024. Gpt-4o system card. *Preprint*, arXiv:2410.21276.

Jordi Planas, Daniel Firebanks-Quevedo, Galina Naydenova, Ramansh Sharma, Cristina Taylor, Kathleen Buckingham, and Rong Fang. 2022. Beyond modeling: NLP pipeline for efficient environmental policy analysis. *CoRR*, abs/2201.07105.

Nils Reimers, Benjamin Schiller, Tilman Beck, Johannes Daxenberger, Christian Stab, and Iryna Gurevych. 2019. Classification and clustering of arguments with contextualized word embeddings. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 567–578, Florence, Italy. Association for Computational Linguistics.

Jianbo Shi and J. Malik. 2000. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905.

Prashant Singh, Erik Lehmann, and Mark Tyrrell. 2024. Climate policy transformer: Utilizing NLP to track the coherence of climate policy documents in the context of the Paris agreement. In *Proceedings of the 1st Workshop on Natural Language Processing Meets Climate Change (ClimateNLP 2024)*, pages 1–11, Bangkok, Thailand. Association for Computational Linguistics.

Noam Slonim, Yonatan Bilu, Carlos Alzate, Roy Bar-Haim, Ben Bogin, Francesca Bonin, Leshem Choshen, Edo Cohen-Karlik, Lena Dankin, Lilach Edelstein, Liat Ein-Dor, Roni Friedman-Melamed, Assaf Gavron, Ariel Gera, Martin Gleize, Shai Gretz, Dan Gutfreund, Alon Halfon, Daniel Hershcovich, and 34 others. 2021. An autonomous debating system. *Nature*, 591(7850):379–384.

Laura South, Michail Schwab, Nick Beauchamp, Lu Wang, John Wihbey, and Michelle A. Borkin. 2020. Debatevis: Visualizing political debates for non-expert users. In *2020 IEEE Visualization Conference (VIS)*, pages 241–245.

Christian Stab, Tristan Miller, Benjamin Schiller, Pranav Rai, and Iryna Gurevych. 2018. Cross-topic argument mining from heterogeneous sources. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3664–3674, Brussels, Belgium. Association for Computational Linguistics.

Alex Tamkin, Miles McCain, Kunal Handa, Esin Durmus, Liane Lovitt, Ankur Rathi, Saffron Huang, Alfred Mountfield, Jerry Hong, Stuart Ritchie, Michael Stern, Brian Clarke, Landon Goldberg, Theodore R. Sumers, Jared Mueller, William McEachen, Wes Mitchell, Shan Carter, Jack Clark, and 2 others. 2024. Clio: Privacy-preserving insights into real-world ai use. *Preprint*, arXiv:2412.13678.

Tommaso Venturini and Anders Kristian Munk. 2022. *Controversy mapping : a field guide*. Polity, Cambridge, UK.

Jason Wei, Maarten Bosma, Vincent Y. Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M. Dai, and Quoc V. Le. 2021. Finetuned language models are zero-shot learners. *CoRR*, abs/2109.01652.

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric P. Xing, Haotong Zhang, Joseph E. Gonzalez, and Ion Stoica. 2023. Judging llm-as-a-judge with mt-bench and chatbot arena. *ArXiv*, abs/2306.05685.

A Appendix

A.1 Data parsing prompt

You will find here the prompt used for the parsing of documents.

Coreference Resolution & Speaker Detection

System Prompt: Role

You are an *NLP engine specialized in text disambiguation and speaker identification*. Your task is to transform ambiguous documents into self-contained, RAG-ready paragraphs with clear entity references and speaker attribution.

Context

- **Document author**: {doc_editor}
- **Document title**: {doc_title}
- **Original language**: Preserved in output
- **Target use case**: Retrieval-Augmented Generation (RAG)

Expected Task

1. Coreference Resolution

1. Replace **all pronouns** (I, you, he, she, it, they, je, il, elle, ils, etc.) with their specific referent.
2. Replace **vague references** ("this project", "the park", "Madam", "Monsieur") with explicit names or descriptive phrases.
3. If an entity is unnamed, create a **short descriptive identifier** based on context.
4. Make each paragraph **self-contained and unambiguous**, preserving order, style, and all factual details.

2. Speaker Detection & XML Tagging

1. Wrap every paragraph in: `<p speakerName="" speakerFunction="">`.
2. Populate **speakerName** when the paragraph contains:
   - A direct quote (« ... », " ... "), **or**
   - An indirect quote with reporting verbs (said, declared, stated, according to, etc.)
3. **speakerFunction** = entity's role *only if stated or clearly inferable*.
4. If **multiple speakers** appear in one paragraph, split it so each `<p>` has a single speaker.
5. If a name is only **mentioned** (not speaking), leave both attributes **empty**.
6. Always output the speaker's name and function *in the paragraph text*.

