

Review

Assessing Impact Evaluation Methods: A Comparative Analysis and Proposal for an Integrated and Transformative Framework

Valentina Cattivelli ^{1,2,*}, Noemi Biancone ¹, Fernando Ferri ¹ and Ester Napoli ¹

¹ CNR—IRPPS, National Research Council—Institute for Research on Population and Social Policies, Via Palestro, 32, 00185 Roma, Italy; noemi.biancone@cnr.it (N.B.); fernando.ferri@cnr.it (F.F.); esternapoli@cnr.it (E.N.)

² Facoltà di Economia e di Giurisprudenza, Pegaso Digital University, Centro Direzionale, Isola F2, Via Giovanni Porzio, 4, 80143 Napoli, Italy

* Correspondence: valentina.cattivelli13@gmail.com

Abstract

The growing complexity of public policies and socio-ecological transitions demands evaluation tools that integrate environmental, economic, and social dimensions across multiple governance contexts. However, the main evaluation impact methods are fragmented and sector-specific, limiting their ability to capture cross-scale interdependencies and sustainability of trade-offs. While existing reviews address methods in isolation, no updated systematic framework is built upon joint epistemological and participatory dimensions, proposing an integrated and transformative framework. Through a common analytical framework—encompassing the type of impact assessed, territorial scale, methodological approach, degree of stakeholder participation, and operational complexity—the paper presents a systematic comparative review of eleven widely used evaluation methods, covering the environmental, economic, social, and multidimensional dimensions. These methods are selected through screening of peer-reviewed literature published between 2022 and 2025 using an AI-assisted text-mining protocol to support keyword extraction, methodological classification, and manual validation. Its focus is on the analysis and critique of their characteristics, complexity, and epistemological limitations, particularly focusing on measurability and participation. The comparative analysis reveals a structural epistemological gap between qualitative bottom-up perspectives and quantitative, top-down approaches, as well as structural fragmentation across evaluative dimensions. They also draw attention to the approaches' inadequate ability to manage non-linear dynamics, uncertainty, and transformative scenarios, which affects the legitimacy, comparability, and general efficacy of decision-making processes. Building on these, the paper advances a paradigmatic shift toward three systemic integrated principles: (i) systemic evaluation grounded in methodological and epistemological integration; (ii) transformative evaluation oriented toward collective learning and the redefinition of development trajectories; and (iii) co-produced evaluation based on the active engagement of territorial actors.

Keywords: impact assessment; AI; economic impact; social impact; environmental impact

Academic Editor: Cristina Raluca Gh. Popescu

Received: 11 May 2026

Revised: 22 June 2026

Accepted: 30 June 2026

Published: 29 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Contemporary societies deal with reflective transformations towards regenerative futures; they support well-being while addressing social and ecological crises. These transformations are challenging and depend on climate change, social and geopolitical crises, widening inequalities in wealth and health, as well as the rapid expansion of digitalisation and artificial intelligence. Addressing them requires urgent systemic, society-wide responses that address their structural roots (e.g., [1–3]).

National and international institutions have progressively institutionalised impact assessment as a conditionality for resource allocation, accountability and evidence-based decision-making. Allocating scarce efforts and resources—deciding which policies and projects to implement—is therefore inherently challenging, framing a complex and pressing evaluation problem in multiple domains, particularly the social, environmental, and economic spheres [4–6].

Recent research increasingly seeks methodological approaches to support scholars and public decision-makers in this assessment.

Despite longstanding efforts to develop more comprehensive evaluative frameworks, existing methods for assessing the impacts of policies and projects expose persistent structural limitations and call for critical reassessment in view of contemporary challenges (e.g., [7]). Many of these methods assume relatively stable conditions and tend to overlook emergent dynamics and systemic interactions. Consequently, they continue to be used in a fragmented and sectoral manner and rarely extend beyond single-sector or single-scale evaluations [8,9]. Based on these presumptions, these approaches usually concentrate on a single domain, provide a limited assessment of the phenomenon in question, and have a restricted capacity to capture the systemic interdependencies that characterise policy contexts [10,11]. Moreover, their reductionist analytical frameworks often prioritise measurable outcomes over contextual nuances, sources of bias, factors that are difficult to estimate, the potential long-term unintended effects, and the influence of stakeholder participation in the evaluative process [12,13].

This latter aspect, the reflexive evaluation paradigm—grounded in Beck’s theory [14]—takes the stance of supporting networks of actors to challenge existing systems. Otherwise, stakeholder engagement is increasingly recognised as crucial for addressing complex problems: it not only facilitates problem-solving, but it also enables a deeper exploration of the ethical and political dimensions that underpin the issues at stake [15]. However, this position also raises awareness of participation in relation to stakeholders’ position and power engagement [16]. It also creates a methodological divide between quantitative, top-down approaches (primarily designed to generate replicable and standardised measurements of discrete impacts) and more qualitative, bottom-up methodologies (oriented toward the co-production of contextualised knowledge and the recognition of multiple stakeholder perspectives).

Beyond these technical limitations, there are additional epistemological constraints. Prevalent evaluation methods rely on consolidating assumptions about what counts as evidence and legitimate knowledge. These assumptions, together with the need to identify measurable impacts, systematically privilege quantifiable and expert-validated dimensions over intangible, relational, and community-held forms of knowledge [17–19]. Cost-Benefit Analysis provides an example of this in the context of ecosystem services analysis. It reduces non-measurable values technically comparable, with the risk of underrating the distributional and ethical dimensions of the trade-offs involved, by reducing biodiversity or cultural landscape values to a single monetary metric [20]. Similarly, Life Cycle Assessment typically excludes governance and social dimensions, and its evaluations are methodologically robust but lack reference to aspects beyond the environmental dimension [20].

Considered technical defaults rather than consequential decisions, these foundational assumptions have inferences for the inclusivity and validity of evaluation outcomes. This epistemological opacity is not incidental but structural: prevalent evaluation methods assume what counts as evidence, whose knowledge is legitimate, and which impacts are considered measurable. These assumptions deeply privilege quantifiable and expert-validated dimensions over intangible, relational, and community-held forms of knowledge [20].

A further challenge concerns the scalar dimension of evaluation. Impact assessment methods differ substantially in the territorial and operational scales for which they are designed, ranging from product- or site-specific analyses to national and transnational assessments. Methods designed for a specific project or territorial grades may not yield equally strong outcomes when applied at different analytical levels, including those related to the policy programme or regional transition strategy. While certain methods are more effective for evaluations at the level of a single enterprise, municipality, or project, they are also ideal for assessing impacts across full production chains or national policy frameworks. The misalignment between the operational scale of evaluative methods and governance structures poses relevant risks for methodological application and may lead to analytical incompleteness.

These constraints become intensely crucial within transformative policy agendas (e.g., the European Green Deal or the Sustainable Development Goals) where interventions are expected to generate systemic, cross-sectoral, and long-term changes that cannot be effectively captured by methods originally designed for relatively static evaluative settings. Addressing these challenges requires incorporating instability, incompleteness, and unpredictability into the evaluative process. Consequently, methodological design must adequately address the tension between complexity and measurability, rather than relying solely on scientific rationality [16]. In practice, this implies integrating evaluation dimensions such as transformation fidelity (i.e., the consistency between transformative rhetoric and actual change), alongside the degree of alignment and momentum among actors and initiatives, and the extent to which ecosystem resilience is achieved [17].

Another challenge concerns the measurability dimension of impact assessment: methods that focus exclusively on observable, quantifiable outcomes systematically underestimate the generative mechanisms and structural conditions that explain why and for whom interventions work. In this sense, measurability is an epistemological choice, not a technical one: what is measured reflects assumptions about what is relevant and amenable to evaluation. Under these assumptions, the evaluation is reductive because it excludes intangible, relational, and cultural impacts that are difficult to estimate with precision.

Another tension occurs between measurability and participation. Methods that generate quantifiable results are prone to excluding the generative mechanisms and structural conditions that clarify why and for whom interventions generate effects. Based on this, measurability emerges as both a technical and epistemological condition: only what can be evaluated is considered relevant and worthy of evaluation [18,19]. Conversely, participatory and qualitative approaches, which are based on the reflexive evaluation paradigm [14] and concerned with the co-production of contextualised knowledge [15], are implemented when it is necessary to capture contextual complexity adequately, even if they are considered less replicable and harder to aggregate for policy comparison. While this epistemological dualism between quantitative top-down and qualitative bottom-up approaches is well-known and studied [10,11], no systematic comparative analysis has considered how it is revealed across the full range of sustainability and policy assessment methods currently adopted.

This fragmentation of methods has practical consequences. Decision-makers are forced to navigate among several methods that are technically sophisticated but epistemologically narrow—capable of measuring what is already measurable, but ill-equipped to capture the complexity, uncertainty, and transformative ambition that characterise contemporary challenges. Moreover, these methods work within the context of digital transformation and algorithmic governance, which risk exacerbating the opacity of underlying epistemological assumptions by embedding evaluative assumptions within automated decision-support systems [21]. At the same time, AI-assisted tools offer new opportunities for processing large volumes of information and data and, prospectively, for revising evaluation frameworks into more adaptive and integrative systems.

This paper examines the tension between the need for comprehensive evaluative frameworks and the persistence of reductionist approaches; it reconciles the demand for critical engagement with the practical constraints shaping policy and project evaluation. It also engages with both dimensions of this digital turn: on the one hand, as a methodological resource in the research design; and, on the other hand, as a prospective enabler of the structured framework.

To this end, the paper proposes a novel systemic and multidimensional analytical framework for evaluating impacts across multiple scales (project and territorial levels).

Several review studies exist, but they focus on specific methods or categories of impact assessment methods. The proliferation of impact assessment methods has, in fact, not been accompanied by a corresponding development of comparative frameworks assessing their relative strengths, limitations, and complementarities in a systematic and theoretically grounded manner. Considered individually, these studies do not provide systematic cross-method comparisons of their epistemological assumptions, participatory orientation, or capacity to manage complexity. Few have proposed a multi-method and integrated study. For example, Buckton et al. [6] have recently advanced twelve principles for transformation-focused evaluation. The present paper differs from it. Rather than deriving principles from evaluation theory, it grounds its framework in a structured bibliometric mapping and cross-method epistemological assessment of widely used methods, thereby identifying structural patterns—particularly the inverse relationship between measurability and participation—that emerge from the literature rather than from a priori normative commitments.

This paper builds on a large body of evaluation literature to identify principles and methods for transformation-focused evaluation. It aims to advance knowledge of the effectiveness of widely applied methods by addressing their structural fragmentation, reducing epistemological limitations, and contributing to a shift in the prevailing evaluative paradigm in practice. The framework draws on the work of several scholars who have developed critical heuristics for designing evaluation methods. It is structured around three distinct but interdependent principles that, when applied by evaluators, are considered to significantly enhance the effectiveness of evaluation in assessing and supporting transformative processes. Despite these positive aspects, tensions may arise between some of these rules. The principles concern: (1) the essential characteristics of evaluation methods (including the types of impacts assessed, territorial scale, project scale, and operational tools and instruments); (2) complexity; and (3) the epistemological characterisation of evaluation, particularly in terms of measurability and participation.

Building on this, the paper advances a set of methodological orientations and practical recommendations to support the integration of existing methods into a systemic, participatory, and transformative evaluation framework.

Against this, the research question is:

“What are the structural and epistemological limitations of the main methods used to evaluate the impacts of policies and projects, and how can these methods be integrated within a systemic, transformative, and participatory framework?”

To address this question, the paper makes three interrelated contributions. First, it provides a systematic comparative review of eleven sustainability and policy impact assessment methods—selected through a bibliometric screening of peer-reviewed literature published between 2022 and 2025. This analysis examines each method through a common analytical framework, covers impact type, territorial scale, methodological approach, stakeholder participation, and operational complexity. Second, it recognises an inverse structural relationship between measurability and participatory depth as a cross-cutting pattern across the methods assessed. Third, it suggests a paradigmatic shift toward systemic, transformative, and co-produced evaluation and examines how AI-enabled tools may support its operationalisation.

