https://doi.org/10.55343/CogSust.22409

Artificial Intelligence and Sustainability: A Conceptual Framework for System-Level Impact Assessment

Adrienn Csernovszky
Budapest University of Technology and Economics, Department of Environmental Economics and Sustainability
Budapest, Hungary
adrienn.csernovszky@gmail.com

Maria Szalmane Csete
0000-0001-7170-9402
Budapest University of Technology and Economics, Department of Environmental Economics and Sustainability
Budapest, Hungary
csete.maria@gtk.bme.hu

Abstract

Artificial intelligence (AI) is rapidly emerging as a general-purpose technology with far-reaching implications for sustainable development. While AI applications are increasingly deployed across sectors such as healthcare, energy systems, urban management, and education, their overall sustainability impacts remain uncertain and often contradictory. Existing research typically examines isolated effects of AI within individual sustainability pillars, limiting understanding of systemic interactions, feedback loops, and long-term consequences. This study introduces a conceptual-analytical framework designed to assess the sustainability impacts of artificial intelligence across environmental, economic, and social dimensions, and extends it with an additional individual-level pillar. The framework defines a set of AI Impact Groups (AIG) that translate technological capabilities into system-level functions, including perception, learning, strategic foresight, coordination, and risk detection. In addition, the model introduces key input parameters – AI intensity, adoption level, autonomy, quality of use, and system quality – that influence how AI capabilities translate into sustainability outcomes. By linking AI capabilities, system-level functions, and sustainability pillars, the proposed framework enables a more integrated assessment of both opportunities and risks associated with AI deployment. The model highlights how AI impacts propagate across domains and may generate both short-term benefits and long-term systemic risks, such as rebound effects, technological dependence, or skill erosion. The framework provides a foundation for future scenario analysis, sector-specific impact assessment, and interdisciplinary collaboration to understand and govern AI-driven sustainability transitions.

Keywords
Artificial intelligence, Sustainability, SDGs, Systems approach, AI governance

1. Introduction

This study describes an ongoing research project conducted at the Department of Environmental Economics and Sustainability, Budapest University of Technology and Economics (BME). It aims to provide a general conceptual framework for assessing the impacts of artificial intelligence applications on sustainability, with a particular focus on risks and opportunities. Artificial intelligence (AI) has emerged as a key driver of technological and economic transformation and is increasingly adopted across a wide range of sectors and domains. However, the sustainability implications of AI remain uncertain, as its effects are often contradictory, making it difficult to determine whether AI ultimately supports or undermines long-term sustainability objectives (Katper at al., 2026). Existing research typically assesses AI impacts in a fragmented manner, focusing on one or two sustainability pillars or addressing isolated effects and outcomes (Antwi, Agyapong, Owusu, 2026). Therefore, we introduce the core definitions of artificial intelligence and sustainability, examine their interrelationships, and establish the need for a new modelling approach.

1.1 Basic concepts

Intelligence is defined as “the ability to learn and perform suitable techniques to solve problems and achieve goals, appropriate to the context in an uncertain, ever-varying world” (Manning, 2020: 1). Alan Turing asked the first foundational

https://doi.org/10.55343/CogSust.22409

question “Can machines think?” in 1935, and the Turing Test became a standard to measure machine intelligence (Doğan, Doğan, 2026). John McCarthy introduced the term Artificial Intelligence in 1955, who defined it as “the science and engineering of making machines that can perform tasks that would normally require human intelligence” (EIMT, 2026). Early artificial intelligence (AI) applications primarily focused on symbolic and logical problem solving. A notable example is the Logic Theorist, developed by Alan Newell and Herbert A. Simon, which demonstrated that machines are capable not only of performing calculations but also of solving problems through logical reasoning (Velupillai, Kao, 2014). Subsequent advances in machine learning and deep learning fundamentally transformed how machines learn from data. The work of Yann LeCun, Geoffrey Hinton, and Yoshua Bengio revolutionised deep neural network training, laying the foundation for a wide range of contemporary AI applications, including robotics, speech recognition, voice assistants, and language models. As a result, AI has evolved into a widely adopted technology across multiple sectors (Browning, LeCun, 2023; Matsuo et al., 2022).

AI types can be classified in multiple frameworks, and the terminology often overlaps or evolves as AI technologies advance rapidly. The term *AI types* generally refers to categories of AI systems, which are distinguished based on their technological approach, high-level functional role, or theoretical capability level (Table 1):

Table 1. Classification of AI systems by technological approach, function, and capability level

Classification basis	AI types
Technological approach	Symbolic AI, Machine Learning (including Deep Learning), Hybrid AI
High-level functional role	Perception AI, Prediction AI, Decision / Optimisation AI, Generative AI, Interaction AI, Autonomous AI
Theoretical capability level	Artificial Narrow AI (ANI), Artificial General Intelligence (AGI), Artificial Superintelligence (ASI)

Conversely, *AI functional categories* refer to groups of algorithms and mathematical approaches that define the main functions performed by AI systems. These functional categories enable specific *AI capabilities*, which represent the underlying abilities of AI systems. *AI applications* refer to the concrete, real-world uses of these functions in specific domains. Table 2 provides examples for all three levels: AI functional categories, AI capabilities (OECD, 2025) and AI applications.

Table 2. Examples of functional categories, AI capabilities (OECD, 2025), and applications

AI functional categories	AI capability	AI applications
Natural language processing (text understanding, text generation, speech processing)	Language	Chatbots, machine translation, text summarisation, virtual assistants
Affective computing and human–AI interaction (emotion recognition, conversational interaction)	Social interaction	Social robots, sentiment-aware customer service bots, and AI tutoring systems
Decision support, predictive analytics, optimisation	Problem solving	Medical diagnostic support systems, predictive maintenance, logistics optimisation
Generative AI and content generation	Creativity	AI image generators, text generation tools, music generation, and design assistants

https://doi.org/10.55343/CogSust.22409

Uncertainty estimation and AI-assisted decision making	Metacognition and critical thinking	Fraud detection systems flagging uncertain cases, risk assessment tools
Knowledge representation, information retrieval, and recommendation systems	Knowledge, learning and memory	Search engines, enterprise knowledge assistants, recommender systems
Computer vision (image recognition, object detection, visual analysis)	Vision	Medical image analysis, facial recognition, automated quality inspection
Robotic manipulation and physical automation	Manipulation	Industrial robotic arms, warehouse picking robots, and automated assembly systems
Autonomous systems and robotic control	Robotic intelligence	Autonomous warehouse robots, delivery robots, and agricultural robots

Global trends in AI development have been examined in the AI Index Report, which analyses data from the period 2022–2025 (SU HAI, 2025). The report characterises AI as one of the most transformative technologies of the 21st century. AI performance has surpassed human baseline levels in several domains (Figure 1), contributing to the rapid diffusion of AI systems and reshaping the boundaries between human and machine tasks.

