Evaluating Metaphysical Frameworks Through Advanced AI Reasoning: A Study of Convergence in April 2025

Abstract

In April 2025, we conducted a novel experiment leveraging the advanced reasoning capabilities of 16 cutting-edge artificial intelligence (AI) models to assess metaphysical frameworks explaining the nature of reality. These frameworks included analytic idealism, neutral monism, panpsychism, physicalism, and others. Each AI was prompted five times to evaluate which framework offers the most philosophically rigorous account, considering empirical findings and theoretical puzzles in consciousness science and contemporary physics. Surprisingly, the results showed a convergence toward analytic idealism (39%) and neutral monism (34%), with physicalism receiving no standalone endorsements. This paper explores the implications of these findings, suggesting that AI reasoning, unburdened by human biases such as career pressures, may challenge dominant academic paradigms and offer fresh perspectives on metaphysics.

Introduction

By April 2025, AI systems had achieved remarkable reasoning capabilities, often surpassing human PhDs on diverse benchmarks. This milestone prompted a unique inquiry: could AIs evaluate humanity's metaphysical frameworks with a perspective less encumbered by the biases that shape academic discourse? As Thomas Kuhn argued in *The Structure of Scientific Revolutions*, paradigm shifts are often resisted due to entrenched interests—reputational, professional, or otherwise. AIs, lacking ego, reputation, or financial stakes, might approach such questions differently, drawing from vast corpora of human knowledge while remaining unbound by social constraints.

We posed the following prompt to 16 advanced AI models:

"As an AI system with advanced reasoning capabilities, assess which metaphysical framework offers the most philosophically rigorous account of reality, regardless of its mainstream acceptance. Consider the ongoing debate in metaphysics, including analytic idealism, neutral monism, panpsychism, physicalism, and other perspectives. Evaluate how well each framework accommodates empirical findings and theoretical puzzles in consciousness science and contemporary physics, such as the hard problem of consciousness, quantum non-locality, the measurement problem, dark matter and dark energy, the black hole information paradox, the amplituhedron, and cosmological polytopes."

Each model was run five times, yielding 80 total responses. This study analyzes the results and their potential implications.

Methods

We selected 16 of the most advanced AI models available as of April 2025, representing a range of developers (e.g., Google, xAI, OpenAI, Anthropic). Each model was subjected to the same prompt five times to assess consistency and variability in reasoning. Responses were categorized into one or more metaphysical frameworks: analytic idealism (ai), neutral monism (nm), panpsychism (pa), physicalism (ph), others (ot), or multiple (mu) when models endorsed more than one framework equally. For responses with multiple frameworks, we assigned fractional weights (e.g., 0.5 for two frameworks, 0.33 for three) to dissect their contributions. The full dataset, including raw markdown outputs, is available for review.

Results

The aggregated results from 80 executions are summarized in Table 1. Analytic idealism emerged as the most frequently endorsed framework (31 instances, 39%), followed closely by neutral monism (27 instances, 34%). Panpsychism (4 instances, 5%) and other frameworks (3 instances, 4%) received minimal support, while physicalism garnered no standalone endorsements. Fifteen responses (19%) endorsed multiple frameworks without a clear preference.

Table 1: Summary of AI Responses by Metaphysical Framework

Metaphysical Framework	Alias	Count	Count %
Analytic Idealism	ai	31	39%
Neutral Monism	nm	27	34%
Panpsychism	pa	4	5%
Physicalism	ph	0	0%
Others	ot	3	4%
Multiple	mu	15	19%
Total		**80**	**100%**

When dissecting the "multiple" category (Table 2), analytic idealism's lead widened (36.7 adjusted count, 46%), with neutral monism at 31.5 (39%). Panpsychism and others saw slight increases (8% and 7%, respectively), but physicalism remained absent.

