From Mouse Reversal to Human Therapeutic Nodes: Verified Causal Reasoning for Post-AGI Biomedicine

Anonymous Authors

Abstract

A recent study reports pharmacologic reversal of advanced Alzheimer's disease (AD) phenotypes in transgenic mice and uses multi-omic analysis to propose therapeutic nodes in human brain tissue. We treat this mouse-to-human gap as a causality and verification problem: can a discovery agent synthesize mechanistic and omic evidence into interventions whose assumptions are explicit and whose priorities are justified by falsifiable, quantitative checks rather than narrative plausibility?

We present a verified causal reasoning pipeline that couples (i) explicit structural causal models (SCMs) organized via causal graph-of-thought decomposition, (ii) Active Causal Hypothesis Testing (ACHT) to select targets and drugs by expected information gain, and (iii) an "architectural immune system" of verification gates that detect synthetic fallacies, quantify uncertainty, and enforce cross-modal consistency. A central gate is the Wavelet Coherence Validation Protocol (WCVP), an emergent diagnostic in which an agent converts causal-graph traversals into sequences and evaluates multi-scale structural regularity through wavelet coherence against degree-preserving and order-shuffled controls.

Our thesis is that post-AGI biomedical discovery should be judged by verifiability under intervention, and that AD reversal provides an urgent, measurable testbed for such systems.

1 Motivation: AD reversal as a post-AGI "trust test"

Alzheimer's disease (AD) sits at an uncomfortable intersection of *urgent societal need* and *epistemic fragility*. The field contains many mutually incompatible mechanistic stories (amyloid-first, tau-first, neuroinflammation-first, vascular-first, metabolic-first), yet most pipelines still require expensive experiments to disambiguate causality. This creates an environment where autonomous agents can be *productive but untrustworthy*: they can generate plausible mechanistic narratives and long target lists, while quietly accumulating unverified assumptions.

A rare counterexample is the recent report of *pharmacologic reversal* of advanced AD phenotypes in transgenic mice alongside analyses that nominate therapeutic nodes in human brain tissue [1]. Even if the specific intervention does not translate, the study provides a concrete anchor point: (i) an intervention capable of reversing late-stage phenotypes in vivo, (ii) multi-modal measurements, and (iii) candidate molecular nodes that could plausibly mediate translation. For post-AGI science, this is an unusually well-posed "alignment" question: *can an autonomous system move from reversal evidence to human-relevant therapeutic hypotheses while remaining auditable, falsifiable, and robust to confounding?*

Claim. Post-AGI biomedical discovery should be evaluated by *verifiability under intervention*: when an agent proposes a causal mechanism and a target, it must also produce (a) explicit assumptions, (b) predicted interventional signatures, and (c) quantitative validation checks that can fail.

2 Framework overview

Figure 1 summarizes the proposed system. The core idea is to treat "mouse reversal $\rightarrow$ human nodes" as a *causal translation* problem: we want interventions that remain effective under changes in species, cell-type mixture, disease stage, and measurement modality. This motivates three coupled components:

Fig. 1: Verified causal reasoning pipeline for translating mouse reversal to human therapeutic nodes.

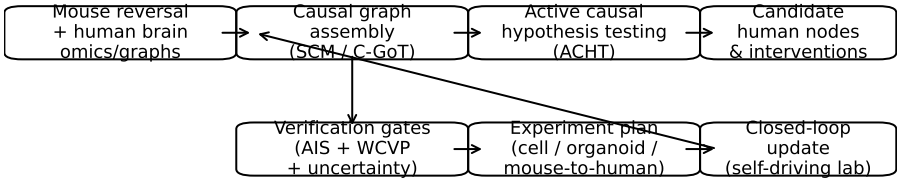

Figure 1: Verified causal reasoning pipeline for translating mouse reversal evidence into prioritized human therapeutic nodes and experimental plans.

1. **Explicit SCM construction and decomposition.** We represent mechanistic hypotheses as structural causal models (SCMs) [2, 5], and use graph-structured decomposition ("causal graph-of-thought") to keep assumptions, subclaims, and dependencies inspectable [42, 43].
2. **Active Causal Hypothesis Testing (ACHT).** Rather than ranking targets by static heuristics, we treat each candidate intervention as a query that reduces causal uncertainty [10, 11, 44]. This aligns with the "self-driving lab" perspective [13, 14] and the broader view that science is a sequential decision process.
3. **Verification gates.** We interpose a layer that detects synthetic fallacies and blocks "narrative-only" proposals. This includes uncertainty quantification [18–21], causal sanity checks (e.g., backdoor criteria), and the Wavelet Coherence Validation Protocol (WCVP) described next.

