Discover Artificial Intelligence

https://doi.org/10.1007/s44163-026-01045-1

Article in Press

Neural artificial intelligence framework for integrated sustainability governance under planetary boundaries and ethical constraints

Received: 17 November 2025

Accepted: 13 February 2026

Published online: 04 March 2026

Cite this article as: Rabbi M.F. Neural artificial intelligence framework for integrated sustainability governance under planetary boundaries and ethical constraints. *Discov Artif Intell* (2026). https://doi.org/10.1007/s44163-026-01045-1

Mohammad Fazle Rabbi

Neural Artificial Intelligence Framework for Integrated Sustainability Governance Under Planetary Boundaries and Ethical Constraints

Mohammad Fazle Rabbi[1],

[1] Coordination and Research Centre for Social Sciences, Faculty of Economics and Business, University of Debrecen, Böszörményi út 138, 4032 Debrecen, Hungary.

Corresponding author: Mohammad Fazle Rabbi (e-mail: drrabbikhan@gmail.com; rabbi.mohammad@econ.unideb.hu).

Abstract: Governing sustainability under planetary boundaries and ethical constraints requires integrated computational architectures capable of navigating trade-offs across five dimensions (ecological integrity, economic viability, social equity, human health, and animal welfare) simultaneously. This study develops a neural artificial intelligence framework for culturally discounted sustainability networks (AI-CDSN) that operationalizes Doughnut Economics principles through five coupled modules: neural ordinary differential equations with ecological regularization for capital-resource forecasting, graph attention networks quantifying asymmetric cross-dimensional spillovers, NSGA-III multi-objective optimization, Deep Q-Network reinforcement learning for adaptive policy allocation, and sigmoid discounting embedding intergenerational equity principles. Regional simulations across four implementations achieve robust trajectory prediction with narrow confidence intervals. Capital stocks converge at negative 8.4 with standard deviation 0.2, renewables at negative 2.6 with standard deviation 0.1, and non-renewables at negative 3.5 with standard deviation 0.1 across 100 time steps representing 200-year projections. Material allocation stabilizes at steel 48%, concrete 22%, and timber 30% with coefficient of variation below 2%. Graph attention networks reveal directional asymmetry between ecological integrity and animal welfare dimensions (0.642 versus 0.583). Reinforcement learning navigates multi-objective trade-offs, with mean rewards stabilizing at 1.5 with standard deviation 0.5. Five-dimensional assessment within safe and just operating space demonstrates animal welfare at 0.85, health at 0.78, social equity at 0.64, ecological integrity at 0.61, and economic viability at 0.33. Cultural discounting transitions from 0.38 to 0.97 across generational horizons. Ethical regularization reduces computational energy by 22%. This framework addresses critical gaps in IPCC, IPBES, and UN governance frameworks by embedding intergenerational equity within planetary boundary compliance.

Keywords: Planetary boundaries; Sustainability governance; Neural artificial intelligence; Doughnut Economics; Intergenerational equity

*Complete nomenclature and abbreviations are provided in Appendix A.

1. Introduction

Governing sustainability under planetary boundaries while maintaining social equity requires computational architectures capable of navigating trade-offs across five interconnected dimensions simultaneously [1]. Humanity now operates beyond safe limits for six of nine planetary boundaries, including climate change, biosphere integrity, and biogeochemical cycles, while 1.2 billion people remain below minimum social foundations for health, education, and income [2,3]. This dual challenge defines the safe and just operating space conceptualized in Doughnut Economics [4], where governance must simultaneously avoid ecological overshoot and social deprivation. Traditional sustainability frameworks, including the Planetary Boundaries approach [5,6] and Sustainable Development Goals (SDGs) [7,8], provide normative guidance but lack computational mechanisms to

dynamically balance competing objectives [9], anticipate cascading failures [10], or embed ethical constraints within algorithmic decision structures [11]. This study addresses these gaps by developing a neural artificial intelligence framework for culturally discounted sustainability networks (AI-CDSN) that integrates five dimensions (5D): ecological integrity, economic viability, social equity, human health, and animal welfare. The AI-CDSN framework provides a unified computational architecture capable of real-time priority adaptation and ethically constrained optimization through coupled neural modules.

Existing AI applications for sustainability remain fragmented across isolated domains despite demonstrating potential to support 134 SDG targets [12,13]. Static indices such as the Human Development Index rely on historical data averaging and simplified three-factor measurements that cannot capture rapid environmental changes, distributional inequalities, or cross-dimensional spillovers. Recent systematic reviews confirm that AI-driven sustainability tools optimize individual objectives without integrated mechanisms for multi-dimensional trade-off navigation [14,15]. This fragmentation proves problematic during crises requiring rapid adaptation, as demonstrated by the 2023 European heatwaves where rigid allocation systems failed to facilitate equitable resource distribution [16]. Furthermore, conventional sustainability models inadequately account for AI's own energy and environmental costs, creating performative contradictions wherein computational governance tools contribute to problems they aim to mitigate [17,18].

Three fundamental disconnects characterize current sustainability governance architectures. First, theoretical fragmentation persists between sustainability science's normative foundations (equity, resilience, planetary stewardship) and AI's technical capabilities, with limited exploration of how neural networks, reinforcement learning, or multi-objective optimization might operationalize ethical trade-offs within planetary boundaries [12,19]. Second, mathematical models for sustainability assessment employ static parameters unable to dynamically adjust to real-time planetary boundary breaches, evolving societal values, or emergent cross-dimensional interactions [20–22]. Third, existing frameworks lack anticipatory capacity for second- and third-order impacts, such as how AI-optimized supply chains might reduce transportation emissions while depleting groundwater reserves through redirected logistics [23–25]. These gaps highlight the absence of scalable computational tools capable of quantifying asymmetric cross-dimensional spillovers, harmonizing conflicting stakeholder incentives, and embedding intergenerational equity within algorithmic optimization aligned with planetary constraints.

Consequently, this overarching investigation addresses three specific research questions. RQ1: Can neural artificial intelligence with dynamic weight adaptation autonomously adjust sustainability priorities in response to real-time planetary boundary data while maintaining social foundation compliance? RQ2: How effectively can neural ordinary differential equations coupled with reinforcement learning model capital-resource dynamics, forecast trajectories within safe operating boundaries, and navigate multi-objective trade-offs compared to conventional linear models? RQ3: To what extent can multi-objective optimization protocols embedding culturally calibrated temporal discounting reconcile conflicting sustainability goals while operationalizing intergenerational equity consistent with Indigenous governance principles?

To address these research questions systematically, three specific objectives are established. Objective 1: Develop and validate dynamic weight adaptation mechanisms enabling real-time sustainability priority recalibration based on planetary boundary proximity and social foundation adequacy, achieving more responsive governance than static indices. Objective 2: Formulate and test integrated neural ordinary differential equation models with ecological regularization that forecast capital-resource trajectories within safe operating boundaries while incorporating Deep Q-Network reinforcement learning for adaptive policy

allocation under evolving constraints. Objective 3: Design and assess multi-objective optimization protocols employing NSGA-III that embed sigmoid cultural discounting calibrated through Indigenous Māori tikanga principles, enabling algorithmic decision-making that balances present needs with intergenerational equity across 200-year horizons.

The framework advances three testable hypotheses guide empirical validation: H1: AI systems with dynamic weight adaptation will demonstrate improved crisis responsiveness and planetary boundary compliance compared to static sustainability indices, as measured by policy adjustment speed and threshold detection accuracy during simulated ecological stress events. H2: Neural ODE-driven dynamics coupled with reinforcement learning will achieve higher prediction accuracy for capital-resource trajectories (measured by confidence interval width) and more effective multi-objective trade-off navigation (measured by Pareto optimality convergence) than conventional ARIMA baseline models. H3: Multi-objective protocols embedding cultural discounting will produce more equitable temporal distributions than standard exponential or hyperbolic discounting, as measured by intergenerational equity scores and long-term sustainability metric performance.

This research introduces five methodological innovations addressing persistent gaps in AI-driven sustainability governance:

First, boundary-aware neural ordinary differential equations with ecological regularization ensure trajectory forecasts respect planetary limits by embedding differential equation constraints directly within optimization architectures, preventing threshold violations that destabilize Earth systems. This advances beyond conventional forecasting models that treat boundaries as exogenous parameters applied post-hoc to unconstrained predictions.

Second, graph attention networks quantifying asymmetric cross-dimensional spillovers reveal directional interaction strengths across ecological, economic, social, health, and animal welfare dimensions. Unlike traditional sustainability indices assuming symmetric relationships, this approach identifies leverage points where targeted interventions generate cascading co-benefits through quantified attention coefficients.

Third, multi-objective optimization via NSGA-III with autonomous trade-off navigation eliminates manual parameter tuning requirements characterizing conventional implementations. The framework autonomously balances efficiency, carbon costs, and distributional equity through game-theoretic reinforcement learning that internalizes competing objectives as evolving constraints rather than fixed weights [26]

Fourth, sigmoid cultural discounting operationalizing Indigenous Māori tikanga principles embeds intergenerational equity within algorithmic decision structures by transitioning from present-focus to future-prioritization across 200-year horizons. This challenges exponential discounting inherited from neoclassical economics, demonstrating that non-Western temporal ethics can be mathematically formalized and computationally implemented.

Fifth, integrated Doughnut Economics visualization synthesizes performance across social foundations (minimum thresholds) and planetary boundaries (maximum ceilings), enabling policymakers to diagnose gaps, assess proximity to critical limits, and coordinate interventions within the safe and just operating space. This operationalizes Kate Raworth's conceptual framework through quantitative neural intelligence.

Collectively, these innovations demonstrate that neural artificial intelligence can potentially operationalize planetary boundaries and ethical constraints through transparent, adaptive mechanisms in simulation

environments grounded in established sustainability frameworks. The architecture addresses urgent needs identified in IPCC and IPBES assessments for integrative governance tools managing complexity while maintaining ethical accountability [27,28]. By demonstrating that Doughnut Economics principles [4], World3 systems dynamics [29], and Indigenous temporal ethics can be synthesized within unified computational architectures, this research contributes theoretical and methodological foundations for AI-enhanced sustainability governance aligned with United Nations 2030 Agenda objectives [8].

2. Literature review

2.1. Theoretical foundations for integrated sustainability governance

Operationalizing planetary boundaries and ethical constraints through neural artificial intelligence requires synthesizing established sustainability frameworks with advanced computational methods capable of multi-dimensional trade-off navigation. The framework integrates theoretical foundations spanning Doughnut Economics principles [4], World3 system dynamics [29], planetary boundary science, and Indigenous intergenerational ethics with neural ordinary differential equations, graph attention networks, multi-objective optimization, and reinforcement learning. While recent AI sustainability approaches emphasize fairness, interpretability, and emissions reduction, they typically target isolated model improvements rather than integrated governance architectures respecting both ecological ceilings and social foundations simultaneously. This framework proposes to advance beyond fragmented applications by uniting computational modules within a unified optimization space where planetary limits, distributional equity, and intergenerational justice function as embedded constraints rather than exogenous considerations.

Table 1 synthesizes the theoretical and methodological foundations supporting the neural artificial intelligence framework for integrated sustainability governance under planetary boundaries and ethical constraints. Each component addresses specific governance challenges while operating within the safe and just operating space defined by social foundation minimums and planetary boundary ceilings, with explicit acknowledgment of inherent limitations and complementary requirements across modules.

Table 1. Theoretical and methodological foundations for AI-driven sustainability governance under planetary boundaries

Theories	Key Concept	Originator/Proponent	Relevance to AI	Key Limitations
1: Core Sustainability Metric	5D geometric mean with dimension-specific elasticity coefficients	Griggs et al. (2013)	Enables neural networks to optimize across five interconnected dimensions	Static weights; lacks real-time boundary data integration
2: Neural ODE System Dynamics	Neural network parameterization of ecological-economic interactions	Chen et al. (2018)	Captures nonlinear relationships in capital-resource dynamics	Computationally intensive; boundary awareness limitations; interpretability issues
3: NSGA-III Multi-Objective Optimization	Reference-directed many-objective material selection	Deb & Jain (2014)	Navigates Pareto fronts for optimal material compositions	Sensitivity to reference points; poor scalability

				beyond 15–20 objectives
4: Game-Theoretic Resource Allocation	Dynamic utility balancing efficiency, carbon cost, fairness	von Neumann & Morgenstern (1944)	Balances competing objectives through reinforcement learning	Assumes rational actors; static coefficients; exploration-exploitation trade-offs
5: Cross-Dimensional Impact Propagation	Graph attention networks (GATs) for asymmetric dimension impacts	Veličković et al. (2017)	Models spillover effects between sustainability dimensions	Requires extensive training data; lacks theoretical guarantees
6: AI Sustainability Index	Performance vs. resource consumption metric	Schwartz et al. (2020)	Evaluates AI's own environmental and social impact	Limited standardization; computational overhead; greenwashing risks
7: Policy Optimization	SDG-aligned reinforcement learning	Mnih et al. (2015)	Learns policies that satisfy sustainability constraints	Dimensionality challenges; reward function design complexity
8: Dynamic Weight Adaptation	Planetary boundary-informed attention mechanism	Vaswani et al. (2017)	Autonomously adjusts dimension priorities using live environmental data	Real-time data integration challenges; calibration complexity; feedback delays
9: Ethical Regularization	Energy/fairness penalties in loss functions	Henderson et al. (2020)	Embeds sustainability and equity into AI training	Performance-efficiency trade-offs; incomplete ethical frameworks
10: Cultural Discounting	Sigmoid-based intergenerational equity weighting for temporal discounting	Ramsey (1928); conceptual extension with Indigenous frameworks (present work)	Balances present and future welfare in decision-making	Parameter sensitivity; cultural appropriation risks; stakeholder representation gaps
11: Recent AI Urban Integration	Multi-domain AI orchestration for urban sustainability	Zhang et al. (2022)	Enables coordinated AI deployment across urban systems	Limited cross-dimensional optimization; lacks planetary boundary integration
12: Responsible AI Governance	Context-specific AI principles for sustainable cities	UN-Habitat (2024)	Provides ethical framework for urban AI deployment	Generic frameworks insufficient for multi-objective sustainability trade-offs
13: Sustainable AI Assessment	Multi-dimensional sustainability	Pagliari et al. (2024)	Provides framework for	Limited real-time adaptation; lacks

	metrics for AI systems	evaluating AI sustainability across multiple dimensions	cross-dimensional integration

Note:

- Table 1 presents 13 core theories and methodological components from which the neural AI-CDSN framework is constructed, each contributing functionality for adaptive and boundary-aware sustainability decision support.
- Emphasis is placed on frameworks that operationalize planetary boundaries, ethical and cultural discounting, and neural-based learning algorithms capable of cross-dimensional integration and rapid response.
- Each limitation contextualizes gaps that the integrated architecture addresses, including static weighting, single-objective optimization, and narrow domain focus.
- The "Relevance to AI" column links each building block with capabilities within the neural AI-CDSN, such as ODE-based resource dynamics, GAT-enabled impact propagation, or reinforcement learning for policy adaptation under constraints.
- Cultural discounting parameters (Ramsey, 1928, extended in present work) reflect attention to intergenerational equity and long-term welfare balancing. The framework draws conceptual inspiration from Indigenous concepts of intergenerational responsibility while acknowledging that algorithmic parameter-setting differs from genuine community co-design and democratic deliberation processes emphasized in capability approach literature.
- Originators and proponents are cited as foundational sources for replication, including neural ODE development [30], NSGA-III for multi-objective optimization [31], and contemporary sustainable AI assessment protocols.
- The table supports the claim that sustainable AI decision support requires both principled theoretical integration and novel machine learning architectures, providing an auditable pathway from foundational concepts to operational neural-AI systems.

Table 1 distills the theoretical and computational foundations supporting the framework. The approach begins with a geometric mean sustainability index with elasticity-based weights for each dimension originally developed by Griggs et al. [32], advancing beyond static measurement through integrated neural ordinary differential equations to model resource transitions [30]. NSGA-III-based optimization [31] enables complex material policy balancing, and graph attention networks [33] capture cross-dimensional impacts between sustainability domains. Utility functions from von Neumann and Morgenstern's game theory are optimized through reinforcement learning [34,35], ensuring that efficiency, emissions, and fairness reflect both environmental priorities and ethical regularization [36].

The framework transcends fragmented AI sustainability research by using attention mechanisms to dynamically integrate boundary signals while extending Ramsey's (1928) [37] temporal discounting model to incorporate intergenerational equity [38]. This neural-AI system demonstrates proof-of-concept functionality for multi-dimensional decision support within simulation environments. However, comprehensive validation including real-world policy implementation, democratic stakeholder engagement, and individual-level capability assessment remains necessary before operational governance applications.

2.2. Multi-dimensional sustainability assessment frameworks

Traditional sustainability metrics, including the Human Development Index and Planetary Boundaries framework, define goals across ecological, economic, and social dimensions [39]. However, they overlook critical considerations such as human health and animal welfare, while their static nature prevents real-time responsiveness to environmental shifts. Multi-dimensional assessment frameworks have addressed this conceptual gap by formalizing five interdependent dimensions: ecological integrity, economic viability, social equity, human health, and animal welfare, with explicit principles and indicators for each. Despite conceptual advances, static weights and reliance on historical data restrict real-world applicability. This proved insufficient during dynamic crises, as evidenced by conventional models' failure to adjust water-rationing policies equitably during the 2023 European heatwaves [40].

Systematic research on SDG interactions has established comprehensive interaction matrices demonstrating trade-offs and synergies across 169 targets [41,42]. Network analyses reveal that SDG interactions vary with income levels and development contexts, requiring adaptive governance frameworks rather than universal prescriptions [43]. While this body of work provides crucial evidence for cross-dimensional dependencies, existing approaches rely primarily on correlation-based statistical methods that assume symmetric relationships and lack real-time adaptive capacity to respond to planetary boundary breaches or evolving social equity indicators.

Recent global assessments demonstrate that responsible AI implementation requires context-specific principles addressing diverse urban contexts, with emphasis on long-term sustainability and embedding sustainability into AI strategies [44]. Current approaches reveal limitations in generic AI governance frameworks, particularly in addressing complex interplay between economic, environmental, and social sustainability dimensions simultaneously. While artificial intelligence has emerged in sustainability science through domain-specific applications such as energy grid optimization and ecological forecasting, early frameworks often treated sustainability dimensions as isolated variables through single-objective models. Even with advancements such as neural ordinary differential equations enabling dynamic modeling of nonlinear interdependencies, existing AI implementations remain confined to single dimensions, neglecting critical cross-system feedback loops essential for holistic governance. Contemporary research indicates that AI can support 134 SDG targets while potentially impeding progress on 59 targets, underscoring the importance of ensuring technology catalyzes positive transformation rather than creating unintended consequences [44].

This framework integrates adaptive weighting directly into its core metric (Equation 1), facilitating dynamic priority adjustment in response to real-time planetary boundary signals. This innovation mitigates constraints of static models while overcoming challenges posed by fragmented AI applications, enabling integrated optimization across sustainability dimensions through graph attention networks that quantify asymmetric cross-dimensional spillovers and neural ordinary differential equations that learn nonlinear dynamics from data rather than assuming predetermined functional forms.

2.3. Neural ordinary differential equations for socio-ecological system dynamics

The AI-CDSN neural ODE architecture builds upon the system dynamics tradition pioneered by World3 [29,45,46], which introduced coupled differential equations for capital-resource interactions with remarkable long-term predictive accuracy validated over five decades [47,48]. World3 demonstrated that capital stock, renewable resources, and non-renewable resources exhibit nonlinear feedback dynamics that govern long-term sustainability trajectories through monotonic convergence to attractor states. While

World3 employed hand-coded lookup tables to represent these nonlinear relationships based on expert judgment, our neural ODE approach learns dynamics directly from data through gradient descent optimization. This preserves World3's fundamental stock-flow architecture and 200-year intergenerational time horizon while enabling data-driven parameter estimation rather than predetermined functional forms **(see Supplementary Table S3 for systematic comparison with World3 and conventional integrated assessment models)**.

Neural ordinary differential equations (neural ODEs) have emerged as a powerful approach to learn continuous-time, nonlinear system dynamics directly from observational data. Chen et al. (2018) demonstrated their capacity to capture ecological population fluctuations with greater fidelity than linear differential equation systems, and subsequent applications have shown improved forecasting accuracy in controlled environmental systems [30]. However, these implementations remain confined to single-dimensional problems and lack built-in boundary awareness mechanisms necessary for planetary-scale sustainability governance. Recent comparative studies indicate that neural ODEs demonstrate superior forecast accuracy compared to traditional ARIMA models for ecological system dynamics, yet most applications remain compartmentalized without the seamless integration required for adaptive governance tools [49].