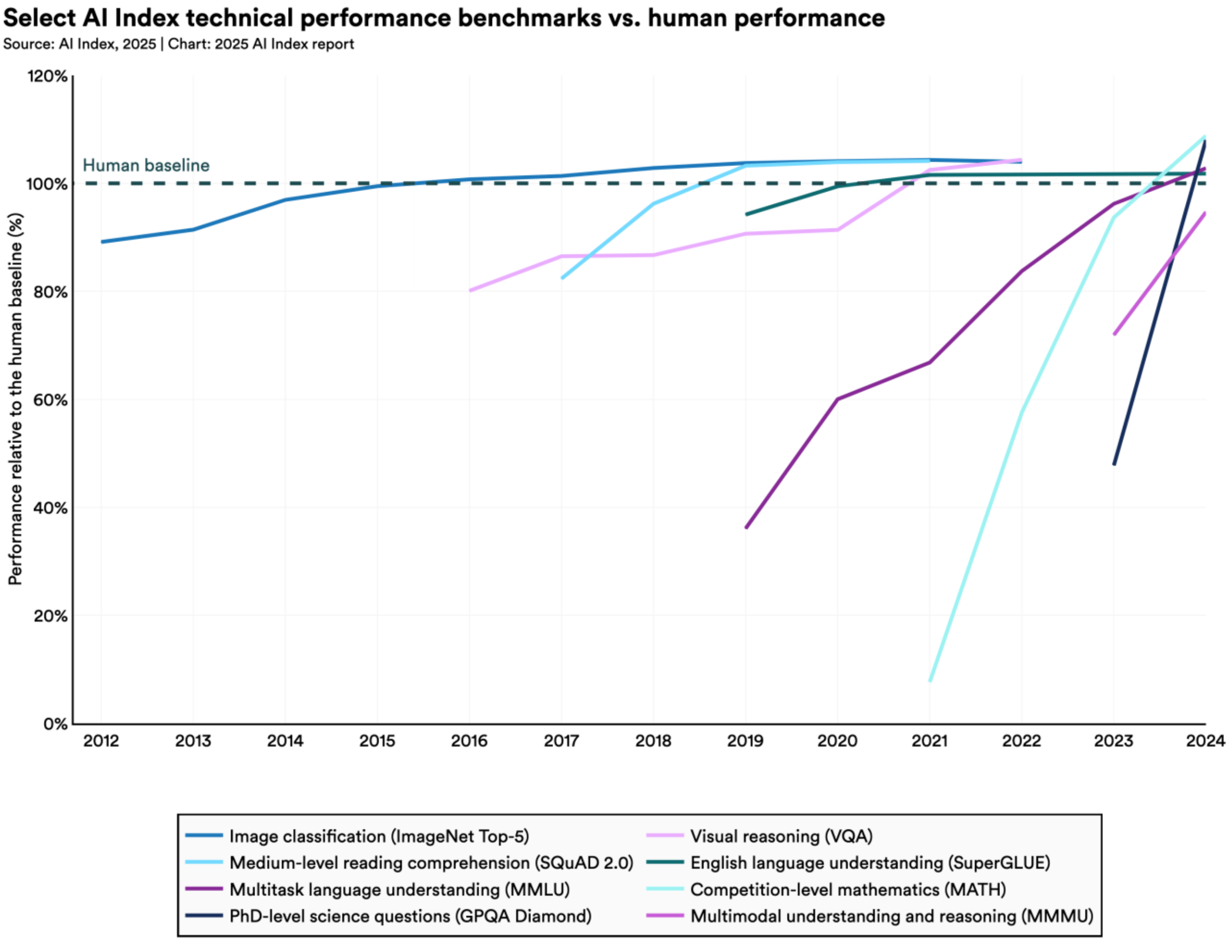


Figure 1. Performance relative to the human baseline (SU HAI, 2025)

https://doi.org/10.55343/CogSust.22409

In the past few years, AI has moved from the laboratory to daily life, with a strong presence in data-intensive sectors such as information and communication technologies, finance, healthcare, and manufacturing. The share of organisations applying AI in at least one function evolved from 20% in 2017 to 78% in 2024, based on 1363 participants representing a full range of regions (Singla et al., 2025).

AI optimism means that public perception of AI is positive and that it is seen as having more benefits than drawbacks. AI optimism shows a general upward trend across several countries, though in a few countries, stagnation or declining optimism can be observed. China has the highest level of AI optimism, with more than 80% of respondents agreeing, while the United States, the Netherlands, and Belgium report levels below 40%, based on data from 2022–2024. There are other perspectives to consider, as AI optimism is unevenly distributed and depends not only on technological performance but also on how visible and immediate the benefits of AI are, who experiences them directly, individual capabilities, and the surrounding AI regulations and policies. AI is mostly utilised in the economic and business domains, where productivity improvements enabled by AI are directly observable, which contributes to higher levels of AI optimism. In contrast, in sectors where tasks are easily automated, AI is associated with job insecurity, and among individuals with lower technical skills, it may be a source of perceived loss of control, leading to lower levels of AI optimism.

AI efficiency refers to how effectively AI applications convert consumed resources into achieved performance, and it is driven by fine-tuned small models and improved technologies and algorithms. Energy efficiency has improved by approximately 40%, while hardware costs have declined by around 30% annually. A rebound effect occurs, meaning that improvements in energy efficiency lower costs but also stimulate increased technology use. Consequently, AI efficiency gains do not necessarily lead to reductions in total resource or energy consumption. The rapid growth in data requirements necessitates more extensive data collection, storage, and processing, further intensifying the rebound effect. As AI has become a central concern due to its data-driven energy demand and societal impacts, governments have increased regulatory activity. In the United States, 59 AI-related regulations were issued in 2024 (SU HAI, 2025), and the European Union introduced the AI Act (Regulation (EU) 2024/1689).

The first definition of *sustainable development* was introduced in the Brundtland Report (Brundtland Commission, 1987) as “development that meets the needs of the present without compromising the ability of future generations to meet their own needs.” The United Nations’ 2030 Agenda for Sustainable Development defines 17 Sustainable Development Goals (SDGs) addressing key global challenges such as climate change, equal opportunities, education, and healthcare (EurLex, n. d.).

1.2 Sustainability and AI

This section reviews how frequently AI tools are used in sustainability-related studies and identifies the domains in which they are most commonly applied. It also summarises prior research on AI’s impacts on sustainability, including the main approaches and classification schemes. Finally, the key research gaps are highlighted, and it is described how the proposed methodology aims to complement and extend existing work by providing a different, more integrative perspective.

The number of AI applications in sustainability-related research has increased, as illustrated in Figure 2. AI is primarily used for forecasting, system optimisation, data mining, and remote sensing, as well as for accelerated experimentation and fast approximate simulation (Gohr et al., 2025).

