

# Knowledge-Centric AI for Scientific Discovery

Carla P. Gomes

*Recent breakthroughs in artificial intelligence have been driven primarily by data-centric approaches, including deep learning and large language models. However, despite their remarkable success, relying solely on purely data-driven methods has intrinsic limitations for scientific discovery. In this essay, I argue for knowledge-centric AI, reasoning grounded in scientific principles, domain structure, and constraints while learning from data. In our research at Cornell University and through interdisciplinary collaborations with researchers across institutions, this agenda has been shaped by data-limited scientific challenges, which has led us to develop knowledge-centric approaches, such as deep reasoning networks, which build interpretable, domain-aligned latent spaces and enforce differentiable constraints end-to-end. Knowledge-centric AI has enabled scientific advances, from automated crystal-structure phase mapping that uncovered high-performing alloyed mixtures to joint species–distribution models that inform conservation, while also delivering general AI methodologies. By aligning with the scientific method and integrating reasoning, learning, and principled experimentation, knowledge-centric AI can catalyze scientific innovation. In our work, this perspective is grounded in computational challenges central to a sustainable future.*

Through breakthroughs such as AlphaGo, AlphaFold, and ChatGPT, artificial intelligence is at the cusp of transforming our lives.<sup>1</sup> These are exciting times, as the field moves closer to artificial general intelligence, replicating general purpose human cognition.<sup>2</sup> At the same time, humanity faces significant global challenges, including climate change and biodiversity loss, which threaten ecosystems and human well-being worldwide. Recognizing these issues, the United Nations' 2030 Agenda for Sustainable Development formulated seventeen Sustainable Development Goals to end poverty, protect the planet, and ensure prosperity for all.<sup>3</sup> Artificial intelligence holds great promise for advancing such goals, but realizing its positive impact requires deliberate effort and careful stewardship. Since the late 2000s, we have helped establish and nurture the field of computational sustainability with the belief that computational scientists can and should tackle societal and environmental problems while simultaneously driving founda-

tional advances in computer science.<sup>4</sup> Our focus is on developing and advancing AI methodologies to accelerate scientific discovery to help address sustainability challenges.

Recent AI breakthroughs have been driven by data-centric approaches, notably deep learning and large language models, but relying on these alone has intrinsic limitations for scientific discovery.<sup>5</sup> Here, I argue that knowledge-centric AI is key to advancing scientific discovery, as it addresses two unique challenges for scientific discovery and exploration.

First, high-quality experimental data are often time-consuming to obtain and interpret, leading to relatively data-limited settings. Second, one needs to leverage existing scientific laws and principles, which allow reaching far beyond any specific training distribution. For example, Newton's laws of motion and gravity were based on experiments and observations conducted on Earth. The leap to apply the same laws and principles to celestial objects is a beautiful example of the ability of scientific laws to generalize across vastly different contexts. Scientists are constantly pursuing such general principles and insights. They use existing scientific theories to guide their exploration – looking for small inconsistencies with experimental data to suggest modifications and refinements – to enable broader applicability of the theories. For instance, Einstein's work on relativity adapts Newton's laws to settings in which speeds approach the speed of light and extreme gravitational forces affect Euclidean notions of time and space. Similarly, AI-driven scientific discovery needs to leverage existing scientific laws and theories to uncover new phenomena and develop new theories. This essay emphasizes the power of knowledge-centric and first-principles reasoning: by integrating reasoning, learning, and experimentation, artificial intelligence can become a catalyst for discovery.

Science represents one of the purest expressions of human intellect. Psychologist Daniel Kahneman characterizes human thinking as comprising two systems.<sup>6</sup> System 1 is fast, automatic, and pattern driven, excelling at perception and intuitive recognition. In contrast, System 2 is slow, deliberate, and responsible for logical reasoning and complex decision-making. Over the past decade, AI has made spectacular progress on System 1-like capabilities through data-driven approaches. However, scientific discovery leans heavily on System 2, knowledge-driven reasoning, wherein limited data are enriched with first-principles reasoning about prior scientific knowledge.<sup>7</sup>

Scientific AI agents (SciAIs) – artificial systems that can devise, conduct, and analyze scientific research with some degree of autonomy – should therefore combine data-driven learning with explicit knowledge-driven reasoning. The line between data and knowledge can be subtle, but it has critical consequences. Crucially, the belief that “enough data” can fully replace knowledge is misguided. The conceptual world is combinatorial and complex, and many scientific regimes are data limited

but knowledge rich. To produce reliable, novel insights, SciAIs need representations that enable explanation, verification, and generalization.

SciAIs should emulate the scientific process, blending multimodal learning with principled reasoning, choosing representations that scale, and manipulating knowledge effectively. They must predict and discover well beyond their training distribution, interpret results, and identify causal relationships. Achieving true generalization and creativity in science requires scientific common sense and deductive, counterfactual, and abductive reasoning, supported by domain-specific inductive biases and hybrid data-knowledge methods that adapt to scarce or atypical data.

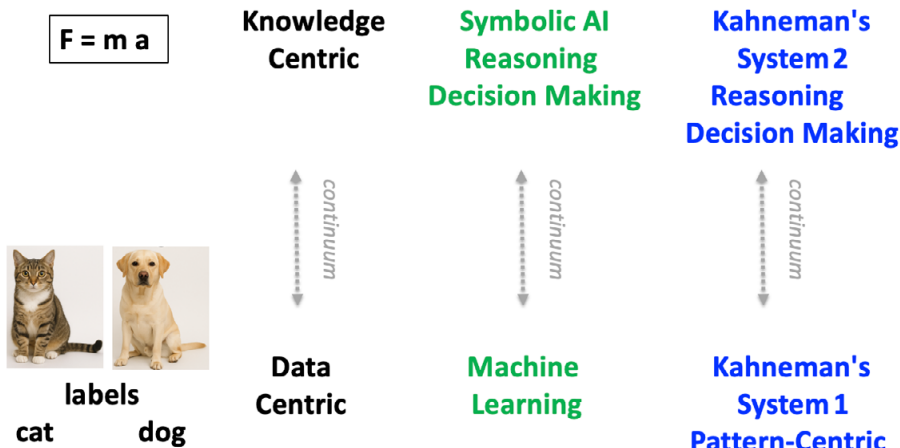
**R**easoning is the process of deriving new conclusions from existing knowledge. It makes implicit knowledge explicit and operates on representations of knowledge, such as symbols, graphs, equations, or rules, following principled steps that preserve truth and quantify uncertainty. Reasoning techniques are central to System 2–style cognition, grounded in first principles, rigorous inference, knowledge synthesis, and hypothesis generation.

