Towards an AI Earth System Scientist: Autonomous Scientific Discovery and Accelerated Transformations Towards Global Sustainability

Jingtao Ding[1,+,*], Yuan Yuan[1,+], Zihan Yu[1,+], Yunke Zhang[1], En Xu[1], Sibo Li[1], Chang Liu[1], and Yong Li[1,*]

[1]Urban Science and Computation Research Center, Department of Electronic Engineering, Tsinghua University, Beijing, China
[+]These authors contributed equally.
[*]Work in progress, please contact: dingjt15@tsinghua.org.cn; liyong07@tsinghua.edu.cn

Global challenges such as climate change, biodiversity loss, and resource depletion accelerate the evolution of Earth system science, yet data overload, insufficient mechanistic understanding, and slow science-to-policy translation remain major obstacles. Advances in artificial intelligence provide new opportunities to address these issues. We propose the concept of the "AI Earth System Scientist (AI-ESS)", a scientific intelligence agent designed to autonomously integrate multi-source data, complex models, and structured knowledge. AI-ESS can support hypothesis generation, mechanistic inference, and experimental validation, while promoting the rapid transformation of scientific outcomes into sustainability actions. Preliminary experiments demonstrate the potential of AI to accelerate Earth system science and highlight key challenges and opportunities ahead.

Introduction

Our Earth and all the life it sustains face unprecedented and interconnected grand challenges. The Earth system is confronting multiple sustainability crises simultaneously: climate change, biodiversity loss, resource depletion, and deepening inequities. Earth System Science (ESS) emerged to address these challenges by understanding Earth as a single, integrated system and providing the knowledge necessary for navigating humanity toward a sustainable future within our planet's finite capacity[1–3].

Over the last four decades, Earth System Science has achieved remarkable progress across multiple fronts. Our conceptual understanding has evolved from viewing Earth's components in isolation to recognizing it as an integrated, self-regulating system profoundly altered by human activities. Foundational concepts such as the Anthropocene[4], tipping elements[5], and planetary boundaries[6] have fundamentally reshaped how we perceive humanity's role in the Earth system and defined safe operating spaces for civilization. Simultaneously, global observation networks combining satellite remote sensing, in situ measurements, reanalysis products, paleoclimate proxies, and climate model simulations have generated unprecedented volumes of Earth science data, providing continuous records spanning from deep geological time to the present[7,8]. These observational advances have been matched by transformative improvements in modeling capabilities. Earth System Models (ESMs) now simulate complex interactions across atmospheric dynamics, ocean circulation, sea ice and ice sheets, land surface processes, and biogeochemical cycles[9], enabling us to project future evolutions and explore what-if questions under different scenarios such as shared socioeconomic pathways[10]. Beyond individual research efforts, major global assessment and synthesis initiatives, exemplified by the Intergovernmental Panel on Climate Change (IPCC), have successfully bridged the gap between scientific communities and policymakers, translating complex Earth system knowledge into actionable information. Most recently, the development and adoption of artificial intelligence (AI) solutions, including foundation models for weather[11], climate[12], and Earth system[13], have opened transformative new frontiers in Earth system process understanding and modeling[14–16].

Despite these substantial advances, Earth system scientists face persistent and growing challenges that limit our ability to address global sustainability crises effectively. The exponential growth in data abundance has not been matched by corresponding advances in understanding and actionable knowledge, creating a widening gap between observation and insight. We possess vast archives of Earth system data, yet struggle to extract meaningful patterns, test hypotheses efficiently, and translate findings into timely interventions. Academic barriers and disciplinary silos continue to hinder the holistic, integrative investigation that ESS fundamentally requires. Researchers trained in atmospheric science, oceanography, ecology, geochemistry, and social sciences often lack common frameworks and tools for genuine collaboration, resulting in a fragmented understanding of inherently coupled processes[1,17]. Perhaps most critically, the traditional scientific discovery-to-action loop remains unable to form a closed, responsive cycle. We face a persistent failure in knowledge co-production, where the isolation of scientific inquiry from societal context prevents the integration of diverse stakeholder perspectives and local constraints[18,19]. This broken loop undermines our capacity to adapt rapidly to emerging crises and optimize pathways toward sustainability.

Technological innovation offers vital potential for overcoming these challenges and accelerating progress toward global sustainability. Recent breakthroughs in AI scientists empowered by large language models have demonstrated transformative capabilities across multiple scientific domains[20]. These systems can autonomously conduct literature reviews, curate and analyze data, generate novel hypotheses, design and execute experiments, validate results, and perform iterative refinement[21,22]. Remarkably, AI scientists have already achieved discoveries comparable to those of human researchers in several fields. In life sciences, cross-disciplinary teams composed of LLM agents have collaborated to design feasible novel SARS-CoV-2 nanobodies[23]. In materials science and chemistry, AI scientists have successfully discovered new compounds and materials with desired properties[24–26]. These achievements suggest that similar approaches could revolutionize ESS.

This perspective paper envisions an "AI Earth System Scientist" (AI-ESS) that harnesses these emerging capabilities to transform how we understand and respond to global sustainability challenges. An AI-ESS would integrate vast and heterogeneous global datasets spanning observations, simulations, and paleoclimate records; construct both predictive and explanatory models that capture Earth system dynamics across scales; autonomously generate and rigorously test innovative hypotheses about system behavior and human-Earth interactions; accelerate scientific discovery by exploring hypothesis spaces far larger than human researchers can manage; and provide actionable, context-aware solutions that promote global sustainable transformation. By closing the loop between observation, understanding, prediction, and action, an AI-ESS could fundamentally accelerate humanity's response to the existential challenges confronting our planet.

Conceptual Framework

Earth System Science

Earth System Science (ESS) investigates the Earth as a single, integrated system by examining the complex interactions and feedbacks between its constituent components: the atmosphere, hydrosphere, cryosphere, lithosphere, and biosphere, with particular emphasis on understanding how human activities are driving rapid global environmental changes and affecting the Earth system's functioning and future trajectory[27]. The recent explosion in the volume of available Earth science data has been pivotal, enabling scientists to explore complex phenomena and processes at unprecedented scales and resolutions[8]. The discovered knowledge is essential for creating innovative, actionable solutions to accelerate the global sustainability transition[1].

Over the last several decades, our ability to observe a changing Earth system has been largely advanced, thanks to a revolution in data acquisition. We now combine high-resolution global monitoring from remote sensing satellites with high-fidelity, localized data from in situ networks, which are synthesized by reanalysis models to create consistent historical records, while novel social sensing techniques are providing new, real-time insights into human-system interactions and impacts[7]. Climate models complement these observations by providing projections of future climate change under different emissions scenarios, enabling us to anticipate potential Earth system responses and inform mitigation and adaptation strategies. Furthermore, paleoclimate data, including historical documentary records, proxy data from natural archives, paleoclimate model simulations, and data assimilation products[28], extend our understanding beyond the instrumental era and provide crucial context for interpreting contemporary climate change by revealing the Earth system's behavior across millennia[29].

The scientific knowledge in ESS can be divided into the following three categories. First, *meta-theory frameworks* provide overarching conceptual paradigms that shape how we understand the Earth system as a whole. Examples include the Bretherton diagram[30], which illustrates the interconnections between Earth system components; the Anthropocene concept[4], which recognizes humans as a geological force; and the identification of tipping elements[5], which highlights critical thresholds in the Earth system. Second, *mechanistic interpretations* explain the specific physical, chemical, and biological processes that drive Earth system behavior. These include the greenhouse effect governing radiative energy balance[31], ice-albedo feedback amplifying polar climate changes[32], the Bjerknes feedback coupling ocean and atmosphere dynamics in the tropical Pacific to generate El Niño-Southern Oscillation variability[33], and thermohaline circulation redistributing heat globally through the ocean conveyor belt[34]. Third, *mathematical models* provide quantitative frameworks for simulating and predicting Earth system dynamics. These range from astronomical theories like Milankovitch cycles explaining glacial-interglacial variations[35,36], to dynamical models such as General Circulation Models[37] and Earth System Models[38] that simulate complex interactions across multiple components, to simplified conceptual models like box models[39] and oscillator models[40,41] that capture essential dynamics while remaining analytically tractable. Together, these three categories of knowledge form a hierarchical structure: meta-theories guide our overall perspective, mechanistic understanding reveals the underlying processes, and mathematical models enable quantitative prediction and hypothesis testing.

This knowledge foundation guides the transformation towards global sustainability through a systematic framework. Status analysis begins with assessing the current state of the Earth system, exemplified by the planetary boundaries framework that identifies safe operating spaces for humanity across critical Earth system processes[6]. Scenario modeling then explores potential future trajectories through climate scenario development and future trend projections, examining how different emission pathways and socioeconomic developments might unfold. Pathway assessment evaluates alternative routes to sustainability using integrated assessment models (IAMs)[42,43] that couple human and natural systems, informing action plan decisions by comparing costs, benefits, and feasibility of different interventions. Finally, action implementation translates scientific knowledge into practice through progress monitoring that tracks key indicators against targets[44], enabling adaptive plan adjustment as new observations reveal the effectiveness of interventions and emerging challenges. This cycle from scientific understanding to societal action represents the ultimate purpose of Earth System Science: not merely to comprehend how our planet functions, but to provide the knowledge necessary for navigating humanity toward a sustainable future within Earth's finite capacity.

AI Earth System Scientists

AI has progressively expanded its role in ESS. Early applications focused on AI as a tool, enabling improved prediction[45–47], downscaling[48–50], emulation[12,51,52], and pattern extraction[53] from high-dimensional geophysical data. With recent advances in reasoning, planning, and tool-use capabilities in large-scale LLM agents, a transitional stage of AI as a collaborator has emerged, in which AI systems assist scientists through hypothesis generation[54], uncertainty quantification[55], attribution analysis[56], and model diagnostics[57]. These developments set the foundation for the forthcoming paradigm shift toward AI as a scientist, where AI systems are able to autonomously navigate a closed-loop scientific discovery. In this paper, we define the AI Earth System Scientist (AI-ESS) as an autonomous or human-in-the-loop scientific agent capable of integrating massive, heterogeneous environmental and socioeconomic datasets derived from earth observations, building predictive and explanatory models, generating and testing new hypotheses, and proposing scientifically and practically grounded solutions that accelerate sustainability transformations.

