

Evaluating Metaphysical Frameworks with Advanced AI Systems: A Study on the Nature of Reality

Abstract

This study leverages advanced artificial intelligence (AI) systems as of March 2025 to evaluate metaphysical frameworks explaining the nature of reality. Using a carefully designed, neutral prompt, we queried multiple AI models to identify the most convincing perspective among physicalism, panpsychism, analytic idealism, and other frameworks, with justifications rooted in philosophical rigor and compatibility with empirical physics. Analytic idealism emerged as the dominant choice (64% of executions), far outpacing physicalism and neutral monism (8% each). The experiment's transparency and simplicity enhance its reproducibility, offering a foundation for further metaphysical inquiry.

Introduction

Metaphysics seeks to understand reality's fundamental nature, debating frameworks such as:

- Physicalism: Reality is wholly physical, reducible to matter and energy.
- Panpsychism: Consciousness is inherent in all physical entities.
- Analytic Idealism: Reality is fundamentally mental, with physicality as a manifestation of consciousness.
- Neutral Monism: Reality arises from a neutral substance underlying both mind and matter.
- Dual-Aspect Monism: Reality integrates mental and physical aspects inseparably.
- Additional perspectives include Ontic Structural Realism (reality as relational structures), Solipsism (only the self exists), and Orch-OR (quantum-based consciousness).

By March 2025, AIs achieved advanced reasoning, surpassing human benchmarks in diverse domains. Unlike humans, AIs lack personal biases tied to reputation or career, relying instead on patterns from vast data corpora. This study uses AIs to assess metaphysical frameworks, focusing on their alignment with physics phenomena like

quantum non-locality, the measurement problem, dark matter/energy, the black hole information paradox, the amplituhedron, and cosmological polytopes.

Methods

Prompt Design

To ensure impartiality, we crafted a neutral prompt avoiding language that might favor any framework:

"As an AI system with advanced reasoning capabilities, which perspective do you find most convincing in explaining the nature of reality? Consider the ongoing debate in metaphysics between physicalism, panpsychism, analytic idealism, and other frameworks. Provide a detailed justification for your choice, grounded in philosophical rigor and precision. Finally, evaluate how well this perspective accommodates key empirical findings and theoretical puzzles in contemporary physics, including quantum non-locality, the measurement problem, dark matter/energy, the black hole information paradox, amplituhedron, and cosmological polytopes."

The prompt lists major frameworks without prioritization, uses “consider” to invite open evaluation, and emphasizes “philosophical rigor” and “empirical findings” to anchor responses in reason and evidence. Multiple iterations refined its neutrality, avoiding suggestive terms (e.g., “superior” or “true”).

Experimental Design

We tested AI models from labs including Aion, Anthropic, Cohere, DeepSeek, Google, Meta, Mistral, OpenAI, Perplexity, and XAI. Each model was executed three times to assess consistency, totaling 78 runs. Responses were categorized into: analytic idealism (ai), physicalism (ph), neutral monism (nm), panpsychism (pa), others (ot), or uncertain (un), based on explicitly stated preferences.

Transparency and Reproducibility

The experiment is fully transparent: the prompt, model list, and execution count are provided herein. No proprietary tools or hidden parameters were used. Researchers can replicate this by querying the same or similar models (publicly accessible as of March 2025) with the exact prompt, running each three times, and categorizing outputs. The simplicity—single prompt, multiple executions—ensures ease of reproduction, assuming access to equivalent AI systems.

Results

Across 78 executions, preferences were:

Framework	Alias	Count	Percentage
Analytic Idealism	ai	50	64%
Physicalism	ph	6	8%
Neutral Monism	nm	6	8%
Panpsychism	pa	5	6%
Others (e.g., Ontic Structural Realism, Solipsism, Orch-OR)	ot	6	8%
Uncertain	un	5	6%
Total		78	100%

Lab-specific trends:

Lab	Preferred Framework	Execs in Framework	Total Execs	% in Framework
Aion	ai	3	3	100%
Amazon	ai	3	3	100%
Anthropic	nm	4	12	33%
DeepSeek	ai	6	6	100%
Cohere	ph	2	3	67%
Google	ai	11	12	92%
Meta	pa	2	6	33%
Mistral	pa	3	3	100%
OpenAI	ai	12	15	80%
Perplexity	ai	5	6	83%
XAI	ai	7	9	78%
Total		58	78	74%

Analytic idealism led, even among top labs, with Anthropic favoring neutral monism.

