Ankita Gupta
ankitagupta@cs.umass.edu

A wide range of social settings, such as public policy, law, and health, require *intelligent text processing at scale*. For instance, democratic processes rely on collective decision-making, which further depends on *institutions* that can incorporate citizens' feedback into lawmaking, and on *citizens* who can access information about their legal and social situations. Health assessments require clinicians to understand patient narratives and transcripts. However, it is impossible to manually read all the relevant documents, as they can span thousands of public comments, clinical transcripts, and millions of judicial opinions.

Automated text analysis systems can assist in such diverse social information processing needs by discovering patterns from large text corpora at scale. However, making these systems work in high-stakes settings is challenging, as real-world data is messy and requires developing careful methods. I develop **natural language processing** (NLP), **machine learning** (ML), and **computational social science** (CSS) methods, advancing them along three complementary dimensions:

1. **Methods to understand rich semantic structures in text**: Social text data, such as public feedback, legal documents often richly intertwine **arguments** (opinions with supporting reasons) through personal **narratives**. How do we extract such structures from texts to enable more accurate insight into social data? To this end, I extract *signed graphs* among text propositions, as semi-structured semantic text representations from a collection of unstructured texts.

2. **Methods for trustworthy and reliable information processing at scale**: These text corpora can include millions of documents and are useful to inform high-stakes decision-making. How do we process these texts and compute quantities of interest at scale with approaches that generalize across multiple problems? I employ large language models (LLMs) for text processing, which have recently demonstrated strong performance across many NLP tasks. While using them, I also address their technical challenges such as *reliability*, *efficiency*, and *transparency*.

3. **Human-centered interdisciplinary method design**: Effective methods must translate theoretical assumptions from a discipline into computational models and be designed around end-user needs and workflows. How do we develop methods that are **informed by domain expertise** and address important domain questions? I draw insights from and collaborate with researchers in linguistics, public policy, law, and health to guide the design of my methods.

These are recurring themes across many projects, including methods for: a) helping institutions understand public feedback (§1), and b) assisting legal professionals in research to expand citizens' access to legal information (§2), and c) sociocultural discourse analyses (§3). These methods also apply broadly to other areas (§4), such as clinical assessment of patient communication abilities and curating annotations for LLM supervision. In future work (§5), I look forward to continuing to contribute methodological advances in computer science while addressing substantive domain questions.

Broader Impacts: My research has been published in major NLP and CSS venues (e.g., ACL'24b; ACL'25a; ACL'25b; EACL'23; LREC-COLING'24; IC^{2}S^2'24) and interdisciplinary journals (JLC'25, AJSLP'25). I received the 2023 IBM PhD Fellowship for my research on improving the robustness of LLMs. My work on public comments summarization was selected for a plenary talk at IC^{2}S^2'24 (2.8% of submissions). Beyond academic venues, my research with colleagues on studying US Supreme Court oral argument practices (JLC'25) gained attention from press outlets covering legal and judicial matters, including Axios, Balls and Strikes, and Strict Scrutiny, and my work on legal argument stances (ACL'25b) is being used by Thomson Reuters to evaluate legal AI systems.

1 Methods for understanding argumentative public feedback in rulemaking

Comprehending large volumes of public feedback is crucial for civic decision-making. For instance, U.S. federal agencies must solicit and respond to public comments for rulemaking under the Administrative Procedures Act, 1946. Prior automated content analysis methods (e.g., topic modeling, sentiment analysis) can help uncover the main beliefs held by the public. However, public comments are often argumentative, where commenters not only state beliefs but also provide supporting reasons.

I advance methods for a deeper analysis of such public feedback to help uncover the public beliefs and arguments at scale. I developed a *zero-shot LLM-based method* to understand unstructured argumentative comments (**Gupta** et al., 2024b). It involves prompting LLMs with name references to Toulmin (1958), a pedagogically popular argumentation theory, to decompose an argument into its structure and implicit reasoning. I then applied this method to visualize arguments across $10K$ comments on COVID-19 vaccine approval for children. The resulting corpus-level argument hypergraph (Figure 1), shows arguments recurring across multiple comments, which are also consistent with prior studies of

Ankita Gupta
ankitagupta@cs.umass.edu

social media discussions (Wawrzuta et al., 2021). This work demonstrates how generative AI can aid interpretive work for policymakers and, more broadly, in computational social science.