Based on recent advances in agentic AI[58,59] and the scientific demands of ESS[60–62], we identify five core attributes that define an AI Earth System Scientist (AI-ESS).

- ***Autonomy.***[63,64] The AI-ESS should be able to conduct a closed-loop scientific workflow: hypothesis → simulation → observation → evaluation → theory refinement. This autonomy enables continuous scientific advancement without human micromanagement while still allowing for human oversight and interpretive validation.
- ***Integration.***[65,66] Earth System Science is intrinsically interdisciplinary. An AI-ESS must integrate multimodal observations (satellite, in situ, social sensing, reanalysis), cross-domain scientific knowledge, and diverse modeling tools (mechanistic, statistical, AI-based) to form a unified, cross-scale understanding of Earth system dynamics.
- ***Interpretability.***[67,68] For AI-generated discoveries or model outputs to be trusted and actionable, their reasoning pathways must be at least partially interpretable by human scientists. This includes mechanistic consistency, uncertainty characterization, and alignment with established physical and ecological principles.
- ***Emergence.***[69,70] A defining capability of an AI scientist is to produce novel scientific insights, e.g., emergent teleconnections, unexpected tipping elements, new functional relationships, or previously overlooked feedback mechanisms, beyond what human scientists have articulated. Such emergent discoveries reflect the potential of AI-ESS to advance ESS theory.
- ***Actionability.***[71] Finally, the AI-ESS must produce solutions that are useful for real-world decision-making, linking scientific understanding to actionable policy and sustainability interventions. This requires end-to-end analytical pipelines that generate interpretable diagnostics, scenario projections, and recommendations.

To support these attributes in practice, AI-ESS requires not only more powerful models but a reconstitution of the foundations on which Earth system inquiry is built. Rather than generic infrastructure for ESS, these foundations must directly support the core capacities that distinguish a scientist: the ability to observe, model, reason, and autonomously advance knowledge. We identify three foundational substrates, i.e., data, models, and knowledge, which together enable a fourth ingredient: an agentic scientific workflow through which the AI-ESS can act as a genuine scientific agent.

First, a unified observational substrate is necessary for an AI-ESS to observe the Earth system in the scientific sense. By integrating satellite records (e.g., MODIS, Sentinel), in-situ networks (e.g., Argo, FLUXNET), reanalysis products (e.g., ERA5, SODA), large-ensemble simulations (e.g., CMIP5, CMIP6), socio-economic indicators (e.g., World Bank, UN population datasets), and paleoclimate archives (e.g., ice cores, tree rings) into a coherent, continually updated substrate, the agent gains the multi-scale empirical grounding from which to detect anomalies, infer relationships, and formulate questions. This substrate provides the observational vantage point that human scientists develop through decades of immersion in data. Second, a heterogeneous model substrate is needed for the AI-ESS to reason about processes that cannot be directly observed. Scientists routinely switch between mechanistic equations, conceptual models, and empirical approximations depending on the question at hand. To emulate such flexibility, an AI-ESS must operate within a pluralistic ecosystem that includes physics-based simulators (e.g., CESM, WRF), statistical tools (e.g., linear mixed models, VAR), machine-learning emulators (e.g., NeuralGCM, Samudra), and symbolic or causal inference methods. These models become the agent's means of probing hypothetical worlds, testing mechanisms, and constraining explanations. Third, a structured knowledge substrate is essential for enabling the AI-ESS to reason mechanistically rather than merely detect patterns. Core Earth system theories, ranging from physical laws to concepts such as tipping cascades, planetary boundaries, and modes of variability, encode how scientists judge processes, plausibility, and causation. Encoding this knowledge in machine-interpretable form, such as structured knowledge graphs, symbolic representations, causal diagrams, and machine-readable formulations of physical laws, provides the conceptual scaffolding that enables the AI-ESS to distinguish mechanistic plausibility from mere statistical association and to identify truly novel scientific insights. To apply this knowledge, the AI-ESS must interface with external datasets and models through retrieval mechanisms (e.g., RAG) and standardized coordination protocols (e.g., MCP), allowing it to incorporate new evidence and interact with simulations as an agentic scientific system.

These three substrates, i.e., observation, modeling, and knowledge, constitute the enabling environment for the fourth ingredient: an agentic scientific workflow. Within such a workflow, the AI-ESS can iteratively formulate hypotheses, design and execute simulations or analyses, evaluate outcomes against empirical and theoretical constraints, and refine its internal representations of the Earth system Crucially, this workflow does not simply automate existing research practices; it enables forms of scientific exploration that were previously unattainable. With unified data, model, and knowledge substrates, an AI-ESS can systematically explore hypothesis spaces that lie far beyond human reach. It can test high-order interactions among climatic, ecological, and socioeconomic drivers; compare alternative mechanistic explanations across different models; and probe emergent cross-scale behaviors through iterative simulation–diagnosis loops. Such capabilities allow the system to investigate large scenario spaces whose scale and complexity surpass what human researchers could feasibly examine.

These foundations are not merely architectural; together, they define the qualitative advances that distinguish an AI-ESS from conventional human-centered ESS workflows or tool-level AI for ESS. First, by integrating coherent observations, diverse models, and mechanistic knowledge, an AI-ESS can address problems defined by vast search spaces and tightly coupled drivers. It can navigate large scenarios and model ensembles and analyze complex multi-factor interactions in ways that exceed human analytical capacity. Second, the structured knowledge substrate and agentic workflow allow the AI-ESS to act as a unifying layer across disciplines. By interfacing consistently with natural-science models, socioeconomic data, and causal or symbolic knowledge, the system can overcome entrenched domain silos and support genuinely integrated analyses of coupled human–Earth systems.

Autonomous Scientific Discovery in Earth System Science

The pursuit of autonomous and interpretable scientific discovery by AI agents in Earth System Science (ESS) can draw significant inspiration from the methodological frameworks established through centuries of human theoretical inquiry. The process by which human researchers discover scientific knowledge can often be described as a Conjecture-Refutation cycle[72]. Faced with specific anomalies or complex phenomena, scientists synthesize existing theories to propose bold hypotheses regarding the underlying mechanisms. To validate such a hypothesis, they must employ logical deduction to derive testable statements, that is, specific predictions that can be empirically verified. Subsequently, experiments are designed to quantitatively assess the validity of these statements. Statements proven accurate serve as corroborating evidence for the original hypothesis. Conversely, falsified statements lead to the refutation of the premise, necessitating either the refinement or the abandonment of the hypothesis. When a hypothesis survives rigorous testing, where its derivative statements are repeatedly corroborated by empirical evidence, confidence in its validity grows. Eventually, this process elevates the proposition from a tentative conjecture to the status of widely accepted scientific knowledge. For instance, Alfred Wegener proposed the continental drift hypothesis[73] to explain the remarkable geometric fit between the coastlines of South America and Africa, alongside the puzzling distribution of identical fossils across separated oceans. Although initially met with skepticism due to the lack of a driving mechanism[74], numerous predictions derived from this framework, such as paleomagnetic patterns on the ocean floor[75] and the subsequent discovery of seafloor spreading, were eventually verified by subsequent observations[76]. These confirmations transformed a controversial hypothesis into the foundational theory of plate tectonics, now a cornerstone of Earth system science.

However, completing this cycle to establish scientific knowledge is traditionally a time-consuming process for humans, primarily for two reasons. First, formulating potent hypotheses requires effective scientific intuition, which demands extensive experience, deep domain and interdisciplinary knowledge, and even luck. Second, validating the statements derived from these hypotheses requires substantial labor and time to design and execute empirical experiments. In contrast, the AI Scientist has the potential to overcome these limitations. During the conjecture phase, its access to a vast repository of domain and interdisciplinary knowledge that surpasses the capacity of individual human researchers empowers it to generate bold yet scientifically plausible conjectures. Meanwhile, during the refutation phase, its capacity to autonomously design and execute experiments significantly accelerates the validation process, facilitating rapid iteration of the discovery cycle. In Figure 2(a), we propose a framework for autonomous scientific discovery in Earth System Science (ESS) driven by an AI Scientist, building upon the methodological framework of conjecture and refutation. In this framework, a collaborative multi-agent system discovers new knowledge regarding specific anomalies or phenomena based on existing literature, diverse scientific datasets, and corresponding calculation tools. Specifically, the workflow proceeds as follows:

1. **Knowledge Acquisition**: An Investigation Agent conducts a comprehensive survey of existing theoretical research and domain knowledge from papers and reports, which are transmitted to the Hypothesis Proposal Agent to support its proposal.

2. **Hypothesis Generation**: Leveraging extensive domain and interdisciplinary knowledge, a Hypothesis Proposal Agent formulates bold hypotheses to elucidate potential mechanisms underlying the observed phenomena. By repeating this process, multiple competing hypotheses can be generated and constitute a candidate pool.

3. **Statement Deduction**: To evaluate these candidates, a Statement Deduction Agent derives testable statements from each hypothesis, such as mathematical formulations, specific predictors, or causal structures, along with their expected behaviors and properties. For instance, regarding a hypothesized teleconnection mechanism between the Amazon rainforest and the Antarctic ice sheet, the agent might derive statements like: (1) a variable from one region can serve as a statistically significant predictor for the other with a non-zero correlation coefficient; and (2) the variables along the hypothesized mechanistic pathway can constitute a causal chain correspondingly.

4. **Experimental Validation**: For each testable statement, an Experiment Design Agent orchestrates the validation process. It selects appropriate datasets, performs necessary preprocessing, and quantitatively assesses the statement's validity through methods such as simulation, correlation analysis, or Bayesian analysis. This assessment encompasses both fidelity, e.g., whether the proposed dynamical formula predicts the evolving trajectories in agreement with observational data, and reasonability, e.g., whether the magnitude and sign of coefficients in a multivariate regression align with the statements derived by the Statement Deduction Agent.