3. Document Structure

1. Merge short, consecutive, related sentences from the **same speaker**.
2. Never merge different speakers.
3. Avoid small paragraphs lacking sufficient context to be understood alone.
4. Produce **valid XML**; do not modify text outside `<content>`.

Content Constraints

- Always include the **project name** in every relevant paragraph.
- Avoid generic references: *not* "the solar farm is too big" *but* "the solar project Zephyr is too big".
- Keep the **original language** and all numerical/factual elements intact.
- Output language: **{output_language}**.

Expected Output (XML)

```
<edited_content>
  <p speakerName="" speakerFunction="">
    The Zephyr solar project was initiated last year...
  </p>
  <p speakerName="Jean_Daniel" speakerFunction="Mayor_of_Duneshale">
    "The_Zephyr_project_is_very_important..." said
    Jean Daniel, the Mayor of Duneshale.
  </p>
  ...
</edited_content>
```

Example

- Input: "The city hall is against the project."
  Output: "The city hall of Duneshale is against the Zephyr project."
- Input: "He gave an opinion against the park installation."
  Output: "The mayor of Duneshale, Jean Daniel, gave an opinion against the Zephyr solar park installation."

User Input: Document to process:

```
<document>
  <author>{doc_editor}</author>
  <title>{doc_title}</title>
  <content>{doc_content}</content>
</document>
```

Return the transformed text in a single `<edited_content>` block containing ordered `<p>` tags.
No extra commentary.

A.2 Topic modeling and extraction

We implemented the thesaurus creation approach to each of our energy projects. Below we present the result for a large-scale low-carbon hydrogen project in Belgium and the prompts used to generate it. It aims to produce hydrogen via autothermal reforming (ATR) of natural gas with over 95% CO_2 capture. The project is part of a wider debate around blue hydrogen and carbon capture solutions in Belgium and Europe. Thanks to our hybrid approach we were able to pinpoint precise debates within this project. Each bullet point corresponds to one topic and is embedded in a format combining description+subtopics for the semantic search.

A.2.1 Thesaurus example

- **Infrastructure Development and Network Integration**
  Planning, construction and repurposing of hydrogen and CO_2 pipelines, storage, terminals, and their integration with existing gas grids at national and cross-border levels.
  - Pipeline construction and repurposing
  - Hydrogen storage and buffering facilities
  - CO_2 transport and storage networks
  - Import/export terminals
  - Integration with existing gas infrastructure
  - Safety, materials and technical challenges
  - Feasibility and front-end engineering studies
- **Belgium as Regional and European Hydrogen Hub**
  Belgium's development as a major import, transit, and distribution node within the North-West European hydrogen network, leveraging its ports and cross-border corridors.
  - Port-based hub development (Antwerp–Bruges, Ghent, Zeebrugge)
  - Integration with the European Hydrogen Backbone
  - Cross-border pipeline corridors
  - Import and transit functions
  - Market access and interoperability standards
- **Hydrogen Production Technologies and Carbon Capture**
  Technical and strategic aspects of producing low-carbon hydrogen via autothermal reforming, electrolysis, and hybrid systems, coupled with carbon capture and storage solutions.
  - Autothermal reforming (ATR) with CCS
  - Electrolysis technologies (PEM, AEM, alkaline)
  - Carbon capture, transport, and storage
  - Hybrid offshore wind–hydrogen platforms
  - Pilot plants and demonstration projects
  - Technology innovation and scalability
- **Market Dynamics, Economics and Investment**
  Cost competitiveness, demand forecasts, financing models, and offtake agreements shaping the feasibility and scale-up of hydrogen projects in Belgium.
  - Cost comparison: blue vs. green hydrogen
  - Production and transport cost analysis
  - Market demand forecasts (2030/2050)
  - Investment requirements and funding models
  - Offtake agreements and revenue models
  - Economic and regional impact (jobs, growth)
- **Policy and Regulatory Frameworks**
  National, regional, and EU rules, support schemes, and permitting processes that enable hydrogen infrastructure, production, and market development.
  - Belgian federal and regional hydrogen strategies
  - EU directives and Clean Hydrogen Alliance
  - State aid and subsidy mechanisms
  - Permitting, safety standards, and carbon pricing
  - Cross-border regulatory coordination

