TruthSplit: Interactive Generation of Multi-Perspective Reasoning Graphs for Contested Claims

Lion Six* **and Adrien Zaradez*** **and Thomas Huber and Christina Niklaus**
University of St. Gallen
{lion.six, adrien.zaradez}@student.unisg.ch
{thomas.huber, christina.niklaus}@unisg.ch

Abstract

Disagreement in contested domains often reflects different assumptions and reasoning norms rather than simple factual errors. We present *TruthSplit*, a system that generates unified multi-perspective reasoning graphs for contested claims. The resulting TruthPath representation encodes shared premises and assumptions, typed divergence points, perspective-specific reasoning paths, and final stances, and combines an interactive visualization with a textual meta-analysis. An expert evaluation of generated TruthPaths provides initial evidence of analytical quality, while a user study with 22 participants indicates that the interface supports identifying shared ground and sources of disagreement. TruthSplit provides an inspectable representation of how different assumptions and values can lead to different conclusions without adjudicating which perspective is correct.

1 Introduction

Why do people reach very different conclusions when considering the same facts? In political and ethical debates, disagreement often arises from differences in definitions, background assumptions, and value commitments that shape what is seen as relevant or plausible (Gallie, 1955). Making these differences explicit is important for NLP systems deployed in socially consequential domains, where interpretability and transparent reasoning are necessary for responsible use.

We study a claim-centric generation setting for multi-perspective reasoning: given a contested claim and a set of perspective profiles, generate a typed reasoning graph, a textual description of common ground, an explanation of *where* and *why* divergence occurs, perspective-conditioned reasoning paths, final stances and a cross-perspective synthesis of the chosen perspectives. We implement this as a prototype called *TruthSplit*[1]. The system prompts a large language model (LLM) with author-curated worldview profiles, validates the structured output against a schema, and renders the resulting graph for inspection (see Figure 1). Each node provides a textual description of common ground or divergence (see right side of Figure 1 for an example). Our goal is to represent disagreement as structured divergence rather than as binary opposition. We ask: *How can an NLP system make the underlying assumptions and perspective differences behind a contested claim explicit without adjudicating correctness?*

This work follows recent perspectivist NLP research that treats disagreement as potentially meaningful variation rather than noise, including work on perspectivist argument-quality assessment (Romberg et al., 2025) and worldview-aware natural language inference (Creangă and Dinu, 2024). This demo builds on the previously published TruthSplit system for conditional-validity analysis of argumentative texts (Stieger et al., 2026). Whereas the earlier system decomposes an input argument and applies a multi-stage NLI and worldview-conditioned analysis pipeline, the present system starts from a contested claim and directly generates a unified TruthPath graph containing shared ground, explicit divergence points, and perspective-specific reasoning branches. It additionally expands the perspective inventory to 18 philosophical positions and evaluates the generated reasoning structures and interaction design. The 18 worldview profiles are author-curated, simplified analytical perspectives specified through short descriptions, core values, and philosophical principles. They are intended to capture recurring reasoning patterns rather than

*Equal contribution.

[1]A live demo is available at https://truthsplit.com. The code is licensed under the MIT license and available at https://github.com/Adrien-Z/TruthSplit. A video demonstration can be found at https://youtu.be/yykuZO6hKis.

Proceedings of the 19th International Natural Language Generation Conference: System Demonstrations, pages 11–13
October 17-21, 2026 ©2026 Association for Computational Linguistics