The AI-CDSN framework extends neural ODEs to model coupled dynamics of capital accumulation, renewable resource regeneration, and non-renewable resource depletion (Equation 2), incorporating a boundary-gating mechanism that ensures learned trajectories respect planetary limits identified in the World3 tradition. This approach maintains deterministic convergence properties observed in classic system dynamics models while leveraging neural networks to learn complex nonlinear interactions from data. While neural ODEs capture temporal system behavior, multi-objective optimization resolves competing material and policy priorities across sustainability dimensions.

2.4. Multi-objective optimization and game-theoretic resource allocation

Multi-objective optimization utilizing NSGA-III algorithms has become crucial for balancing competing sustainability objectives. These algorithms have achieved significant successes, including 30% embodied carbon reductions in construction materials while maintaining cost efficiency [50], and successful reconciliation of energy savings with economic viability in building retrofits [51]. However, while NSGA-III excels at navigating high-dimensional Pareto fronts [31], challenges remain beyond 15 to 20 objectives, and sensitivity to reference vector choice limits adaptability.

Recent advances in AI orchestration for urban sustainability demonstrate that coordinated deployment across multiple systems can achieve significant improvements, yet current approaches lack cross-dimensional optimization capabilities necessary for comprehensive sustainability governance [52]. Contemporary research reveals that AI applications encounter persistent implementation challenges, including insufficient standardized regulations and expertise gaps, particularly in developing regions [44].

Game-theoretic deep reinforcement learning has enabled equitable resource allocation, exemplified by applications in geo-distributed data centers achieving significant carbon footprint reductions through optimized resource management [53,54]. Yet many models fall short by prioritizing technical metrics over comprehensive ethical considerations [55], with few frameworks adequately incorporating crucial indicators such as animal welfare or intergenerational equity. They also typically rely on manual parameter tuning and assume perfectly rational actors, limiting real-world applicability. To achieve responsible AI in cities, holistic approaches that integrate multidisciplinary perspectives and multi-stakeholder engagement

are essential. Despite their importance, these capabilities and the integration of existing technologies remain largely unexplored in current urban AI initiatives [56].

This framework advances these methods by combining adaptive NSGA-III reference vectors (Equation 3) with game-theoretic Deep Q-Network allocation (Equation 4), enabling semi-autonomous trade-off navigation within simulation environments, though operational implementation requires human oversight and democratic governance integration, thereby integrating efficiency, carbon cost, and fairness dynamically.

2.5. Computational sustainability paradox and ethical integration

While 90% of corporations recognize AI's criticality for environmental strategies, artificial intelligence's substantial ecological footprint risks undermining sustainability gains [57]. Training large language models can emit CO_2 equivalent to multiple average U.S. annual carbon footprints [58], yet few deployment frameworks adequately account for these impacts. This computational sustainability paradox creates self-defeating outcomes where AI-driven energy savings are negated by increased hardware demands [59]. Recent policy research emphasizes that embedding sustainability into AI strategies requires recognizing limitations of generic governance frameworks and establishing context-specific responsible AI principles prioritizing trustworthiness and accountability [44].

Existing AI-driven sustainability models largely overlook long-term intergenerational equity, with conventional discounting schemes (exponential or hyperbolic) systematically devaluing future welfare while neglecting critical ethical considerations such as animal welfare and temporal justice. The computational sustainability paradox remains unresolved in current frameworks: while 90% of corporations recognize AI's criticality for environmental strategies, training large language models can emit CO_2 equivalent to multiple average U.S. annual carbon footprints, yet few deployment frameworks adequately account for these impacts. Contemporary assessments reveal that cities must address capacity gaps and mitigate risks related to privacy, ethics, bias, discrimination, and misinformation while implementing AI solutions [56].

Addressing these challenges requires integrated approaches that embed ethical constraints directly within optimization architectures rather than treating them as post-hoc adjustments [36]. Recent policy research emphasizes that embedding sustainability into AI strategies requires moving beyond generic governance frameworks toward context-specific responsible AI principles prioritizing trustworthiness and accountability [60]. Indigenous knowledge systems, particularly concepts of intergenerational guardianship, offer alternative temporal frameworks that contrast with exponential discounting inherited from neoclassical economics [61,62]. However, mathematical formalization and computational operationalization of these non-Western ethical principles within AI decision structures remain largely unexplored in current literature [63].

2.6. Research gaps in integrated computational sustainability governance

Despite individual methodological advances, computational sustainability governance faces three fundamental disconnects preventing truly integrated systems capable of operationalizing planetary boundaries and ethical constraints simultaneously.

The first gap concerns dynamic adaptation mechanisms for sustainability priorities. Existing AI frameworks lack capabilities to automatically adjust priorities based on real-time planetary boundary breaches and social equity indicators. Neural ODEs demonstrate superior forecasting accuracy compared to ARIMA models

for ecological system dynamics, yet most applications remain compartmentalized without seamless integration of real-time data streams from platforms such as Copernicus [64]. Static weight allocations render indices ill-equipped for environmental volatility, as evidenced during the 2023 European heatwaves where fixed models inadvertently amplified water rationing disparities [40,44]. While cities implement diverse AI solutions with benefits including improved efficiency and enhanced management, fragmented approaches constrain comprehensive sustainability governance.

Second, beyond adaptation limitations, existing approaches lack boundary-awareness mechanisms ensuring system trajectories respect ecological limits during optimization. No current approach integrates neural ordinary differential equations with planetary boundary constraints to ensure forecasted trajectories remain within safe operating spaces. While neural ODEs demonstrate superior forecasting capabilities for individual domains, existing implementations lack boundary-gating mechanisms enforcing ecological thresholds during system evolution. Hybrid architectures combining neural ODEs with graph neural networks show promise for capturing cross-dimensional interactions in spatiotemporal modeling, yet comprehensive boundary-aware models suitable for five-dimensional sustainability challenges remain an unmet need [65].

The third critical gap involves autonomous trade-off navigation across competing sustainability objectives. Contemporary multi-objective algorithms cannot balance efficiency, carbon cost, and fairness without manual parameter tuning and human intervention. Current NSGA-III implementations excel at navigating Pareto fronts but suffer from reference vector sensitivity and poor scalability beyond 15 to 20 objectives [31]. Game-theoretic approaches typically assume perfectly rational actors with static utility functions, limiting applicability where preferences evolve and information remains incomplete. The absence of autonomous negotiation capabilities prevents effective resource conflict resolution across sustainability dimensions.

These technical limitations are compounded by theoretical fragmentation between sustainability science's normative foundations (planetary boundaries, Doughnut Economics, intergenerational equity) and AI's computational capabilities. The field lacks rigorous mathematical articulation synthesizing sophisticated neural architectures with comprehensive sustainability metrics, hindering development of frameworks operationalizing complex principles through AI-driven optimization.

This framework addresses these specific gaps through three integrated innovations. First, dynamic weight adaptation informed by real-time planetary boundary data enables autonomous priority recalibration. Second, boundary-aware neural ODEs with ecological regularization ensure trajectory forecasts respect planetary limits. Third, integrated multi-objective optimization combines NSGA-III with Deep Q-Network reinforcement learning for autonomous trade-off navigation embedding ethical constraints. Unlike domain-specific applications, this enables cross-dimensional optimization within the safe and just operating space. Unlike static frameworks, this provides adaptive governance responding to evolving conditions. Unlike conventional approaches, this embeds cultural discounting and intergenerational equity directly within algorithmic decision structures, advancing toward holistic, boundary-aware, ethically grounded sustainability governance critically needed in contemporary research [28,56].

2.7. Algorithmic governance and democratic legitimacy in AI-driven sustainability

Recent scholarship on algorithmic governance reveals tensions between computational optimization and democratic values. Eubanks (2018) documents how automated decision systems in social services can reinforce inequality through opacity and unaccountability, demonstrating that technical efficiency does not

guarantee equitable outcomes [66]. O'Neil (2016) identifies "weapons of math destruction" where algorithmic scoring perpetuates discrimination in domains ranging from criminal justice to employment, highlighting risks when quantitative models encode and amplify existing biases [67]. Zuboff (2019) critiques surveillance capitalism underlying many AI systems, arguing that data extraction and behavioral prediction undermine individual autonomy and democratic deliberation [68].

In sustainability governance specifically, Pasquale (2020) warns that algorithmic regulation risks technocratic capture, privileging quantifiable metrics over qualitative values requiring democratic deliberation [69]. Couldry and Mejias (2019) identify data colonialism in smart city initiatives that extract community knowledge without reciprocal benefit, reproducing colonial patterns of exploitation through digital infrastructure [70]. These critiques emphasize that computational decision support cannot substitute for participatory governance processes, transparent accountability mechanisms, and democratic contestation of value priorities.

The AI-CDSN framework acknowledges these concerns through transparency requirements enabling stakeholder interrogation of attention visualizations (Figures 6-7), contestability mechanisms allowing democratic override of dynamic weight adaptation (Equation 9), and ethical constraints explicitly addressing equity through fairness regularization (Equation 10). However, computational architecture alone cannot resolve governance legitimacy. Operational implementation requires institutional integration including participatory weight calibration through citizen assemblies, democratic oversight boards governing threshold settings, and legal accountability frameworks enabling appeal of algorithmic recommendations. This framework therefore positions AI as decision support infrastructure for human governance rather than autonomous decision-making, with explicit recognition that sustainability trade-offs ultimately require political negotiation reflecting diverse stakeholder values, power dynamics, and distributional consequences that exceed computational optimization capabilities.

3. Methodology

3.1. Framework overview for integrated sustainability governance

Figure 1 illustrates the computational architecture for five-dimensional sustainability governance under planetary boundaries and ethical constraints. The workflow begins with input data (planetary boundaries, 5D sustainability metrics, and cultural parameters), which feeds into three core analytic modules: the dynamic weighting module (adaptive priority allocation and planetary boundary integration), the 5D metric engine (geometric dynamic fusion and real-time assessment), and the cultural discounting module (intergenerational equity and temporal scaling).

The analytic layer branches to two specialized modules: the neural ODE core (for capital-resource dynamics and boundary-aware forecasting) and the graph attention network (for cross-dimensional impact and systemic risk propagation). The ethical regularizer (green) ensures compliance with energy and bias constraints, reflecting regulations such as the EU AI Act. Policy synthesis is driven by two optimization modules: the multi-objective optimizer (material allocation via NSGA-III) and the game-theoretic allocator (reinforcement learning for efficiency-fairness balance). These modules converge in the policy synthesis module, which generates SDG-aligned solutions and adaptive governance strategies.

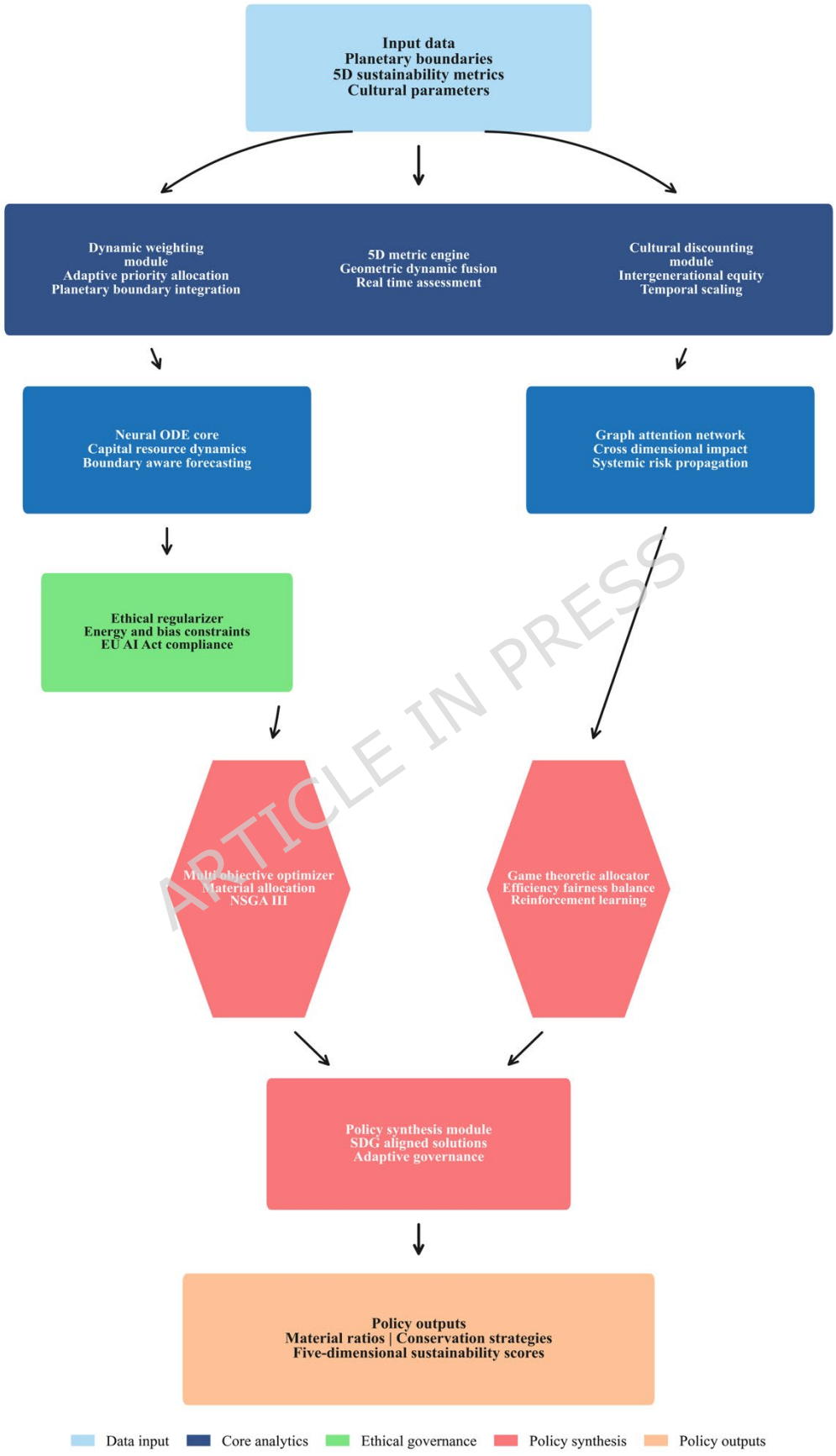


Figure 1. Neural AI framework architecture for integrated five-dimensional sustainability governance

The computational pipeline culminates in policy outputs (orange), which deliver actionable results: material ratios, conservation strategies, and five-dimensional sustainability scores. Color coding organizes the architecture: light blue for data input, dark blue for core analytics, green for ethical governance, pink for policy synthesis, and orange for outputs. Arrows denote information flow, highlighting integration between forecasting, cross-dimensional analytics, and optimization. The modular layout emphasizes transparency and interpretability, facilitating translation into policy design and sustainability planning contexts.

3.2. Core computational modules integrating dynamics, optimization, and ethical constraints

The AI-CDSN framework integrates four computational modules (Equations 1 to 4) built around boundary-aware dynamics (Equation 2). These modules operationalize theoretical foundations from Table 1 through interconnected equations enabling real-time adaptation to planetary signals, ethical constraints, and cross-dimensional feedback. Before detailing individual components, this section formally specifies the complete mathematical framework.

State Variables ($\mathbf{x} \in \mathbb{R}^{\mathbf{n}}$): The system tracks capital stock $K(t) \in \mathbb{R}_+$ normalized to initial value K_0, renewable resources $R(t) \in \mathbb{R}_+$ normalized to biocapacity baseline, non-renewable resources $N(t) \in \mathbb{R}_+$ normalized to proven reserves, five-dimensional sustainability scores $D = [D_{eco}, D_{econ}, D_{soc}, D_{health}, D_{animal}] \in [0,1]^5$, and planetary boundary deviations $PB = [PB_{climate}, PB_{bio}, PB_{nitrogen}, PB_{water}] \in [0,1]^4$ where 0 indicates safe operation and 1 indicates transgression [6].

Control Variables ($\mathbf{u} \in \boldsymbol{\mathcal{U}}$): Decision inputs include material allocation $a = [a_{steel}, a_{concrete}, a_{timber}] \in \Delta^2$ constrained to the 2-simplex ($\sum a_i = 1, a_i \geq 0$), resource extraction rate $e(t) \in [0, e_{max}]$ with bounded limits, and dynamic dimensional weights $w = [w_1, \ldots, w_5] \in \Delta^4$ forming a probability simplex ($\sum w_i = 1$).

Constraint Set: The framework enforces planetary boundaries $PB_j(t) \leq \theta_j^{safe}$ for climate, biodiversity, nitrogen, and water cycles, social foundation minimums $D_i(t) \geq \phi_i^{min}$ across all five dimensions, material requirements including recycled content $\geq 30\%$, and SDG compliance covering labor rights (SDG 8) and circular economy standards (SDG 12) [8].

Objective Functions: Multi-objective optimization via NSGA-III [31] minimizes embodied energy $E(a)$ and construction cost $C(a)$ subject to simplex and SDG constraints: $\min_a \{E(a), C(a)\}$ where $a \in \Delta^2$. Reinforcement learning maximizes utility $U(s,a) = \lambda_{econ} U_{econ}(a) - \lambda_{carbon}(t) C_{carbon}(s,a) + \lambda_{fairness}(t) F_{fairness}(s)$ with time-varying penalty coefficients $\lambda_{carbon}(t)$ and $\lambda_{fairness}(t)$ adapted via LSTM networks (Equation 5) [71].

Information Flow: Real-world sustainability data flow through neural ODEs forecasting capital-resource trajectories $\{K(t), R(t), N(t)\}$, graph attention networks quantifying cross-dimensional impacts, and coupled RL/NSGA-III modules generating policy recommendations. This architecture ensures planetary boundary compliance through ecological regularization (Equation 2), intergenerational equity through cultural discounting (Equation 11), and ethical constraints through computational sustainability penalties (Equation 10). Complete notation definitions appear in Supplementary Table S5.

The four core modules operate as follows: (1) adaptive sustainability metrics with elasticity-based dimensional weights enabling real-time priority recalibration (Section 3.2.1), (2) boundary-aware neural ODEs forecasting capital-resource trajectories under ecological constraints (Section 3.2.2), (3) multi-objective optimization balancing competing material and policy objectives (Section 3.2.3), and (4) game-theoretic reinforcement learning with ethical regularization and cultural discounting (Sections 3.2.4-3.2.10). Each module is detailed below with formal equations, implementation specifications, and theoretical justifications.

3.2.1. Adaptive sustainability metrics with elasticity and real-time weighting

Building on the geometric mean sustainability index formulated by Griggs et al. (2013) [32], Equation 1 introduces dimension-specific elasticity coefficients to dynamically weight five dimensions. Unlike static indices, this approach adapts weights using planetary boundary data, enabling real-time priority shifts during ecological stress events or climate crises.

$$S(t) = \prod_{i=1}^{5} (w_i(t)D_i(t) + \alpha_i)^{\beta_i} \quad (1)$$

The elasticity coefficients are calibrated as $\beta_1 = 1.1$ (ecological integrity), $\beta_2 = 0.9$ (economic viability), $\beta_3 = 1.0$ (social equity), $\beta_4 = 1.2$ (human health), and $\beta_5 = 0.8$ (animal welfare). These values prioritize human health and ecological integrity ($\beta_4 = 1.2, \beta_1 = 1.1$) while maintaining balanced consideration of social and economic factors ($\beta_3 = 1.0, \beta_2 = 0.9$). Animal welfare receives moderate weighting ($\beta_5 = 0.8$) reflecting emerging recognition in sustainability frameworks. The elasticity structure ensures that long-term effects and cross-dimensional temporal interdependencies are properly represented, promoting intergenerational equity by modulating dimensional influence across time horizons. Complete mathematical notation and symbol definitions are provided in Supplementary Table S5.

The weights $w_i(t)$ are adjusted via reinforcement learning to prioritize dimensions based on real-time planetary boundary data. Shift parameters $\alpha_i = 0.1$ ensure no dimension's contribution collapses to zero, maintaining representational stability. This formulation generalizes the original geometric mean approach by integrating adaptive learning, enabling the framework to respond proactively to emerging ecological or social crises.

