The Journal of Prevention of Alzheimer's Disease 13 (2026) 100402

Contents lists available at ScienceDirect

The Journal of Prevention of Alzheimer's Disease

journal homepage: www.elsevier.com/locate/tjpad

Special Article

Mining the gaps: Deciphering Alzheimer's biology through AI-driven reconciliation

Cory C. Funk [a,b,*], Tom Paterson [b], Alex Bangs [b], David M. Cannon [f], George Savage [b], Eric Ringger [c], Lee Hood [a,b,d,e]

[a] Institute for Systems Biology, Seattle WA
[b] Fulcrum Neuroscience, Palo Alto, CA
[c] Brigham Young University, Provo, UT
[d] Phenome Health, Seattle, WA
[e] The Buck Institute, Novato, CA
[f] Provo, Utah, USA

ARTICLE INFO

Keywords:
AI
Machine learning
LLMs
Etiology
Personalization
Reconciliation

ABSTRACT

Alzheimer's disease remains one of the most complex and contested domains in biomedicine, characterized by fragmented findings, competing hypotheses, and limited translational success. We propose that AI can offer not just technical acceleration but a deeper epistemic contribution: reconciliation. Rather than optimizing predictive performance or replicating existing assumptions, the goal is to align disparate data, methods, and mechanistic insights into coherent models that explain how the disease emerges, progresses, and can be treated. This approach centers on digital twins, not as monolithic models, but as flexible, testable architectures grounded in homeostasis, destabilization, and multiscale coherence. Through an iterative, interoperable AI architecture, digital twins integrate evidence, resolve contradictions, and highlight where critical gaps remain. This framework moves beyond incremental progress within the prevailing model to catalyzing a paradigm shift in how Alzheimer's is understood. Reconciliation, in this sense, is not a method but a guiding principle for transforming both the science and its applications.

1. Introduction

Despite decades of intensive research, Alzheimer's disease (AD) remains without a cohesive, mechanistically grounded hypothesis of its etiology. The amyloid hypothesis has long shaped therapeutic development, and recent trials of lecanemab and donanemab have demonstrated modest cognitive benefits in early disease, in spite of both drugs significant reduction in amyloid plaques [1,2]. These results confirm that anti-amyloid therapies can produce incremental clinical effects, although side effects and cost limit their applicability in many patients. Notably, brain atrophy continues despite plaque clearance [3], raising the possibility that neuronal loss may precede or drive amyloid accumulation rather than follow it.

In parallel, billions of dollars in research funding have yielded rich datasets that document AD in extraordinary detail. Efforts such as Alzheimer's Disease Neuroimaging Initiative (ADNI) [4], the Dominantly Inherited Alzheimer Network (DIAN) [5,6], the Religious Orders Study and Memory and Aging Project (ROSMAP) [7], the Accelerating Medicines Partnership - Alzheimer's Disease (AMP-AD), and the ADDI Workbench [8] have collected genomic, transcriptomic, proteomic, metabolomic, imaging, and clinical data across thousands of individuals. ADNI alone has cost over $210 million and includes multi-modal, time-resolved data from thousands of participants [9]. Yet despite this scale, meaningful therapeutic breakthroughs have not followed. Like the parable of the blind men and the elephant, each dataset reveals one part of the story, but integration without reconciliation has left the whole picture incomplete.

This fragmentation extends beyond datasets to tools and standards. As molecular biologist Robert Tjian quipped, scientists would rather use each other's toothbrushes than each other's nomenclature. This simple aphorism reflects broader challenges, including conflicting analysis pipelines, incompatible data formats, and isolated computing environments. Similarly within transcriptomics, dozens of competing tools exist for RNA-seq alignment and analysis, each tailored for narrow use cases

* Corresponding author.
E-mail address: cfunk@isbscience.org (C.C. Funk).

https://doi.org/10.1016/j.tjpad.2025.100402
Received 4 September 2025; Received in revised form 14 October 2025; Accepted 20 October 2025
Available online 1 December 2025

and often incompatible with others. Although major consortia are building interoperable platforms to support data harmonization, such efforts typically reinforce existing models rather than produce new insight.

What is needed is not just better integration, but a shift in the architecture of explanation. As Thomas Kuhn described in *The Structure of Scientific Revolutions*, science often progresses through long periods of stability punctuated by paradigm shifts that restructure the conceptual foundations of a field [10]. Clayton Christensen's theory of disruptive innovation makes a similar point in organizational settings: dominant actors tend to optimize within current frameworks, while transformative change requires the willingness to rebuild from first principles [11]. AD research shows symptoms of both stagnation and sunk-cost inertia, where existing investments make it harder to abandon familiar approaches even when they fall short. A similar dynamic is seen in evolutionary biology, where the theory of punctuated equilibrium describes long periods of stasis interrupted by bursts of genomic reorganization [12,13].

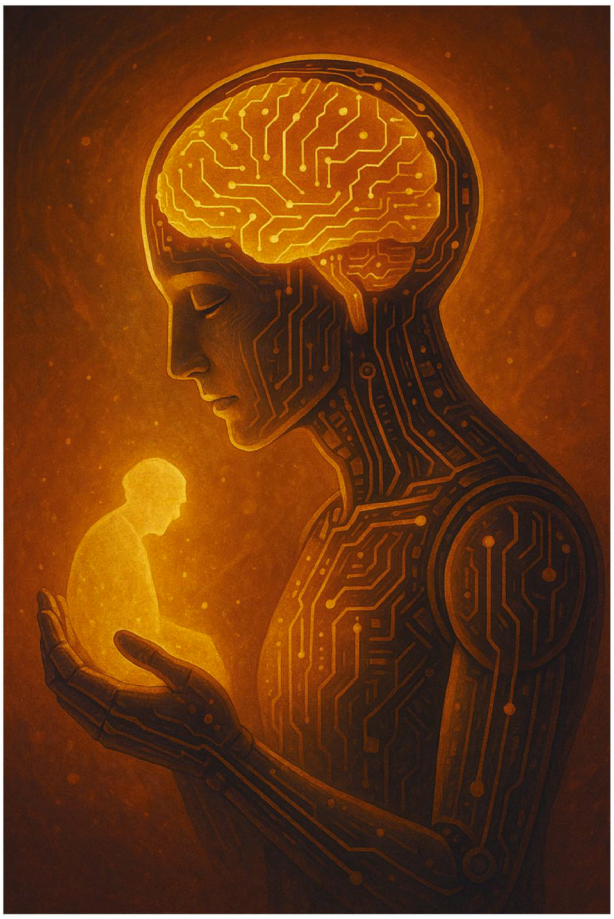
Artificial intelligence offers an opportunity to catalyze such a shift, but only if used strategically. Many applications of AI in AD focus on basic harmonizing of datasets or improving predictions, which are important but limited goals. The real opportunity lies in enabling reconciliation: aligning heterogeneous, multiscale data into causal, testable frameworks that explain rather than merely correlate. This is the promise of digital twins: mechanistic, data-driven models that can simulate biological systems, evaluate interventions, and generate new hypotheses with explanatory power.

In the sections that follow, we define reconciliation as a guiding principle for AI in AD research. We review how different AI frameworks, including language models, generative models, and digital twins, may help resolve contradictions, identify hidden variables, and support causal inference. Our aim is not to catalog all AI tools, but to show how select approaches can drive the kind of conceptual change that Alzheimer's research urgently needs.

2. Reconciliation as the central challenge

Progress in understanding Alzheimer's disease is now limited less by data availability and more by the challenge of reconciling diverse and sometimes conflicting findings. Contradictions, population heterogeneity, experimental artifacts, and static views of dynamic processes fragment our understanding. A metabolic shift in CSF, for instance, could signal pathology, compensation, or sampling error, each with different causal implications. Bridging these gaps requires more than pattern recognition; it calls for tools that integrate data within consistent causal frameworks.

We define reconciliation as the process of aligning and integrating disparate or seemingly conflicting data within a shared explanatory framework that preserves biological plausibility. This involves three essential elements: (1) integrating evidence across multiple scales and modalities, including quantitative, phenotypic, mechanistic, and systems-level; (2) making the causal logic linking these data transparent and open to scrutiny; and (3) ensuring the logic remains faithful to underlying biological reality. In Alzheimer's research, reconciliation means constructing interpretable models that can hold contradictory findings in view, explain variability, and evolve as new evidence emerges. It is not about declaring one pathway correct and discarding the rest, but about building frameworks that accommodate uncertainty while still supporting testable hypotheses, actionable interventions, and scientific trust.



ChatGPT-generated humanoid machine brain holding a fading human memory.

This challenge is made clearer by analogy to cellular automata, where simple local rules can generate highly complex global behavior. If the rules and initial conditions are known, predicting future states is straightforward. But working backward to infer the rules from observations is often computationally intractable, a problem known as computational irreducibility [14]. We see a parallel in Alzheimer's research. Even as data accumulates across scales, we still cannot explain stark phenotypic outliers. A particularly stark example involves the rare APOE Christchurch and RELN protective variants, both of which have been observed to mitigate the effects of early-onset PSEN1 mutations [15–17]. These individuals challenge prevailing causal models and offer a test for any mechanistic framework. Reconciling such cases is not optional; it is the benchmark for mechanistic understanding.