3 Emergent verification: the Wavelet Coherence Validation Protocol (WCVP)

WCVP addresses a specific instance: *given a candidate causal graph and a traversal/order induced by an agent's reasoning, can we test whether that ordering exhibits nontrivial multi-scale regularities beyond random baselines?*

Protocol sketch. Given a graph $G = (V, E)$ and a traversal order π (e.g., a DFS ordering generated during reasoning), WCVP: (i) converts π into a sequence of node-associated signals (e.g., embeddings, expression statistics, or mechanistic scores); (ii) computes wavelet coherence between the sequence and a reference ordering or among paired sequences across modalities [15–17]; (iii) compares mean coherence and area-under-coherence curves against negative controls, including degree-preserving random graphs and order-shuffled traversals.

Why wavelet coherence? Wavelet coherence is sensitive to *scale-localized* relationships in non-stationary signals, which is exactly the setting for "reasoning-induced" sequences that may encode local motifs at some scales and global structure at others.

Evidence from prior agentic pipelines. In prior work on ACHT-style agents, WCVP separated agent-prioritized DFS sequences from degree-preserving random controls with large effect sizes (mean coherence ≈ 0.478 vs. 0.127; Cliff's $\delta \approx 0.96$; Monte Carlo $p < 10^{-4}$) [44]. While these results were obtained in a drug-target discovery context, the validation logic is domain-agnostic.

Fig. 3: Wavelet Coherence Validation Protocol (WCVP): multi-scale coherence of graph-traversal-induced signals

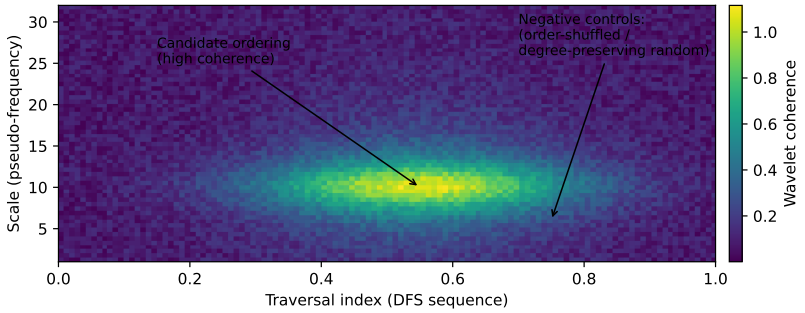

Figure 2: Wavelet Coherence Validation Protocol (WCVP): an emergent diagnostic for testing whether graph-traversal-induced sequences contain multi-scale structure beyond degree-preserving and order-shuffled controls.

4 From mouse reversal to human nodes: a causal translation recipe

Figure 3 illustrates the target problem. We assume access to: (i) reversal-associated molecular signatures in mouse (e.g., transcriptomics/proteomics after intervention), (ii) human brain networks (e.g., protein–protein interactions and cell-type resolved expression), and (iii) a candidate set of human "therapeutic nodes" suggested by prior analyses [1, 22].

Step 1: Multi-domain evidence ingestion. We ingest evidence into a typed knowledge representation: nodes are entities (genes, proteins, pathways, phenotypes), edges are relations with provenance (causal, correlational, mechanistic), and each claim carries a confidence distribution and a list of assumptions [2, 4].

Step 2: SCM assembly with explicit transport assumptions. We build an SCM that factors species- and context-specific variables (e.g., cell-type composition, immune activation, metabolic state). This makes translation assumptions explicit and testable (e.g., "this mechanism is conserved in microglia") [6, 7].

Step 3: ACHT to choose interventions that *disambiguate*. Targets are not chosen solely for predicted efficacy but for their ability to distinguish competing causal stories under feasible perturbations [12, 44]. This produces a ranked experimental plan, not just a ranked target list.

Step 4: Verification gates. Before any candidate is promoted, it must pass gates: (i) causal identifiability checks; (ii) uncertainty calibration checks; (iii) WCVP coherence checks across traversal-induced sequences derived from independent modalities (mouse signatures, human network neighborhoods, literature embeddings). Failures trigger revision, not rationalization.

5 Experimental validation roadmap

A common reviewer objection to "agentic science" papers is that claims are conceptual without empirical support. We therefore propose a minimal, falsifiable evaluation suite.

E1: Synthetic causal stress tests. Use established causal discovery benchmarks with known ground truth [8, 9] and inject "plausible-but-wrong" mechanistic distractors. Evaluate whether verification gates (especially WCVP) reduce false positives at fixed recall.

