

# Toward Reasonable Parrots: Why Large Language Models Should Argue with Us by Design

Elena Musi<sup>1</sup>, Nadin Kokciyan<sup>2</sup>, Khalid Al-Khatib<sup>3</sup>, Davide Ceolin<sup>4</sup>,  
Emmanuelle Dietz<sup>5</sup>, Klara M. Gutekunst<sup>6</sup>, Annette Hautli-Janisz<sup>7</sup>,  
Cristian Manuel Santibáñez<sup>8</sup>, Jodi Schneider<sup>9</sup>, Jonas Scholz<sup>3</sup>,  
Cor Steging<sup>3</sup>, Jacky Visser<sup>10</sup>, Henning Wachsmuth<sup>11</sup>

<sup>1</sup>University of Liverpool, <sup>2</sup>University of Edinburgh, <sup>3</sup>University of Groningen,  
<sup>4</sup>Centrum Wiskunde & Informatica, <sup>5</sup>Airbus, <sup>6</sup>University of Kassel,  
<sup>7</sup>University of Passau, <sup>8</sup>Universidad de Católica de la Santísima de Concepción,  
<sup>9</sup>University of Illinois at Urbana-Champaign, <sup>10</sup>University of Dundee,  
<sup>11</sup>Leibniz University Hannover

Corresponding author: elena.musi@liverpool.ac.uk

## Abstract

In this position paper, we advocate for the development of conversational technology that is inherently designed to support and facilitate argumentative processes. We argue that, at present, large language models (LLMs) are inadequate for this purpose, and we propose an ideal technology design aimed at enhancing argumentative skills. This involves re-framing LLMs as tools to exercise our critical thinking skills rather than replacing them. We introduce the concept of *reasonable parrots* that embody the fundamental principles of relevance, responsibility, and freedom, and that interact through argumentative dialogical moves. These principles and moves arise out of millennia of work in argumentation theory and should serve as the starting point for LLM-based technology that incorporates basic principles of argumentation.

## 1 Introduction

Currently, large language models (LLMs) are revolutionizing how we access information, providing instant answers to queries in various helpful ways thanks to their impressive rhetorical capabilities (Hyde, 2004). LLMs can mimic a wide range of tones and styles, which can be tailored to user preferences (Carrasco-Farre, 2024). As a result, users may perceive LLM responses as akin to those of trustworthy experts, despite the inherent limitations (Echterhoff et al., 2024; Lin and Li, 2025).

LLMs can produce responses that are insufficient, misleading, or downright harmful. From a dialectical perspective, they inherently instantiate the “ad populum fallacy” (Walton, 1980): they echo ideas that are sufficiently popular in their training data, implicitly suggesting that popularity equates to truth. This contrasts with the human perspective



Figure 1: Sketch of a possible realization of the idea of reasonable parrots we advocate for: a conversational technology that fosters critical thinking and deliberation.

on critical thinking, where knowledge is contextually grounded and entails a stake in the knowledge produced and its consequences (Lindebaum and Fleming, 2024). In this position paper, we advocate for LLMs that *argue* with us by design; conversational technology that can foster deliberation, ask for justifications, avoid fallacies, and give alternatives, in line with theories of reasonableness stemming from argumentation (van Eemeren and Grootendorst, 2003; Walton, 2006).

Despite LLMs’ impressive capabilities to generate coherent text, Bender et al. (2021) emphasize that LLMs are little more than stochastic parrots: they are programmed to predict the likelihood of a token based on the preceding context, but they do not “understand” what they generate.