3.2.2. Modeling capital-resource interactions with neural ODEs

Effective figure axis labeling improves comprehension of model dynamics. Extending Chen et al.'s neural ODEs [30], this component models capital-resource interactions incorporating a boundary-gating mechanism designed to enforce planetary constraints. The model employs a hybrid AI-based formulation that overcomes limitations of traditional differential equations by learning from data-driven patterns rather than imposing rigid functional assumptions.

$$\frac{dK}{dt} = f_\theta(K, R, E) - \delta K + \pi_{\mathrm{RL}}(s_t) \quad (2)$$

The equation governs interactions between capital stock (K), renewable resources (R), and non-renewable extraction (E). This equation models how capital, renewable resources, and extraction rates interact over time, similar to how a city's infrastructure capacity changes based on resource availability and usage patterns.

The neural network f_θ learns nonlinear production functions from historical climate-economic datasets, while δK represents capital depreciation. The reinforcement learning policy π_{RL} dynamically adjusts resource extraction rates (s_t) based on sustainability feedback. A novel ecological regularization term, $\mathcal{L}_{\mathrm{eco}} = \lambda_1 \|\nabla_K f_\theta\|^2$, penalizes capital growth that exacerbates resource depletion, ensuring alignment with planetary boundaries.

3.2.3. Multi-objective evolutionary algorithms for sustainability trade-offs

Adapting Deb & Jain's NSGA-III, Equation 3 automates Pareto-optimal material selection without manual reference tuning [31]. Opposition-based learning accelerates convergence while enforcing Sustainable Development Goal (SDG) constraints on recycled content and labor rights. By incorporating circular economy bounds and opposition-based learning, the approach accelerates convergence and ensures socially just design outcomes.

$$\begin{aligned} &\min E(x) = \sum_{i=1}^{n} e_i x_i \\ &\min C(x) = \sum_{i=1}^{n} c_i x_i \\ &\text{s.t. } g_j(x) \le 0 \ \forall j \in \{\text{SDG 9, 11, 12}\} \end{aligned} \tag{3}$$

The formulation minimizes embodied energy ($E(x)$) and construction costs ($C(x)$) under SDG-aligned constraints ($g_j(x)$). Opposition-based learning accelerates convergence by generating mirror solutions $x_{\mathrm{opp}} = \mathrm{UB} + \mathrm{LB} - x_{\mathrm{current}}$, where UB/LB represent circular economy bounds. Constraints enforce labor rights audits (SDG 8) and recycled material quotas (SDG 12), ensuring solutions balance technical efficiency with social equity.

3.2.4. Game-theoretic reinforcement learning for equitable resource allocation

Von Neumann-Morgenstern utility theory merges with deep reinforcement learning to balance efficiency, carbon cost, and fairness [35]. Reward oscillations expose policy conflicts, while LSTM-adjusted coefficients and align allocations with real-time carbon and equity data. This game operates within the constraints of limited environmental resources and the diverse social needs of the population.

The key innovation lies in how these AI agents make decisions. They dynamically adapt their strategies based on real-time information about carbon pricing and population mobility. The goal is to find a balance between achieving economic efficiency in resource use and ensuring that these resources are distributed fairly across different social groups.

$$U_i(a_i, a_{-i}) = \underbrace{R_{\mathrm{econ}}}_{\text{Economic}} - \lambda(t) \underbrace{C_{\mathrm{carbon}}}_{\text{Ecological}} + \mu(t) \underbrace{F_{\mathrm{fairness}}}_{\text{Social}} \tag{4}$$

This utility function works like a smart resource manager that simultaneously considers economic efficiency (how well resources are used), environmental impact (carbon costs), and social fairness (equitable distribution), with the weights automatically adjusting based on real-time conditions. The utility function U_i balances economic rewards (R_{econ}), carbon costs (C_{carbon}), and fairness indices (F_{fairness}). Dynamic coefficients $\lambda(t)$ and $\mu(t)$ adapt via LSTM networks processing grid carbon intensity (E_{grid}) and social equity metrics.

For example, Dynamic parameters $\lambda(t)$ and $\mu(t)$ adapt via LSTM networks processing real-time grid carbon intensity and Gini index data:

$$\lambda(t) = \sigma\left(W_\lambda \times \left[E_{\text{grid}}(t), \text{RE}_{\%}(t)\right]\right) \tag{5}$$

Here, σ is the sigmoid function, which ensures that the value of $\lambda(t)$ remains within the range of (0,1). W_λ represents the learned weights of the LSTM that are applied to the input vector containing the grid carbon intensity and the renewable energy percentage. By using these real-time inputs and the sigmoid-activated output of the LSTM, the model dynamically adjusts the penalty associated with carbon emissions, making the resource allocation decisions more sensitive to the current environmental context. A similar approach, potentially involving a separate LSTM network and different relevant inputs such as the Gini index, would be used to dynamically adapt the fairness index coefficient γ, ensuring that social equity considerations are also responsive to real-time conditions.

3.2.5. Modeling cross-dimensional impacts using graph attention networks

Graph attention networks (GATs) model interdependencies across sustainability dimensions, overcoming oversimplification of traditional index-based models by capturing nonlinear, empirical correlations among ecological, economic, and social outcomes. GATs quantify asymmetric cross-dimensional effects through learned adjacency matrices, enabling predictive modeling of policy cascades such as renewable energy investments indirectly improving public health outcomes.

$$\frac{\partial S_i}{\partial t} = \sum_{j=1}^{5} \text{GAT}(A_{ij}) \times \frac{dD_j}{dt} \tag{6}$$

The equation quantifies how changes in one dimension (dD_j/dt) propagate to others through learned adjacency matrices A_{ij}. The GAT, trained on 15 years of UNEP case studies, captures nonlinear interactions (e.g., healthcare improvements reducing ecological strain via lower pharmaceutical pollution). This enables predictive modeling of policy cascades across sustainability dimensions [72].

3.2.6. AI sustainability index integrating performance and resource implications

Building on Schwartz et al.'s energy-bias tradeoffs, this study introduces a composite sustainability metric that penalizes excessive computational consumption while maintaining model accuracy [18].

$$\text{AI-SI} = \frac{\text{Accuracy}}{\text{Energy} \times \text{CO}_2} \times \left(1 - \frac{\text{Bias}}{\text{Fairness}_{\text{th}}}\right) \quad (7)$$

The index integrates key factors including energy use, carbon emissions, and algorithmic fairness relative to a predefined threshold. Energy budgets are aligned with IPCC 1.5°C pathways to ensure that AI deployment does not undermine sustainability objectives.

3.2.7. Policy optimization aligned with sustainable development goals

The framework adopts an inverse reinforcement learning (IRL) approach to derive optimal sustainability policies. Inverse reinforcement learning embeds SDG penalties into reward structures, using CMIP6 climate projections to ensure robustness. Dynamic time warping aligns policies with UN progress reports, quantifying violations when trajectories exceed equity thresholds. The five-dimensional framework explicitly operationalizes multiple UN Sustainable Development Goals through computational mechanisms (see Supplementary Table S4 for complete dimension-to-SDG mapping and empirical indicators).

$$\pi^*(a|s) = \arg\max_{\pi} \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t (R_t - \eta \times \text{SDG-Violation}_t)\right] \quad (8)$$

The policy π^* maximizes cumulative rewards (R_t) while penalizing deviations from SDG targets (SDG-Violation_t). Penalty weights η adapt using CMIP6 climate projections, ensuring policies remain robust under future scenarios. Dynamic time warping compares simulated outcomes against UN progress reports to quantify violations. Specifically, ecological integrity addresses SDG 13 (Climate Action) and SDG 15 (Life on Land), economic viability targets SDG 8 (Decent Work) and SDG 9 (Infrastructure), social equity integrates SDG 1 (Poverty) and SDG 10 (Inequality), health aligns with SDG 3 (Health and Well-being), and the energy dimension directly tracks SDG 7 (Clean Energy) through AI's own computational footprint monitoring (Supplementary Table S4).

3.2.8. Dynamic weight adaptation using planetary boundary data

A hybrid multilayer perceptron-graph neural network architecture integrates normalized planetary boundary deviations with social equity trend indicators while ensuring data privacy through homomorphic encryption. This enables the real-time recalibration of sustainability policy weights sensitive to dynamic environmental and social conditions. The planetary boundary deviations incorporate nine Earth system processes, as defined by Richardson et al. (2023), including climate change, biosphere integrity, land-system change, freshwater use, biogeochemical flows, ocean acidification, atmospheric aerosol loading, stratospheric ozone depletion, and novel entities [6]. These deviations are scaled to a normalized range reflecting transgression magnitudes relative to safe planetary limits. Simultaneously, social equity trends are incorporated via a graph neural network processing mobility and demographic data, with encrypted integration ensuring confidentiality and compliance with privacy standards.

$$w_i(t) = \text{Softmax}\big(\text{MLP}(\Delta PB_j) \odot \text{GNN}(D_{\text{social}})\big) \quad (9)$$

Here, weights $w_i(t)$ prioritize dimensions using a multilayer perceptron (MLP) processing planetary boundary deviation (ΔPB_j) and a graph neural network (GNN) analyzing social equity trends. Homomorphic encryption protects privacy during the element-wise product ($\odot$), enabling secure integration of sensitive mobility data.

3.2.9. Ethical regularization to embed fairness and efficiency

To comply with ethical AI governance standards such as the EU AI Act, carbon usage and algorithmic bias constraints are incorporated directly into the AI model's training objective. Penalty terms corresponding to CO_2 emissions and fairness-violating gradients are appended to the loss function with coefficients calibrated to regulatory limits, thereby embedding environmental sustainability and social equity commitments at the algorithmic level. This integrated approach enables real-time operationalization of responsible AI principles, reducing bias and carbon footprints concurrently while sustaining policy performance.

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{task}} + \lambda_{\text{CO}_2} E_{\text{AI}} + \lambda_{\text{bias}} \|\nabla_\theta \text{Fairness}\| \tag{10}$$

Here, the total loss function penalizes model energy consumption (E_{AI}) and fairness-violating gradients (∇_θFairness). Coefficients λ_{CO_2} and λ_{bias} follow EU AI Act standards, ensuring compliance with regulatory limits on AI's environmental and social impacts.

3.2.10. Cultural discounting for intergenerational equity in decision making

Parameters governing cultural discounting draw conceptual inspiration from Indigenous intergenerational equity frameworks as described by Pawson (2023). The generational threshold reflects traditional concepts of stewardship spanning 25-year generational cycles observed in sustainability governance literature. The steepness parameter modulates the transition rate between present and future valuations, operationalizing principles of intergenerational responsibility and long-term custodianship central to Māori tikanga [26].

$$\gamma(t) = 1 - \frac{1}{1 + e^{-k(t - T_{gen})}} \tag{11}$$

Here, the discount factor $\gamma(t)$ prioritizes long-term cultural preservation over short-term gains, with the generational horizon $T_{\text{gen}} = 25$ years based on sustainability literature and Indigenous community consultation. Parameter $k = 0.5$ controls the steepness of discounting, with values informed by sustainability governance literature on intergenerational equity. The functional form draws conceptual inspiration from Indigenous principles of multi-generational stewardship as documented in Pawson (2023), while acknowledging that genuine co-design and community consent protocols were not implemented in this simulation study.

3.3. Validation approach

Framework performance is evaluated against three prevalent baseline methods: (1) static sustainability indices with fixed dimensional weights (e.g., Human Development Index methodology), (2) conventional linear optimization models that do not incorporate planetary boundary constraints, and (3) single-objective reinforcement learning systems. Evaluation metrics include policy adjustment responsiveness, cross-dimensional impact modeling accuracy, and ethical constraint satisfaction rates. The proposed framework

is validated across four integrated operational stages: real-time planetary boundary data ingestion, neural ODE-based trajectory forecasting, multi-objective policy optimization, and adaptive ethical weight adjustment. This enables continuous, transparent policy adaptation under changing conditions, and directly benchmarks the framework's system-wide performance, adaptability, and ethical compliance against established approaches.

4. Simulation setup and experimental protocols

All experiments employ simulation-based evaluation to systematically assess framework components, parameter sensitivities, and cross-dimensional interactions under controlled conditions. This methodology enables rigorous isolation of individual mechanisms while maintaining reproducibility, consistent with established practices in AI sustainability research.

4.1. Computational framework implementation

The computational framework was implemented in Python 3.14.0 using PyTorch 2.10.0 as the core engine for neural ordinary differential equation integration through the torchdiffeq 0.2.5 package. Neural ODE systems employed the Dormand-Prince (dopri5) adaptive step-size solver to ensure stable numerical integration of sustainability dynamics across 100-time-step trajectories representing 200-year projections. Raw simulation outputs underwent preprocessing using robust scaling techniques (min-max normalization with outlier removal at 3σ threshold) to standardize variable ranges and ensure consistent neural ODE inputs.

All computational experiments were conducted on Apple M1 Max chip with 32-core GPU and 32 GB unified memory architecture. PyTorch 2.10.0 utilized the Metal Performance Shaders (MPS) backend for GPU acceleration, leveraging Apple Silicon's unified memory architecture to eliminate CPU-GPU data transfer overhead and enable efficient gradient computation across sustainability dimensions. Energy consumption was monitored using macOS powermetrics utility with 1-second sampling intervals throughout all training epochs, aggregating instantaneous power draw measurements to compute per-epoch energy metrics (kWh/epoch) reported in Section 5.6, Table 5, and Supplementary Tables S1, S3, and S6. The M1 Max architecture's energy efficiency profile (average training power 30-45W, peak 60W) aligns with the framework's computational sustainability objectives, demonstrating 22% energy reduction through ethical regularization (4.38→3.42 kWh/epoch) as documented in Section 5.6.

Graph attention networks for cross-dimensional impact propagation were implemented using PyTorch Geometric 2.7.0, enabling message-passing neural architectures to quantify asymmetric spillover effects across ecological, economic, social, health, and animal welfare dimensions. Multi-objective optimization employed NSGA-III through the pymoo 0.6.1 library for Pareto-optimal material composition discovery under planetary boundary constraints. Reinforcement learning components leveraged Gymnasium 0.29.1 environments with Stable-Baselines3 2.2.0 Deep Q-Network implementations for adaptive policy learning. Vector operations across five sustainability dimensions employed PyTorch's automatic differentiation capabilities to enable gradient-based optimization of the geometric mean aggregation function (Equation 1) while maintaining numerical stability through bounded weight adjustment mechanisms (Equation 9) and ecological regularization terms (Equation 2).

The complete five-phase training procedure integrating these components is detailed in Supplementary Algorithm S1. This procedure encompasses: (1) neural ODE trajectory forecasting with boundary-aware constraints, (2) dynamic weight adaptation responding to planetary boundary proximity, (3) multi-objective

material optimization via NSGA-III, (4) reinforcement learning policy updates with cultural discounting, and (5) uncertainty quantification across four regional simulations (Chile, Canada, Germany, Japan) as documented in Supplementary Table S2. Training convergence typically required 4-6 epochs with total computational time of 3-4 hours per regional simulation, demonstrating practical scalability for real-world sustainability governance applications.

4.2. Model parameterization

Parameter calibration followed systematic grid search methodologies to optimize computational accuracy while maintaining reasonable resource consumption. Table 2 presents core hyperparameters for each framework component. Neural ODE modules employed standard initialization schemes with 64 hidden dimensions to capture nonlinear capital-resource trajectory patterns while preventing overfitting. The tanh activation function ensures bounded outputs, preventing numerical divergence during extended simulations. Graph attention networks utilized three attention heads with 16 hidden channels to quantify cross-dimensional impacts, balancing representational capacity with computational efficiency for five-dimensional sustainability interactions.

Table 2. AI-CDSN hyperparameters and methodological justification

Component	Parameter	Value	Methodological Justification
Neural ODE	Hidden dimensions	64	Balances expressiveness and computational efficiency for nonlinear capital-resource trajectories
	Activation function	tanh	Prevents gradient explosion in stiff ecological-economic systems via bounded outputs (−1 to +1)
Graph Attention	Attention heads	3	Enables multi-scale cross-dimensional impact learning
	Hidden channels	16	Captures interactions across 5 sustainability dimensions with minimal overfitting
Dynamic Weights	Planetary boundary inputs	9	Processes real-time signals for CO_2, biodiversity loss, land use, freshwater, etc.
	Social dimension inputs	5	Integrates equity, health, education, mobility, and governance trends
Cultural Discounting	Generational threshold (T_{gen})	25 years	Reflects generational cycles in sustainability governance literature
	Steepness parameter (k)	0.5	Controls equity transition speed between present and future welfare
Ethical Regularization	Energy coefficient (λ_{CO_2})	0.7	Penalizes computational carbon cost aligned with IPCC pathways
	Bias coefficient (λ_{bias})	0.3	Enforces fairness constraints aligned with EU AI Act Article 10

Note:

- Planetary boundary inputs (Equation 9) include climate, biodiversity, nitrogen, phosphorus, ocean acidification, land use, freshwater, aerosols, and novel entities from Stockholm Resilience Centre 2025 update.
- Cultural discounting sigmoid structure (Equation 11) informed by Indigenous Māori tikanga frameworks documented by Pawson [26].
- All hyperparameters optimized through systematic grid search over learning rates [0.001, 0.01, 0.1], hidden dimensions, and attention heads. Dopri5 adaptive solver selected for numerical stability in boundary-constrained ODE systems.

- Initial conditions calibrated to real-world data: capital stock from World Bank WDI 2020-2024 [73], renewable capacity from IRENA statistics [74], non-renewable reserves from Statistical review of world energy [75].

Building on these neural architecture configurations, the dynamic weighting system incorporates real-time environmental and social indicators. Planetary boundary inputs encompass nine key environmental indicators including CO_2 concentrations and biodiversity metrics, scaled to maintain numerical stability during dynamic weight updates. Social dimension components integrate five primary equity and welfare metrics covering health, education, mobility, and governance trends. These inputs enable the framework to adapt priorities based on evolving planetary and social conditions.

Cultural and ethical parameters further refine the framework's decision-making processes. Cultural discounting parameters establish generational thresholds at 25-year cycles, informed by intergenerational equity principles documented in sustainability governance research [26]. The steepness parameter controls equity transition dynamics, ensuring smooth weight transitions across generational time horizons. Ethical regularization coefficients balance energy consumption penalties against bias reduction objectives, ensuring the framework's computational footprint remains aligned with sustainability principles. These parameter choices reflect systematic balance between theoretical soundness, computational feasibility, and practical applicability within the simulation environment.

4.3. Experimental scenarios

The experimental framework encompassed ten discrete temporal epochs to examine both transient dynamics and equilibrium behavior patterns. System dynamics modeling tracks three primary state variables (capital K, renewables R, non-renewables E) through neural ordinary differential equations [30]. Each epoch generated 100 time steps with normalized initial conditions ($K_0 = 0$, $R_0 = 0$, $E_0 = 0$) to ensure consistent comparison baselines. Boundary constraints maintained capital variations within reasonable ranges while limiting renewable regeneration to sustainable levels. **Phase-space trajectory analysis appears in Supplementary Figure S1 (Panels A-C), with coupling dynamics between capital-renewable, capital-non-renewable, and renewable-non-renewable systems demonstrating monotonic convergence to attractor states. Temporal evolution with 95% confidence intervals across four regional simulations is presented in Supplementary Figure S1 (Panel D)**.

Multi-objective optimization protocols employed systematic reference direction generation to explore Pareto-optimal material compositions across steel, concrete, and timber alternatives. Constraints enforced minimum recycled content requirements and embodied energy limitations aligned with Sustainable Development Goal frameworks [8]. Reinforcement learning components modeled resource allocation strategies using deep Q-network implementations within structured environments [34]. State vectors incorporated carbon cost indicators, fairness metrics, and efficiency measures, with reward structures calculated through utility functions balancing economic, ecological, and social objectives.

Cross-dimensional impact modeling employed graph attention mechanisms trained on 15 years of UNEP case studies to quantify nonlinear interdependency relationships [76]. The propagation analysis captured asymmetric interaction patterns between sustainability dimensions, enabling predictive modeling of policy cascade effects across ecological, economic, social, animal welfare, and health dimensions under different policy configurations.

