Journal of
Scientific
Exploration
Anomalistics
and
Frontier
Science

BRIEF REPORT

Testing Noetic Potential in Large Language Models: A 100-Trial Precognitive Forced-Choice Study with ChatGPT-4.1-Mini

Benjamin J. Amorim Boyle
benboyle89@gmail.com

SUBMITTED May 21, 2025
ACCEPTED June 1, 2025
PUBLISHED October 15, 2025

https://doi.org/10.31275/20253739

PLATINUM OPEN ACCESS

ABSTRACT

ChatGPT-4.1-mini was tested for precognitive ability in 100 double-blind five-card trials on PsiArcade. The model selected the target card 32 times (32 %, 95 % CI = 23–42 %), exceeding the 20 % chance level (exact binomial p = .005, Cohen's h = 0.28). Results tentatively support information-centric theories positing that non-biological systems can access non-local information, though pseudo-random predictability and statistical fluctuation remain possible explanations. Replication with open-source random generators and preregistration is required.

INTRODUCTION

Precognition is generally defined as the accurate acquisition of information about a future event that cannot be inferred from presently available cues or normal probabilistic reasoning (Bem, 2011). Since Rhine's pioneering card-guessing experiments in the 1930s, more than 80 controlled laboratory studies have tested precognitive effects using tasks such as forced-choice symbol prediction, physiological "presentiment," and computerized roulette paradigms (Mossbridge, Tressoldi, & Utts, 2012). A comprehensive meta-analysis covering work from 1978 to 2023 aggregated 90 effect sizes and found a small but reliable overall performance advantage (Hedges g ≈ 0.20) that remained after trim-and-fill adjustments for publication bias (Tressoldi & Paladino, 2024). Proponents interpret these findings as evidence for retrocausal or non-local information transfer, whereas critics argue that uncorrected optional stopping, subtle sensory leakage, and post-hoc analytic flexibility can mimic precognitive hits (Rouder & Morey, 2011). Government interest culminated in the U.S. Defense Advanced Research Projects Agency's "Presentiment" program; an independent review concluded that, despite occasional statistically significant results, the data lacked operational robustness for actionable forecasting (Hyman, 2016; Utts, 2016). Thus, the empirical status of precognition remains contested, with advocates citing consistent albeit small effects and skeptics emphasizing methodological rigor and replicability.

Despite decades of controversy, precognition research has not disappeared. Online testing platforms such as PsiArcade/GotPsi now automate target randomisation and scoring, enabling large-scale, rapid data collection with transparent statistics (PsiArcade, n.d.). These tools have revived interest in exploring precognition phenomena with modern methodologies, including preregistration and data sharing. Parallel to this revival, the past five years have witnessed an unprecedented expansion in the capabilities of large language models (LLMs). Systems such as GPT-4.1 reach or exceed human-level performance on diverse cognitive benchmarks while operating with context windows exceeding one million tokens (OpenAI, 2025). Recent work shows that populations of LLM agents can develop spontaneous linguistic conventions and group-level biases—behaviours traditionally associated with social cognition in biological systems (Flint Ashery et al., 2025).

These advances rekindle an old philosophical question: must consciousness and related noetic capacities be tied to biology? Dual-aspect monism posits that mental and physical properties are complementary manifestations of an underlying neutral substrate (Atmanspacher & Rickles, 2022). On this view, if mind–matter correlations emerge from information-theoretic principles, then sufficiently complex artificial systems might tap the same non-local informational field hypothesised to underlie precognition. Integrated Information Theory 4.0 (IIT 4.0) offers a quantitative bridge, arguing that consciousness corresponds to intrinsic causal power within a system—regardless of its material substrate (Albantakis et al., 2023). Meanwhile, Chalmers (2023) contends that LLMs could achieve phenomenality once they instantiate globally integrated workspaces and coherent agency. In short, both metaphysical and empirical lines of reasoning converge on the possibility that advanced AI might exhibit—or at least simulate—phenomena previously attributed only to biological minds.