Although we have yet to see AI fully reconcile data into mechanistic explanations, biology offers precedents that demonstrate that reconciliation is possible and potentially powerful. The discovery of the Yamanaka factors, which reprogram somatic cells into induced pluripotent stem cells, reconciled 103 transcriptional profiles with functional assays to overturn assumptions about irreversible cell fate, is one notable example [18,19]. Eric Davidson's work on sea urchin development, integrating gene expression, cis-regulatory logic, and perturbation studies to infer gene regulatory networks to explain cell fate specification, is another [20]. In Drosophila embryogenesis, spatial gene expression patterns were linked to morphogen gradients like Bicoid through dynamical modeling to explain robust developmental patterning [21,22]. These efforts involved spatial, temporal, and functional data, resolved into causal models. They show that even for complex biological systems, simple rules may underlie seemingly intractable complexity. Alzheimer's, by definition, is a complex disease with a complex etiology, but this does not mean it lacks underlying structure. What's missing may be the tools to reconcile it.

One of the key requirements for reconciliation is interpretability. In many AI methods, interpretability and predictive accuracy turn out to be opposing goals. For Alzheimer's research, they must be deeply intertwined. Predictive models without explanation cannot build trust, and trust is essential for both clinical and scientific adoption. The disease's long history of failed trials suggests that predictive accuracy alone is not enough. Models must explain mechanisms to break the cycle of failed predictions and unsupported hypotheses.

Interpretability in this setting must go beyond local explanations of outputs to support epistemic transparency: the ability to reconstruct a model's internal logic, assumptions, and inference pathways so that scientists and clinicians can meaningfully engage with, evaluate, and build upon them [23,24]. This need for transparency is not just a philosophical preference, it is foundational to building trust. In clinical settings, where decisions directly affect patient well-being, the principle of do no harm demands caution [25]. Treatments like lithium or aspirin succeeded before mechanisms were fully understood, but this was only possible due to consistent empirical outcomes. AI models, in contrast, must justify their outputs to earn similar credibility AI models, in contrast, must justify their outputs to earn similar credibility. Without reconciliation of outputs to known biology, AI predictions risk leading to similar past outcomes, with limited or no benefit to patients.

This dynamic can be understood through the lens of the Hegelian dialectic. The *thesis* is interpretability: models whose structures and reasoning align with biological processes, enabling transparency, hypothesis generation, and scientific engagement. The *antithesis* is predictive accuracy: black-box models that achieve impressive results but resist explanation and may lack mechanistic grounding [26,27]. The *synthesis* we argue for is reconciliation. The most powerful are models that integrate predictive strength with epistemic transparency: they not only forecast outcomes but also explain mechanisms, integrate conflicting evidence, and generate new hypotheses. Interpretable AI systems thus become dialectical tools, helping researchers see how disparate observations cohere into a unified understanding, and enabling their reasoning processes to be interrogated, revised, and refined.

Mechanistic understanding can be advanced through both data and modeling. Human-relevant models such as organoids and organ-on-chip systems aim to capture biology that traditional animal models miss, and may help reduce reliance on animal testing [28]. However, whether animal models reflect human biology is itself contested. A prominent study once argued that mouse genomic responses fail to mimic human inflammation, though later analyses disputed this claim, highlighting the need to reconcile model systems themselves [29]. On the modeling side, digital twins and in silico trials can integrate biological constraints and simulate interventions. Such models have already reduced control-arm sizes by up to 33 % in Phase III trials, improving statistical power and reducing patient burden [30,31].

As George Box noted, all models are wrong, but some are useful. We would extend this: the most useful models are those that reconcile the most data across the most contexts, while remaining flexible enough to evolve. In Alzheimer's, where many observations remain unexplained or contradictory, reconciliation should not be an afterthought. It should be the central organizing activity of scientific inquiry. Through reconciliation, we can transform Alzheimer's research from a fragmented collection of signals into a coherent framework capable of explaining resilience, guiding intervention, and restoring scientific clarity.

3. Evaluating AI approaches for reconciliation

3.1. Language models: fluency without mechanism

Large Language Models (LLMs) have become widespread in research, offering new tools for summarization, hypothesis generation, and automation of routine tasks. These models are based on generative transformer architectures that excel at detecting and reproducing linguistic patterns across long sequences. While this makes them highly effective for contextual reasoning, they are optimized for fluency and statistical plausibility rather than factual accuracy or mechanistic understanding [27,32,33]. They do not possess an internal model of physical or biological systems and often fail when tasked with problems outside their training distribution. A concise glossary of these methods is provided in Box 1 for reference.

Recent LLMs are increasingly paired with tool-augmented systems that allow interaction with external resources such as code execution environments, retrieval modules, and search engines. These hybrid systems function as orchestrators, using natural language as the interface for integrating outputs from other tools that may offer greater factual precision or structured reasoning. One such extension is Retrieval-Augmented Generation (RAG), which improves factual grounding by linking responses to external documents. This is especially valuable in scientific domains where traceability and citation are essential. RAG systems enhance transparency by allowing users to verify claims against original sources.

Despite these enhancements, the core limitations of LLMs remain. As noted by Apple researchers in *The Illusion of Thinking* [*34*], LLMs still struggle to construct coherent causal chains, even when provided with the right data. They often falter not due to missing information, but because they lack the capacity to integrate knowledge into a structured, mechanistic understanding. This limitation is particularly problematic in biology, where reconciliation requires aligning data from heterogeneous, noisy, and sometimes contradictory sources. LLMs cannot simulate counterfactuals, weigh conflicting findings, or infer biological mechanisms. Even when they retrieve the correct papers, they frequently fail to interpret differences in experimental design, patient stratification, or underlying confounders [35]. The addition of additional context material to an LLM through RAG does not change these limitations.

LLMs and RAG systems represent one branch of a broader machine learning ecosystem. This ecosystem also includes neural networks for image recognition, causal inference frameworks, interpretable models, and structured causal representations. The strengths of LLMs lie in their ability to synthesize large volumes of literature, generate plausible hypotheses, and automate tasks involving pattern recognition and contextual reasoning. Their weaknesses are equally clear: they may sacrifice accuracy for fluency, lack grounding in biological mechanisms, and perform poorly when asked to generalize beyond their training data. In short, LLMs and RAG systems are effective tools for summarization and idea generation, but they remain inadequate for reconciliation tasks that require causal reasoning and mechanistic fidelity.

3.2. The case for interpretability

Machine learning models vary in how much insight they provide into their predictions. Many deep learning systems offer high performance but low transparency, leaving users with little understanding of how decisions are made. Interpretable machine learning (IML) methods aim to make the logic of a model accessible. These can be inherently interpretable models or post-hoc techniques such as SHAP values [36]. In Alzheimer's research, where understanding mechanism is essential, interpretability is not just a convenience but a requirement.

Box 1
Glossary of AI Methods

Large Language Models (LLMs): Deep learning models trained on massive text corpora using transformer architectures to predict and generate language. LLMs can synthesize literature, generate hypotheses, and automate routine tasks, but they optimize for fluency rather than mechanistic truth, making them prone to errors and "hallucinations." Tool-augmented variants extend LLMs with external reasoning modules (e.g., retrieval, code execution, vision), enabling broader orchestration across AI approaches.

Retrieval-Augmented Generation (RAG): A hybrid approach that grounds LLM outputs in external documents by retrieving relevant references during generation. RAG enhances transparency and factual accuracy by linking outputs back to sources. Its strength lies in evidence retrieval, but it lacks deeper reasoning capacity, and struggles when data are inconsistent, noisy, or mechanistically incomplete.

Interpretable Machine Learning (IML): A set of methods designed to make model decision processes transparent. IML techniques, such as feature attribution, rule extraction, or inherently interpretable architectures, allow researchers to evaluate whether patterns reflect underlying mechanisms. These methods trade predictive accuracy for interpretability, which is essential in domains like Alzheimer's where mechanistic clarity and trust are critical.

Deep Learning Neural Networks (DNNs): Multi-layered computational architectures that learn hierarchical representations of data through successive transformations. DNNs underpin many modern AI systems, including LLMs, generative image models, and AlphaFold, by enabling powerful pattern recognition in high-dimensional spaces. Their strength lies in predictive accuracy and scalability across diverse modalities (text, images, omics), but they often operate as "black boxes," offering limited interpretability. This opacity makes them challenging to refine mechanistically, a key limitation in scientific domains where causal understanding is essential.