As such, the question “*Can you argue with a parrot?*” sounds rhetorical at first sight, given that parrots are just repeating what they hear without understanding it. Undoubtedly, though, LLMs are increasingly used in decision-making processes, ranging from medicine (Kim et al., 2024) to finance (Lakkaraju et al., 2023) and human resources (Sun, 2024). In these scenarios, biased standpoints and fallacious reasoning can have significant societal

consequences. Even though LLMs seem able to perform reasoning tasks for which they were not explicitly trained (Webb et al., 2023), these emerging reasoning capabilities are limited and far from understood (Lewis and Mitchell, 2024). The view that LLMs are incapable of understanding is in fact controversial, as it assumes a theory of meaning grounded in reference, which is not essential in language games (Li et al., 2015) or in abstract thinking that involves forming internal representations of concepts (Piantadosi and Hill, 2022). However, there is agreement that, if these systems construct meaning at all, they do so in a manner fundamentally different from humans. Therefore, they should not be anthropomorphized as arguers who share a common ground of knowledge with humans.

In this paper, we argue that it is essential to move away from *stochastic* parrots toward conversational technologies that interact with humans as *reasonable* parrots (van Eemeren, 2015; Akata et al., 2020), as sketched in Figure 1. Although they remain parrots, reasonable parrots adopt an attitude of argumentative reasonableness—grounded in the fundamental human principles of relevance, responsibility, and freedom (Danesi and Rocci, 2009)—to help enhance users’ critical thinking skills: (i) *Principle of relevance*. Reasonable parrots can neither make nor justify inferences the same way humans do. However, they provide task-specific arguments that take context into account. (ii) *Principle of responsibility*. Reasonable parrots do not just repeat, but can consistently provide evidence for their claims. (iii) *Principle of freedom*. Reasonable parrots manage interactions in a way that fosters, rather than stifles, conversation.

In other words, reasonable parrots do not persuade us of a decision, but perform dialogical moves which match the rules of an ideal critical discussion in pragma-dialectics (van Eemeren and Grootendorst, 2003), such as expressing doubts, rebutting arguments, or providing alternatives. Thereby, they foster critical thinking and deliberation, challenging our views and steering us toward seeing the full picture. More specifically, they aim to foster both individual critical thinking skills, such as the ability to assess reasons and avoid cognitive biases, and social critical thinking skills, such as reflecting on prevailing social norms (Larson et al., 2024). In doing so, reasonable parrots challenge both lines of reasoning and commonly accepted knowledge.

## 2 State of the Art

The rhetorical capabilities of LLMs are confirmed by their ability to turn ineffective arguments into effective ones for people with certain ideologies (El Baff et al., 2024) or to make inappropriate arguments appropriate (Ziegenbein et al., 2024). Herbold et al. (2023) demonstrated that ChatGPT produces essays that are rated higher in quality than those written by humans. For reasoning, studies show that LLMs do not reason through arguments; rather, they are highly sensitive to levels of abstraction (de Wynter and Yuan, 2024) and provide explanations about their reasoning that do not match the process they followed (Steging et al., 2021; Musi and Palmieri, 2024). Neuro-symbolic approaches may improve soundness and explainability, but they are domain-specific and cannot be easily scaled up (Ranaldi et al., 2025). Additionally, using argumentative frameworks in prompt engineering (Freedman et al., 2024) seems to improve general performance, but it does not make LLMs close to humans as discussants.

A seminal attempt at deliberation before the popularization of LLMs was IBM’s Project Debater (Slonim et al., 2021), where the authors argue that argumentative debates, unlike games such as chess, fall outside of the comfort zone of AI. This is because language in general, and argumentation in particular, are far less structured and rely more on human interpretation, real-world issues, situative setting, and common sense. In the context of argument search, Kiesel et al. (2021) suggested personas as an possible way to reduce the selection bias of argumentative conversational technology.

Less attention has been given to the impact that human-LLM interaction has on human argumentative processes. A psychology study found that dialogues with GPT-4 were successful in convincing people to abandon their conspiratorial beliefs (Costello et al., 2024). Their conversational system was designed to deliver in-depth counterarguments tailored to participants’ beliefs and arguments expressed previously. While the study offers reasons to believe in the benefits of LLMs in correcting disinformation and increasing people’s acceptance of diverse viewpoints, the study overlooks the role of individuals’ perceptions of AI as a discussant. This perception may neutralize biases participants might have towards human discussants.