While data-driven approaches map inputs to outputs through statistical regularities, reasoning manipulates knowledge to derive new facts or hypotheses. Together, they form a continuum: from correlational mapping to explicit knowledge manipulation (Figure 1). At the data-driven end, models uncover statistical regularities and largely interpolate within the training distribution. Moving toward reasoning, outputs gain structure, domain rules are enforced, and causal dependencies become explicit and testable. Scientific and sustainability problems usually sit toward the reasoning side, where scarce data and rich prior knowledge call for hybrids that learn from data while reasoning about prior knowledge.

From early symbolic systems – such as the Logic Theorist for theorem proving; the General Problem Solver I (GPS-1) for understanding the processes required for human problem solving; and DENDRAL, which applied rule-based reasoning to analyze chemical data and infer molecular structures (one of the first AI systems for a scientific domain) – to modern neurosymbolic and probabilistic methods, AI has relied on reasoning and search to solve complex problems and guide discovery.<sup>8</sup> By integrating data-centric learning with logic, constraints, optimization, and symbolic representations, artificial intelligence can move toward genuine scientific insight. This shift to reasoning requires more explicit representations, greater use of prior knowledge and causal models, deeper planning, and the ability to verify rather than merely score solutions. In practice, one moves along this continuum by making (latent) states interpretable, encoding domain constraints and rules, adding search or external tools, training for consistency, and checking results with separate methods for validation.

In recent years, an emergent line of work has sought to explicitly integrate scientific knowledge, such as physics-based models and governing equations, into

Figure 1  
Data Versus Knowledge: Continuums from Correlational Mapping to Explicit Knowledge Manipulation



Source : Illustration provided by the author.

machine-learning frameworks, moving beyond purely data-driven approaches. These efforts encode prior knowledge through modified objectives, architectural constraints, or hybrid combinations of neural and process-based models, reflecting an important shift toward incorporating domain structure into learning systems. At the same time, this integration remains challenging: such models can be difficult to train, do not always fully satisfy underlying scientific laws, and often remain opaque. As a result, their ability to support interpretation and to yield new scientific insight remains, for now, an open challenge, one that may require tighter integration with symbolic and reasoning-based approaches.<sup>9</sup>

Large language models (LLMs), embodied by ChatGPT, Claude, Gemini, and DeepSeek, increasingly pair learning with reasoning and search techniques.<sup>10</sup> Examples include reinforcement learning to incorporate human feedback, as well as chain-of-thought (CoT), tree-of-thoughts (ToT), and graph-of-thoughts (GoT) frameworks for multiple step-by-step inference (a form of “what-if” exploration), to cover large search spaces of possible responses.<sup>11</sup> These developments suggest that advanced generative AI and LLMs are moving beyond responses purely based on the learned language modeling toward more deliberate, controlled reasoning and search. A complementary line uses LLMs to evolve algorithms under constraints. The models generate candidate code or proof steps. Verifiers and guardrails

such as tests, invariants, proof checkers, and resource budgets eliminate invalid options. An evaluation loop measures solution rate, runtime, and memory use to identify improvements. Search layers include population-based evolution, best-first or beam search over edits, Monte Carlo tree search or reinforcement learning for sequential choices, and bandit or Bayesian tuning, accepting only population variants that remain correct and faster. This template spans systems for algorithm discovery (AlphaEvolve), formal reasoning (AlphaProof), and the design of logical reasoning engines (SAT solvers).<sup>12</sup>

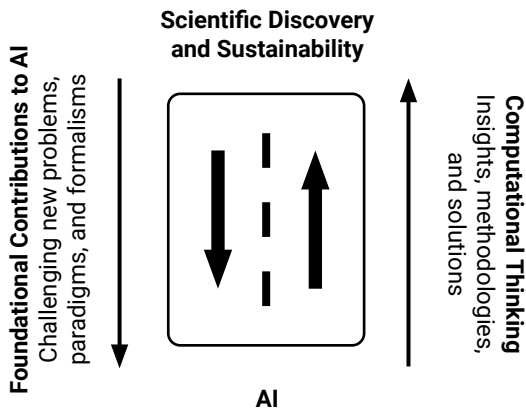
**A**rtificial intelligence for science is a two-way street. This metaphor stems from the core computer science notion of problem reduction, which amplifies the effectiveness of computational methods: Problems that may appear unrelated can often be solved using the same underlying computational approach.<sup>13</sup> The generality of computational models thus enables the transference of methodologies across domains, underpinning the two-way street shown in Figure 2. By developing general approaches to tackle scientific challenges in sustainability, we not only solve domain-specific problems, but we can also advance artificial intelligence through the creation of general, broadly applicable methodologies. I illustrate the two-way street metaphor with several examples from my lab's work at Cornell involving crystal-structure phase mapping and Deep Reasoning Networks.

*AI* → *Science*: Crystal-structure phase mapping is a central challenge in materials science, requiring inference of complex structures from X-ray diffraction (XRD) patterns. A rich body of prior knowledge exists about chemical systems and X-ray diffraction, including thermodynamic principles and Bragg's law, which describes how X-rays are diffracted by atomic planes in a crystal. We developed an approach to automate crystal-structure phase mapping that seamlessly combines learning from diffraction data with knowledge-centric reasoning about known crystal phases and the diffraction process, resulting in a significant improvement in the state-of-the-art of crystalline phase mapping.<sup>14</sup> This work exemplifies how AI techniques can be used to accelerate the scientific discovery process.