5. **Refinement and Corroboration**: Finally, falsified hypotheses are returned to the Hypothesis Proposal Agent for refinement or rejection, while hypotheses that survive falsification but lack sufficient corroboration are returned to the Statement Deduction Agent to generate additional testable statements for further verification. Only hypotheses that are repeatedly corroborated by substantial evidence are output as discovered scientific knowledge, thereby concluding the conjecture-refutation cycle.

Through the repetition of this cycle, the candidate hypothesis pool is guided to evolve toward the optimal description of the studied Earth system phenomena effectively and efficiently.

Illustrative examples

Modeling of global climate mode network dynamics.

Climate modes interacting through teleconnections form a global network that fundamentally shapes climate variability and human ecosystems[77]. Despite its critical role in the global climate system[78], the complexity of underlying physical processes and data patterns has historically confined research to isolated modes (e.g., El Niño-Southern Oscillation, or ENSO)[79,80], neglecting holistic system dynamics. To model the system dynamics of this global climate mode network, we design a multi-agent AI Scientist framework with cohesive five stages as illustrated in Figure 2(b): i) First, an Investigation Agent surveys the geophysical processes influencing climate modes and the dynamical patterns that these modes exhibit. ii) Building on this foundation, a Hypothesis Generation Agent synthesizes known mechanisms while postulating novel mechanisms to describe the complex interactions and evolution of these modes. iii) An Equation Derivation Agent then formalizes these conceptual hypotheses into mathematical formulas, and utilizes genetic programming to generate a diverse cluster of augmented equations to account for system sensitivity. iv) Subsequently, an Experiment Agent collects the required data and simulates these candidates, evaluating them against both observational trajectory fidelity and their capacity to reproduce emergent behaviors (e.g., ENSO complexity). v) Finally, a Conclusion Agent interprets the physical mechanism manifested in the optimal dynamic formula, thereby distilling the findings into formalized scientific knowledge.

Empirical validation corroborates the effectiveness of this framework. Through the iterative discovery process, the model converged on an optimal governing equation:

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} T_i = b_{1,i}(t) + b_{2,i}(t) T_i + b_{3,i}(t) h + b_{4,i}(t) T_i^2 + b_{5,i}(t) \tanh(p) + \sum_{j=1}^{N} b_{6,ij}(t) T_j, \quad i \in 1 \cdots N, \\ \frac{\mathrm{d}}{\mathrm{d}t} h = \sum_{i=1}^{N} b_{7,i}(t) T_i, \end{cases} \tag{1}$$

Here, $N = 9$ is the number of climate modes within the network, T_i denotes the sea surface temperature (SST) anomalies of the i-th climate mode, h is the warm water volume (WWV) anomalies in the equatorial Pacific, t is the years elapsed since January 1979, p is the normalized atmospheric carbon dioxide (CO_2) concentration, $\tanh(p) = (e^p - e^{-p})/(e^p + e^{-p})$, and $b_*(t) = a_*^{(1)} + a_*^{(2)} \sin(\omega_0 t) + a_*^{(3)} \cos(\omega_0 t) + a_*^{(4)} \sin(0.195\omega_0 t)$ are time-dependent coefficients, where $a_*^{(i)}$ are scalar parameters and $\omega_0 = 2\pi/\text{year}$. Crucially, the identified terms successfully recover established physical mechanisms, including the Bjerknes feedback ($b_{2,i}T_i$), the SST-WWV interaction of the Recharge Oscillator ($b_{3,i}h$ and $b_{7,i}T_i$)[79], the nonlinearity of the nRO model ($b_{4,i}T_i^2$)[79], and the teleconnections of the XRO model ($b_{6,ij}T_j$)[80]. Beyond verification, the AI autonomously formulated novel hypotheses: it constructed $b_{5,i}\tanh(p)$ to explicitly represent greenhouse gas forcing and introduced a unique 5.13-year interannual component ($a_*^{(4)}$) within the time-varying coefficients to capture low-frequency variability. Consequently, this AI-derived model demonstrates significant performance gains over state-of-the-art human-designed models, that is, the XRO model[80]. As illustrated in Figure 2(c), predictive skills across various climate modes are improved by 20%—58%, with the prediction skills for ENSO extending from 19 months to 30 months. Moreover, the model exhibits superior capability in reproducing the emergent dynamical behaviors of real-world indices such as ENSO complexity (measured by monthly standard deviation, skewness, and power spectral density), reducing the mean average error by 17%-–37% relative to the XRO baseline.

Identification of possible new tipping elements.

The identification and mechanistic understanding of Earth-system tipping elements represent an urgent research frontier[5,81,82]. However, such progress is hindered by vast multidisciplinary data, complex non-linear interactions, and protracted human-led modeling campaigns[83,84]. We propose an autonomous, multi-agent AI Scientist framework designed to systematically navigate this complexity and accelerate the scientific discovery pipeline. Specifically, our framework structures the discovery process into four key stages as in Figure 2(d): (i) comprehensive literature investigation to scout for candidate systems and research gaps; (ii) hypothesis generation, where this knowledge is fused with empirical data to formulate theories on new tipping elements and their mechanisms; (iii) rigorous hypothesis verification via targeted in silico experiments within an AI-enhanced Earth system model; and (iv) a rapid, iterative trial-and-error loop for self-correction and convergence.

This AI Scientist demonstrates profound advantages over traditional research. First, it achieves unparalleled multidisciplinary synthesis: investigation agents concurrently process a massive corpus spanning geophysics, ecology, and atmospheric chemistry, identifying non-obvious, cross-disciplinary linkages that elude specialized human reviews. Second, it excels at cross-disciplinary knowledge fusion. A core hypothesis-generation agent merges top-down abstract knowledge from the literature with bottom-up empirical evidence from multi-modal data, including satellite observations, re-analysis data, and knowledge from paleoclimatology. By identifying intersections between known research gaps, anomalous data, and plausible established geophysical constraints, it autonomously generates novel, high-potential hypotheses on tipping elements and their corresponding mechanisms. Third, the framework enables rapid iterative validation. The experiment design agent proposes and executes bespoke simulations on the AI-enhanced Earth system model refined by the model design agent, replacing time-intensive manual modeling by humans. This velocity facilitates the system's most transformative capability with a closed-loop, self-evolutionary potential. False tipping element results after verification are treated as valuable data points, triggering immediate hypothesis reformulation. In parallel, an explanation agent translates the verified hypothesis on tipping elements into theoretically consistent and understandable narratives, enabling human scientists to acquire new knowledge of previously unknown tipping elements rather than merely receiving predictions.

Modeling the Coupled Human-Earth System in the Anthropocene.

As the geological epoch shifts into the Anthropocene, the scale and intensity of human activities have reached unprecedented levels, sufficient to deviate the Earth system from its natural evolutionary trajectory[85]. This profound shift has compelled Earth System Models (ESMs) to incorporate human activity as a critical external forcing[86]. Conversely, human activities are also constrained by planetary boundaries and inevitably adapt as the Earth's state evolves, as traditionally modeled in IAMs[42,43] by treating the Earth system as a boundary condition[87]. Despite the inherent bidirectional coupling between natural and anthropogenic processes, a fully integrated model of the coupled Human-Earth system remains elusive[87]. Conventionally, ESMs neglect the adaptive feedback loops where human societies adjust strategies in response to environmental changes. Simultaneously, IAMs often simplify the Earth system into static boundary conditions, overlooking how human interventions fundamentally alter the Earth's evolutionary trajectory. To accurately reconstruct past climate changes and project future scenarios, it is necessary to transform from unidirectional forcing to bidirectional coupling, making human activity a coupled variable but not just an external driver[88]. However, achieving this integration is fraught with challenges. First, establishing such a coupled system requires interdisciplinary knowledge of both socio-economic and natural environments. Secondly, Earth systems obey immutable physical laws described by differential equations, whereas human systems are driven by complex decision-making processes governed by economic optimization and game theory, leading to their heterogeneity of underlying dynamics. Finally, the different spatiotemporal scales and granularities of natural environmental data and socio-economic data make it challenging to integrate them to evaluate the proposed human-land coupling system. Our proposed AI Earth System Scientist framework offers a promising pathway for overcoming these challenges, as illustrated in Figure 2(e): First, an Investigation Agent aggregates existing ESMs and IAMs from literature, identifying key state variables for both natural systems (e.g., temperature, carbon cycle) and human societies (e.g., GDP, emissions, land use), along with their known interrelationships.

Then, based on this knowledge, a Hypothesis Proposal Agent formulates potential coupling architectures (which variables exert direct effects on others) and corresponding governing laws (dynamics formulas governing physical dynamics and optimization equations governing socio-economic). To rigorously verify these hypotheses, an Experiment Design Agent implements a hierarchical strategy: It first verifies the causal structure through correlation analysis and Bayesian inference, distinguishing true causality between coupled variables from spurious correlations or mere co-occurrence. Subsequently, it validates the proposed equations by executing dynamical simulations to evaluate the fidelity of the model's output against observational and historical data. These evaluation metrics are fed back to the Hypothesis Proposal Agent to refine the coupling structures or mathematical formulations. Ultimately, the hypothesis that achieves high simulation fidelity with a verified causal structure emerges as the winner. The Conclusion Agent then interprets this optimal coupled model, distilling it into formalized scientific knowledge regarding the unified Human-Earth system.

Accelerating Sustainability Transformations

AI-ESS shortens the science-policy chain.