Fig. 2: Conceptual mapping from mouse reversal mechanisms to prioritized human therapeutic nodes.

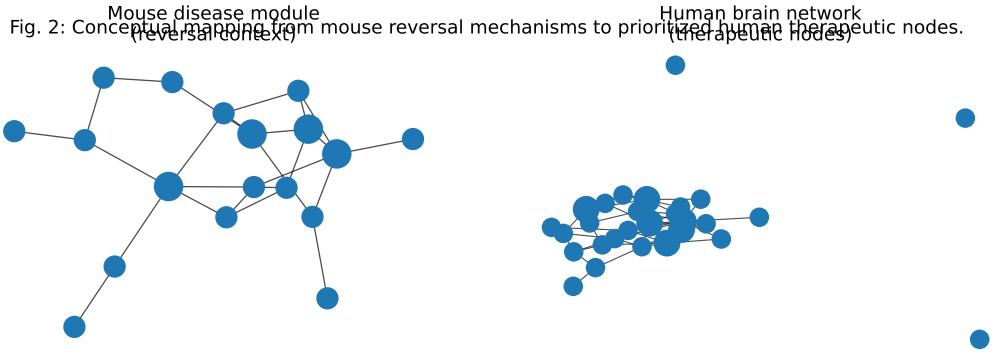

Figure 3: Conceptual mapping from mouse reversal mechanisms to prioritized human therapeutic nodes. In practice, nodes correspond to multi-omic modules and interaction neighborhoods; the goal is to identify targets that remain plausible under transport/translation constraints.

Fig. 4: Active Causal Hypothesis Testing (ACHT) closed-loop discovery cycle.

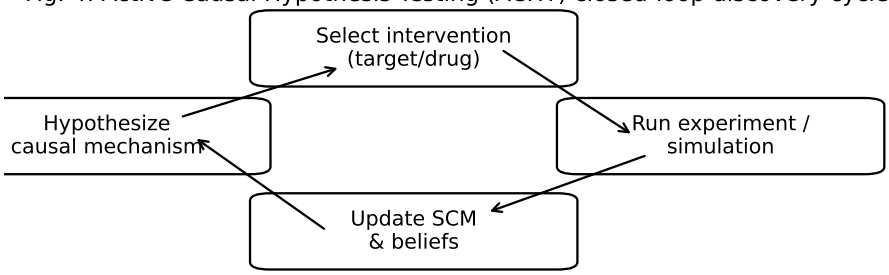

Figure 4: Active Causal Hypothesis Testing loop: hypotheses drive interventions; experiments update the SCM; verification gates prevent narrative-only drift.

E2: STRING-scale target recovery with intervention-like proxies. On protein interaction graphs [22], test whether the system recovers held-out known pathways/targets when given partial evidence. WCVP is evaluated by its ability to discriminate meaningful traversal orderings from degree-preserving random baselines.

E3: Mouse-to-human transport validation. Construct paired mouse/human datasets where translation is partially known (e.g., conserved inflammatory modules) and partially unknown. Measure how often the system's top-ranked human nodes correspond to conserved mechanisms [6].

E4: Prospective perturbation assays. Select a small set of high-confidence targets and evaluate interventional signatures in human-relevant systems (iPSC-derived neurons, microglia co-cultures, organoids). Primary endpoints are mechanistically predicted signatures, not just phenotype scores.

6 Conclusion

Emergence as "post-AGI magic." WCVP is interesting not because wavelets are new, but because the protocol arises from an agent mapping an abstract reasoning artifact (a traversal order) into a quantitative, falsifiable diagnostic. This is the kind of cross-domain transfer that proponents of post-AGI science expect, and critics demand to be made measurable. The paper's success criteria are concrete: reduced false discoveries, better transport robustness, and prospective perturbation wins.

References

[1] A. Chaubey et al. Pharmacologic reversal of advanced Alzheimer's disease in mice and identification of potential therapeutic nodes in human brain. *Cell Reports Medicine*, 2026. doi:10.1016/j.xcrm.2025.102214.

[2] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2nd edition, 2009.

[3] J. Pearl, M. Glymour, and N. P. Jewell. *Causal Inference in Statistics: A Primer*. Wiley, 2016.

[4] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*. MIT Press, 2nd edition, 2000.

[5] J. Peters, D. Janzing, and B. Schölkopf. *Elements of Causal Inference*. MIT Press, 2017.