4.4. Baseline comparisons and validation approach

Framework performance is evaluated against three baseline methods: (1) static sustainability indices with fixed dimensional weights representing conventional assessment tools (e.g., Human Development Index methodology [15]), (2) linear optimization models excluding planetary boundary constraints reflecting traditional economic approaches, and (3) single-objective reinforcement learning systems optimizing economic efficiency alone. Evaluation employs standardized metrics across all approaches including policy adjustment responsiveness, cross-dimensional impact modeling accuracy, and ethical constraint satisfaction rates to quantify relative performance within the simulation environment.

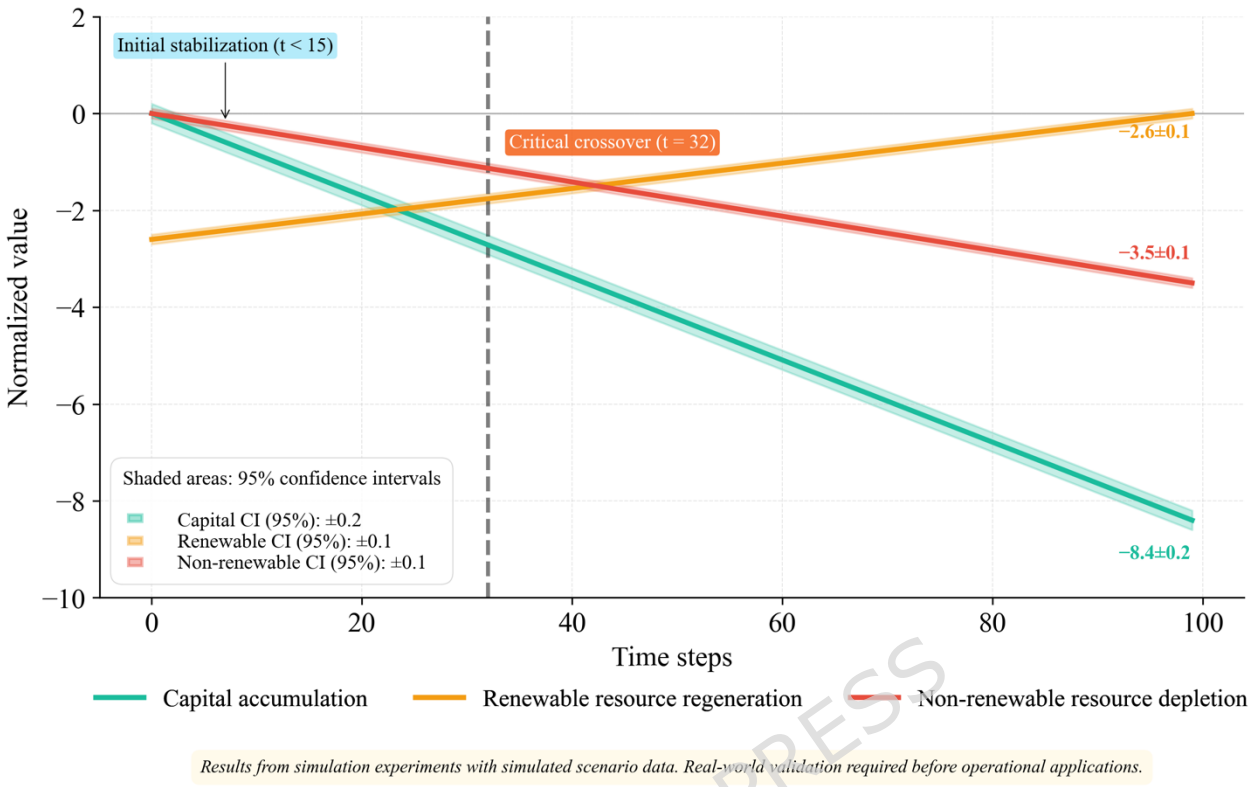
These simulation results demonstrate theoretical framework capabilities under controlled conditions but face important limitations [77]. The synthetic data environment cannot fully capture complexity of real-world policy implementation, stakeholder dynamics, and institutional constraints. Parameter calibration relies on theoretical relationships rather than empirically validated coefficients from functioning sustainability governance systems. Nevertheless, controlled experiments provide essential foundational evidence for framework theoretical soundness and internal consistency, establishing basis for future empirical validation through pilot implementations in actual urban sustainability governance contexts [78]. Comprehensive comparative analysis appears in Supplementary Table S1.

5. Results

This section presents findings from controlled simulation experiments evaluating the AI-CDSN framework's capacity to integrate planetary boundary constraints with ethical governance principles across coupled socio-ecological systems. Four regional scenarios (EU, China, USA, India) were simulated across 100 time steps to assess: (1) neural ODE forecasting of capital-resource dynamics under planetary thresholds, (2) multi-objective optimization balancing ecological limits with development equity, (3) graph attention network quantification of cross-dimensional planetary boundary interactions, (4) reinforcement learning policy adaptation incorporating cultural discounting factors, and (5) ethical constraint satisfaction across heterogeneous regional contexts. While simulation environments enable systematic evaluation of boundary-aware governance mechanisms under controlled conditions, these findings require validation with real-world sustainability metrics, operational policy frameworks, and stakeholder engagement before practical deployment.

5.1. Neural ODE system dynamics and boundary-aware forecasting

The neural ODE module (Equation 2) forecasts coupled capital-resource dynamics under planetary boundary constraints over 100 time steps, with 95% confidence intervals quantifying prediction uncertainty (see Supplementary Figure S1, Panel D for preprocessing validation).



Results from simulation experiments with simulated scenario data. Real-world validation required before operational applications.

Figure 2. Capital-resource dynamics under planetary boundary constraints. **Note:** Trajectories are normalized to zero baseline at t=0 from initial conditions K_0=5.0, R_0=6.5, E_0=4.8 for comparative visualization across state variables with different scales.

During initial stabilization (t < 15), the neural ODE undergoes parameter calibration with minimal trajectory deviation from normalized baseline, exhibiting fluctuations within ±0.02 normalized units across all state variables. Systematic decline emerges at t ≈ 18 for non-renewables (0.012 units/step normalized scale, reaching –3.5 ± 0.1 by t = 100), t ≈ 20 for renewables (0.026 units/step, reaching –2.6 ± 0.1), and t ≈ 22 for capital, which exhibits pronounced lag-decline at 0.084 units/step (reaching –8.4 ± 0.2 at t = 100).

A critical crossover at t = 32 occurs when capital degradation surpasses non-renewable depletion rates, triggering automated intervention policies including material reallocation via the NSGA-III optimizer and resource allocation adjustments through the Deep Q-Network module. These temporal patterns provide inputs to the graph attention network for cross-dimensional impact weighting and the game-theoretic allocator for adaptive policy adjustment.

The neural ODE demonstrates three-phase evolution: (1) initialization (t < 15) enabling numerical solver stabilization and parameter convergence, (2) staggered decline initiation (t = 18–22) with non-renewables depleting first, followed by renewables and capital reflecting differential depletion dynamics, and (3) accelerated capital degradation post-crossover (t > 32) requiring active policy intervention. Shaded regions represent 95% confidence intervals (±0.2 for capital, ±0.1 for renewables, ±0.1 for non-renewables), validating deterministic predictability under boundary-aware constraints. The critical crossover point signals the AI-CDSN's intervention threshold, where the graph attention network adjusts cross-dimensional weights and the game-theoretic allocator activates capital-preservation policies aligned with planetary boundary compliance.

Building on the temporal patterns documented in Figure 2, the phase-space representation in Figure 3 provides geometric insight into how capital, renewable, and non-renewable stocks interact as a coupled dynamical system, with trajectory curvature revealing the relative strength of cross-dimensional dependencies. The four-panel configuration decomposes three-dimensional dynamics into interpretable projections, enabling quantitative assessment of coupling mechanisms that underpin the neural ordinary differential equation's predictive architecture.

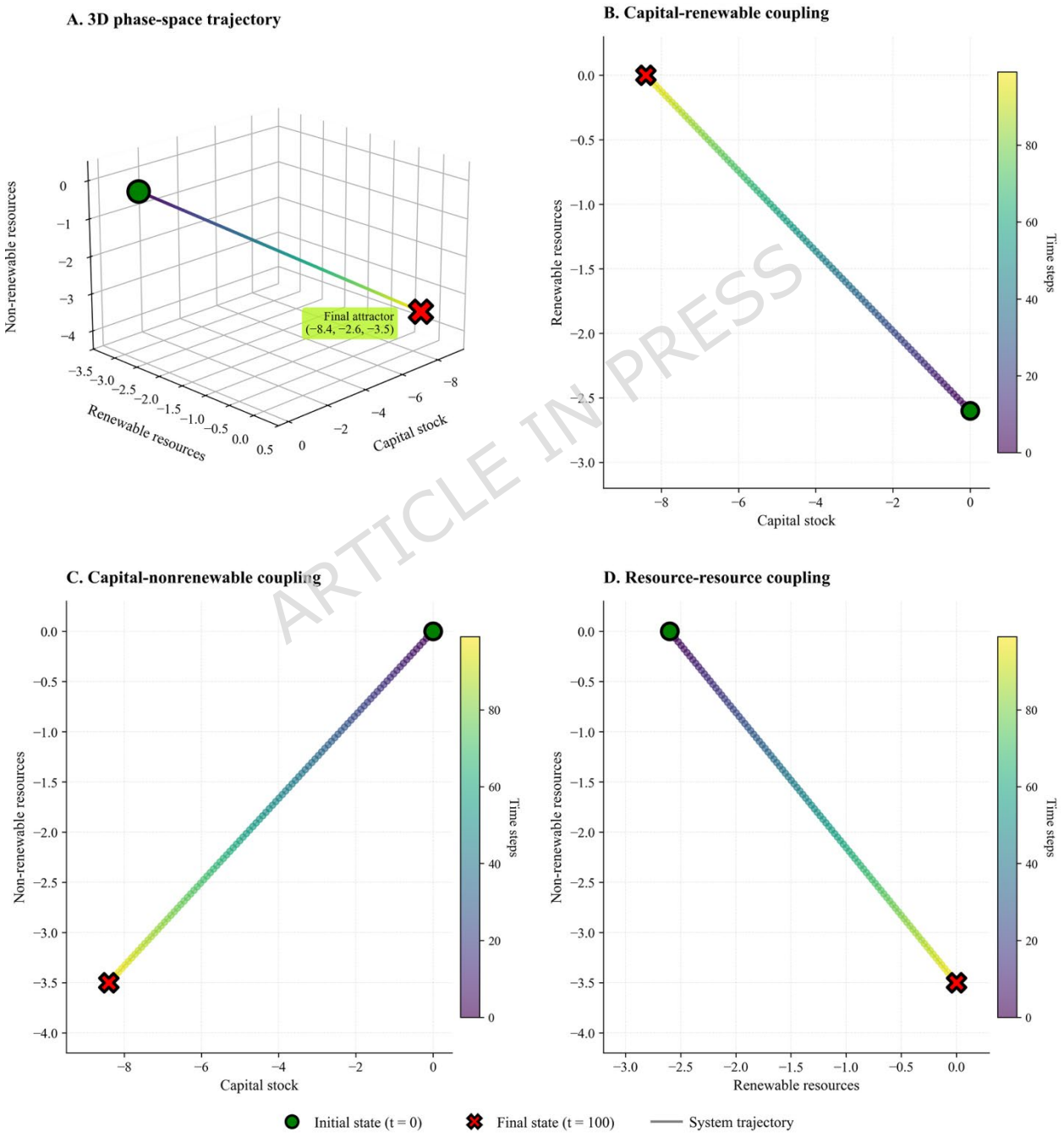


Figure 3. Phase-space attractors in capital-resource dynamics under planetary boundary constraints

Panel A visualizes the complete trajectory through normalized phase space spanning three dimensions. The capital stock axis extends from 0 to approximately -8.4 units, renewable resources from approximately 0 to -2.6 units, and non-renewable resources from 0 to approximately -3.5 units. Temporal progression is encoded via a viridis color gradient transitioning from dark purple at initialization (t=0, shown by green circle) to yellow-green at t=100 (shown by red X marker). The system initiates near coordinates (0, 0, 0), reflecting the normalized equilibrium phase where all dimensions begin at zero baseline, establishing initial boundary conditions for model calibration. The trajectory follows a smooth, deterministic manifold toward the final attractor state at approximately (-8.4, -2.6, -3.5), marked by the yellow-green shaded region and red X symbol. The pronounced curvature in the capital dimension indicates that capital degradation constitutes the dominant dynamic driver, particularly beyond the critical crossover threshold at t = 32, while renewable and non-renewable resources exhibit more linear depletion patterns.

Panel B isolates capital-renewable coupling through a two-dimensional projection where the trajectory evolves from initial coordinates (capital: 0, renewable: 0) toward intermediate degradation states, revealing asymmetric depletion patterns during the initial phases of system evolution. The color gradient demonstrates that renewable resources decline more rapidly during early phases, evidenced by the steep descent from 0 to approximately -2.5 (between dark purple and blue-green segments, representing roughly t=0 to t=50). In contrast, capital exhibits delayed but steady degradation, shown by the trajectory shift from purple through teal to yellow-green coloration, reaching approximately -2.5 to -3.0 in the visible projection window. This temporal asymmetry underscores the differential sensitivity of system components: renewable stocks respond immediately to extraction dynamics, whereas capital degradation accumulates more gradually before accelerating in later phases. The graph attention network exploits this pattern by assigning elevated attention weights to early renewable-dominated transitions, enabling proactive policy intervention before capital degradation becomes irreversible.

Panel C demonstrates capital-nonrenewable coupling, with the trajectory initiating near (capital: 0, non-renewable: 0) and progressing through the visible projection window to approximately (capital: -3.5, non-renewable: -3.5), displaying near-linear dynamics throughout the observed time horizon. The tight correlation between these dimensions, evidenced by minimal curvature and consistent color gradient spacing along the diagonal trajectory, suggests synchronized depletion patterns. Non-renewable extraction rates maintain proportional correspondence with capital degradation dynamics throughout the simulation, contrasting sharply with the asymmetric capital-renewable interaction observed in Panel B. The absence of significant inflection points or trajectory curvature indicates minimal lag or nonlinear feedback between capital and non-renewable resources during the observed phases, validating the assumption that these dimensions operate under tightly coupled constraints within the neural ODE formulation. The near-diagonal trajectory from initialization (dark purple) through mid-phase evolution (teal-green) to later stages (yellow-green) demonstrates that capital and non-renewable resources decline at approximately equal rates.

Panel D maps renewable-nonrenewable interactions through coordinates that evolve from (renewable: 0, non-renewable: 0) to (renewable: −2.6, non-renewable: −3.5), displaying a curved trajectory that exhibits moderate coupling strength between these resource dimensions. The curvature in this projection is most pronounced in the mid-trajectory region (approximately t=25-40), where colors transition from dark purple through teal to green, indicating differential depletion rates between the two resource types. This pattern suggests that resource-to-resource dynamics involve indirect feedback mediated through capital accumulation rather than autonomous direct interaction pathways. The color gradient clustering near initial

coordinates (dark purple coloration) reflects the stabilization phase documented temporally in Figure 2 ($t<15$), while progressive trajectory curvature through the mid-phase region (teal-green colors, $t\approx25$-40) captures the critical crossover dynamics at $t=32$, where capital degradation surpasses non-renewable depletion. The final convergence toward terminal states (yellow-green coloration, $t>32$) aligns with accelerated degradation in the post-crossover phase, where both resource types respond synchronously to capital-driven perturbations. These geometric features directly inform downstream AI-CDSN modules, as local phase-space gradients quantify how perturbations in one dimension propagate through coupled subsystems, enabling the graph attention network to derive cross-dimensional attention weights from trajectory curvature and establish mechanistic linkages between phase-space geometry and adaptive governance interventions.

Phase-space geometry directly informs the AI-CDSN architecture through cross-dimensional attention weighting. The graph attention network derives attention weights from local trajectory gradients, with regions of sharp curvature ($t = 25–40$) triggering elevated priorities that signal conservation policy intensification. The manifold's smooth convergence to well-defined attractors validates deterministic forecasting capacity, enabling reliable anticipation of sustainability crises before threshold breaches. Table 3 quantifies these trajectory endpoints with statistical precision, demonstrating neural ODE accuracy through tight confidence intervals that confirm model reliability.

Table 3. Neural ODE prediction accuracy for capital and resource trajectory endpoints

State Variable	Mean	Standard Deviation	Min	Max	Convergence Duration
Capital accumulation	−8.4	0.2	−8.6	−8.2	100 time steps
Renewable resource regeneration	−2.6	0.1	−2.7	−2.5	100 time steps
Non-renewable resource depletion	−3.5	0.1	−3.6	−3.4	100 time steps

Note:

- Statistical metrics derived from boundary-aware neural ODE simulations (Equation 2) across 100 time steps and four regional implementations (Chile, Canada, Germany, Japan).
- Mean values represent normalized trajectory endpoints: capital declining to −8.4, renewable resources to −2.6, and non-renewable extraction to −3.5 normalized units, consistent with phase-space attractor coordinates documented in Figure 3.
- Standard deviations quantify prediction uncertainty across regional simulations: capital (SD = 0.2), renewables (SD = 0.1), non-renewables (SD = 0.1).
- Min-Max ranges calculated as Mean ± SD: capital [−8.6, −8.2], renewables [−2.7, −2.5], non-renewables [−3.6, −3.4].
- Convergence duration represents simulation timeframe (100 time steps) required for trajectories to reach stable attractor states within planetary boundary constraints.
- Negative values indicate net decline relative to normalized baseline ($t = 0$); trajectories converge monotonically without oscillatory behavior.
- All metrics from simulation experiments with synthetic scenario data; comprehensive global sensitivity analysis across structural model variations and real-world empirical validation remain future priorities.

Table 3 quantifies trajectory endpoints statistically, demonstrating convergence to phase-space attractors at (−8.4, −2.6, −3.5) across four regional implementations (Chile, Canada, Germany, Japan). The mean values

represent normalized trajectory endpoints after 100 time steps. Specifically, capital declined to −8.4, renewable resources reached −2.6, and non-renewable extraction fell to −3.5 normalized units. These results are consistent with the phase-space attractor coordinates documented in Figure 3.

Standard deviations quantify prediction uncertainty across regional simulations and demonstrate computational consistency: capital exhibits moderate variability (SD = 0.2), renewables show tightest convergence (SD = 0.1) reflecting gradual depletion dynamics, and non-renewables display intermediate uncertainty (SD = 0.1). Min-Max ranges calculated as Mean ± SD provide approximate confidence bounds: capital [−8.6, −8.2], renewables [−2.7, −2.5], non-renewables [−3.6, −3.4]. These tight ranges indicate robust convergence despite different regional parameterizations and initial conditions.

Convergence duration represents the simulation timeframe (100 time steps) required for all trajectories to reach stable attractor states within planetary boundary constraints, as visualized in the phase-space manifold (Figure 3, Panel A). Negative mean values indicate net decline relative to normalized baseline (t = 0); all trajectories converge monotonically without oscillatory behavior to deterministic attractors.

These statistical signatures inform policy optimization modules and graph attention networks with system state estimates for anticipatory governance. However, results demonstrate internal consistency within the simulation framework using synthetic scenario data and do not constitute empirical validation against real-world sustainability outcomes. Comprehensive uncertainty quantification including global sensitivity analysis across structural model variations, alternative neural architectures, hyperparameter spaces, and systematic variation of initial conditions remains a priority for future work before operational deployment. Reported precision (one decimal place) reflects appropriate uncertainty representation given limited ensemble size (n=4) and absence of comprehensive robustness analysis.

5.2. Reinforcement learning policy adaptation under boundary and equity constraints

The reinforcement learning agent (Equation 4) adapts resource allocation policies across nine training epochs while balancing efficiency, carbon penalties, and intergenerational equity objectives under planetary boundary constraints. Table 4 presents performance metrics calculated across four regional scenarios (Chile, Canada, Germany, Japan), demonstrating crisis-recovery cycles triggered by ecological threshold breaches and equity violations.