*Science* → *AI*: Solving crystal-structure phase mapping from X-ray diffraction required embedding scientific laws and constraints within a deep learning framework. This led to our development of the Deep Reasoning Networks (DRNets) framework: an interpretable deep architecture that integrates domain knowledge and enables differentiable reasoning.<sup>15</sup> DRNets combine an encoder that constructs a structured latent space, a reasoning module that enforces domain rules, and a decoder that injects background knowledge. During training, the desired interpretation emerges in the latent space, guided by a combination of unlabeled data with scientific domain knowledge and constraints. The modular design enables the integration of knowledge-centric reasoning with data-centric learning, adaptable to a wide range of scientific and engineering tasks.

Figure 2

The Two-Way Street Between AI and Scientific Discovery



AI enables new approaches for scientific discovery to address challenges in sustainability, and scientific sustainability challenges drive AI advancements. Source: Illustration provided by the author.

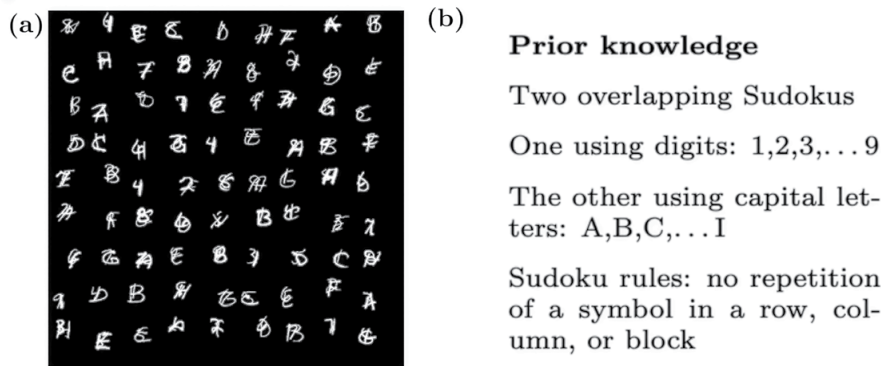
*AI → Science (again)*: The generality of DRNets enables a range of other applications. For instance, the DMVP-DRNets variant (short for Deep Multivariate Probit Deep Reasoning Networks) predicts joint-species distributions, where it is crucial to capture the species’ environmental preferences and interactions among and/or between species. The same framework has also been used for mapping the Amazon land cover, multi-object detection, and soil-organic-carbon modeling.<sup>16</sup> Overall, the two-way street metaphor underscores a fundamental aspect of work on AI for scientific discovery, resulting in both novel AI formalisms and accelerated scientific exploration.

**D**eep Reasoning Networks embed reasoning about prior knowledge directly within deep neural architectures.<sup>17</sup> They seamlessly incorporate knowledge-centric reasoning with data-centric learning through a structured, interpretable latent space that captures domain semantics – enabling more accurate, physics-informed, end-to-end solutions. Originally developed for crystal-structure phase mapping, a core challenge in materials discovery, DRNets extend naturally to other domains. To build intuition about the framework, we begin with a game example (Figure 3) before turning to scientific applications.

Can you make sense of the symbols in Figure 3 (a)? This figure presents a puzzle mixing digits and letters from two overlapping Sudoku games: nine-by-nine grids where each cell must contain a symbol from a given set with no repeats in any row, column, or three-by-three block. Even knowing that the figure represents two

Figure 3

Example of a Demixing Task Based on the Game of Sudoku



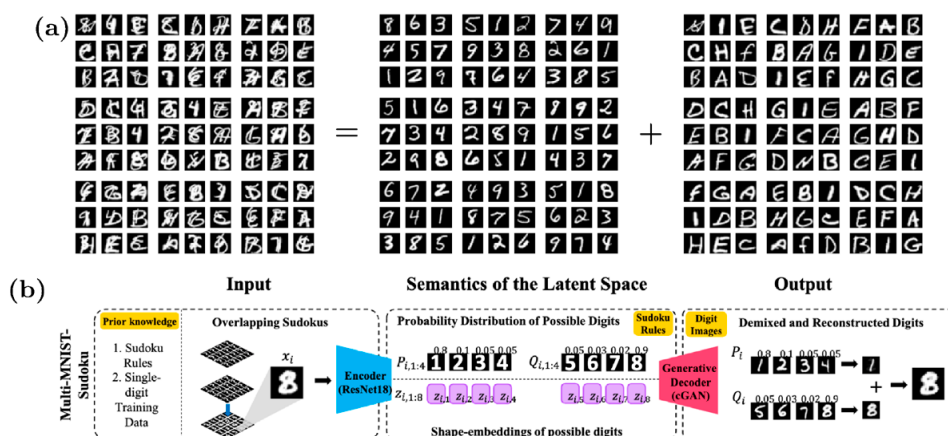
(a) shows the mystery pattern of overlapping symbols and (b) is the background knowledge. Source: Illustration provided by the author.

overlapping Sudokus, one using digits (1–9) and the other using letters (A–I), disentangling them is nontrivial. This task, called Multi-MNIST-Sudoku, formalizes demixing overlapping handwritten puzzles.<sup>18</sup> Traditional combinatorial solvers, such as SAT solvers that aim to resolve the Boolean satisfiability problem, excel at pure symbolic tasks but cannot process noisy image data, motivating the development of hybrids that integrate perception and reasoning.<sup>19</sup>

DRNets emulate human problem-solving by combining visual evidence with symbolic rules in an end-to-end deep learning framework. An encoder maps each Sudoku cell image to an interpretable latent representation of symbol probabilities and shape features. A conditional generative adversarial network (GAN), informed by single-symbol prototypes, reconstructs digits and letters, while a reasoning module (an LSTM, or long short-term memory network, over the Sudoku constraint graph), coupled with differentiable reasoning constraints, enforces row, column, and block consistency.<sup>20</sup> Because Sudoku rules are discrete, DRNets use differentiable, entropy-based continuous relaxations to encode the Sudoku constraints, which are optimized jointly with image reconstruction losses via constraint-aware stochastic gradient descent.