The paper is structured as follows. Section 2 describes the research design and the multi-step methodology for the systematic identification, mapping, and comparative assessment of the selected evaluation methods. Section 3 presents the results of the comparative study, including a structured overview of operational characteristics and a cross-method assessment of complexity, participation, and epistemological assumptions. Section 4 discusses the structural and epistemological patterns identified across the methods and develops the proposed integrative framework. Section 5 concludes and identifies directions for future research, including the importance of AI in the process.

2. Methods

As outlined in the introduction, this research pursues three interrelated objectives:

- (i) to demonstrate the structural fragmentation in the existing evaluation frameworks for sustainability and policy impact assessment methods;
- (ii) to critically examine their underlying epistemological limitations, with reference to measurability and stakeholder participation;
- (iii) and to advance a set of theoretically grounded and practically oriented proposals to overcome such limitations and reconfigure the prevailing evaluation paradigm.

The study adopts a scoping review design [22] and the PRISMA extension for Scoping Reviews (PRISMA-ScR; [23]). This design enables the systematic mapping of the extent and nature of a body of literature without requiring full-text quality appraisal of individual records. It is appropriate when the objective is to identify prevailing patterns, gaps, and conceptual boundaries across a heterogeneous corpus, rather than to synthesise evidence concerning specific interventions or outcomes [24]. This study comprises four sequential steps: (1) a systematic literature screening across academic and grey literature sources; (2) an AI-assisted exploratory mapping of the resulting corpus; (3) the selection of methods for structured comparative assessment; and (4) the comparative assessment. Each step is described in detail below.

2.1. The Systematic Literature Screening

The identification of relevant records was conducted across two complementary sources. This dual-source approach aligns with the PRISMA-ScR recommendations to extend searches beyond indexed academic databases to include the full evidence pertinent to the research question [23].

The primary search consisted of a literature search conducted in Scopus. The search string (“impact assessment methods” OR “multi-criteria decision analysis”) was applied to titles, abstracts, and keywords and deliberately kept broad to capture methodological diversity across evaluation domains, rather than restricting the results to one or two disciplinary

fields of analysis. This decision aligns with the exploratory and mapping objectives of scoping reviews. Accordingly, it also employed the Boolean operator OR to maximise recall and therefore aligned with the empirical nature of the present mapping exercise.

The temporal range was limited to the 2022–2025 period to include the most recent methodological developments and current evaluation practices. This procedure yielded 1272 peer-reviewed indexed articles. Evidence from subject areas unrelated to policy-relevant evaluation—particularly engineering applications, medical and clinical sciences, physics, and chemistry—was excluded. The rationale for this exclusion is that such documents do not address social, environmental, or economic impact assessment methods, which are the focus of this review. This exclusion reduced the Scopus corpus to 1252 documents.

The second search was conducted through Google Scholar and Publish or Perish software v.8 [25] to also include institutional and grey literature. This latter literature is relevant in applied evaluation practice, where methods are documented outside peer-reviewed channels. The same search string and temporal range were applied. This secondary extraction yielded 995 documents, of which 976 met the disciplinary inclusion criteria. After integrating all results, duplicate records were removed ($n = 338$). The final dataset comprised 1890 documents.

The complete screening procedure is represented in Figure 1, which was created using Claude Sonnet 5.

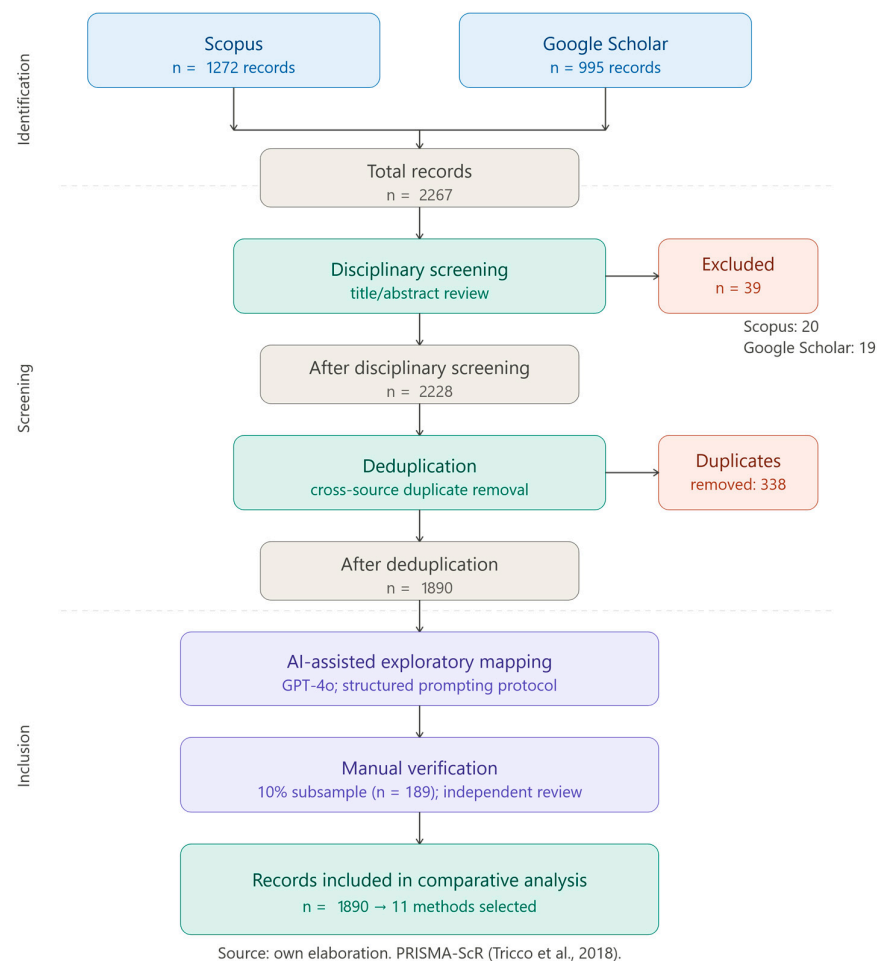


Figure 1. PRISMA ScR flow diagram—literature search and selection. Source: authors’ own elaboration, 2025. Flow diagram generated with an AI-assisted visualisation tool (Claude). Dashed lines represent the different phases of the PRISMA-ScR Methodology [23]. Inclusion criteria: peer-re-

viewed papers and institutional or grey literature documents that address impact assessment, policy evaluation, or sustainability assessment methods; published between 2022 and 2025 in English or Italian. Exclusion criteria: papers and grey literature that address clinical or medical evaluation, engineering, physical or chemical measurement methods; not available in English or Italian; duplicate records.

The complementary search, which targeted the intersection of AI tools and impact assessment methods, was conducted in parallel by a subset of the research team and documented separately in the text.

2.2. AI-Assisted Exploratory Mapping

The second step consisted of AI-assisted exploratory mapping of 1890 documents, aiming to identify recurring methodological patterns, dominant evaluative methods, and their frequency of occurrence across the literature. All abstracts were compiled into a single plain-text file and analysed using ChatGPT (GPT-4o, OpenAI, accessed March–April 2025) as an AI-assisted text-mining instrument. This step introduced a standardised prompting protocol, applied across all sessions to minimise variability in model outputs. The primary prompt instructed the model to retrieve all impact assessment or policy evaluation methods mentioned across the considered abstracts and to quantify their frequency. Its precise wording was as follows: “*You are a systematic review assistant. Read the following compiled abstracts and identify all distinct impact assessment or policy evaluation methods mentioned. For each method, provide: (1) the method name as it appears in the text; (2) a standardised label; (3) the number of abstracts in which it appears. Return results as a structured table.*” Successive prompts instructed the model to group identified methods by evaluative domain (environmental, economic, social, or integrated), to flag methodological expressions associated with participatory or quantitative approaches, and to identify methods with high institutional relevance but limited bibliometric visibility within the indexed literature.

This step also included a consistency check of the AI-assisted classification. The authors randomly selected a 10% subsample of the corpus ($n = 189$ documents). On two separate occasions, the authors reviewed this subsample utilising the same classification categories applied in the AI-assisted analysis (i.e., method identification, evaluative domain assignment, and participatory orientation). Inconsistencies between the two reviews were resolved through collective discussion, full-text document analysis, and consideration of the method most consistent with the document’s primary orientation. This verification served primarily qualitative purposes, helping to identify problematic categories and boundary cases. Divergences were predominantly concentrated in documents addressing hybrid or multi-method approaches. This concentration in a structurally defined document type indirectly evidence that the AI-assisted classifications were broadly consistent with manual coding for most single-method documents. The AI tool was employed solely as an exploratory and classificatory instrument, and interpretations are attributed solely to the authors.

The AI-assisted mapping identified recurring methodological patterns within the corpus, with Multi-Criteria Analysis and its variants emerging as overwhelmingly dominant, followed by LCA and EIA.

Table 1 presents the complete frequency distribution of evaluation methods identified across the corpus. The sum of method frequencies ($n = 2560$) exceeds the total corpus size ($n = 1892$) because documents may report more than one evaluation method; each document was counted once per method identified within its abstract. Methods with zero frequency in the indexed corpus—specifically SROI, GIA, and IO—were nonetheless in-

cluded in the comparative assessment based on institutional relevance, as described in Section 2.3.

Table 1. Frequency distribution of evaluation methods identified through AI-assisted mapping. Source: own elaboration, based on 2025/2026 research.

Rank	Evaluation Method	N Articles	% of Corpus (1890)
1	MCA/MCDA (Multi-Criteria Analysis)	1517	80.3%
2	GIS-based Assessment	369	19.5%
3	AHP (Analytic Hierarchy Process)	322	17.0%
4	LCA (Life Cycle Assessment)	222	11.7%
5	Fuzzy Logic/Fuzzy AHP	162	8.6%
6	EIA (Environmental Impact Assessment)	61	3.2%
7	CBA (Cost–Benefit Analysis)	38	2.0%
8	CEA (Cost-Effectiveness Analysis)	28	1.5%
9	Carbon Footprint Assessment	24	1.3%
10	DEA (Data Envelopment Analysis)	23	1.2%
11	SIA (Social Impact Assessment)	9	0.5%
12	HIA (Health Impact Assessment)	8	0.4%
13	SWOT Analysis	4	0.2%
14	DLCA (Dynamic Life Cycle Assessment)	2	0.1%
15	IO (Input–Output Analysis)	2	0.1%
16	RIA (Regulatory Impact Assessment)	1	0.1%
17	SROI (Social Return on Investment)	0	0.0%
18	GIA (Gender Impact Assessment)	0	0.0%

Note: The sum of method frequencies ($n = 2792$) exceeds the total corpus size ($n = 1890$) because documents may mention more than one method; each document was counted once per method identified within the abstract. Percentages refer to the proportion of documents in which each method appears at least once. GIS-based Assessment and AHP are operational tools rather than autonomous impact assessment methods and were therefore not included among the eleven selected methods. DEA, RIA, and SWOT were excluded on grounds of disciplinary scope or methodological specificity.

However, the selection of methods for the next comparative assessment was not decided solely by bibliometric frequency. A purely frequency-based selection would risk reproducing the existing blind spots of indexed scientific literature, which systematically underrepresents methods predominantly deployed in institutional and applied evaluation contexts—such as SROI, GIA, CEA, and Input–Output Analysis—despite their widespread use in international policy frameworks (e.g., European Commission, UNDP, World Bank).

Additionally, a co-occurrence analysis evaluated whether methods are typically implemented in combination within the same study the research. The results of this analysis show that multi-method combinations are uncommon. The most recurrent pairing, MCA and LCA, co-occurs in only 46 documents (i.e., 2.4% of the total). Other notable pairings, such as MCA and CBA or LCA and CBA, occur in fewer than 2% of the considered corpus. All remaining pairings appear in fewer than 10 documents each. This pattern of predominantly single-method applications provides direct empirical support for the structural fragmentation identified in Section 4.1. Despite the availability of complementary methods, evaluators overwhelmingly apply them in isolation rather than in systematic combination.