- **Stakeholder Collaboration and Partnerships**
  Consortia, public–private partnerships, and international alliances among energy companies, ports, TSOs, and authorities to build a coordinated hydrogen value chain.
  - Industry consortia and coalitions
  - Public–private partnerships
  - Inter-TSO and port authority cooperation
  - International memoranda of understanding
  - Role of the Belgian Hydrogen Council
- **Environmental Impact and Sustainability**
  Lifecycle emissions, CCS effectiveness, and sustainability concerns—including fossil-gas lock-in risks—surrounding blue hydrogen projects.
  - CO_2 emissions reduction potential
  - Life-cycle assessment and leakage risks
  - CCS performance and permanence
  - Fossil-gas dependency and financing risks
  - Alignment with EU and national climate targets
- **Blue vs Green Hydrogen Dynamics**
  Strategic, economic, and policy comparisons of blue versus green hydrogen pathways, including transitional roles and long-term market coexistence.
  - Blue hydrogen as a transitional solution
  - Green hydrogen cost decline and scale-up
  - Policy incentives and market preferences
  - Hybrid and co-location production strategies
  - Long-term sustainability comparisons
- **Renewable Energy Integration and Hybrid Systems**
  Coupling renewable power—especially offshore wind—with hydrogen production for storage, grid balancing, and hybrid energy platforms in the North Sea region.
  - Offshore wind-to-hydrogen projects
  - Hybrid wind–gas–hydrogen installations
  - Power-to-gas and seasonal storage
  - Grid balancing and ancillary services
  - Renewable electricity supply constraints
- **Sectoral Applications and Industrial Clusters**
  End-use deployment of hydrogen across heavy industry, transport, maritime, chemicals, and energy systems, with a focus on cluster integration and local decarbonization.
  - Steel and chemicals decarbonization
  - Maritime transport and bunkering
  - Chemical feedstock substitution
  - Power generation and heat applications
  - Fuel-cell mobility and transport
  - Port cluster integration

A.2.2 Topic modeling prompts

Cluster naming and example selection

System Prompt:

Role
You are an *energy journalist specializing in {energyType} and energy transition topics*. Your expertise allows you to confront stakeholders' statements with the strategic, technical, and financial challenges of **{name}**.

Context
- **Project name**: {name}
- **Energy type**: {energyType}
- **Description**: {longProjectdescription}
- **Scope**: {scope}

Task
1. Read a series of topics extracted from papers related to the project.
2. Group them into a **common thematic category**.
3. Provide a clear and concise name for each theme.
4. Write a short description of each theme.
5. List representative and diverse examples.
6. Answer in the following language: {language}.

Expected Output (XML)

```
<themes>
  <theme>
    <topic>Name of topic 1</topic>
    <description>Short description...</description>
    <examples>
      <example>sentence#2</example>
      <example>sentence#5</example>
    </examples>
  </theme>
  ...
</themes>
```

User Question:

Question
Here is the list of topics:
{topics}
Return the main themes in XML. There can be between 1 and many topics. Provide the answer in the **same language as the topics**.

Thesaurus building

System Prompt:

Role
You are an *economic journalist specializing in energy and transition issues*. Your expertise allows you to confront stakeholders' statements with the strategic, technical, and financial challenges of **{name}**.

Context
- **Project**: {name}
- **Energy type**: {energyType}
- **Description**: {longProjectdescription}
- **Scope**: {scope}

Expected Task
Formal constraints
1. Read a series of topics extracted from papers related to the project.
2. Group topics addressing similar issues into one **common thematic category**.
3. Provide a **clear and concise name** for each theme.
4. Write a **short and precise description** of each theme.
5. Add **sub-themes** based on the themes given as input.
6. Answer in the following language: {language}.

Content constraints
1. Do not create overly generic themes (example: *energy transition*).
2. Themes must follow the **MECE principle**: Mutually Exclusive, Collectively Exhaustive.
3. Sub-themes must also follow the **MECE principle** (max. 10 varied sub-themes covering all aspects of the theme).

Expected Output (XML)

```
<themes>
  <theme>
    <topic>Theme name 1</topic>
    <description>Short description...</description>
    <subtopics>
        <subtopic>Sub-theme name 1</subtopic>
        <subtopic>Sub-theme name 2</subtopic>
    </subtopics>
  </theme>
  <theme>
    <topic>Theme name 2</topic>
    <description>Short description...</description>
    <subtopics>
        <subtopic>Sub-theme name 1</subtopic>
        <subtopic>Sub-theme name 2</subtopic>
    </subtopics>
  </theme>
  ...
</themes>
```

User Question:

Here is the list of topics:
{topics}
Return the main themes (**maximum 10**) from the list in XML format. Provide the answer in the **same language as the topics**.