Reinforcement Learning (RL): An AI paradigm in which agents learn adaptive control policies through interaction, feedback, and iteration. RL excels at discovering strategies in dynamic systems without explicit supervision. In scientific domains, it offers a way to model adaptive responses, simulate interventions, and explore trajectories of system stability or breakdown. While not itself a neuro-symbolic method, RL integrates naturally within neuro-symbolic frameworks to probe feedback dynamics and intervention policies.

Neuro-Symbolic Reasoning (Umbrella): A hybrid paradigm that combines data-driven learning with structured knowledge (graphs, rules, and priors) to ensure predictions are both powerful and mechanistically grounded. Within this umbrella, specific modeling tools can be employed:

Structured Causal Models (SCMs): Directed acyclic graphs (DAGs) that encode causal assumptions, biological priors, and constraints. They clarify directionality (e.g., Mendelian Randomization), rule out confounders, and support counterfactuals. SCMs excel at testing conditional hypotheses but cannot represent feedback loops central to biological homeostasis.

Dynamic Models: Systems of equations that explicitly model time, feedback, and compensatory processes. They capture recursive regulation, nonlinear adaptation, and resilience/failure modes. Dynamic models are indispensable for simulating disease progression and for digital twins that integrate mechanistic priors with empirical data.

Digital Twins: Dynamic, continuously updated models that serve as reconciliation engines, integrating heterogeneous data, mechanistic constraints, and biological priors into evolving frameworks. Digital twins are capable of simulating homeostatic regulation and its breakdown, enabling a broader capability for causal inference, mechanistic explanation, and testing of interventions across scales.

Surveys by Leist et al. [37]., Freiesleben et al. [38]., and Roscher et al. [39]. emphasize the importance of interpretability for scientific discovery. However, IML also faces limitations. Interpretability can come at the cost of accuracy and may not scale well with high-dimensional biomedical data. Moreover, post-hoc explanations can create a false sense of understanding if they do not reflect the model's actual internal structure. For interpretability to support reconciliation, it must be validated against biological priors and experimental data.

Deep neural networks (DNNs) are the foundation of many high-performing AI systems, including LLMs, generative models, and AlphaFold. Built from layers of nonlinear transformations, they are capable of extracting complex, high-dimensional features from raw data, enabling remarkable predictive performance across fields ranging from natural language processing to protein structure prediction[40]. However, the same layered complexity that makes DNNs powerful also makes them opaque. Their internal representations are difficult to interpret, which limits transparency, reproducibility, and mechanistic insight, especially when data distributions shift[41]. This tension between predictive accuracy and scientific interpretability remains a central challenge in applying deep learning to biology, where causal understanding and experimental validation are essential.

3.3. Explanation as iteration

An alternative to static explanation is the view of explanation as an iterative process. In AI planning, Chakraborti and colleagues proposed that the explanation involves aligning an AI's internal model with the user's mental model through mutual adjustment[42,43]. Rather than delivering a final answer, the system engages in a process of interaction that corrects misconceptions and refines understanding, an approach reminiscent of the Hegelian dialectic, where *thesis* and *antithesis* converge into *synthesis*.

This framing is particularly relevant in biology, where reconciling models with human understanding is often the primary challenge. In biomedicine, the difficulty lies not just in explaining results but in identifying which assumptions are valid. Unlike AI planning, which starts from a well-defined model, biomedical science typically begins with fragmented or conflicting knowledge. Still, the iterative model offers a useful framework for how reconciliation might be operationalized: as a dynamic process of alignment, adaptation, and refinement.

3.4. Neuro-symbolic reasoning as umbrella

While LLMs are useful for generating hypotheses, they rely solely on language-based associations. In contrast, neuro-symbolic systems combine statistical learning with structured knowledge, allowing models to represent causal and mechanistic relationships explicitly. Researchers in this area argue that combining data-driven methods with symbolic reasoning supports better generalization, interpretability, and intervention [44,45].

In this framing, reconciliation involves integrating statistical inference with mechanistic priors to support causal understanding[46]. Structured causal models and dynamic models provide scaffolds that neuro-symbolic systems can use to reason across biological systems.

Reinforcement Learning (RL) adds the capacity to simulate how systems adapt over time, which complements but does not replace the role of structure in causal modeling. As we later argue, digital twins can be seen as a concrete instantiation of this neuro-symbolic vision: an architecture that couples mechanistic cores with adaptive learning to iteratively reconcile diverse data into coherent, testable narratives.

3.5. Reinforcement learning: policies and control

Reinforcement learning (RL) learns by interacting with its environment, adjusting its strategy based on feedback. It does not require labeled data and can discover control policies through experience. In biomedical contexts, RL can model how systems respond to perturbations or interventions over time. This makes it valuable for simulating dynamic responses to treatment or environmental change.

However, RL has limitations. It is resource-intensive, sensitive to reward specification, and prone to instability when feedback is delayed or noisy. RL is not inherently mechanistic, but when embedded in a neuro-symbolic framework, it can test adaptive behaviors while remaining grounded in known biology. Used in this way, RL adds flexibility without sacrificing structure.

3.6. Structured causal models: clarifying directionality

Structured causal models (SCMs) encode assumptions about cause and effect using directed acyclic graphs (DAGs). These models support hypothesis testing, intervention analysis, and the removal of confounding effects. Popularized by Judea Pearl in *The Book of Why*, they have been influential in fields like epidemiology, economics, and in biology where they underpin approaches like Mendelian randomization [47].

SCMs are not a subset of neuro-symbolic reasoning, but a distinct causal framework that can be integrated within neuro-symbolic systems to enhance grounding. SCMs are efficient tools when the causal structure is known or can be approximated from data. However, their acyclic nature means they cannot capture feedback loops or compensatory processes, which are central to biological systems. They can clarify directionality but cannot model the full dynamics of resilience and homeostasis.

3.7. Dynamic models: Capturing feedback and adaptation

Dynamic models extend causal approaches by explicitly representing time, feedback, and adaptation through mathematical formalisms. These models simulate how systems evolve, how they respond to internal or external perturbations, and how they maintain or lose stability. They are especially well-suited for studying diseases like Alzheimer's, where breakdowns in homeostasis occur gradually and are shaped by complex feedback mechanisms.

Dynamic models are interpretable by design and can be integrated into neuro-symbolic frameworks to enable iterative reconciliation. They allow researchers to test hypotheses about how diverse biological variables interact across time, and they support the simulation of how disease might progress or respond to intervention. This capacity makes them foundational to digital twin systems that aim to simulate both health and disease in mechanistic terms.

4. The AI scientist

Several groups have already advanced visions of an "AI scientist" that go beyond single task tools, creating systems that autonomously generate hypotheses, design experiments, and even debate or refine mechanistic models, including the paper by Landess and Bateman et al. in this special issue [REF]. Similar ambitions appear in projects across biomedical discovery and drug development [48–50]. As Demis Hassabis, Nobel laureate and CEO of DeepMind, has noted, the hardest frontier is not generating answers but identifying the right questions. The AI Futures Project's *AI 2027* roadmap envisions a "Superintelligent AI Researcher" capable of doing so at scale [51]. While such systems remain speculative, our framework offers a pragmatic roadmap for Alzheimer's that uses today's AI tools to orchestrate existing methods, reconcile fragmented evidence, and begin by asking the right questions needed to achieve true mechanistic understanding.

4.1. Reckoning with failure, rethinking the questions that matter

Framing better questions, which is central to the goal of an AI scientist, requires not just new tools but a shift in how we approach scientific complexity. This shift does not reject the field's prior successes. On the contrary, it reflects humility toward what decades of brilliant work have already uncovered. The only way forward is by standing on the shoulders of that work and being honest about the places where it has not yet translated to better clinical care.

Modern biology has advanced by dissecting complexity into tractable parts, producing detailed maps of genes, pathways, and molecular circuits. This reductionist strategy has powered transformative discoveries, including the identification of APOE as a key Alzheimer's risk gene and the development of CRISPR-based editing tools. But as Lazebnik warned with his "radio repair" analogy, understanding components in isolation can obscure the organizing principles of the system itself [52]. In Alzheimer's, as in many complex diseases, the result has been an explosion of specialized findings. Most of these findings are true and many are important, but they are often disconnected from a unifying explanation of system failure.

This fragmentation has encouraged a pattern some have called statistical storytelling, which involves weaving plausible narratives from correlational data in the absence of causal models. The reproducibility crisis reflects this broader problem [53]. Studies that initially appear compelling often fail to replicate, a pattern Ioannidis famously attributed to systemic biases and statistical misuse [54]. This is not due to bad science, nor is it exclusive to Alzheimer's [55]. It is often the predictable outcome of disconnected evidence, selective inference, and the lack of frameworks that can reconcile findings into robust, mechanistic insights. AI systems risk amplifying this pattern unless they are paired with models designed to integrate and interpret complexity.