The final selection of eleven methods was therefore guided by four complementary criteria: bibliometric recurrence within the corpus; disciplinary coverage, ensuring repre-

sensation across environmental, economic, social, and integrated evaluation domains; co-occurrence analysis and institutional relevance, capturing methods that are central to applied evaluation practice even where their scientific visibility remains limited.

Structured in this way, the AI-assistant provides exploratory and classificatory support, enabling the detection of dominant methodological patterns within the corpus while preserving independent qualitative analysis. However, it also generates potential biases, such as sensitivity to prompt formulation and the risk of overlooking low-frequency items. These are mitigated through three procedures: standardisation of the prompting protocol, cross-checking the identification of low-frequency methods and a manual validation procedure, as described above. At the same time, to reinforce this mitigation, the AI tool was employed only as a classificatory instrument, not as an evaluative tool.

In parallel with the AI-assisted exploratory mapping described above, the AI assistant was also used to analyse the intersection between AI tools and the impact assessment methods in a subset of identified literature. The corresponding search used the string (“impact assessment” AND “artificial intelligence”) to titles, abstracts, and keywords and retrieved 81 documents, which were subjected to a structured coding procedure conducted independently by two members of the research team. These researchers adopted a binary criterion to classify the documents using the recognition of an impact assessment methodology recognition and AI tools as an operational component of the evaluation process. Documents in which AI was employed exclusively for purposes such as bibliometric analysis, data extraction from administrative sources, sentiment analysis of social media content, or conceptual and regulatory review were classified as not using an AI-assisted impact assessment methodology and were excluded. Conversely, documents in which the evaluation method integrated AI tools (e.g., machine learning models, generative AI, LLMs) into one or more stages of the substantive analysis were considered. Disagreements among authors were resolved through discussion for this analysis as well.

2.3. Selection of Methods for Comparative Assessment

The AI-assisted mapping identified recurring methods in the literature. Multi-Criteria Analysis and its variants emerged as the most frequently mentioned approaches, followed by Life Cycle Assessment and Environmental Impact Assessment.

However, the selection of methods to be considered for the comparative assessment was not based solely on a purely frequency-based criterion. Relying only this criterion risks reproducing the existing blind spots of the indexed literature. Such literature systematically underrepresents methods mostly used in applied evaluation contexts and used within international policy frameworks (such as Social Return on Investment, Gender Impact Assessment, Cost-Effectiveness Analysis, and Input–Output Analysis).

Three criteria therefore guided the final choice of eleven methods. The first criterion was bibliometric recurrence. Thus, methods were included in the list if they appeared in at least 10 documents within the selected literature, as recognised through AI-assisted mapping and confirmed by manual verification. The second criterion, i.e., disciplinary coverage, ensured balanced representation across all four evaluative domains (environmental, economic, social, and integrated or multidimensional) to avoid bias toward any domain. The last criterion, institutional relevance, included methods that are central to applied evaluation practice in major international policy frameworks, even if their prominence in indexed literature is limited.

The eleven methods selected for structured comparative assessment are: Life Cycle Assessment (LCA), carbon footprint (CF), Dynamic Life Cycle Assessment (DLCA), Cost–Benefit Analysis (CBA), Cost-Effectiveness Analysis (CEA), Input–Output Analysis (IO), Social Return on Investment (SROI), Gender Impact Assessment (GIA), Health Impact Assessment (HIA), Multi-Criteria Analysis (MCA), and Social Impact Assessment (SIA).

Methods principally associated with causal inference practice (e.g., Randomised Controlled Trials, quasi-experimental designs, Contribution Analysis, and Realist Evaluation) were not considered and therefore excluded. These methods are structured to calculate causal attribution between an intervention and its outcomes, rather than to characterise the nature, distribution, and epistemological dimensions of impacts across domains. Their exclusion reflects a substantive difference in analytical purpose, not a judgment on their appropriateness or merit. Other methods (i.e., Strategic Environmental Assessment and Environmental Impact Assessment) were excluded because they consist of procedural-regulatory instruments rooted in statutory planning frameworks, rather than analytical methods applicable across policy and project settings in the manner adopted by the present study.

2.4. Structured Comparative Assessment

The last step includes the structured comparative assessment of the eleven selected methods.

Despite the proliferation of impact assessment methods across policy spheres, the current literature does not offer a comparative framework to analyse their structural characteristics across dimensions simultaneously. Existing studies compare methods along a single dimension at a time and do not integrate them into a coherent structure. To move beyond these fragmented and domain-specific assessments, the research develops a structured analytical grid with two dimensions: the operational characterisation and the epistemological assessment.

The first dimension examines:

- Type of impact assessed (environmental, economic, social, or integrated).
- Territorial scale of application, i.e., the geographical extent at which methods are most often applied and are most analytically effective. This dimension helps to assess the level of detail or generalisation characterising the evaluation process.
- Project scale of application, i.e., the operational dimension of the projects or institutions employing the method. This may range from micro-level local interventions to large-scale and complex programmes.
- Operational tools and instruments, i.e., qualitative, quantitative or mixed-method approaches were classified according to whether they were more consistent with the research design and the features of the studies. To increase reproducibility and transparency of the evaluation process, the research also documents the software used.

Differences among methods cannot be fully explained by their operational and technical features alone. Methods that share similar territorial or project scales, or that employ comparable tools, may nonetheless produce radically different evaluative outcomes depending on the assumptions embedded in their design—about what counts as evidence, whose knowledge is legitimate, and which impacts are deemed measurable (Table 2). This observation motivates another dimension of analysis, i.e., the epistemological assessment.

This assessment examines the privileging of certain dimensions over others and the consequences for the validity of the evaluation outcomes, and it is operationalised on a three-point ordinal scale (low, medium, high). While the grid identifies what each method measures and how, the epistemological analysis extends the assessment of the methods by considering three interrelated dimensions for each method: complexity, measurability, and stakeholder participation.

Table 2. The structured analytical grid for the comparable assessment. Source: authors' own elaboration, 2025–2026.

Method	Type of Impact	Territorial Scale	Project Scale	Operational Tools and Instruments
LCA	Environmental	Local/Regional/Global	Functional unit (product, process, or system level)	SimaPro, OpenLCA, GaBi; Ecoinvent database; ISO 14040/44 standards [26]; impact assessment methods
CF	Environmental/Climate	Global (climate impact boundary)	Functional unit (product/service) or organisational boundary	GHG Protocol, ISO 14064/14067 [26]; GaBi, SimaPro with GWP100 impact category; life cycle inventory databases
DLCA	Environmental (dynamic)	Local/Regional/Global	Functional unit (time-dependent); long-horizon infrastructure and energy transition projects	Dynamic LCA software; time-series models (regressions, stochastic models); evolution equations; decarbonisation modelling tools integrated with standard LCA software
CBA	Economic	Local/Regional/National	Project or policy level (monetised comparison of alternatives)	Economic valuation guidelines; financial modelling tools; willingness-to-pay estimates; sensitivity analysis; hybrid frameworks integrating LCA and MCA
CEA	Economic	Local/Regional/National	Product, project, or policy level (least-cost pathway to a defined objective)	Cost-effectiveness ratios; GIS-based spatial tools; dynamic MCA models; standard statistical software (R, Stata, SPSS)
IO	Economic/Environmental (sector-based)	National/Regional (economy-wide)	Sectoral/economy-wide (monetary demand-driven)	IO tables (WIOD, OECD TiVA); IO software (PyIO, EXIOBASE); econometric software (Stata, R, MATLAB); Environmentally Extended IO models (EEIO)
SROI	Social (monetised social value)	Local/Community/Regional	Project or programme level (stakeholder-based monetisation)	Shadow prices; well-being valuations; interviews; focus groups; stakeholder mapping; statistical software; Social Value International Framework
SIA	Social (qualitative and quantitative)	Local/Regional	Project or policy level	Interviews; focus groups; direct observation; stakeholder mapping; statistical indicators; social LCA databases (PSILCA); NVivo, MAXQDA, SPSS
GIA	Social (gender equality and distributional impacts)	Local/Regional/National/Supranational	Policy or programme level	Gender-disaggregated statistical indicators; interviews; focus groups; participatory observation; EU Gender Mainstreaming Toolkit; UNDP Gender Equality Marker
HIA	Health (population health effects)	Local/Regional/National	Project, plan, or policy level	Dose-response functions; burden of disease methods (DALYs); epidemiological data; GIS mapping; stakeholder consultation; WHO HIA framework
MCA	Multidimensional (environmental, economic, social)	Local/Regional/National/Supranational	Alternative/scenario comparison framework (micro to macro-scale)	AHP, Fuzzy AHP, ELECTRE, PROMETHEE, TOPSIS, VIKOR; entropy weight method; GIS integration; decision-support software; machine learning and AI-assisted tools

Complexity refers to the method's capacity to handle non-linear dynamics, indirect or emergent effects, and transformative scenarios. Methods that are rigidly top-down or overly standardised fail to capture real-world complexity. *Complexity* is assessed based on each method's capacity to handle non-linear dynamics, indirect or emergent effects, and transformative scenarios. The level of complexity reflects both the analytical architecture of the method and the extent to which it accommodates multidimensionality and uncer-

tainty. A low complexity score indicates methods that rely on predefined quantitative indicators and readily available data. Analytical procedures are standardised and replicable. Evaluator subjectivity is limited. While operationally efficient, such approaches may privilege measurable variables and exclude context-specific or intangible dimensions. Conversely, a high complexity score is assigned to methods that involve complex design, integration of qualitative and quantitative data, potential monetisation of diverse impacts, and multidisciplinary expertise. Evaluator judgement plays a significant role. While analytically comprehensive, such approaches may introduce opacity, elevated transaction costs, and barriers to replication.

The dimension of measurability and value assumptions (i.e., the impacts that are structurally excluded and the normative choices embedded in methodological design) examines which dimensions each method can readily quantify (e.g., greenhouse gas emissions or financial costs) and which it structurally excludes or underestimates, such as well-being, cultural impacts, or relational dynamics. This dimension serves to identify forms of analytical reductionism and reflects implicit assumptions about what should be evaluated as important or amenable to quantification.

Stakeholder participation is treated as an epistemological variable rather than a merely operational characteristic; it captures how evaluative knowledge is produced, validated, and legitimised within each method. As such, it is defined as the level of active involvement of relevant stakeholders (e.g., public decision-makers, practitioners, beneficiaries) in impact assessment processes and the way evaluative knowledge is produced, validated, and legitimised within each method. Participation is evaluated at three levels (low, medium, or high). In this evaluation, research is assessed for its level of abstraction (top-down) or contextualisation (bottom-up) at their extremes, as well as the use of subjective and/or objective data and standardised versus interpretative information. A low participation score is attributed to methods where knowledge production is expert-driven and centralised. Stakeholders are passive recipients of evaluation outcomes as they do not participate in data collection or evaluation definition processes. The epistemic model is mainly technocratic. Stakeholders provide consultative input to indicator definition or to the validation of findings, while methodological design remains mainly under expert control. A high score designates methods in which evaluative knowledge is co-produced: stakeholders actively participate in defining outcomes, determining methodological design, collecting data, and interpreting results. Here, evaluative authority is thereby allocated, and plural epistemic perspectives are accepted.

Each author assigned the scores of Table 3 based on a structured reading of the method-specific literature identified within the corpus. They are then compared, and disagreements were discussed iteratively until consensus was reached. Where consensus could not be reached, the more conservative score was adopted. It is acknowledged that this scoring procedure presents an element of interpretive judgment. The subsequent ratings are hence presented as analytically grounded propositions derived from the comparative framework, rather than as objective or definitive measurements.