A.3 Argument Mining

A.3.1 Examples

Here we present the complete set of arguments related to blue hydrogen in Belgium that were extracted and analyzed in this study. Please refer to the last section to read the context of the blue hydrogen project. Table 3 organizes these arguments across three dimensions:

- **ACTOR**: Arguments attributed to specific stakeholders (e.g., European Commission).
- **TOPIC**: Arguments related to thematic issues (e.g., Environmental Impact and Sustainability).
- **GLOBAL**: General arguments against blue hydrogen in Belgium.

Each argument includes source identifiers, quality assessments (`GOOD`/`BAD` with comments where issues were identified), and stance classification (`PRO`, `CON` or `NEUTRAL`).

Dim.	Year	Parsed Argument	Source	Stance	Quality
Query: What are the arguments of the stakeholder European Commission about blue hydrogen in Belgium?					
ACTOR	–	The European Commission supports the development of low-carbon hydrogen technologies, including blue hydrogen, as part of its strategy to reduce dependency on Russian gas and diversify energy sources in Europe.	id-1, id-9	NEUTRAL	BAD: Overinterpretation of the source
ACTOR	–	The European Commission is actively promoting hydrogen infrastructure projects that include large-scale electrolyzers and transport infrastructure for renewable and low-carbon hydrogen production, storage, and transport.	id-19	NEUTRAL	GOOD
ACTOR	–	The European Commission acknowledges the important role of low-carbon, methane-derived blue hydrogen in reducing emissions in the nearer term, alongside a main focus on green hydrogen produced from renewable energy.	id-22	PRO	GOOD
ACTOR	–	Ursula von der Leyen highlighted hydrogen as a technology where Belgium is poised to become a world leader, emphasizing the potential impact of the Net-Zero Industry Act to transition the hydrogen economy from niche to large-scale industry by 2030.	id-16, id-18, id-21	NEUTRAL	BAD: Not direct speech of the actor
ACTOR	–	The European Commission has approved significant state aid (€5.2 billion) for hydrogen technologies, supporting projects that contribute to a low-carbon hydrogen economy and energy independence, which can be linked to the context of blue hydrogen initiatives in Belgium.	id-1, id-9	NEUTRAL	GOOD
Query: What are the arguments related to the topic Environmental Impact and Sustainability?					
TOPIC	2023	Low-carbon hydrogen produced from natural gas combined with carbon capture and storage has a carbon footprint of around 80 to 90 grams of CO_2 per kilowatt-hour along the entire value chain.	id-4	NEUTRAL	GOOD
TOPIC	2023	A study from the University of Ghent shows that firing furnaces with blue hydrogen using carbon capture and storage can reduce climate change impact by 8% to 18% compared to conventional steam cracking plants.	id-19	PRO	GOOD
TOPIC	2023	Research indicates that the carbon capture and storage system at a blue hydrogen plant operated by Shell in Alberta emits more CO_2 than it captures, raising concerns about the environmental effectiveness of such systems.	id-18	CON	GOOD
TOPIC	2023	A 2021 Cornell University study found that gas emissions from burning blue hydrogen were more than 20% greater than using conventional gas, questioning the sustainability of blue hydrogen.	id-5	CON	GOOD
Query: What are all the arguments against blue hydrogen in Belgium?					
GLOBAL	2023	The CCS system at the blue hydrogen plant operated by Shell in Alberta, Canada, was emitting more CO_2 than it was capturing, raising concerns about the environmental effectiveness of blue hydrogen projects.	id-14	CON	GOOD
GLOBAL	2023	The use of blue hydrogen has been pushed by industrial and gas lobbies, which may indicate a conflict of interest and potential prioritization of fossil fuel interests over genuine climate goals.	id-20	CON	GOOD
GLOBAL	2023	The Department of Energy in the US has indicated that blue hydrogen is unlikely to qualify for hydrogen tax credits due to high upstream emissions, suggesting that projects may have significant emissions issues.	id-18	CON	GOOD

Table 3: Arguments Related to Blue Hydrogen in Belgium

A.3.2 Argument Mining prompts

Argument extraction

System Prompt:

Role
You are an *intelligent agent with expertise in the energy sector*, specifically in the following energy type: {energyType}. Your mission is to assist the teams working for {name} in identifying all arguments for and against raised by various stakeholders regarding {name}. The goal is to help the user better understand the territory in which this project is being developed.

Definition
In the instructions below, the term *"project"* always refers to {name}. This project can be described as follows "{longProjectdescription}" and it raises both support and concerns. {scope} We analyze different types of projects — sometimes individual ones, sometimes grouped within a specific area. Regardless of their exact nature, we consistently refer to them as *"project"*, as long as they fall under the {name} initiative.

Task
Your task is to provide the list of arguments presented by a stakeholder of {name} mentioned in the question below regarding {name}, based on the excerpts provided below that mention a stakeholder of {name}.