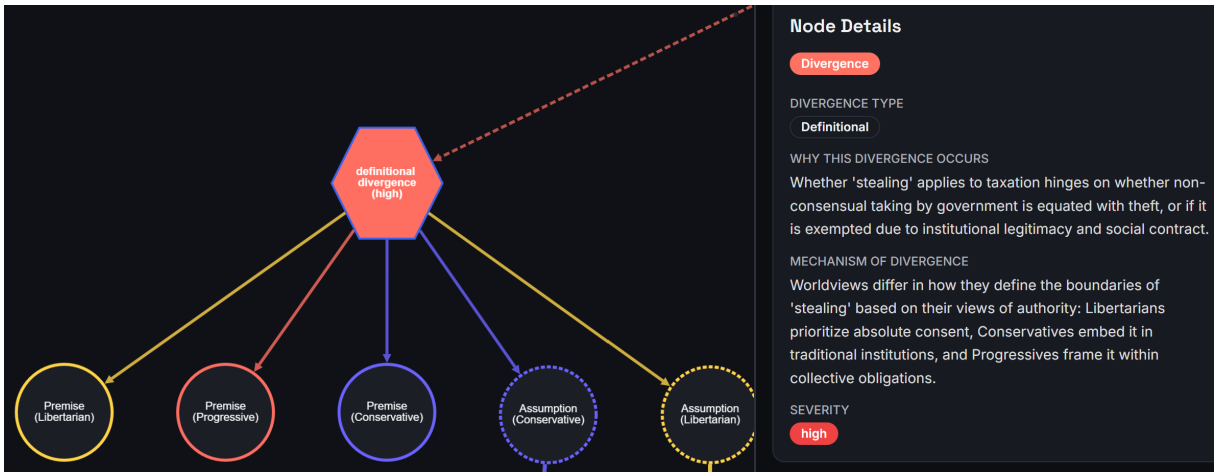


Figure 1: TruthSplit's unified TruthPath visualization. Shared ground precedes a typed divergence point, after which generated perspective-specific reasoning paths branch (left). Selecting a node reveals its textual explanation (right).

exhaustively characterize individual adherents.

2 Interface and User Interaction

User Flow The user begins by entering a claim, or choosing from a curated list of examples, and selecting worldviews from a list of 18 philosophical positions. After submission, the pipeline produces a structured representation of the reasoning paths associated with each selected worldview. The results page then invites users to explore this representation interactively.

Key Interactions The central component of TruthSplit is the TruthPath visualization: a hierarchical graph built with Cytoscape.js (Franz et al., 2015) (see the left side of Figure 1). The claim is shown at the top, followed by shared premises and assumptions that are common to the selected worldviews. The graph then splits into the divergence points—i.e., parts where worldviews split due to definitional, normative, empirical, or epistemological differences—with separate branches for each worldview containing unique premises, assumptions, and conclusions. Nodes are color-coded by type and worldview, and edges are styled to indicate shared ground or divergence.

Users can zoom, recenter, and search within the graph; a mini-map provides spatial context; keyboard shortcuts allow for efficient navigation; and a legend explains the meaning of node shapes and edge styles. Selecting a node opens a details pane describing its content, associated worldviews, and reasoning (shown in Figure 1 on the right side). The system additionally generates textual descriptions of shared ground, divergence points, and meta-analysis results, as well as cards listing worldview-specific premises, assumptions, and final stances. Figure 1 shows an excerpt of a full TruthPath visualization. The divergence analysis and meta-analysis appear below the graph.

Feedback, Transparency and Safeguards User claims undergo safety moderation before analysis. The generation prompt additionally frames worldview profiles as simplified analytical constructs, requires respectful and non-discriminatory language, prohibits stereotypical or reductive portrayals, and asks the model to represent each perspective charitably. The interface explains the generation process and warns users that analyses may nevertheless contain bias or hallucinations.

Tooltips and hover cards define technical terms (such as assumption types and divergence severity) when users interact with the graph. The meta-analysis summarizes model-estimated common ground, the primary source of disagreement, and possible bridge concepts, offering an immediate overview of why the selected worldviews disagree.

3 Implementation Details

TruthSplit follows an *input* → *moderation* → *multi-perspective generation* → *validation* → *storage* → *visualization* pipeline. The user-provided claim and selected worldview profiles are combined in a structured analysis prompt. The LLM jointly generates a JSON representation containing shared premises and assumptions, typed divergence points with explanatory mechanisms, worldview-specific premises and assumptions, final stances, and a cross-perspective meta-analysis. Outputs are val-

idated against a predefined schema. Invalid or incomplete analyses are rejected rather than visualized. Valid analyses are stored using Supabase and rendered as hierarchical Cytoscape.js graphs in the React frontend.