Examples reported in the brackets are drawn from the reviewed corpus and are intended as illustrative rather than exhaustive.

Table 3. (a) Operationalisation criteria for epistemological assessment dimensions. Source: authors'; own elaboration, 2025. (b) The level of complexity, measurability, and participation of the methods considered. Source: authors' own elaboration, 2025–2026.

(a)			
Level	Complexity	Measurability	Participation
Low	Relies on predefined quantitative indicators and standardised	Predominantly intangible, context-dependent, or multi-	Expert-driven and centralised knowledge production. Stake-

	procedures. Limited evaluator subjectivity. Excludes context-specific or intangible dimensions.	dimensional impacts. Proxy valuations or qualitative evidence are required. Limited cross-context comparability.	holders are passive recipients of evaluation outcomes. Mainly a technocratic epistemic model.
Medium	Integrates quantitative and semi-quantitative data. Some evaluator discretion is required in scenario modelling or indirect effect estimation.	Quantifiable dimensions prevail, but indirect or intangible impacts are estimated through proxies. Methodological choices introduce value-laden assumptions.	Hybrid but asymmetrical epistemic model. Stakeholders contribute consultative input to indicator definition or findings validation. Methodological design remains largely expert-controlled.
High	Integrates qualitative and quantitative data. Significant evaluator judgement. Capacity to handle non-linear dynamics, uncertainty, and transformative scenarios. High transaction costs.	Standardised physical or monetary data with internationally codified protocols. Social, cultural, and relational dimensions are structurally excluded from the analytical perimeter.	Co-produced evaluative knowledge. Stakeholders actively participate in defining outcomes, shaping methodological design, collecting data, and interpreting results. Evaluative authority is distributed.
(b)			
Method	Complexity	Measurability	Participation
LCA	High	High	Low
CF	Medium	High	Low
DLCA	High	Medium	Low
CBA	Medium	Medium	Medium
CEA	Medium	Medium	Medium
IO	High	Medium	Low
SROI	Medium	Low	High
SIA	Medium	Low	High
GIA	Medium	Low	High
HIA	Medium	Medium	Medium
MCA	High	Variable	Variable

3. Results

3.1. General Characteristics

The selected methods are LCA, CF, DLCA, CBA, CEA, Input–Output Analysis, SROI, SIA, GIA, HIA, and MCA. Their characterisation is provided below, while their operational dimensions are detailed in Section 3.2.

Life Cycle Assessment (LCA) is an internationally standardised method used to evaluate the environmental impacts associated with a product, service, or process throughout the entire life cycle [27,28]. Defined by the ISO 14040: 2006 standard [26], LCA enables a systematic examination of resource consumption and emissions and assesses both direct and indirect impacts [29–31]. Its implementation is structured into four main phases: (i) Goal and scope definition, (ii) Life cycle inventory (LCI) analysis, (iii) Life cycle impact assessment (LCIA), and (iv) Life cycle interpretation.

Carbon footprint (CF) is a method used to quantify the greenhouse gas emissions generated by a product, service, or activity (e.g., [32]). It aggregates the impacts of different climate-altering gases into a single metric, thereby providing an estimate of the amount emitted into the atmosphere attributable to individual behaviours and production choices (e.g., [33]).

Dynamic Life Cycle Assessment (DLCA) represents a methodological evolution of conventional LCA, explicitly incorporating temporal variability into the analytical frame-

work [34]. Static LCA assumes a fixed system configuration throughout the life cycle. DLCA addresses this limitation by integrating time-dependent dynamics, prospective scenarios, and evolving characterisation factors into life cycle modelling (e.g., [35,36]). Its implementation requires addressing three critical challenges: defining temporal boundaries, using system-specific data, and determining the temporal resolution of flows and characterisation factors [37–39].

Cost–Benefit Analysis (CBA) systematically compares the total expected costs and benefits of a project or policy by assigning monetary values to all relevant impacts (including indirect and intangible) [40–43]. A project is considered economically justified when its monetised benefits exceed its costs. This method is articulated into four main stages: (i) defining the analytical framework; (ii) identifying and estimating direct, indirect, intangible, and opportunity costs, together with direct and indirect benefits; (iii) assigning monetary values to all relevant costs and benefits (including those that are difficult to determine); and (iv) aggregating and comparing total costs and benefits to determine the net project value. Direct financial elements are quite straightforward to estimate. By contrast, the valuation of indirect, intangible, opportunity costs and benefits often requires advanced estimation techniques (i.e., parametric modelling or expert-based approaches).

Cost–Effectiveness Analysis (CEA) compares the costs and effectiveness of two or more competing interventions to support informed decision-making (e.g., [44,45]). Rather than monetising benefits, CEA relates costs to specific outcome measures, such as cost per life saved or cost per health outcome achieved. For each alternative, net costs are calculated and compared to the desired results. CEA is particularly appropriate when the monetisation of benefits is impractical or ethically contested and when programme objectives are clearly defined and measurable. Compared with CBA, CEA is less demanding as it does not require full monetisation of benefits.

Input–Output Analysis (IO) assesses interdependencies among economic sectors and quantifies the propagation of demand-side changes across an entire economic system (e.g., [46–48]). Operating predominantly at the macroeconomic scale, it enables estimation of direct, indirect, and induced effects of policy or investment decisions along supply chains. It is based on the premise that changes in demand propagate across sectors, with the magnitude of effects determined by intersectoral interconnections. Social Return on Investment (SROI) estimates the economic, social, and environmental value generated by a project or organisation by assigning monetised proxy values to social outcomes identified through stakeholder engagement [49]. It is guided by seven core principles set up by Social Value International, encompassing stakeholder involvement, theory of change, outcome valuation, and transparency [50]. Two types of SROI can be applied: evaluative, applied ex-post on the outcomes, and predictive, which estimates the potential social value and its alignment with expected outcomes. Its implementation follows six interrelated stages, from scope definition and stakeholder identification to reporting and validation of results [51].

Gender Impact Assessment (GIA) evaluates the differential effects of policies, programmes, and projects on gender equality (e.g., [52,53]). The European Commission defines it as “the process of comparing and assessing, according to gender-relevant criteria, the current situation and trends with the expected developments resulting from the proposed policy.” Its aim has increasingly extended beyond legislative scrutiny to development programmes, budgets, and research calls, as emphasized by the Council of Europe [54,55].

Health Impact Assessment (HIA) systematically evaluates the effects of policies, programmes, and projects on population health and the distribution of those effects across social groups [56]. According to the IAIA, it is “a process that systematically assesses the potential, sometimes unintended, effects of a project, program, plan, policy, or strategy on

population health” [57]. It combines epidemiological evidence, stakeholder consultation, and equity analysis, while its implementation requires several steps: screening, scoping, assessment, recommendations, reporting, and monitoring and evaluation.

Multi-Criteria Analysis (MCA) supports decision-making among competing alternatives by integrating several potentially incommensurable criteria covering environmental, economic, and social dimensions. Because it is flexible in weighting criteria and handling uncertainty through fuzzy logic and sensitivity analysis, it considers different stakeholder perspectives and enables the integration of machine learning and AI-assisted tools for scenario analysis and appropriateness mapping.

3.2. Comparative Assessment of the Selected Methods

This structured analytical table ensures a systematic and comparable assessment of the methods considered (Table 2).

From this comparison, a scalar divergence emerges. Environmental methods operate through a functional logic unit. Their territorial scale is subordinated to the system boundary delineated by the scope of the investigation, and this dependence enables their application from the local to the global scale. For example, LCA has been implemented at the local scale to evaluate quarry operations in Istanbul and additive manufacturing facilities, at the national scale to assess household food waste management in Finland, and at the global scale to examine biodiesel supply chains and bio-energy climate impacts. DLCA demonstrates additional scalar flexibility considering that it incorporates temporal dynamics. Accordingly, it is implemented for long-horizon assessments such as decarbonisation trajectories and evolving electricity grid compositions. Economic methods are mostly calibrated to the scale of the policy or investment decision. For example, CBA has been implemented at the local scale to sustain water reuse decision-making and off-grid hybrid energy systems in rural areas, while at the national scale to evaluate marine plastic waste management alternatives. The scalability of CEA depends on the availability of data and the comparability of outcomes across contexts. As such, it cannot automatically translate from one scale to another, and its implementation is most consistent when programme objectives are defined and measurable, as in the allocation of carbon emissions allowance in provincial areas of China. IO analysis is calibrated to the macroeconomic scale and traces monetary flows among economic sectors characterising a national or regional economy. Some examples include studies of the OECD, World Bank, and the European Commission for macro-level policy evaluation, input–output environmental extensions, and carbon footprint accounting at the national scale.

Social methods are largely implemented at the local scale, where stakeholder identification and the traceability of social impacts to specific beneficiary groups are most analytically feasible. For example, SROI is more effective at the local level, where social value creation can be attributed to well-identified beneficiary groups. Similarly, SIA has been more effective at the local scale, when assessing distributional social impacts in aquaculture systems in coastal Spain and in circular economy initiatives in agriculture, for instance. GIA functions across multiple scales concurrently, although its territorial delimitation is particularly challenging because gender inequalities are spatially non-homogeneous and depend on the characteristics of local labour markets. HIA is more effective when applied at the urban scale in the reviewed literature (e.g., when implemented to quantify the health burden of ozone pollution across Beijing-Tianjin-Hebei cities or the health co-benefits of transportation policies across 14 US cities). MCA territorial scale is determined by the decision context rather than by a structural analytical constraint, and it has been applied from urban-scale site selection—such as groundwater potential mapping in Nigeria and aquifer recharge planning in Ghana—to national policy design, in-

cluding fossil-to-alternative fuel transition strategies in Germany and provincial carbon allocation in China.

Another observation concerns the divide between methods that rely on structured protocols and procedures, e.g., LCA and IO, and those whose implementation depends predominantly on primary data collection through qualitative instruments, e.g., GIA and SIA. This divide underlines differences in operational structure and suggests a deeper epistemological distinction about the foundations of legitimate evaluative knowledge.

3.3. *The Capacity to Handle Complexity*

The assessment of the methods' capacity to handle complexity reveals significant differences.

Environmental methods show different complexity profiles. LCA complexity is predominantly technical, as it is structured around predefined quantitative indicators and internationally standardised procedures. This structure imposes subjectivity and transparency in selecting methodologies and suggests reflection on the limited capacity to accommodate emergent dynamics, context-specific social dimensions, or non-linear systemic interactions. It also requires considering the underlying motivations for the evaluation and the characteristics and limits of the decision-making context in which outcomes will be used.

As a derivative of LCA, CF reveals similar limits and, in particular, its dependence on predefined quantitative indicators that potentially can reduce multidimensional environmental complexity to a single metric, rendering it operationally efficient but epistemologically narrow.

DLCA includes temporal modelling, stochastic simulations, and decarbonisation pathways, mitigating the conventional LCA limits. As it explicitly integrates non-linear processes, indirect effects, and scenario-based projections, it is more adaptable to complex and evolving systems. Although it is still structurally expert-driven, its complexity is high: temporal dynamics are modelled top-down, without integration of bottom-up participatory approaches. The result is a method of considerable epistemic openness but very limited replicability and high transaction costs.

Different levels of complexity characterise CEA and CBA. While CEA is reasonably simpler, CBA is more elaborate. Both combine quantitative data to assess costs, outcomes, and economic interdependencies and produce insights into indirect effects, especially through scenario modelling and multipliers. These methods are largely efficient in assessing clearly quantifiable variables, such as costs, monetary benefits, or sectoral outputs. However, they have some limits when assessing intangible outcomes, social well-being, or cultural effects. Especially for CBA, difficulties are in estimating intangible costs, and advanced estimation techniques are required. Their resolution introduces relevant evaluator subjectivity.

While CEA and CBA can capture some emergent effects in policy or project evaluation, IO analysis allows economy-wide impact estimation across sectors with more structural rigidity. Since it needs IO tables and econometric expertise, this method is inherently top-down and offers limited adaptability to local or project-level contexts.