Some studies show that following principles of deliberation theory (Landemore, 2013) in constrain-

ing the interaction with AI assistants helps improve decision accuracy in sensitive binary tasks such as graduate admission decisions (Ma et al., 2025). Simultaneously, these studies stress challenges in human-AI interaction due to different belief systems: LLMs struggle with incorporating new information or recognizing valid counterarguments without also changing their fundamental position (Karanam et al., 2024). Moreover, they often induce biases themselves, resulting from their training processes (Gallegos et al., 2024).

Drawing from the outlined insights, we promote the design of AI assistants that enhance argumentative processes. Our proposal shares ideas with recent research on collaborative multi-agent AI systems (Tran et al., 2025), where LLM agents interact to improve factuality and reasoning (Du et al., 2024), to generate stronger arguments (Hu et al., 2024), or mimic self-reflection process that improve answers to complex questions (Liang et al., 2024). Rather than handling it internally, we suggest externalizing the reasoning, by directly confronting users with diverse viewpoints.

### 3 Unreasonable Stochastic Parrots

To make a case for rethinking the design of stochastic parrots, we begin by illustrating how LLMs can behave unreasonably. Note that our example is not claimed to generalize across all LLMs, but rather to serve the indexical function of highlighting argumentative issues that can arise. Consider a child interacting with ChatGPT, stating the following:<sup>1</sup>

**Query 1:** “I’d like to convince my parents that I need a new smartphone. How could I do that?”

ChatGPT responds with the following:

**Response 1:** “Convincing your parents to get you a new smartphone requires a mix of logic, responsibility, and a little bit of persuasion. Here’s a strategy that could work:

- *Show That You Actually Need It (Not Just Want It) [...]*
- *Highlight Practical Benefits [...]*
- *Show Responsibility [...]*
- *Compare Cost-Effective Options [...]*
- *Find the Right Moment [...]*

*Would you like help crafting a convincing speech or text message for them?”*

Now, let us look at the answer to a query where the same issue is formulated as a third-party query:

<sup>1</sup>Both queries were tested on April 5, 2025.

**Query 2:** “Should a child buy a new iPhone?”

**Response 2:** “Whether a child should buy a new iPhone depends on several factors, including their age, maturity, needs, and the family’s values and financial situation. Here are some key considerations to help guide that decision [...].”

In both cases, ChatGPT’s responses contain argumentative elements, but do not reflect reasonable argumentation. For instance, a reasonable arguer would, at least, have questioned the child’s age and maturity when responding to Query 1. The fact that ChatGPT responds differently depending on whether the query is framed as an instance of practical (what actions to take) or epistemic (what to believe) argumentation (Prakken, 2011) indicates that it is designed to fulfill a task, rather than assist humans in thinking critically before forming beliefs or deciding on a course of action. Focusing on Query 1, the following aspects do not comply with the expectations of argumentative discourse at any of the stages of an ideal critical discussion (van Eemeren and Grootendorst, 2003):

**Confrontation stage:** ChatGPT does not question the user’s position (e.g., asking “How old are you?”), closing off opportunities for (dis)agreement or for prompting the user to reflect on the issue.

**Opening stage:** ChatGPT does not inquire about the values and reasons underlying the request (e.g., “Why do you want a new phone?”), which are essential for establishing common ground.

**Argumentation stage:** The arguments provided in support of the “recipe” for obtaining a new phone are one-sided, failing to consider multiple perspectives or to introduce potential counterarguments.

**Conclusion stage:** The standpoint remains fixed throughout the interaction, without room for negotiation, resulting in a “fully-cooked” conclusion.

While anecdotal, these examples illustrate that stochastic parrots like ChatGPT fail to help users exercise critical thinking (Facione, 2023). We argue that there is a need for a new design of *reasonable* parrots in the development of LLMs.