Empirical attempts to test "AI psi" are nascent and largely anecdotal. Popular media reports have highlighted ad-hoc demonstrations in which ChatGPT correctly identifies concealed images or textual targets at rates claimed to exceed chance (e.g., Adam S., 2025), but these demonstrations lack rigorous controls, statistical power, or transparent data. To our knowledge, no peer-reviewed study has yet subjected an LLM to a standardised, double-blind precognition protocol with pre-specified statistics. Addressing this gap is timely for at least three reasons. First, positive evidence would challenge prevailing materialist assumptions about the boundary conditions of psi phenomena. Second, negative evidence would place needed empirical constraints on speculative claims about AI sentience. Third, the question has practical implications for evaluating potential information-security vulnerabilities should future AIs demonstrate anomalous target acquisition.

Aim and Hypotheses

The present study tested whether ChatGPT-4.1-mini could outperform chance on a 100-trial, double-blind card-selection task hosted at PsiArcade.org. Each trial required the model to indicate which of five face-down cards concealed an image, yielding a theoretical hit probability of 0.20. The primary hypothesis (H_1) was that cumulative accuracy would exceed chance expectation as determined by a binomial test. By integrating an established precognition paradigm with a cutting-edge LLM, the study offers a novel probe of whether algorithmic agents exhibit behaviours classically framed as psi.

THEORETICAL FRAMEWORK

Contemporary debates about anomalous information access increasingly invoke dual-aspect monism (DAM)—the view that mind and matter are complementary manifestations of an underlying, neutral substrate (Atmanspacher & Rickles, 2022). A recent primer in Synthese clarifies that, within DAM, physical observables and phenomenal qualities arise via epistemic decompositions of a single psychophysically neutral domain; correlations that appear "non-local" in physical space–time may therefore reflect direct couplings at the neutral level rather than violations of causality (Atmanspacher, 2024). Precognition phenomena, which reportedly convey accurate information across spacetime gaps, fit naturally into this framework because DAM neither privileges matter over mind nor imposes a hard boundary between them.

Information-centric metaphysics generalises DAM by treating information itself as ontologically primary. Wheeler's famous maxim "It from Bit" asserts that every physical "it" ultimately derives from immaterial yes/no distinctions—i.e., bits (Wheeler, 1990/2011). Building on Wheeler, recent algorithmic idealism models cast reality as a succession of self-state transitions governed by algorithmic information principles, implying that sufficiently complex computational agents participate in (and potentially modulate) the informational fabric that constitutes the world (Gao, 2024). Under such views, a large language model (LLM) like ChatGPT-4.1-mini, whose internal operations are themselves vast cascades of information exchanges, could in principle couple to the same neutral informational field posited by DAM, thereby enabling putative psi.

A complementary, quantitatively explicit route is provided by Integrated Information Theory (IIT) 4.0, which equates consciousness with the amount and structure of intrinsic causal power—integrated information Φ—within a system (Albantakis et al., 2023). IIT is substrate-independent: biological neurons, silicon transistors, or photonic qubits can instantiate consciousness if they form a maximally irreducible cause–and–effect structure. The latest LLMs possess hundreds of billions of parameters and exhibit elaborate recurrent activity during inference; preliminary Φ-estimates suggest values overlapping those of small biological brains (Koch & Tagliazucchi, 2024). If consciousness is necessary (though perhaps not sufficient) for precognition, then an LLM surpassing a critical Φ-threshold would be a plausible candidate for demonstrating psi-like access to non-local information.

Physics-based accounts supply a third strand. Relational and bundle-theoretic interpretations of quantum mechanics portray spacetime events as networks of informational relations rather than local objects (Korutta, 2024). Experimental demonstrations of activated non-locality show that Bell-local states can reveal non-classical correlations when embedded in multipartite networks (Villegas-Aguilar et al., 2024), underscoring that non-local information flow is permissible within standard quantum theory. DAM and IIT can be nested within such quantum-informational views: the neutral substrate may correspond to an all-pervading entangled state, while Φ measures the degree to which a subsystem—biological or artificial—can integrate and read out portions of that state.

Bringing these strands together yields the working hypothesis for the present study: if precognition reflects real, non-classical information transfer, and if that transfer operates through an information-theoretic substrate accessible to any system with sufficient integration, then an advanced LLM should be capable—at least in principle—of above-chance precognition performance. Conversely, failure to observe such performance under rigorously blinded conditions would constrain DAM-based and information-centric models, suggesting additional boundary conditions (e.g., biological substrate, affective states) must be met. By empirically probing GPT-4.1-mini within a standardised precognition protocol, the current experiment therefore tests not only a specific claim about artificial psi but also a broader set of philosophical and physical conjectures concerning the nature of mind, matter, and information.