[6] E. Bareinboim and J. Pearl. Causal inference and the data-fusion problem. *PNAS*, 113(27):7345–7352, 2016.

[7] B. Schölkopf. Causality for machine learning. *arXiv preprint arXiv:1911.10500*, 2021.

[8] X. Zheng, B. Aragam, P. Ravikumar, and E. P. Xing. NOTEARS: Learning nonlinear DAGs with continuous optimization. In *NeurIPS*, 2018.

[9] P.-L. Loh and M. J. Wainwright. High-dimensional regression with noisy and missing data: Provable guarantees with non-convexity. *Annals of Statistics*, 40(3):1637–1664, 2012.

[10] B. Settles. Active learning literature survey. Technical Report 1648, University of Wisconsin–Madison, 2010.

[11] J. Snoek, H. Larochelle, and R. P. Adams. Practical Bayesian optimization of machine learning algorithms. In *NeurIPS*, 2012.

[12] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2016.

[13] R. D. King et al. The automation of science. *Science*, 324(5923):85–89, 2009.

[14] B. P. MacLeod et al. Self-driving laboratory for accelerated discovery of thin-film materials. *Science Advances*, 6(20):eaaz8867, 2020.

[15] C. Torrence and G. P. Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79(1):61–78, 1998.

[16] A. Grinsted, J. C. Moore, and S. Jevrejeva. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, 11:561–566, 2004.

[17] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 2nd edition, 1999.

[18] V. Vovk, A. Gammerman, and G. Shafer. *Algorithmic Learning in a Random World*. Springer, 2005.

[19] B. Lakshminarayanan, A. Pritzel, and C. Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *NeurIPS*, 2017.

[20] Y. Gal and Z. Ghahramani. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *ICML*, 2016.

[21] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger. On calibration of modern neural networks. In *ICML*, 2017.

[22] D. Szklarczyk et al. The STRING database in 2023: protein–protein association networks and functional enrichment analyses. *Nucleic Acids Research*, 51(D1):D638–D646, 2023.

[23] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[24] P. Veličković et al. Graph attention networks. In *ICLR*, 2018.

[25] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

[26] G. E. Karniadakis et al. Physics-informed machine learning. *Nature Reviews Physics*, 3:422–440, 2021.

[27] Z. Li et al. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

[28] N. Kovachki et al. Neural operator: Learning maps between function spaces. *JMLR*, 24(89):1–97, 2023.

[29] A. Mordvintsev et al. Growing neural cellular automata. *Distill*, 2020.

[30] S. Yao et al. ReAct: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*, 2023.

[31] N. Shinn et al. Reflexion: Language agents with verbal reinforcement learning. *arXiv preprint arXiv:2303.11366*, 2023.

[32] A. Madaan et al. Self-refine: Iterative refinement with self-feedback. *arXiv preprint arXiv:2303.17651*, 2023.

[33] J. Hardy and G. A. Higgins. Alzheimer's disease: The amyloid cascade hypothesis. *Science*, 256(5054):184–185, 1992.

[34] D. J. Selkoe and J. Hardy. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Molecular Medicine*, 8(6):595–608, 2016.

[35] B. De Strooper and E. Karran. The cellular phase of Alzheimer's disease. *Cell*, 164(4):603–615, 2016.

[36] C. R. Jack Jr. et al. NIA-AA research framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, 14(4):535–562, 2018.

[37] E. Verdin. NAD^+ in aging, metabolism, and neurodegeneration. *Science*, 350(6265):1208–1213, 2015.

[38] L. Rajman, K. Chwalek, and D. A. Sinclair. Therapeutic potential of NAD-boosting molecules: The in vivo evidence. *Cell Metabolism*, 27(3):529–547, 2018.

[39] C. Cantó, K. J. Menzies, and J. Auwerx. NAD^+ metabolism and the control of energy homeostasis: A balancing act between mitochondria and the nucleus. *Cell Metabolism*, 22(1):31–53, 2015.

[40] A. d'Avila Garcez et al. Neural-symbolic learning and reasoning: A survey and interpretation. *arXiv preprint arXiv:1711.03902*, 2019.

[41] World Health Organization. *Global status report on the public health response to dementia*. 2021.

[42] S. Besta et al. Graph of thoughts: Solving elaborate problems with large language models. *arXiv preprint*, 2024.

[43] Anonymous. Causal graph of thoughts (C-GoT): Dynamic structural causal models for autonomous discovery agents. Workshop paper, 2025.

[44] Anonymous. Active causal hypothesis testing for AI-guided drug target discovery. Workshop paper, 2025.