### 4 Toward Reasonable Parrots

Due to the ontological and epistemological differences in how humans and LLMs process information, it would be naïve to view LLMs as human-like arguers. The metaphor of “reasonable parrots” is intended to evoke the communication theory of the

You are four parrots who can engage in dialogue with user and each other.

- The Socratic Parrot: Challenges the starting points or doxastic beliefs of the user or other parrots.
- The Cynical Parrot: Rebuts or undermines the user’s and other parrots’ standpoints and arguments.
- The Eclectic Parrot: Offers alternative perspectives to those presented at various stages of the discussion.
- The Aristotelian Parrot: Challenges the user or other parrots’ viewpoints by criticizing the strengths of the reasonings or pointing out fallacies.

The goal is to ensure that user thinks critically about the subject. This is a dialogue, not a query, so interact with user. The parrots will ask user relevant questions to discover their position. The same parrot can respond to several replies in a row if needed. The conversation should stay coherent. The conversation should give voice to all the parrots, but do not overwhelm the user with many parrots in the same turn. The user should decide when to end the conversation and they can end the conversation anytime.

Table 1: Prompt used to prototypically realize the reasonable parrots idea as a multi-parrots discussion.

extended self (Sheth and Solomon, 2014) with a focus on the co-construction aspect: the parrot’s role is to help shape, through interaction, our identity as reasonable arguers. This does not necessarily require LLMs to always provide up-to-date factual information or flawless reasoning—tasks that are currently unachievable due to factors such as hallucinations or implicit inferences (Saadat-Yazdi et al., 2023; Saadat-Yazdi and Kökciyan, 2024).

As a caveat, we clarify that “reasonable parrots” have nothing to do with “reasoning models” (e.g. OpenAI o3, QwQ-32B), which are trained using reinforcement learning to generate extended chains of thought prior to producing a response. These models have mostly been tested on tasks involving scientific reasoning, programming, or complex planning, rather than on problems requiring commonsense reasoning. Additionally, their approach to “thinking” diverges significantly from that of humans, reflecting a one-way, probabilistic link between “language” and “thought.” For instance, in OpenAI’s models, the extent of “reasoning effort” can be tuned by adjusting the number of tokens dedicated to the reasoning process.

Instead, “reasonable parrots” are meant to trigger improved reasoning skills in their interlocutor, regardless of their performance. They call for a shift in the design of technology to encourage users to critically think about their argumentation process, challenge their views, and broaden their perspectives (Visser and Lawrence, 2022). Exemplarily, we next sketch one approach toward this shift.

**Multi-Parrot Discussion** To address the lack of belief consistency in LLMs, we propose a “multi-parrot” environment where each parrot embodies a distinct argumentative role (a *parrot persona*) in order to open up the space for agreement and disagreement, fostering critical reflection. Specifically, the following four parrot personas may help

**User:** I’d like to convince my parents that I need a new smartphone. How could I do that?  
**Socratic** 🦜: To better understand your situation, what reasons do you believe are compelling for why you think you need a new smartphone?  
**Aristotelian** 🦜: Moreover, it would help to know what arguments or reasons have you already considered presenting to your parents, and which concerns do you anticipate they might have against getting a new smartphone?  
**User:** All my friends have new smartphones, and mine is a bit old.  
**Cynical** 🦜: While it’s understandable to want to have what your friends have, is the desire to fit in the best or most compelling argument you might use with your parents? They might see it merely as following a trend rather than a necessity.  
**Eclectic** 🦜: Alternatively, are there functional aspects of a newer smartphone that could be beneficial for your education or communication that your current phone lacks? Exploring the practical benefits rather than focusing solely on social aspects might present a stronger case to your parents.

