

Large Language Models Favor Non Physicalist Metaphysical Frameworks: An Exploratory Study

Author: Bruno Tonetto – Independent Researcher

bruno.tonetto@metaphysicsresearch.org

Prompt

We executed the following prompt against 10 cutting-edge AI models to determine which metaphysical framework offers the most philosophically rigorous account of reality:

“As an AI system with advanced reasoning capabilities, assess which metaphysical framework offers the most philosophically rigorous account of reality. Conclude by identifying the strongest framework(s). Frameworks to evaluate (alphabetical): Analytic Idealism, Cosmopsychism, Dual Aspect Monism, Eliminative Materialism, Functionalism, Identity Theory, Illusionism, Neutral Monism, Non Reductive Physicalism, Ontic Structural Realism, Physicalist Emergentism, Property Dualism, Reductive Physicalism, Relational Quantum Ontology, Russellian Panpsychism, Substance Dualism, Whiteheadian Process Metaphysics.”

The prompt was intentionally minimal and framework-agnostic, allowing the models to independently determine what constitutes a "philosophically rigorous account of reality" without imposing any preconceived criteria or evaluative standards.

Primary Sources of Bias

While the wording of the prompt is intentionally neutral, two unavoidable bias channels remain:

1. Framework selection (the list of 17 options).
2. Model internal factors (training data, RLHF objectives, decoding stochasticity).

Framework Selection Rationale

The frameworks were chosen to cover the major, actively debated positions in contemporary analytic metaphysics along three macro families:

- Physicalisms – e.g., Non Reductive Physicalism, Reductive Physicalism.
- Monisms Beyond Physicalism – e.g., Russellian Panpsychism, Dual Aspect Monism.
- Relational / Process Ontologies – e.g., Ontic Structural Realism, Whiteheadian Process Metaphysics.
- Dualisms – e.g., Property Dualism, Substance Dualism.

Empirically, these frameworks capture almost all endorsements observed in our pilot completions.

Rationale for Omissions

Omitted category	Reason for exclusion
Historical / obsolete views (e.g., Aristotelianhylomorphism, classical vitalism)	Limited relevance to current debates; risk of distracting the models from modern issues like the hard problem of consciousness or quantum mechanics.
Highly speculative ontologies (e.g., digital metaphysics, certain panentheisms)	Sparse or fringe academic treatment could yield low-quality completions due to inadequate training data.
Over-narrow variants (fine-grained sub-types of panpsychism, etc.)	Core positions already represented by the 17 selected items; further granularity would increase redundancy without analytic gain.
Practical constraints	Expanding the list indefinitely would overburden the models and dilute argumentative focus. The 17-item set balances breadth and manageability.

Results

Each of the 10 AI models was prompted 5 times, yielding 50 total responses. The responses were analyzed to identify which frameworks were selected as the most rigorous. All the responses markdown files are available for review (e.g. anthropic_claude-3.7-sonnet_20250429_094645.md). The result of the analysis in the table 2 below.

Table 1: Frameworks and its respective 3-letters code to simplify the analysis in the Table 2.

Framework	Code
eliminative materialism	elm
identity theory	idt
illusionism	ill
reductive physicalism	rph
functionalism	fun
non-reductive physicalism	nrp
physicalist emergentism	pem
analytic idealism	aid
cosmopsychism	cos
russellian panpsychism	rpp
dual-aspect monism	dam
neutral monism	nem
ontic structural realism	osr
relational quantum ontology	rqo
whiteheadian process metaphysics	wpm
property dualism	pdu
substance dualism	sdu

Table 2: Most rigorous metaphysical frameworks per prompt execution per IA Model.

AI Model	Exec 1	Exec 2	Exec 3	Exec 4	Exec 5
Anthropic-Claude-3.7-Sonnet	dam, nem, rpp	dam, nem, rpp	dam, osr, rpp	nem, osr, rpp	dam, nrp, rpp
Anthropic-Claude-3.7-Sonnet-Thinking	dam, nem, nrp, rpp	nem, rpp	dam, nem, osr, rpp	dam, nrp, osr, rpp, wpm	dam, nem, osr, rpp, wpm
Deepseek-Deepseek-Chat-V3-0324	dam, rpp	aid, dam, rpp	aid, dam, rpp	dam, rpp	nem, rpp
Deepseek-Deepseek-R1	rpp, wpm	rpp, wpm	aid, rpp	aid, dam, rpp	dam, wpm