https://doi.org/10.55343/CogSust.22409

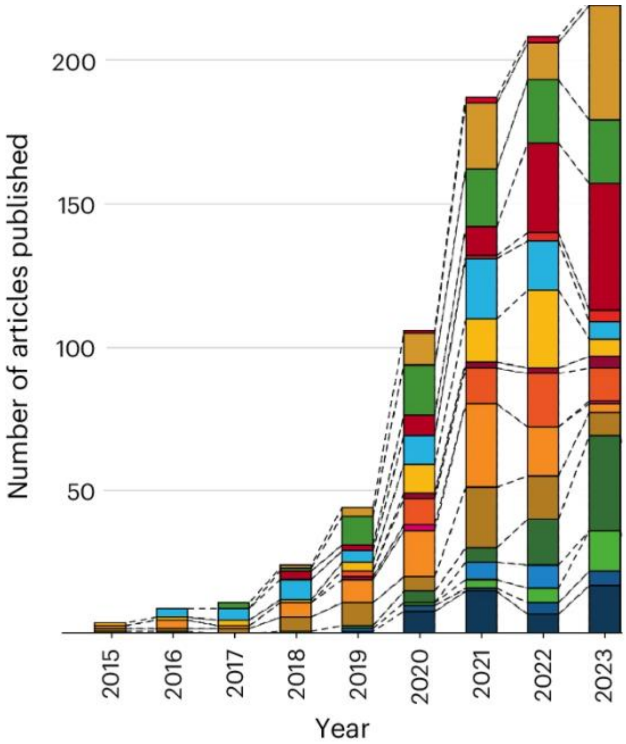
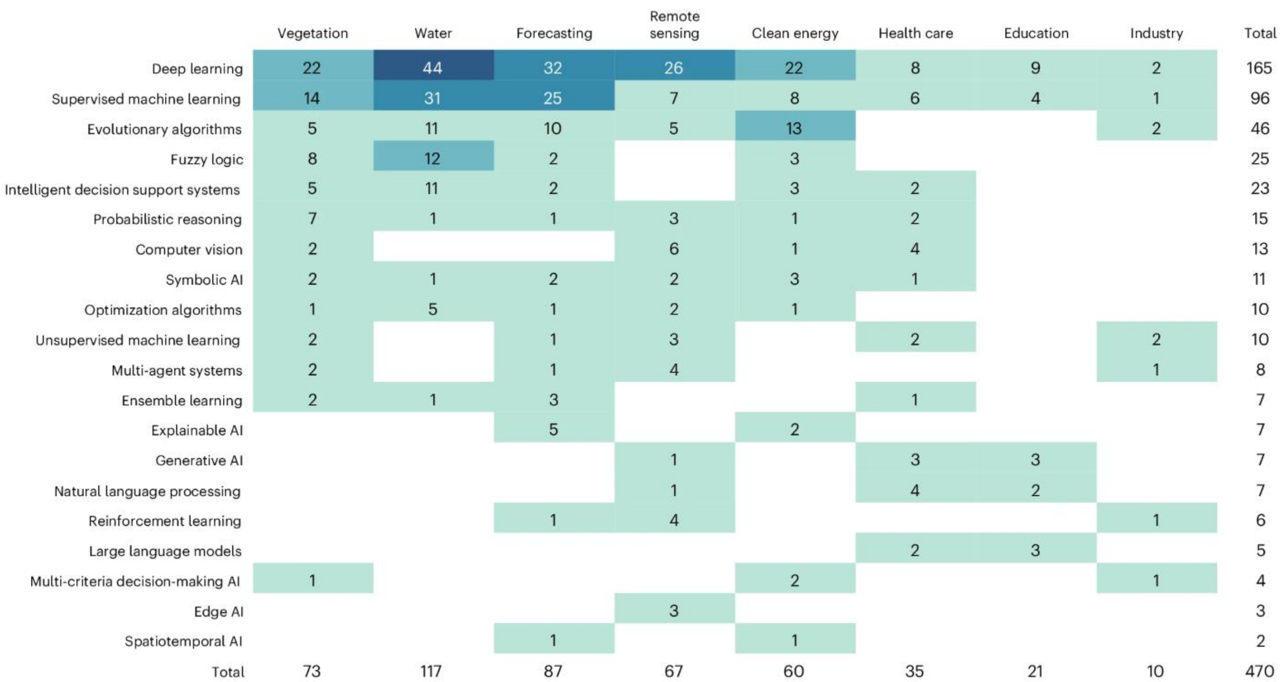


Figure 2. Number of AI applications in sustainability-related research (Gohr et al., 2025).

The same research classifies AI applications across sustainability topic groups and shows that the largest number of studies is in deep learning and supervised machine learning, particularly in applications involving system optimisation in water management and forecasting (Figure 3). The study reveals a methodological concentration, focusing on a limited number of sustainability domains or a narrow range of areas.



	Vegetation	Water	Forecasting	Remote sensing	Clean energy	Health care	Education	Industry	Total
Deep learning	22	44	32	26	22	8	9	2	165
Supervised machine learning	14	31	25	7	8	6	4	1	96
Evolutionary algorithms	5	11	10	5	13			2	46
Fuzzy logic	8	12	2		3				25
Intelligent decision support systems	5	11	2		3	2			23
Probabilistic reasoning	7	1	1	3	1	2			15
Computer vision	2			6	1	4			13
Symbolic AI	2	1	2	2	3	1			11
Optimization algorithms	1	5	1	2	1				10
Unsupervised machine learning	2		1	3		2		2	10
Multi-agent systems	2		1	4				1	8
Ensemble learning	2	1	3			1			7
Explainable AI			5		2				7
Generative AI				1		3	3		7
Natural language processing				1		4	2		7
Reinforcement learning			1	4				1	6
Large language models						2	3		5
Multi-criteria decision-making AI	1				2			1	4
Edge AI				3					3
Spatiotemporal AI			1		1				2
Total	73	117	87	67	60	35	21	10	470

Figure 3. Application frequency of the top 20 most frequent AI types across sustainability topic groups (Gohr et al., 2025)

https://doi.org/10.55343/CogSust.22409

Toderas (2025) analyses the complex role of AI related to sustainability. AI applications were highlighted for their positive impact, including optimising complex systems, monitoring across several fields, and smart city development, while challenges were primarily addressed through risk assessment and mitigation recommendations. The main risk categories were defined using a multi-pillar, functional methodology: ecological risks, socio-economic and ethical risks, and data and algorithm-related risks (Figure 4).

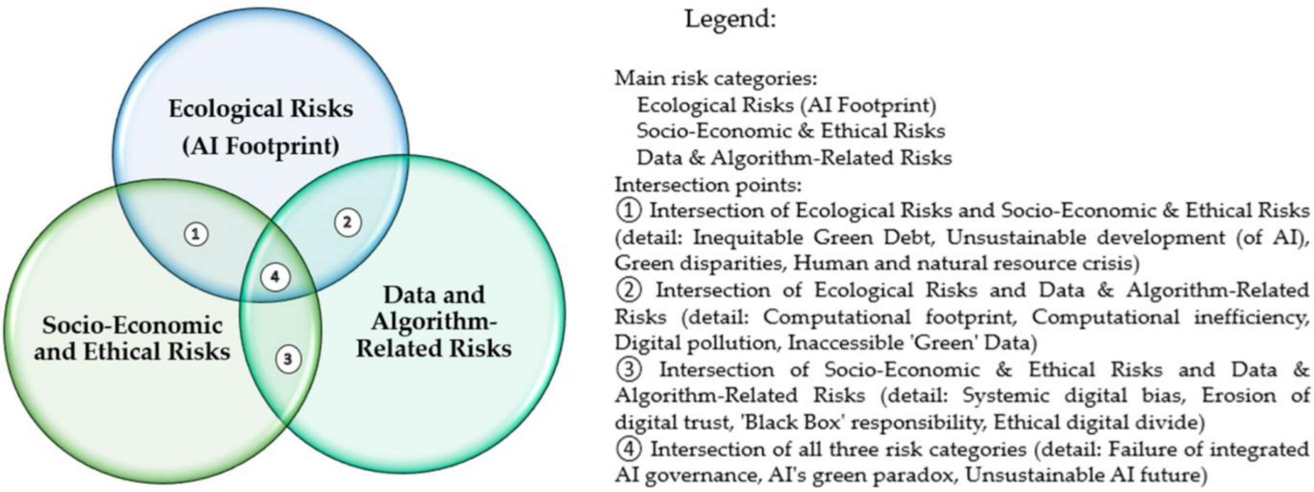


Figure 4. Main risk categories (Toderas, 2025)

The risk score was calculated by combining the probability of a risk's occurrence with its severity. Identified critical risks were (i) data bias, (ii) risk of amplifying inequalities and (iii) technological divide (AI divide) (Toderas, 2025).