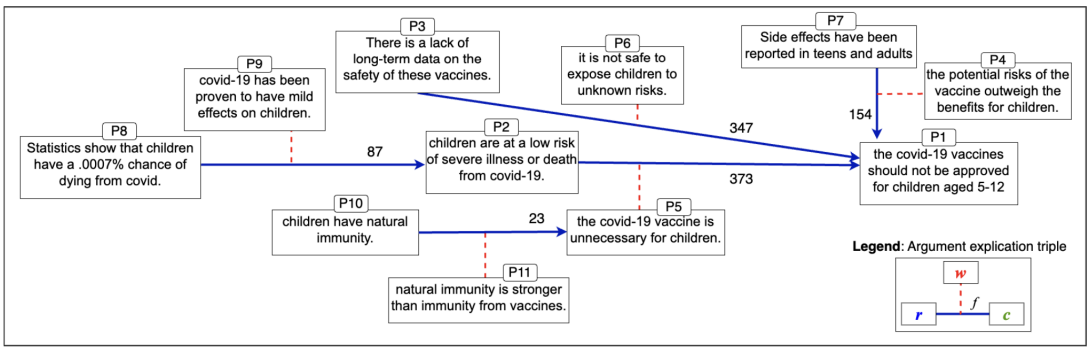

Figure 1: A portion of the corpus-level argument hypergraph we automatically extract from public comments submitted on regulations.gov on whether to approve a COVID-19 vaccine for children. Each node is a cluster of propositions extracted from comments. An argument is a triple of nodes, ⟨(c)laim, (r)eason, (w)arrant⟩, visualized as solid blue and dotted red arrows connecting the reason and warrant (r, w) to the claim (c). f is the triple's corpus frequency.

2 Dataset and method for understanding legal argumentation

Law is an important area with significant information overload, where conducting research demands significant time and expertise, as professionals must sift through vast amounts of supporting and opposing evidence to construct well-reasoned arguments. I develop computational tools to make legal research more tractable for professionals, with the broader goal of expanding citizens' access to legal information.

In one project, I developed a legal stance classifier (**Gupta** et al., 2025b) to determine whether a retrieved *case summary* strengthens or weakens a *legal argument* (polarity) and to what extent (intensity). For this task, LLMs can be computationally expensive at scale, and training smaller models requires substantial manual annotation effort. Instead, I used legal writing conventions, as prescribed in the *Bluebook* citation guide, to semi-automatically extract stance annotations at scale. The resulting dataset, δ-`Stance`, contains over 2M+ naturally occurring stance triples annotated by the US judges and grounded in real-world historical legal practice (Figure 2). Using this dataset, I trained *small and cost-efficient models* that outperform larger models on this task.

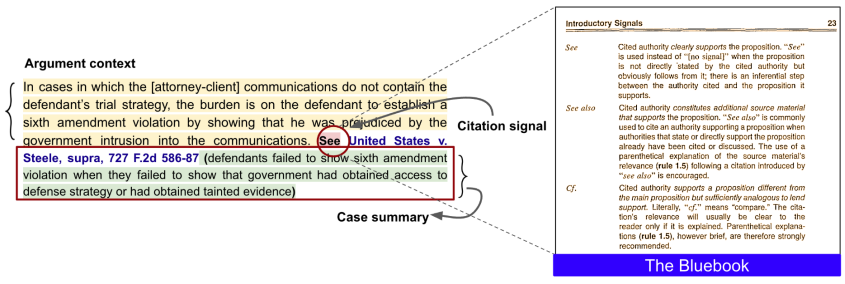

Figure 2: A judicial opinion paragraph, containing an argument context, a cited case with a citation signal, and a case summary. We map each citation signal to a stance value based on its definition in the Bluebook, yielding stance-annotated triples in our δ-`Stance` dataset.

The general public seeking legal advice could also benefit from my work. One challenge is that laypersons describe their situations in everyday language while legal research requires mapping these narratives to legal terminology. In an ongoing project, I am developing a *reasoning-augmented retrieval tool* to bridge this gap by mapping layperson narratives to relevant legal issues and thus prior cases, using a judicial opinion's structure to curate synthetic training data.

3 Understanding how social factors shape argumentative discourse.

The methods I have developed can also be applied to obtain other semi-structured text representations of how people argue (e.g., who cites whom), which, when aggregated over large text corpora, can enable us to better understand our society.