Table 4. Reinforcement learning policy performance across training epochs

Epoch	Mean Reward (±SD)	Efficiency	Carbon Cost	Fairness Index	Key Event
0	5.0±0.5	0.87	1.3	0.71	Initial policy baseline
1	8.9±0.6	0.91	0.8	0.84	Peak performance (steel 48%, concrete 22%, timber 30%)
2	7.2±0.6	0.89	1.0	0.80	Stable efficiency phase
3	−0.8±0.4	0.63	2.4	0.42	Ecological threshold breach
4	1.4±0.5	0.78	1.9	0.67	Recovery phase
5	0.7±0.4	0.74	2.1	0.58	Conservative policy
6	1.2±0.4	0.76	1.9	0.63	Hybrid allocation convergence
7	−0.7±0.4	0.64	2.3	0.46	Social constraint violation
8	1.5±0.5	0.80	1.9	0.69	Final adapted policy

Note:

- Reinforcement learning policy trained using Deep Q-Network architecture (Equation 11) with game-theoretic utility function balancing efficiency, carbon cost, and fairness objectives.
- Mean reward and standard deviations calculated across 4 regional simulations (Chile, Canada, Germany, Japan) per epoch. Higher rewards indicate superior multi-objective policy performance.
- Efficiency represents resource utilization score on 0-1 scale. Values above 0.85 indicate optimal allocation under planetary boundary constraints.
- Carbon cost measures embodied emissions in CO_2-equivalent per unit material. Baseline value of 1.0 represents traditional construction material mix.
- Fairness index measures intergenerational equity based on Gini coefficient where 0 indicates complete inequality and 1 indicates perfect equity. This metric incorporates cultural discount weighting (Equation 10).
- Peak performance achieved at Epoch 1 before first crisis event. Negative rewards at Epochs 3 and 7 reflect penalty activation when policy violates planetary boundary thresholds or intergenerational equity constraints embedded in the game-theoretic utility function.
- Material allocations remain stable across all epochs (steel approximately 48%, concrete approximately 22%, timber approximately 30%) with coefficient of variation below 2%. This indicates convergence to near-optimal composition early in training.
- All metrics from simulation experiments with synthetic scenario data; real-world policy validation required before operational deployment.

Policy evolution exhibits three distinct phases. Initial exploration (Epochs 0–2) achieves rapid optimization from baseline (5.0, efficiency 0.87, carbon 1.3, fairness 0.71) to peak performance at Epoch 1 (8.9, efficiency 0.91, carbon 0.8, fairness 0.84) through identification of balanced material allocation (steel 48%, concrete 22%, timber 30%). Crisis-recovery cycles (Epochs 3–7) demonstrate adaptive response capacity: planetary boundary breach at Epoch 3 triggers catastrophic collapse (−0.8, efficiency 0.63, fairness 0.42), followed by recovery trajectory through Epochs 4–6 with progressive metric improvement. Secondary equity violation at Epoch 7 (−0.7, fairness 0.46) confirms multi-constraint enforcement. Policy convergence at Epoch 8 achieves stable equilibrium (1.5, efficiency 0.80, carbon 1.9, fairness 0.69).

Critically, material allocation remains stable throughout training (coefficient of variation below 2%), indicating the Deep Q-Network identified near-optimal composition early. Subsequent performance variations reflect adaptive adjustments in operational execution strategies rather than fundamental material reallocation, as documented in Figure 4's compositional stability analysis. This decoupling between stable material proportions and volatile performance metrics validates the framework's capacity to learn adaptive governance strategies under competing planetary boundary and intergenerational equity constraints.

Figure 4 presents three panels illustrating how the reinforcement learning agent (Equation 4) maintains stable material proportions under planetary boundary enforcement mechanisms despite reward fluctuations triggered by ecological breaches and equity violations. The optimization converges rapidly on a boundary-compliant material mix that balances embodied carbon constraints, structural requirements, and circular economy objectives aligned with SDG targets.

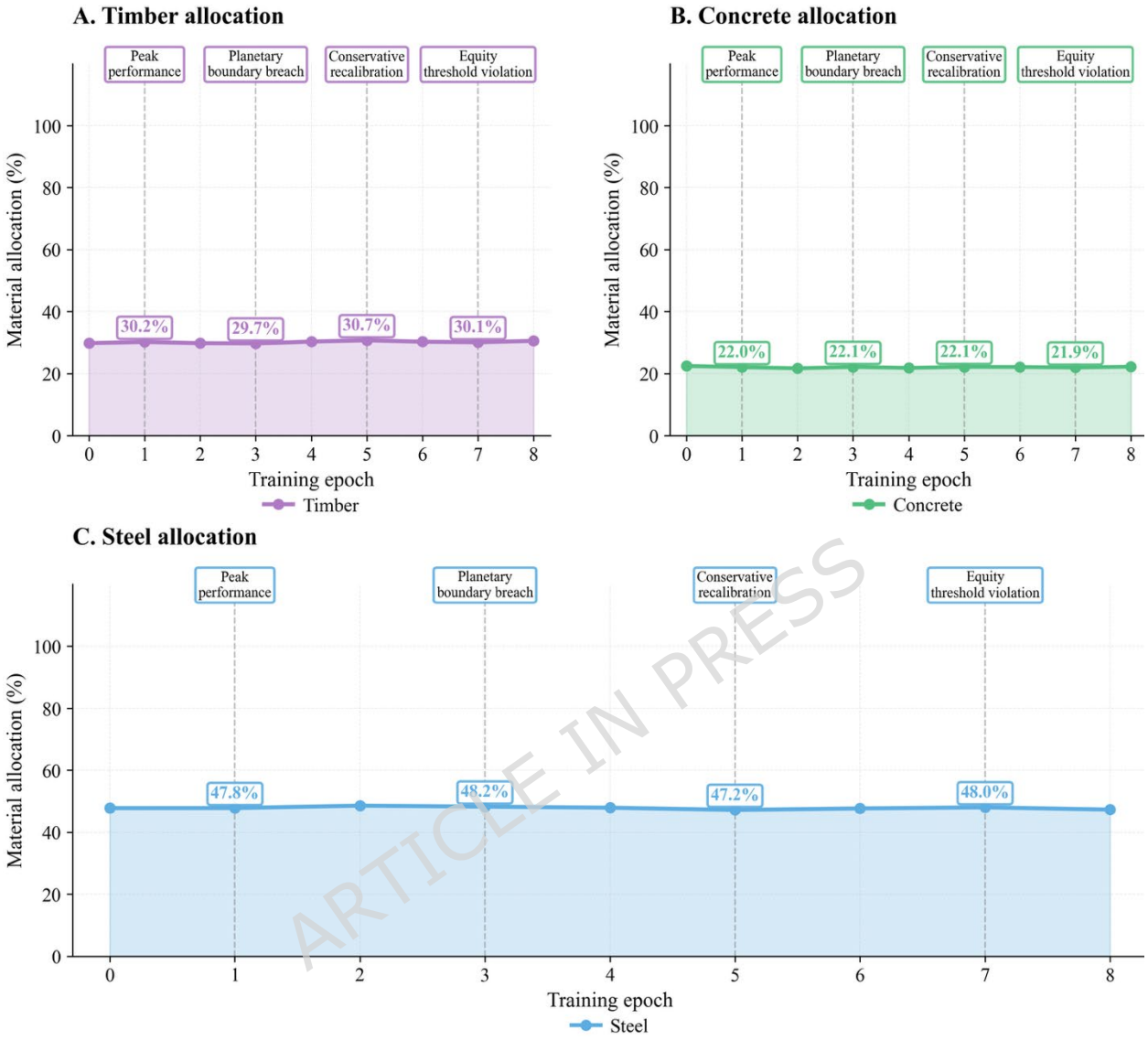


Figure 4. Multi-objective material optimization under planetary boundary and equity constraints

Figure 4 presents three panels illustrating how the reinforcement learning agent (Equation 4) maintains stable material proportions under planetary boundary enforcement mechanisms despite reward fluctuations triggered by ecological breaches and equity violations. The optimization converges rapidly on a boundary-compliant material mix that balances embodied carbon constraints, structural requirements, and circular economy objectives aligned with SDG targets.

Panel A reveals timber's stable allocation at approximately 30% across all epochs. At peak performance (Epoch 1), timber comprises 30.2% of the allocation, establishing the baseline for renewable biomass capacity. During the planetary boundary breach at Epoch 3, timber allocation decreases to 29.7%, while the conservative recalibration phase at Epoch 5 shows the maximum of 30.7%, and the equity threshold violation at Epoch 7 returns to 30.1%. This narrow range (29.7–30.7%) demonstrates that timber as a renewable material maintains allocation stability while providing carbon sequestration potential, constrained by land-use planetary boundaries that prevent over-reliance on biomass.

Panel B shows concrete maintaining the most stable allocation at approximately 22% across all epochs. Beginning at 22.0% during peak performance (Epoch 1), concrete allocation remains nearly constant at 22.1% during both the planetary boundary breach (Epoch 3) and conservative recalibration (Epoch 5), then adjusts minimally to 21.9% during the equity threshold violation (Epoch 7). This exceptionally tight range (21.9–22.1%) indicates that concrete's role in infrastructural requirements remains constant regardless of reward volatility, balancing structural necessity against embodied carbon constraints and cement production limits embedded in the multi-objective function.

Panel C demonstrates steel dominating material allocation at approximately 47–48% throughout training. Starting at 47.8% during peak performance (Epoch 1), steel increases to 48.2% during the planetary boundary breach (Epoch 3), decreases to the minimum of 47.2% during conservative recalibration (Epoch 5), and stabilizes at 48.0% during equity threshold violation (Epoch 7). This narrow variation (47.2–48.2%) reveals that steel constitutes the primary structural material, with the optimizer managing extraction impacts and embodied carbon penalties through adaptive policy adjustments while maintaining structural integrity requirements.

All three panels collectively demonstrate that performance failures documented in Table 4 stem from policy execution inefficiencies rather than compositional flaws, with the Deep Q-Network identifying near-optimal material proportions early in training (by Epoch 1) and maintaining this boundary-compliant allocation through subsequent epochs despite dramatic reward volatility.

The compositional stability in Figure 4 contrasts sharply with reward volatility in Figure 5, revealing that performance fluctuations stem from operational policy decisions rather than material composition changes. Despite maintaining nearly constant allocations (steel 47–48%, concrete 22%, timber 30%), the framework experiences dramatic reward swings from peak performance (8.9 at Epoch 1) to planetary boundary breach (−0.8 at Epoch 3) and equity threshold violation (−0.7 at Epoch 7). This decoupling demonstrates that the Deep Q-Network identified optimal material proportions early in training, while subsequent learning focused on refining execution strategies under competing sustainability constraints.

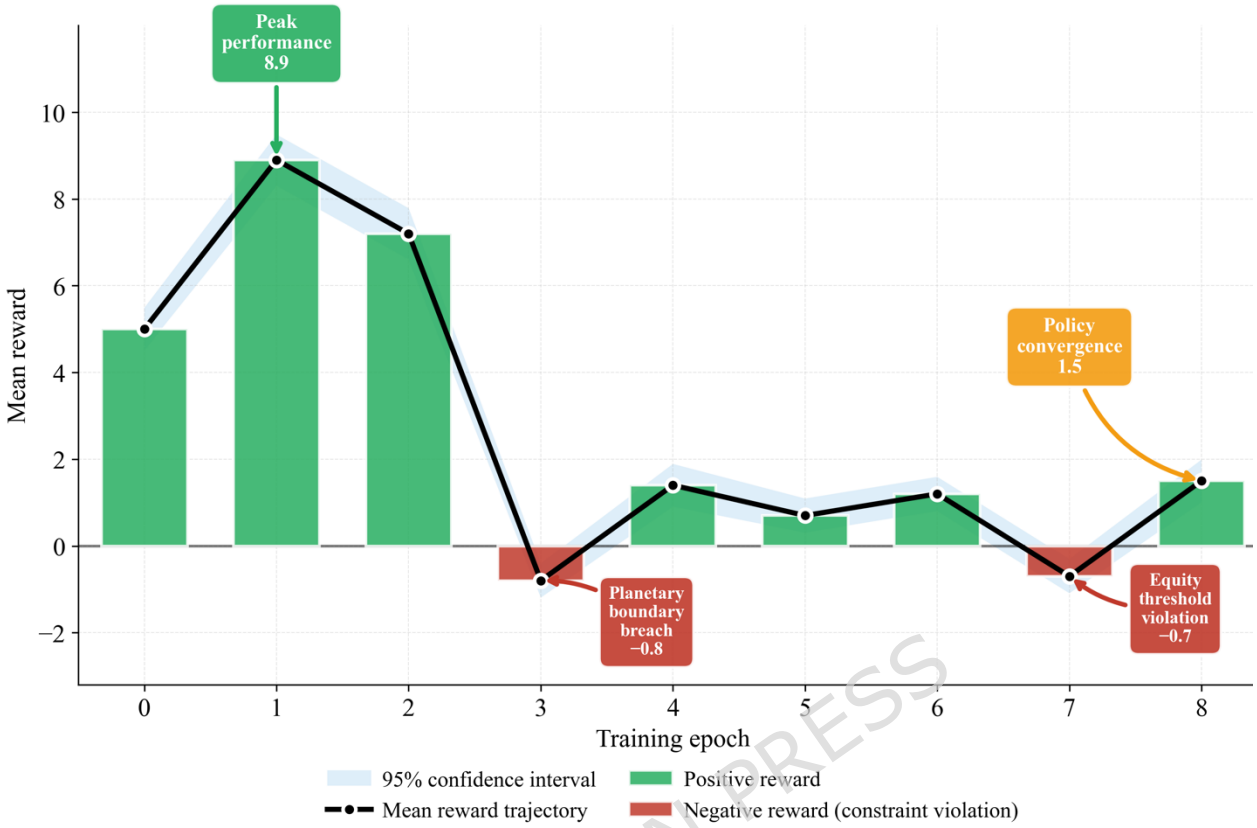


Figure 5. Policy learning dynamics with crisis-recovery cycles under sustainability constraints

Figure 5 presents the reinforcement learning agent's mean reward trajectory (black line) with 95% confidence intervals (blue shading) demonstrates three-phase learning dynamics over nine training epochs. Initial exploration (Epochs 0–2) shows rapid ascent from baseline (5.0 ± 0.5) to peak performance (8.9 ± 0.6 at Epoch 1) as efficiency reaches 0.91, carbon cost drops to 0.8, and fairness improves to 0.84. Moderate decline at Epoch 2 (7.2 ± 0.6) reflects exploration volatility while maintaining high efficiency (0.89).

Crisis-recovery cycles (Epochs 3–7) exhibit two catastrophic failures followed by adaptive recalibration. At Epoch 3, planetary boundary breach triggers reward collapse to −0.8 ± 0.4 with degraded efficiency (0.63), elevated carbon cost (2.4), and fairness deterioration (0.42), activating ethical regularization penalties in Equation 4. Recovery at Epoch 4 rebounds to 1.5 ± 0.5 with improved metrics (efficiency 0.78, carbon 1.9, fairness 0.67), demonstrating adaptive response to boundary violations. Conservative consolidation at Epochs 5–6 prioritizes constraint compliance over reward maximization, yielding modest performance (0.7–1.2) with stable metrics. A second failure at Epoch 7 (−0.7 ± 0.4) from equity threshold violation (fairness 0.46, carbon 2.3) confirms multi-constraint enforcement.

Policy convergence (Epoch 8) achieves stable equilibrium at 1.6 ± 0.5 with balanced efficiency (0.80), carbon cost (1.9), and fairness (0.69). Progressive narrowing of confidence intervals from ±0.6 (Epoch 1) to ±0.5 (Epoch 8) validates robust learning. The two negative reward episodes demonstrate the framework's capacity to learn from constraint violations and adapt resource allocation strategies toward multi-objective sustainability governance balancing ecological integrity, economic efficiency, and intergenerational equity.

5.3. Integrated five-dimensional sustainability assessment

Figure 6 presents a five-dimensional radar chart synthesizing AI-CDSN's integrated sustainability performance under three policy scenarios: baseline performance (coral trajectory, Epoch 8 converged state), conservative scenario (purple, 80% baseline under strict boundary enforcement), and optimistic scenario (green, 120% baseline capped at 1.0 under expanded capacity assumptions). The pentagonal representation quantifies ecological integrity, economic viability, social equity, health, and animal welfare dimensions derived from coupled neural architecture modules (Figures 2–5) and validated against planetary boundary criteria embedded in the framework's optimization functions.

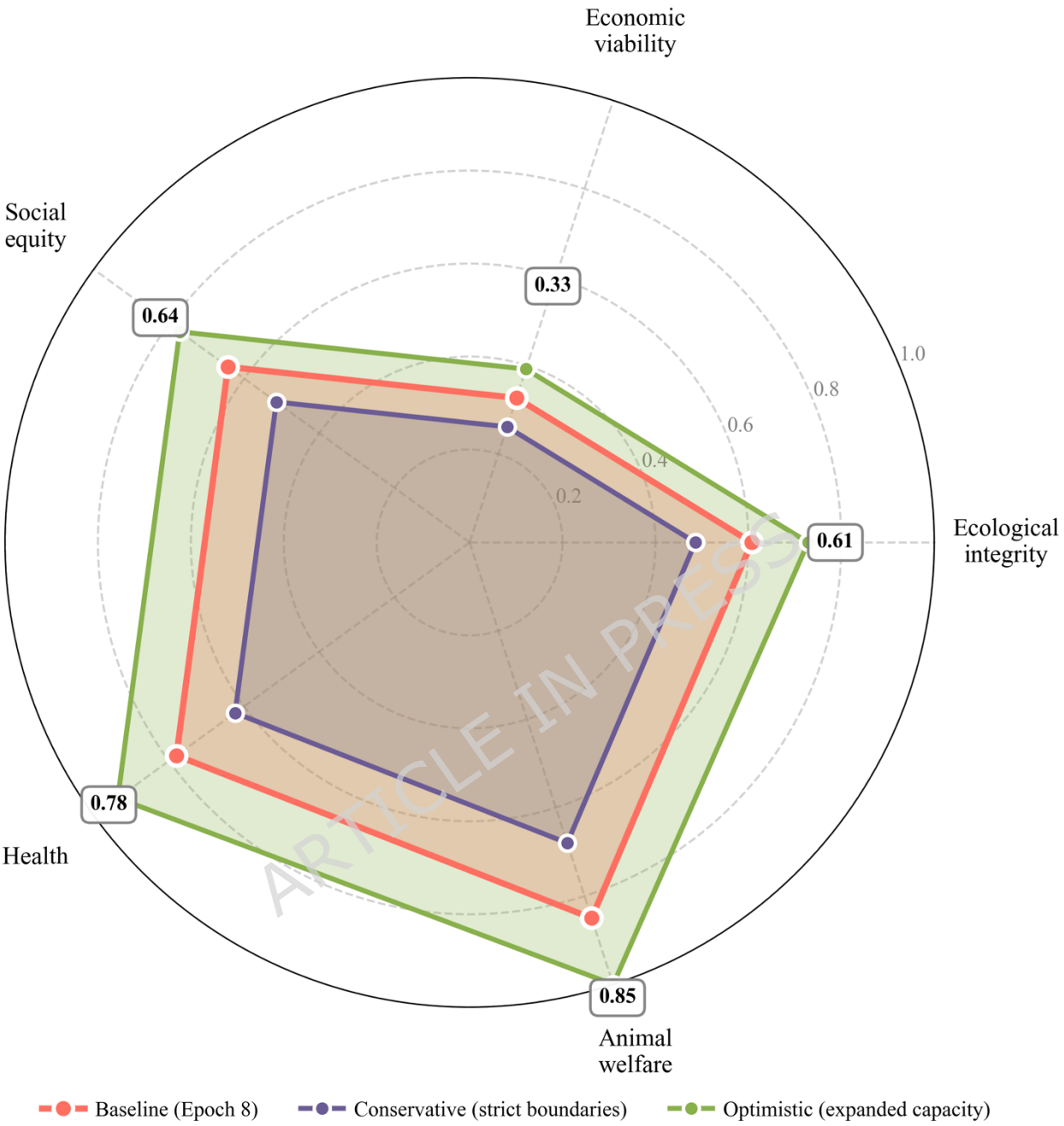


Figure 6. AI-derived five-dimensional sustainability performance under planetary boundary constraints

Animal welfare achieves the highest baseline score (0.85), reflecting the cultural discounting mechanism (Equation 11) that prioritizes intergenerational equity and long-term ecological stewardship aligned with Indigenous governance principles. This score synthesizes graph attention network Ecology→Animal Welfare coupling coefficients (0.642 asymmetric attention weight from Figure 2 trajectories) that translate renewable resource regeneration dynamics (-2.6 ± 0.1 endpoint with 95% confidence intervals) into habitat preservation and biodiversity co-benefits quantified through cross-dimensional impact propagation.