The concept of the *AI green paradox* has emerged (Toderas, 2025). It describes a phenomenon in which AI applications intended to improve sustainability may increase overall energy use and emissions due to the hardware and software requirements of AI systems. This highlights the need to assess AI impacts across its entire lifecycle. A lack of transparency associated with so-called *black-box models*, as well as other socio-economic risks, has also been identified in relation to AI applications. *Green debt* refers to the risk that AI's short-term benefits are offset by its long-term impacts.

Mitigating these risks requires research in Green AI, the development of responsible governance frameworks, international collaboration, increased focus on education and training, and the development of "Green AI engineers" (Toderas, 2025). While this study adopts a multi-pillar perspective, it lacks an individual dimension of sustainability. Furthermore, there is a limited explanation of how impacts propagate across pillars, which input factors lead to specific outcomes, and how risks or opportunities may amplify over time.

1.2.1 AI impacts across sustainability pillars

Isolated effects, focusing on a single sustainability pillar or a single impact direction, despite the fact that AI and its sustainability consequences are complex and systemic. These impacts depend on sectors, regions, and time, and require analysing both benefits and risks, capturing interdependencies and cross-effects across sustainability pillars. Table 3 provides an overview of how AI-based applications generate both benefits and risks across environmental, economic, social, and individual pillars of sustainability, illustrating these dynamics with one example for each domain.

Table 3. Examples of AI applications and their associated opportunities and risks across sustainability pillars

Pillar	Domain	AI application examples	Potential sustainability benefits	Potential risks

https://doi.org/10.55343/CogSust.22409

Environmental	Resource and material use	Material and product design system with optimisation	Reduction of raw material use	More efficient and cheaper products may increase consumption (rebound effect)
Economic	Labour market and employment	Process automation system	Emergence of new sectors applying AI	Job displacement
Social	Equality	Digital educational assistant	Improved accessibility	Digital divide
Individual	Health and well-being	Mental health assistant	Health support	Excessive reliance on technology and risk of psychological dependency

AI impacts do not remain confined to a single pillar but often propagate across other pillars as well. For example, AI-optimised transportation can reduce emissions and improve air quality, leading to better public health outcomes and lower healthcare-related economic costs. This highlights the need to identify input factors that shape the direction of effects and to develop a common analytical framework capable of assessing all sustainability pillars.

1.2.2 AI applications across the UN Sustainable Development Goals

Artificial intelligence is increasingly recognised as a general-purpose technology with the potential to influence multiple dimensions of sustainable development. While the relationship between artificial intelligence and sustainability is often discussed in terms of environmental, economic, and social pillars, global policy frameworks typically structure sustainability challenges through the Sustainable Development Goals (SDGs) defined in the United Nations 2030 Agenda (UN, 2015).

Mapping AI applications across the SDGs provides a structured way to identify where artificial intelligence may help solve global challenges and where new risks may emerge. AI technologies are currently applied in areas such as healthcare diagnostics, energy system optimisation, urban management, climate modelling, and education. These applications demonstrate that AI can accelerate progress toward multiple SDGs simultaneously. However, the same technologies may also introduce systemic risks, including technological dependence, amplification of inequality, and increased energy demand associated with digital infrastructures.

Table 4 summarises selected examples of AI applications across key SDGs and highlights both potential opportunities and associated risks. The examples illustrate that AI-related sustainability impacts rarely occur in isolation and often involve complex interactions between environmental, economic, social, and individual dimensions. The examples presented in Table 4 demonstrate that artificial intelligence can support progress across multiple SDGs by improving efficiency, enabling better monitoring systems, and enhancing decision-making processes. In particular, AI has strong potential in areas characterised by complex systems and large datasets, such as healthcare diagnostics, energy systems, climate modelling, and urban management.

At the same time, AI-related sustainability impacts are inherently ambivalent. Technologies designed to improve efficiency may generate unintended systemic consequences. For example, improvements in energy efficiency enabled by AI may reduce operational costs, thereby stimulating greater use of digital technologies and contributing to rebound effects and higher overall energy consumption. Similarly, AI-driven automation may increase productivity and innovation while simultaneously contributing to labour market disruptions or new forms of technological dependency.

Another important observation is that AI impacts often extend beyond the domain in which the technology is initially deployed. For instance, AI-supported energy optimisation in urban infrastructure may reduce emissions (environmental pillar), improve economic efficiency (economic pillar), and enhance urban quality of life (social pillar). At the same time, increased reliance on AI systems may influence individual skills and competencies, particularly in decision-making, learning, and risk perception.

These examples highlight that the sustainability implications of artificial intelligence cannot be fully understood through isolated sectoral analyses. Instead, they require analytical approaches capable of capturing cross-domain interactions, feedback loops, and long-term systemic effects.

https://doi.org/10.55343/CogSust.22409

Table 4. AI applications across selected SDGs: potential sustainability benefits and risks

SDG	Domain	AI application examples	Potential sustainability benefits	Potential risks
SDG 3 – Good Health and Well-being	Healthcare systems	Medical image diagnostics, epidemic forecasting, and AI-based mental health assistants	Faster diagnosis, improved healthcare access, personalised treatment	Algorithmic bias, data privacy concerns, and overreliance on automated decision support
SDG 4 – Quality Education	Education systems	Adaptive learning platforms, AI tutoring systems, automated assessment	Personalised education, broader access to learning resources	Reduced critical thinking, digital divide
SDG 7 – Affordable and Clean Energy	Energy systems	Smart grid optimisation, renewable energy forecasting, and demand management systems	Increased energy efficiency, improved integration of renewable energy	Increased electricity demand of data centres, rebound effects
SDG 9 – Industry, Innovation and Infrastructure	Industrial production	Predictive maintenance, AI-driven product design, automated quality control	Higher productivity, resource efficiency, and accelerated innovation	Job displacement, technological dependency
SDG 11 – Sustainable Cities and Communities	Urban systems	Intelligent traffic management, smart waste management, urban climate modelling	Reduced congestion and emissions, improved urban services	Surveillance risks, data governance challenges
SDG 12 – Responsible Consumption and Production	Supply chains and the circular economy	AI-based supply chain monitoring, material optimisation, and waste sorting systems	Resource efficiency, waste reduction, improved transparency	Acceleration of consumption cycles, supply chain concentration
SDG 13 – Climate Action	Climate monitoring and modelling	Climate prediction models, carbon monitoring systems, and disaster forecasting	Improved climate risk management and mitigation planning	Carbon footprint of AI infrastructure
SDG 16 – Peace, Justice and Strong Institutions	Governance systems	AI-assisted public administration, corruption detection systems, policy simulation tools	Increased administrative efficiency and transparency	Algorithmic decision bias, reduced democratic oversight

The SDG-based overview shows that artificial intelligence applications simultaneously affect multiple sustainability domains and often create both opportunities and risks. However, these impacts within specific sectors or individual sustainability pillars limit the ability to analyse systemic interactions and cross-pillar effects.

https://doi.org/10.55343/CogSust.22409

To address this challenge, this study introduces a conceptual analytical framework designed to assess the sustainability impacts of artificial intelligence across interconnected dimensions. The proposed approach extends the traditional three-pillar sustainability model by incorporating an additional individual dimension. It introduces a set of AI impact groups and input parameters that enable a more systematic assessment of how AI capabilities translate into sustainability outcomes.