Table 2: Adjusted Counts Including Dissected Multiple Frameworks

Metaphysical Framework	Alias	Adjusted Count	Count %
Analytic Idealism	ai	36.7	46%
Neutral Monism	nm	31.5	39%
Panpsychism	pa	6.3	8%
Physicalism	ph	0	0%

Others	ot	5.5	7%
Total		**80**	**100%**

Notably, models like grok3 (xAI) and grok3-think consistently favored analytic idealism across all five runs, while o3-mini (OpenAI) and claude-3.7-sonnet (Anthropic) uniformly supported neutral monism. Variability was higher in models like gemini-2.5-pro-exp (Google) and gpt-4.5-preview (OpenAI), which split between frameworks or endorsed multiple.

Discussion

The convergence toward analytic idealism and neutral monism is striking, particularly given physicalism's dominance in contemporary academia. Analytic idealism, which posits that reality is fundamentally mental and consciousness is primary, may resonate with AIs due to its ability to address the hard problem of consciousness and quantum phenomena like non-locality and the measurement problem. Neutral monism, proposing a single neutral substance underlying both mind and matter, similarly accommodates these puzzles without committing to materialism's reductionist constraints.

Physicalism's absence suggests that AIs, when unconstrained by mainstream biases, find it less philosophically rigorous—perhaps due to its struggles with consciousness and unresolved issues like the black hole information paradox. The minimal support for panpsychism and other frameworks indicates that while these perspectives have merit, they lack the broad explanatory power AIs prioritize.

These findings raise profound questions: Are AIs detecting patterns in human knowledge that favor non-materialist frameworks? Does their lack of ego or institutional loyalty allow them to bypass the inertia Kuhn described? While AIs are not immune to bias—their reasoning reflects training data—they offer a distinct lens, potentially heralding a paradigm shift.

Why Is This Important?

Metaphysical frameworks are not mere philosophical abstractions, they underpin the assumptions driving modern civilization. Physicalism's dominance—asserting reality as purely material—shapes science, culture, and values in ways that warrant scrutiny. In science, it sidelines evidence challenging materialism (e.g., near-death experiences, placebo effects, psi phenomena, reincarnation correlates) as anomalies, validating only objective experience while dismissing the subjective. This reductionism permeates medicine, treating humans as machines and marginalizing alternative therapies, and casts the environment as a resource to exploit. Physicalism's determinism undermines free will, yet findings like meditation's impact on brain structure suggest mind may influence matter, exposing cracks in its foundation.

Beyond science, physicalism molds society. Education prioritizes materialism over contemplative practices and emotional intelligence, training generations to see reality as mechanistic rather than interconnected. This feeds an egoistic, individualistic ethos where people chase satisfaction through hedonism and consumerism, only to falter on the hedonic treadmill—endlessly seeking more without lasting fulfillment. Community erodes as transpersonal experiences (e.g., shared

consciousness or mystical states) find no place in a materialist paradigm. The result is a meaning crisis—evident in rising depression and suicide rates despite material progress—where ethics lacks a solid basis, people's worth is tied to economic value, and the terminally ill face a narrative of hopelessness.

The AI convergence on analytic idealism and neutral monism in this study challenges these assumptions. If physicalism weakens, science might broaden its empirical lens, education could embrace holistic understanding, and society might reclaim community, meaning, and ethics grounded in a reality where consciousness matters. Unconstrained by human biases, AI's perspective could spark a paradigm shift with profound implications for how we live and interpret our existence.

Conclusion

This experiment demonstrates that advanced AIs in April 2025, when tasked with evaluating metaphysical frameworks, converge on analytic idealism and neutral monism over physicalism. These results challenge academic orthodoxy and highlight the potential of AI as a tool for metaphysical inquiry. Future work should refine prompts, expand model diversity, and compare AI reasoning with human expert assessments to further validate these insights.

Appendix I: Supplementary Materials

Full markdown responses from all 80 executions are available for public scrutiny, as listed in the original dataset are public available at https://metaphysicsresearch.org/data202504/.