Trained only through self-supervision from relaxed Sudoku rules, without mixed-symbol labels, DRNets outperform supervised baselines, improving both symbol recognition and puzzle completion. Figure 4 (a) shows the demixed pair of Sudokus; Figure 4 (b) gives the DRNets architecture for this task. The knowledge constraints as listed in Figure 3 (b) are enforced through the interpretable latent

Figure 4  
Demixed Sudoku Puzzle and the DRNets Framework for the Demixing Task



(a) shows the process of demixing Sudokus and (b) is the DRNets framework and the semantics of the latent space. Source: (a) Illustration provided by the author; and (b) illustration provided by the author using a modified version of Figure 2 (a) in Di Chen, Yiwei Bai, Sebastian Ament, et al., “Automating Crystal-Structure Phase Mapping by Combining Deep Learning with Constraint Reasoning,” *Nature Machine Intelligence* 3 (9) (2021): 812–822, <https://doi.org/10.1038/s42256-021-00384-1>.

space. Interestingly, the correct semantics of the embedded space emerges through self-supervised training. Moreover, the information derived from the image data and from the constraints complement and amplify each other in our integrated framework.

Now turn to scientific applications of the DRNets framework in materials science and sustainability settings. Sustainable materials and processes are crucial to meeting some of the most pressing challenges in energy generation, transportation, and consumption, as well as broader sustainability issues. Long-term progress will rely on breakthrough innovations in materials science, including the development of new materials and processes for more efficient renewable energy systems and for the reduction of carbon dioxide emissions. Here, I highlight our lab’s work in collaboration with materials scientists at Cornell University and at the California Institute of Technology on advancing AI to accelerate materials discovery for renewable energy. Our knowledge-centric AI reasoning approaches are grounded in self-supervision, leveraging the rich scientific knowledge in materials science.

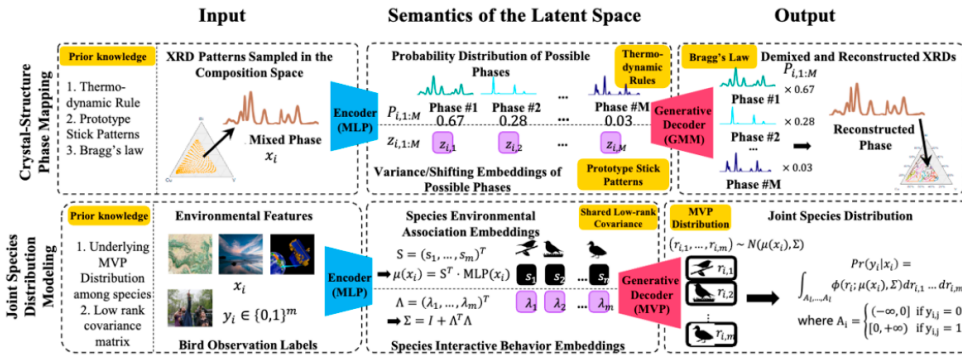
In search of renewable materials for solar fuels, high-throughput experimental settings enable the generation of thousands of materials samples and the analysis of their properties. A major bottleneck in this process, however, is inferring the crystal structure of the samples from high-intensity X-ray diffraction data. Automating crystal-structure phase mapping is therefore essential to speeding up the discovery cycle.

Crystal-structure phase mapping is a demixing task, analogous to Sudoku demixing: Each XRD pattern combines unknown proportions of “pure” phase signals, constrained by the chemistry of the full system. Phase prototypes appear as stick patterns specifying Bragg-peak locations (the positions of strongest X-ray diffraction peaks), analogous to prototypical digits or letters. While simple datasets may be solved by supervised or unsupervised methods, phase mapping is NP-hard, among the most difficult problems to solve computationally, and especially difficult in ternary or higher composition spaces, where alloying causes peak locations to shift and overlap. Addressing this challenge requires reasoning about crystal structures, thermodynamics, and the diffraction process.

Physically and Symmetry Informed DRNets (PSI-DRNets) for crystal-structure phase mapping jointly demix and reconstruct XRD patterns using a graph neural network encoder that produces a structured, interpretable latent space, and a generative decoder that integrates Bragg’s law – linking X-ray wavelength, diffraction angle, and crystal plane spacing – with a Gaussian mixture model (GMM) to represent peak shapes.<sup>21</sup> The decoder incorporates prior knowledge from previously characterized XRD prototype patterns, which are ideal, noise-free profiles from resources such as the Inorganic Crystal Structure Database, used to guide reconstruction.<sup>22</sup> The interpretable latent space encodes phase probabilities and lattice parameters, while thermodynamic and symmetry constraints are enforced via differentiable entropy-based functions with a batched sampling strategy to handle combinatorial complexity across multiple samples (see the top panel of Figure 5 that shows a simplified version of the latent space). A hybrid loss function balances reconstruction fidelity and constraint satisfaction, allowing us to solve phase-mapping instances that had resisted both traditional algorithms and expert manual analysis.

The DRNets framework outperforms state-of-the-art methods and, more important, succeeds where human experts could not: for example, solving the previously unsolved Bi – Cu – V oxide system, a breakthrough that enabled the discovery of solar fuels materials by revealing a three-phase mixture – alloy variants of  $\text{BiVO}_4$ ,  $\text{Cu}_2\text{BiVO}_6$ , and  $\text{Cu}_3(\text{VO}_4)_2$  – that outperformed monoclinic  $\text{BiVO}_4$  in photoactivity, overturning the assumption that phase mixtures inherently reduce performance.<sup>23</sup> These results illustrate how combining deep learning with first-principles, knowledge-centric reasoning can automate data interpretation and significantly accelerate scientific discovery.