Ankita Gupta
ankitagupta@cs.umass.edu

In one project, I studied how authors with different political ideologies selectively cite different entities when building arguments (**Gupta** et al., 2022). I proposed *epistemic stance detection* task to extract entities cited by authors to support/oppose a claim. Analyzing 370 political opinion books using a supervised model trained on this task, we found that liberal authors frequently cite technocratic authorities, while conservatives cite founders and express derision for the media.

In another project, colleagues and I developed a *causal inference method* that helps estimate the influence of gender on communication dynamics in Supreme Court oral arguments (Cai et al., 2025). We examined nearly four decades of oral arguments transcript data and found that female advocates are interrupted more frequently by justices than male advocates during oral arguments (Figure 3).

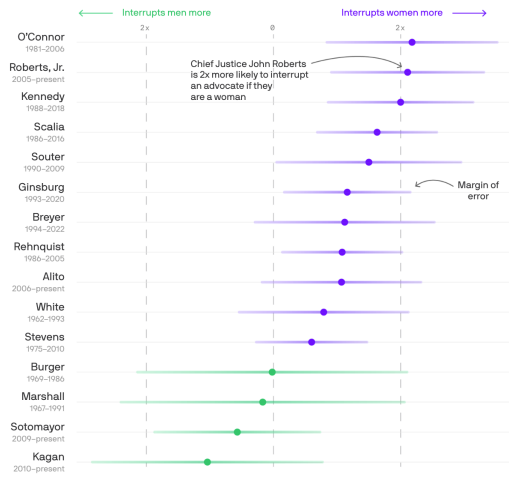

Figure 3: Effect of gender on interruption from Cai et al. (2025), as reproduced in the press coverage by Axios.

4 Cross-cutting methods for broader applications

One of the common underlying themes in my work is the use of semi-structured text representations for understanding discourse in complex real-world settings. This approach applies broadly to other areas, which I illustrate through the following studies.

Narrative summarization for clinical assessment of aphasia. My work on argument summarization (§1) uses LLMs to obtain proposition tuples and rigorously uses this mathematical structure for corpus visualization. This approach is also extremely useful in clinical assessment of aphasia, where patients retell short stories, and clinicians evaluate their communication abilities based on whether they mention *main concepts* (MCs)—an ordered list of propositions capturing the story's essential elements (Kurland et al., 2025). Building on my argument summarization work (§1), I developed a method to automatically generate MCs (**Gupta** et al., 2025a) for new stories, by prompting open-source LLMs with pedagogically popular story retelling strategies (e.g., five-finger retell). I further developed a *prompt ensemble method* to address LLMs' prompt sensitivity issues. These generated *MCs* can be reviewed/revised by clinicians, and serve as a checklist to evaluate patient retells. I am currently extending this work to **procedural texts**, where evaluating *temporal ordering* among MCs is essential. More broadly, this work can also apply to other settings rich with personal narratives, such as citizens' lived experiences in public comments or summarizing litigant narratives.

Annotation analyses. As I have worked on diverse applications, annotated datasets remain crucial for future work. For instance, developing modern LLMs requires millions of annotated examples for accurate supervision (Singh et al., 2024). My work has made advances towards open-source tools for collecting annotations for complex linguistic tasks (**Gupta** et al., 2023; Rogers et al., 2024).

Another challenge is dealing with *annotation disagreements*. Prior work has advocated for preserving them during model training to reflect pluralistic values. However, distinguishing genuine differences from annotation noise or unclear guidelines remains a manual process that is difficult and time-consuming. To address this, in an ongoing work with Microsoft Research, I am developing an *LLM-based* method to generate disagreement explanations in multi-annotator text datasets. I am further exploring sparse autoencoders (SAEs) for this task for more interpretable analyses. Looking forward, I aim to explore other types of annotations used across different LLM training phases (e.g., pairwise preference judgments).

5 Future Work

I have developed computational models to uncover rich semantic structures (e.g., arguments, narratives) from large collections of unstructured texts across diverse social contexts, that end-users can verify and build upon. Looking ahead, I will broaden this trajectory into several long-horizon thrusts that advance methods and evaluations for serving *social information processing needs*.