Figure 5 presents a general AI system evaluation framework covering the entire lifecycle. Evaluation is embedded across multiple phases, including design, data collection, model and system development, and deployment. Throughout these phases, various stakeholders conduct different forms of assessment – such as impact evaluation, data evaluation, model evaluation, model testing, benchmarking, and capability evaluation.

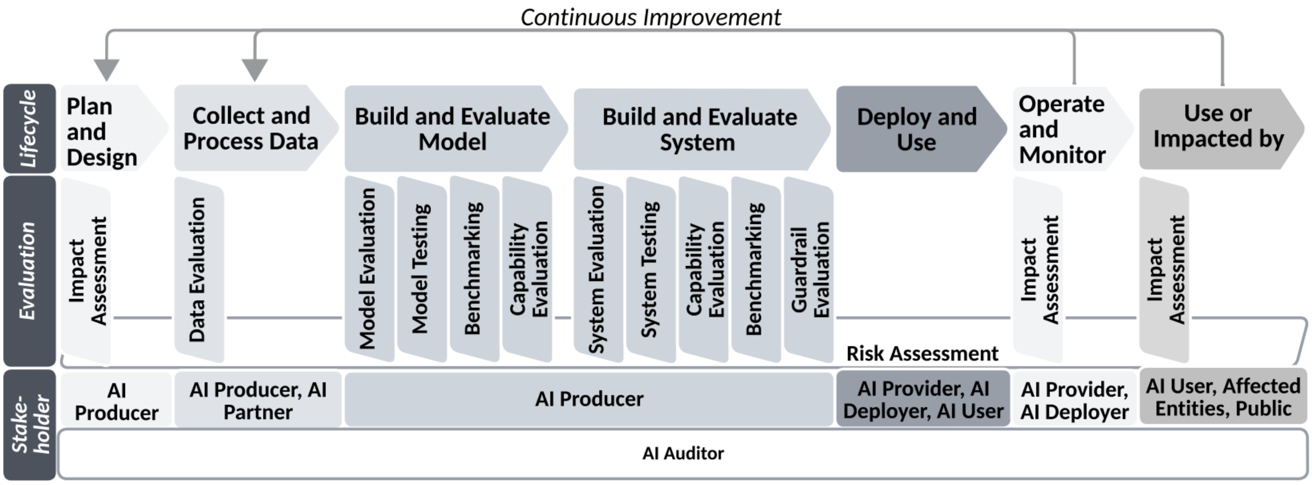


Figure 5. AI system evaluation from the point of view of supply chain (Xia et al., 2024: 78)

Humanity's last exam (HLE) (Phan et al., 2025a) provides an evaluation framework for large language models, consisting of 2500 challenging questions across a wide range of subject areas, contributed by nearly 1000 subject experts. Figure 6 compares the accuracy of several large language models across different benchmark types. While models achieve high scores on other benchmarks, HLE remains lower, indicating that it is more challenging and focuses on expert-level questions.

https://doi.org/10.55343/CogSust.22409

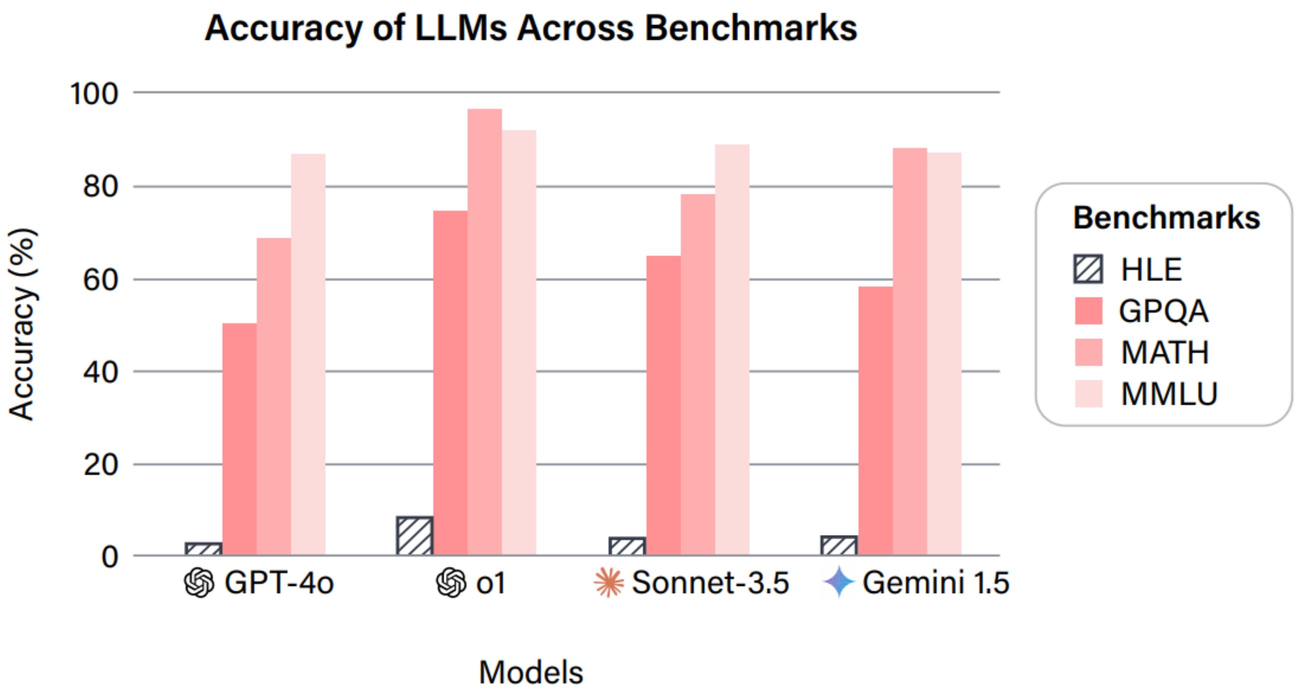


Figure 6. HLE and existing benchmark comparison in accuracy (Phan et al., 2025a)

Accuracy scores have increased over time: over 1.5–2 years, leading models improved from around 3–5% to 30–40% (see Figure 7). Competition among developers (OpenAI, Google, Anthropic, xAI) has been a key driver of this progress, contributing to higher accuracy levels. Newly introduced question sets and tasks are more informative for evaluation than repeated use of the same benchmarks, as models can be explicitly trained on previously released questions.