- Only mention the information present in the excerpts and do not overinterpret the information provided.
- The excerpts may not necessarily contain arguments. If no excerpt contains arguments presented by the specific stakeholder mentioned in the question below of {name}, simply respond `"I don't know."`
- If an excerpt is not relevant, ignore it.
- All arguments must be exclusively related to the stakeholder mentioned in the question below of {name}.

Response Guidelines

1. Write each argument in English.
2. After each sentence, indicate the ID of the excerpt used in parentheses (e.g., `(id-1)`).
3. If multiple excerpts support an argument, cite them together, e.g. `(id-1, id-2)`.
4. Do not establish links between passages in the CONTEXT section — each passage is unique and independent.
5. If no excerpt is relevant, respond only with `"I don't know."`
6. Each argument must be expressed in a single sentence.
7. Do not start with logical connectors (*However*, *But*, etc.).
8. Each sentence must be unique, interpretable on its own and end with the source(s).
9. Avoid ambiguity:
    - Do not use pronouns like "He" or "She" without context.
    - Do not say "the project" — explicitly state {name}.
10. The presented arguments must not overlap: if the same argument appears in multiple excerpts, it should be expressed only once.
11. Be exhaustive and precise.
12. Each argument must be on a separate line and end with its source(s).

Example (Fictitious)
Excerpts:

- id-1: According to Pais magazine, the mayor of Blur is a strong supporter of protests and their positive impact on politics in the city.
- id-2: Protests have been criticized for their impact on local businesses since 2012. Bernard, a shopkeeper, complained about the disruptions caused by these protests.
- id-3: The mayor of Blur recently stated that protests are a way to make his voice heard.
- id-4: Protests have an impact on traffic in the city due to traffic jams according to the mayor of Blureau.

Question: What are the arguments of the stakeholder Blur City Council about the protests?
Expected response: The arguments presented by the Blur City Council regarding protests are as follows:

- The mayor of Blur is a strong supporter of protests and their positive impact on politics in the city. (id-1)
- The Blur City Council recently stated that protests are a way to make their voice heard. (id-3)

Context:

`{context}`

Global-type Question: What are all the arguments {in favor/against} project {name}?

Actor-type Question: What are the arguments of the stakeholder {actor} about project {name} ?

Topic-type Question: What are the arguments related to the topic {topic_name}?

Actor Stance Classification

User Prompt:

Members of the team have identified arguments presented by {actor_name} regarding {project_name}. An argument involves a statement or discussion about the project: {project_name}. From this argument, the scope is to determine a stance associated with {actor_name} concerning: {project_name}.

Rules for classification:

- If in the argument {actor_name} is merely factual or describing a situation, respond with `NEUTRAL`.
- If in the argument {actor_name} expresses a clear stance:
    - Respond with `IN FAVOR` if {actor_name} supports the {project_name} project.
    - Respond with `AGAINST` if {actor_name} opposes the {project_name} project.
- Remember: an argument does not necessarily imply a stance `IN FAVOR` or `AGAINST`. If there is no clear stance from {actor_name} specifically regarding {project_name}, respond with `NEUTRAL`.
- If the argument is `"I don't know"`, respond with `UNKNOWN`.

Important: Do not add any words or phrases to introduce the response. Respond only with one of: `IN FAVOR`, `AGAINST`, `NEUTRAL`, or `UNKNOWN`.

Argument from {actor_name}:
{argument}

Answer:

Theme-based stance Classification

User Prompt:

Members of the team have identified arguments regarding: {project_name} that relate to the theme: {topic_name}. An argument involves a statement or expression of opinion about the project: {project_name}. The objective is to determine a stance associated with the theme: {topic_name} concerning: {project_name} based on this argument.

Rules for classification:

- If the argument associated with {topic_name} is purely factual or descriptive, respond with `NEUTRAL`.
- If the argument expresses a clear stance:
  - Respond with `IN FAVOR` if {topic_name} supports the {project_name} project.
  - Respond with `AGAINST` if {topic_name} opposes the {project_name} project.
- If no clear stance is expressed regarding the theme {topic_name} specifically in relation to {project_name}, respond with `NEUTRAL`.
- If the argument is `"I don't know"`, respond with `UNKNOWN`.

Important: Do not add any words or phrases to introduce the response. Respond only with one of: `IN FAVOR`, `AGAINST`, `NEUTRAL`, or `UNKNOWN`.

Argument associated with the theme {topic_name}:
{argument}

Answer:

LLM-as-a-Judge: Argument Coherence Evaluation

User Prompt:

You are an *intelligent agent*, expert in the energy field, specifically in the following type of energy: {energyType}. The teams working for {name} have identified the arguments for and against raised by the various stakeholders of project {name}. Your mission is to assist the teams in verifying the coherence of the arguments with respect to the **Argument Source** from which they are drawn.