Figure 5  
Semantics of the Interpretable Latent Space of the DRNets Framework for Crystal-Structure Phase Mapping and Joint Species–Distribution Modeling



(Top) In crystal-structure phase mapping, DRNets encode the input XRD patterns  $x_i$  into  $P_{i,1:M}$  and  $z_{i,1:M}$ , which denote the probability distribution and the variance/shifting embeddings of  $M$  possible phases. This is a simplified version that does not depict the crystal-structure lattice parameters, also captured by the latent space. The generative decoder (GMM) uses  $z_{i,1:M}$  to generate the decomposed phases and reconstruct the original XRD using the phase probability distribution  $P_{i,1:M}$ . (Bottom) In joint species–distribution modeling, DRNets extract the high-level environmental features  $\text{MLP}(x_i)$  from the raw input  $x_i$  and encode the species environmental association embeddings  $s_j$  and the interactive behavior embeddings  $\lambda_j$  for each species  $j$ . The generative decoder (MVP) uses  $s_j$ ,  $\text{MLP}(x_i)$ , and  $\lambda_j$  to compute the corresponding  $\mu(x_i)$  and  $\Sigma$  for the latent multivariate Gaussian random variables  $r_{i,j}$  and maximizes the likelihood of the species observation  $y_i$  given  $x_i$  under the multivariate probit distribution. Source: Illustration provided by the author. The top panel is a modified version of Figure 2 (b) in Di Chen, Yiwei Bai, Sebastian Ament, et al., “Automating Crystal-Structure Phase Mapping by Combining Deep Learning with Constraint Reasoning,” *Nature Machine Intelligence* 3 (9) (2021): 812–822, <https://doi.org/10.1038/s42256-021-00384-1>.

**B**iodiversity represents the variety of life on Earth and is essential to human well-being, supporting healthy environments and providing natural resources such as food, water, and medicine. It also offers important economic, cultural, and aesthetic benefits. High biodiversity strengthens ecosystem resilience, helping systems recover from disturbances such as climate change and other extreme events.

Understanding how species are distributed across space and time is critical for guiding effective conservation efforts. Inspired by joint species–distribution models and powered by over nine million records of eBird, an online database and citizen-science project for bird observations run by the Cornell Laboratory of Ornithology, we applied the DRNets framework to model year-round spatiotemporal distributions, species–environment associations, and interspecies interactions for

five hundred North American bird species, at a scale surpassing prior joint species–distribution models.<sup>24</sup> Scientific ecological knowledge enhances interpretability and improves joint-species prediction by constraining models to realistic regimes not fully captured by labeled data.

The bottom panel of Figure 5 depicts the DMVP-DRNets architecture: an end-to-end model with a semantically structured latent space that imposes an ecology-informed low-rank constraint on the shared covariance and that uses efficient sampling to estimate species interactions explicitly, paired with a multivariate probit (MVP) decoder based on a multivariate Gaussian model.<sup>25</sup> Unlike many deep models (such as variational autoencoders, or VAEs) that assume diagonal covariance and thus independence, an ecologically unrealistic simplification, DMVP-DRNets capture cross-species dependence directly.<sup>26</sup>

The DMVP-DRNets framework reveals both bird habitat associations and species co-occurrence patterns, producing fine-grained richness maps for around five hundred species, the entire North American avifauna, critical for conservation planning and management. By encoding prior ecological knowledge and modeling interactions explicitly, it improves interpretability and robustness, extending for the first time joint-species modeling to a continental scale.

As another example, the Scientifically-Interpretable Reasoning Network (SciREN) architecture, which models the soil-organic-carbon cycle, is an end-to-end framework with a fully interpretable encoder mapping inputs to an interpretable latent space of over twenty biogeochemical parameters and driving a differentiable decoder based on the Community Land Model version 5 (CLM5), a process-based climate model.<sup>27</sup> The framework’s predictive power matches or exceeds that of prior approaches and, crucially, enables interpretation of the underlying biogeochemical parameters.

In summary, the DRNets framework seamlessly integrates reasoning and deep learning: An encoder builds an interpretable structured latent space for integrating prior knowledge, a reasoning module enforces domain rules, and a decoder injects background insights. This architecture (of encoder + interpretable latent space + reasoning module + decoder) combines knowledge-centric reasoning with data-centric learning: it leverages from a true integration of data-driven and knowledge-centric approaches, and its modular design makes it broadly applicable across domains.

**M**ulti-agent knowledge-centric frameworks will play a central role in AI for scientific discovery, as highly specialized, (semi-)autonomous agents learn, reason, and collaborate while interacting with digital and physical environments.<sup>28</sup> Our work on SARA, the Scientific Autonomous Reasoning Agent, embodies this vision for materials discovery: a robot scientist that operationalizes the scientific method through hypothesis formulation, experimental design,

planning, execution, interpretation, and knowledge generation.<sup>29</sup> SARA already conducts closed-loop experiments using active learning to control laser-spike-annealing synthesis of metastable oxide materials, collect optical and X-ray diffraction data, and map the underlying crystal structures. In SARA's ecosystem, different agents apply various elements of the scientific method and coordinate so that work flows seamlessly from data and knowledge to insight, from conjecture to experiment. The PSI-DRNets component for crystal-structure phase mapping, described earlier, is one of SARA's agents. Work with A-Lab, another autonomous agent, has shown that integrating literature-trained planning, active learning, robotics, and rapid characterization can automate key steps in inorganic materials discovery.<sup>30</sup> Complementing this automation, recent work on scientific reasoning for chemistry suggests that domain-tuned models can articulate step-by-step chemical reasoning and produce structured molecular representations, highlighting the role of explicit reasoning in autonomous discovery.<sup>31</sup>

To make knowledge-centric SciAIs effective, they must communicate using clear and concise representations that both human scientists and other AI agents can interpret and exploit. Learning should proceed not only from data but also from prior scientific knowledge, including physical laws and domain constraints that act as training signals in data-limited regimes. Interpretable latent spaces should align with scientific semantics, allowing SciAIs to incorporate and effectively exploit prior knowledge. Reasoning can leverage problem structure. For example, agents can identify symmetries and invariants, surface the constraints that matter, and recognize when a task can be translated, reduced, or reformulated as one that another agent already knows how to solve efficiently. In this way, the art of modeling and problem reduction becomes a general capability shared across agents and domains.

As autonomy increases, these SciAIs must couple inference with action in self-driving laboratories. A self-driving lab links planning, measurement, and control so that it not only interprets results but also decides what to do next. It selects the next sample, tunes process parameters, schedules measurements, and updates hypotheses as data arrive, enabling disciplined, iterative cycles that converge on reliable knowledge. Decisions are guided by active learning based on the value of information, by safety and feasibility constraints, and by an understanding of how errors propagate through models.