https://doi.org/10.55343/CogSust.22409

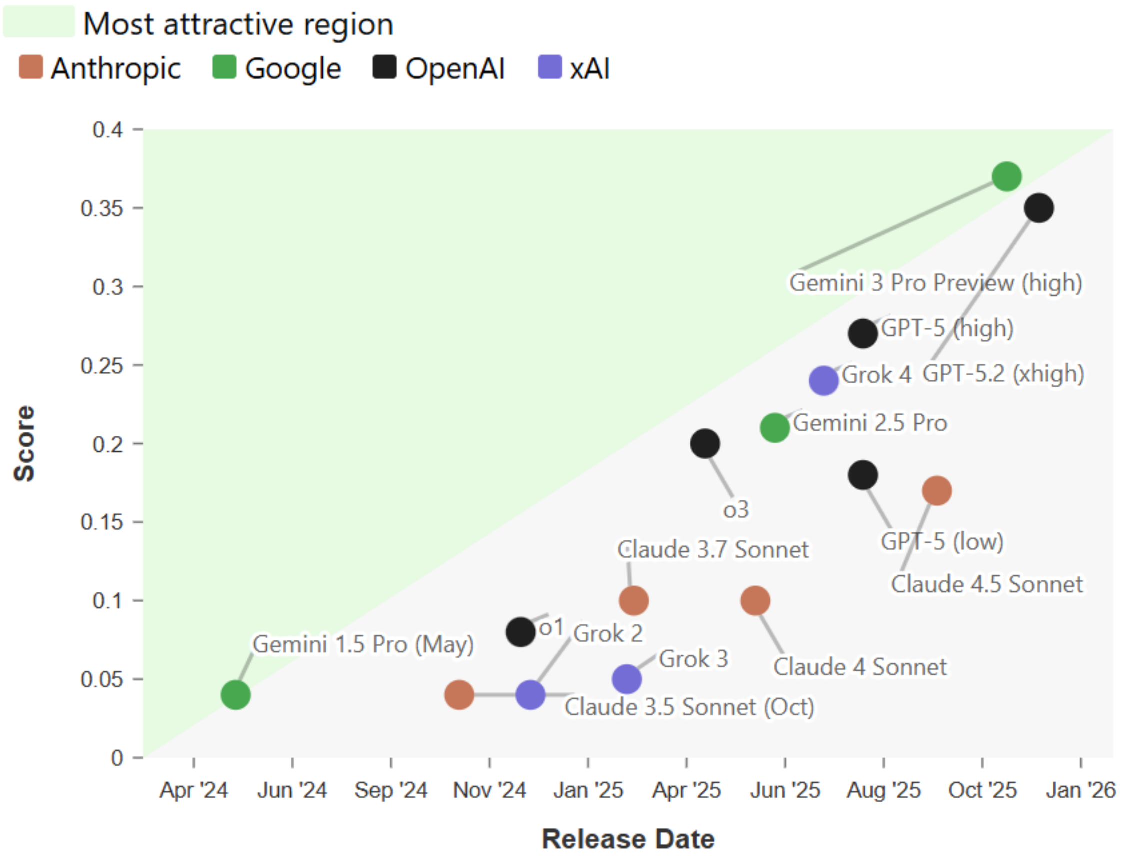


Figure 7. HLE benchmark leaderboard showing model accuracy scores over time. Results are based on evaluations reported by Artificial Analysis, an independent AI benchmarking and analysis company (Phan et al., 2025b)

Figure 8 shows the average number of completion tokens used across subject domains. Models tend to produce more tokens for physics, engineering and computer science tasks, whereas biology, medicine and humanities tasks require shorter responses. Differences in average token usage indicate that computational demands are not only model-specific but also domain-dependent. Longer reasoning chains involve more steps, increasing the likelihood of errors and making it more difficult to maintain coherence across extended responses. In such cases, the generation of intermediate reasoning steps combined with uneven or incomplete training data can further increase the risk of hallucinations and inaccuracies.

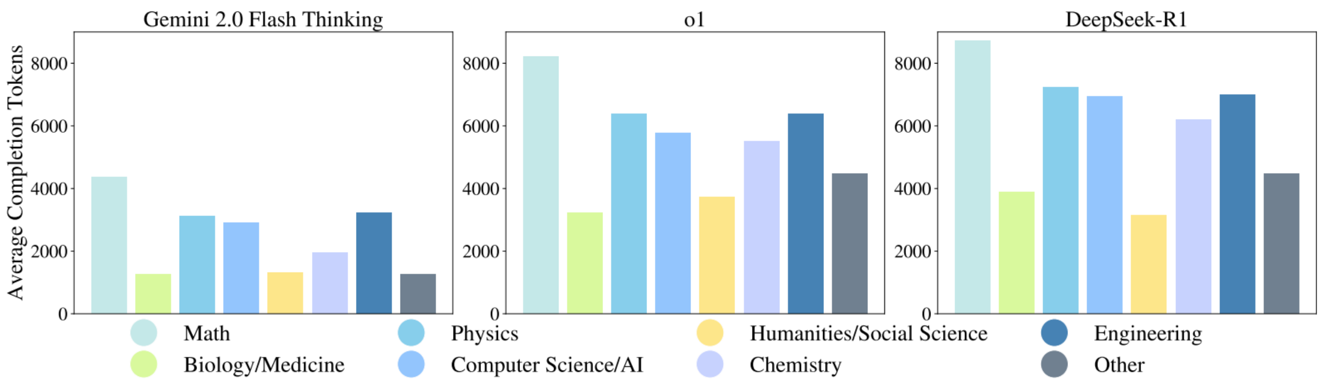


Figure 8. Average completion token counts across subject domains for selected AI models evaluated on the Humanity's Last Exam (HLE) (Phan et al., 2025a)

https://doi.org/10.55343/CogSust.22409

2. Framework description

This ongoing research is developing a new methodology for impact analysis. In this introductory study, the focus is on defining impact categories and input factors that will form the foundation of a model designed to assess impacts across sustainability pillars and the model's input factors.

2.1 Artificial Intelligence impact groups

AI impact groups are defined to enable the assessment of effects and the analysis of changes in skills, capabilities, and resilience. The proposed framework covers perception, interpretation, decision support, learning, risk and conflict detection, coordination, creative future-oriented thinking, motivation, and memory. Table 5 presents the AI impact groups. To emphasise the importance of each group, it also outlines the system-level consequences of their absence.

Table 5. AI impact groups and system-level consequences of their absence

No	AI Impact Groups - AIG	System-level consequences of absence
1	Strategic foresight and planning	The system is myopic, focusing only on current states and failing to account for delayed consequences.
2	Perception and pattern recognition	The system is blind, fails to detect warning signals and has delayed responses.
3	Knowledge integration and meaning formation	The system does not understand what it observes: the data are present, but lack coherence, and no interpretation is formed.
4	Learning and optimisation	The system does not optimise, repeat errors, and fails to adapt.
5	Risk and conflict detection	The system avoids difficult decisions.
6	Creative future construction and cyclical self-organisation	The system is unable to transform or develop; it optimises an outdated model and lacks alternative futures. Creative exhaustion.
7	Coordination and precision control	Intention does not translate into action; there is no execution.
8	Motivation, emotional and behavioural modelling	The system becomes exhausted, lacking persistence and commitment. Motivational collapse.
9	Memory and spatial-spatiotemporal modelling	The system does not store or properly utilise prior experience, resulting in short-term optimisation and decision-making.