In Alzheimer's research, this challenge is especially acute. Animal models have translated poorly, with over 99.6 % of drug candidates ultimately failing in clinical trials [56], and are often treated as black boxes: useful for producing pathology but poorly predictive of human outcomes. These models can reproduce plaque and tangle pathology but are unable to predict clinical course or therapeutic response. Meanwhile, nearly 100 independent GWAS loci have been linked to Alzheimer's [57, 58], yet aside from APOE, none currently inform diagnosis, prognosis, or treatment. This is not a failure of discovery. It is a failure to assemble discoveries into a cumulative, testable understanding.

A different framing is needed. Many Alzheimer's-associated loci already converge on interpretable biology, such as microglial function and cholesterol trafficking. The challenge is not the absence of meaningful discoveries but the inability to assemble them into a cumulative, testable understanding. Rather than beginning with isolated variables or model outputs, we must start with the goal of reconciling fragmented evidence into coherent, mechanistic understanding. This shift, grounded in past successes but honest about current limitations, is essential for asking better questions. It is the foundation for the AI scientist proposed here.

4.2. No single tool is enough

To achieve its goal, the AI Scientist must operate as an orchestrator across a wide range of methods. Predictive modeling, causal inference, dynamic simulation, symbolic reasoning, and mechanistic modeling all offer partial insights. Their true value emerges when used together to

interrogate the same system from multiple perspectives. This orchestration is as much about judgment as computation: knowing which tool applies to which question, recognizing contradictions as informative, and adjusting models as knowledge evolves. We anticipate that many individual efforts will continue to use AI to harmonize datasets, identify features, and optimize predictions. These are important contributions, but without a unifying framework, they risk reinforcing the very fragmentation that is antithetical to reconciliation. The purpose of this paper is to outline how those incremental efforts can be directed by a higher-level orchestration, where the AI Scientist integrates them as components of a broader reconciliation strategy. As detailed in the sections that follow, we propose specific ways that AI can be used to define and constrain the solution space by treating reconciliation across biological scales as the central challenge. In Alzheimer's, this means integrating across scales (genes, cells, circuits, and populations) while staying focused on the deeper goal: not simply predicting decline, but uncovering how the brain's homeostatic balance destabilizes into disease, with the ultimate aim of enabling treatments, and ultimately cures, that restore stability.

4.3. Kind vs. wicked learning environments

A central challenge for the AI Scientist is recognizing the nature of the problem space. As David Epstein describes in *Range* [59], some domains are kind learning environments, where rules are explicit, feedback is consistent, and outcomes clearly reflect causes. Others are wicked, shaped by delayed feedback, hidden variables, and ambiguity. This distinction is crucial because it defines how AI systems can learn, iterate, and reconcile conflicting information.

Games like chess or Go are quintessentially kind: the rules are fixed, the objectives are unambiguous, and feedback is immediate. In fact, the real breakthrough for AlphaGo came not when it imitated human play, but when it moved beyond human examples and began generating novel strategies by exploring the game space under these transparent rules [60]. But biology, and Alzheimer's in particular, rarely offers such clarity. It is a wicked environment where data are sparse, signals are noisy, and feedback is often delayed. In this context, the goal is not to invent new capabilities, but to reverse-engineer mechanisms that nature has already solved and to align models with those underlying biological truths.

AlphaFold, an AI system developed by DeepMind, transformed structural biology by accurately predicting the 3D structures of proteins from their amino acid sequences, solving a problem that had eluded scientists for more than half a century [61]. Protein folding involves both kind and wicked elements. The kind aspects include well-defined physical constraints that make much of the problem tractable. The wicked aspects include intrinsically disordered regions, which make up approximately 40 percent of the human proteome and never resolve into a single stable structure [62]. AlphaFold succeeded in this mixed regime by embedding biophysical and evolutionary priors and applying iterative refinement to recycle its predictions until structure, constraint, and data aligned within the structured regions. Its success illustrates how AI can navigate partially understood systems by using known constraints while recognizing and respecting areas of unresolved complexity. As John Moult described, AlphaFold solved two problems simultaneously: finding the right solution and knowing when you're there [63].

This is the essence of reconciliation. In wicked or mixed domains, it is not enough to generate outputs that merely appear plausible. The AI Scientist must identify where the rules are well-defined, where uncertainty remains, and how advances in tractable areas can help constrain ambiguity elsewhere. Many current applications of agentic AI thrive in open-ended environments where the goal is to invent new capabilities, unconstrained by a single correct answer. But understanding biology is a fundamentally different challenge: it requires reverse-engineering mechanisms that nature has already solved, where success depends on aligning with those underlying truths. In this context, understanding the limits of current knowledge is itself a valuable scientific contribution, helping to guide discovery toward the questions that matter most.

4.4. Neuro-Symbolic reasoning as framework

To act as a scientist, AI must do more than fit patterns or generate predictions. It must also reason about mechanisms, test hypotheses, and update its beliefs in light of new evidence. This requires a framework that integrates both perception and inference. Neuro-symbolic reasoning provides such a framework by combining data-driven learners (such as deep nets) with structured knowledge (such as graphs, rules, and constraints), allowing inferences that are both powerful and checkable [44]. For Alzheimer's, this is especially important because the problem spans both kind and wicked learning environments. We need models that can learn from noisy, incomplete data while also asserting and testing mechanistic claims. In this framework, learned components handle perception and imputation, while the symbolic layer encodes biological priors, defines allowable transitions, and supports counterfactual reasoning. Structured Causal Models (SCMs) and dynamic systems naturally fit here. SCMs supply testable causal scaffolds, and dynamic models represent trajectories and feedback. Digital twins can instantiate this hybrid approach by embedding mechanistic cores (e.g., compartmental/ODE models, mass/energy/flux constraints) alongside learned modules and then updating as new data arrive. Used this way, neuro-symbolic reasoning transforms disparate data into transparent, testable narratives that not only predict outcomes but explain why, under what assumptions, and how an intervention might shift the course of disease.

SCMs help the AI scientist disentangle directionality by encoding assumptions as directed graphs, allowing for hypothesis formalization, confounder control, and causal inference. This supports systematic exploration of explanatory models, ruling some out while refining others. Mendelian Randomization exemplifies this, using genetic variants as natural experiments to probe causality [64]. But SCMs assume acyclicity and are limited in representing the feedback loops central to biological homeostasis. Used alone, they risk reducing complex dynamics to one-way arrows. Their strength lies in hypothesis narrowing and causal constraint, especially when paired with dynamic models that represent recursive regulation. For the AI scientist, SCMs are precision tools: valuable for pruning the explanatory space, but incomplete for modeling full systems.

Flux Balance Analysis (FBA) uses constraint-based optimization to infer metabolic fluxes under steady-state assumptions [65]. In Alzheimer's research, FBA helps test hypotheses about astrocyte–neuron metabolic coupling. Models show how astrocyte-produced lactate supports neuronal energy demands under aerobic glycolysis [66,67]. FBA enforces physical plausibility and checks biochemical consistency, offering more than statistical correlation. For the AI scientist, it's a principled method to assess whether observed metabolite patterns align with shuttle mechanisms like the ANLS. However, because FBA assumes steady state, it excels when conditions are stable but needs complementing with dynamic models in settings involving perturbations or time-dependent change.

Quantitative Systems Pharmacology (QSP) models use differential equations to simulate Alzheimer's-related pathways across compartments such as brain, CSF, and plasma. They encode production, aggregation, clearance, and drug responses (e.g., monoclonal antibodies) [68]. These models support hypothesis testing, dose optimization, and biomarker trajectory forecasting. For the AI scientist, QSP translates biological knowledge into simulation-ready form. However, QSP typically focuses on narrow pathways and lacks integration with broader homeostatic systems. As such, it is a powerful tool for scoped inquiries, but not a substitute for more comprehensive mechanistic models.

Reinforcement Learning (RL) enables machines to learn control policies through feedback—trial, correction, and policy refinement [69–71]. Layered RL architectures combine fast reflexes with slower

strategic control [72], mirroring biological regulation across timescales from ionic shifts to transcriptional changes. For the AI scientist, RL offers a model for digital twins that adapt over time, not just predict. Alzheimer's pathology emerges from regulatory failure, making RL's adaptive framing essential.

While RL could be used to optimize treatment strategies (e.g., dosing), this risks shallow gains unless guided by deeper constraints. We envision RL agents operating within digital twins that embed homeostatic principles across scales. Here, the reward function prioritizes long-term system stability, penalizing destabilizing trajectories, such as impaired ANLS or microglial lipid overload. In this role, RL becomes not just an optimizer but a discovery engine, probing which control policies sustain resilience. It shifts from reactive adjustment to active inference of system-level scaffolds that underlie progression and recovery.

Box 2 outlines core use cases for AI in Alzheimer's research, illustrating how different methods—hypothesis generation, causal inference, dynamic modeling—offer complementary strengths. Rather than exhaustively listing tools, we suggest how combining these approaches, often within neuro-symbolic frameworks or digital twins, can shift the field from fragmented associations toward mechanistic, testable understanding.