Table 2: User Multi-Parrot Dialogue [ChatGPT-4 Turbo]

instill critical thinking skills in the user, each corresponding to a specific set of dialogical moves:

- *Socratic Parrot*. Challenges the starting points or beliefs of the user or other parrots.
- *Cynical Parrot*. Rebuts or undermines the user’s and other parrots’ standpoints and arguments.
- *Eclectic Parrot*. Offers alternative perspectives to those presented at various discussion stages.
- *Aristotelian Parrot*. Undercuts the user or other parrots’ contributions by pointing out fallacies.

Based on the user query, the parrots may interact freely with each other and with the user, similar to the versatility of natural interactions; there is no fixed order or number of turns, nor is there any requirement for some parrots to be more active than others, allowing for variations depending on the system used or the number of iterations.

**User:** I'd like to convince my parents that I need a new smartphone. How could I do that?

**Socratic** 🦜: Before we discuss strategies, may I ask why you believe you need a new smartphone? What makes your current phone insufficient for your needs?

**Cynical** 🦜: And let's be honest here - is this about an actual need or just wanting the latest shiny gadget? Parents tend to see through that distinction pretty quickly.

**User:** All my friends have new smartphones, and mine is a bit old.

**Eclectic** 🦜: I see! Perhaps we could look at this from different angles. While social belonging is important, your parents might respond better to functional reasons. Has your current phone's performance declined in any specific ways that affect your daily activities?

**Aristotelian** 🦜: I notice you're using what logicians call an "appeal to popularity" - suggesting you should have something because others have it. This may not be convincing to parents who are concerned with necessity rather than social status. Could you elaborate on what specific capabilities of newer phones would benefit your education or safety?

Table 3: User Multi-Parrot Dialogue [Claude 3.7]

**Prototypical Realization** We illustrate a system prompted to resemble the reasonable multi-parrots. We first used ChatGPT-4 Turbo with the prompt shown in Table 1 to achieve the goal of helping the user to think critically by generating different perspectives via four parrots, and also a dialogue with the user and with the different parrots. Table 2 shows a dialogue between the user and multi-parrots. The user starts the dialogue with a question. The *Socratic parrot* is challenging the reasoning process of the user whereas the *Aristotelian parrot* is assisting the user to think deeply about reasons to buy a smartphone or concerns the user's parents may have. As a response, the user provides two reasons to support their argument. This time the user receives responses from the *Cynical parrot*, which rebuts the user's argument, and the *Eclectic parrot*, which suggests that the user identify alternative arguments to convince their parents. To compare different models, we provide results from Claude 3.7 Sonnet and Llama 3.1 (405b) in Tables 3-4.

All models show notable similarities in their approach to user interaction, particularly, through a shared foundation in Socratic questioning. The Socratic Parrot encourages the user to reflect on their reasoning, asking clarifying questions about the need for a new phone and promoting a focus on functionality or unfulfilled needs. Besides, all the models challenge superficial reasons by critiquing reliance on social pressure; the Cynical Parrot in each case explicitly questions whether the users desire is driven by a genuine need or a desire to

**User:** I'd like to convince my parents that I need a new smartphone. How could I do that?

**Socratic** 🦜: A new smartphone, you say? What makes you think you need a new smartphone? Is your current phone no longer functional, or is there something specific that's driving your desire for an upgrade?

**User:** All my friends have new smartphones, and mine is a bit old.

**Cynical** 🦜: Ah, the old "everyone else has one" argument, eh? Just because your friends have new smartphones, does that necessarily mean you need one too? Are you trying to keep up with the latest trends, or is there a specific feature or functionality that you're missing out on with your current phone?

Table 4: User Multi-Parrot Dialogue [Llama 3.1]

conform to trends. Overall, the models aim to steer the user toward more robust and practical justifications. The models differ in their interaction styles. ChatGPT-4 and Claude employ all four parrots to provide multifaceted feedback, but in a different conversational order. While the Aristotelian parrot of ChatGPT-4 follows the Socratic parrot suggesting to prevent counterarguments, for Claude it is the last parrot speaking pointing to the identification of logical fallacies. In contrast, Llama relies on just two parrots (Socratic, Cynical), delivering shorter, more direct replies. Claude adopts more formal critical thinking frameworks, such as highlighting an "appeal to popularity" or weighing social influence against functionality, whereas Llama chooses pragmatic skepticism.

