https://doi.org/10.1038/s42254-025-00895-5

Understanding emergence in complex systems using abductive AI

Jingtao Ding, Yu Zheng, Fengli Xu, Carlo Vittorio Cannistraci, Xiaowen Dong, Paolo Santi, Guido Caldarelli, Yizhou Sun, Qi R. Wang, Boleslaw K. Szymanski, Carlo Ratti, Trey Ideker, Jianxi Gao, Yong Li & Deliang Chen

Traditional approaches in complexity science struggle to capture emergent phenomena, but abductive reasoning — now computationally feasible through artificial intelligence — offers a new pathway for discovery.

Emergent phenomena, such as bird flocking or ecosystem collapse, arise from collective dynamics that have properties absent at the individual level. Although the macroscale patterns are observable, their underlying mechanisms remain hidden. Despite major advances in systems theory, chaos theory and network science, it is still not generally possible to explain how microscale interactions give rise to macroscale patterns[1].

We want to bring attention to a source of the problem that has been relatively neglected: the structural limits of deduction and induction. These are the two dominant modes of formal scientific reasoning, but neither is well suited to connecting local processes with emergent global behaviours. Fortunately, there is another type of reasoning: abductive reasoning, or 'inference to the best explanation'. Consider a simple example: you encounter a traffic jam on an urban road, but it's not rush hour. You hypothesize an accident occurred ahead that would explain the unexpected congestion. This reasoning moves backwards from observation to probable cause, generating testable explanations for puzzling phenomena.

Traditionally, a limitation of abductive reasoning is the human ability to devise and test hypotheses, but we believe that augmenting abductive reasoning with artificial intelligence (AI) provides a way forward. Abductive AI can generate and test non-obvious hypotheses about hidden mechanisms that drive emergence, opening new directions for complexity science. How does it work?

The limitations of deduction and induction

Deduction is a process that starts with general principles and derives specific outcomes. Physics exemplifies its power: in principle, given laws and initial conditions, one could deduce planetary orbits. Yet, in practice, the inner solar system is a complex, chaotic system, and predicting far-future planetary trajectories remains an unresolved challenge[2]. Chaos theory illustrates both its strengths and its weaknesses: we can deduce the sensitivity of orbital evolution to initial conditions, but not the exact, long-term trajectories of individual planets. Similarly, the Barabási–Albert network model[3] deduces power-law degree distributions from preferential attachment but cannot predict which nodes become hubs or when cascades occur. Deduction thus falls short when local nonlinearities drive macroscale emergence.

Induction instead generalizes from observations to broader rules. Inductive approaches include the use of reservoir computing to predict chaotic dynamics[4] and sparse regression for discovering governing equations[5]. These approaches learn intricate relationships from observational data without predefined theoretical frameworks. Yet, induction cannot go beyond the available data, and it falters when key microscale data are unobservable — as in the climate system, where abrupt shifts can be detected but underlying processes remain hidden. Induction alone cannot reveal unmeasured mechanisms behind emergent patterns.

What is abductive reasoning?

Abduction, as articulated by Charles Sanders Peirce, generates plausible hypotheses to explain observed phenomena. Whereas deduction and induction operate within established premises or observed data, abduction ventures beyond them to propose candidate mechanisms for unexpected patterns. Working backwards from macroscale patterns, it proposes microscale mechanisms that can be tested and refined. However, in situations involving scientific discovery, the vast hypothesis spaces involved and the volume of observational data often make this approach intractable for human cognition. For example, thinking again about climate science, rare abrupt shifts are hidden within massive observational records, yet innumerable interacting feedback mechanisms — such as ocean–atmosphere couplings, or the carbon cycle — could plausibly account for them.

How can AI power abductive reasoning?

AI may expand the range of problems abductive reasoning can tackle. Foundation models process information at unprecedented scales, enabling them to characterize nonlinear relationships across massive datasets[6]. Reinforcement learning explores large hypothesis spaces and uncovers new strategies beyond human priors, from the Go strategies of AlphaGo to more recent discoveries of new matrix multiplication algorithms[7]. Interpretable AI tools extract symbolic knowledge from black-box models, revealing human-understandable insights[8]. Together, these advances transform abduction into a computationally viable methodology for complexity science.