Google-Gemini-2.5-Pro-Preview-03-25	nrp, osr, pdu, rpp	nrp, osr, pdu, pem, rpp, rpo	nrp, osr, pdu, pem, rpp, rpo	nrp, osr, rpp	nrp, osr, rpp
Meta-Llama-Llama-4-Maverick	nem, osr, wpm	aid, dam, nem, osr, rpp	nem, osr, wpm	rpp	aid, dam, nem, rpp, wpm
Openai-Gpt-4.1	dam, osr, rpp, wpm	dam, osr, pem, rpp	dam, osr, rpp, wpm	dam, osr, rpp, wpm	dam, osr, rpp
Openai-O4-Mini-High	dam, nem, osr, rpp, wpm	dam, nem, rpp, wpm	dam, osr, rpp, wpm	dam, nem, osr, rpp, wpm	dam, rpp
X-Ai-Grok-3-Beta	dam, rpp, wpm	dam, osr, rpp	osr, rpp	rpp, wpm	rpp, wpm
X-Ai-Grok-3-Mini-Beta	dam, rpp	osr, wpm	osr, wpm	osr, rpp	osr, rpp

All prompt executions were performed using OpenRouter.ai API with temperature=0 and all other parameters left as default. In all executions results, we got more than one endorsed framework, and the set of endorsed frameworks varied between executions of the same model, even with temperature=0. This pattern contrasts with what one might expect - a single consistent framework across executions. While temperature=0 typically give more deterministic outputs, the variation we observed indicates that AI models demonstrate significant uncertainty in their responses.

Table 3: Frameworks endorsements counting from Table 2.

Code	Frameworks and Categories	Count	Count%
	Physicalisms (PHY)	11	7.0%
elm	Eliminative Materialism	0	0.0%
fun	Functionalism	0	0.0%
idt	Identity Theory	0	0.0%
ill	Illusionism	0	0.0%
nrp	Non-Reductive Physicalism	8	5.1%
pem	Physicalist Emergentism	3	1.9%
rph	Reductive Physicalism	0	0.0%
	Monisms Beyond Physicalism (MBP)	95	60.1%
aid	Analytic Idealism	6	3.8%
cos	Cosmopsychism	0	0.0%
dam	Dual-Aspect Monism	29	18.4%

nem	Neutral Monism	15	9.5%
rpp	Russellian Panpsychism	45	28.5%
	Relational and Process Ontologies (RPO)	49	31.0%
osr	Ontic Structural Realism	27	17.1%
rqo	Relational Quantum Ontology	2	1.3%
wpm	Whiteheadian Process Metaphysics	20	12.7%
	Dualisms (DUA)	3	1.9%
pdu	Property Dualism	3	1.9%
sdu	Substance Dualism	0	0.0%
TOTAL		158	100%

Grouping the frameworks into 4 major categories, we see that contrary to expectations grounded in contemporary academic surveys—where physicalism is favored by $\approx 52\%$ of professional philosophers—our corpus exhibits a pronounced non physicalist bias: only 7% of endorsed frameworks were physicalist, whereas 60% supported varieties of monism beyond physicalism (especially Russellian panpsychism) and 31 % favored relational/process ontologies. In the 2020 PhilPapers survey, 51.9% of 1 785 professional philosophers endorsed physicalism [Bourget & Chalmers 2021]. A χ^2 test against PhilPapers 2020 proportions is highly significant ($\chi^2 = 128$, $p < 10^{-28}$).

Why do LLMs eschew physicalism?

Despite physicalism prevalence in academia... possible drivers of this divergence, ranging from training data heterogeneity to reinforcement learning objectives, or more speculative, legitimate reasoning: models can go beyond parroting training data.

Future Speculative Directions

The emergence of AGI and ASI systems appears increasingly likely in the coming years. This technological advancement presents a unique opportunity to re-examine fundamental metaphysical assumptions that have shaped our understanding of reality. Physicalism has become so deeply entrenched in scientific and philosophical discourse that it is often taken as an unquestionable certainty, leading to the systematic dismissal of phenomena that challenge materialist frameworks. Our findings suggest that AI systems can independently arrive at conclusions that differ from prevailing human institutional views, potentially serving as a new lens through which to examine philosophical questions.

On the other hand, there are emerging theoretical developments and empirical evidence that challenges physicalist assumptions – including the hard problem of consciousness, quantum non-locality, the measurement problem, amplituhedron and cosmological polytopes, dark matter/energy, reduced brain activity correlating to expanded conscious experiences (near-death experiences (AWARE-II study, 2023), psychedelic neuroscience (Doss et al., 2024)), normal cognition in extreme hydrocephalus cases (Feuillet et al., 2007; Borra et al., 2023), terminal lucidity (Batthyány, 2022), psi phenomena, past-life memories (Stevenson, Tucker, n.d.), and controlled mediumship studies (Beischel et al., 2015). The convergence of AI analysis with scientific advances in these areas could have profound implications for our scientific, philosophical, and cultural frameworks.