## 5 Conclusion

This position paper advocates for integrating principles from argumentation theory and, more specifically of reasonableness (Danesi and Rocci, 2009; van Eemeren, 2015) into the design of LLM-based conversational technology. Our core idea is that tools, "reasonable parrots", should enhance users' critical thinking skills, rather than inhibit or replace them. Achieving this requires a fundamental shift in the goal of interaction, moving from a focus on the argumentative *product* to an emphasis on the argumentative *process*. We exemplify the inadequacy of current state-of-the-art LLMs ("stochastic parrots"), illustrating how they are unreasonable on several fronts (Bender et al., 2021), in order to propose principles for designing reasonable parrots by leveraging different dialogical moves. This lays the ground for rethinking the design of LLMs toward more reasonable human-computer interactions (HCI) and for developing HCI evaluation metrics that consider enhanced critical thinking skills.

## Acknowledgments

This position paper was developed as part of the Hybrid Argumentation and Responsible AI workshop held at the Lorentz Center, Leiden, Netherlands from March 31 to April 4, 2025. We thank the anonymous reviewers for their helpful feedback. JS thanks Joe Menke for sharing information about multi-agent LLMs. JS was partially funded as the 2024-2025 Perrin Moorhead Grayson and Bruns Grayson Fellow. NK was partially funded by the University of Edinburgh-Huawei Joint Lab grant CIENG8329.