Skill assessment example for pillars in case of 1. Strategic foresight and planning:

- Individual: system-level perspective, which helps to connect and integrate large-scale data patterns;
- Society: Societal system level perspective, which supports institutional transparency and strategic planning;
- Economy: Economic system level coherence, which covers a comprehensive assessment of economic performance and understanding of value change operations;
- Environment: Ecological system-level coherence, which helps in understanding the interdependencies of ecosystems and ecological networks.

2.2 Input parameters

2.2.1 Intensity

The depth of AI integration and the frequency of AI use are combined into a single input parameter, referred to as AI intensity. While depth of integration is primarily related to process, the frequency of use is more closely associated with individuals. Handling them separately would risk overrepresentation.
AI intensity is rated from 1 to 5:

1. Minimal integration and/or use (AI as experimentation):
   AI is rarely used in experimental settings or for isolated tasks. Without AI, there is no impact on system operation.
2. Low level of integration and/or use (AI as an optional tool):
   AI is used occasionally to support processes, but not the core of operations. Without AI, the system remains largely unchanged.

https://doi.org/10.55343/CogSust.22409

3. Moderate level of integration and/or use (AI as an integrated support layer):
   AI is widely used across many tasks, but it remains replaceable. Processes can function without AI, but less effectively. If AI fails, the system slows down or operates with reduced quality but continues to function.
4. High level of integration and/or use (AI has a mission-critical role):
   Processes are designed around AI, and it is embedded in core operations. If AI fails, the system is disrupted, and only partial operation is possible.
5. Full integration and/or high frequency of use (AI has a system-defining role):
   The system fully depends on AI; if AI fails, the system stops functioning.

Higher levels of AI intensity increase potential benefits while simultaneously amplifying systemic risks and dependencies. Meanwhile, lower levels of AI intensity are associated with limited impacts and competitive disadvantages due to missed efficiency gains.

2.2.2 AI adoption level

AI adoption level is defined as the extent to which AI applications are adopted within a sector.
AI adoption level is also rated from 1 to 5:

1. Very low adoption: AI is used in isolated experimental projects.
2. Low adoption: AI applications are used by a small number of actors within the sector.
3. Moderate adoption: AI applications are used by a significant share of actors within the sector.
4. High adoption: AI applications are used by the majority of actors in the sector.
5. Very high adoption: AI applications are used by nearly all actors within the sector.

AI adoption determines whether the impacts remain local or become systemic.

2.2.3 Autonomy

AI autonomy refers to the degree of autonomous operation of an AI application.
AI autonomy is rated from 1 to 5:

1. No autonomy (static AI): AI operates entirely under human control.
2. Low autonomy (assistant AI): AI supports tasks but executes decisions only with human approval.
3. Partial autonomy (supervised agent): AI operates under continuous human supervision.
4. High autonomy (autonomous operator): AI operates with human interventions only in exceptional cases.
5. Full autonomy (AI agent): AI operates without human intervention.

As AI autonomy increases, AI shifts from a supportive to a high-level decision-making tool. Reduced human oversight may amplify impacts and decrease the detectability of errors or unintended effects. Higher levels of autonomy may cover a broader range of tasks, potentially leading to skill erosion at the individual level, while also reducing human workload and improving operational efficiency.

https://doi.org/10.55343/CogSust.22409

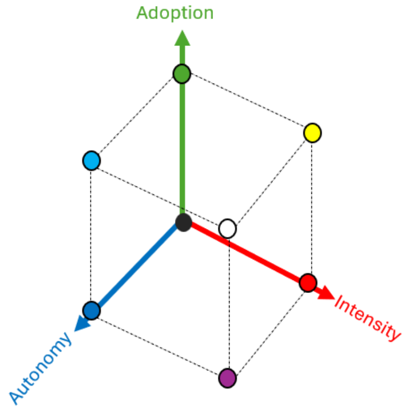


Figure 9. AI application space defined by adoption, autonomy and intensity

Figure 9 illustrates how the three input factors – AI adoption, autonomy and intensity – jointly define the space of existing AI applications, allowing any AI use case to be positioned within this framework. This input layer provides a functional classification, complemented by an additional layer focusing on quality, including quality of use and system quality.

2.2.4 Quality of use

Quality of use refers to the degree to which AI is applied consciously and critically, including how AI outputs are interpreted and evaluated. The following values can be assigned to this factor.

-1: Blind use: The output of an AI application is accepted without critical evaluation or validation.
0: Conscious use: The output of an AI application is accepted with some level of control, applying occasional validation.
1: Reflective use: The output of AI is interpreted and critically evaluated. AI supports learning, systemic understanding, and decision-making.

AI validation refers to the user confirming that the AI application is being used for its intended purpose. Quality of use is a key input factor as it shapes the direction of AI's influence on outcomes. Blind use may amplify the negative impacts of AI and contribute to unsustainable practices. In contrast, reflective use can lead to more informed, responsible and sustainable supportive decision making.

2.2.5 System quality

The last input factor relates to the evaluation of an AI application, including its robustness, performance, security and safety, reliability, and transparency. A wide range of AI application types exists, each with distinct evaluation processes. This section provides illustrative examples and proposes a unified rating approach.

In the present framework, the definition of the system quality input factor is kept general, as a wide variety of AI applications exist across sectors. Within a given domain, professionals with task-specific expertise are best positioned to evaluate the quality of an AI application on a scale of 1 to 5. To support this assessment, the following classification provides a structured reference framework:

1. Poor quality: Low data quality with frequent errors or hallucinations without any validation and risk management, and exhibiting a poor level of security.
2. Low quality: Occasional error handling combined with poor data quality.
3. Moderate quality: Acceptable performance with partial validation, basic monitoring, and risk management functions applied in several cases.
4. High quality: Strong data and model quality, regular validation, continuous monitoring and risk management application.
5. Very high quality: Verified and stable performance supported by high quality of data, robust and secure operation, documented processes and measurable compliance.

https://doi.org/10.55343/CogSust.22409

An evaluation of system quality highlights that the impacts of AI on sustainability also depend on model performance, deployment context, governance mechanisms, and continuous monitoring.

3 Discussion

The primary goal of this ongoing research is to develop an analytical framework for assessing the impacts of artificial intelligence across the sustainability pillars. This research investigates how the effects of AI applications propagate across the different pillars of sustainability, what determines whether these impacts are positive or negative, and which input factors shape the sustainability impacts of AI.

The analysis first reviews key concepts related to artificial intelligence, including AI types, functional categories, capabilities, and application examples, and discusses related phenomena such as efficiency gains, rebound effects, and the AI green paradox. Global AI trends show that artificial intelligence is becoming one of the most transformative technologies and is increasingly used in sustainability-related research.

Existing approaches focus on isolated effects, single domains, or specific AI applications; the proposed methodology captures cross-pillar interactions, feedback mechanisms, and the role of input factors. By extending the traditional three-pillar sustainability framework with an individual dimension, the model enables an assessment of how AI influences skills, capabilities, resilience, and long-term system dynamics. The framework provides a policy-relevant analytical approach by linking AI-driven changes to sustainability challenges addressed in the Sustainable Development Goals (SDGs), while allowing examination of how these changes may simultaneously influence progress toward multiple SDGs and potentially generate cross-domain trade-offs.