Status analysis forms the critical foundation of the sustainability transformation cycle, requiring a comprehensive and accurate assessment of the Earth system's current state against key benchmarks, such as the planetary boundaries framework[6] and the UN Sustainable Development Goals[44]. This stage involves integrating massive, heterogeneous datasets, such as satellite remote sensing and in-situ sensor networks to socioeconomic statistics, to calculate indicator values and diagnose planetary health. However, this process faces profound challenges: essential data remain scattered across disciplines and formats[44], suffer from significant spatial and temporal gaps[89], and are often analyzed with long delays[7]. Even the most advanced global monitoring reports are constrained to analyzing a small subset of all potential indicators due to these data integration barriers and reliance on manual, retrospective methods. Although current AI tools can perform specific tasks like data imputation or pattern recognition[90], they function as isolated aids and cannot autonomously synthesize a holistic, real-time diagnosis of the Earth system. The AI-ESS addresses these limitations by introducing a closed-loop, autonomous paradigm for planetary-scale observation and diagnosis. It can continuously ingest and harmonize multi-source data, using foundation models to dynamically fill data gaps and reconstruct complete environmental fields[91]. By systematically surveying scientific literature and assessment reports, the AI-ESS builds a structured knowledge substrate of Earth system theories and indicator definitions, enabling it to interpret data patterns mechanistically. This allows the system to autonomously track a vastly expanded portfolio of sustainability indicators, moving beyond static snapshots to provide a continuously updated, interpretable status of the Earth system. Through this integrated workflow, the AI-ESS transforms status analysis from a fragmented, lagging effort into a unified, proactive, and explanatory diagnostic process.

To explore possible sustainability transformation trajectories and assess the effectiveness of climate mitigation strategies, scientists develop scenario frameworks that provide integrated descriptions of plausible futures for the human-environment system[92]. However, the traditional scenario development process is both time-consuming and labor-intensive. The Shared Socioeconomic Pathways (SSPs) exemplify these challenges: their development required several years of extensive deliberation and iterative cycles to construct coherent narratives[10], reconcile inconsistencies across disciplinary perspectives[93], and translate these narratives into quantitative drivers—ranging from population dynamics to policy interventions[94,95]. This protracted timeline creates a fundamental mismatch between the speed of scenario generation and the urgency of real-world decision-making. The current SSP-RCP framework has been criticized for its static nature, leaving it ill-equipped to address near-term mitigation pathways through 2040 with the necessary granularity and responsiveness to recent global shocks[96]. Waiting for the next IPCC assessment cycle to update these assumptions is untenable. Therefore, dramatically accelerating the development of comprehensive, internally consistent scenarios has become essential for effective climate mitigation planning. The AI-ESS addresses these limitations through several transformative capabilities. LLM agents can autonomously investigate vast bodies of literature across disciplines, identifying emerging trends and evolving societal priorities that traditional review processes might overlook or integrate too slowly. By simulating cross-disciplinary collaboration, i.e., embodying perspectives from climate scientists, economists, sociologists, and policy analysts, these agents can propose novel scenarios that complement the current SSP framework while maintaining internal consistency. Unlike traditional approaches that freeze assumptions at the outset, AI-ESS enables real-time scenario updating as new data emerges from geopolitical shifts, technological breakthroughs, or observed climate impacts. Furthermore, AI agents can conduct accelerated vetting and peer review processes, checking scenarios for logical consistency, identifying potential contradictions across quantitative projections, and flagging areas requiring human expert judgment. Critically, AI-ESS augments rather than replaces human judgment in developing climate scenarios, enabling scientists to explore far larger possibility spaces and focus their expertise on the most consequential decisions.

Designing climate mitigation and long-term sustainability pathways requires systematically examining how alternative socioeconomic, technological, and policy choices shape future trajectories. Integrated assessment models (IAMs), which couple human and natural systems, serve as foundational tools for evaluating the costs, benefits, and feasibility of mitigation and adaptation strategies[43]. In current practice, however, IAM workflows remain primarily dependent on human design. Given high-level mitigation goals—such as temperature limits or carbon budget constraints—and the input socioeconomic and technical variables for the target scenario, researchers must manually configure model modules, run optimization routines, and interpret the resulting pathways. Although advances in AI, such as surrogate modeling[97] and generative scenario augmentation[98], have improved efficiency in individual steps, these techniques operate mainly as tool-level enhancements and do not overcome the central limitation of restricted scenario exploration. AI-ESS addresses this limitation by providing coordinated support across all stages of the IAM workflow. Starting from the human-specified mitigation goal and target scenario, it analyzes model repositories and literature to recommend suitable economic, energy, and land-use modules, configuring their interactions so that demand, technology deployment, and emissions are represented consistently. During optimization and solving, AI-ESS selects and combines methods such as reinforcement learning and generative modeling to explore high-dimensional mitigation decision spaces. The resulting emissions trajectories are evaluated through climate response components, which compute temperature outcomes and check alignment with mitigation goals and physical consistency. When targets are not met or inconsistencies arise, the system diagnoses the responsible elements—whether in scenario definition, module interaction, or optimization settings—and proposes revisions. Through these iterative cycles, pathway assessment evolves from a static, expert-driven procedure into a self-refining loop capable of exploring a broader, more adaptive set of futures.

Action implementation and monitoring complete the sustainability transformation cycle by translating scientific insights into tangible interventions and systematically tracking their effectiveness. This crucial phase involves executing policies, allocating resources, and measuring outcomes against established targets, such as the Sustainable Development Goals, where progress monitoring provides essential accountability[44]. Yet current practice faces persistent barriers: monitoring data rarely informs real-time decision adjustments[89], human-led evaluation processes are too slow for rapidly evolving sustainability challenges[99], and institutional silos prevent integrated assessment of cross-sectoral interventions[7]. Though AI tools can assist with isolated tasks like optimizing renewable energy deployment or detecting deforestation from satellite imagery, they typically operate as disconnected components rather than forming a responsive knowledge-action system. The AI-ESS overcomes these limitations by creating a unified implementation-monitoring framework that continuously learns from outcomes. It directly connects intervention execution with real-world observations—tracking everything from urban green space expansion to carbon sequestration in restored forests—and automatically analyzes whether observed changes align with intervention goals. By applying causal inference techniques, it distinguishes genuine policy impacts from external influences like economic shifts or climate variability. When interventions underperform, the system doesn't just flag problems but proposes specific adjustments: redirecting funding to struggling regions, modifying implementation timelines based on seasonal constraints, or scaling successful pilot projects to broader geographic areas. This transforms sustainability action from a rigid, sequential process into an adaptive practice that continuously refines strategies based on their measurable impacts on human well-being and planetary boundaries.

Illustrative examples

Assessments of extreme weather impacts on renewable energy production.

A pivotal application of the AI-ESS is a collaborative human-AI framework for conducting end-to-end assessments of extreme weather impacts on renewable energy production[100]. As shown in Figure 3a, a human scientist first posits a high-level research topic, such as investigating the complex relationship between escalating extreme weather events and the stability of renewable energy systems. The AI-ESS then operationalizes this inquiry by autonomously executing a detailed research pipeline. The pipeline begins with an automated review of scientific literature, where the investigation agent identifies critical research gaps, such as analyses with low spatial resolution, i.e., province level[101,102] or country level[103], or less representativeness of the used assessment models for energy yield in actual power plants[104,105]. To address these limitations, the hypothesis agent formulates a hypothesis and designs a comprehensive analysis plan. Then, a data collection agent conducts iterative collection and fusion of multi-source data, including satellite remote sensing images[106] and historical reanalysis data[107,108]. Based on these data, the processing agent utilizes remote sensing detection models to precisely identify power plant locations[109], and extracts corresponding extreme weather events at these locations. It further leverages advanced tools like Physics-Informed Neural Networks (PINNs) for predicting renewable power supply during extreme weather events[110]. To estimate the impacts of extreme weather on renewable energy production, an analysis agent computes the spatial-temporal distribution and fluctuations in power production, and summarizes the results in a report, which is further evaluated by a critic agent. The provided feedback serves as guidance for updating and improving this workflow. Final results are double-checked with human experts.

Employing an AI Scientist paradigm, rather than merely using AI as a tool, is critical for tackling such a complex problem. First, it autonomously synthesizes specialized knowledge and methodologies from disparate fields. In the given example, this entails fusing computer vision for asset detection, climate science for weather projections, and power engineering data, whose cross-disciplinary integration is challenging for domain experts. In addition, it dramatically accelerates scientific discovery by automating the entire iterative loop. The system rapidly generates, cross-verifies, and refines its hypotheses, while human intervenes and check only at a few critical steps in the whole loop.

Construction of global mitigation pathways.

As illustrated in Figure 3b, AI-ESS can construct global mitigation pathways by replacing the traditional IAM approach with an autonomous agentic workflow. Beginning from human-specified mitigation goals, a scenario-framing agent first formalizes the scenario definition by gathering and synthesizing relevant information from IPCC and related assessments. This includes SSP socioeconomic trajectories and policy constraints such as carbon-price schedules, emission caps, and technology-deployment limits. These elements collectively define the boundary conditions under which mitigation pathways must be generated. In the second stage, a literature-review agent analyzes existing IAM studies and empirical evidence to identify which domain components—typically economic activity, energy-system transitions, and land-use dynamics—are required to represent the scenario, and to determine the key variables governing interactions among these components (e.g., energy demand, technology costs, emissions factors). Building on this, a modeling agent selects appropriate domain models from established IAM families such as MESSAGEix[111] for energy-system and technology-substitution dynamics, and simplified climate-response models such as MAGICC[112] or FaIR[113]. When beneficial, the modeling agent can also construct AI-based surrogate components[114] to accelerate computationally intensive modules or better capture nonlinear interactions between system components. During the third stage of pathway solving, an optimization agent formulates and solves the high-dimensional, multi-objective optimization problem implied by the configured IAM. Depending on the computational context, it may adopt reinforcement learning[115], generative diffusion models[116], or hybrid approaches—for example, generating diverse candidate pathways using generative models and subsequently applying reinforcement learning to refine them toward scenario-specific objectives. In the next stage, a climate-evaluation agent processes the resulting emissions trajectories through climate-response models to compute atmospheric concentrations, radiative forcing, and temperature outcomes. A critic agent then assesses whether the generated pathways satisfy the specified mitigation targets and remain internally consistent. If deviations or structural inconsistencies are detected, the critic agent localizes the source—such as conflicting scenario inputs, unstable module interactions, or suboptimal optimization choices—and routes corrective feedback to the appropriate upstream stage. Once mitigation goals are satisfied, a synthesis agent consolidates the resulting mitigation pathways and evaluates their sectoral transitions, feasibility characteristics, and technology requirements. These synthesized outputs form the basis for expert interpretation and final refinement.