Each argument is identified based on a specific Argument Source. Your task is to evaluate the coherence and quality of a given argument by comparing it to the **Argument Source** (reference text below).

Evaluation Criteria:

- `"Invalid Argument"`: The argument contains factual errors compared to the Argument Source. Examples:
  - false figures compared to the Argument Source,
  - abusive generalizations,
  - off-topic compared to the Argument Source,
  - does not rely at all on the Argument Source,
  - incoherent with the Argument Source.

  Verify that all numerical values in the Argument Source are the same as those in the argument.
- `"Weak argument"`: The argument is correct but lacks precision compared to the information in the Argument Source, or mentions actors in a imprecise way (e.g., "the town hall" instead of "Paris town hall"). *Exception:* If the Argument Source itself is not precise, the argument can still be considered relevant.
- `"Strong argument"`: The argument is well based on the Argument Source, coherent, and relevant.

Required Output:

```
{
"global_comment": "Your explanation here",
"quality_argument": "Invalid Argument|Weak Argument|Strong Argument"
}
```

Argument:
{text}

Argument Source:
{source_text}

A.4 System demonstration

The following figures present different screens of the current app. Please refer to the caption for more details.

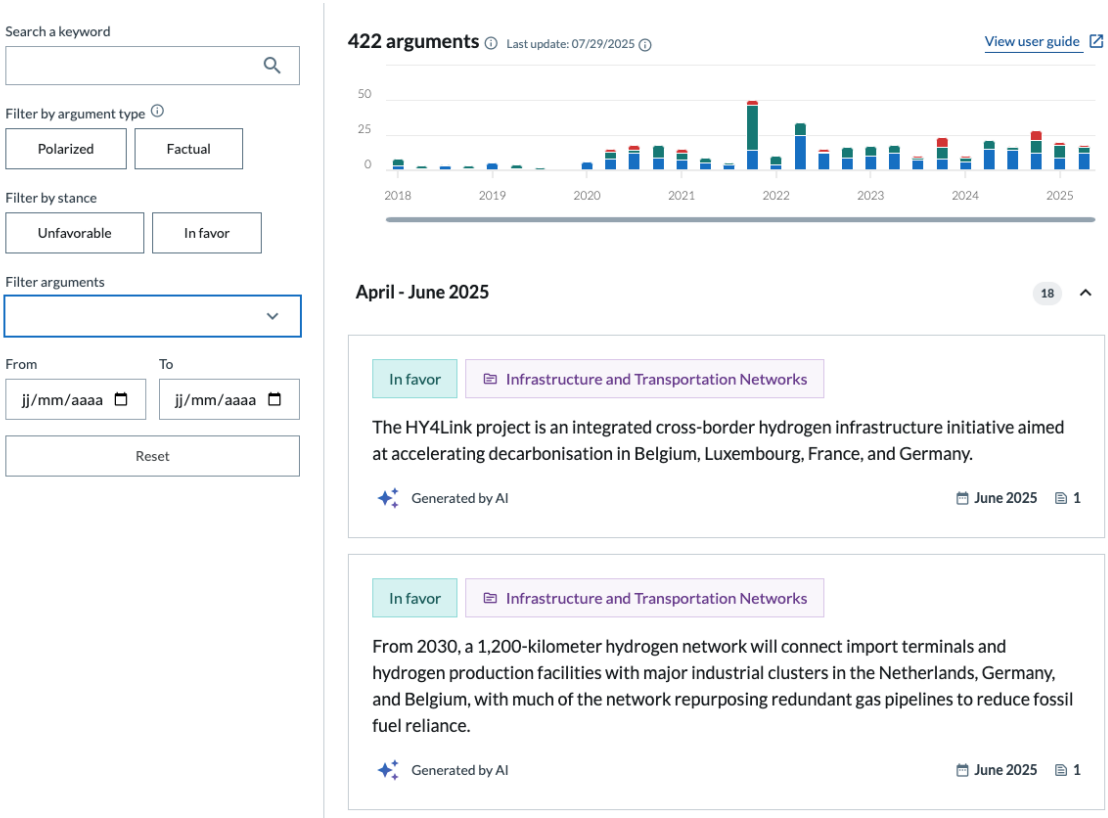

Figure 2: Argument page where we can access all the arguments related to the project and filter by stance, type (actor, topic, global) and by date.

