

# A Multi-Perspective AI Framework for Mitigating Disinformation Through Contextual Analysis and Socratic Dialogue

Manuel Delaflor<sup>1</sup> and Carlos Toxtli<sup>2</sup>

<sup>1</sup>Metacognition Institute, Derbyshire, U.K.

<sup>2</sup>Clemson University, Clemson, USA

## ABSTRACT

The proliferation of digital information channels has created an unprecedented challenge in discerning credible information from sophisticated disinformation campaigns. Traditional fact-checking methods, often relying on binary true/false classifications, struggle to address the complexity, context-dependency, and nuanced nature of many claims circulating online. This limitation underscores the urgent need for advanced tools that empower individuals to critically evaluate information from multiple angles. Our AI-driven framework, Belief Explorer combines persistent contextual memory with Socratic dialogue and a three-lens analytical pipeline to foster deeper understanding and resilience against manipulation. As users interact, inputs are segmented and stored. Each claim is evaluated by Empirical, Logical, and Pragmatic LLM arbiters, synthesized into metrics: Verifact Score (evidence strength), Model Diversity Quotient (inter-arbiter agreement), Contextual Sensitivity Index (scenario appropriateness), and Reflective Index (exposed assumptions). A Perspective Generator crafts alternatives, promoting epistemic humility. Belief Explorer aims to shift users from seeking definitive “truth” to evaluating the pragmatic utility and coherence of information models, a vital skill in an era of amplified, often weaponized, narratives.

**Keywords:** Disinformation, Critical thinking, Artificial intelligence, Model dependent ontology, Multi-perspective analysis, Epistemic humility

## INTRODUCTION

The 21st century is characterized by an information environment of unprecedented scale and complexity, yet paradoxically, this abundance has precipitated what many term a “post-truth” era (Lewandowsky, Ecker, & Cook, 2017). This condition is marked by the erosion of trust in established epistemic authorities, the ascendancy of emotional appeals over verifiable evidence, and the strategic deployment of “alternative facts” by political actors, particularly those advancing populist and absolutist agendas. These agendas thrive by presenting simplistic, unassailable “truths” while systematically discrediting mediating institutions that traditionally contextualize and validate information (Waisbord, 2018).

Traditional fact-checking mechanisms, while valuable, often operate on a binary true/false model that proves insufficient against the deluge of nuanced disinformation, decontextualized narratives, and weaponized communication. They struggle with the sheer volume and velocity of information, and their pronouncements can be easily dismissed or politicized in highly polarized environments. Compounding this, the digital public sphere is increasingly dominated by an “amplifier effect,” where the loudest voices, often those with substantial resources, artificially inflated followings, or an intent to deceive, drown out reasoned discourse. This perversion of “free speech” allows weaponized narratives, including those propagated by sophisticated AI chatbots designed to manipulate public opinion or incite conflict by entrenching artificial “opposites,” to gain undue prominence (Bradshaw & Howard, 2019).

This paper argues that navigating this treacherous epistemic landscape requires more than just better fact-checking, it demands a fundamental shift in how individuals engage with information. An epistemic framework, called Model Dependent Ontology (MDO), articulated by Delaflor (2024), offers a robust philosophical grounding for such a shift. MDO posits that all human understanding is constituted by models, and these models are evaluated not by their correspondence to an elusive objective “truth,” but by their pragmatic utility and other metrics.

Our Belief Explorer is then an AI-driven tool designed to foster critical thinking and epistemic humility. Belief Explorer employs a multi-perspective analytical pipeline and Socratic dialogue to guide users in examining the coherence, contextual appropriateness, and underlying assumptions of information models. This paper details the epistemic justification for Belief Explorer, rooted in MDO, and argues for its urgent societal value in empowering individuals to critically navigate the complexities of the contemporary information ecosystem.

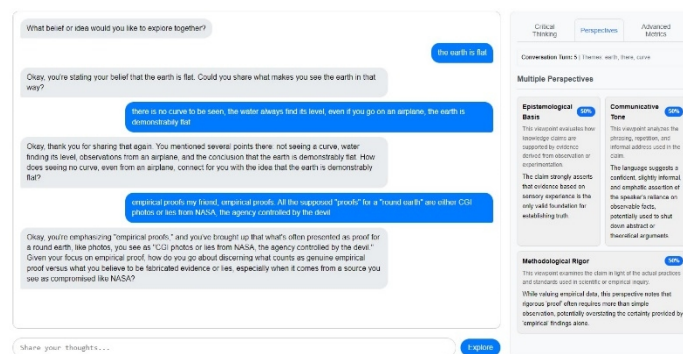


Figure 1: Sample of interaction, belief explorer prototype, 2025.