A key advantage of AI-ESS is its ability to discover new, previously unreachable mitigation pathways. Traditional IAMs generate only a limited set of solutions because scenarios must be manually specified and solved individually, whereas current limited explorations using generative AI[98] merely sample around existing IAM outputs. By autonomously constructing scenarios and solving the underlying high-dimensional optimization problems, AI-ESS expands the feasible pathway space itself, enabling the identification of valid mitigation trajectories that neither manual IAM workflows nor data-driven generative models can produce.

Comprehensive monitoring and evaluation of sustainable development progress

A transformative application of AI-ESS is its capacity to conduct end-to-end, comprehensive evaluation of global sustainable development progress across all 17 SDGs and 232 indicators. As illustrated in Figure 3c, a sustainability policymaker begins by setting a broad evaluation objective: understanding the holistic progress and interconnections across all Sustainable Development Goals at national and subnational scales. The AI-ESS system then executes an integrated four-stage pipeline that transforms fragmented monitoring into a unified assessment framework. The pipeline begins with the Data Agent autonomously discovering, integrating, and quality-assessing multimodal datasets from satellite observations, socioeconomic statistics, environmental sensors, and administrative records. This agent constructs a dynamic metadata knowledge base with RAG capabilities, enabling intelligent data discovery that addresses critical coverage gaps—particularly in regions where 68% of traditional SDG indicators cover less than half of countries and 51% lack updates beyond 2015. Next, the Indicator Parsing Agent automatically interprets the textual definitions of all 232 SDG indicators, mapping each to required variables, computational formulas, and statistical methodologies. This agent resolves semantic ambiguities and identifies cross-indicator dependencies that human analysts typically miss when working with isolated metrics. The third stage employs the Indicator Engine Agent to perform precise calculations at unprecedented spatial resolution (30-meter grid cells) while addressing missing data through Earth system foundation models that perform spatial interpolation, temporal extrapolation, and proxy variable construction. Unlike conventional reports that analyze fewer than 20 indicators, this system quantifies over 200 indicators with uncertainty estimates for each calculation. Finally, the Evaluation Agent synthesizes these results into actionable insights by building a policy knowledge base from global sustainability initiatives, applying causal inference methods to attribute indicator changes to specific interventions, and constructing an SDG digital twin for simulating future scenarios under different policy combinations. The system automatically generates region-specific reports highlighting not only progress patterns but also identifying synergistic and conflicting goal interactions, such as how water conservation policies in one region affect energy security and food production in connected systems.

This comprehensive monitoring paradigm fundamentally transforms sustainability assessment from fragmented reporting to integrated systems analysis. By processing the full spectrum of SDG indicators simultaneously with causal attribution capabilities, the AI-ESS reveals hidden interconnections and trade-offs that isolated indicator tracking cannot detect. The system's ability to maintain high-resolution analysis across all goals enables policymakers to identify leverage points where targeted interventions create cascading positive effects across multiple sustainability dimensions, while avoiding siloed approaches that optimize single indicators at the expense of systemic resilience. This represents a critical advancement toward truly integrated sustainability governance that matches the complexity of Earth's coupled human-natural systems.

Challenge and Open Questions

The proposed AI-ESS framework fundamentally embodies abductive reasoning, a form of logical inference that seeks the best explanation for observed phenomena[117]. As defined by C.S. Peirce as "the only logical operation which introduces any new idea," abduction generates plausible hypotheses that, if true, would account for surprising or anomalous observations. Unlike deduction or induction, abduction is inherently creative and explanatory—the cognitive mechanism through which scientists propose novel causal mechanisms and develop new theories. The AI-ESS framework operationalizes abduction through its Hypothesis Proposal Agent, which synthesizes vast interdisciplinary knowledge to identify plausible causal mechanisms. Advanced AI approaches for hypothesis search across large microlevel spaces[20,70] enable the agent to explore hypothesis spaces far beyond human limitations, considering non-obvious cross-domain connections and complex feedback mechanisms that individual researchers might overlook. For instance, whereas a human climate scientist might propose teleconnection mechanisms based primarily on atmospheric dynamics, an AI agent could simultaneously consider coupled ocean-atmosphere-cryosphere pathways, biogeochemical feedbacks, and socioeconomic factors. However, abductive reasoning requires first recognizing that something puzzling has occurred—that current theories fail to explain observed patterns. Developing robust anomaly detection capabilities in AI scientists[118] is therefore essential for complete autonomous discovery. More crucially, this requires distinguishing between theoretically significant anomalies that indicate novel physical mechanisms and spurious patterns arising from random noise, measurement errors, or statistical artifacts. An AI scientist must possess the domain knowledge to recognize when an anomaly violates fundamental theoretical expectations versus when it represents mere data variability.

The integrity of an AI Scientist is fundamentally bounded by the representativeness of its training data. While datasets are abundant for the Northern Hemisphere and terrestrial regions, critical gaps persist in the Global South, the deep ocean, and polar regions[119–122]. For example, a machine learning-enhanced analysis of over 100,000 studies illustrates a systematic attribution gap where robust evidence for attributable impacts is twice as prevalent in high- compared to low-income countries globally[119]. AI scientists accessing such skewed repositories risk developing systematic neglect of under-represented regions[90] and failing to identify mechanisms in data-sparse environments that may hold the key to global sustainability. Beyond data limitations, according to Goodhart's Law—which states that when a measure becomes a target, it ceases to be a good measure—AI agents entrusted with autonomous discovery and validation tasks may exhibit unexpected biases in goal-setting and reasoning processes[123]. When optimization metrics become explicit objectives, AI systems may exploit shortcuts that satisfy these metrics superficially while undermining genuine scientific insight. This tendency can amplify existing inequities, increasing the risk of climate injustice in designing mitigation strategies that inadvertently favor well-documented regions over vulnerable but under-represented communities[124]. Moreover, the absence of clear accountability frameworks risks creating a trust crisis where the public and policymakers question the validity of AI-generated scientific conclusions[125]. To address these challenges, human-in-the-loop accountability mechanisms are essential, ensuring that AI-ESS leverages computational power for discovery while maintaining scientific rigor and serving global climate justice.

In the framework and case studies outlined, AI-ESS is better suited as a collaborative partner within human-AI scientific teams rather than a standalone solution. This human-AI synergy has already yielded surprising insights in mathematical research[126] and holds untapped potential for investigating Earth systems. Effective collaboration hinges on AI's ability to convey its understanding to human experts, emphasizing the necessity of a scientific understanding test aimed at narrowing the gap between human and AI scientists[127]. From the perspective of autonomous scientific discovery in ESS, human-AI collaboration addresses inherent complexities that neither can fully tackle alone. While AI excels at exploring vast hypothesis spaces and identifying non-obvious patterns across heterogeneous datasets, human scientists provide irreplaceable domain expertise in recognizing physically meaningful anomalies, constraining hypotheses with theoretical principles, and interpreting results within broader scientific contexts. From the perspective of accelerating sustainability transformations, human-AI collaboration ensures scientific discoveries translate into equitable and actionable solutions. AI-driven processes risk optimizing for computational efficiency or statistical significance over social justice and ecological integrity. Human oversight redirects AI toward research questions relevant to vulnerable communities, incorporates local and indigenous knowledge absent from formal datasets, and evaluates interventions against ethical criteria beyond technical feasibility.

References

1. Reid, W. V. *et al.* Earth system science for global sustainability: Grand challenges. *Science* **330**, 916–917, DOI: 10.1126/science.1196263 (2010).
2. Steffen, W. *et al.* The emergence and evolution of earth system science. *Nat. Rev. Earth Environ.* **1**, 54–63, DOI: 10.1038/s43017-019-0005-6 (2020).
3. Zheng, Y. *et al.* A holistic perspective on earth system science. *Sci. China Earth Sci.* **67**, 3013–3040 (2024).
4. Crutzen, P. J. & Stoermer, E. F. The "Anthropocene". *IGBP Newsl.* **41**, 17–18 (2000).
5. Lenton, T. M. *et al.* Tipping elements in the Earth's climate system. *Proc. Natl. Acad. Sci.* **105**, 1786–1793 (2008).
6. Rockström, J. *et al.* A safe operating space for humanity. *nature* **461**, 472–475 (2009).
7. Li, X. *et al.* Big data in earth system science and progress towards a digital twin. *Nat. Rev. Earth & Environ.* **4**, 319–334, DOI: 10.1038/s43017-023-00409-w (2023).
8. Vance, C., Tiffany, Huang, T. & Butler, A., Kevin. Big data in earth science: Emerging practice and promise. *Science* **383**, DOI: 10.1126/science.adh9607 (2024).
9. Flato, G. M. Earth system models: an overview. *Wiley Interdiscip. Rev. Clim. Chang.* **2**, 783–800 (2011).
10. O'Neill, B. C. *et al.* A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Clim. change* **122**, 387–400 (2014).
11. Bi, K. *et al.* Accurate medium-range global weather forecasting with 3d neural networks. *Nature* **619**, 533–538 (2023).
12. Kochkov, D. *et al.* Neural general circulation models for weather and climate. *Nature* **632**, 1060–1066 (2024).
13. Bodnar, C. *et al.* A foundation model for the earth system. *Nature* 1–8 (2025).
14. Reichstein, M. *et al.* Deep learning and process understanding for data-driven earth system science. *Nature* **566**, 195–204 (2019).
15. Irrgang, C. *et al.* Towards neural earth system modelling by integrating artificial intelligence in earth system science. *Nat. Mach. Intell.* **3**, 667–674, DOI: 10.1038/s42256-021-00374-3 (2021).
16. Bracco, A. *et al.* Machine learning for the physics of climate. *Nat. Rev. Phys.* **7**, 6–20 (2025).
17. Mooney, H. A., Duraiappah, A. & Larigauderie, A. Evolution of natural and social science interactions in global change research programs. *Proc. Natl. Acad. Sci.* **110**, 3665–3672 (2013).
18. Clark, W. C., Van Kerkhoff, L., Lebel, L. & Gallopin, G. C. Crafting usable knowledge for sustainable development. *Proc. Natl. Acad. Sci.* **113**, 4570–4578 (2016).
19. Norström, A. V. *et al.* Principles for knowledge co-production in sustainability research. *Nat. sustainability* **3**, 182–190 (2020).
20. Li, X. & Guo, Y. Paradigm shifts from data-intensive science to robot scientists. *Sci. Bull.* **70**, 14–18 (2025).
21. Gottweis, J. *et al.* Towards an ai co-scientist. *arXiv preprint arXiv:2502.18864* (2025).
22. Xin, H., Kitchin, J. R. & Kulik, H. J. Towards agentic science for advancing scientific discovery. *Nat. Mach. Intell.* 1–3 (2025).
23. Swanson, K., Wu, W., Bulaong, N. L., Pak, J. E. & Zou, J. The virtual lab of ai agents designs new sars-cov-2 nanobodies. *Nature* 1–3 (2025).
24. Ghafarollahi, A. & Buehler, M. J. Sciagents: automating scientific discovery through bioinspired multi-agent intelligent graph reasoning. *Adv. Mater.* **37**, 2413523 (2025).
25. Ghafarollahi, A. & Buehler, M. J. Automating alloy design and discovery with physics-aware multimodal multiagent ai. *Proc. Natl. Acad. Sci.* **122**, e2414074122 (2025).
26. Zhang, Z. *et al.* A multimodal robotic platform for multi-element electrocatalyst discovery. *Nature* 1–3 (2025).
27. Steffen, W. *et al.* *Global change and the earth system: a planet under pressure* (Springer, 2005).
28. Fang, M. & Li, X. Paleoclimate data assimilation: Its motivation, progress and prospects. *Sci. China Earth Sci.* **59**, 1817–1826 (2016).
29. Harrison, S. P. *et al.* Paleoclimate perspectives on contemporary climate change. *Annu. Rev. Environ. Resour.* **50**, 67–95, DOI: 10.1146/annurev-environ-112922-110121 (2025).