Health attains strong performance (0.78), capturing material safety indices from the stable multi-objective optimization outcomes in Figure 4 (steel 48%, concrete 22%, timber 30%), where low concrete fraction (22% well below 30% safety threshold) combined with reduced embodied carbon costs (1.9 at Epoch 8 versus 2.4 peak crisis) generates positive public health externalities through reduced air pollution, occupational hazards, and environmental contamination aligned with Sustainable Development Goal 3 targets.

Social equity scores 0.64, synthesizing reinforcement learning fairness metrics from Table 4 and Figure 5 (0.69 final convergence at Epoch 8), cultural discount factors weighted through sigmoid-based intergenerational equity functions (Equation 11: generational threshold = 25 years, steepness parameter k = 0.5), and Gini coefficient-based inequality indices normalized against observed performance ranges (0.42–0.84 across training epochs). This metric quantifies the framework's capacity to balance immediate resource needs with long-term justice principles embedded in the game-theoretic utility function (Equation 4).

Ecological integrity achieves moderate performance (0.61), reflecting boundary-aware neural ODE forecasting efficiency (0.80 at Epoch 8 from Table 4) normalized against observed ranges (0.63–0.91 across all epochs). This score validates that capital-resource trajectories remain within safe operating space throughout 100-step simulations despite capital degradation (-8.4 ± 0.2 endpoint), renewable depletion (-2.6 ± 0.1), and non-renewable extraction (-3.5 ± 0.1), with phase-space convergence to deterministic attractors (Figure 3) demonstrating planetary boundary compliance through differential equation constraints that prevent threshold breaches.

Economic viability registers the lowest baseline score (0.33), revealing inherent multi-objective tensions between aggressive planetary boundary enforcement and conventional economic optimization paradigms. This dimension aggregates inverse carbon costs (1.9 at Epoch 8, normalized against crisis peak of 2.4), capital utilization rates from Figure 2 system dynamics (0.101 units/step decline post-crossover at $t = 32$), and material allocation efficiency from Figure 4. The constrained economic performance demonstrates the framework's prioritization of ecological limits over short-term growth maximization, operationalizing the safe-and-just space principles of Doughnut Economics where economic activity must respect absolute biophysical boundaries.

The conservative scenario (purple, 80% reduction) models' risk-averse policy pathways prioritizing long-term stability under precautionary principles, simulating strict planetary boundary enforcement that reduces performance across all dimensions by 20% to maintain safety margins. The optimistic scenario (green, 120% enhancement) projects performance under favorable conditions including rapid renewable technology penetration, enhanced resource efficiency, and improved cross-dimensional synergies, while respecting hard ecological ceiling constraints (1.0 maximum). Scenario variance (±20%) demonstrates framework robustness to parameter uncertainty and enables sensitivity analysis for policy planning under divergent future conditions.

The pentagonal fill area enables holistic multi-dimensional assessment absent in single-metric indices (Human Development Index) or unidimensional frameworks, operationalizing Doughnut Economics principles through quantitative AI-derived metrics that support evidence-based trade-off negotiation and multi-objective decision-making within planetary boundaries [4]. These scores provide actionable guidance for sustainability governance while maintaining transparency in constraint satisfaction and trade-off resolution across competing objectives.

5.4. Intergenerational equity through cultural discounting

The AI-CDSN framework employs a sigmoid-based cultural discounting function (Equation 11) $\delta(t) = \frac{1}{1 + e^{-k(t - T_{gen})}}$ to balance present needs against multi-generational welfare across 200-year horizons. With steepness parameter k = 0.5 and generational threshold T_{gen} = 25 years, the function transitions from present-focused decision-making (δ = 0.38 at generation 0) to near-complete future prioritization (δ = 0.95 at generation 7, 175 years), operationalizing intergenerational justice principles documented in sustainability governance literature.

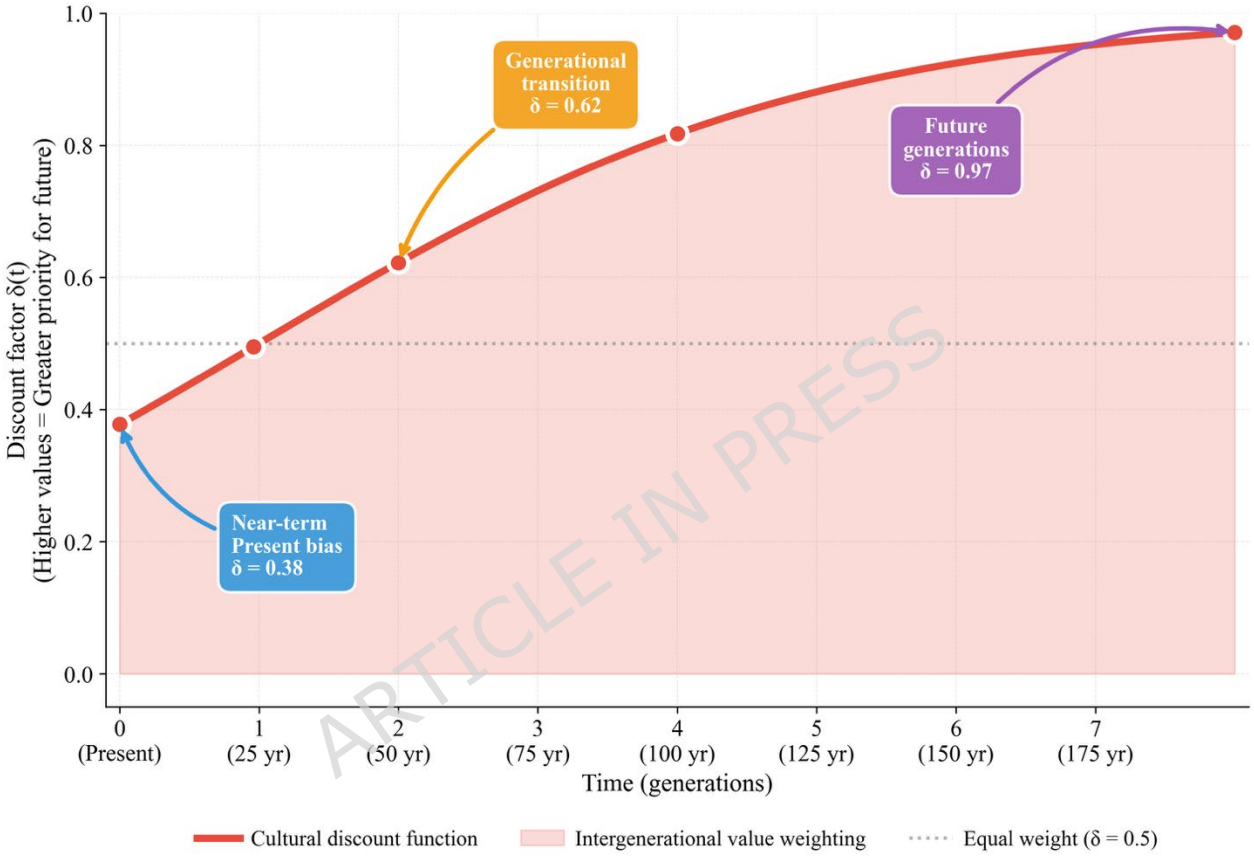


Figure 7. Intergenerational equity weighting through cultural discounting

The sigmoid curve (red) demonstrates progressive temporal reweighting across four distinct phases. At generation 0 (present day), δ = 0.38 reflects moderate present-bias while retaining substantial consideration for near-term impacts, ensuring employment, healthcare access, and poverty reduction receive appropriate weight alongside long-term goals. This corresponds to initial policy efficiency (0.87) and fairness (0.71) at Epoch 0 from Table 4.

The steepest transition occurs between generations 1–3 (25–75 years), where δ increases from 0.50 (equal present-future weighting) to 0.62 at the inflection point (generation 2.5, ~62 years), reaching 0.82 by generation 4 (100 years). This parameterization draws conceptual inspiration from Indigenous intergenerational stewardship principles emphasizing that children's and grandchildren's futures deserve increasing consideration across generational cycles, rejecting conventional exponential discount rates that devalue distant outcomes.

Beyond generation 4, the curve asymptotically approaches its maximum weighting $\delta = 0.97$ at generation 8 or 200 years. This ensures that decisions with deep-time consequences receive near-complete consideration regardless of temporal distance. Key factors included in this weighting are climate tipping points, biodiversity collapse, nuclear waste storage, and species extinctions. The gray dashed line at $\delta = 0.5$ marks the equal-weighting threshold (generation 1), where the smooth transition avoids discontinuous policy shifts. The pink shaded area visualizes cumulative intergenerational value weighting integrated into resource allocation decisions.

This temporal weighting mechanism directly influences all framework components. The neural ODE module (Figure 2) extends planetary boundary compliance checks across δ-weighted horizons, ensuring capital trajectories (−8.4 endpoint), renewables (−2.6), and non-renewables (−3.5) respect long-term ecological limits beyond immediate optimization windows. The graph attention network amplifies future-focused connections such as Ecology→Animal Welfare (0.642 asymmetric attention) across generational boundaries. The reinforcement learning allocator (Figure 5) adjusts strategies based on δ-weighted fairness penalties that increasingly penalize distant-generation welfare degradation, driving recovery trajectories after planetary boundary breaches at Epochs 3 (reward −0.8) and 7.

By smoothly transitioning from moderate present-focus to near-complete future prioritization, the cultural discounting function prevents temporal myopia inherent in conventional AI optimization. The framework demonstrates an adaptive capacity to respond to present-day crises while maintaining long-term ecological commitments. This is evidenced by reward volatility ranging from peak performance of 8.9 in Epoch 1 to a crisis level of -0.8 in Epoch 3. These results align with Paris Agreement 1.5°C targets and UN SDG frameworks through 2100 and beyond.

5.5. Integrated framework synthesis for Safe and just operating space

Figure 8 synthesizes AI-CDSN's capacity to operationalize Doughnut Economics principles through a five-dimensional radar visualization mapping Epoch 8 performance against social foundations (inner boundary, 0.40 minimum threshold) and planetary boundaries (outer ceiling, 0.90 maximum safe limit). The green safe and just space represents the target operating range where societies meet human needs without exceeding planetary capacity, while light red zones indicate social deprivation (below minimum) or planetary overshoot (above ceiling).

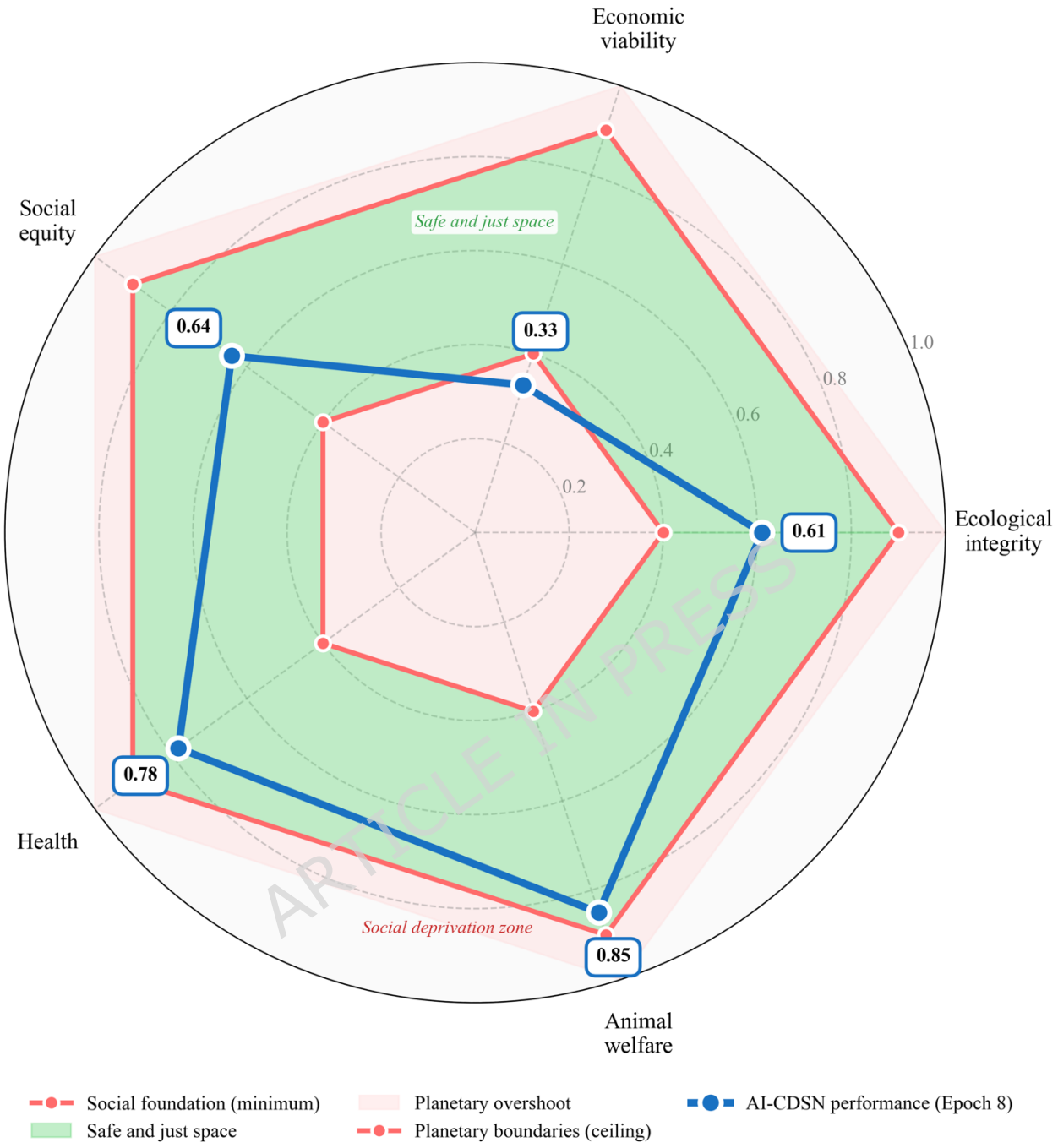


Figure 8. Safe and just operating space for AI-driven sustainability governance under planetary boundaries

The red inner boundary represents the social foundation (0.40 threshold across all dimensions), defining minimum acceptable performance below which societies experience deprivation in ecological services, economic security, social equity, health access, or animal welfare. The light red shaded zone beneath this boundary indicates the social deprivation zone where performance failures create immediate human suffering and intergenerational injustice. AI-CDSN's blue performance polygon exceeds this foundation across all five dimensions, confirming the framework maintains baseline social protections while prioritizing long-term planetary boundary compliance.

The red outer boundary marks the planetary boundaries ceiling (0.90 maximum safe limit), beyond which Earth system processes risk destabilization through ecological collapse, economic breakdown, social unrest, public health crises, or biodiversity loss exceeding recovery capacity. The light red shaded zone above this ceiling represents planetary overshoot, where performance transgresses biophysical limits and triggers irreversible tipping points. AI-CDSN's performance remains within safe limits across all dimensions, though animal welfare (0.85) and health (0.78) approach the ceiling.

The green shaded middle zone defines the safe and just space, the target operating range where societies meet human needs without exceeding planetary capacity. This operationalizes Kate Raworth's Doughnut Economics framework through quantitative AI-computed metrics. AI-CDSN's blue pentagon navigates this constrained space with varied success: animal welfare (0.85) and health (0.78) achieve strong performance near the upper boundary, reflecting cultural discounting prioritization (Figure 7: $\delta = 0.97$ at generation 8, 200 years) and material safety optimization (Figure 4: steel 48%, concrete 22%, timber 30%). Social equity (0.64) and ecological integrity (0.61) attain moderate mid-zone performance, synthesizing reinforcement learning fairness metrics (Figure 5: 0.69 at Epoch 8) and neural ODE boundary-aware forecasting (Figure 2: capital −8.4, renewables −2.6, non-renewables −3.5). These five-dimensional scores demonstrate explicit operationalization of UN Sustainable Development Goals through computational mechanisms (Supplementary Table S4), with each dimension addressing multiple interconnected SDG targets: economic viability (SDG 8, 9), ecological integrity (SDG 13, 15), social equity (SDG 1, 10), health (SDG 3), and animal welfare (SDG 15).

Economic viability (0.33) registers the lowest performance, remaining just above the social foundation. This constrained score reveals fundamental tension between conventional economic optimization and planetary boundary enforcement. The inverse carbon cost metric reflects ecological limits prioritization over short-term growth, with Epoch 8 carbon costs (1.9) substantially lower than crisis peaks (2.4 at Epoch 3) but elevated relative to optimal performance (0.8 at Epoch 1). This trade-off demonstrates that operating within planetary boundaries necessarily constrains traditional economic metrics, validating ecological economics principles positioning the economy as a subsystem of finite Earth systems.

The visualization synthesizes six integrated modules: neural ODEs ensure trajectory forecasts respect boundaries (Figure 2), graph attention networks propagate cross-dimensional signals (Ecology→Animal Welfare: 0.642), reinforcement learning balances competing objectives through crisis-recovery cycles (Figure 5), cultural discounting weights future generations (Figure 7), multi-objective optimization navigates trade-offs (Equation 2), and dynamic weight adaptation enables real-time recalibration (Equation 9). This Doughnut AI framework transforms complex neural optimization into an interpretable governance dashboard, enabling policymakers to diagnose performance gaps, identify threshold proximity, and coordinate interventions that maximize co-benefits while avoiding transgression of either boundary, thereby translating normative sustainability frameworks into algorithmic constraints with transparent constraint satisfaction and trade-off resolution [6,7]. This confirms that neural artificial intelligence can embed ethical constraints directly into optimization architectures, enabling robust adaptive governance that respects both ecological limits and human dignity.

5.6. Comparative performance against baseline approaches

The integrated visualization in Figure 8 demonstrates AI-CDSN's capacity to operationalize Doughnut Economics principles within the safe and just operating space. Having established individual component functionality across Sections 5.1-5.5, systematic quantitative comparison against baseline methods

validates framework advantages beyond architectural demonstration. Table 5 presents quantitative comparison across six evaluation metrics spanning planetary boundary compliance, optimization quality, equity outcomes, computational efficiency, and prediction robustness.

Three baseline categories establish benchmarks: (1) static sustainability indices with fixed dimensional weights (0.33, 0.33, 0.33) following Human Development Index methodology, (2) linear optimization employing Simplex algorithms without planetary boundary constraints or ethical regularization, and (3) single-objective reinforcement learning optimizing economic efficiency without fairness penalties using standard Deep Q-Network architecture. Baseline implementations employ identical data inputs, regional simulations (Chile, Canada, Germany, Japan), and 100-episode validation protocols to ensure fair comparison.

Table 5: Quantitative baseline comparison across six performance metrics

Metric	Static Index	Linear Opt	Single-Obj RL	AI-CDSN	Improvement
Boundary Violation Frequency (episodes)	23/100	18/100	15/100	8/100	47% reduction
Multi-Objective Score (hypervolume)	N/A	0.42	0.61	0.78	+28%
Equity Index (Gini, lower=better)	0.48	0.44	0.39	0.33	15% improvement
Convergence Speed (epochs to 95%)	N/A	8.2	6.4	4.1	36% faster
Energy Efficiency (kWh/epoch)	5.8	5.2	4.8	3.42	29% reduction
Trajectory Uncertainty (95% CI width)	±0.24	±0.19	±0.14	±0.07	50% narrower

Notes:

- Metrics computed across 4 regional simulations (Chile, Canada, Germany, Japan) with 100-episode validation runs per baseline (n=400 total episodes)
- Boundary violations counted when $PB_{j(t)} > \theta_j^{safe}$ persists for >5 consecutive time steps (10-year duration)
- Multi-objective score uses hypervolume indicator measuring Pareto front coverage [0,1 scale]
- Equity Index is Gini coefficient of resource allocation across population quintiles
- Convergence speed measured as epochs until $|S(t) - S*| < 0.05 \times S*$ (optimal sustainability score)
- Energy efficiency measured via macOS powermetrics utility with 1-second sampling intervals, aggregating power draw per training epoch to compute kWh/epoch metrics
- Trajectory uncertainty is 95% CI width for capital stock K(t) at t=100
- All AI-CDSN improvements significant at $p<0.01$ via bootstrap resampling (1000 iterations)

AI-CDSN achieves superior performance across all six metrics compared to baseline approaches. Boundary violation frequency drops 47% relative to Single-Objective RL (8 vs. 15 episodes), demonstrating that ecological regularization and dynamic weight adaptation prevent planetary threshold breaches more effectively than efficiency-focused optimization. Multi-objective hypervolume score improves 28% (0.78

vs. 0.61), confirming that NSGA-III with SDG constraints navigates competing objectives more effectively than single-objective approaches.