Towards improved modeling of semantic structures embedded in social data. Social texts such as clinical notes, political books, or judicial opinions frequently document human behavior. Rich

Ankita Gupta
ankitagupta@cs.umass.edu

semantic structures (e.g., arguments, narratives, connotation frames, stances) embedded in such texts can help *diagnose human biases*, as demonstrated by my prior work (**Gupta** et al., 2022; Cai et al., 2025). Such semantic structures extracted from unstructured data can also *improve broader equity and access to information*, as demonstrated by my work on aiding professionals and the general public in navigating complex information landscapes (**Gupta** et al., 2024b, 2025b). While LLMs show great promise for such semantic information extraction tasks and can significantly reduce manual effort, they still struggle with many challenges. For instance, my ongoing work shows that LLMs struggle to *model narrative components* (e.g., character desires, environment descriptions) (**Gupta** et al., 2025a), as well as reason over *temporally ordered procedural steps*. Moving forward, I will develop systematic benchmarks and improved methods for extracting and reasoning over semantic structures in social data. LLMs also struggle with modeling long-range dependencies found in *long-form arguments and narratives* (e.g., judicial opinions, news articles, policy documents). Building on techniques such as self-verification, synthetic data generation, and step-level reward modeling, I will explore how to evaluate and improve the long-form reasoning capabilities of LLMs and use them to better analyze social data.

Argumentation for improved reasoning and interpretability of LLMs and agents. My prior work has focused on using *LLMs for understanding arguments* (public comments (**Gupta** et al., 2024b), judicial opinions (**Gupta** et al., 2024b)). Conversely, *argumentation* can be useful *for improving LLMs*. Currently, evaluating and improving the reasoning capabilities of LLMs relies on extensive human preference judgments to train reward models. On the other hand, decades of scholarly work on argumentation provide systematic theories and frameworks for understanding reasoning (Toulmin, 1958; Walton, 1996; Lawrence and Reed, 2019). Building on my prior work, I will explore how reasoning schemas grounded in argumentation theories (e.g., Toulmin's theory) can analyze and evaluate LLM reasoning, inform training objectives that reward well-formed arguments, and enhance interpretability by representing LLM reasoning chains as explicit argument graphs. Another direction I am very excited about involves *multi-agent systems (MAS)*, where LLM-based agents interact and challenge each other to improve reasoning while mitigating single-agent failure modes like degeneration-of-thought (Liang et al., 2024). However, designing effective multi-agent interactions remains challenging and lacks principled frameworks. Building on techniques drawn from computational argumentation, I plan to design, evaluate, and improve reasoning in both single-agent and multi-agent systems, as well as model sophisticated agent interactions (e.g., consensus building, conflict resolution).

Argumentation for detecting AI-generated social engineering attacks. Besides human-generated arguments, LLMs can also generate arguments that have been shown to be as persuasive as humans in controlled experimental settings (Bai et al., 2025). These capabilities can be exploited by adversaries for various cybersecurity attacks, such as generating highly persuasive spear-phishing messages. For instance, prior work has reported that LLMs prompted to assist with spear-phishing attacks can produce click-through rates of 50%, comparable with human experts, on unknowing participants, suggesting their likely effectiveness in the real world (Heiding and Lermen, 2025). Building on my work on analyzing argumentative structures (**Gupta** et al., 2024b), I plan to develop methods to detect linguistic fingerprints in AI-generated social engineering attacks, such as characterizing argument structures in LLM-generated persuasive messages and using them to inform the development of interpretable social engineering attack detection systems. I will seek to collaborate with researchers in cybersecurity to design, develop, and validate these detection methods in simulated and realistic threat scenarios.

Broadening access to reliable information in high-impact domains. More broadly, I am committed to developing trustworthy AI models that address the specific needs of specialized fields through collaboration with domain experts.

For instance, I look forward to expanding my research in **Legal NLP**, which is an exciting emerging area, and I am actively engaged with this growing interdisciplinary community. I participated in the CS&Law 2025 conference, which connected me with scholars across disciplines and countries, opening several research directions. My prior work has laid the groundwork for analyzing legal arguments at scale (**Gupta** et al., 2025b). Moving forward, I plan to advance more *reliable* legal LMs through techniques like retrieval augmented generation, tool use, and long-form argument generation for more accurate LM response generation grounded in authoritative knowledge sources. My ongoing work on retrieving relevant precedents from millions of prior cases is a first step towards this goal. I further plan to develop *efficient* LMs through techniques like parameter-efficient fine-tuning and knowledge distillation to develop small open LMs for resource-constrained settings (e.g., legal aid). Additionally, I plan to expand this research to *other languages and countries*, providing insights into the cross-cultural and cross-lingual capabilities of generative AI. Finally, I intend to address the broader challenge of reducing administrative burden in

Ankita Gupta
ankitagupta@cs.umass.edu

accessing public services (e.g., housing, social services), a key challenge in access to justice (Burnett and Sandefur, 2024). Building on my prior work, this direction will involve developing new NLP tools, such as simplifying legalese and extracting key information from policy documents.