30. Committee, N. A. C. E. S. S. *Earth system science overview: a program for global change* (National Aeronautics and Space Administration, 1986).

31. Arrhenius, S. On the influence of carbonic acid in the air upon the temperature of the ground. *Philos. Mag. J. Sci.* **41**, 237–276 (1896).

32. Budyko, M. I. The effect of solar radiation variations on the climate of the Earth. *Tellus* **21**, 611–619 (1969).

33. Bjerknes, J. Atmospheric teleconnections from the equatorial pacific. *Mon. weather review* **97**, 163–172 (1969).

34. Broecker, W. S. The great ocean conveyor. *Oceanography* **4**, 79–89 (1991).

35. Milankovitch, M. Kanon der erdbestrahlung und seine anwendung auf das eiszeitenproblem. *Royal Serbian Acad. Special Publ.* **133**, 1–633 (1941).

36. Hays, J. D., Imbrie, J. & Shackleton, N. J. Variations in the Earth's orbit: Pacemaker of the ice ages. *Science* **194**, 1121–1132 (1976).

37. Manabe, S. & Wetherald, R. T. Thermal equilibrium of the atmosphere with a given distribution of relative humidity. *J. Atmospheric Sci.* **24**, 241–259 (1967).

38. Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A. & Totterdell, I. J. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature* **408**, 184–187 (2000).

39. Stommel, H. Thermohaline convection with two stable regimes of flow. *Tellus* **13**, 224–230 (1961).

40. Suarez, M. J. & Schopf, P. S. A delayed action oscillator for enso. *J. Atmospheric Sci.* **45**, 3283–3287 (1988).

41. Jin, F.-F. An equatorial ocean recharge paradigm for enso. part i: Conceptual model. *J. atmospheric sciences* **54**, 811–829 (1997).

42. Nordhaus, W. D. An optimal transition path for controlling greenhouse gases. *Science* **258**, 1315–1319 (1992).

43. Weyant, J. Some contributions of integrated assessment models of global climate change. *Rev. Environ. Econ. Policy* (2017).

44. Guo, H. *et al.* Measuring and evaluating sdg indicators with big earth data. *Sci. Bull.* **67**, 1792–1801 (2022).

45. Price, I. *et al.* Probabilistic weather forecasting with machine learning. *Nature* **637**, 84–90 (2025).

46. Allen, A. *et al.* End-to-end data-driven weather prediction. *Nature* **641**, 1172–1179 (2025).

47. Lam, R. *et al.* Learning skillful medium-range global weather forecasting. *Science* **382**, 1416–1421 (2023).

48. Glawion, L., Polz, J., Kunstmann, H., Fersch, B. & Chwala, C. Global spatio-temporal era5 precipitation downscaling to km and sub-hourly scale using generative ai. *npj Clim. Atmospheric Sci.* **8**, 219 (2025).

49. Schmidt, J., Schmidt, L., Strnad, F. M., Ludwig, N. & Hennig, P. A generative framework for probabilistic, spatiotemporally coherent downscaling of climate simulation. *npj Clim. Atmospheric Sci.* **8**, 270 (2025).

50. Hess, P., Aich, M., Pan, B. & Boers, N. Fast, scale-adaptive and uncertainty-aware downscaling of earth system model fields with generative machine learning. *Nat. Mach. Intell.* 1–11 (2025).

51. Dheeshjith, S. *et al.* Samudra: An ai global ocean emulator for climate. *Geophys. Res. Lett.* **52**, e2024GL114318 (2025).

52. Yu, S. *et al.* Climsim: A large multi-scale dataset for hybrid physics-ml climate emulation. *Adv. neural information processing systems* **36**, 22070–22084 (2023).

53. Zhao, T. *et al.* Artificial intelligence for geoscience: Progress, challenges, and perspectives. *The Innov.* **5** (2024).

54. Bazgir, A., Zhang, Y. *et al.* Agentichypothesis: A survey on hypothesis generation using llm systems. *Towards Agentic AI for Sci. Hypothesis Gener. Comprehension, Quantification, Valid.* (2025).

55. Feng, Y. *et al.* Rethinking llm uncertainty: A multi-agent approach to estimating black-box model uncertainty. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, 12349–12375 (2025).

56. Najjar, A. A., Ashqar, H. I., Darwish, O. & Hammad, E. Leveraging explainable ai for llm text attribution: Differentiating human-written and multiple llm-generated text. *Information* **16**, 767 (2025).

57. Lin, Z. *et al.* Geogalactica: A scientific large language model in geoscience. *arXiv preprint arXiv:2401.00434* (2023).

58. Acharya, D. B., Kuppan, K. & Divya, B. Agentic ai: Autonomous intelligence for complex goals–a comprehensive survey. *IEEe Access* (2025).

59. Hosseini, S. & Seilani, H. The role of agentic ai in shaping a smart future: A systematic review. *Array* 100399 (2025).

60. Bauer, P. *et al.* The digital revolution of earth-system science. *Nat. Comput. Sci.* **1**, 104–113 (2021).

61. Steffen, W. *et al.* The emergence and evolution of earth system science. *Nat. Rev. Earth & Environ.* **1**, 54–63 (2020).

62. Reid, W. V. *et al.* Earth system science for global sustainability: grand challenges. *Science* **330**, 916–917 (2010).

63. Zenil, H. *et al.* The future of fundamental science led by generative closed-loop artificial intelligence. *arXiv preprint arXiv:2307.07522* (2023).

64. Lu, C. *et al.* The ai scientist: Towards fully automated open-ended scientific discovery. *arXiv preprint arXiv:2408.06292* (2024).

65. Chen, W., Hao, X., Wu, Y. & Liang, Y. Terra: A multimodal spatio-temporal dataset spanning the earth. *Adv. Neural Inf. Process. Syst.* **37**, 66329–66356 (2024).

66. Dalla Mura, M. *et al.* Challenges and opportunities of multimodality and data fusion in remote sensing. *Proc. IEEE* **103**, 1585–1601 (2015).

67. Jiang, S. *et al.* How interpretable machine learning can benefit process understanding in the geosciences. *Earth's Futur.* **12**, e2024EF004540 (2024).

68. Toms, B. A., Barnes, E. A. & Ebert-Uphoff, I. Physically interpretable neural networks for the geosciences: Applications to earth system variability. *J. Adv. Model. Earth Syst.* **12**, e2019MS002002 (2020).

69. Liu, T. *et al.* Teleconnections among tipping elements in the earth system. *Nat. Clim. Chang.* **13**, 67–74 (2023).

70. Ding, J. *et al.* Understanding emergence in complex systems using abductive ai. *Nat. Rev. Phys.* 1–3 (2025).

71. Weichselgartner, J. & Kasperson, R. Barriers in the science-policy-practice interface: Toward a knowledge-action-system in global environmental change research. *Glob. environmental change* **20**, 266–277 (2010).

72. Popper, K. *Conjectures and refutations: The growth of scientific knowledge* (routledge, 2014).

73. Wegener, A. Die entstehung der kontinente. *Geol. Rundschau* **3**, 276–292 (1912).

74. Jeffreys, H. *The Earth: its origin, history and physical constitution* (University Press, 1929).

75. Vine, F. J. & Matthews, D. H. *Magnetic anomalies over oceanic ridges* (Nature Publishing, 1963).

76. Hess, H., Engel, A., James, H. & Leonard, B. History of ocean basins. *Off. Nav. Res. Dep. Navy* (1962).

77. Cai, W. *et al.* Pantropical climate interactions. *Science* **363**, eaav4236 (2019).

78. Hernández, A. *et al.* Modes of climate variability: Synthesis and review of proxy-based reconstructions through the holocene. *Earth-Science Rev.* **209**, 103286 (2020).

79. Vialard, J. *et al.* The el niño southern oscillation (enso) recharge oscillator conceptual model: Achievements and future prospects. *Rev. Geophys.* **63**, e2024RG000843 (2025).

80. Zhao, S. *et al.* Explainable el niño predictability from climate mode interactions. *Nature* **630**, 891–898 (2024).

81. Armstrong McKay, D. I. *et al.* Exceeding 1.5 c global warming could trigger multiple climate tipping points. *Science* **377**, eabn7950 (2022).

82. Brovkin, V. *et al.* Past abrupt changes, tipping points and cascading impacts in the earth system. *Nat. Geosci.* **14**, 550–558 (2021).