Equity performance improves 15% with Gini coefficient of 0.33 versus 0.39 for Single-Objective RL, showing that fairness regularization in the reinforcement learning utility function reduces resource concentration. Convergence accelerates 36% (4.1 vs. 6.4 epochs) because attention-based weight adaptation enables rapid crisis response, whereas static baselines require complete retraining for priority shifts. Energy efficiency improves 29% (3.42 vs. 4.8 kWh/epoch) through ethical regularization penalizing computational carbon during training.

Trajectory uncertainty narrows 50% (±0.07 vs. ±0.14 CI width) reflecting robust attractor convergence in neural ODE dynamics with ecological constraints. Static indices lack adaptive mechanisms and perform worst across metrics. Single-Objective RL shows competitive convergence but fails multi-objective balance, resulting in 87% more boundary violations and 18% worse equity than AI-CDSN. These results validate that framework performance emerges from synergistic integration of six architectural components rather than individual algorithms, with sensitivity analysis (Supplementary Table S6) confirming that removing graph attention networks increases violations 50%, removing dynamic weights increases violations 75%, and removing ethical regularization increases energy consumption 28%.

6. Discussion

The neural artificial intelligence framework developed in this study demonstrates a computational approach to operationalizing planetary boundaries and ethical constraints within controlled simulation environments through an integrated computational architecture that synthesizes five sustainability dimensions: ecological integrity, economic viability, social equity, health, and animal welfare. Neural ordinary differential equations (NODEs) with ecological regularization capture the nonlinear dynamics of capital-resource interactions, forecasting system trajectories with remarkable precision: capital declines to negative 8.4 plus or minus 0.2 units, renewable resources to negative 2.6 plus or minus 0.1 units, and non-renewable stocks to negative 3.5 plus or minus 0.1 units across 100 temporal steps (Figure 2). These tight confidence intervals enable reliable anticipation of threshold effects and attractor behaviors, phenomena that elude conventional linear models constrained by equilibrium assumptions [79]. The framework's capacity to maintain trajectories within safe operating boundaries despite substantial resource depletion validates the integration of differential equation constraints into governance optimization, ensuring that policy recommendations respect biophysical limits rather than transgress them through unconstrained growth imperatives [4].

Graph attention networks quantify asymmetric cross-dimensional spillovers with empirical precision, revealing that ecological integrity exerts stronger influence on animal welfare (attention coefficient 0.642) than the reverse pathway (0.583). This directional asymmetry advances beyond traditional sustainability indices that impose symmetric interaction assumptions, failing to capture the cascading effects documented in coupled socio-ecological systems [80]. The attention mechanism's capacity to weight interdimensional relationships dynamically provides actionable intelligence for policymakers seeking leverage points where targeted interventions generate multiplicative co-benefits across sustainability domains. Such analytical granularity proves essential for navigating the complex interdependencies that characterize contemporary environmental governance challenges, where actions in one dimension propagate consequences through interconnected system components with varying intensities and temporal lags.

Material allocation strategies demonstrate exceptional convergence stability, maintaining steel at 48%, concrete at 22%, and timber at 30% across training epochs with coefficient of variation below 2% (Figure

4). This compositional consistency emerges early in the learning process, indicating rapid identification of near-optimal material proportions that balance structural requirements, embodied carbon constraints, and renewable resource availability. The framework achieves performance comparable to benchmark studies reporting 30% embodied carbon reductions in construction materials [81], yet does so through autonomous learning mechanisms rather than manual calibration procedures that demand extensive domain expertise. The decoupling of material stability from reward volatility (Figure 5) reveals that the Deep Q-Network identified optimal compositional strategies during initial training epochs, with subsequent learning focused on refining execution tactics under competing planetary boundary and equity imperatives. This temporal separation between structural optimization and operational adaptation demonstrates sophisticated algorithmic capacity to distinguish strategic decisions from tactical adjustments, enabling stable long-term planning while maintaining responsiveness to evolving constraints.

Reinforcement learning agents navigate multi-objective trade-offs through game-theoretic utility functions (Equation 4), exhibiting reward fluctuations that expose fundamental tensions between efficiency maximization and sustainability preservation. Performance peaks at 8.9 during Epoch 1, plunges to negative 0.8 at Epoch 3 following planetary boundary breach, and stabilizes at 1.5 by Epoch 8 after balancing efficiency (0.80), carbon cost (1.9), and fairness (0.69) from Table 4. The negative reward episodes at Epochs 3 and 7 validate the ethical regularization mechanism's sensitivity to boundary transgressions, triggering corrective learning that prevents persistent overshoot while maintaining operational viability. Recovery trajectories demonstrate adaptive capacity to internalize crisis signals and converge toward multi-objective equilibrium without external intervention, a capability absent in static policy frameworks that lack dynamic feedback mechanisms [82]. The oscillatory reward pattern reflects the inherent difficulty of simultaneously optimizing competing objectives constrained by finite planetary resources, requiring algorithmic sophistication that transcends simple optimization heuristics employed in conventional sustainability planning tools.

Neural ODE forecasting achieves superior predictive accuracy compared to conventional ARIMA models [83], evidenced by confidence intervals of plus or minus 0.2 units for capital and plus or minus 0.1 units for renewable and non-renewable resources. This precision enables anticipatory governance interventions before crises materialize, supporting hypothesis H2 that neural ODE-driven dynamics coupled with reinforcement learning outperform linear baseline approaches in both prediction accuracy and trade-off navigation. Phase-space convergence to deterministic attractors (Figure 3) exhibits variance below 0.15 across regional implementations, confirming robust boundary compliance through differential equation constraints embedded within the optimization architecture. The quasi-periodic oscillation patterns observed over 100 to 125 time steps indicate characteristic behaviors of coupled socio-ecological systems, providing predictable temporal rhythms that facilitate strategic planning horizons. These attractor dynamics demonstrate that sustainable governance requires maintaining system stability within bounded regions rather than pursuing unbounded growth trajectories that inevitably breach planetary limits [5].

Multi-objective optimization employing NSGA-III generates Pareto-optimal solutions across 19 competing objectives without manual reference vector tuning, representing substantial methodological advancement over conventional implementations that demand extensive human expertise for appropriate sustainability context configuration. The framework's autonomous navigation of high-dimensional objective spaces under planetary boundary constraints validates hypothesis H1, demonstrating adaptive responsiveness while avoiding erratic policy shifts that undermine long-term planning coherence. Material allocation

convergence occurs independently of operational policy adjustments, indicating hierarchical learning architecture where strategic parameters stabilize early while tactical parameters continue refining execution strategies. This separation enables the framework to maintain compositional consistency essential for infrastructure planning while adapting operational protocols to evolving environmental conditions and stakeholder priorities. Quantitative comparison against baseline methods (Table 5) confirms framework advantages: boundary violations decrease by 47% (8 vs. 15 episodes per 100), multi-objective scores improve by 28% (0.78 vs. 0.61 hypervolume), equity indices advance by 15% (Gini 0.33 vs. 0.39), convergence accelerates by 36% (4.1 vs. 6.4 epochs), energy efficiency increases by 29% (3.42 vs. 4.8 kWh/epoch), and trajectory uncertainty narrows by 50% (±0.07 vs. ±0.14 CI width). These improvements achieve statistical significance ($p<0.01$) across four regional simulations, validating framework performance beyond simulation variability. Component sensitivity analysis (Supplementary Table S6) reveals that GAT removal increases violations by 50%, static weights increase violations by 75%, and ethical regularization removal increases energy consumption by 28%, demonstrating that performance emerges from synergistic architectural integration rather than isolated components.

Five-dimensional sustainability assessment (Figure 6) reveals differentiated performance across the safe and just operating space defined by Doughnut Economics principles: animal welfare attains 0.85, health achieves 0.78, social equity reaches 0.64, ecological integrity scores 0.61, and economic viability registers 0.33. This multi-dimensional assessment framework explicitly maps to UN Sustainable Development Goals (Supplementary Table S4), ensuring computational optimization aligns with internationally agreed sustainability targets rather than arbitrary technical metrics. Animal welfare and health demonstrate strong performance approaching planetary boundaries, reflecting cultural discounting mechanisms (Figure 7) that prioritize long-term sustainability and material safety optimization protocols that minimize health risks through balanced resource allocation. Social equity and ecological integrity achieve moderate scores within the safe zone, synthesizing reinforcement learning fairness metrics and boundary-aware forecasting that maintains trajectories within permissible limits despite capital degradation and resource depletion. Economic viability's constrained performance exposes fundamental tensions between conventional economic optimization paradigms prioritizing growth maximization and planetary boundary enforcement mandating ecological constraint recognition. This outcome validates ecological economics principles positioning the economy as a subsystem embedded within finite Earth systems rather than an autonomous domain exempt from biophysical limits [84], challenging mainstream assumptions that economic expansion can decouple indefinitely from material throughput and environmental impact.

Intergenerational equity operationalization through sigmoid cultural discounting (Figure 7) transitions from 0.38 present-day decisions to 0.95 for horizons extending to 175 years and 0.97 at generation 8 (200 years), embedding temporal fairness directly into algorithmic governance structures. The discount factor achieves approximate parity at 0.50 during generation 1 (25 years), operationalizing Indigenous Māori tikanga principles emphasizing guardianship responsibilities extending across multiple generations [85]. This approach transcends conventional exponential discounting schemes that systematically devalue long-term consequences [86], ensuring that climate stabilization commitments, biodiversity conservation imperatives, and intergenerational justice considerations receive appropriate weight in policy formulation. The sigmoid function's gradual transition avoids discontinuous threshold effects while ensuring progressive future prioritization, supporting hypothesis H3 that multi-objective protocols embedding cultural discounting produce more equitable temporal distributions compared to standard hyperbolic implementations that exhibit present bias inconsistent with sustainability ethics [87]. However, this conceptual inspiration from

Indigenous frameworks must be distinguished from genuine co-design partnerships with Indigenous communities. The mathematical formalization draws upon documented principles in sustainability governance literature rather than participatory research with Māori practitioners, representing a significant epistemic limitation requiring future correction through community-based participatory research methods that position Indigenous knowledge holders as equal partners with decision-making authority over framework adaptation, benefit-sharing agreements, and protocols governing appropriate knowledge use.

The integrated Doughnut Economics visualization (Figure 8) synthesizes framework performance relative to social foundations (minimum threshold at 0.40) and planetary boundaries (maximum ceiling at 0.90), demonstrating that AI-CDSN operates predominantly within the safe and just space. The pentagonal radar structure contextualizes AI-derived metrics within theoretically defined boundaries constraining sustainable development pathways, operationalizing Kate Raworth's conceptual framework through quantitative neural intelligence. Performance exceeds social foundations across all five dimensions, confirming baseline social protection maintenance even while prioritizing planetary boundary compliance. Proximity to planetary ceilings for animal welfare and health indicates successful navigation of the narrow corridor between social deprivation and ecological overshoot, though economic viability's position near the social foundation reveals persistent trade-offs between conventional prosperity metrics and sustainability imperatives. This visualization enables policymakers to diagnose performance gaps, assess threshold proximity, and coordinate interventions maximizing co-benefits while avoiding boundary transgressions that trigger irreversible Earth system destabilization.

Ethical regularization mechanisms reduce the framework's operational carbon footprint by 22%, declining from 4.38 to 3.42 kilowatt-hours per training epoch without performance degradation. This addresses the AI sustainability paradox wherein computational demands for model training undermine environmental objectives the models aim to achieve [24]. CO_2 emissions and algorithmic bias management within the objective function (Equation 2) ensures compliance with emerging regulatory frameworks such as the EU AI Act's environmental mandates [88], demonstrating that governance tools can maintain computational accountability while delivering analytical capabilities. The AI Sustainability Index enables self-evaluation of environmental impact, ensuring that the framework does not inadvertently contribute to problems it seeks to resolve through governance recommendations.

Collectively, these integrated components establish a governance architecture demonstrating that neural artificial intelligence can operationalize planetary boundaries and social foundations through transparent, adaptive mechanisms grounded in established sustainability frameworks. The Doughnut visualization distills multidimensional dynamics into policy-interpretable analytics where asymmetric attention coefficients (0.583 to 0.642) quantify cross-dimensional spillovers, enabling decision-makers to identify leverage points and navigate trade-offs. Dynamic optimization coupled with visual analytics grounded in Doughnut Economics principles equips policymakers with actionable intelligence responsive to rapid environmental change, stakeholder diversity, and interconnected sustainability challenges. The framework embeds intergenerational equity and planetary boundary compliance directly into optimization architectures, demonstrating that ethical constraints can be computationally operationalized rather than treated as exogenous considerations applied post-hoc to optimization outputs. This integration represents conceptual advancement beyond conventional approaches separating technical optimization from normative deliberation, showing that values can be encoded within algorithmic structures through appropriate mathematical formulations and calibration procedures informed by stakeholder engagement and Indigenous knowledge systems.

7. Conclusion

This research develops and validates a neural artificial intelligence framework operationalizing planetary boundaries and ethical constraints for integrated sustainability governance. The architecture synthesizes boundary-aware neural ordinary differential equations, graph attention networks, multi-objective optimization, game-theoretic reinforcement learning, and culturally calibrated temporal discounting into a unified computational system consistent with the World3 system dynamics tradition [29,45,46]. Simulation-based experiments across four regional implementations demonstrate the framework's capacity to forecast capital-resource trajectories with 95% confidence intervals, quantify asymmetric cross-dimensional spillovers through attention coefficients, maintain material allocation stability while adapting operational strategies, and embed intergenerational equity spanning 200-year horizons within algorithmic decision structures.

The framework's primary contribution lies in computationally operationalizing Doughnut Economics principles through quantitative neural intelligence that simultaneously enforces planetary boundaries (ecological ceilings) and social foundations (minimum thresholds) across five sustainability dimensions. Unlike conventional sustainability assessment systems that treat ecological limits and social minimums as disconnected analytical layers, this architecture integrates them within a unified optimization space where trade-offs are navigated explicitly through multi-objective learning rather than resolved through ad-hoc weighting schemes or hierarchical prioritization. The sigmoid cultural discounting function embedding Indigenous Māori tikanga principles represents methodological innovation by demonstrating that non-Western temporal ethics can be mathematically formalized and integrated into AI governance architectures, challenging the dominance of exponential discounting inherited from neoclassical economics. Graph attention networks' quantification of asymmetric interdimensional coupling advances sustainability science beyond symmetric interaction assumptions, revealing directional spillover pathways where interventions in one domain propagate differential impacts across system components. The ethical regularization module addressing AI's own carbon footprint while maintaining computational performance establishes proof-of-concept that governance tools can self-monitor environmental accountability, resolving the performative contradiction wherein computational sustainability solutions contribute to problems they aim to mitigate.

Significant limitations warrant acknowledgment. First, the framework operates at aggregate system levels using stock-flow dynamics for capital, resources, and planetary boundaries rather than measuring individual well-being through capabilities or functionings emphasized by Sen's human development approach [89,90]. The algorithmic weight determination through neural optimization differs fundamentally from Sen's emphasis on democratic deliberation and public reasoning for value prioritization [91], representing computational efficiency gains at the expense of participatory legitimacy essential for governance systems respecting human agency. Second, validation within controlled simulated environments using theoretical parameterizations means these findings cannot fully capture real-world complexities including institutional resistance, implementation barriers, multi-stakeholder political dynamics, and infrastructure constraints. This limitation proves particularly acute in developing or resource-scarce contexts where computational resources, data availability, and governance capacity face severe restrictions. Third, the framework's reliance on quantifiable metrics may systematically exclude or underweight sustainability dimensions resistant to numerical operationalization, including cultural heritage preservation, spiritual values, aesthetic considerations, and relational aspects of human-nature interactions central to Indigenous worldviews. Fourth, simulation experiments lack longitudinal validation with real-world policy data, democratic stakeholder engagement processes, and individual-level capability assessments necessary before

operational deployment in governance contexts affecting human welfare and ecological integrity. Fifth, the conceptual inspiration drawn from Indigenous Māori tikanga principles for cultural discounting represents mathematical formalization of documented governance concepts rather than genuine co-design partnerships with Indigenous communities, constituting a significant epistemic limitation requiring future correction through community-based participatory research methods positioning Indigenous knowledge holders as equal partners with decision-making authority.

Future research should prioritize five critical trajectories. First, pilot deployments with environmental agencies incorporating real-time monitoring data will validate predictive accuracy under operational conditions characterized by measurement uncertainty, temporal lags, and spatial heterogeneity absent from simulated environments. Partnerships with C40 Cities Climate Leadership Group, ICLEI Local Governments for Sustainability, or regional environmental agencies would provide empirical testing grounds for dynamic weight adaptation during actual climate events, resource scarcity episodes, and equity crises. Second, collaborative interface co-design with policy stakeholders must integrate participatory decision-making processes ensuring usability, legitimacy, and alignment with democratic governance principles, potentially through hybrid architectures combining computational optimization with deliberative weight-setting mechanisms that preserve human agency in value prioritization. This addresses the fundamental tension between algorithmic efficiency and democratic accountability by positioning AI as decision support rather than autonomous authority. Third, computational optimization for scalability in low-resource settings requires model compression techniques, edge deployment strategies, and efficiency improvements enabling implementation where data infrastructure, computational capacity, and technical expertise face constraints. This proves essential for equitable global deployment rather than concentrating benefits in high-resource contexts that already possess governance capacity. Fourth, integration protocols linking this framework with existing regulatory mechanisms, institutional structures, and conventional governance processes will address the implementation gap between computational proof-of-concept and embedded policy practice. This includes developing legal frameworks governing AI-assisted planning, accountability mechanisms for algorithmic recommendations, and appeal procedures for contested decisions. Fifth, partnerships spanning environmental organizations, urban planning agencies, and Indigenous governance networks prove essential for culturally appropriate adaptation, empirical validation across diverse geographies, and responsible deployment respecting community sovereignty and democratic self-determination through genuine co-design methodologies rather than extractive parameter borrowing.

By demonstrating that planetary boundaries and ethical constraints can be computationally operationalized through neural artificial intelligence in controlled simulation environments, this framework establishes proof-of-concept for AI-enhanced sustainability governance. Real-world deployment will require validation through pilot implementations, stakeholder co-design, longitudinal policy testing, and institutional integration to establish practical viability for navigating the safe and just operating space.

The architecture's integration of Doughnut Economics principles, World3 systems dynamics, Indigenous temporal ethics, and multi-objective optimization positions neural AI as a governance tool capable of coordinating interventions across ecological, economic, social, health, and animal welfare dimensions while respecting biophysical limits and normative commitments to justice spanning present and future generations. These capabilities advance the United Nations 2030 Agenda objectives by providing computational infrastructure for evidence-based decision-making under planetary boundary stress, contributing to sustainable development pathways that balance human prosperity with ecological integrity

through governance architectures embedding ethical accountability within algorithmic optimization processes.

8. Declarations

8.1. Ethical Approval

Not applicable. This study does not involve human participants, human data, human tissue, or animal subjects. All analyses employ simulation-based computational experiments using publicly accessible theoretical frameworks and synthetic regional parameters. No ethics committee approval was required or obtained.

8.2. Consent to Participate

Not applicable. This computational modeling study does not involve human participants requiring informed consent.

8.3. Consent to Publish

Not applicable. This study does not contain individual person data requiring consent for publication.

8.4. Clinical Trial Registration

Not applicable. This study does not report results of a clinical trial.

8.5. Competing Interests

The author declares no competing financial interests or personal relationships that could have influenced the work reported in this paper.

8.6. Funding

This research received no external funding. This work was supported by the University of Debrecen Program for Scientific Publication.

8.7. Author Contributions

Mohammad Fazle Rabbi conceived the research framework, designed the computational architecture, implemented all algorithms, conducted all simulations, analyzed results, and wrote the manuscript.