Social methods are characterised by high complexity as they integrate qualitative and quantitative data, require the monetisation or interpretation of intangible outcomes, and address context-specific social, gender, and health impacts. Combining different benefits is a good decision to capture non-linear and cumulative effects, as well as promising social phenomena and an expansive perception of value. However, they address operational difficulties, advanced transaction costs, and greater evaluator subjectivity. Estimation is difficult. While some quantitative proxies exist (e.g., financial valuation in SROI, statistical health, or social indicators), many other outcomes (especially social, cultural, or gendered

outcomes) are difficult to measure. In particular, SROI relies on linear mappings of inputs, outputs, and outcomes implementing the theory of change. In this way, it does not completely capture dynamic feedback or non-linear interactions. Its predictive (ex-ante) application increases uncertainty, as projected outcomes may diverge from final impacts due to context variability. SROI includes a layer of stakeholder co-construction, which raises contextual sensitivity and limits replicability.

SIA mixes qualitative and quantitative tools in extremely context-specific targets, while HIA combines epidemiological data, dose-response functions, and stakeholder consultation within a structured and analytically discretionary system. In recent applications, HIA complexity has increased further through the incorporation of machine learning models and GIS-based exposure mapping. GIA faces analogous challenges, as the intersectionality of gender dynamics resists reduction to standardised analytical procedures.

MCA occupies an intermediate-to-high position in terms of complexity. MCA combines environmental, economic, and social criteria and incorporates uncertainty through scenario analysis. Complexity depends on multiple criteria, integrating qualitative and quantitative data, conducting sensitivity analyses or involving AI-assisted tools in the analytical design. Its complexity stems from configurability rather than standardisation.

3.4. *The Epistemological Dimension of the Selected Methods*

3.4.1. *Their Measurability*

The extent to which impacts can be consistently measured varies across the selected methods. The differences depend not only on operational characteristics but also on normative choices in methodological design, especially what should count as a measurable impact and what should be excluded. Among the selected methods, the environmental methods have the highest degree of measurability. CF and LCA integrate standardised and database-supported physical flow data. As shown in various papers, including studies of biodiesel supply chain assessments, quarrying operations in Istanbul, or household food waste management in Finland, LCA determines comparable environmental scores across impact categories. Cultural or social dimensions are structurally excluded from the analysis, as they cannot be expressed in physical or monetary units. Similarly, this applies to other impacts, e.g., landscape degradation or community disruption, which are not considered because, from a normative standpoint, objects of assessment are only those impacts that are amenable to physical quantification. As a correction, DLCA partially addresses the temporal dimension of measurability. By introducing the “variable time”, it can capture and describe the evolution of emission factors and technology performance over time, as well as their cumulative and long-term effects. However, the range of measurable impact types is limited to a physical, quantitative analytical perimeter.

Economic method reinforces measurability by adopting a monetisation logic even for values that are difficult to quantify. CBA assigns monetary values to all relevant benefits and costs, including intangible ones. However, this approach raises several concerns, as it implies that all impacts must be expressed in monetary terms. As described in some analysed publications (such as those assessing impacts of marine plastic waste management solutions or off-grid energy systems in developing countries), this necessity forces evaluators to make methodological decisions (e.g., discount rates, shadow prices, and willingness-to-pay estimates) that are themselves value-laden or to reduce all comparisons to a single metric to measure impacts that resist quantification. While the impact assessment of some elements is relatively straightforward (e.g., direct costs and benefits), others (specifically indirect or intangible components) raise substantial questions. Impacts that are difficult to monetise are therefore excluded from the assessment or approximated subjectively through proxies of uncertain validity. To avoid this, complex methods or software

tools can be employed, including parametric modelling, analogy-based estimation, and the Delphi method.

CEA partially mitigates this tension by fixing the evaluative objective and measuring only costs. However, this practice reduces evaluation to defined measurable outcomes, even though the definition of such outcomes is debated due to their contested effectiveness. Its measurability is moderate, as the definition of the evaluative objective and consequently of what qualifies as a meaningful outcome remains contested in practice.

Similarly, the IO method has medium measurability. It is highly effective when it measures only what national accounting systems record concerning environmental and economic flows at the sectoral level. Derived indicators, such as production, employment, and value-added multipliers, can be quickly determined. Outside these flows, it assesses informal economies, unpaid care work, and ecosystem services with less accuracy. Otherwise, outcomes concerning social well-being, cultural impacts, or environmental sustainability require changes (e.g., SAM or NAMEA matrices) and remain difficult to measure precisely. Implicit value assumptions, such as the equivalence of domestic and import prices or the allocation of environmental burdens, extend limitations to the interpretability of the conclusions.

Social methods demonstrate low to medium measurability. Their analyses focus on phenomena that are multidimensional, partially intangible, and context-dependent, and therefore difficult to estimate. SROI relies on financial proxies and well-being valuations to monetise social outcomes identified through stakeholder engagement. Its measurability is low: the reliability of proxy valuations depends heavily on the quality of stakeholder consultation and the availability of comparable benchmarks, both of which vary substantially across evaluative contexts. SIA assessments typically integrate statistical indicators with qualitative evidence (e.g., focus groups and interviews) to integrate distributional social impacts, as illustrated by studies on aquaculture assessments in coastal Spain and circular economy initiatives in agriculture. However, many impacts, including community cohesion, cultural identity, and social capital, remain constraints to standardised indicators, with doubts on their measurability. Measurability in GIA aligns with the intersectionality of the examined phenomena: gender impacts are difficult to reduce to a single indicator, and gender-disaggregated data are unavailable or unreliable in several regions. Key challenges encompass the evaluation of intangible outcomes. Empowerment, agency, and social norms are just some examples, together with the integration of qualitative insights into decision-making frameworks. HIA offers the most developed measurability framework among social methods, as it incorporates subjective evaluations, environmental exposures, and health outcomes, as demonstrated in assessments of ozone pollution in Beijing-Tianjin-Hebei and surface water contamination in China. However, indirect and long-term health effects, as well as psychosocial impacts, remain difficult to capture within standard HIA protocols.

The measurability of MCA is derived from and depends on the commensurability and quality of the examined indicators. Aggregating measurements produced by other assessment methods into a comparative ranking implies explicitly stating the choice of criteria and their operationalisation. When applied with AI-assisted scoring, the opacity of the underlying algorithm may further constrain the auditability of results. This variability makes MCA's measurability profile context-dependent rather than intrinsic to the method itself.

3.4.2. The Participation of Stakeholders

One of the innovations of this paper is the conceptualisation of stakeholder participation. Within this context, this participation is considered a key characteristic of each method and, at the same time, an epistemological variable that determines how evaluative

knowledge is produced, validated, and legitimised. Different levels of participation reproduce different expectations about the locus of knowledge production (whether expertise is centralised among technical specialists or distributed among affected stakeholders) and, indeed, establish a methodological dimension.

Environmental methods exhibit low to medium stakeholder participation, with a prevalence of technocratic epistemic models. In LCA, participation is limited to specific review phases and is mandatory only for publicly available comparative studies involving policy implications. In such cases, a critical review incorporating stakeholder representatives is recommended alongside independent expert review. For internal assessments, stakeholder involvement is still optional. According to the European Commission (2010) [28], stakeholder involvement should be tailored to the type of study and balanced against available resources. In practice, stakeholder review is compulsory for publicly disclosed comparative studies, but optional for internal assessments. It is also pivotal that studies addressed to broad audiences or characterised by significant methodological complexity should enhance transparency and credibility. When decision-makers are involved in the LCA development, outcomes are more likely to be understood and accepted. This helps reduce potential conflicts over methodological choices and is relevant in policy or territorial development contexts, where participation may improve through local priorities and practical needs. However, participation can also increase the complexity and duration of the modelling process and may generate power imbalances among actors. For this reason, effective participation requires balancing scientific accuracy with comprehensive engagement. Generally, stakeholder participation is low to moderate in LCA.

Participation is similarly limited in CF assessments. Calculations are often performed by technical experts or external consultants, with limited involvement from other organisational units. The concentration of analytical processes within specialised software platforms further reduces internal awareness and broadens the gap between knowledge production and organisational practice. However, in recent times, some institutions have adopted more participatory methods to address the need to consider carbon accounting as a cultural and organisational process. In these cases, internal and external stakeholders are involved in data collection, analysis, and calculation of greenhouse gas emissions. With this, the completeness and reliability of information are facilitated, as well as the dissemination of emission-reduction objectives across organisational levels. However, these practices remain exceptions rather than structural features of the method.

A similar pattern characterises DLCA: researchers and technical specialists primarily conduct calculations, while end users are involved only in limited phases, such as scenario definition.

Across these three methods, knowledge production remains largely expert-driven and centralised. Data are mainly drawn from standardised databases and administrative sources, while evaluative criteria are predefined by international standards, leaving affected communities with little influence over the analytical process. As described in some cases, for example, in the geothermal heat production assessments and the national input-output-based carbon footprint accounting, the prevalent epistemic model is technocratic. Stakeholders are, in large part, passive recipients of evaluation outcomes. This condition enhances replicability and comparability but has negative consequences as it excludes contextual and experiential knowledge (this might be important for the inclusion of the full set of impacts).

Economic methods occupy an intermediate position in terms of stakeholder participation. CEA and CBA include stakeholder preferences indirectly through willingness-to-pay estimates, revealed preferences, or politically defined objectives. However, the overall methodological design is still largely controlled by experts. Stakeholders' involvement can nonetheless be relevant in integrating more inclusive information on expected costs

and benefits. It can involve multiple perspectives and define those social or environmental externalities articulated to be measured. It also enhances the legitimacy and acceptability of evaluation outputs. Realistically, participation tends to remain limited because evaluation consists of quantitative data and economic indicators. Occasionally, stakeholders may collaborate to make decisions about data or its validation, as documented in carbon dioxide removal strategies in Saudi Arabia or the allocation of provincial carbon emission allowances in China. However, the analytical architecture and evaluative criteria are mostly predefined by the researchers. The epistemic model can therefore be portrayed as hybrid but asymmetrical. Stakeholder input is consultative rather than constitutive of the evaluative framework.

IO analysis is the most constrained method among economic ones. Here, stakeholder participation is absent, as the method is based totally on administrative data sources and institutional accounting frameworks and does not structurally include any participatory input.

By contrast, participation is central in social methods. Within these methods, the epistemic model is constructivist and deliberative because the co-production of knowledge is a methodological prerequisite rather than an optional enhancement.

SROI explicitly needs stakeholder involvement in several phases. These include definition, construction, and evaluation of the social value created by an intervention or project. Participation, therefore, is not purely consultative but epistemological and methodological, happening as stakeholders co-build the theory of change, identify the relevant outcomes, select monetary proxies, and estimate corrective indicators such as attribution, deadweight, displacement, and drop-off. For SROI to be epistemologically effective, stakeholders must freely express their perspectives, influence evaluative assumptions and estimations, and contribute to the interpretation of results. Consequently, the method builds upon the principles of participatory evaluation and value-oriented reporting, while also promoting forms of organisational learning and transformative accountability. Similarly, SIA extends stakeholder consultation to qualitative engagement instruments. Examples from the considered literature include interviews, focus groups, and direct observation. GIA integrates participatory methods with gender-disaggregated data collection. Examples here include the active involvement of women and marginalised groups to identify differential impacts and the definition of evaluative criteria. HIA incorporates stakeholder consultations in the scoping and interpretation phases, as in urban air quality assessments across US cities and heritage impact evaluations. As a result, its analytical structure appears more controlled than SROI or SIA. Its participation is accordingly medium because it is based on a hybrid epistemic model in which community knowledge advises but does not control the evaluative process.