## The Epistemic Crisis: Navigating Amplified Narratives in a Post-Truth Landscape

The “post-truth” condition is not merely an absence of truth, but a state where truth itself becomes a contested and often secondary consideration in

public discourse. This environment is particularly fertile for populist agendas that rely on constructing a direct, unmediated relationship with “the people,” often by demonizing critical voices and established knowledge systems as elitist or corrupt (Mudde & Rovira Kaltwasser, 2017).

The ideal of free speech, essential for democratic deliberation, is paradoxically invoked in the post-truth era to shield the propagation of disinformation. What often results is not a vibrant marketplace of ideas, but a skewed information ecosystem where volume, repetition, and emotional resonance, amplified by algorithmic processes and coordinated inauthentic behavior, overshadow verifiable claims and nuanced perspectives. Recent studies indicate that false news stories reach 1,500 people six times quicker than accurate stories and are 70% more likely to be retweeted (Vosoughi et al., 2018). The capacity to “hear” in this digital cacophony is often determined by the power to amplify one’s voice, leaving reasoned but less amplified models of understanding marginalized. This dynamic is readily exploited by state and non-state actors seeking to manipulate public opinion (Benkler, Faris, & Roberts, 2018).

### **Sophisticated Disinformation and AI-Driven Cognitive Warfare**

The challenge posed by disinformation has evolved significantly. Beyond crude falsehoods, we now face highly sophisticated, contextually aware, and emotionally resonant narratives crafted to exploit cognitive biases. The emergence of advanced AI, particularly Large Language Models (LLMs), introduces a new vector for this “cognitive warfare.” Chatbots can be designed not only to generate convincing disinformation at scale but also to engage in seemingly authentic dialogues intended to subtly shift users’ perspectives, entrench them in echo chambers, or provoke unproductive conflict between artificially constructed “opposites,” thereby paralyzing constructive public conversation. Recent research demonstrates that AI-generated content is becoming increasingly indistinguishable from human-created content, with accuracy rates for detection dropping below 60% in many cases (Kreps & McCain, 2019). Traditional fact-checkers, often focused on discrete claims, are ill-equipped to counter these pervasive, dynamic, and psychologically targeted campaigns.

### **The Fragility of Truth-Centric Paradigms**

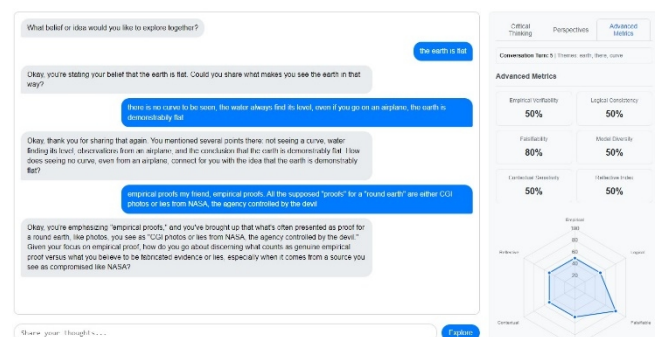
The inherent limitations of strictly truth-centric paradigms become stark in this context. Many manipulative narratives are not easily classifiable as simply “true” or “false.” They may involve selective presentation of facts, decontextualization, insinuation, or the strategic framing of issues to evoke specific emotional responses. Research on motivated reasoning indicates that individuals often evaluate evidence in a biased manner when it conflicts with their existing beliefs, leading to belief perseverance even in the face of contradictory evidence (Klayman & Ha, 1987). Clinging to a binary epistemological framework can be ineffective when the primary goal of a communication is not factual assertion but affective mobilization or cognitive disruption. This underscores the necessity for an epistemic framework that

can accommodate complexity, context-sensitivity, and the evaluation of claims beyond their immediate factual status.