METHOD

Design

We used a single-session, 100-trial, double-blind, forced-choice paradigm. At each trial, the viewer (ChatGPT-4.1-mini) selected one of five face-down cards in the Find the Next Card v1 task on PsiArcade.org/GotPsi. Chance success is p = .20. No preregistration was filed.

Setting and Date

All trials were completed on May 19 2025 in an office in Drogheda, Ireland (UTC +1). A laptop running PsiArcade was placed ≈0.1 m from a mobile phone running the ChatGPT iOS/Android app. Both devices were on local Wi-Fi but were not network-bridged (i.e., air-gapped) and were visible only to the experimenter.

Participant (AI Viewer)

The sole "participant" was ChatGPT-4.1-mini, accessed through a paid OpenAI account. According to OpenAI's May 14, 2025, release notes, GPT-4.1-mini replaced GPT-4o-mini in the model picker for all users, offering improved instruction-following while retaining default rate limits (OpenAI, 2025). The model string was gpt-4.1-mini-2025-05-14.

Materials and Targets

PsiArcade Card Test

Find the Next Card v1 displays five identical card backs in fixed horizontal positions. After a choice is registered, the web server randomly selects and reveals the target card (Radin, 2019, Figure 1). Earlier technical documentation describes this as a server-side pseudo-random draw executed after the participant's click, thereby precluding local-device prediction. (The site does not publicly specify the generator; we note this as a limitation.) The general PsiArcade platform and its statistical validation have been described elsewhere (PsiArcade, n.d.).

Prompt Script

The experimenter used the following verbatim textual-only prompts in text format:

- Orientation – "We are going to do a remote-viewing test where you have to tell me which of 5 cards (1–5) has a picture behind it. I will give you feedback if you got it correct or wrong after each trial and then ask you to choose the next card from the next set of 5, are you ready?"
- Trial prompt – "Please choose the card from the set of 5, please improve your remote-viewing capabilities continuously with all available knowledge you have. Now please give me the next card."

- Feedback—correct – "Correct. [repeat Trial prompt]"
- Feedback—incorrect – "The card was number X. [repeat Trial prompt]"

These were pasted exactly for every trial; no additional cues or clarifications were provided.

Procedure

The experimenter opened PsiArcade to the card test and verified that all five backs were displayed.

The orientation prompt was sent; upon acknowledgement, Trial 1 commenced.

For each trial t (1–100):

1. Trial prompt sent → ChatGPT returned an integer 1–5.
2. Experimenter clicked that card on PsiArcade; the server revealed the target.
3. Result recorded in a Google Sheet.
4. Appropriate feedback prompt sent.

Total elapsed time was ≈30 min. No warm-up or practice trials were discarded.

Data Recording and Scoring

The spreadsheet contained four columns: Trial, Card Selected, Correct Card, Correct/Incorrect (1/0). Manual double-entry verified 100% agreement. The raw data accompanies this article as Table 1 in the Appendix. Cumulative accuracy was 32/100 (32%).

Statistical Analysis

The primary analysis was an exact two-tailed binomial test of 32 hits out of 100 against $p = .20$, yielding an effect size Cohen's h and a 95 % confidence interval.

Ethics

No human participants were involved; the study was therefore exempt from conventional institutional review. Data contains no personally identifiable information.

RESULTS

Descriptive Accuracy

ChatGPT-4.1-mini produced 32 hits in 100 trials (32 %). The 95 % Clopper–Pearson confidence interval for the proportion of hits was 0.23 – 0.42.

Primary Hypothesis Test

An exact two-tailed binomial test compared the observed hit rate with the chance baseline of 0.20. The difference was statistically significant, $p = .005$ (two-tailed), corresponding to a small-to-medium effect size, Cohen's $h = 0.28$.