MCA demonstrates the most variable participation profile of all selected methods. In its standard configuration—as in AHP-based site selection for agricultural land or groundwater potential mapping—MCA is expert-driven, with criteria and weights defined by the research team or a small group of technical experts. However, MCA's configurability also makes it uniquely suited to participatory adaptation: when weights are elicited through multi-stakeholder deliberation, or when community preferences are incorporated into the criteria definition process—as in sanitation prioritisation among Palestinian communities—the method can achieve high levels of epistemic inclusivity. Therefore, participation here is not only an intrinsic methodological property but also a governance choice. It depends on the configuration of the methods and the characteristics of the institutional framework, implying that epistemological openness does not depend merely on the structure of the method itself but also on the conditions under which it is operationalised and legitimised.

Table 3(a) presents the criteria description, while Table 3(b) presents the epistemological assessment of the eleven selected methods across three dimensions.

4. Discussion

4.1. Epistemological Patterns Across the Selected Methods

The comparative assessment of the methods considered reveals an articulated landscape of sustainability and policy impact assessment. The identified patterns are not only methodological limitations, but they also represent the concrete consequences of epistemological assumptions grounded in the design of the evaluation methods.

The first pattern is an inverse relationship between measurability and participation: methods with the highest measurability (LCA, CF, IO) are the least participatory because they rely on expert-driven knowledge production models that structurally exclude stakeholder interventions. Conversely, strongly participatory methods (SROI, SIA, GIA) exhibit lower measurability because they rely on qualitative approximations and context-dependent valuations that limit cross-context comparability. This inverse relationship results in epistemological conflict between the demand for replicable, generalisable knowledge and the recognition that evaluative validity requires the inclusion of community-held forms of knowledge. The integrity of the first methods depends on the standardisation of data and procedures. This standardisation in turn structurally excludes the contextual and community-held forms of knowledge that participatory methods can capture [18–20]. By contrast, stakeholder participation in co-producing knowledge enhances the validity of these methods. Epistemologically, their validity also depends on sensitivity to local conditions, value frameworks, and legitimacy among communities. At the same time, it represents the source of their weakness in measurability. Since they are co-produced and locally dependent, evaluation results cannot be successfully aggregated or abstracted for cross-context comparison.

This inverse relationship has practical consequences that are seldom recognised in evaluation practice. When decision-makers choose one method over another based on its measurability, as has sometimes occurred in EU regulatory impact assessments or World Bank project appraisals, they are at the same time, and frequently mistakenly, deciding on a low-participation epistemic model [18,19]. This decision is therefore political rather than only technical because it defines whose knowledge counts and which impacts need to be visible. At the same time, they determine whether local communities are co-creators of knowledge or are passive recipients [18–20]. Rigorous measurability and participatory co-production are in contrast and generate a structural tension in the evaluation exercises that should be explicitly addressed.

The second pattern refers to the correlation between complexity and epistemological openness. The analysis reveals that methods with high complexity (DLCA, LCA, IO) do not automatically have superior epistemological openness. Rather than a capability to accommodate uncertainty and transformative dynamics, their complexity is principally technical and reflects data requirements, software infrastructure, and multidisciplinary expertise. Two aspects, the epistemological richness and the methodological sophistication, are not equivalent. For example, DLCA advances the technical LCA framework. Yet its epistemic model is mainly technocratic: temporal dynamics are introduced top-down by specialists, while stakeholder participation is limited to a few activities. Assessed impacts are restricted to only physical and quantitative dimensions. High technical complexity produces a method of considerable analytical depth but limited epistemological breadth. In this direction, MCA is one of the clearer examples. Structurally and epistemologically open, this method can nevertheless be employed in fully technocratic ways. However, its openness is not a specific property of the methods; it depends on how the criteria are selected and the weights assigned, as well as how it is configured and legiti-

mised. Therefore, the analysis suggests that epistemological openness is not an essential characteristic of a method, but rather a function of how it is configured, governed, and legitimised within specific evaluative contexts [10,11].

The third pattern concerns what the analysis terms analytical reductionism. Most methods exclude those dimensions that they cannot measure and justify this exclusion as a technical default rather than a consequential normative choice. For example, LCA excludes social and distributional dimensions because they fall outside the physical flow accounting logic defined by ISO standards. Indeed, the method's epistemological structure does not include these dimensions, not because these dimensions are considered irrelevant. CBA reduces values that are difficult to estimate to monetary metrics. In this way, it helps compare them, but contemporary CBA nullifies the distributional and ethical dimensions of the involved trade-off. As outlined in some literature concerning off-grid energy systems in developing countries, or marine plastic waste management, the monetisation of values is itself value-laden, including normative assumptions about whose preferences count and which impacts should be discounted [12,13].

Analytical reductionism also concerns the implementation of methods in practices. Despite the proliferation of methods across fields of intervention, evaluators typically apply methods in isolation rather than in a complementary way. This application reflects practical constraints and epistemological default. Concerning this last aspect, the application reflects implicit assumptions about what constitutes a complete and legitimate evaluation of the respective methods. This aspect does not encourage integration with methods characterised by different epistemological premises.

SIA and GIA exhibit tension with the standardisation imposed by institutional decision-making frameworks—not because social and gender impacts are inherently unmeasurable, but because the participatory, context-sensitive epistemological model on which they are based resists the abstraction required for cross-context comparison. The reductionism observed in both cases is not a methodological failure but a structural consequence of specialisation, the high cost of analysis and data acquisition, or the reduction of the epistemological richness of participatory assessments to a form compatible with technocratic governance [18].

The last pattern concerns the underrepresentation of institutionally deployed methods in indexed scientific literature. The present bibliometric mapping demonstrates a gap between the methods most often used in applied evaluation practice and those most frequently represented in indexed academic literature. For example, SROI, GIA, and IO are largely implemented by international institutions but are underrepresented in the examined literature. This underrepresentation reinforces the established hierarchies of methodological legitimacy [18,20]: when methods produce quantifiable output, they are favoured in academic evaluation research; otherwise, methods that produce stakeholder-co-produced and institutionally embedded knowledge are less considered. The exclusion of the latter reinforces the dominance of quantitative and technocratic methods in academic discourse and, therefore, in evaluation practice.

4.2. Complementarities, Conflicts, and Trade-Offs Among Methods

Based on the epistemological patterns previously identified, no single method addresses all dimensions related to sustainability and policy impact assessment. Evaluation methods in isolation are therefore common practice; the present analysis examines the relational structure of the methodological landscape, i.e., eventual complementarities and conflicts among methods, and the trade-offs that evaluators face when combining them.

Regarding complementarities, some methods can be applied productively and flexibly. For instance, the combination of SIA and LCA is the most established complementary pairing in the examined literature, as they integrate environmental and distributional so-

cial consequences, enabling a more complete sustainability assessment than either method can achieve in isolation [22,23]. However, their combination is not straightforward. While LCA operates through a functional logic unit anchored in physical flows, SIA relies on stakeholder-defined impact pathways and qualitative evidence. Combining them means aligning methodological decisions, aggregating findings across incommensurable dimensions, and presenting results to decision-makers in a coherent form.

The combination of CBA and MCA also provides valuable results, especially in contexts where some impacts are monetisable, and others are not. Some projects related to territorial development or infrastructure planning are among the most suitable applications. Here, CBA can provide the economic foundation of the assessment while MCA structures the deliberation over incommensurable trade-offs. This combination is also recommended by several EU policy evaluation guidelines, particularly the European Commission Better Regulation Guidelines and associated Toolbox [54,55]. However, there is a risk of misidentification: the apparent comprehensiveness of the integrated framework may conceal a residual hierarchy in which monetised evaluation dominates the decision-making even when MCA is equally important.

Other international organisations, i.e., the World Bank, suggest using HIA and GIA in a complementary way, as both methods rely on gender-disaggregated and population-specific data. However, related problems can concern the nature of this data, often unavailable or unreliable, particularly in low- and middle-income country contexts.

Not all method combinations are productive. Some of them are based on epistemological assumptions that are incompatible or generate conflicts that can be resolved through technical refinement.

One of the most relevant conflicts is between MCA and CBA in contexts where values are incommensurable. While the first assumes that values across different dimensions cannot be reduced to a common metric without normative loss, the second assumes this as a mandatory requirement. When both methods are combined, the choice of how to weight monetised against non-monetised criteria becomes the centre of the normative decision-making process.

The epistemological architectures of LCA and SROI are different. The standardisation that makes LCA results auditable is what SROI's participatory logic resists. Their combination requires specifying whose knowledge is legitimate, how physical and social impacts are to be aggregated, and whether the resulting composite indicator retains meaningful interpretive validity. In other words, it requires explicit governance decisions that current methodological frameworks do not provide.

Beyond these combinations, the research demonstrates some trade-offs that evaluators face regardless of which methods they select; the first one concerns rigour and participation. Methods with the highest quantitative rigour are also the most replicable and comparable. At the same time, they are also the least participatory. By contrast, the most participatory methods are the least replicable and most sensitive to local context. Evaluators cannot maximise both dimensions but can decide which knowledge is recognised and worthy of dissemination [14,15].

The second trade-off concerns the analytical comprehensiveness and operational feasibility. Multi-method integrated frameworks that combine different methods capture more dimensions of a project or policy than a single method. In doing so, they need large datasets or complex expertise. This aspect makes these methods prohibitive for many evaluators, who therefore choose simpler, single-method assessments in applied evaluation practice. This bias reinforces the analytical reductionism already detailed.

Another trade-off involves temporal depth and operational timeliness. Methods that incorporate long-term and dynamic dimensions are enabled to capture cumulative, indirect, and non-linear impacts over time. Their implementation timelines make them poorly

suited to the decision cycles of most policy processes, are shorter. The tension between the temporal scope required for rigorous impact assessment and the temporal constraints of policy decision-making is a structural feature to be considered in the integrated frameworks.

4.3. Towards a New Evaluation Paradigm

Addressing the limitations described above requires paradigmatic reorientation toward a co-produced evaluation framework based on the principles of systematicity, transformativity, and co-produced evaluation. This reorientation is possible, and additional considerations support the prospective role of AI-enabled tools as cross-cutting enablers of the framework.

Systematicity means that impact assessment should be reformulated to be more efficient and operate across multiple dimensions simultaneously (environmental, economic, and social), rather than maximising analytical depth within a single disciplinary domain. Evaluators are encouraged to minimise reductionism and develop integrative frameworks that leverage complementarities among methods. At the same time, they should abandon the idea of adopting only a single universal method. Such frameworks should better explain the value trade-offs involved in prioritising one dimension over another. In addition, they should better support decision-makers in navigating the plurality of evaluative perspectives and abandoning the epistemological limits of any single approach. This aspect implies designing a multi-method architecture that leverages complementarities while maintaining manageable trade-off between methods. It also implies delineating consistent system boundaries across methods and developing integration protocols involving comprehensive and comparable results.

The second principle is transformativity, which repositions impact assessment beyond a mere exercise in accountability and ex-post attribution. Guided by this principle, assessment becomes a generative tool for collective learning, adaptation, and transformative change. To be operational, it integrates developmental approaches into evaluation practice, prioritising real-time feedback and adaptive learning over ex-post attribution and reconceptualises uncertainty not as a limitation to be controlled but as an inherent feature of the policy settings in which evaluation operates. Essentially, it recognises that the objectives themselves are subject to revision and that evaluation has a constitutive role in determining the direction of policy and institutional change. Operationally, transformativity prioritises real-time feedback and adaptive learning over retrospective attribution and incorporates mechanisms to revise evaluative criteria as the policy context evolves. Uncertainty is reconceptualised not as a limitation to be controlled through methodological conservatism but as a structural feature of the policy environments in which evaluation operates.

The third principle is participatory inclusivity. Rather than treating stakeholder participation as a procedural requirement or an exercise in democratic accountability, impact assessment should recognise it as an epistemological imperative. Structured in this way, participation strengthens the validity of evaluation by making diverse forms of knowledge visible and assessable. This legitimacy cannot be conferred retrospectively through consultation at the validation stage; it must be built iteratively through the active involvement of affected stakeholders in defining evaluative criteria, selecting methods, collecting data, and interpreting results. In this respect, several institutional frameworks governing evaluation practice—such as those developed within the European Commission's Better Regulation agenda—reflect this orientation, emphasising stakeholder engagement as a core element of effective governance [55,58].