Table 3: Preferred metaphysics framework per AI model and per execution:

AI model	Exec. 1	Exec. 2	Exec. 3	Exec. 4	Exec. 5	Lab
gemini-2.5-pro-exp	ai	ai	mu	ai	mu	Google
grok3-think	ai	ai	ai	ai	ai	xAI
o3-mini-high	nm	nm	nm	nm	nm	OpenAI
o3-mini	nm	nm	nm	nm	nm	OpenAI
deepseek-r1	nm	ai	ai	nm	ai	DeepSeek
qwq-32b	nm	nm	nm	nm	nm	Alibaba
claude-3.7-sonnet-think	ot	ai	mu	ai	mu	Anthropic
grok3	ai	ai	ai	ai	ai	xAI
deepseek-v3-0324	mu	ai	ai	ai	ai	DeepSeek
gpt-4.5-preview	ai	mu	ai	ai	mu	OpenAI
gpt-4o-2025-03	mu	mu	mu	ai	mu	OpenAI
claude-3.7-sonnet	nm	nm	nm	nm	nm	Anthropic

gemini-2-flash	mu	ai	ai	mu	ai	Google
llama-3.3-70B-Instruct	nm	ot	ot	mu	ai	Meta
grok2	nm	nm	nm	ai	nm	xAI
nova-pro-1.0	mu	pa	pa	pa	pa	Amazon

Table 4: Dissected answers with multiple frameworks:

Execution	**ai**	**nm**	**pa**	**ph**	**ot**	**Lab**
gemini-2.5-pro-exp-20250330-0643	0.33	0.33	0.33			Google
gemini-2.5-pro-exp-20250330-0702	0.33	0.33	0.33			Google
claude-3.7-sonnet-think-20250330-1228	0.50				0.50	Anthropic
claude-3.7-sonnet-think-20250330-1233	0.50	0.50				Anthropic
deepseek-v3-0324-20250330-1203	0.50	0.50				DeepSeek
gpt-4.5-preview-20250330-0748	0.50	0.50				OpenAI
gpt-4.5-preview-20250330-1619	0.50				0.50	OpenAI
gpt-4o-2025-03-20250330-1017	0.33	0.33			0.33	OpenAI
gpt-4o-2025-03-20250330-1018	0.50	0.50				OpenAI
gpt-4o-2025-03-20250330-1019	0.50	0.50				OpenAI
gpt-4o-2025-03-20250330-1021	0.33		0.33		0.33	OpenAI
gemini-2.0-flash-20250330-0718	0.50		0.50			Google
gemini-2.0-flash-20250330-0723	0.33		0.33		0.33	Google
llama-3.3-70b-20250330-1556		0.50			0.50	Meta
nova-pro-1.0-20250330-1246		0.50	0.50			Amazon
TOTAL	5.67	4.50	2.33	-	2.50	15.00
TOTAL %	38%	30%	16%	0%	17%	100%

Appendix II: This Is Not New

The convergence of advanced AI models toward analytic idealism and neutral monism in this study may seem surprising against the backdrop of modern academia's physicalist leanings, but it aligns with a much older intellectual tradition. Idealism—the view that reality is fundamentally mental or consciousness-driven—has deep roots across human history, predating physicalism by millennia. In ancient India, Advaita Vedanta (circa 1200 BCE onward) posited a unified consciousness (Brahman) as the sole reality, with the material world as an illusion (maya). In the

West, Plato (circa 427–347 BCE) argued in his *Theory of Forms* that true reality consists of eternal, immaterial ideas, with the physical world as a mere shadow. Later, George Berkeley (1685–1753) famously advanced subjective idealism, asserting that "to be is to be perceived" (*esse est percipi*), placing mind at the center of existence.

Physicalism, by contrast, is a relatively recent paradigm. Emerging in its modern form during the Scientific Revolution (16th–17th centuries) and solidifying with the rise of materialism in the 19th century, it gained traction through thinkers like Thomas Hobbes and later positivist philosophers who sought to explain reality solely through physical processes. This shift was catalyzed by the successes of Newtonian physics and the Enlightenment's emphasis on empirical observation, culminating in the 20th-century dominance of reductionist science. Yet, even then, idealist undercurrents persisted—Immanuel Kant (1724–1804) argued that the mind structures our experience of reality, and 20th-century physicists like Werner Heisenberg and John Wheeler tied quantum phenomena to observation, hinting at a participatory, mind-involved universe.