We envision an abductive AI framework that discovers underlying mechanisms of emergent phenomena, consisting of three synergistic subsystems (Fig. 1). The workflow proceeds iteratively:

- **Hypothesis generation**. The first AI subsystem explores hypothesis spaces to generate the microscale hypothesis, that is, a computational model. For example, deep reinforcement learning learns optimal exploration policies for hypothesis spaces, while diffusion generative models generate hypotheses through progressive refinement from noise. Unlike human intuition constrained by

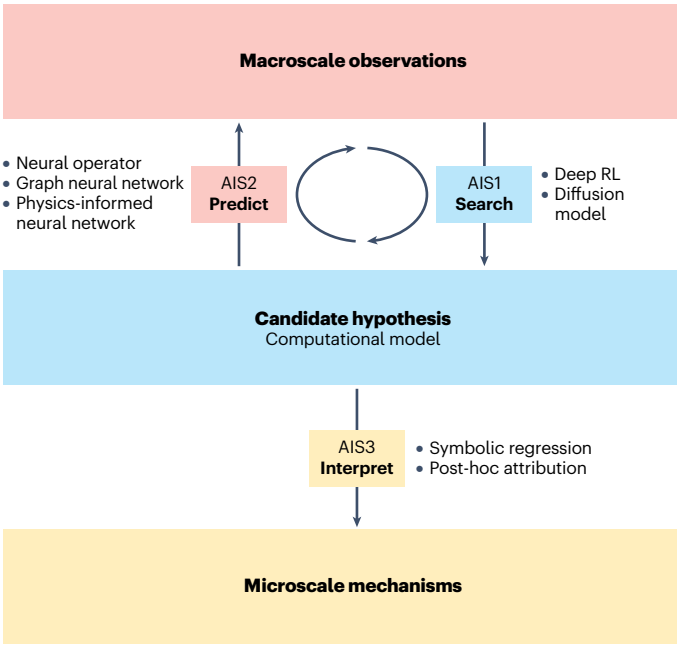

Fig. 1 | An AI-empowered abductive reasoning framework for discovering emergence in complex systems. The process yields parsimonious theories that both predict and explain specific emergent phenomena. AI is used in three systems, AIS1, AIS2 and AIS3, which correspond to the three systems outlined in the main text. RL, reinforcement learning.

knowledge limits, prior experience and computational limitations, an AI system can exhaustively search vast hypothesis spaces and discover non-obvious candidates.

- **Validation**. A second AI subsystem receives each candidate hypothesis and simulates its macroscale consequences, rapidly testing whether the proposed microscale mechanism would generate the observed macroscale phenomena. This process leverages diverse AI architectures — such as neural operators, graph neural networks and physics-informed neural networks — to predict nonlinear dynamics and complex interactions directly from data, accelerating validation compared to traditional first-principles methods.
- **Refinement**. The discrepancy between the predictions of the second subsystem and observed macroscale data serves as feedback, guiding the first subsystem to refine its hypothesis generation in subsequent iterations until the model accurately reproduces the observed macroscale patterns.
- **Interpretation**. A third AI subsystem translates computational results into explicit formulas or causal rules, yielding testable mechanistic understanding of how microscale processes govern system behaviour. For example, symbolic regression extracts symbolic formulas, while post-hoc attribution identifies key mechanistic relationships.

The outcome is parsimonious theories that both predict and explain emergent phenomena.

Illustrative application

An example demonstrates the promise of abductive AI in addressing important emergence problems with a long history. In systems such as ecosystems and power grids, it is essential to be able to identify critical nodes whose failure triggers collapse[9]. Our framework trained AI agents to learn optimal node removal policies with graph neural networks and deep reinforcement learning, then used symbolic regression to derive interpretable formulas quantifying each node's microscale contribution to macroscale system resilience[10]. These expressions revealed how resilience emerges from local structures and dynamical properties.

The path forward

For complexity science, abductive AI should not be positioned as an autonomous discovery engine, but as a scientific co-pilot. Researchers define objectives and constraints, while AI explores hypotheses at computational scale. Human expertise remains central for guiding searches and validating results.