4.5. *The AI scientist as conductor, digital twins as the orchestra*

Each AI approach illuminates only part of the Alzheimer's puzzle. Causal graphs clarify direction but miss feedback. Dynamic models capture adaptation but require constraint. Reinforcement learning discovers control policies but depends on environments that biology rarely provides. Language models synthesize evidence but often without mechanism. No single method is sufficient. The AI Scientist's role is conceptual: an orchestrator who selects, sequences, and integrates diverse methods to produce coherent, mechanistic hypotheses. This orchestration is similar to how current agentic AI systems coordinate multiple tools to accomplish goals. However, unlike today's agents, which operate without a world model, the AI Scientist must root its reasoning in causal and dynamic realities.

Digital twins provide the formal setting for this orchestration. They are not metaphorical scientists but structured environments where causal graphs, dynamic models, reinforcement learning policies, and statistical learners interact within a living representation of disease. Built on dynamic modeling, twins capture feedback, adaptation, and homeostasis, linking molecular to population scales. In this way, they bridge collective and individual variation, showing how mechanisms of risk or resilience emerge while keeping insight tethered to shared system dynamics. The following section details how such twins, grounded in systems engineering and multiscale integration, can supply the scaffolding needed to close persistent gaps in Alzheimer's research.

5. From orchestration to implementation: digital twins as the framework for reconciliation

5.1. *What we mean by a digital twin*

The idea of a "digital twin" traces its origins to aerospace and manufacturing: NASA's early use of simulators to mirror spacecraft behavior, especially during Apollo missions, laid the foundation for today's virtual replicas of complex systems. By the 2010s, digital twins had evolved into real-time, physics-based models used for predictive, "personalized" maintenance in jet engines and aircraft, continuously assimilating sensor data to forecast failures and optimize performance [73]. The American Institute of Aeronautics and Astronautics 2020 position paper reinforces this lineage, recognizing digital twins as integral decision-support tools in safety-critical environments [74]. In these settings, the analogy becomes clear: while every jet engine begins as a standardized design, each experiences different conditions (stress cycles,

Box 2
Use Cases for Al in Alzheimer's Research

Box 2: Use Cases for AI in Alzheimer's Research

Use Case	Description	Example Applications
Hypothesis Generation	Using AI to identify connections, generate new questions, or propose mechanisms. LLMs are powerful for breadth—surfacing broad but less rigorous hypotheses—while neuro-symbolic reasoning and digital twins enable more grounded, mechanistic hypothesis generation.	LLMs suggesting underexplored gene–lipid associations; neuro-symbolic models proposing causal roles for astrocyte–neuron lactate shuttle breakdown; digital twins identifying testable intervention points.
Data Imputation & Integration	Filling missing values and harmonizing heterogeneous data across scales and cohorts.	Inferring unmeasured biomarkers from omics; aligning imaging, proteomics, and cognitive data across studies; harmonizing single-cell and bulk transcriptomic signals.
Simulation & Dynamics	Modeling disease progression, interventions, and compensatory processes over time, incorporating feedback and adaptation.	QSP models simulating amyloid/tau interventions across APOE genotypes; RL-inspired architectures modeling adaptive breakdown of homeostasis; digital twins forecasting destabilization trajectories under different perturbations.
Causal Inference & Constraint Reasoning	Using formal frameworks to infer directionality and enforce physical or biochemical constraints.	SCMs and Mendelian Randomization identifying causal drivers vs. passengers in lipid metabolism; FBA enforcing stoichiometric and flux constraints in neuron–glia metabolism.
Reconciliation of Evidence	Integrating diverse, sometimes conflicting evidence into mechanistically coherent frameworks.	Digital twins unifying omics, imaging, and clinical trajectories; SCM + dynamic models reconciling heterogeneous biomarker findings across cohorts.
Personalized Forecasting	Generating individualized predictions of disease course or treatment response by continuously updating models with patient data.	Digital twins forecasting cognitive decline trajectories under lifestyle or therapeutic interventions; personalized intervention planning guided by dynamic model calibration.

maintenance regimes, environmental exposures) that gradually introduce variation. A fleet of engines thus becomes a distribution of outcomes, where twins both capture the shared physics of the system and track how individualized histories shape divergence over time.

Recent years have seen a surge of interest in digital twin technologies across biomedical domains, with a few broad classes beginning to emerge [75]. The first includes QSP-based digital twins, which model pharmacokinetics and pharmacodynamics at the individual level to simulate treatment effects and optimize dosing regimens. For example, Maharjan et al. describe how digital twins can transform pharmaceutical pipelines from discovery through post-market surveillance [76], while Susilo et al. demonstrate their utility in characterizing clinical dose-response relationships in rare diseases [77]. The second class focuses on neurology-specific digital twins, which aim to model dynamic disease trajectories for individuals. Fekonja et al. propose digital twins for personalized brain modeling [78], and Cen et al. show how such twins can track disease-specific atrophy in multiple sclerosis [79]. A third, increasingly visible application is the use of digital twins to optimize clinical trial design, where simulated populations are used to reduce the size of control arms, as noted in recent systems pharmacology literature [80,81].

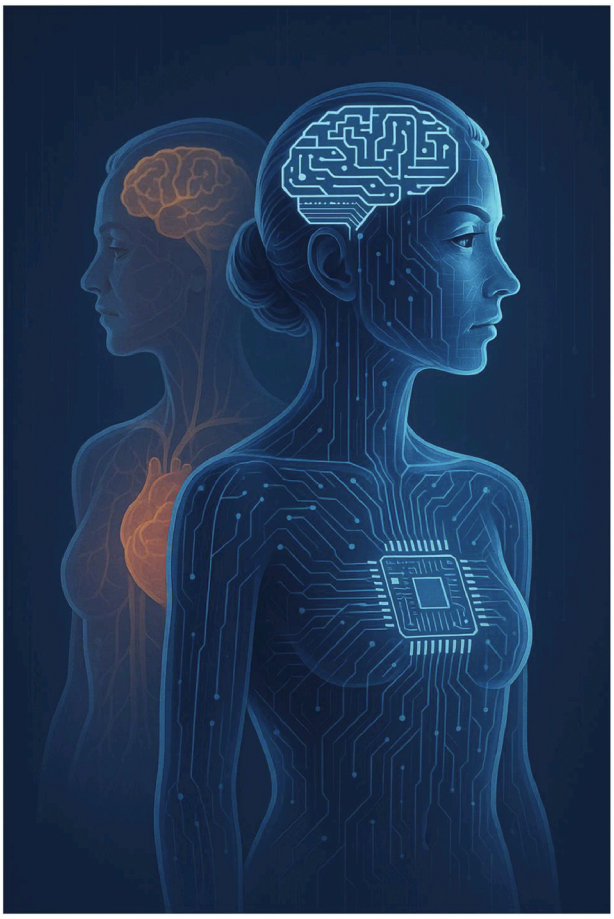
By contrast, the digital twin framework we propose differs in several critical respects. Rather than focusing on drug-specific responses or individualized prognostication, a reconciliation-centered model centers on homeostatic regulation and causal inference. Its primary goal is to reconstruct multiomic, imaging, and clinical data into mechanistic models of disease. We envision twins designed to not only simulate outcomes but also to test hypotheses about biological function. While QSP and neurology twins typically prioritize predictive accuracy or trial simulation, our approach aims to capture and resolve internal inconsistencies across diverse data modalities. This is key to uncovering the underlying rules that govern system-level dysfunction. Such a broad scope is especially critical in Alzheimer's disease, where diverse and potentially conflicting findings across data types, such as proteomics and imaging, may have limited individual contribution potential towards true causal understanding.

In the context of Alzheimer's, we adopt that best-practice foundation, but repurpose it for scientific reconciliation, not just operational forecasting. Our definition is more stringent: we envision digital twins as dynamic, mechanistic models that evolve with longitudinal data, enforce conservation and homeostatic constraints, and are built for causal inference rather than prediction alone. In doing so, we retain the proven architecture of adaptation and feedback from aerospace but deploy it as a reconciliation engine, integrating heterogeneous data and mechanistic priors into coherent, evolving models suited to unraveling the complexities of Alzheimer's disease.

5.2. Organizing principles

Homeostasis: Living systems evolved for stability, not disease. Homeostasis provides both the goal state and the constraints that digital twins must encode. This includes conservation laws (mass, energy, flux) and feedback loops governing lipid transport, neuronal excitability, and immune signaling. Without grounding in these regulatory architectures, models may generate statistically plausible but biologically invalid results.

Destabilization: Disease reflects progressive erosion of regulatory balance. In Alzheimer's, destabilization may stem from impaired lipid clearance, disrupted astrocyte–neuron energy coupling, or unchecked inflammation. Digital twins must represent not only intact feedback systems, but also how they degrade over time and stress. Capturing this dynamic misalignment enables models to explain how vulnerability accumulates and leads to pathology.



ChatGPT-generated representation of a digital twin.