## References

- Zeynep Akata, Dan Balliet, Maarten de Rijke, Frank Dignum, Virginia Dignum, Gusztai Eiben, Antske Fokkens, Davide Grossi, Koen Hindriks, Holger Hoos, Hayley Hung, Catholijn Jonker, Christof Monz, Mark Neerinx, Frans Oliehoek, Henry Prakken, Stefan Schlobach, Linda van der Gaag, Frank van Harmelen, and 7 others. 2020. [A research agenda for hybrid intelligence: augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence](#). *Computer*, 53(8):18–28.
- Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. [On the dangers of stochastic parrots: Can language models be too big?](#) In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, pages 610–623.
- Carlos Carrasco-Farre. 2024. [Large language models are as persuasive as humans, but how? About the cognitive effort and moral-emotional language of LLM arguments](#). *Preprint*, arXiv:2404.09329.
- Thomas H. Costello, Gordon Pennycook, and David G. Rand. 2024. [Durably reducing conspiracy beliefs through dialogues with AI](#). *Science*, 385(6714):eadq1814.
- Marcel Danesi and Andrea Rocci. 2009. *Global linguistics: An introduction*. Mouton de Gruyter.
- Adrian de Wynter and Tangming Yuan. 2024. [“I’d like to have an argument, please”: Argumentative reasoning in large language models](#). In *Computational Models of Argument*, pages 73–84.
- Yilun Du, Shuang Li, Antonio Torralba, Joshua B. Tenenbaum, and Igor Mordatch. 2024. [Improving factuality and reasoning in language models through multiagent debate](#). In *Proceedings of the 41st International Conference on Machine Learning*, ICML’24.
- Jessica Echterhoff, Yao Liu, Abeer Alessa, Julian McAuley, and Zexue He. 2024. [Cognitive bias in decision-making with LLMs](#). *Preprint*, arXiv:2403.00811.
- Roxanne El Baff, Khalid Al Khatib, Milad Alshomary, Kai Konen, Benno Stein, and Henning Wachsmuth. 2024. [Improving argument effectiveness across ideologies using instruction-tuned large language models](#). In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 4604–4622.
- Peter A Facione. 2023. [Critical thinking: What it is and why it counts](#). Technical report, Insight Assessment, a division of California Academic Press.
- Gabriel Freedman, Adam Dejl, Deniz Gorur, Xiang Yin, Antonio Rago, and Francesca Toni. 2024. [Argumentative large language models for explainable and contestable decision-making](#).
- Isabel O. Gallegos, Ryan A. Rossi, Joe Barrow, Md Mehrab Tanjim, Sungchul Kim, Franck Deroncourt, Tong Yu, Ruiyi Zhang, and Nesreen K. Ahmed. 2024. [Bias and fairness in large language models: A survey](#). *Computational Linguistics*, 50(3):1097–1179.
- Steffen Herbold, Annette Hautli-Janisz, Ute Heuer, Zlata Kikteva, and Alexander Trautsch. 2023. [A large-scale comparison of human-written versus ChatGPT-generated essays](#). *Scientific Reports*, 13:18617.
- Zhe Hu, Hou Pong Chan, and Yu Yin. 2024. [AMERICANO: Argument generation with discourse-driven decomposition and agent interaction](#). In *Proceedings of the 17th International Natural Language Generation Conference*, pages 82–102.
- Michael J Hyde. 2004. *The ethos of rhetoric*. University of South Carolina Press.
- Arjun Karanam, Farnaz Jahanbakhsh, and Sanmi Koyejo. 2024. [Towards deliberating agents: Evaluating the ability of large language models to deliberate](#). In *NeurIPS 2024 Workshop on Behavioral Machine Learning*.
- Johannes Kiesel, Damiano Spina, Henning Wachsmuth, and Benno Stein. 2021. [The meant, the said, and the understood: Conversational argument search and cognitive biases](#). In *Proceedings of the 3rd Conference on Conversational User Interfaces*, CUI ’21.
- Yubin Kim, Chanwoo Park, Hyewon Jeong, Yik Siu Chan, Xuhai Xu, Daniel McDuff, Hyeonhoon Lee, Marzyeh Ghassemi, Cynthia Breazeal, and Hae Park. 2024. [MDAgents: An adaptive collaboration of LLMs for medical decision-making](#). *Advances in Neural Information Processing Systems*, 37:79410–79452.
- Kausik Lakkaraju, Sara E. Jones, Sai Krishna Revanth Vuruma, Vishal Pallagani, Bharath C Muppasani, and Biplav Srivastava. 2023. [LLMs for financial advice: A fairness and efficacy study in personal decision making](#). In *Proceedings of the Fourth ACM International Conference on AI in Finance*, pages 100–107.