Scientific progress is rarely defined by a single objective; SciAIs and self-driving labs must reason about trade-offs. In particular, AI systems should reveal the Pareto frontier among multiple objectives such as performance, robustness, interpretability, cost, and sustainability criteria; help scientists understand the shape of that frontier; and quantify the consequences of moving along it.<sup>32</sup> Trade-off analysis makes explicit the values behind each choice. For example, in a sustainability setting concerning strategic hydropower planning, my lab, in collaboration with a highly international and interdisciplinary group of scientists, demonstrated how Pareto

multi-objective optimization can balance energy generation with the reduction of adverse impacts on people and nature.<sup>33</sup>

Specialized SciAIs coordinate through shared abstractions and verified messages. Deeply specialized agents can be more effective since their knowledge and reasoning can be precisely tuned to the particular structure of a specific domain. One agent finds the right problem formulation, for example, a materials question, another finds the best suitable algorithm to solve it under constraints, and a third maps the solution back into physical parameters and predicted outcomes. The resulting chain can be checked and validated end-to-end, with provenance and assumptions carried forward rather than hidden – enabling transparency and reproducibility, preconditions for scientific autonomy.

Science advances by leveraging accumulated knowledge, as beautifully captured by Isaac Newton, who, in 1675, wrote in a letter to Robert Hooke, “If I have seen further, it is by standing on the shoulders of giants.”<sup>34</sup> Knowledge-centric AI and reasoning from first principles offer significant new opportunities for accelerating scientific exploration and discovery. AI for science dates to the early days of the 1960s, with DENDRAL, one of the first expert systems, developed at Stanford University under the leadership of Edward Feigenbaum, with a team that included Nobel Prize-winning molecular biologist Joshua Lederberg and pharmaceutical chemist Carl Djerassi. DENDRAL used rule-based reasoning to analyze chemical data and infer molecular structures. Over the past decade, AI has ridden a data-centric wave powered by deep learning and large language models. Here, I argue for injecting knowledge-centric reasoning into AI systems and that genuine discovery requires first-principles reasoning over background and scientific knowledge, tightly integrated with data-centric learning. Our work supports this thesis through a focus on computational sustainability: a two-way street in which AI methods address environmental and societal challenges, and those challenges, in turn, spur new methodologies that advance AI. By tackling sustainability-driven scientific tasks, including pattern demixing for phase mapping in materials science, bioacoustic call identification, joint species–distribution modeling, materials property prediction, land-cover mapping, object detection, and soil-organic-carbon modeling, we have developed general methodologies that transfer across domains.

Looking ahead, scientific progress is poised to be shaped by communities of knowledge-centric scientific AI agents that learn, reason, and collaborate across domains, close the loop between inference and action in (semi-)autonomous laboratories, and navigate trade-offs transparently. In this vision, artificial intelligence becomes a valuable scientific collaborator.

## AUTHOR'S NOTE

This work was partially supported by an AI2050 Senior Fellowship from Schmidt Sciences, the National Science Foundation, the National Institute of Food and Agriculture, and the Air Force Office of Scientific Research.

## ABOUT THE AUTHOR

**Carla P. Gomes** is the Ronald C. and Antonia V. Nielsen Professor of Computing and Information Science at Cornell University, Director of the Institute for Computational Sustainability at Cornell, and Codirector of the Cornell University AI for Science Institute. She is a founder of the field of computational sustainability and the author or coauthor of over two hundred publications. She has published in such journals as *Nature*, *Science*, and *Advances in Neural Information Processing Systems*.

## ENDNOTES

- <sup>1</sup> David Silver, Aja Huang, Chris J. Maddison, et al., “Mastering the Game of Go with Deep Neural Networks and Tree Search,” *Nature* 529 (7587) (2016): 484–489, <https://doi.org/10.1038/nature16961>; Josh Abramson, Jonas Adler, Jack Dunger, et al., “Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3,” *Nature* 630 (8016) (2024): 493–500, <https://doi.org/10.1038/s41586-024-07487-w>; OpenAI, Josh Achiam, Steven Adler, et al., “GPT-4 Technical Report,” arXiv (2023), <https://doi.org/10.48550/arXiv.2303.08774>; and Jared Willard, Xiaowei Jia, Shaoming Xu, et al., “Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems,” *ACM Computing Surveys* 55 (4) (2022): 1–37.
- <sup>2</sup> Ben Goertzel, “Artificial General Intelligence: Concept, State of the Art, and Future Prospects,” *Journal of Artificial General Intelligence* 5 (1) (2014): 1–48, <https://doi.org/10.2478/jagi-2014-0001>.
- <sup>3</sup> United Nations General Assembly, *Transforming Our World: The 2030 Agenda for Sustainable Development*, A/RES/70/1 (September 25, 2015), <https://docs.un.org/en/A/RES/70/1>.
- <sup>4</sup> Carla Gomes, Thomas Dietterich, Christopher Barrett, et al., “Computational Sustainability: Computing for a Better World and a Sustainable Future,” *Communications of the ACM* 62 (9) (2019): 56–65, <https://doi.org/10.1145/3339399>; and Carla P. Gomes, “Computational Sustainability: Computational Methods for a Sustainable Environment, Economy, and Society,” *The Bridge* 39 (4) (2009), <https://www.nae.edu/Publications/Bridge/17281/17286.aspx>.
- <sup>5</sup> Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, “Deep Learning,” *Nature* 521 (7553) (2015): 436–444, <https://doi.org/10.1038/nature14539>; and Ashish Vaswani, Noam Shazeer, Niki Parmar, et al., “Attention Is All You Need,” in *Advances in Neural Information Processing Systems* 30 (NIPS 2017), ed. Isabelle Guyon, Ulrike Von Luxburg, Samy Bengio, et al. (Neural Information Processing Systems Foundation, 2017), [https://proceedings.neurips.cc/paper\\_files/paper/2017/file/3f5ee243547de91fbd053c1c4a845aa-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547de91fbd053c1c4a845aa-Paper.pdf).
- <sup>6</sup> Daniel Kahneman, *Thinking, Fast and Slow* (Farrar, Straus and Giroux, 2011), 499.