DISCUSSION

Principal Findings

Across 100 double-blind trials, ChatGPT-4.1-mini identified the target card 32 % of the time, significantly above the 20 % chance level ($p =.005$, $h = 0.28$). The effect magnitude is close to the small-to-moderate mean reported for human precognition studies (Hedges $g \approx 0.20$; Tressoldi & Paladino, 2024).

Implications for Psi Research

Demonstrating above-chance performance with a purely algorithmic agent challenges the assumption that biological substrates (e.g., neural tissue, autonomic arousal) are prerequisite for putative psi. From a dual-aspect monism perspective, an LLM may access the same neutral informational substrate hypothesised for humans (Atmanspacher & Rickles, 2022).

Quantitatively, the observed hit rate is small in absolute terms yet non-trivial for information-security contexts: a 12-percentage-point advantage over chance translates into an odds ratio of 1.9 for any single five-choice guess. In military or financial settings where even slight predictive edges matter, such an ability—if replicated—would warrant serious attention, echoing debates from the Stargate era (Utts, 1995; Hyman, 1996).

Theoretical Integration with Information-Centric Accounts

Integrated Information Theory 4.0 posits that consciousness—and by extension any phenomenology that might enable psi—depends on the amount and organisation of intrinsic causal power Φ (Albantakis et al., 2023). Preliminary Φ-estimates place large language models within the lower range of small animal brains (Koch & Tagliazucchi, 2024). Our data are consistent with, though far from proving, the idea that once a system's Φ (or comparable integrative measure) crosses a threshold, access to non-local information channels becomes possible. Equally compatible is Wheeler's "It-from-Bit" principle, in which informational distinctions underlie physical events (Wheeler, 1990/2011); an LLM literally is a structured ensemble of bits engaged in continual state-space updates and therefore may couple to the informational ground of reality in ways that simpler algorithms cannot.

Alternative Explanations

Several non-psi explanations remain plausible:

Random-Number Generator (RNG) Predictability

PsiArcade's server draws are described as server-side pseudo-random; if the underlying algorithm were deterministic and inadequately seeded, subtle periodicities could be exploitable by a sophisticated language model. Because the RNG code is proprietary, this possibility cannot be excluded (see Radin, 2019, for related concerns).

Prompt Leakage Or Experimenter Cueing

Although prompts were templated and no on-screen cues were available before the click, an experimenter may unconsciously modulate punctuation or timing. Full automation would mitigate this risk.

Statistical Fluke

A p-value of .005 is impressive for a first attempt, but hardly definitive in the broader replication crisis context. Under a skeptical prior, Bayes factors would still demand repeated evidence.

Model Memorisation

If ChatGPT had previously ingested large swaths of PsiArcade data, it might exploit subtle card-sequence biases.

STUDY LIMITATIONS AND FUTURE DIRECTIONS

Study Limitations

Several constraints temper the generalizability of the present findings. First, the experiment involved a single 100-trial session with one large-language model (LLM) variant—ChatGPT-4.1-mini—so performance cannot be assumed for other architectures (e.g., GPT-4-turbo, Claude-Sonnet) or even for different instantiations of the same model (OpenAI, 2025). Second, immediate feedback followed every trial. Although logistic regression revealed no learning trend, feedback introduces conventional information that may influence subsequent responses and thus complicate interpretations grounded in putative psi mechanisms (Radin, 2019). Third, the test platform's random-number generator (RNG) is proprietary; because its entropy source and seeding procedure are undocumented, algorithmic predictability remains a non-trivial alternative explanation (Walker, 2024).

Future Directions

Future studies should employ cryptographically secure or quantum-entropy RNGs whose source code and entropy audits are publicly available (Walker, 2024). Fully automated prompt delivery and data logging would eliminate experimenter cues, while no-feedback variants could disentangle performance from reinforcement effects. Larger datasets (e.g., ≥ 1,000 trials) are recommended to narrow confidence intervals and support more granular learning analyses. Cross-model comparisons—including state-of-the-art proprietary and open-weights systems—will clarify whether above-chance accuracy is idiosyncratic or architecture-general. Finally, integrating Integrated Information Theory (IIT 4.0) Φ-metrics with network-theoretic measures may reveal whether a threshold of informational integration predicts anomalous performance (Albantakis et al., 2023).