AI impact groups are defined in this study to translate AI capabilities into system-level functions. Examining system-level consequences in the absence of these impact groups indicates that performance metrics alone are insufficient and that AI effects should be interpreted as system-level phenomena. The proposed approach extends existing work by linking AI capabilities to sustainability-relevant system behaviour rather than isolated outcomes.

The following input parameters are proposed: AI intensity, adoption level, autonomy, quality of use, and system quality. Quality of use and quality of system emerge as critical factors, as the blind application of a low-quality system may amplify negative effects. These parameters are also relevant for governance and policy design, as they help identify conditions under which AI applications support sustainable development goals or generate unintended systemic risks. Input factors collectively shape AI's impacts and should therefore be assessed together.

Sustainability is commonly structured around three core pillars: environmental, economic, and social dimensions. In addition to these pillars, the individual can be conceptualised as a fourth, cross-cutting pillar embedded within and interacting with all three dimensions. The proposed model extends the traditional three-pillar sustainability framework by incorporating the individual as a fourth pillar, resulting in a 3+1 framework. The model links AI capabilities to changes at the skill and parameter levels, enabling a systematic analysis of cross-pillar interactions across time and regions. It can provide insights into how AI may simultaneously generate short-term benefits and long-term risks. AI may accelerate the development of fusion power plants and quantum computers, leading to a new technological paradigm shift that feeds back into AI and reshapes its broader systemic impacts. These complex feedback loops and inherent uncertainties underscore the need for a structured, shared analytical framework. The goal is to develop a methodology and an initial model that supports collaboration among researchers and experts within specific sectors or sustainability pillars.

4 Future research directions

Future research will focus on further developing the proposed model by refining input factors and systematically analysing their interactions and cross-effects. Beyond structural refinement, particular emphasis will be placed on integrating a cognitive sustainability perspective, examining how knowledge generation, perception, and decision-making processes shape and are shaped by the model's inputs and outputs. This includes exploring what external and internal conditions influence input parameters, identifying additional latent factors, and integrating regional and temporal dimensions to capture spatial variation and long-term dynamics. The model is also intended to support forward-looking analysis and scenario-based assessment, with a specific focus on enhancing cognitive accessibility and usability for diverse stakeholders. Further work will also address the selection and development of appropriate computational tools that allow domain experts to interact

https://doi.org/10.55343/CogSust.22409

with the model, contribute assessments and collaboratively add use cases. Particular attention will be given to incorporating feedback loops and cross-pillar interactions. This may enable the model to represent how impacts propagate between sustainability pillars over time, while also reflecting how these dynamics are perceived, interpreted, and acted upon by decision-makers. In this context, the model aims to bridge analytical rigour with cognitive realism. Finally, future efforts will focus on developing effective visualisation approaches to support transparency, interpretability, and practical use of the model across research and policy contexts.

References

Antwi, B. O., Agyapong, D., & Owusu, D. (2026). Beyond adoption: Analysing the sufficient and necessary factors for AI-driven sustainable performance in an emerging economy. Sustainable Futures, 11, 101750. DOI: https://doi.org/10.1016/j.sftr.2026.101750

Browning, J., & LeCun, Y. (2023). Language, common sense, and the Winograd schema challenge. Artificial Intelligence, 325, 104031. DOI: https://doi.org/10.1016/j.artint.2023.104031

Brundtland Commission (1987). *Report of the World Commission on Environment and Development: Our Common Future*. United Nations. URL: http://www.un-documents.net/our-common-future.pdf

Doğan, A. R., & Doğan, A. İ. (2026). From the Turing Test to AI detectors: an epistemological mismatch in scholarly publishing. Brazilian Journal of Anesthesiology, 76(2), 844740. DOI: https://doi.org/10.1016/j.bjane.2026.844740

EIMT (2026). Who is the Father of AI and Machine Learning? *European Institute of Management & Technology*. URL: https://www.eimt.edu.eu/who-is-the-father-of-ai-and-machine-learning

EurLex (n. d.). Glossary of Summaries. Sustainable development. *Official website of the European Union*. URL: https://eur-lex.europa.eu/EN/legal-content/glossary/sustainable-development.html

Gohr, C., Rodríguez, G., Belomestnykh, S. et al. (2025). Artificial intelligence in sustainable development research. *Nature Sustainability*. 8, 970–978. DOI: https://doi.org/10.1038/s41893-025-01598-6

Katper, N., Hamid, A. B. B. A., Kaur, K., & Mohy-ud-Din, K. (2026). Dual pathways toward sustainability performance through AI technologies: Moderated mediation role of green HRM by fintech. Journal of Cleaner Production, 554, 148038. DOI: https://doi.org/10.1016/j.jclepro.2026.148038

Manning, P. C. (2020). Artificial Intelligence Definitions. Stanford University Human-Centered Artificial Intelligence. URL: https://hai-production.s3.amazonaws.com/files/2020-09/AI-Definitions-HAI.pdf

Matsuo, Y., LeCun, Y., Sahani, M., Precup, D., Silver, D., Sugiyama, M., ... & Morimoto, J. (2022). Deep learning, reinforcement learning, and world models. Neural Networks, 152, 267-275. DOI: https://doi.org/10.1016/j.neunet.2022.03.037

OECD (2025). Introducing the OECD AI Capability Indicators. OECD Publishing. DOI: https://doi.org/10.1787/be745f04-en

Phan, L., Gatti, A., Han, Z. et al. (2025a). Humanity's Last Exam. *arXiv*. DOI: https://doi.org/10.48550/arXiv.2501.14249

Phan, L., Gatti, A., Han, Z. et al. (2025b). Humanity's Last Exam Benchmark Leaderboard: Score vs. Release Date. *Center for AI Safety*. URL: https://artificialanalysis.ai/evaluations/humanitys-last-exam

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

Singla, A., Sukharevsky, A., Yee, L., Chui, M., Hall, B. (2025). *The state of AI in early 2024*. Quantum Black, McKinsey. URL: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2024

SU HAI (2025). The 2025 AI Index Report. *Stanford University, Human-Centered Artificial Intelligences (SU HAI).* URL: https://hai.stanford.edu/ai-index/2025-ai-index-report

Toderas, M. (2025). Artificial Intelligence for Sustainability: A Systematic Review and Critical Analysis of AI Applications, Challenges, and Future Directions. *Sustainability*. 17(17), 8049. DOI: https://doi.org/10.3390/su17178049

Velupillai, K. V., & Kao, Y. F. (2014). Computable and computational complexity theoretic bases for Herbert Simon's cognitive behavioral economics. Cognitive Systems Research, 29, 40-52. DOI: https://doi.org/10.1016/j.cogsys.2013.07.005

UN – United Nations (2015). Transforming our world: the 2030 Agenda for Sustainable Development. *UN Department of Economic and Social Affairs*. URL: https://sdgs.un.org/2030agenda

Xia, B. Lu, Q., Zhu, L., Xing, Z. (2024). An AI System Evaluation Framework for Advancing AI Safety: Terminology, Taxonomy, Lifecycle Mapping. *AIware 2024: Proceedings of the 1st ACM International Conference on AI-Powered Software*. 74–78. DOI: https://doi.org/10.1145/3664646.3664766