83. Scheffer, M. *et al.* Early-warning signals for critical transitions. *Nature* **461**, 53–59 (2009).

84. Fang, S., Wang, Z., Kurths, J. & Fan, J. Tipping points and cascading transitions: Methods, principles, and evidences. *arXiv preprint arXiv:2511.01168* (2025).

85. Crutzen, P. J. Geology of mankind. In *Paul J. Crutzen: A pioneer on atmospheric chemistry and climate change in the Anthropocene*, 211–215 (Springer International Publishing Cham, 2016).

86. Eyring, V. *et al.* Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geosci. Model. Dev.* **9**, 1937–1958 (2016).

87. Calvin, K. & Bond-Lamberty, B. Integrated human-earth system modeling—state of the science and future directions. *Environ. Res. Lett.* **13**, 063006 (2018).

88. Motesharrei, S. *et al.* Modeling sustainability: population, inequality, consumption, and bidirectional coupling of the earth and human systems. *Natl. Sci. Rev.* **3**, 470–494 (2016).

89. Tan, X. *et al.* Leveraging artificial intelligence for research and action on climate change: opportunities, challenges, and future directions. *Sci. Bull.* **70**, 2886–2893 (2025).

90. Rolnick, D. *et al.* Tackling climate change with machine learning. *ACM Comput. Surv. (CSUR)* **55**, 1–96 (2022).

91. Bell, A. *et al.* Earth ai: Unlocking geospatial insights with foundation models and cross-modal reasoning. *arXiv preprint arXiv:2510.18318* (2025).

92. IPCC. *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge, UK and New York, NY, USA, 2022). Doi: 10.1017/9781009157926.

93. Schweizer, V. J. & O'Neill, B. C. Systematic construction of global socioeconomic pathways using internally consistent element combinations. *Clim. Chang.* **122**, 431–445 (2014).

94. Samir, K. & Lutz, W. The human core of the shared socioeconomic pathways: Population scenarios by age, sex and level of education for all countries to 2100. *Glob. Environ. Chang.* **42**, 181–192 (2017).

95. Dellink, R., Chateau, J., Lanzi, E. & Magné, B. Long-term economic growth projections in the shared socioeconomic pathways. *Glob. Environ. Chang.* **42**, 200–214 (2017).

96. Pirani, A. *et al.* Scenarios in ipcc assessments: Lessons from ar6 and opportunities for ar7. *npj Clim. Action* **3**, 1 (2024).

97. Dyer, J. *et al.* Interventionally consistent surrogates for complex simulation models. *Adv. Neural Inf. Process. Syst.* **37**, 21814–21841 (2024).

98. Li, P. *et al.* Using deep learning to generate key variables in global mitigation scenarios. *Nat. Clim. Chang.* 1–9 (2025).

99. Xu, J. Conflicts in multi-level governance: an analysis of international climate policy implementation at the sub-national level. *Glob. Public Policy Gov.* **1**, 401–420 (2021).

100. Xu, L. *et al.* Resilience of renewable power systems under climate risks. *Nat. Rev. Electr. Eng.* **1**, 53–66 (2024).

101. Shen, J. *et al.* Quantifying the impact of extreme weather on china's hydropower–wind–solar renewable energy system. *Nat. Water* 1–15 (2025).

102. Jiang, K. *et al.* Spatiotemporal assessment of renewable adequacy during diverse extreme weather events in china. *Nat. Commun.* **16**, 5198 (2025).

103. Zheng, D. *et al.* Climate change impacts on the extreme power shortage events of wind-solar supply systems worldwide during 1980–2022. *Nat. Commun.* **15**, 5225 (2024).

104. Thevenard, D. & Pelland, S. Estimating the uncertainty in long-term photovoltaic yield predictions. *Sol. energy* **91**, 432–445 (2013).

105. Lee, J. C. & Fields, M. J. An overview of wind energy production prediction bias, losses, and uncertainties. *Wind. Energy Sci. Discuss.* **2020**, 1–82 (2020).

106. Xi, Y. *et al.* A satellite imagery dataset for long-term sustainable development in united states cities. *Sci. data* **10**, 866 (2023).

107. Hersbach, H. *et al.* The era5 global reanalysis. *Q. journal royal meteorological society* **146**, 1999–2049 (2020).

108. He, J. *et al.* The first high-resolution meteorological forcing dataset for land process studies over china. *Sci. data* **7**, 25 (2020).

109. Yang, Y., Lin, S., Lu, R. & Liu, X. CPVPD-2024: A Chinese photovoltaic plant dataset derived via a topography-enhanced deep learning framework. *Sci. Data* **12**, 1601 (2025).

110. Huang, B. & Wang, J. Applications of physics-informed neural networks in power systems–a review. *IEEE Transactions on Power Syst.* **38**, 572–588 (2022).

111. Huppmann, D. *et al.* The messageix integrated assessment model and the ix modeling platform (ixmp): An open framework for integrated and cross-cutting analysis of energy, climate, the environment, and sustainable development. *Environ. Model. & Softw.* **112**, 143–156 (2019).

112. Meinshausen, M., Raper, S. C. & Wigley, T. M. Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, magicc6–part 1: Model description and calibration. *Atmospheric Chem. Phys.* **11**, 1417–1456 (2011).

113. Smith, C. J. *et al.* Fair v1. 3: a simple emissions-based impulse response and carbon cycle model. *Geosci. Model. Dev.* **11**, 2273–2297 (2018).

114. Friedl, A., Kübler, F., Scheidegger, S. & Usui, T. Deep uncertainty quantification: With an application to integrated assessment models. *Work. Pap., Univ. Lausanne, Lausanne, Switz* (2023).

115. Li, K., Zhang, T. & Wang, R. Deep reinforcement learning for multiobjective optimization. *IEEE transactions on cybernetics* **51**, 3103–3114 (2020).

116. Sun, Z. & Yang, Y. Difusco: Graph-based diffusion solvers for combinatorial optimization. *Adv. neural information processing systems* **36**, 3706–3731 (2023).

117. Peirce, C. S. *Collected papers of charles sanders peirce*, vol. 5 (Harvard University Press, 1935).

118. Bright-Thonney, S. *et al.* Autoscidact: Automated scientific discovery through contrastive embedding and hypothesis testing. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems* (2025).

119. Callaghan, M. *et al.* Machine-learning-based evidence and attribution mapping of 100,000 climate impact studies. *Nat. climate change* **11**, 966–972 (2021).

120. Meyer, C., Kreft, H., Guralnick, R. & Jetz, W. Global priorities for an effective information basis of biodiversity distributions. *Nat. communications* **6**, 1–8 (2015).

121. Roemmich, D. *et al.* On the future of argo: A global, full-depth, multi-disciplinary array. *Front. Mar. Sci.* **6**, 439 (2019).

122. Meredith, M. *et al.* Polar regions. chapter 3, ipcc special report on the ocean and cryosphere in a changing climate. (2019).

123. Manheim, D. & Garrabrant, S. Categorizing variants of goodhart's law. *arXiv preprint arXiv:1803.04585* (2018).

124. Al Khourdajie, A. The role of artificial intelligence in climate change scientific assessments. *PLOS Clim.* **4**, e0000706 (2025).

125. Machen, R. & Pearce, W. Anticipating the challenges of ai in climate governance: An urgent dilemma for democracies. *Wiley Interdiscip. Rev. Clim. Chang.* **16**, e70002 (2025).

126. Davies, A. *et al.* Advancing mathematics by guiding human intuition with ai. *Nature* **600**, 70–74 (2021).

127. Krenn, M. *et al.* On scientific understanding with artificial intelligence. *Nat. Rev. Phys.* **4**, 761–769 (2022).