### Belief Explorer: System Architecture and Implementation

Because all the aforementioned problems, we started to model a tool to avoid or minimize disinformation. Model Dependent Ontology (MDO), as articulated by Delaflor (2024), offers a solid epistemic framework that helps to improve our ability to discern what is valuable or factually right in a sea of cacophonies. MDO radically reorients our understanding of belief, knowledge, and existence by positing the primacy of models in all human cognition and experience. MDO's foundational theses include the primacy of models, where all human beliefs, perceptions, and experiences are constructed through and operate entirely within “models,” structured sets of rules, narratives, conceptual frameworks, and inferential processes.

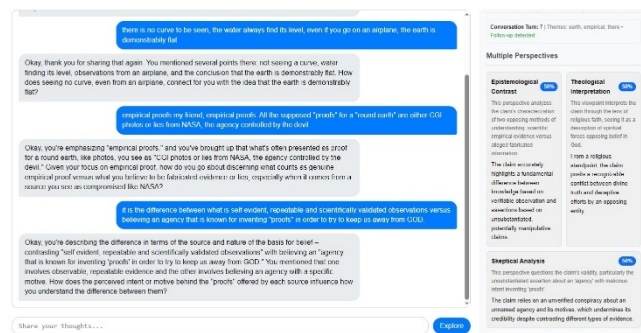
Belief Explorer is an AI-driven software framework conceived as a practical application of MDO principles, designed to help users navigate the complexities of online information. Its design philosophy, as articulated in its user-facing materials, emphasizes a non-judgmental, exploratory approach: it “doesn’t judge or force, it simply listens and gently reveals the underlying premises,” aiming to “provoke your own internal reasoning” like a “friend helping you see your path more clearly.” The Socratic Dialogue Engine is central to Belief Explorer, encouraging users to articulate and examine the internal coherence, assumptions, and implications of their own and presented information models. This aligns with MDO's emphasis on understanding models from within and evaluating their structure, rather than simply accepting or rejecting them based on external authority. The Socratic interaction aims to make the “model-ness” of beliefs explicit.



**Figure 2:** Sample of the advanced metrics, belief explorer prototype, 2025.

Multi-Perspective Analysis via LLM Arbiters processes each claim (an atomic information model) input into Belief Explorer, analyzing it in parallel by three specialized LLM arbiters that reflect different dimensions of pragmatic utility. The Empirical Arbiter assesses the claim's consistency with data from curated repositories and its observational coherence. From an MDO perspective, “observations” are themselves model-laden but serve as

crucial pragmatic checks on the utility of a claim-model. The Logical Arbiter examines the internal coherence, logical consistency, and presence of fallacies within the claim-model, addressing a key MDO criterion for model utility: internal consistency. The Pragmatic Arbiter directly embodies MDO's core evaluative principle by weighing potential outcomes, utility for specific goals, and situational appropriateness of accepting the claim-model.



**Figure 3:** Sample of the multiple perspectives analysis, belief explorer prototype, 2025.

The arbiters' analyses are synthesized into user-facing metrics visible in the UI mockups as Interpretable Metrics acting as MDO Indicators. The Verifact Score (Evidence Strength) is a composite measure reflecting the claim-model's empirical and logical support, indicating its immediate utility regarding consistency with observations and internal structure, displayed as a percentage to make it accessible to users without technical backgrounds. The Model Diversity Quotient (MDQ, measuring Inter-Arbiter Agreement) is particularly salient from an MDO viewpoint, quantifying the degree of agreement or divergence among the three arbiters, where a high MDQ (low agreement) signals that the claim-model is complex, contested, or viewed differently depending on the evaluative lens (empirical, logical, pragmatic), prompting deeper inquiry rather than premature judgment and highlighting the model-dependent nature of evaluation itself. The Contextual Sensitivity Index (CSI, Scenario Appropriateness) reflects MDO's emphasis on the context-dependency of a model's utility, recognizing that a claim-model useful in one context may be useless or even harmful in another. The Reflective Index (Exposed Assumptions) encourages users to identify and consider the unstated assumptions underpinning a claim-model, fostering a deeper understanding of its structure and potential vulnerabilities, a core aspect of MDO-based model evaluation.

Visual Analytics incorporate a radar chart visualization that provides an immediate visual representation of how a claim performs across all metrics, operationalizing MDO's multi-dimensional approach to model evaluation and allowing users to quickly grasp the strengths and weaknesses of different information models across various evaluative dimensions. The Perspective Generator actively crafts counter-arguments or alternative viewpoints to a user's stated belief or an analyzed claim, directly promoting epistemic

humility by demonstrating that any given model is one among many possibilities and encouraging users to consider the utility and structure of alternative models. The system generates these perspectives by drawing from diverse philosophical, ethical, and practical frameworks, reflecting MDO's pluralistic approach.