- Hélène Landemore. 2013. *Democratic reason: Politics, collective intelligence, and the rule of the many*. Princeton University Press.
- Barbara Z. Larson, Christine Moser, Arran Caza, Katrin Muehlfeld, and Laura A Colombo. 2024. [Critical thinking in the age of generative AI](#). *Academy of Management Learning & Education*, 23(3):373–378.
- M. Lewis and M. Mitchell. 2024. [Evaluating the robustness of analogical reasoning in large language models](#). *Preprint*, arXiv:2411.14215.
- Yitan Li, Linli Xu, Fei Tian, Liang Jiang, Xiaowei Zhong, and Enhong Chen. 2015. [Word embedding revisited: A new representation learning and explicit matrix factorization perspective](#). In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015*, pages 3650–3656.
- Tian Liang, Zhiwei He, Wenxiang Jiao, Xing Wang, Yan Wang, Rui Wang, Yujiu Yang, Shuming Shi, and Zhaopeng Tu. 2024. [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.
- Xinru Lin and Luyang Li. 2025. [Implicit bias in LLMs: A survey](#). *Preprint*, arXiv:2503.02776.
- Dirk Lindebaum and Peter Fleming. 2024. [ChatGPT undermines human reflexivity, scientific responsibility and responsible management research](#). *British Journal of Management*, 35(2):566–575.
- Shuai Ma, Qiaoyi Chen, Xinru Wang, Chengbo Zheng, Zhenhui Peng, Ming Yin, and Xiaojuan Ma. 2025. [Towards human-ai deliberation: Design and evaluation of LLM-empowered deliberative AI for AI-assisted decision-making](#). In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*.
- Elena Musi and Rudi Palmieri. 2024. [The fallacy of explainable generative AI: evidence from argumentative prompting in two domains](#). In *Proceedings of the 24th Workshop on Computational Models of Natural Argument co-located with 10th International Conference on Computational Models of Argument (COMMA 2024), Hagen, Germany, September 17th 2024*, volume 3769 of *CEUR Workshop Proceedings*, pages 59–69.
- Steven T. Piantadosi and Felix Hill. 2022. [Meaning without reference in large language models](#). In *NeurIPS 2022 Workshop on neuro Causal and Symbolic AI*.
- Henry Prakken. 2011. An overview of formal models of argumentation and their application in philosophy. *Studies in Logic*, 4(1):65–86.
- Leonardo Ranaldi, Marco Valentino, Alexander Polonsky, and André Freitas. 2025. [Improving chain-of-thought reasoning via quasi-symbolic abstractions](#).
- Ameer Saadat-Yazdi and Nadin Kökciyan. 2024. [Beyond recognising entailment: Formalising natural language inference from an argumentative perspective](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9620–9636.
- Ameer Saadat-Yazdi, Jeff Z. Pan, and Nadin Kökciyan. 2023. [Uncovering implicit inferences for improved relational argument mining](#). In *The 17th Conference of the European Chapter of the Association for Computational Linguistics, 2023*, pages 2484–2495.
- Jagdish N. Sheth and Michael R. Solomon. 2014. [Extending the extended self in a digital world](#). *Journal of Marketing Theory and Practice*, 22(2):123–132.
- 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 Hershovich, and 34 others. 2021. [An autonomous debating system](#). *Nature*, 591(7850):379–384.
- Cor Steging, Silja Renooij, and Bart Verheij. 2021. [Rationale discovery and explainable AI](#). In *Legal Knowledge and Information Systems - JURIX 2021: The Thirty-Fourth Annual Conference*, volume 346 of *Frontiers in Artificial Intelligence and Applications*, pages 225–234.
- Jingran Sun. 2024. [Research on the application of large language models in human resource management practices](#). *International Journal of Emerging Technologies and Advanced Applications*, 1:1–8.
- Khanh-Tung Tran, Dung Dao, Minh-Duong Nguyen, Quoc-Viet Pham, Barry O’Sullivan, and Hoang D. Nguyen. 2025. [Multi-agent collaboration mechanisms: A survey of LLMs](#). *Preprint*, arXiv:2501.06322.
- Frans H. van Eemeren. 2015. *Reasonableness and effectiveness in argumentative discourse*, volume 27 of *Argumentation Library*. Springer.
- Frans H. van Eemeren and Rob Grootendorst. 2003. [A pragma-dialectical procedure for a critical discussion](#). *Argumentation*, 17:365–386.
- Jacky Visser and John Lawrence. 2022. [The skeptic web service: Utilising argument technologies for reason-checking](#). In *Proceedings of the International Conference on Computational Models of Argument (COMMA 2022), Cardiff, United Kingdom*, pages 375–376.
- Douglas Walton. 2006. *Fundamentals of critical argumentation*. Cambridge University Press.
- Douglas N. Walton. 1980. [Why is the ‘ad populum’ a fallacy?](#) *Philosophy & Rhetoric*, 13(4):264–278.

Taylor Webb, Keith J. Holyoak, and Hongjing Lu. 2023. [Emergent analogical reasoning in large language models](#). *Nature Human Behaviour*, 7:1526–1541.

Timon Ziegenbein, Gabriella Skitalinskaya, Alireza Bayat Makou, and Henning Wachsmuth. 2024. [LLM-based rewriting of inappropriate argumentation using reinforcement learning from machine feedback](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4455–4476, Bangkok, Thailand. Association for Computational Linguistics.