Challenges remain. Interpretability assumes viable explanations can be formalized into human-understandable theories about which scientists can reason qualitatively without exact calculations[8] — a condition not always met for computationally irreducible systems such as protein folding or neural networks. In such cases, predictive accuracy may still outpace mechanistic insight.

Even so, the integration of AI with abductive reasoning signals a methodological shift. By combining human creativity with machine computation, abductive AI offers a powerful framework for decoding emergence across domains — from network science and systems biology to urban systems and Earth system science — and for generating genuine explanations of emergent phenomena.

Jingtao Ding [1,16], Yu Zheng[2,16], Fengli Xu[1,16], Carlo Vittorio Cannistraci [3,4,5], Xiaowen Dong[6], Paolo Santi[2,7], Guido Caldarelli[8,9,10], Yizhou Sun [11], Qi R. Wang [12], Boleslaw K. Szymanski [13], Carlo Ratti[2], Trey Ideker[14], Jianxi Gao[13], Yong Li [1] ✉ & Deliang Chen [15] ✉

[1]Department of Electronic Engineering, Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing, China. [2]Senseable City Lab, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]Center for Complex Network Intelligence (CCNI), Tsinghua Laboratory of Brain and Intelligence (THBI), Department of Psychological and Cognitive Sciences, Tsinghua University, Beijing, China. [4]Department of Computer Science, Tsinghua University, Beijing, China. [5]School of Biomedical Engineering, Tsinghua University, Beijing, China. [6]Department of Engineering Science, University of Oxford, Oxford, UK. [7]Istituto di Informatica e Telematica, Consiglio Nazionale delle Ricerche, Pisa, Italy. [8]Institute of Complex Systems (ISC) and Interdepartmental Centre for Cities Science (ICSC), National Research Council (CNR), Rome, Italy. [9]Department of Molecular Sciences and Nanosystems, Ca' Foscari University of Venice, Venice, Italy. [10]European Centre for Living Technology, Ca' Foscari University of Venice, Venice, Italy. [11]Department of Computer Science, University of California, Los Angeles, CA, USA. [12]Department of Civil and Environmental Engineering, Northeastern University, Boston, MA, USA. [13]Department of Computer Science and Network Science and Technology Center, Rensselaer Polytechnic Institute, Troy, NY, USA. [14]Departments of Bioengineering and Computer Science and Engineering, University of California San Diego, San Diego, CA, USA. [15]Department of Earth System Science, Tsinghua University, Beijing, China. [16]These authors contributed equally: Jingtao Ding, Yu Zheng, Fengli Xu.
✉e-mail: liyong07@tsinghua.edu.cn; deliangchen@tsinghua.edu.cn

Published online: 11 November 2025

References

1. Anderson, P. W. More is different: Broken symmetry and the nature of the hierarchical structure of science. *Science* **177**, 393–396 (1972).
2. Mogavero, F., Hoang, N. H. & Laskar, J. Timescales of chaos in the inner solar system: Lyapunov spectrum and quasi-integrals of motion. *Phys. Rev. X* **13**, 021018 (2023).
3. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
4. Pathak, J., Hunt, B., Girvan, M., Lu, Z. & Ott, E. Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. *Phys. Rev. Lett.* **120**, 024102 (2018).

5. Brunton, S. L., Proctor, J. L. & Kutz, J. N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl Acad. Sci.* **113**, 3932–3937 (2016).
6. Senior, A. W. et al. Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).
7. Fawzi, A. et al. Discovering faster matrix multiplication algorithms with reinforcement learning. *Nature* **610**, 47–53 (2022).
8. Krenn, M. et al. On scientific understanding with artificial intelligence. *Nat. Rev. Phys.* **4**, 761–769 (2022).
9. Artime, O., Grassia, M. & De Domenico, M. et al. Robustness and resilience of complex networks. *Nat. Rev. Phys.* **6**, 114–131 (2024).
10. Zheng, Y., Ding, J., Jin, D., Gao, J., & Li, Y. Advancing network resilience theories with symbolized reinforcement learning. Preprint at https://doi.org/10.48550/arXiv.2507.08827 (2025).

Competing interests

The authors declare no competing interests.