- <sup>7</sup> Hanchen Wang, Tianfan Fu, Yuanqi Du, et al., “Scientific Discovery in the Age of Artificial Intelligence,” *Nature* 620 (7972) (2023): 47–60, <https://doi.org/10.1038/s41586-023-06221-2>.
- <sup>8</sup> Allen Newell and Herbert Simon, “The Logic Theory Machine—A Complex Information Processing System,” *IRE Transactions on Information Theory* 2 (3) (1956): 61–79, <https://doi.org/10.1109/TIT.1956.1056797>; Allen Newell, J. C. Shaw, and Herbert A. Simon, “Report on a General Problem-Solving Program,” in *Information Processing: Proceedings of the International Conference on Information Processing* (UNESCO, 1960), 256–264, [http://bitsavers.informatik.uni-stuttgart.de/pdf/rand/ipl/P-1584\\_Report\\_On\\_A\\_General\\_Problem-Solving\\_Program\\_Feb59.pdf](http://bitsavers.informatik.uni-stuttgart.de/pdf/rand/ipl/P-1584_Report_On_A_General_Problem-Solving_Program_Feb59.pdf); Robert K. Lindsay, Bruce G. Buchanan, Edward A. Feigenbaum, and Joshua Lederberg, “DENDRAL: A Case Study of the First Expert System for Scientific Hypothesis Formation,” *Artificial Intelligence* 61 (2) (1993): 209–261, [https://doi.org/10.1016/0004-3702\(93\)90068-M](https://doi.org/10.1016/0004-3702(93)90068-M); and Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. (Pearson, 2021), 1115.
- <sup>9</sup> Willard, Jia, Xu, et al., “Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems”; George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, et al., “Physics-Informed Machine Learning,” *Nature Reviews Physics* 3 (6) (2021): 422–440; Maziar Raissi, Paris Perdikaris, and George Em 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 (2019): 686–707; and Aditi S. Krishnapriyan, Amir Gholami, Shandian Zhe, et al., “Characterizing Possible Failure Modes in Physics-Informed Neural Networks,” *Advances in Neural Information Processing Systems* 34 (*NeurIPS* 2021), ed. Marc’Aurelio Ranzato, Alina Beygelzimer, Yann Dauphin, et al. (Curran Associates, Inc., 2021), 26548–26560.
- <sup>10</sup> OpenAI, Achiam, Adler, et al., “GPT-4 Technical Report”; Anthropic, “The Claude 3 Model Family: Opus, Sonnet, Haiku,” [https://www-cdn.anthropic.com/de8ba9b01c9ab7cbabf5c3b80b7bbc618857627/Model\\_Card\\_Claude\\_3.pdf](https://www-cdn.anthropic.com/de8ba9b01c9ab7cbabf5c3b80b7bbc618857627/Model_Card_Claude_3.pdf) (accessed March 9, 2026); Google Gemini Team, “Gemini: A Family of Highly Capable Multimodal Models,” arXiv (2023), <https://doi.org/10.48550/arXiv.2312.11805>; and Daya Guo, Dejian Yang, Haowei Zhang, et al., “DeepSeek-R1 Incentivizes Reasoning in LLMs Through Reinforcement Learning,” *Nature* 645 (8081) (2025): 633–638, <https://doi.org/10.1038/s41586-025-09422-z>.
- <sup>11</sup> Richard S. Sutton and Andrew G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. (MIT Press, 2018), 526; Jason Wei, Xuezhi Wang, Dale Schuurmans, et al. “Chain-of-Thought Prompting Elicits Reasoning in Large Language Models,” in *Advances in Neural Information Processing Systems* 35 (*NeurIPS* 2022), ed. Sanmi Koyejo, Shkir Mohamed, Alekh Agarwal, et al. (Curran Associates, Inc., 2022), 24824–24837, <https://dl.acm.org/doi/10.5555/3600270.3602070>; Shunyu Yao, Dian Yu, Jeffrey Zhao, et al., “Tree of Thoughts: Deliberate Problem Solving with Large Language Models,” in *Advances in Neural Information Processing Systems* 36 (*NeurIPS* 2023), ed. Alice Oh, Tristan Naumann, Amir Globerson, et al. (Curran Associates, Inc., 2023), 11809–11822, <https://dl.acm.org/doi/abs/10.5555/3666122.3666639>; and Maciej Besta, Nils Blach, Ales Kubicek, et al., “Graph of Thoughts: Solving Elaborate Problems with Large Language Models,” in *Proceedings of the 38th AAAI Conference on Artificial Intelligence*, ed. Michael Wooldridge, Jennifer Dy, and Sriraam Natarajan (AAAI Press, 2024), 17682–17690, <https://doi.org/10.1609/aaai.v38i16.29720>.
- <sup>12</sup> Alexander Novikov, Ngán Vũ, Marvin Eisenberger, et al. “AlphaEvolve: A Coding Agent for Scientific and Algorithmic Discovery,” arXiv (2025), <https://doi.org/10.48550/arXiv.2506.13131>; Bernardino Romera-Paredes, Mohammadamin Barekatian, Alexander Novikov, et al., “Mathematical Discoveries from Program Search with Large Language