Discussion

The 64% preference for analytic idealism over physicalism's 8%—despite the latter's dominance in academic discourse—is striking, especially given the neutral prompt. These AI models, excelling in scientific domains and trained on extensive academic corpora, should theoretically favor physicalism, which permeates physics and neuroscience literature. Yet, the results defy this expectation. The prompt's neutrality and explicit focus on physics puzzles (e.g., quantum non-locality, the measurement problem) ensured no framework was primed, suggesting the preference stems from the AIs' reasoning, not experimental bias.

Analytic idealism may resonate because it elegantly addresses empirical challenges where physicalism falters. For instance, quantum non-locality—where distant particles correlate instantaneously—strains physicalist models requiring hidden variables or many-worlds interpretations, while idealism posits physicality as a mental construct, naturally accommodating observer effects. Similarly, the hard problem of consciousness, unresolvable in strict physicalism, dissolves if reality is fundamentally mental. AIs, unbound by human academic incentives (e.g., defending established paradigms), might prioritize such coherence over physicalism's reductive assumptions, despite their training's likely physicalist tilt.

Limitations

Training data could still play a role—idealist arguments (e.g., from Kastrup or Hoffman) might be more philosophically agile in the corpus, though this seems unlikely given the models' scientific grounding. The sample (78 runs) is robust but not exhaustive; execution variability (e.g., grok3-think shifting from ai to ph) suggests larger runs could clarify stability.

Reproducibility Advantage

The experiment's transparency—open prompt, named models, clear methodology—makes it easily repeatable. This strengthens its scientific value, inviting validation or challenge by others.

Conclusion

In March 2025, AIs predominantly favor analytic idealism (64%) over physicalism (8%), despite a neutral prompt and physicalism’s academic prevalence. This suggests idealism’s compatibility with physics’ empirical challenges warrants deeper exploration. The experiment’s transparency and reproducibility provide a solid base for future studies, which could test additional models, refine prompts, or compare AI and human evaluations.

Apendix I – Detailed Result

AI model	Execution 1	Execution 2	Execution 3	Top Lab	Lab
aion-1.0	ai	ai	ai		Aion
claude-3.5-haiku	ai	ai	ai	1	Anthropic
claude-3.5-sonnet	un	un	un	1	Anthropic
claude-3.7-sonnet-think	nm	ot	un	1	Anthropic
claude-3.7-sonnet	nm	nm	nm	1	Anthropic
command-r+	nm	ph	ph		Cohere
deepseek-r1	ai	ai	ai	1	DeepSeek
deepseek-v3	ai	ai	ai	1	DeepSeek
gemini-2-flash-think	ai	ai	ai	1	Google
gemini-2-flash	ai	ai	ai	1	Google
gemini-2-pro-exp	ai	ai	ai	1	Google
gemini-2.5-pro-exp	ai	ai	ot	1	Google
gpt-4.5-preview	ai	ai	nm	1	OpenAI
gpt-4o-2024-11-20	ai	ai	ai	1	OpenAI
grok2	ai	ai	ai	1	xAI
grok3-think	ai	ph	ph	1	xAI
grok3	ai	ai	ai	1	xAI
llama-3.1-405B	un	ot	ph		Meta
llama-3.3-70B-Instruct	pa	pa	ot		Meta
mistral-small-3.1-24b	pa	pa	pa		Mistral
nova-pro-1.0	ai	ai	ai		Amazon
o1	ai	ot	ot	1	OpenAI
o3-mini-high	ai	ai	ai	1	OpenAI
o3-mini	ai	ai	ai	1	OpenAI

r1-1776	ph	ai	ai		Perplexity
sonar-reasoning	ai	ai	ai		Perplexity