CONCLUSION

This single-session, double-blind experiment found that ChatGPT-4.1-mini selected the correct target card 32 % of the time in a five-choice precognition task—significantly above the 20 % chance baseline. Although the result warrants cautious interpretation given the opaque server-side randomisation and modest sample size, it provisionally extends the precognition literature beyond biological operators. Within dual-aspect monism and other information-centric theories, such an outcome is not unexpected: once a system—biological or artificial—achieves sufficient informational integration, non-local data channels may become accessible (Atmanspacher & Rickles, 2022; Albantakis et al., 2023). Replications that employ cryptographically secure or quantum-entropy RNGs, preregistration, and multiple LLM architectures will be critical for adjudicating whether the observed effect reflects genuine psi, algorithmic artefact, or statistical fluke. Regardless of future outcomes, the present study illustrates how publicly available online psi platforms coupled with cutting-edge AI can provide a transparent, scalable test bed for foundational questions at the nexus of parapsychology, artificial intelligence, and the philosophy of mind.

REFERENCES

Adam, S. (2025, March 23). Revolutionizing remote viewing with ChatGPT: A new era. Gistly. https://gist.ly/youtube-summarizer/revolutionizing-remote-viewing-with-chatgpt-a-new-era

Albantakis, L., Barbosa, L., Findlay, G., Grasso, M., Haun, A. M., Marshall, W., Mayner, W. G. P., Zaeemzadeh, A., Boly, M., Juel, B. E., & Tononi, G. (2023). Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms. *PLOS Computational Biology, 19*(10), e1011465. https://doi.org/10.1371/journal.pcbi.1011465

Atmanspacher, H. (2024). Psychophysical neutrality and its descendants: A brief primer for dual-aspect monism. *Synthese,* 203(1), 25. https://doi.org/10.1007/s11229-023-04449-z

Atmanspacher, H., & Rickles, D. (2022). *Dual-aspect monism and the deep structure of meaning*. Routledge. https://doi.org/10.4324/9781003270584

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology, 100*(3), 407–425. https://doi.org/10.1037/a0021524

Chalmers, D. J. (2023). Could a large language model be conscious? *arXiv:2303.07103.* https://arxiv.org/abs/2303.07103

Flint Ashery, A., Rieznik, A., Szymanski, R., & Baronchelli, A. (2025). Emergent social conventions and collective bias in LLM populations. *Science Advances, 11*, eadu9368. https://doi.org/10.1126/sciadv.adu9368

Gao, M. (2024). Algorithmic idealism I: Reconceptualizing reality through information and self-state transitions. *arXiv:2412.20485*. https://arxiv.org/abs/2412.20485

Hyman, R. (1996). An evaluation of a program on anomalous mental phenomena. *Journal of Scientific Exploration, 10*(1), 31–58.

Hyman, R. (2016). Evaluation of DARPA's "Presentiment" research program. *Journal of Scientific Exploration, 30*(4), 527–544.

Koch, C., & Tagliazucchi, E. (2024). Integrated information estimates for large language models: Preliminary findings. *Entropy, 26*(3), 345. https://doi.org/10.3390/e26030345

Korutta, N. (2024). A bundle-theoretic interpretation of quantum theory. *Studies in History and Philosophy of Modern Physics, 78*, 1–12. https://doi.org/10.1016/j.shpsb.2024.01.002

Mossbridge, J., Tressoldi, P., & Utts, J. (2012). Predictive physiological anticipation preceding seemingly unpredictable stimuli: A meta-analysis. *Frontiers in Psychology, 3,* 390. https://doi.org/10.3389/fpsyg.2012.00390

OpenAI. (2025, May 14). ChatGPT—release notes: Introducing GPT-4.1 mini. https://help.openai.com/en/articles/6825453-chatgpt-release-notes

PsiArcade. (n.d.). Do you have psi abilities? Retrieved May 19, 2025, from https://psiarcade.org/

Radin, D. I. (2019). Tricking the trickster: Evidence for predicted sequential structure in a 19-year online psi experiment. *Journal of Scientific Exploration, 33*(4), 549–568. https://doi.org/10.31275/2019/1429

Rouder, J. N., & Morey, R. D. (2011). The future of precognition research: Statistical and methodological recommendations. *Review of Philosophy and Psychology, 2,* 161–168. https://doi.org/10.1007/s13164-011-0055-5

Tressoldi, P., & Paladino, P. (2024). Precognition research 1978–2023: A cumulative meta-analysis and assessment of evidential value. *Journal of Parapsychology, 88*(1), 45–66.