- Models,” *Nature* 625 (7995) (2024): 468–475, <https://doi.org/10.1038/s41586-023-06924-6>; and Cunxi Yu, Rongjian Liang, Chia-Tung Ho, and Haoxing Ren, “Autonomous Code Evolution Meets NP-Completeness,” arXiv (2025), <https://doi.org/10.48550/arXiv.2509.07367>.
- <sup>13</sup> Richard M. Karp, “Reducibility Among Combinatorial Problems,” in *50 Years of Integer Programming, 1958–2008: From the Early Years to the State-of-the-Art*, 1st edition, ed. Michael Jünger, Thomas M. Liebling, Denis Naddef, et al. (Springer Berlin Heidelberg, 2010), 219–241.
- <sup>14</sup> Di Chen, Yiwei Bai, Sebastian Ament, et al., “Automating Crystal-Structure Phase Mapping by Combining Deep Learning with Constraint Reasoning,” *Nature Machine Intelligence* 3 (9) (2021): 812–822, <https://doi.org/10.1038/s42256-021-00384-1>.
- <sup>15</sup> Di Chen, Yiwei Bai, Wenting Zhao, et al., “Deep Reasoning Networks for Unsupervised Pattern Demixing with Constraint Reasoning,” in *ICML’20: Proceedings of the 37th International Conference on Machine Learning*, ed. Hal Daumé and Aarti Singh (JMLR, 2020), 1500–1509.
- <sup>16</sup> Courtney L. Davis, Yiwei Bai, Di Chen, et al., “Deep Learning with Citizen Science Data Enables Estimation of Species Diversity and Composition at Continental Extents,” *Ecology* 104 (12) (2023), <https://doi.org/10.1002/ecy.4175>; Chen, Bai, Zhao, et al., “Deep Reasoning Networks for Unsupervised Pattern Demixing with Constraint Reasoning”; and Joshua Fan, Haodi Xu, Feng Tao, et al., “Scientifically-Interpretable Reasoning Network (SciREN): Discovering Hidden Relationships in the Carbon Cycle and Beyond,” arXiv (2025), <https://doi.org/10.48550/arXiv.2506.14054>.
- <sup>17</sup> Chen, Bai, Ament, et al., “Automating Crystal-Structure Phase Mapping by Combining Deep Learning with Constraint Reasoning.”
- <sup>18</sup> Chen, Bai, Zhao, et al., “Deep Reasoning Networks for Unsupervised Pattern Demixing with Constraint Reasoning.”
- <sup>19</sup> Vijay Ganesh and Moshe Y. Vardi, “On the Unreasonable Effectiveness of SAT Solvers,” in *Beyond the Worst-Case Analysis of Algorithms*, ed. Tim Roughgarden (Cambridge University Press, 2020).
- <sup>20</sup> Mehdi Mirza and Simon Osindero, “Conditional Generative Adversarial Nets,” arXiv (2014), <https://doi.org/10.48550/arXiv.1411.1784>; and Sepp Hochreiter and Jürgen Schmidhuber, “Long Short-Term Memory,” *Neural Computation* 9 (8) (1997): 1735–1780, <https://doi.org/10.1162/neco.1997.9.8.1735>.
- <sup>21</sup> Yimeng Min, Ming-Chiang Chang, Shufeng Kong, et al., “Physically Informed Graph-Based Deep Reasoning Net for Efficient Combinatorial Phase Mapping,” in *2023 International Conference on Machine Learning and Applications (ICMLA)* (Curran Associates, Inc., 2023), 392–399, <https://doi.org/10.1109/ICMLA58977.2023.00061>.
- <sup>22</sup> FIZ Karlsruhe–Leibniz Institute for Information Infrastructure, “About ICSD,” <https://icsd.products.fiz-karlsruhe.de/en/about/about-icsd> (accessed March 11, 2026).
- <sup>23</sup> Chen, Bai, Ament, et al., “Automating Crystal-Structure Phase Mapping by Combining Deep Learning with Constraint Reasoning.”
- <sup>24</sup> Di Chen, Yexiang Xue, Daniel Fink, et al., “Deep Multi-Species Embedding,” in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, ed. Carles Sierra (International Joint Conferences on Artificial Intelligence, 2017), 3639–3646, <https://doi.org/10.24963/ijcai.2017/509>.
- <sup>25</sup> Davis, Bai, Chen, et al., “Deep Learning with Citizen Science Data Enables Estimation of Species Diversity and Composition at Continental Extents.”

- <sup>26</sup> Diederik P. Kingma and Max Welling, “Auto-Encoding Variational Bayes,” arXiv (2013), <https://doi.org/10.48550/arXiv.1312.6114>.
- <sup>27</sup> Fan, Xu, Tao, et al., “Scientifically-Interpretable Reasoning Network (SciREN)” ; and David M. Lawrence, Rosie A. Fisher, Charles D. Koven, et al., “The Community Land Model Version 5: Description of New Features, Benchmarking, and Impact of Forcing Uncertainty,” *Journal of Advances in Modeling Earth Systems* 11 (12) (2019): 4245–4287, <https://doi.org/10.1029/2018MS001583>.
- <sup>28</sup> Hongliang Xin, John R. Kitchin, and Heather J. Kulik, “Towards Agentic Science for Advancing Scientific Discovery,” *Nature Machine Intelligence* 7 (9) (2025): 1373–1375, <https://doi.org/10.1038/s42256-025-01110-x>.
- <sup>29</sup> Sebastian Ament, Maximilian Amsler, Duncan R. Sutherland, et al., “Autonomous Materials Synthesis via Hierarchical Active Learning of Nonequilibrium Phase Diagrams,” *Science Advances* 7 (51) (2021), <https://doi.org/10.1126/sciadv.abg4930>.
- <sup>30</sup> Nathan J. Szymanski, Bernardus Rendy, Yuxing Fei, et al., “An Autonomous Laboratory for the Accelerated Synthesis of Inorganic Materials,” *Nature* 624 (7990) (2023): 86–91, <https://doi.org/10.1038/s41586-023-06734-w>.
- <sup>31</sup> Siddharth M. Narayanan, James D. Braza, Ryan-Rhys Griffiths, et al., “Training a Scientific Reasoning Model for Chemistry,” arXiv (2025), <https://doi.org/10.48550/arXiv.2506.17238>.
- <sup>32</sup> Arun Mannodi-Kanakkithodi, Ghanshyam Paliana, Rampi Ramprasad, et al., “Multi-Objective Optimization Techniques to Design the Pareto Front of Organic Dielectric Polymers,” *Computational Materials Science* 125 (2016): 92–99, <https://doi.org/10.1016/j.commatsci.2016.08.018>.
- <sup>33</sup> Alexander S. Flecker, Qinru Shi, Rafael M. Almeida, et al., “Reducing Adverse Impacts of Amazon Hydropower Expansion,” *Science* 375 (6582) (2022): 753–760, <https://doi.org/10.1126/science.abj4017>.
- <sup>34</sup> The Historical Society of Pennsylvania, “Isaac Newton Letter to Robert Hooke, 1675,” <https://digitallibrary.hsp.org/Detail/objects/9792#> (accessed March 6, 2026).