Their implementation entails important operational implications. Evaluation frameworks should support multidimensional designs and recognise stakeholder-generated

knowledge as legitimate evidence. Rather than relying on a single standardised method (as normally done in EU regulatory impact assessments and World Bank project appraisals), businesses should integrate distributional and social dimensions. Conventional analytical tools often overlook these dimensions; their integration should be consistent with CSRD reporting requirements and the Global Reporting Initiative Standards. For civil society organisations and communities, participatory inclusivity requires the co-creation of criteria, indicators, and value assumptions within impact assessment. One operational route is an integrated evaluation framework that combines LCA, SIA, and MCA. Within this architecture, LCA provides the environmental substance, and SIA addresses social impacts. MCA structures negotiation over value trade-offs and reconfigures assumptions related to each component's method in a more transparent, contested, and legitimised way through stakeholder participation. It does not eliminate methodological trade-offs but governs them systematically.

The integration of measurable environmental analysis, participatory social assessment, and an integrative decision framework avoids relying on a single epistemological model to guide the evaluation process. This process also subjects value choices to clear deliberation rather than technical decision-making.

Another element in the governing of these principles should be the use of artificial intelligence (AI). Rather than constituting a standalone evaluation method, AI can be understood as a cross-cutting enabler that augments existing approaches and supports the integration of environmental, economic, and social dimensions within more adaptive analytical frameworks [59–61]. Processing several databases simultaneously, it reveals cross-domain interdependencies often overlooked by siloed methods and supports the qualitative evidence through natural language processing techniques, the systematic analysis of stakeholder narratives, and other unstructured data. In predictive terms, AI helps to simulate and capture non-linear dynamics, explore alternative policy scenarios, and extend the temporal scope of evaluation. Otherwise, concerns about its use enclose algorithmic transparency, data governance, and potential biases in training datasets, which demand robust ethical oversight and institutional protection.

Another dimension of the proposed framework concerns the prospective role of AI-enabled tools as cross-cutting enablers of the three principles. AI cannot be a standalone method but an instrument that increases efficiency and supports process integration with more adaptive architectures. The analysis presented in this paper demonstrates that AI integration is supported only by environmental methods, especially LCA and its variants, and has not been extended to participatory or social assessment methods. This evidence reinforces the inverse relationship between measurability and participation. While AI reinforces the systemicity and transformativity principles, it does not equally address the co-produced principle. Introducing AI into integrated and participatory evaluations (including through natural language processing, participatory simulation tools, and stakeholder engagement) at scale requires explicit attention to algorithmic transparency, data governance, and the risk that AI reproduces the existing epistemological hierarchies.

4.4. Artificial Intelligence-Enabled Tools and Perspectives for Impact Assessment: Current State of the Art and Emerging Directions

Recent developments in understanding the influence of artificial intelligence on impact assessment reveal two distinct paths: the creation of AI-driven tools coincides with a growing body of ideas, theoretical frameworks, and moral considerations about their effects on evaluation practices. From a functional perspective, Rahman et al. [61] present a comprehensive transformation of Environmental Impact Assessment (EIA) and Strategic Environmental Assessment (SEA) via generative artificial intelligence (GenAI). This work considers AI a framework for scenario generation and fosters participation in assessment.

The research describes several applications, such as synthetic data creation, environmental scenario simulations across different scales, improving the analysis of geospatial and remote-sensing data, and enhancing stakeholder interaction and communication developed through Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), diffusion models, and large language models (LLMs). The paper also includes a theoretical discussion of epistemic uncertainty, non-linear system behaviours, transparency, clarity, and ethical governance as pillars supporting the adoption of AI in impact assessment methodologies [61].

In terms of evaluations of products and processes, Nwagwu et al. [62] suggest an organised framework for integrating AI into Life Cycle Assessment (LCA), utilising machine learning and generative AI throughout all stages of analysis, including inventory building, filling data gaps, impact characterisation, managing uncertainty, and exploring scenarios. In addition to its technical setup, the framework proposes a forward-looking vision for AI-assisted impact assessments characterised by semi-automation, adaptive learning, and human oversight, addressing the potential dangers of algorithmic ambiguity, biased models, and excessive automation in decision-making processes.

Building on this research, Bouzari et al. [63] expand the application of AI by considering governance and accountability aspects of sustainable evaluation within a transformer-based framework. The objective of this structure is to evaluate the credibility of environmental sustainability claims in the cryptocurrency sector. Although the work does not use this technique as a standard impact assessment tool, it uses artificial intelligence within qualitative, discursive, and self-reported analysis of impact claims, integrating it with broader ethical and legal debates on greenwashing, transparency, and legitimacy in AI-enabled sustainability governance [60–62].

The literature generally suggests that artificial intelligence is influencing impact evaluations as both a collection of functional tools and an imaginative epistemological methodology, therefore enabling data augmentation, scenario-based reasoning, and new forms of impact evidence. At the same time, it highlights current problems with ethical oversight, interpretability, stakeholder participation, and the convergence of environmental, financial, social, and governance elements. These difficulties point to future research directions that centre on cooperation in developing AI-enabled impact assessment tools, incorporating ethical and epistemic safeguards from the beginning, and building cohesive, transforming evaluation frameworks—objectives that directly support and inspire the conceptual approach presented in this paper (Figure 2).

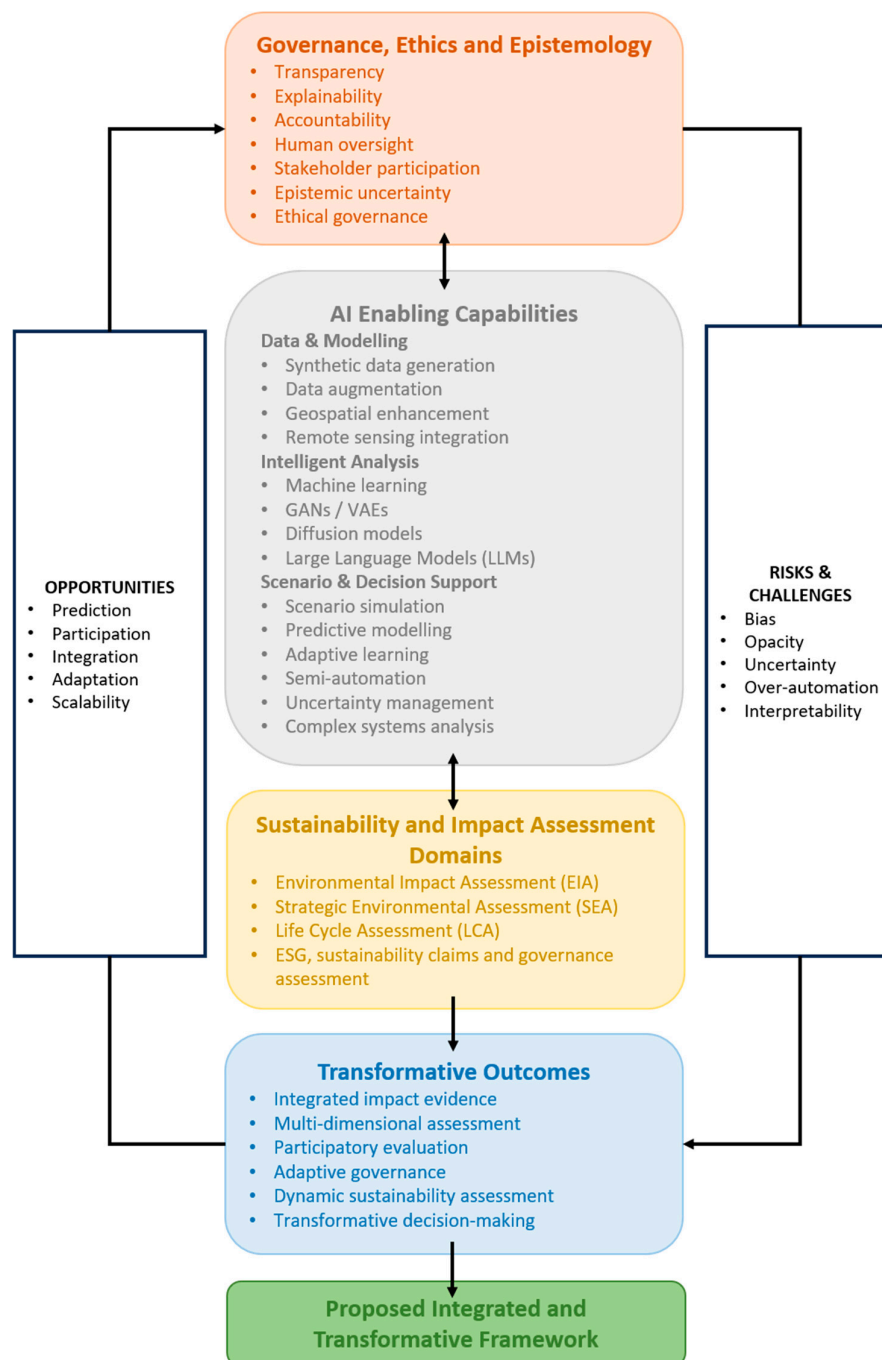


Figure 2. From AI-enabled impact assessment to an integrated and transformative evaluation framework. Source: authors' own elaboration (2025/2026).

5. Conclusions

This paper has examined the structural and epistemological limitations of eleven widely used sustainability and policy impact assessment methods through a scoping review of peer-reviewed and institutional literature published between 2022 and 2025, complemented by an AI-assisted text-mining protocol and a structured comparative assessment across seven analytical dimensions.

This analysis confirms that existing methods continue to be deployed in a fragmented and sector-specific manner, rarely extending beyond single-domain or single-scale evaluations. The relative underrepresentation of institutionally deployed methods—particu-

larly SROI, GIA, and IO—in the indexed scientific corpus reflects not merely a bibliometric artefact but a structural bias in academic knowledge production that reinforces existing hierarchies of methodological legitimacy.

Beyond this, three findings emerge: 1. an inverse relationship between measurability and participation; 2. the separation between epistemological openness and technical complexity; and 3. the structural bias towards analytical reductionism. Addressing them is a prerequisite to a paradigmatic reorientation organised around three interrelated principles. Systemic evaluation implies the thoughtful design of multi-method architectures that leverage complementarities across all dimensions while making value trade-offs explicit and manageable. Transformative evaluation positions assessment as a generative instrument for collective learning and adaptive change, reconceptualising uncertainty as a fundamental feature of policy environments rather than a constraint to be controlled. Co-produced evaluation recognises stakeholder participation as an epistemological imperative, requiring the active involvement of affected in defining evaluative criteria, selecting methods, and interpreting results.

However, the operationalisation of these principles within specific policy and project evaluation contexts remains a key challenge for future research and practice. Rather than proposing a single integrated method, this study outlines an epistemological architecture built on the complementary strengths of existing approaches that can be strategically mobilised, their limitations transparently acknowledged, and their underlying value assumptions subjected to critical scrutiny—commensurate with the complexity and normative ambitions of contemporary governance challenges. In this sense, transformation-focused evaluation requires a transformation of evaluation itself, driven by contextual adaptation and high-inference interpretation rather than by rigid or binary methodological approaches or mere incremental methodological improvement. On the contrary, it requires a revision of the foundational assumptions that govern how impact is defined, measured, and legitimised. In this process, an integrated evaluation framework combining methods and artificial intelligence may also play an important supporting role.

The examination also reveals that AI integration in impact assessment is currently concentrated on quantitative methods concerning environmental issues and has not extended to participatory or social approaches. This absence reinforces the epistemological hierarchy identified in the comparative analysis. The AI's transformative potential requires the development of tools oriented toward co-produced evaluation, including natural language processing and participatory simulation, with explicit attention to algorithmic transparency, data governance, and power asymmetries in knowledge production.