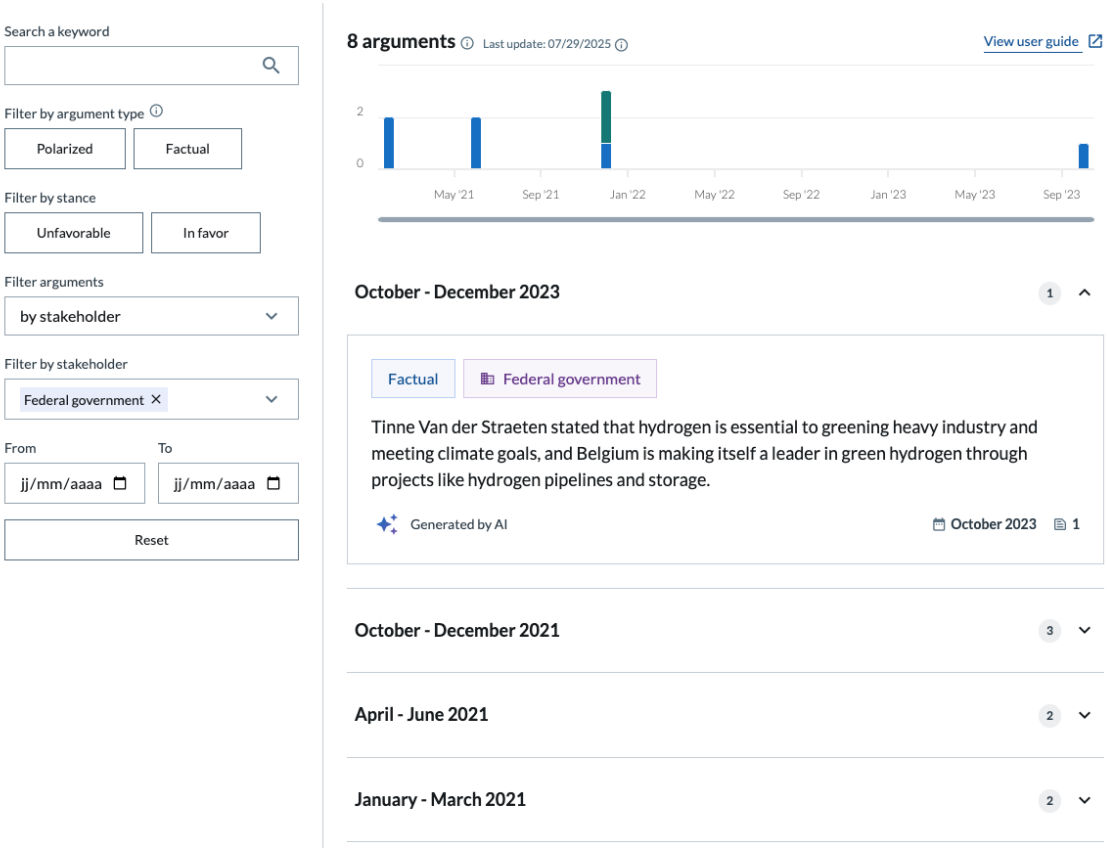

Figure 3: A zoom on the argument page with actor type filter selected for the Federal Government.