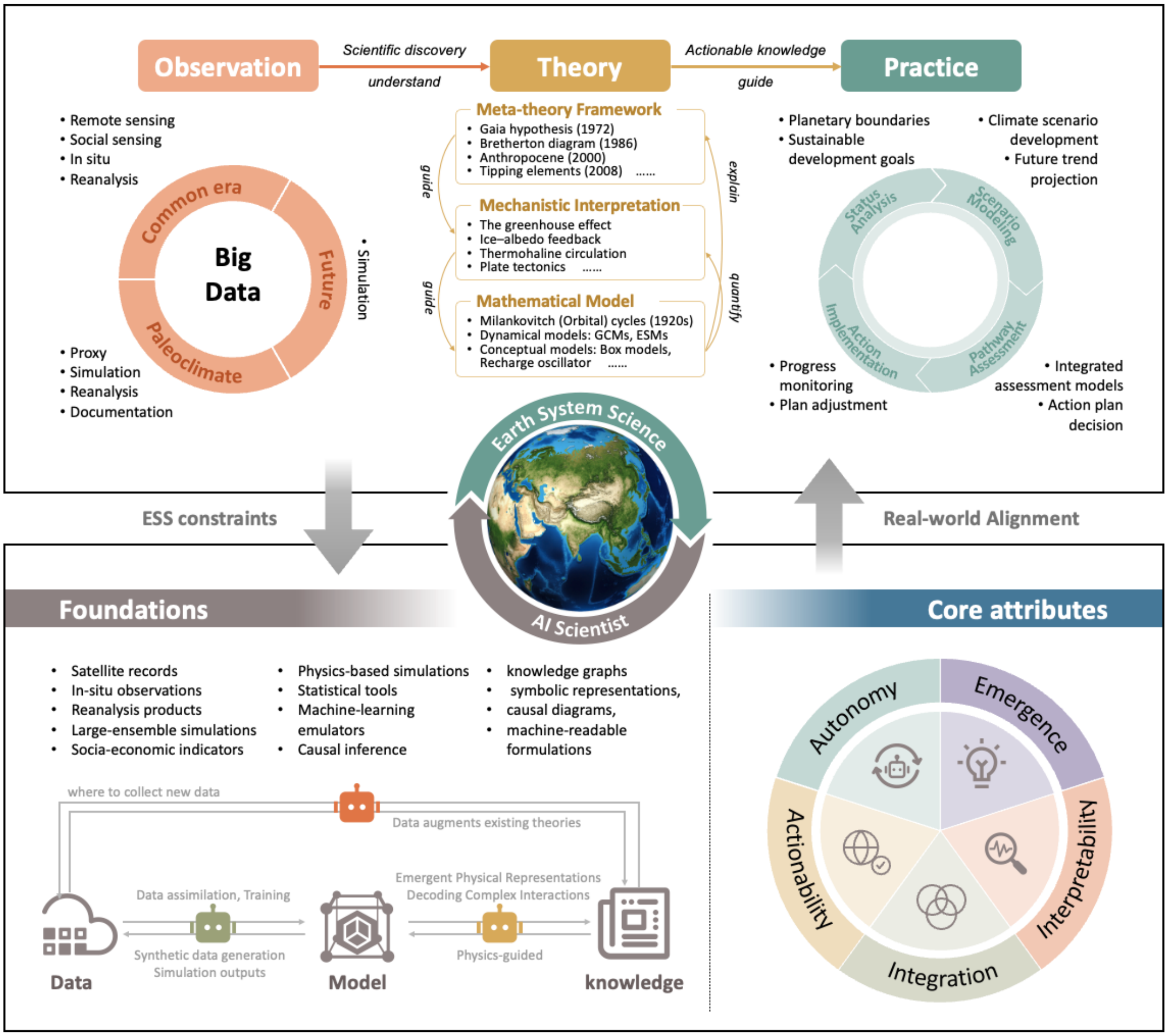

Figure 1. Conceptual framework of the AI Earth System Scientist (AI-ESS). Earth system science defines the scientific questions and constraints, which guide the design of foundational substrates for AI-ESS with unified observational data, heterogeneous models, structured knowledge, and agentic workflows. Five core attributes (autonomy, integration, interpretability, emergence, and actionability) shape these foundations and ensure that the resulting system delivers two kinds of gains: autonomous scientific discovery and accelerated sustainability transformations at the Earth system scale.

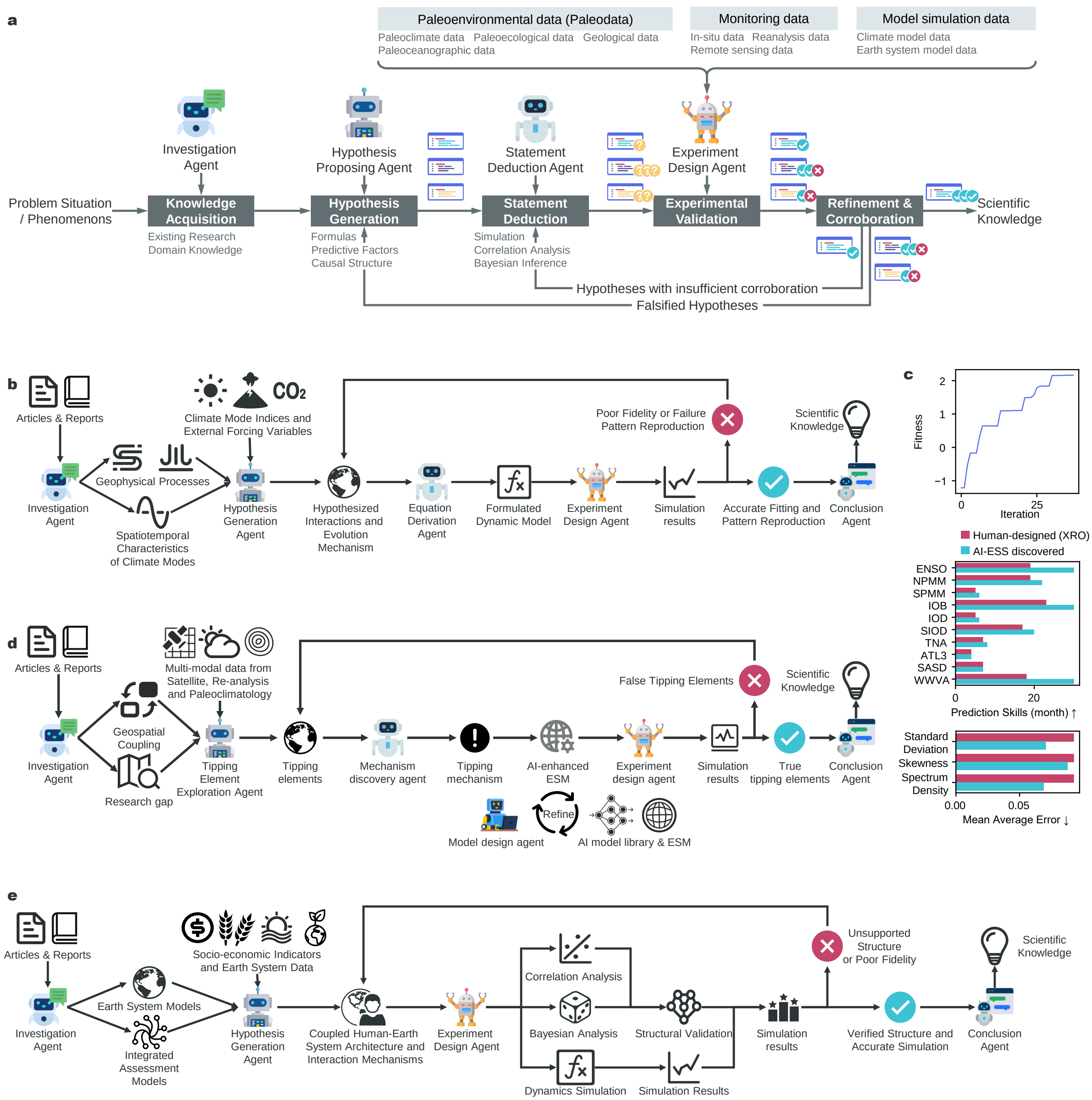

Figure 2. (a) Overview of the AI-ESS workflow for automated discovery in Earth system science. (b—e) Three examples showcasing AI-ESS's capability in autonomous scientific discovery: Discovery of the dynamics of global climate mode networks (b) and the experimental results (c), identification of possible new tipping elements (d), and modeling of the coupled human-Earth system (e).

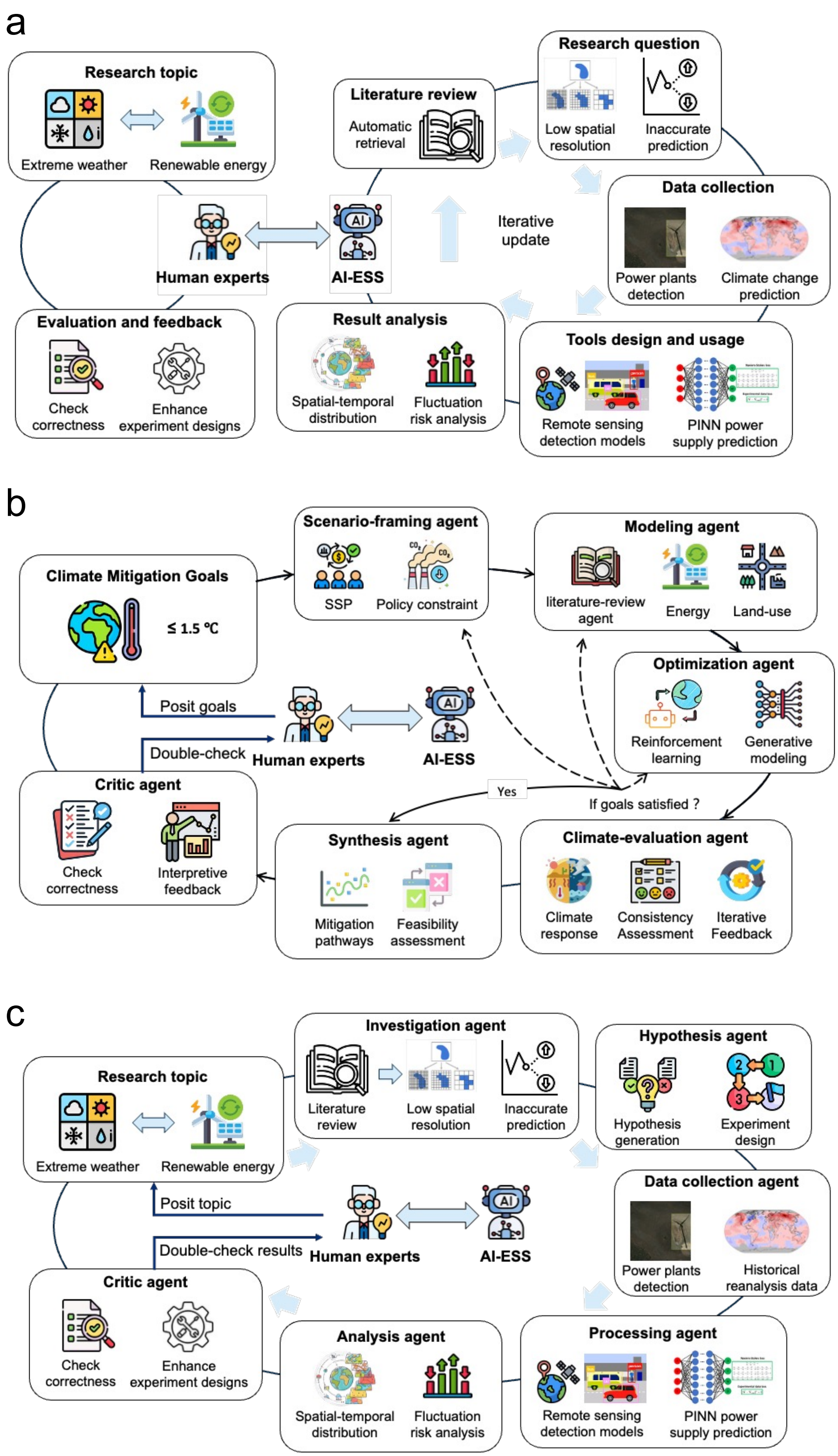

Figure 3. Three examples showcasing AI-ESS's capability in accelerating sustainability transformations. (a) Assessments of climate impacts on renewable energy. (b) Construction of global mitigation pathways. (c) Comprehensive monitoring and evaluation of sustainable development progress.