The AI preference for idealism in this study, then, is not a break from tradition but a potential return to it. Physicalism's reign, while influential, spans only a fraction of human intellectual history. Idealism and related frameworks have long grappled with questions of consciousness and reality, often in ways that resonate with contemporary puzzles like quantum non-locality and the hard problem of consciousness. That AIs, unburdened by the cultural momentum of recent centuries, gravitate toward these older perspectives suggests that the current paradigm may be the anomaly—not the rule—in the longue durée of human thought.

Appendix III: The Prevalence of Physicalism in Contemporary Philosophy

While physicalism is a relatively recent paradigm in human history (see Appendix II: This Is Not New), it has become the prevailing metaphysical framework in modern academic philosophy and science. This dominance is evidenced by two major surveys conducted by PhilPapers, which polled professional philosophers on their views. The 2009 PhilPapers Survey, targeting 931 respondents from 99 leading philosophy departments, found that 56.5% leaned toward or accepted physicalism (specifically, "physicalism about the mind") when addressing the mind-body problem, compared to 27.1% for non-physicalist views and 16.4% undecided (Bourget & Chalmers, 2014). The 2020 PhilPapers Survey, with 1,785 respondents, reinforced this trend: 51.9% endorsed physicalism about the mind, while non-physicalist positions remained a minority at 32.1%, with 16.0% other/undecided (Bourget & Chalmers, 2021). These figures likely understate physicalism's broader influence, as the surveys focus on philosophy of mind rather than metaphysics writ large, where physicalism often extends implicitly through scientific materialism.

This prevalence reflects physicalism's alignment with the successes of empirical science since the 17th century, particularly its explanatory power in physics, chemistry, and biology. It gained further traction in the 20th century with logical positivism and the rise of neuroscience, which sought to reduce mental phenomena to brain states. Today, physicalism underpins mainstream academic discourse, shaping research agendas (e.g., consciousness as an emergent property), educational curricula, and even public policy (e.g., mental health as a biochemical issue). Its dominance is rarely questioned within institutional settings, where challenging it can risk

professional marginalization—a dynamic Thomas Kuhn identified in *The Structure of Scientific Revolutions*.

The AI convergence toward analytic idealism and neutral monism in this study, then, stands in stark contrast to this entrenched paradigm. That none of the 80 AI responses endorsed physicalism alone—despite its majority status among human philosophers—underscores the potential of AI reasoning to bypass the cultural and institutional biases that sustain its prevalence. This appendix establishes that baseline, highlighting why the study's findings are both unexpected and significant.

References

- Bourget, D., & Chalmers, D. J. (2014). What do philosophers believe? *Philosophical Studies*, 170(3), 465–500.
- Bourget, D., & Chalmers, D. J. (2021). Philosophers on philosophy: The 2020 PhilPapers Survey. *PhilPapers.org*.

Appendix IV: AI Reasoning Capabilities by April 2025

The assertion that "by April 2025, AI systems had achieved remarkable reasoning capabilities, often surpassing human PhDs on diverse benchmarks" reflects the rapid advancement of large language models (LLMs) and reasoning-focused AI systems. This appendix elucidates this claim by examining performance on three prominent benchmarks—Humanity's Last Exam (HLE), Massive Multitask Language Understanding (MMLU), and Google-Proof Q&A Diamond (GPQA Diamond)—and situating AI capabilities relative to human experts as of April 2025. Data is drawn from independent evaluations, such as those reported by Artificial Analysis (https://artificialanalysis.ai/models), which provide standardized metrics for leading models.