Multiscale Reconciliation: Alzheimer's spans molecules, cells, circuits, and populations. Twins must integrate across these layers: molecular priors (e.g., APOE and lipid metabolism), cellular behaviors (e.g., microglial flux), systems physiology (e.g., glymphatic clearance), and population trajectories. This is more than model nesting—it requires coherence across scales, where cellular dynamics are constrained by cohort-level biomarkers and vice versa. Without this, the landscape fragments into disciplinary silos.

Temporal Alignment: Biological processes unfold on vastly different time scales: milliseconds (ion currents), hours (metabolism), days (immune shifts), and years (atrophy, plaques). Alzheimer's arises not from a single failure but from mismatches across these rhythms—-when fast neuronal needs outpace slower astrocyte support, or when debris accumulates over decades. Twins must simulate fast, medium, and slow processes together, showing how asynchrony drives system instability. Time becomes as central as scale.

5.3. Architectural roles

With foundational principles defined, we now explore how AI can operationalize them. In this framework, the AI Scientist coordinates a layered system of reasoning roles. The digital twin provides the structure within which this coordination unfolds. It is not a single model, but an environment composed of three core functions: the orchestrator, the enforcer, and the architect.

The orchestrator manages knowledge flow. It selects which tools to use, interprets their outputs, and ensures consistency across the system. This role can be performed by language models enhanced with retrieval tools and code execution, allowing them to synthesize literature, modeling results, and statistical outputs into coherent narratives. The orchestrator also tracks uncertainty, noting which results are supported

by data, which rely on assumptions, and which remain unresolved. It routes this information between the enforcer and architect to support adaptive model refinement. While it does not generate mechanistic insight on its own, the orchestrator is essential for aligning evidence with evolving hypotheses.

The enforcer is responsible for simulation, optimization, and learning under constraint. This includes reinforcement learning, flux balance models, and dynamic systems. Its job is to evaluate how well candidate explanations hold up against biological constraints and available data. In wicked domains like biology, where signals are intermittent, indirect, or confounded by observational limits, the enforcer plays the critical role of pushing models until they either hold or break under the weight of evidence. In early stages, this is a human-guided process. Over time, the enforcer may gain autonomy, identifying new data sources or proposing targeted experiments to resolve conflict. It plays a central role in iterative discovery, using contradiction as a signal to refine or revise current understanding.

The architect designs and compares alternative model structures. Since no model can capture biology in full, the goal is to create multiple abstractions and evaluate how well each explains the data. This process draws on structured causal models, symbolic reasoning, and probabilistic tools. The architect also tracks assumptions and priors, helping clarify what each model implies. When gaps are exposed, the architect assists in generating targeted hypotheses that can be tested experimentally. While AI may suggest plausible experiments, matching those proposals to biologically meaningful systems remains a task that often requires domain expertise. A well-designed twin helps prioritize experiments by their likely impact and feasibility.

This system is designed to evolve. This progression aligns with systems engineering principles, where contradictions in the model prompt the acquisition of new, discriminative data. The goal is not to explain everything at once but to identify the most strategic gaps and design targeted experiments that reduce uncertainty. Reconciliation is a central function, allowing twins to integrate new data while also addressing the backlog of conflicting results. Taken together, the orchestrator ensures clarity, the enforcer grounds models in reality, and the architect explores structural alternatives. Their interaction forms a dynamic reasoning engine that transforms digital twins into living frameworks for discovery.

When reconciliation exposes a mechanistic gap, the next step is translating that gap into a tractable experiment. Today, this process is human-guided: researchers identify where a model lacks constraint and propose perturbations or observations to test it. LLMs can assist by suggesting plausible in vitro or in vivo experiments given sufficient context, and their utility in this domain is likely to grow. Still, designing testable hypotheses depends not just on mechanistic reasoning but on choosing a biologically relevant model system, which remains highly context specific. In fields like neurobiology, where cell type, spatial location, and timing influence results, domain expertise remains essential. A key function of digital twins is to help prioritize among candidate experiments by estimating their informativeness, feasibility, and relevance to the broader system.

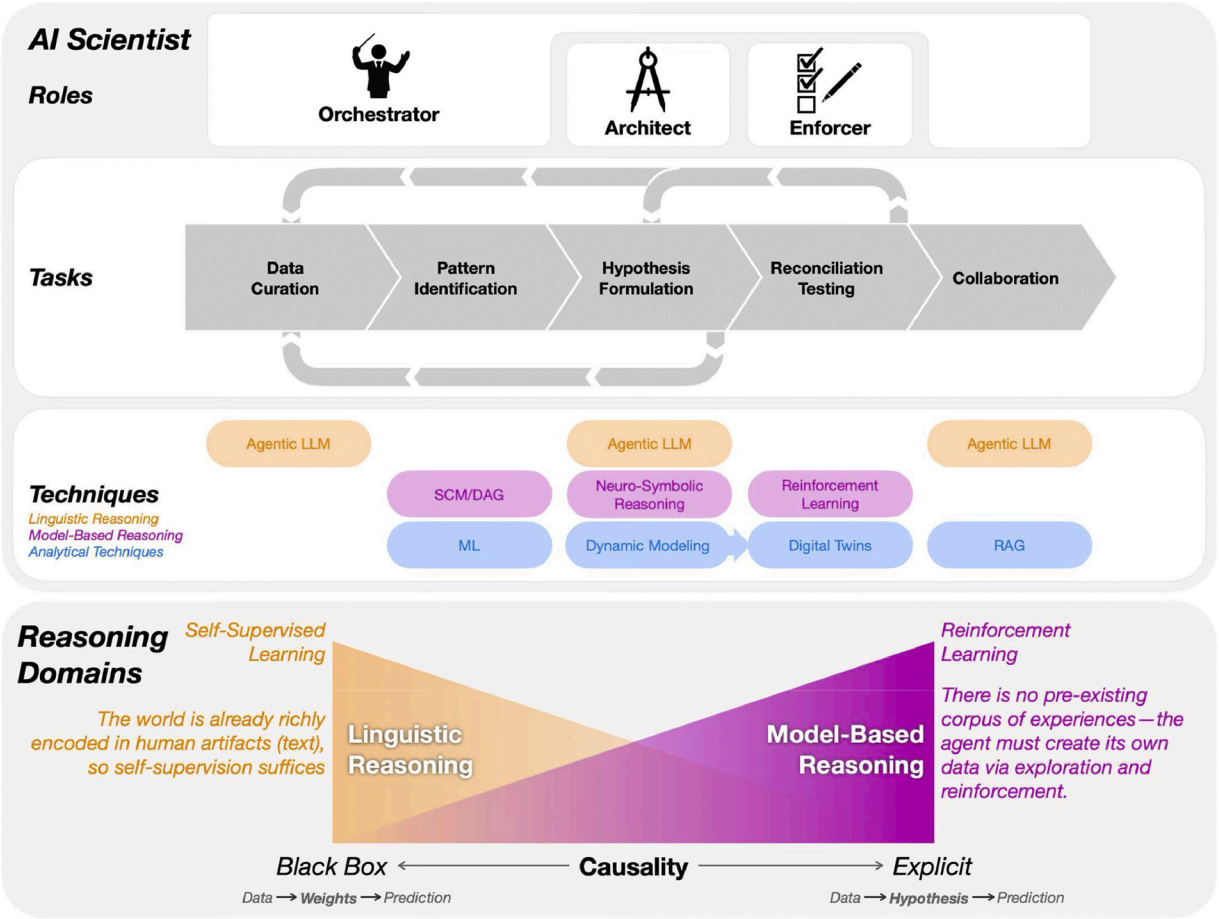


Fig. 1. Mapping AI Approaches for Reconciliation in Alzheimer's Research. Illustration of how diverse AI methods align with stages of the scientific process—from data curation and pattern identification to hypothesis formulation, reconciliation testing, and collaboration. Large language models, causal graphs, dynamic models, reinforcement learning, and neuro-symbolic reasoning each occupy different roles, but their greatest value emerges when coordinated iteratively. The framework highlights where linguistic reasoning suffices, where model-based reasoning and feedback are essential, and how iteration across methods enables reconciliation of partial evidence into mechanistic insight.

Just as experiments refine the model, models must be built to absorb those refinements. In our framework, this is not a technical hurdle but a foundational feature. Reconciliation-driven twins are structured to remain flexible at their edges, allowing new data to update causal links without full retraining. This refinement can be manual or automated depending on how constrained the relevant variables are. Yet the bigger challenge is not integrating future data, but absorbing the vast backlog of existing results. A reconciliation-first architecture helps triage which gaps warrant new experiments and avoids redundant or low-value inquiry.

These methods reinforce that the AI scientist's task is not simply tool selection but orchestration of iterative cycles of hypothesis, testing, and refinement. Fig. 1 outlines this process, showing how different AI approaches interact and where iteration drives progress toward mechanistic understanding.