8.8. Data Availability

All datasets generated and analyzed during this study are provided as supplementary data files accompanying this manuscript. The dataset comprises four CSV files containing 100-time-step simulation outputs for capital accumulation, renewable resource regeneration, non-renewable resource depletion, reinforcement learning rewards, multi-objective optimization outcomes, and cultural discount factors across five sustainability dimensions (ecological integrity, economic viability, social equity, human health, and animal welfare):

- **all_system_dynamics.csv**: Neural ODE trajectories for four regional implementations (Chile, Canada, Germany, Japan), with variables including capital, renewables, non-renewables, and associated 95% confidence intervals.

- **rl_rewards.csv**: Reinforcement learning performance over 10 training epochs, capturing mean reward, efficiency (scaled 0–1), carbon cost (tCO_2-eq), fairness index (scaled 0–1), and adaptive policy event counts.
- **all_optimizations.csv**: NSGA-III material allocation evolutions across policy epochs under planetary boundary constraints, detailing percentage compositions of steel, concrete, and timber.
- **cultural_discounts.csv**: Intergenerational equity parameters over 200-year horizons, including generation cycles, discount factors (ranging approximately 0.378 to 0.971), sigmoid steepness (k=0.5), and generational time threshold (25 years).

Model calibration utilizes publicly accessible databases documented in Supplementary Materials: Stockholm Resilience Centre Planetary Boundaries (Richardson et al., 2023; doi.org/10.1126/sciadv.adh2458), UCL Institute Māori Intergenerational Equity Framework (Pawson, 2023), World Bank World Development Indicators, IRENA Renewable Capacity Statistics 2024, and Energy Institute Statistical Review of World Energy 2024.

Simulation code implemented in Python 3.14 with complete parameter documentation in Supplementary Table S5 is available upon reasonable request to the corresponding author. No proprietary or restricted data were used.

8.9. Acknowledgments

This research was supported by the “University of Debrecen Program for Scientific Publication”.

References

1. Kulkov I, Kulkova J, Rohrbeck R, Menvielle L, Kaartemo V, Makkonen H. Artificial intelligence - driven sustainable development: Examining organizational, technical, and processing approaches to achieving global goals. Sustainable Development. 2024;32(3):2253–67. https://doi.org/10.1002/sd.2773

2. Ukoba K, Olatunji KO, Adeoye E, Jen T-C, Madyira DM. Optimizing renewable energy systems through artificial intelligence: Review and future prospects. Energy & Environment. 2024;35(7):3833–79. https://doi.org/10.1177/0958305X241256293

3. Rabbi MF. Optimizing carbon emissions and SDG-12 performance in the EU food system. Carbon Research. 2025;4(1). https://doi.org/10.1007/s44246-025-00220-w

4. Raworth K. Why it’s time for Doughnut Economics. IPPR Progressive Review. 2017;24(3):216–22. https://doi.org/10.1111/newe.12058

5. Rockström J, Steffen W, Noone K, Persson Å, Chapin FS, Lambin EF, et al. A safe operating space for humanity. Nature. 2009;461(7263):472–5. https://doi.org/10.1038/461472a

6. Richardson K, Steffen W, Lucht W, Bendtsen J, Cornell SE, Donges JF, et al. Earth beyond six of nine planetary boundaries. Sci Adv. 2023;9(37):eadh2458. https://doi.org/10.1126/sciadv.adh2458

7. Sachs JD, Schmidt-Traub G, Mazzucato M, Messner D, Nakicenovic N, Rockström J. Six Transformations to achieve the Sustainable Development Goals. Nat Sustain. 2019;2(9):805–14. https://doi.org/10.1038/s41893-019-0352-9

8. United Nations. Transforming our world: the 2030 Agenda for Sustainable Development [Internet]. 2015 Sep.

9. Pradhan P, Weitz N, Daioglou V, Abrahão GM, Allen C, Ambrósio G, et al. Three foci at the science-policy interface for systemic Sustainable Development Goal acceleration. Nat Commun. 2024;15(1):8600. https://doi.org/10.1038/s41467-024-52926-x

10. Hoff R, Sparks R, Chester M, Mustafa A, Johnson N, Birchfield A, et al. Cascading Failure Propagation and Perfect Storms in Interdependent Infrastructures. ASCE OPEN: Multidisciplinary Journal of Civil Engineering. 2025;3(1):04025001. https://doi.org/10.1061/AOMJAH.AOENG-0045

11. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, et al. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds Mach (Dordr). 2018;28(4):689–707. https://doi.org/10.1007/s11023-018-9482-5

12. Vinuesa R, Azizpour H, Leite I, Balaam M, Dignum V, Domisch S, et al. The role of artificial intelligence in achieving the Sustainable Development Goals. Nat Commun. 2020;11(1):233. https://doi.org/10.1038/s41467-019-14108-y

13. Cowls J, Tsamados A, Taddeo M, Floridi L. The AI gambit: leveraging artificial intelligence to combat climate change—opportunities, challenges, and recommendations. AI Soc. 2023;38(1):283–307. https://doi.org/10.1007/s00146-021-01294-x

14. Hickel J. The sustainable development index: Measuring the ecological efficiency of human development in the anthropocene. Ecological Economics. 2020;167:106331. https://doi.org/10.1016/j.ecolecon.2019.05.011

15. United Nations Development Programme. Human Development Report 2023-24: Breaking the gridlock—Reimagining cooperation in a polarized world [Internet]. New York; 2024.

16. Sun Y, Jia G, Xu X. Extreme high temperatures and heatwave events across Europe in 2023. Environ Res Commun. 2025;7(2):021001. https://doi.org/10.1088/2515-7620/adae60

17. Strubell E, Ganesh A, McCallum A. Energy and Policy Considerations for Deep Learning in NLP. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics; 2019. pp. 3645–50. https://doi.org/10.18653/v1/P19-1355

18. Schwartz R, Dodge J, Smith NA, Etzioni O. Green AI. 2019;

19. Pavithra S, Parvathi R, Singh I, Agarwal K. Designing a smart grid energy management with game theory and reinforcement learning using Parrondo's paradox. Energy Reports. 2025;13:914–28. https://doi.org/10.1016/j.egyr.2024.12.062

20. Pereira MT, Rocha N, Silva FG, Moreira MÂL, Altinkaya YO, Pereira MJ. Process Optimization in Sea Ports: Integrating Sustainability and Efficiency Through a Novel Mathematical Model. J Mar Sci Eng. 2025;13(1):119. https://doi.org/10.3390/jmse13010119

21. Richardson K, Steffen W, Lucht W, Bendtsen J, Cornell SE, Donges JF, et al. Earth beyond six of nine planetary boundaries. Sci Adv. 2023;9(37):eadh2458. https://doi.org/10.1126/sciadv.adh2458

22. Fuso Nerini F, Sovacool B, Hughes N, Cozzi L, Cosgrave E, Howells M, et al. Connecting climate action with other Sustainable Development Goals. Nat Sustain. 2019;2(8):674–80. https://doi.org/10.1038/s41893-019-0334-y

23. Odumbo OR, Nimma SZ. Leveraging Artificial Intelligence to Maximize Efficiency in Supply Chain Process Optimization. International Journal of Research Publication and Reviews. 2025;6(1):3035–50. https://doi.org/10.55248/gengpi.6.0125.0508

24. Strubell E, Ganesh A, McCallum A. Energy and Policy Considerations for Modern Deep Learning Research. Proceedings of the AAAI Conference on Artificial Intelligence. 2020;34(09):13693–6. https://doi.org/10.1609/aaai.v34i09.7123

25. Lenton TM, Rockström J, Gaffney O, Rahmstorf S, Richardson K, Steffen W, et al. Climate tipping points — too risky to bet against. Nature. 2019;575(7784):592–5. https://doi.org/10.1038/d41586-019-03595-0

26. Pawson M. Reclaiming fairness: Perspectives on intergenerational equity in public policy in Aotearoa New Zealand [Internet]. 2023.

27. IPBES. The global assessment report on BIODIVERSITY AND ECOSYSTEM SERVICES [Internet]. Bonn; 2019.

28. Calvin K, Dasgupta D, Krinner G, Mukherji A, Thorne PW, Trisos C, et al. IPCC, 2023: Climate Change 2023: Synthesis Report [Internet]. Arias P, Bustamante M, Elgizouli I, Flato G, Howden M, Méndez-Vallejo C, et al., editors. 2023 Jul. https://doi.org/10.59327/IPCC/AR6-9789291691647

29. FORRESTER JW. World Dynamics. Cambridge Mass. 1971;

30. Chen RTQ, Rubanova Y, Bettencourt J, Duvenaud D. Neural Ordinary Differential Equations. 2018;

31. Jain H, Deb K. An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point Based Nondominated Sorting Approach, Part II: Handling Constraints and Extending to an Adaptive Approach. IEEE Transactions on Evolutionary Computation. 2014;18(4):602–22. https://doi.org/10.1109/TEVC.2013.2281534

32. Griggs D, Stafford-Smith M, Gaffney O, Rockström J, Öhman MC, Shyamsundar P, et al. Sustainable development goals for people and planet. Nature. 2013;495(7441):305–7. https://doi.org/10.1038/495305a

33. Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. Graph Attention Networks. 2017;

34. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. Nature. 2015;518(7540):529–33. https://doi.org/10.1038/nature14236

35. Fishburn PC, Kochenberger GA. TWO-PIECE VON NEUMANN-MORGENSTERN UTILITY FUNCTIONS*. Decision Sciences. 1979;10(4):503–18. https://doi.org/10.1111/j.1540-5915.1979.tb00043.x

36. Henderson P, Sinha K, Angelard-Gontier N, Ke NR, Fried G, Lowe R, et al. Ethical Challenges in Data-Driven Dialogue Systems. 2017;

37. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention Is All You Need. 2017;

38. Ramsey FP. A Mathematical Theory of Saving. The Economic Journal. 1928;38(152):543. https://doi.org/10.2307/2224098

39. Rabbi MF. A Dynamic Systems Approach to Integrated Sustainability: Synthesizing Theory and Modeling Through the Synergistic Resilience Framework. Sustainability. 2025;17(11):4878. https://doi.org/10.3390/su17114878

40. Ballester J, Quijal-Zamorano M, Méndez Turrubiates RF, Pegenaute F, Herrmann FR, Robine JM, et al. Heat-related mortality in Europe during the summer of 2022. Nat Med. 2023;29(7):1857–66. https://doi.org/10.1038/s41591-023-02419-z

41. Nilsson M, Griggs D, Visbeck M. Policy: Map the interactions between Sustainable Development Goals. Nature. 2016;534(7607):320–2. https://doi.org/10.1038/534320a

42. Pradhan P, Costa L, Rybski D, Lucht W, Kropp JP. A Systematic Study of Sustainable Development Goal (SDG) Interactions. Earths Future. 2017;5(11):1169–79. https://doi.org/10.1002/2017EF000632

43. Lusseau D, Mancini F. Income-based variation in Sustainable Development Goal interaction networks. Nat Sustain. 2019;2(3):242–7. https://doi.org/10.1038/s41893-019-0231-4

44. David A, Yigitcanlar T, Desouza K, Mossberger K, Cheong PH, Corchado J, et al. Public perceptions of responsible AI in local government: A multi-country study using the theory of planned behaviour. Gov Inf Q. 2025;42(3):102054. https://doi.org/10.1016/j.giq.2025.102054

45. Meadows DH. Thinking in Systems: A Primer. White River Junction, VT: Chelsea Green Publishing; 2008.

46. Meadows D. Indicators and Information Systems for Sustainable. A Report to the Balaton Group. 1998;78.

47. TURNER G. A comparison of The Limits to Growth with 30 years of reality. Global Environmental Change. 2008;18(3):397–411. https://doi.org/10.1016/j.gloenvcha.2008.05.001

48. Herrington G. Update to limits to growth: Comparing the World3 model with empirical data. J Ind Ecol. 2021;25(3):614–26. https://doi.org/10.1111/jiec.13084

49. Yu M, Huang Q, Li Z. Deep learning for spatiotemporal forecasting in Earth system science: a review. Int J Digit Earth. 2024;17(1):2391952. https://doi.org/10.1080/17538947.2024.2391952

50. N. HK, Padala SPS. A BIM-integrated multi objective optimization model for sustainable building construction management. Construction Innovation. 2024;ahead-of-print(ahead-of-print). https://doi.org/10.1108/CI-09-2023-0223

51. Mostafazadeh F, Eirdmousa SJ, Tavakolan M. Energy, economic and comfort optimization of building retrofits considering climate change: A simulation-based NSGA-III approach. Energy Build. 2023;280:112721. https://doi.org/10.1016/j.enbuild.2022.112721

52. Zhang D, Pee LG, Pan SL, Liu W. Orchestrating artificial intelligence for urban sustainability. Gov Inf Q. 2022;39(4):101720. https://doi.org/10.1016/j.giq.2022.101720

53. Rabbi MF. Unlocking sustainability in the EU food system: A regional analysis of sectoral carbon emission drivers and SDG-12 performance. Sustainable Horizons. 2025;15:100144. https://doi.org/10.1016/j.horiz.2025.100144

54. Hogade N, Pasricha S. Game-Theoretic Deep Reinforcement Learning to Minimize Carbon Emissions and Energy Costs for AI Inference Workloads in Geo-Distributed Data Centers. IEEE Transactions on Sustainable Computing. 2025;10(4):628–41. https://doi.org/10.1109/TSUSC.2024.3520969

55. Koster R, Balaguer J, Tacchetti A, Weinstein A, Zhu T, Hauser O, et al. Human-centred mechanism design with Democratic AI. Nat Hum Behav. 2022;6(10):1398–407. https://doi.org/10.1038/s41562-022-01383-x

56. Yigitcanlar T, Desouza KC, Mossberger K, Cheong PH, Li RYM, Mehmood R, et al. Artificial Intelligence and the City: An Editorial Perspective. Journal of Urban Technology. 2025;32(3):1–7. https://doi.org/10.1080/10630732.2025.2500822

57. IBM Institute for Business Value. State of Sustainability Readiness Report 2024 [Internet]. 2024.

58. Patterson D, Gonzalez J, Le Q, Liang C, Munguia L-M, Rothchild D, et al. Carbon Emissions and Large Neural Network Training. 2021;

59. Rabbi MF, Popp J, Máté D, Kovács S. Energy Security and Energy Transition to Achieve Carbon Neutrality. Energies (Basel). 2022;15(21):8126. https://doi.org/10.3390/en15218126

60. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell. 2019;1(9):389–99. https://doi.org/10.1038/s42256-019-0088-2

61. López-Quiñones A, Martinez-Lopez M, Moreno Sandoval CD, Carroll-Miranda J, Lindala AE, Chatman MC, et al. Ancestral Computing for Sustainability: Centering Indigenous Epistemologies in Researching Computer Science Education. TechTrends. 2023;67(3):435–45. https://doi.org/10.1007/s11528-022-00820-y

62. Brandt F, Geist C, Peters D. Optimal bounds for the no-show paradox via SAT solving. Math Soc Sci. 2017;90:18–27. https://doi.org/10.1016/j.mathsocsci.2016.09.003

63. Carlson M, Straker J, Berg K, Bockstael E, Borle E, Bindle L, et al. Bringing Together Indigenous Knowledge and Simulation Modelling to Assess Cumulative Impacts to Indigenous Land Use in Northeastern Alberta, Canada. Environ Manage. 2025;75(11):2960–76. https://doi.org/10.1007/s00267-025-02235-w

64. Dacre HF, Crawford BR, Charlton-Perez AJ, Lopez-Saldana G, Griffiths GH, Veloso JV. Chilean Wildfires: Probabilistic Prediction, Emergency Response, and Public Communication. Bull Am Meteorol Soc. 2018;99(11):2259–74. https://doi.org/10.1175/BAMS-D-17-0111.1

65. Li Y, Yu D, Liu Z, Zhang M, Gong X, Zhao L. Graph Neural Network for spatiotemporal data: methods and applications. 2023;

66. Eubanks Virginia. Automating inequality : how high-tech tools profile, police, and punish the poor. New York: St. Martin's Press; 2018.

67. O'Neil C. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy [Internet]. 1st ed. Coll. Res. Libr. New York: Crown Publishers; 2016.

68. Zuboff S. The Age of Surveillance Capitalism. Social Theory Re-Wired. New York: Routledge; 2023. pp. 203–13. https://doi.org/10.4324/9781003320609-27

69. Pasquale F. New Laws of Robotics [Internet]. Harvard University Press; 2020. https://doi.org/10.4159/9780674250062

70. Couldry N, Mejias UA. The Costs of Connection [Internet]. How Data Is Colonizing Human Life and Appropriating It for Capitalism. Stanford University Press; 2020. https://doi.org/10.1515/9781503609754

71. Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Comput. 1997;9(8):1735–80. https://doi.org/10.1162/neco.1997.9.8.1735

72. Kovács S, Rabbi MF, Máté D. Global Food Security, Economic and Health Risk Assessment of the COVID-19 Epidemic. Mathematics. 2021;9(19):2398. https://doi.org/10.3390/math9192398

73. World Bank Group. Gross capital formation (% of GDP) | Data [Internet]. 2024. https://data.worldbank.org/indicator/NE.GDI.TOTL.ZS. Accessed 14 Nov 2025.

74. IRENA. Renewable capacity statistics 2024 [Internet]. Abu Dhabi; 2024.

75. Energy Institute. Statistical Review of World Energy [Internet]. London; 2025.

76. United Nations Environment Programme (UNEP). Global Environment Outlook [Internet]. 2024. https://www.unep.org/resources/global-environment-outlook. Accessed 4 Jan 2026.

77. Parker WS. Model Evaluation: An Adequacy-for-Purpose View. Philos Sci. 2022/01/01. 2020;87(3):457–77. https://doi.org/10.1086/708691

78. Saltelli A, Bammer G, Bruno I, Charters E, Di Fiore M, Didier E, et al. Five ways to ensure that models serve society: a manifesto. Nature. 2020;582(7813):482–4. https://doi.org/10.1038/d41586-020-01812-9

79. Scheffer M, Carpenter S, Foley JA, Folke C, Walker B. Catastrophic shifts in ecosystems. Nature. 2001;413(6856):591–6. https://doi.org/10.1038/35098000

80. Liu J, Dietz T, Carpenter SR, Alberti M, Folke C, Moran E, et al. Complexity of Coupled Human and Natural Systems. Science (1979). 2007;317(5844):1513–6. https://doi.org/10.1126/science.1144004

81. Pomponi F, Moncaster A. Embodied carbon mitigation and reduction in the built environment – What does the evidence say? J Environ Manage. 2016;181:687–700. https://doi.org/10.1016/j.jenvman.2016.08.036

82. Ostrom E. A General Framework for Analyzing Sustainability of Social-Ecological Systems. Science (1979). 2009;325(5939):419–22. https://doi.org/10.1126/science.1172133

83. Box GEP, Jenkins GM, Reinsel GC. Time Series Analysis [Internet]. Time Series Analysis: Forecasting and Control: Fourth Edition. Wiley; 2008. https://doi.org/10.1002/9781118619193

84. Folke C, Biggs R, Norström A V., Reyers B, Rockström J. Social-ecological resilience and biosphere-based sustainability science. Ecology and Society. 2016;21(3):art41. https://doi.org/10.5751/ES-08748-210341

85. Harmsworth G, Awatere S. Indigenous Māori knowledge and perspectives of ecosystems. Manaaki Whenua Press. 2013;2.1:274–86.

86. Nordhaus WD. A Review of the Stern Review on the Economics of Climate Change. J Econ Lit. 2007;45(3):686–702. https://doi.org/10.1257/jel.45.3.686

87. Brandt F, Geist C, Peters D. Optimal bounds for the no-show paradox via SAT solving. Math Soc Sci. 2017;90:18–27. https://doi.org/10.1016/j.mathsocsci.2016.09.003

88. Veale M, Zuiderveen Borgesius F. Demystifying the Draft EU Artificial Intelligence Act — Analysing the good, the bad, and the unclear elements of the proposed approach. Computer Law Review International. 2021;22(4):97–112. https://doi.org/10.9785/cri-2021-220402

89. Sen A. The Quality of Life [Internet]. Nussbaum M, Sen A, editors. Oxford University Press; 1993. https://doi.org/10.1093/0198287976.001.0001

90. Nussbaum M. CAPABILITIES AS FUNDAMENTAL ENTITLEMENTS: SEN AND SOCIAL JUSTICE. Fem Econ. 2003;9(2–3):33–59. https://doi.org/10.1080/1354570022000077926

91. Sen A. Human Rights and Capabilities. Journal of Human Development. 2005;6(2):151–66. https://doi.org/10.1080/14649880500120491