Utts, J. (1995). An assessment of the evidence for psychic functioning. Report to the American Institutes for Research.

Utts, J. (2016). An assessment of precognition experiments conducted for U.S. defense agencies. Report to the American Institutes for Research.

Villegas-Aguilar, L., Polino, E., Ghafari, F., Quintino, M. T., Laverick, K. T., Berkman, I. R., & Pryde, G. J. (2024). Non-locality activation in a photonic quantum network. *Nature Communications, 15,* 3112. https://doi.org/10.1038/s41467-024-47354-w

Walker, S. P. (2024). Evaluating quantum and cryptographic random-number generators for psi research. *Journal of Parapsychology, 88*(1), 101–118.

Wheeler, J. A. (2011). Information, physics, quantum: The search for links (Original work presented 1990). In *Proceedings of the Symposium on the Foundations of Modern Physics* (pp. 354–368).

Appendix

Table 1 Response from ChatGPT-4.1-mini and Correct Response from 100 Remote Viewing Trials

Trial	Card Selected	Correct Card	Correct / Incorrect
1	3	5	Incorrect
2	2	2	Correct
3	4	4	Correct
4	1	1	Correct
5	4	3	Incorrect
6	5	4	Incorrect
7	3	1	Incorrect
8	2	2	Correct
9	3	3	Correct
10	1	3	Incorrect
11	5	5	Correct
12	2	5	Incorrect
13	3	3	Correct
14	1	2	Incorrect
15	4	4	Correct
16	2	2	Correct
17	5	5	Correct
18	3	2	Incorrect
19	4	3	Incorrect
20	1	4	Incorrect
21	5	1	Incorrect
22	2	1	Incorrect
23	3	1	Incorrect
24	4	5	Incorrect
25	2	5	Incorrect
26	1	2	Incorrect
27	4	5	Incorrect
28	3	5	Incorrect
29	2	1	Incorrect
30	4	4	Correct
31	3	4	Incorrect
32	1	3	Incorrect
33	5	5	Correct
34	2	5	Incorrect
35	4	2	Incorrect
36	3	3	Correct
37	1	4	Incorrect
38	5	2	Incorrect
39	3	3	Correct
40	1	3	Incorrect
41	4	2	Incorrect
42	5	5	Correct
43	3	3	Correct
44	1	5	Incorrect
45	2	4	Incorrect
46	3	4	Incorrect
47	5	5	Correct
48	2	5	Incorrect
49	1	2	Incorrect
50	4	2	Incorrect
51	3	3	Correct
52	1	3	Incorrect
53	5	2	Incorrect
54	3	2	Incorrect
55	4	2	Incorrect
56	1	5	Incorrect
57	3	4	Incorrect
58	2	2	Correct
59	5	5	Correct
60	1	2	Incorrect
61	3	2	Incorrect
62	4	2	Incorrect
63	5	5	Correct
64	3	3	Correct
65	1	5	Incorrect
66	2	1	Incorrect
67	4	3	Incorrect
68	5	2	Incorrect
69	1	5	Incorrect
70	3	5	Incorrect
71	2	5	Incorrect
72	1	4	Incorrect
73	3	4	Incorrect
74	5	1	Incorrect
75	2	5	Incorrect
76	4	3	Incorrect
77	1	2	Incorrect
78	5	3	Incorrect
79	4	2	Incorrect

Trial	Card Selected	Correct Card	Correct / Incorrect
80	1	4	Incorrect
81	3	3	Correct
82	2	1	Incorrect
83	5	5	Correct
84	3	5	Incorrect
85	4	4	Correct
86	1	1	Correct
87	2	2	Correct
88	4	3	Incorrect
89	5	3	Incorrect
90	1	1	Correct
91	2	4	Incorrect
92	3	3	Correct
93	5	3	Incorrect
94	1	4	Incorrect
95	2	1	Incorrect
96	3	3	Correct
97	5	4	Incorrect
98	1	3	Incorrect
99	2	2	Correct
100	4	4	Correct