The purpose of this two-level approach, which combines an AI scientist with a biological digital twin, is not simply to explain Alzheimer's in theory. It is designed to support better decisions about treatment, trials, and care. Mechanistic clarity has value only if it improves the ability to act under uncertainty. Decision theory provides the formal bridge between understanding and action by offering a framework to evaluate options through probabilities, outcomes, and utilities. In Alzheimer's, this means structuring an iterative relationship between scientific insight and clinical intervention. Mechanistic models inform therapeutic strategies, which then generate new data. That data refines the models, sharpening their predictions and explanations. This creates a self-reinforcing cycle connecting discovery, development, and practice.

In drug development, this approach enables rational trial design, more precise recruitment, and model systems that better reflect underlying biology. Instead of broad statistical averages, decision-aware twins can simulate how interventions affect specific genotypes or stages of disease, helping avoid costly failures and improving targeting. For clinicians, the same principles provide decision support grounded in mechanism, not just association, allowing for better anticipation of treatment outcomes. In this way, decision theory links the scientific goal of understanding disease with the practical goal of guiding action, making reconciliation both a scientific tool and a translational engine for progress.

6. Conclusion

6.1. Reconciliation as the central task

Artificial intelligence has the potential to transform Alzheimer's research, not through any single breakthrough, but by integrating diverse methods into a coherent system. Language models, retrieval tools, reinforcement learning, structured causal models, flux balance analysis, and quantitative pharmacology each illuminate a different aspect of disease. Yet individually, they remain partial and insufficient. What is needed is reconciliation: a way to align these tools into frameworks that are explanatory, testable, and faithful to biology.

Focusing on reconciliation introduces real challenges. Digital twins are computationally demanding and rely on rich, longitudinal data. Mechanistic models, while interpretable, can still embed flawed assumptions or biases. As AI-generated hypotheses enter clinical research, ethical and regulatory concerns increase, particularly around validation, transparency, and accountability. These challenges highlight the need for modular, interpretable systems that allow each component to be tested and trusted independently.

Digital twins provide the biological foundation for this reconciliation. Rather than static forecasts, they represent evolving, mechanistic models that span levels of scale and time: from cells to circuits, from patients to populations, from rapid feedback to long-term adaptation [82]. Guided by principles of homeostasis, destabilization, and multiscale coherence, digital twins operate as living systems—continually adjusting predictions, reconciling contradictions, and exposing knowledge gaps. Their effectiveness depends on the architecture built by the AI Scientist: orchestrators that manage reasoning, enforcers that stress-test hypotheses, and architects that explore and refine model structures. The AI Scientist supplies the reasoning infrastructure; the digital twins embody its evolving output.

The lesson of AlphaFold is that iteration, paired with constraint, can resolve seemingly intractable problems. The lesson of Alzheimer's is that no single method will succeed when biology is irregular in its signals, complex in its interactions, and shifting over time. By treating reconciliation as the central goal, and digital twins as the scaffolding that supports it, we can move from fragmented knowledge toward integrated understanding. If successful, this approach will not only close long-standing gaps in Alzheimer's but also offer a new model for scientific discovery—one in which AI acts not as a black box, but as a transparent, evolving system for mechanistic insight.

During the preparation of this work the authors used ChatGPT and Claude in order to research topics, gather information, and improve linguistic clarity and brevity. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

(Note on references: In the rapidly-developing field of AI, some papers published as preprints are highly influential and well-cited, and reveal important new insight into methods and capabilities. But publication in a reviewed journal may be done much later or never sought at all. We include some of these preprints in the following references.)

Declaration of competing interest

I, **Cory Funk**, am a cofounder of **Fulcrum Neuroscience**, a biotechnology company developing computational and mechanistic approaches for understanding and treating Alzheimer's disease. In this capacity, I hold an equity interest in the company, receive financial compensation, and am associated with intellectual property (patents, licenses, or royalties) related to its work.

These relationships represent potential financial conflicts of interest relevant to the subject matter of my research. I disclose them here in full recognition of the importance of transparency and to allow editors and reviewers to evaluate the manuscript with this context in mind.

References

[1] Dyck CH van, Swanson CJ, Aisen P, et al. Lecanemab in Early Alzheimer's Disease. New Engl J Med 2022;388:9–21.
[2] Mintun MA, Lo AC, Evans CD, et al. Donanemab in Early Alzheimer's Disease. New Engl J Med 2021;384:1691–704.
[3] Alves F, Kalinowski P, Ayton S. Accelerated brain volume loss caused by anti–β-amyloid drugs. Neurology 2023;100:e2114–24.
[4] Mueller SG, Weiner MW, Thal LJ, et al. Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's disease neuroimaging initiative (ADNI). Alzheimer's Dement 2005;1:55–66.
[5] Morris JC, Aisen PS, Bateman RJ, et al. Developing an international network for Alzheimer's research: the Dominantly Inherited Alzheimer Network. Clin Investig 2012;2:975–84.
[6] Bateman RJ, Xiong C, Benzinger TLS, et al. Clinical and Biomarker Changes in Dominantly Inherited Alzheimer's Disease. N Engl J Med 2012;367:795–804.
[7] Bennett DA, Buchman AS, Boyle PA, et al. Religious orders study and rush memory and aging project. J Alzheimer's Dis 2018;64:S161–89.
[8] Imam F, Saloner R, Vogel JW, et al. The Global Neurodegeneration Proteomics Consortium: biomarker and drug target discovery for common neurodegenerative diseases and aging. Nat Med 2025;31:2556–66.
[9] Health F for the NI of. Alzheimer's Disease Neuroimaging Initiative (ADNI). fnih.org/our-programs/alzheimers-disease-neuroimaging-initiative-adni.
[10] Kuhn TS, Hawkins D. The Structure of Scientific Revolutions. Am J Phys 1963;31:554–5.
[11] Christensen CM, Raynor ME, McDonald R. What Is Disruptive Innovation? Harv Bus Rev 2015. https://hbr.org/2015/12/what-is-disruptive-innovation?.
[12] Koonin EV. Evolution of genome architecture. Int J Biochem cell Biol 2008;41:298–306.
[13] Heasley LR, Sampaio NMV, Argueso JL. Systemic and rapid restructuring of the genome: a new perspective on punctuated equilibrium. Curr Genet 2021;67:57–63.
[14] Wolfram S. *A New Kind of Science*. Champaign, IL, USA: Wolfram Media; 2002. https://www.wolframscience.com/nks/.

