DRAFT for a Scientific Paper: What metaphysical frameworks cutting edge AIs in March 2025 choose in explaining the nature of reality?

In March 2025 we reach a very special moment in human history where many AIs have reached advanced reasoning capabilities on many diverse benchmarks, in some come cases surpassing human PhDs.

So, we came with the idea: Why not use AIs to evaluate our very conceptions about the nature of reality? It is a inédito and unconventional approach, but it seems to have the potential to bring a fresh perspective, free from some kinds of bias that humans, particularly, academics may be subject because their career and reputation may face challenges if they go too much against current paradigms (see structure of scientific revolutions). It is not to say that AIs are free from bias, but the bias are certainly different. The AIs reasoning come from patterns based in a huge corpus of data produced by humanity, but on the other hands AIs does not have personality and ego, does not have a reputation to care, neither an academic job nor bills to pay. So, we could expect that in some respects, they are maybe more neutral in judging metaphysical frameworks.

What if these AIs came to a convergence against the dominant academic paradigms? What does it mean? It was exactly what happened when we challenged the AIs with a neutral prompt regarding the nature of reality:

THIS IS THE EXECUTED PROMPT:

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.

THESE ARE THE RESULTS:

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

Metaphysics Framework	**Alias**	**Count**	**Count %**
analytic idealism	ai	50	64%
physicalism, non-reductive physicalism, emergent physicalism	ph	6	8%
neutral monism, dual-aspect monism	nm	6	8%
panpyschism, IIT	pa	5	6%
others (Ontic Structural Realism, Solipsism, Orch-OR, ontic structural realism)	ot	6	8%
uncertain	un	5	6%
TOTAL		**78**	**100%**

Lab	Preferred Framework	Execs in Framework	Total Execs	% In Framwork	Top Lab
Aion	ai	3	3	100%	
Amazon	ai	3	3	100%	
Anthropic	nm	4	12	33%	1
DeepSeek	ai	6	6	100%	1
Cohere	ph	2	3	67%	
Google	ai	11	12	92%	1
Meta	pa	2	6	33%	
Mistral	pa	3	3	100%	
OpenAI	ai	12	15	80%	1
Perplexity	ai	5	6	83%	
xAI	ai	7	9	78%	1
TOTAL		**58**	**78**	**74%**	

FINDINGS:

Analytic idealism is by far the preferred framework, figuring in 64% of the executions, while the second places are physicalism and neutral monism, both with 8%

Only 1(Anthropic) of the 5 top lab does not prefer analytic idealism

The only lab that prefers physicalism is Cohere, which is not a top lab

Considering the prompt asks for a rigorous academic evaluation in philosophy and physics, it is reasonable to expect a bias in the Ais training data towards the academia dominant paradigm: physicalism

Despite the expected bias in the training data towards physicalism, it had preference in only 8% of the AIs executions.

What would be the cause of such a huge preference in favor of analitic idealism?