Another direction I plan to explore is to advance methods in **Civic NLP** to address the broad problem of enabling institutions to *listen* to public feedback *at scale*. While my prior work is on regulatory public comments (**Gupta** et al., 2024b), many other civic media channels (e.g., public town hall meetings, facilitated dialogues) contain rich argumentative public feedback that remains difficult to analyze. These settings will also introduce new technical challenges, such as processing lengthy multi-party deliberations and ensuring faithfulness to original citizen input. Building on my prior work (**Gupta** et al., 2022, 2024a), I will develop novel NLP benchmarks and methods to address these challenges.

Across all of these directions, I have been fortunate to establish strong **interdisciplinary connections** and publish with co-authors from linguistics, political science, public policy, legal studies, and health sciences, and I look forward to cultivating these bonds and developing more. For instance, in an ongoing project, I conducted qualitative interviews with legal professionals to understand their workflow when helping laypeople. In the future, I am committed to **human-centered computing**, and I plan to extend this line of work by seeking to collaborate with human-computer interaction (HCI) researchers to inform the design of my methods. Methodologically, I use a wide variety of **computational and statistical techniques** in my research, such as supervised and unsupervised machine learning algorithms, neural networks, statistical sampling methods, linear and non-linear optimization, distributed computation, databases, etc. These will be a key part of my future work, and I look forward to teaching them to students for use in both research and industry.

Funding sources. I plan to seek diverse funding from public, philanthropic, and industry sources. My prior work has been supported by NSF IIS, NIDCD/NIH, and an IBM Research fellowship. In 2025, my work and contributions to the grant proposal helped secure a three-year, $1.06 million grant from the Patient-Centered Outcomes Research Institute. I also plan to pursue interdisciplinary grants, such as from the Schmidt Sciences Humanities and AI Virtual Institute, and build upon my existing industry collaborations with IBM Research, Microsoft Research, and Thomson Reuters—who contacted me to discuss using the δ-Stance dataset (**Gupta** et al., 2025b) for evaluating their legal AI model.

Ankita Gupta
ankitagupta@cs.umass.edu

References

Hui Bai, Jan G. Voelkel, Shane Muldowney, Johannes C. Eichstaedt, and Robb Willer. LLM-generated messages can persuade humans on policy issues. *Nature Communications*, 16(1):6037, July 2025. doi: 10.1038/s41467-025-61345-5. URL https://doi.org/10.1038/s41467-025-61345-5.

Matthew Burnett and Rebecca L Sandefur. Justice work as democracy work: Reimagining access to justice as democratization. *SCL Rev.*, 76:833, 2024. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5365751.

Erica Cai, **Ankita Gupta**, Katherine A. Keith, Brendan O'Connor, and Douglas Rice. "Let me just interrupt you": Estimating gender effects in supreme court oral arguments. *Journal of Law and Courts*, page 1–22, 2025. URL https://doi.org/10.1017/jlc.2024.7.

Ankita Gupta, Su Lin Blodgett, Justin H Gross, and Brendan O'Connor. Examining political rhetoric with epistemic stance detection. In *Proceedings of the Fifth Workshop on Natural Language Processing and Computational Social Science (NLP+CSS)*, pages 89–104. Association for Computational Linguistics, 2022. URL https://aclanthology.org/2022.nlpcss-1.11/.

Ankita Gupta, Marzena Karpinska, Wenlong Zhao, Kalpesh Krishna, Jack Merullo, Luke Yeh, Mohit Iyyer, and Brendan O'Connor. ezCoref: Towards unifying annotation guidelines for coreference resolution. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 312–330. Association for Computational Linguistics, 2023. URL https://aclanthology.org/2023.findings-eacl.24/.