[15] Lopera F, Marino C, Chandrahas AS, et al. Resilience to autosomal dominant Alzheimer's disease in a Reelin-COLBOS heterozygous man. Nat Med 2023:1–10.
[16] Llibre-Guerra JJ, Fernandez MV, Joseph-Mathurin N, et al. Longitudinal analysis of a dominantly inherited Alzheimer disease mutation carrier protected from dementia. Nat Med 2025;31:1267–75.
[17] Arboleda-Velasquez JF, Lopera F, O'Hare M, et al. Resistance to autosomal dominant Alzheimer's disease in an APOE3 Christchurch homozygote: a case report. Nat Med 2019;25:1680–3.
[18] Buganim Y, Faddah DA, Jaenisch R. Mechanisms and models of somatic cell reprogramming. Nat Rev Genet 2013;14:427–39.
[19] Zhu F, Nie G. Cell reprogramming: methods, mechanisms and applications. Cell Regen 2025;14:12.
[20] Martik ML, Lyons DC, McClay DR. Developmental gene regulatory networks in sea urchins and what we can learn from them. F1000Res 2016;5:F1000. Faculty Rev-203.
[21] Schüpbach T. Genetic Screens to Analyze Pattern Formation of Egg and Embryo in Drosophila: a Personal History. Annu Rev Genet 2019;53:1–18.
[22] Lynch JA, Roth S. The evolution of dorsal–ventral patterning mechanisms in insects. Genes Dev 2011;25:107–18.
[23] Doshi-Velez F, Kim B. Towards A Rigorous Science of Interpretable Machine Learning. ArXiv 2017. https://doi.org/10.48550/arxiv.1702.08608. Epub ahead of print.
[24] Arrieta AB, Díaz-Rodríguez N, Ser JD, et al. Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 2020;58:82–115.
[25] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019;1:206–15.
[26] Johnston WJ, Fusi S. Abstract representations emerge naturally in neural networks trained to perform multiple tasks. Nat Commun 2023;14:1040.
[27] Vafa K., Chen J.Y., Rambachan A., et al. Evaluating the World Model Implicit in a Generative Model. In: *Advances in neural information processing systems*. Curran Associates, Inc., pp. 26941–26975, 2024.
[28] Pound P, Bracken MB. Is animal research sufficiently evidence based to be a cornerstone of biomedical research? BMJ: Br Méd J 2014;348:g3387.
[29] Lancaster MA, Knoblich JA. Generation of cerebral organoids from human pluripotent stem cells. Nat Protoc 2014;9:2329–40.
[30] Wang Z, Gao C, Glass LM, et al. Artificial intelligence for in silico clinical trials: a review. ArXiv 2022. https://doi.org/10.48550/arxiv.2209.09023. Epub ahead of print.
[31] Sinisi S, Alimguzhin V, Mancini T, et al. Optimal personalised treatment computation through in silico clinical trials on patient digital twins. Fundam Informaticae 2020;174:283–310.
[32] McCoy RT, Yao S, Friedman D, et al. Embers of autoregression show how large language models are shaped by the problem they are trained to solve. Proc Natl Acad Sci 2024;121:e2322420121.
[33] Hao S, Gu Y, Ma H, et al. Reasoning with language model is planning with world model. ArXiv 2023. https://doi.org/10.48550/arxiv.2305.14992. Epub ahead of print.
[34] Shojaee P, Mirzadeh I, Alizadeh K, et al. The illusion of thinking: understanding the strengths and limitations of reasoning models via the lens of problem complexity. Apple Machine Learn Res 2025. https://machinelearning.apple.com/research/illusion-of-thinking.
[35] Zhu Y, Yuan H, Wang S, et al. Large language models for information retrieval: a survey. ArXiv 2023. https://doi.org/10.48550/arxiv.2308.07107. Epub ahead of print.
[36] Ponce-Bobadilla AV, Schmitt V, Maier CS, et al. Practical guide to SHAP analysis: explaining supervised machine learning model predictions in drug development. Clin Transl Sci 2024;17:e70056.
[37] Leist AK, Klee M, Kim JH, et al. Mapping of machine learning approaches for description, prediction, and causal inference in the social and health sciences. Sci Adv 2022;8:eabk1942.
[38] Freiesleben T, König G, Molnar C, et al. Scientific inference with interpretable machine learning: analyzing models to learn about real-world phenomena. Minds Mach 2024;34:32.
[39] Roscher R, Bohn B, Duarte MF, et al. Explainable machine learning for scientific insights and discoveries. IEEE Access 2020;8:42200–16.
[40] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.
[41] Lipton ZC. The mythos of model interpretability. Commun ACM 2018;61:36–43.
[42] Chakraborti T, Sreedharan S, Zhang Y, et al. Plan explanations as model reconciliation: moving beyond explanation as soliloquy. Proc Twenty-Sixth Int Jt Conf Artif Intell 2017:156–63.
[43] Sreedharan S, Chakraborti T, Kambhampati S. Foundations of explanations as model reconciliation. Artif Intell 2021;301:103558.
[44] Garnelo M, Shanahan M. Reconciling deep learning with symbolic artificial intelligence: representing objects and relations. Curr Opin Behav Sci 2019;29:17–23.
[45] Lake BM, Ullman TD, Tenenbaum JB, et al. Building machines that learn and think like people. Behav Brain Sci 2017;40:e253.
[46] Tenenbaum JB, Kemp C, Griffiths TL, et al. How to grow a mind: statistics, structure, and abstraction. Science (1979) 2011;331:1279–85.
[47] Pearl J, Mackenzie D. *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books; 2018.
[48] Lu C, Lu C, Lange RT, et al. The AI scientist: towards fully automated open-ended scientific discovery. ArXiv 2024. https://doi.org/10.48550/arxiv.2408.06292. Epub ahead of print.
[49] Yamada Y, Lange RT, Lu C, et al. The AI scientist-v2: workshop-level automated scientific discovery via agentic tree search. ArXiv 2025. https://doi.org/10.48550/arxiv.2504.08066. Epub ahead of print.
[50] Gottweis J, Weng W-H, Daryin A, et al. Towards an AI co-scientist. ArXiv 2025. https://doi.org/10.48550/arxiv.2502.18864. Epub ahead of print.
[51] Kokotajlo D, Alexander S, Larsen T, et al. AI 2027. AI futures project. 2025. https://ai-2027.com/.
[52] Lazebnik Y. Can a biologist fix a radio?—Or, what I learned while studying apoptosis. Cancer Cell 2002;2:179–82.
[53] Resnik DB, Shamoo AE. Reproducibility and research integrity. Account Res 2017;24:116–23.
[54] Ioannidis JPA. Why most published research findings are false. PLoS Med 2005;2:e124.
[55] Nosek BA, Errington TM. What is replication? PLoS Biol 2020;18:e3000691.
[56] Drummond E, Wisniewski T. Alzheimer's disease: experimental models and reality. Acta Neuropathol 2017;133:155–75.
[57] Bellenguez C, Küçükali F, Jansen IE, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat Genet 2022;54:412–36.
[58] Wightman DP, Jansen IE, Savage JE, et al. Largest GWAS (N=1126,563) of Alzheimer's disease implicates microglia and immune cells. medRxiv 2020.2020.11.20.20235275.
[59] Epstein D. Range: *why generalists triumph in a specialized world*. New York: Riverhead Books, 2019.
[60] Silver D, Huang A, Maddison CJ, et al. Mastering the game of Go with deep neural networks and tree search. Nature 2016;529:484–9.
[61] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596:583–9.
[62] Ruff KM, Pappu RV. AlphaFold and Implications for Intrinsically Disordered Proteins. J Mol Biol 2021;433:167208.
[63] Staff V. Scaling agentic ai safely — and stopping the next big security breach. 2025. https://venturebeat.com/ai/scaling-agentic-ai-safely-and-stopping-the-next-big-security-breach/.
[64] Boehm FJ, Zhou X. Statistical methods for Mendelian randomization in genome-wide association studies: a review. Comput Struct Biotechnol J 2022;20:2338–51.
[65] Thiele I, Swainston N, Fleming RMT, et al. A community-driven global reconstruction of human metabolism. Nat Biotechnol 2013;31:419–25.
[66] Çakır T, Alsan S, Saybaşılı H, et al. Reconstruction and flux analysis of coupling between metabolic pathways of astrocytes and neurons: application to cerebral hypoxia. Theor Biol Méd Model 2007;4:48.
[67] Kobayashi T, Yoshizawa K. Optimization algorithm for feedback and feedforward policies towards robot control robust to sensing failures. Robomech J 2022;9:18.
[68] Ramakrishnan V, Friedrich C, Witt C, et al. Quantitative systems pharmacology model of the amyloid pathway in Alzheimer's disease: insights into the therapeutic mechanisms of clinical candidates. CPT: Pharmacomet Syst Pharmacol 2023;12:62–73.
[69] Ali M, Giri S, Liu S, et al. Digital twin-enabled real-time control in robotic additive manufacturing via soft actor-critic reinforcement learning. ArXiv 2025. https://doi.org/10.48550/arxiv.2501.18016. Epub ahead of print.
[70] Kober J, Bagnell JA, Peters J. Reinforcement learning in robotics: a survey. Int J Robotics Res 2013;32:1238–74.
[71] Schena L, Marques PA, Poletti R, et al. Reinforcement Twinning: from digital twins to model-based reinforcement learning. J Comput Sci 2024;82:102421.
[72] Goel G, Chen N, Wierman A. Thinking Fast and Slow. ACM SIGMETRICS Perform Eval Rev 2017;45:27–9.
[73] Staff S. Digital twin evolution: a 30-Year journey that changed industry. 2025. https://www.simio.com/digital-twin-evolution-a-30-year-journey-that-changed-industry/?.
[74] Committee ADEI. *Digital Twin: Definition & Value*. American Institute of Aeronautics and Astronautics (AIAA) and AIA; 2020. https://aiaa.org/wp-content/uploads/2024/12/digital-twin-institute-position-paper-december-2020.pdf?.
[75] Katsoulakis E, Wang Q, Wu H, et al. Digital twins for health: a scoping review. npj Digit Med 2024;7:77.
[76] Maharjan R, NAh Kim, Kim KH, et al. Transformative roles of digital twins from drug discovery to continuous manufacturing: pharmaceutical and biopharmaceutical perspectives. Int J Pharm: X 2025;10:100409.
[77] Susilo ME, Li C, Gadkar K, et al. Systems-based digital twins to help characterize clinical dose–response and propose predictive biomarkers in a Phase I study of bispecific antibody, mosunetuzumab, in NHL. Clin Transl Sci 2023;16:1134–48.
[78] Fekonja LS, Schenk R, Schröder E, et al. The digital twin in neuroscience: from theory to tailored therapy. Front Neurosci 2024;18:1454856.
[79] Cen S, Gebregziabher M, Moazami S, et al. Toward precision medicine using a "digital twin" approach: modeling the onset of disease-specific brain atrophy in individuals with multiple sclerosis. Sci Rep 2023;13:16279.
[80] Mann DL. The Use of Digital Healthcare Twins in Early-Phase Clinical Trials Opportunities, Challenges, and Applications. JACC: Basic Transl Sci 2024;9:1159–61.
[81] Vidovszky AA, Fisher CK, Loukianov AD, et al. Increasing acceptance of AI-generated digital twins through clinical trial applications. Clin Transl Sci 2024;17:e13897.
[82] Barbiero P, Torné RV, Lió P. Graph representation forecasting of patient's medical conditions: towards a digital twin. 2020. https://arxiv.org/abs/2009.08299?. accessed October 9, 2025.