A Persistent Contextual Memory stores interaction history, ensuring that the exploration of information models and their utility is a coherent and cumulative process, allowing for iterative refinement of understanding across extended dialogues. This feature addresses the challenge of maintaining coherent reasoning chains across complex conversations, a key limitation in many current fact-checking approaches.

Belief Explorer aims to shift the user's cognitive approach from a reactive "Is this true?" to a more reflective and analytical "How useful, coherent, and contextually appropriate is this information model? What are its underlying assumptions? What are alternative models?" This MDO-inspired stance emphasizes critical engagement and the pragmatic evaluation of information as a tool for understanding and action, rather than the passive reception of supposed truths.

The system's design carefully balances sophistication with accessibility. While the underlying philosophy is complex, the interface presents information in digestible formats that encourage exploration without overwhelming users. The progression from basic claim analysis to multi-perspective exploration allows users to develop increasingly sophisticated epistemic capabilities over time. The value of an MDO-grounded tool like Belief Explorer extends beyond individual critical thinking, it addresses systemic issues within the current information ecosystem.

By equipping individuals to deconstruct and evaluate information models based on their intrinsic characteristics (coherence, empirical consistency, pragmatic utility, contextual relevance) rather than solely on the prominence or insistence of their source, Belief Explorer offers a cognitive toolkit to resist the "amplifier effect." Users are encouraged to scrutinize the "loudest voice" with the same rigor as any other.

Furthermore, by fostering an awareness of how information models are constructed and what makes them useful or misleading, the framework can build resilience against manipulative AI chatbots. Users become more adept at recognizing flawed logic, hidden agendas, or emotionally exploitative framing within the models presented by such entities. The Socratic dialogue, in particular, can help to dissect and neutralize the simplistic or polarizing models often pushed by weaponized AI.

## CONCLUSION

Belief Explorer, an AI framework inspired by MDO, seeks to translate its epistemic principles into a practical tool. Through its multi-perspective analysis, Socratic dialogue, and novel metrics, it aims to empower users to dissect, understand, and evaluate the information models they encounter, fostering resilience against disinformation and manipulation. The system represents not just a technological solution, but a cognitive and cultural intervention aimed at elevating the quality of public discourse.

In a world increasingly saturated with algorithmically amplified and often intentionally misleading content, the cultivation of an MDO-informed epistemic stance is not merely an academic exercise but a vital necessity for individual autonomy, informed public discourse, and the health of democratic societies. Moving beyond simplistic binaries and artificially imposed “opposites” towards a more nuanced, utility-focused engagement with information is essential for navigating the complexities ahead.

As we face unprecedented challenges to democratic deliberation and shared understanding, tools like Belief Explorer offer hope for a more epistemically resilient future. By operationalizing philosophical insights into practical technologies, we can help individuals develop the cognitive skills necessary to navigate an increasingly complex information landscape while maintaining the intellectual humility essential for productive dialogue across difference.

## REFERENCES

- Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press. <https://doi.org/10.1093/oso/9780190923624.001.0001>
- Bradshaw, S., & Howard, P. N. (2019). *The global disinformation order: 2019 global inventory of organised social media manipulation*. Project on Computational Propaganda, Oxford Internet Institute.
- DeLaflor, M. (2024). *Introduction to Model Dependent Ontology*. ResearchGate <https://doi.org/10.13140/RG.2.2.16809.20325>
- Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, 94(2), 211–228. <https://doi.org/10.1037/0033-295X.94.2.211>
- Kreps, S., & McCain, R. (2019). Not your father’s bots: AI is making fake news look real. *Foreign Affairs*, 98(2), 171–175.
- Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the “post-truth” era. *Journal of Applied Research in Memory and Cognition*, 6(4), 353–369. <https://doi.org/10.1016/j.jarmac.2017.07.008>
- Mudde, C., & Rovira Kaltwasser, C. (2017). *Populism: A very short introduction*. Oxford University Press. <https://doi.org/10.1093/acrade/9780190234874.001.0001>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>
- Waisbord, S. (2018). The elective affinity between post-truth communication and populist politics. *Communication Research and Practice*, 4(1), 17–34. <https://doi.org/10.1080/22041451.2018.1428928>