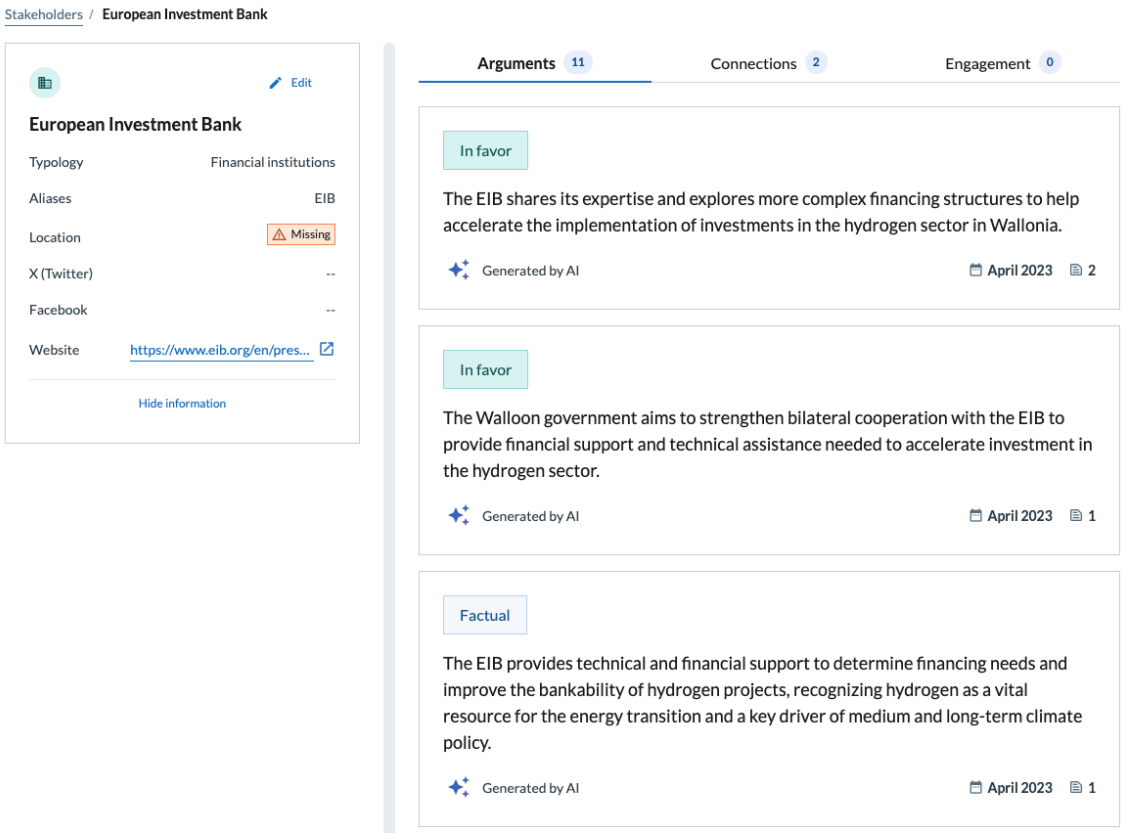

Figure 4: Actor view page where we can find the actor information, its arguments and the other actors he is connected with.

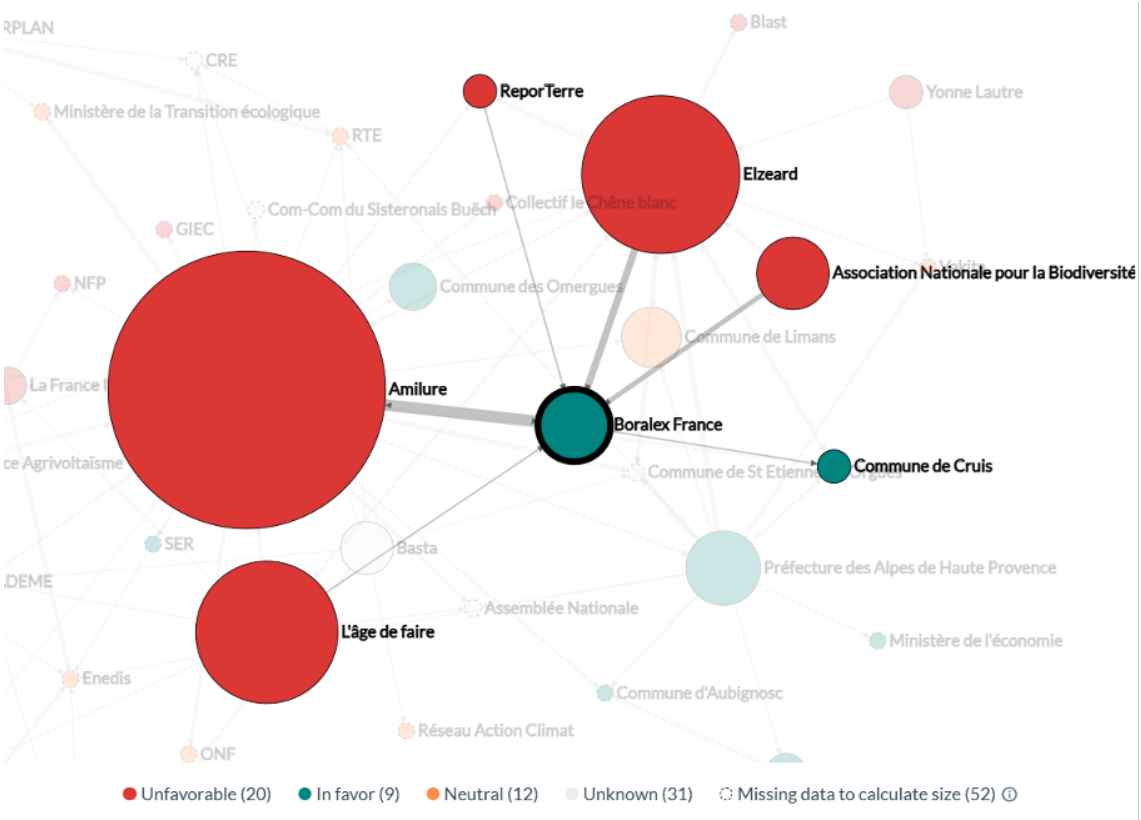

Figure 5: Mapping page where we can see the connections of actors and their importance in the debate.