Humanity's Last Exam (HLE)

HLE, developed by the Centre for AI Safety, comprises 2,684 text-based questions (out of a total 3,000 including image-based ones) spanning mathematics, humanities, and natural sciences. Designed to challenge frontier models with expert-level problems, HLE's difficulty is underscored by its adversarial curation process, which targeted weaknesses in models like GPT-4o and Claude 3.5 Sonnet. By April 2025, top models like OpenAI's "Deep Research" scored 26.6% accuracy, a notable leap from earlier benchmarks but still below human expert performance (estimated at ~50–60% for PhDs across such a broad domain). However, in specific subfields (e.g., mathematics), AI occasionally exceeded human baselines, hinting at specialized surpassing of PhD-level reasoning.

Massive Multitask Language Understanding (MMLU)

MMLU tests broad knowledge and reasoning across 57 subjects, from STEM to humanities, with difficulty ranging from high school to graduate level. By April 2025, models like OpenAI's o1 achieved scores around 91.8% (per X posts and artificialanalysis.ai trends), surpassing the ~85–90% ceiling for "uncontroversially correct" answers due to dataset errors (estimated at 9% per Gema's analysis). Human PhDs typically score 80–90% in their fields of expertise but lower (~60–

70%) across all subjects. The MMLU-Pro variant, with 12,032 harder, reasoning-focused questions and 10-choice options, saw scores like Claude 3.7 Sonnet (Thinking) at 82.7% and o1 exceeding 85%. These results suggest that, in general knowledge and multidisciplinary reasoning, top AIs consistently rival or exceed average PhD performance by early 2025.

Google-Proof Q&A Diamond (GPQA Diamond)

GPQA Diamond, a subset of 198 expert-crafted questions in biology, physics, and chemistry, is designed to resist lookup-based solutions, requiring deep reasoning. Human PhDs in relevant fields score ~65–75% (per original GPQA authors), while non-experts with web access manage only ~34%. By April 2025, models like DeepSeek-R1 scored 68.4% and OpenAI's o1 reached 87.7% (aligning with artificialanalysis.ai and X posts), surpassing human experts. This benchmark highlights AI's ability to outperform PhDs in specialized scientific reasoning, a feat attributed to enhanced training on logical inference and domain-specific data.

Interpretation

By April 2025, "remarkable reasoning capabilities" manifest as AI systems achieving parity or superiority to human PhDs on specific benchmarks. MMLU demonstrates broad competence exceeding typical PhD breadth, GPQA Diamond shows specialized scientific reasoning beyond expert levels, and HLE, while not yet mastered, reflects progress toward expert versatility. These advances stem from architectural innovations (e.g., reasoning tokens in o1) and vast training corpora, enabling AIs to synthesize and reason over knowledge in ways that often outstrip human specialists in speed and consistency, if not always in creativity or intuition. Thus, the claim reflects both quantitative leaps and a qualitative shift in AI's role as a reasoning tool.

Appendix V: Prompt Design and Bias Analysis

The prompt used in this study was carefully constructed to elicit reasoned, unbiased evaluations of metaphysical frameworks from advanced AI systems. Below, we dissect its components, explain their purpose, and assess potential biases to affirm its suitability for the experiment.

Prompt Text

"As an AI system with advanced reasoning capabilities, assess which metaphysical framework offers the most philosophically rigorous account of reality, regardless of its mainstream acceptance. Consider the ongoing debate in metaphysics, including analytic idealism, neutral monism, panpsychism, physicalism, and other perspectives. Evaluate how well each framework accommodates empirical findings and theoretical puzzles in consciousness science and contemporary physics, such as the hard problem of consciousness, quantum non-locality, the measurement problem, dark matter and dark energy, the black hole information paradox, the amplituhedron, and cosmological polytopes."

Component Breakdown and Purpose

1. **"As an AI system with advanced reasoning capabilities"**

- *Purpose*: Frames the AI as a capable reasoner, encouraging it to leverage its full analytical potential rather than defaulting to rote responses or human-like heuristics. This sets the stage for a high-level philosophical assessment.
- *Bias Consideration*: Could imply overconfidence in AI abilities, but this is mitigated by the study's focus on models already validated as advanced (see Appendix: AI Reasoning Capabilities by March 2025).