Ankita Gupta, Chulaka Gunasekara, Hui Wan, Jatin Ganhotra, Sachindra Joshi, and Marina Danilevsky. Evaluating robustness of open dialogue summarization models in the presence of naturally occurring variations. *Proceedings of the 6th Workshop on NLP for Conversational AI (NLP4ConvAI 2024)*, 2024a. URL https://aclanthology.org/2024.nlp4convai-1.4/.

Ankita Gupta, Ethan Zuckerman, and Brendan O'Connor. Harnessing Toulmin's theory for zero-shot argument explication. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, pages 10259–10276. Association for Computational Linguistics, 2024b. URL https://aclanthology.org/2024.acl-long.552/.

Ankita Gupta, Marisa Hudspeth, Polly Stokes, Jacquie Kurland, and Brendan O'Connor. Automated main concept generation for narrative discourse assessment in aphasia. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 24437–24451. Association for Computational Linguistics, 2025a. URL https://aclanthology.org/2025.findings-acl.1255/.

Ankita Gupta, Douglas Rice, and Brendan O'Connor. δ-stance: A large-scale real world dataset of stances in legal argumentation. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics*, pages 31450–31467. Association for Computational Linguistics, 2025b. URL https://aclanthology.org/2025.acl-long.1517/.

Fred Heiding and Simon Lermen. Evaluating large language models' capability to launch fully automated spear phishing campaigns. In *ICML 2025 Workshop on Reliable and Responsible Foundation Models*, 2025. URL https://openreview.net/forum?id=f0uFpuea1s.

Jacquie Kurland, Vishnupriya Varadharaju, Anna Liu, Polly Stokes, **Ankita Gupta**, Marisa Hudspeth, and Brendan O'Connor. Large language models' ability to assess main concepts in story retelling: A proof-of-concept comparison of human versus machine ratings. *American Journal of Speech-Language Pathology*, pages 1–11, 2025. URL https://pubs.asha.org/doi/abs/10.1044/2025_AJSLP-24-00400.

John Lawrence and Chris Reed. Argument mining: A survey. *Computational Linguistics*, 45(4):765–818, 2019. URL https://aclanthology.org/J19-4006/.

Tian Liang, Zhiwei He, Wenxiang Jiao, Xing Wang, Yan Wang, Rui Wang, Yujiu Yang, Shuming Shi, and Zhaopeng Tu. Encouraging divergent thinking in large language models through multi-agent debate. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 17889–17904. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.emnlp-main.992. URL https://aclanthology.org/2024.emnlp-main.992/.

Anna Rogers, Marzena Karpinska, **Ankita Gupta**, Vladislav Lialin, Gregory Smelkov, and Anna Rumshisky. Narrativetime: Dense temporal annotation on a timeline. In *International Conference on Language Resources and Evaluation*, 2024. URL https://aclanthology.org/2024.lrec-main.1054/.

Shivalika Singh, Freddie Vargus, Daniel D'souza, Börje F. Karlsson, Abinaya Mahendiran, Wei-Yin Ko, Herumb Shandilya, Jay Patel, Deividas Mataciunas, Laura O'Mahony, Mike Zhang, Ramith Hettiarachchi, Joseph Wilson, Marina Machado, Luisa Moura, Dominik Krzemiński, Hakimeh Fadaei, Irem Ergun, Ifeoma Okoh, Aisha Alaagib, Oshan Mudannayake, Zaid Alyafeai, Vu Chien, Sebastian Ruder, Surya Guthikonda, Emad Alghamdi, Sebastian Gehrmann, Niklas Muennighoff, Max Bartolo, Julia Kreutzer, Ahmet Üstün, Marzieh Fadaee, and Sara Hooker. Aya Dataset: An open-access collection for multilingual instruction tuning. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, pages 11521–11567. Association for Computational Linguistics, 2024. URL https://aclanthology.org/2024.acl-long.620/.

Stephen E. Toulmin. *The Uses of Argument*. Cambridge University Press, 1958.

Douglas Walton. *Argumentation Schemes for Presumptive Reasoning*. Lawrence Erlbaum Associates, 1996. URL https://doi.org/10.4324/9780203811160.

Dominik Wawrzuta, Mariusz Jaworski, Joanna Gotlib, and Mariusz Panczyk. What arguments against covid-19 vaccines run on facebook in poland: content analysis of comments. *Vaccines*, 9(5):481, 2021. URL https://www.mdpi.com/2076-393X/9/5/481.