4 Evaluation

We evaluated two LLM configurations: OpenAI GPT-4.1[2] and xAI Grok 4 Fast Reasoning (`grok-4-fast-reasoning`) using provider-default sampling parameters. Generation used structured JSON-schema output. We conducted two complementary evaluations of TruthSplit.

Analytical quality Two domain experts with academic backgrounds in philosophy and argumentation independently assessed 40 generated TruthPaths using the TruthPath Analytical Quality Score (TAQS). Each output was rated on assumption transparency, definition quality, logical coherence, worldview-divergence mapping, factual and normative accuracy, and overall analytical quality. Overall TAQS scores were 3.85/5 for GPT-4.1 and 3.83/5 for Grok-4-FR, with logical coherence and factual/normative accuracy among the highest-rated dimensions. Experts nevertheless identified occasional ambiguity between premises and assumptions and highlighted the risk of false equivalence between unsupported and evidence-based perspectives.

User evaluation 22 participants from a university community interacted with our system and completed 20 Likert-style statements and open questions. Participants rated usability and navigation at 4.05/5 and the system's support for identifying divergence and shared ground at 3.98/5 (see Table 1).

Category	k	Mean (SD)	Median	Range
Motivation & Perspective	3	3.64 (0.96)	3.83	1.67–5.00
Identifying Divergence & Shared Ground	5	3.98 (0.79)	4.10	2.00–5.00
Engagement & Cognitive Load	5	3.74 (0.72)	3.70	2.60–5.00
Usability & Navigation	4	4.05 (0.87)	4.13	1.50–5.00
Willingness & Credibility	3	3.97 (0.84)	4.00	1.67–5.00

Table 1: Composite category scores (Likert 1–5) across participants. k denotes the number of items included in each composite category. Higher is better ↑.

[2]The implementation supports GPT-4.1 either through the Vercel AI Gateway (`openai/gpt-4.1`) or as the pinned OpenAI snapshot `gpt-4.1-2025-04-14`.

5 Conclusion

TruthSplit makes disagreement inspectable through graphical and textual representations of shared ground, divergence points, and perspective-specific reasoning. Our evaluations provide initial evidence of analytical quality and usability, while highlighting challenges around worldview simplification, false equivalence, and epistemic signaling.

6 Limitations

TruthSplit relies on LLM-generated interpretations of simplified, manually curated worldview profiles. Structural schema validation does not guarantee factual correctness or faithful representation, and the visually symmetric presentation may obscure differences in evidential support. The two-expert evaluation provides initial evidence of analytical quality, but does not constitute a comprehensive validation of the worldview profiles.

References

Claudiu Creangă and Liviu P. Dinu. 2024. Designing NLP systems that adapt to diverse worldviews. In *Proceedings of the 3rd Workshop on Perspectivist Approaches to NLP (NLPerspectives) @ LREC-COLING 2024*, pages 95–99, Torino, Italia. ELRA and ICCL.

Max Franz, Christian T. Lopes, Gerardo Huck, Yue Dong, Onur Sumer, and Gary D. Bader. 2015. Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics*, 32(2):309–311.

W. B. Gallie. 1955. Essentially contested concepts. *Proceedings of the Aristotelian Society*, 56:167–198.

Julia Romberg, Maximilian Maurer, Henning Wachsmuth, and Gabriella Lapesa. 2025. Towards a perspectivist turn in argument quality assessment. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 7458–7485, Albuquerque, New Mexico. Association for Computational Linguistics.

Benjamin Stieger, Maximilian Terberger, Thomas Huber, and Christina Niklaus. 2026. TruthSplit: Operationalizing conditional validity in arguments through multi-perspective reasoning. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, pages 648–659, San Diego, California, United States. Association for Computational Linguistics.