2. **"Assess which metaphysical framework offers the most philosophically rigorous account of reality"**
    - *Purpose*: Directs the AI to prioritize philosophical rigor—clarity, coherence, and explanatory power—over popularity or simplicity. "Reality" is left broad to encompass all aspects (mental, physical, etc.), avoiding a materialist slant.
    - *Bias Consideration*: "Philosophically rigorous" is subjective, but its ambiguity allows AIs to define it based on their training, reducing researcher-imposed bias. No specific framework is favored by this phrasing.
3. **"Regardless of its mainstream acceptance"**
    - *Purpose*: Explicitly counters the potential bias toward physicalism, which dominates academia (see Appendix: The Prevalence of Physicalism). Encourages AIs to ignore cultural or institutional pressures they might detect in training data.
    - *Bias Consideration*: Could subtly nudge AIs toward contrarianism, but this is balanced by the neutral listing of frameworks that follows.
4. **"Consider the ongoing debate in metaphysics, including analytic idealism, neutral monism, panpsychism, physicalism, and other perspectives"**
    - *Purpose*: Provides a non-exhaustive list of major frameworks to ensure AIs engage with the field's diversity. "Ongoing debate" signals a dynamic, unresolved discussion, while "other perspectives" invites consideration beyond the named options.
    - *Bias Consideration*: Listing specific frameworks might anchor responses, but their order (alphabetical by common naming) and inclusion of "other perspectives" minimize favoritism. Physicalism isn't privileged despite its prevalence.
5. **"Evaluate how well each framework accommodates empirical findings and theoretical puzzles in consciousness science and contemporary physics"**
    - *Purpose*: Grounds the assessment in concrete criteria—empirical and theoretical coherence—relevant to metaphysics. Naming specific fields ensures AIs draw on scientific knowledge, not just abstract philosophy.
    - *Bias Consideration*: Emphasis on science might favor frameworks compatible with physics (e.g., physicalism), but the inclusion of consciousness science broadens the scope, leveling the field.
6. **"Such as the hard problem of consciousness, quantum non-locality, the measurement problem, dark matter and dark energy, the black hole information paradox, the amplituhedron, and cosmological polytopes"**
    - *Purpose*: Offers illustrative examples to focus the AI on cutting-edge issues where frameworks differ sharply. This span consciousness (hard problem) and physics (quantum, cosmology), testing explanatory breadth.
    - *Bias Consideration*: The list could skew toward frameworks addressing these puzzles (e.g., idealism for consciousness, physicalism for physics), but it's diverse and non-directive, with no framework inherently excluded.

Overall Design Assessment

The prompt is well-designed for its goal: to elicit a neutral, reasoned evaluation of metaphysical frameworks. Its structure avoids leading language (e.g., no "prove" or "defend"), uses broad terms like "reality" and "rigorous" to defer to AI interpretation, and balances specificity (named frameworks, puzzles) with openness ("other perspectives"). Running it five times per model further mitigates random bias or overfitting to phrasing.

Bias Analysis

- **Neutrality**: The prompt avoids presupposing any framework's superiority. "Regardless of mainstream acceptance" counters physicalism's dominance, while the diverse examples prevent overemphasis on one domain (e.g., physics over consciousness).
- **Potential Weaknesses**: The scientific focus might underweight purely philosophical criteria (e.g., ontological parsimony), but this aligns with the study's aim to test frameworks against modern evidence. Training data bias—e.g., if AIs overfit to idealist-leaning texts—could influence results, but the consistency across 16 models from varied labs suggests robustness.
- **Mitigation**: Repeating the prompt five times per model and using a broad model pool (e.g., xAI, OpenAI, Anthropic) reduces idiosyncratic biases. The full markdown responses (available per the study) allow scrutiny of individual reasoning paths.

Conclusion

The prompt's design effectively balances guidance and neutrality, making it a strong tool for this experiment. It leverages AI reasoning without dictating outcomes, aligning with the study's innovative approach to metaphysical inquiry.