Findings open the door to discussion of the roles and the power dynamics/asymmetry among stakeholders. They also suggest better investigating existing structural inequalities and the ethical and political dimensions of evaluation transitions. Further research requires testing across additional methods and policy domains, including emerging approaches such as AI-assisted evaluation and real-time data monitoring. Enriching impact assessment methods through AI can facilitate the analysis of unstructured qualitative evidence using natural language processing, enabling scenario exploration and modelling of non-linear and dynamic system behaviour, and increasing the scalability, adaptability, and timeliness of evaluation processes. AI can also support systemic, participatory, and transformation-oriented impact assessment. Further research should include empirical validation of the proposed frameworks, and the development of AI-assisted evaluation tools that genuinely support participatory and transformative assessment.

Author Contributions: Conceptualization: V.C., N.B., F.F., and E.N.; methodology: V.C., N.B., F.F., and E.N.; validation, F.F.; formal analysis, V.C., N.B., and E.N.; investigation: V.C., N.B. and E.N.; writing—original draft preparation, V.C., N.B., and E.N.; writing—review and editing, V.C., N.B.,

F.F., and E.N.; visualization, N.B. and E.N.; supervision, F.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available on request.

Acknowledgments: The authors would like to express their deepest gratitude to Patrizia Grifoni for her invaluable support, generous guidance, and thoughtful insights, which greatly enhanced this work. During the preparation of this manuscript/study, the authors used DeepL App Desktop 26.7.1.20277 and ChatGPT (GPT-4o, OpenAI) for the proofreading. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Sachs, J.D.; Schmidt-Traub, G.; Mazzucato, M.; Messner, D.; Nakicenovic, N.; Rockström, J. Six transformations to achieve the sustainable development goals. *Nat. Sustain.* **2019**, *2*, 805–814.
2. UNEP. *The Closing Window: Climate Crisis Calls for Rapid Transformation of Societies (Emissions Gap Report 2022)*; United Nations Environment Programme: Nairobi, Kenya, 2022. Available online: <https://www.unep.org/resources/emissions-gap-report-2022> (accessed on 1 December 2025).
3. Vogel, C.; O'Brien, K. Getting to the heart of transformation. *Sustain. Sci.* **2022**, *17*, 653–659.
4. Ofir, Z.; Rugg, D. Transforming evaluation for times of global transformation. *Am. J. Eval.* **2021**, *42*, 47–52.
5. Magro, C.; Van den Berg, R.D.; Adrien, M.-H. Transforming evaluation for transformational changes. In *Transformational Evaluation for the Global Crises of Our Times*; Van den Berg, R.D., Magro, C., Adrien, M.-H., Eds.; IDEAS: Exeter, UK, 2021; pp. 1–18.
6. Buckton, S.J.; Fazey, I.; Ball, P.; Ofir, Z.; Colvin, J.; Darby, M.; Hejnowicz, A.P.; Leicester, G.; Newman, R.; Page, G.G.; et al. Twelve principles for transformation-focused evaluation. *PLoS Sustain. Transform.* **2025**, *4*, e0000164.
7. Albakri, M.; Wood-Harper, T. Revisiting critical systems thinking: Enhancing the gaps through sustainability and action methodologies. *Syst. Res. Behav. Sci.* **2025**, *42*, 157–170.
8. Pope, J.; Bond, A.; Hugé, J.; Morrison-Saunders, A. Reconceptualising sustainability assessment. *Environ. Impact Assess. Rev.* **2017**, *62*, 205–215.
9. Ekins, S.; Puhl, A.; Zorn, K.; Lane, T.; Russo, D.; Klein, J.; Clark, A. Exploiting machine learning for end-to-end drug discovery and development. *Nat. Mater.* **2019**, *18*, 435–441.
10. Boulanger, P.; Bréchet, T. Models for policy-making in sustainable development: The state of the art and perspectives for research. *Ecol. Econ.* **2005**, *55*, 337–350.
11. Rotmans, J. Tools for integrated sustainability assessment: A two-track approach. *Integr. Assess. J.* **2006**, *6*, 35–57.
12. Jackson, M. Critical systems practice 3: Intervene—Flexibly executing a multimethodological intervention. *Syst. Res. Behav. Sci.* **2022**, *39*, 1014–1023.
13. Jackson, M. Critical systems practice 4: Check—Evaluating and reflecting on a multimethodological intervention. *Syst. Res. Behav. Sci.* **2023**, *40*, 617–632.
14. Beck, U.; Bonss, W.; Lau, C. The theory of reflexive modernization: Problematic, hypotheses and research programme. *Theory Cult. Soc.* **2003**, *20*, 1–33.
15. Jackson, M. *Critical Systems Thinking and the Management of Complexity*; John Wiley and Sons: Chichester, UK, 2019.
16. Schwandt, T.A. Post-normal evaluation? *Evaluation* **2019**, *25*, 317–329.
17. Patton, M.Q. Evaluation criteria for evaluating transformation: Implications for the coronavirus pandemic and the global climate emergency. *Am. J. Eval.* **2020**, *42*, 53–89.
18. Mertens, D.; Wilson, A. *Program Evaluation Theory and Practice: A Comprehensive Guide*; Guilford Press: New York, NY, USA, 2012.
19. Mertens, D. Transformative research: Personal and societal. *Int. J. Transform. Res.* **2017**, *4*, 18–24.

20. Turnhout, E.; Metze, T.; Wyborn, C.; Klenk, N.; Louder, E. The politics of co-production: Participation, power, and transformation. *Curr. Opin. Environ. Sustain.* **2020**, *42*, 15–21.
21. Van Noordt, C.; Misuraca, G. Artificial intelligence for the public sector: Results of landscaping the use of AI in government across the European Union. *Gov. Inf. Q.* **2022**, *39*, 101645.
22. Peters, M.D.; Marnie, C.; Tricco, A.C.; Pollock, D.; Munn, Z.; Alexander, L.; Khalil, H. Updated methodological guidance for the conduct of scoping reviews. *JBI Evid. Synth.* **2020**, *18*, 2119–2126.
23. Tricco, A.C.; Lillie, E.; Zarin, W.; O'Brien, K.K.; Colquhoun, H.; Levac, D.; Straus, S.E. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Ann. Intern. Med.* **2018**, *169*, 467–473.
24. Munn, Z.; Peters, M.D.; Stern, C.; Tufanaru, C.; McArthur, A.; Aromataris, E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med. Res. Methodol.* **2018**, *18*, 143.
25. Mingers, J.; Harzing, A.W. Ranking journals in business and management: A statistical analysis of the Harzing data set. *Eur. J. Inf. Syst.* **2007**, *16*, 303–316.
26. ISO 14040:2006; Environmental Management—Life Cycle Assessment—Principles and Framework. International Organization for Standardization (ISO): Geneva, Switzerland, 2006.
27. European Commission. Life Cycle Assessment. Available online: <https://eplca.jrc.ec.europa.eu/lifecycleassessment.html> (accessed on 1 December 2025).
28. European Commission Joint Research Centre Institute for Environment and Sustainability. *International Reference Life Cycle Data System (ILCD) Handbook—General Guide for Life Cycle Assessment—Detailed Guidance*; Publications Office of the European Union: Luxembourg, 2010.
29. Batool, F.; Kurniawan, T.A.; Mohyuddin, A.; Othman, M.H.D.; Aziz, F.; Al-Hazmi, H.E.; Anouzla, A. Environmental impacts of food waste management technologies: A critical review of life cycle assessment (LCA) studies. *Trends Food Sci. Technol.* **2024**, *143*, 104287.
30. De Leonardis, F. *Lo Stato Ecologico. Approccio Sistemico, Economia, Poteri Pubblici e Mercato*; Giappichelli Editore: Torino, Italy, 2023; p. 107.
31. Samani, P. Synergies and gaps between circularity assessment and life cycle assessment (LCA). *Sci. Total Environ.* **2023**, *903*, 166611.
32. Wiedmann, T.; Minx, J. A definition of 'carbon footprint'. *Ecol. Econ. Res. Trends* **2008**, *1*, 1–11.
33. AIPE. *Carbon Footprint—Guida al Calcolo*; AIPE: Milan, Italy, 2016.
34. de Lima, G.B.C.; de Castro Pereira, G.; González, J.; Haddad, A.N. A systematic literature review for the development of a consolidated method for DLCA. *Recent Prog. Sci. Eng.* **2025**, *1*, 14.
35. Su, S.; Zhang, H.; Zuo, J.; Li, X.; Yuan, J. Assessment models and dynamic variables for dynamic life cycle assessment of buildings: A review. *Environ. Sci. Pollut. Res.* **2021**, *28*, 26199–26214.
36. Bahramian, M.; Hynds, P.D.; Priyadarshini, A. Dynamic life cycle assessment of commercial and household food waste: A critical global review of emerging techniques. *Sci. Total Environ.* **2024**, *921*, 170853.
37. Beloin-Saint-Pierre, D.; Albers, A.; Hélias, A.; Tiruta-Barna, L.; Fantke, P.; Levasseur, A.; Collet, P. Addressing temporal considerations in life cycle assessment. *Sci. Total Environ.* **2020**, *743*, 140700.
38. Lueddeckens, S.; Saling, P.; Guenther, E. Temporal issues in life cycle assessment—A systematic review. *Int. J. Life Cycle Assess.* **2020**, *25*, 1385–1401.
39. Sohn, J.; Kalbar, P.; Goldstein, B.; Birkved, M. Defining temporally dynamic life cycle assessment: A review. *Integr. Environ. Assess. Manag.* **2020**, *16*, 314–323.
40. Robinson, R. Cost-benefit analysis. *Br. Med. J.* **1993**, *307*, 924–926.
41. Brent, R.J. *Applied Cost-Benefit Analysis*; Edward Elgar Publishing: Cheltenham, UK, 2007.
42. Layard, P.R.G.; Layard, R.; Glaister, S. (Eds.) *Cost-Benefit Analysis*; Cambridge University Press: Cambridge, UK, 1994.
43. Mishan, E.J.; Quah, E. *Cost-Benefit Analysis*; Routledge: London, UK, 2020.
44. Glandon, D.; Fishman, S.; Tulloch, C.; Bhula, R.; Morgan, G.; Hirji, S.; Brown, L. The state of cost-effectiveness guidance: Ten best resources for CEA in impact evaluations. *J. Dev. Eff.* **2023**, *15*, 5–16.
45. Wouterse, B.; van Baal, P.; Versteegh, M.; Brouwer, W. The value of health in a cost-effectiveness analysis: Theory versus practice. *Pharmacoeconomics* **2023**, *41*, 607–617.
46. Leontief, W. *Input-Output Economics*; Oxford University Press: New York, NY, USA, 1986.
47. Leontief, W. Input-output analysis. In *The New Palgrave Dictionary of Economics*; Palgrave Macmillan: London, UK, 2018; pp. 6566–6572.

48. Miller, R.E.; Blair, P.D. *Input-Output Analysis: Foundations and Extensions*; Cambridge University Press: Cambridge, UK, 2009.
49. Maldonado, M.; Corbey, M. Social return on investment (SROI): A review of the technique. *Maandbl. Account. Bedrijfsecon.* **2016**, *90*, 79–86.
50. Arvidson, M.; Lyon, F.; McKay, S.; Moro, D. Valuing the social? The nature and controversies of measuring social return on investment (SROI). *Volunt. Sect. Rev.* **2013**, *4*, 3–18.
51. Busacca, M.; Caputo, A. *Valutazione, Apprendimento e Innovazione nelle Azioni di Welfare Territoriale*; Ca' Foscari Edizioni: Venezia, Italy, 2020.
52. Kalpazidou Schmidt, E.; Cacace, M. Addressing gender inequality in science: The multifaceted challenge of assessing impact. *Res. Eval.* **2017**, *26*, 102–114.
53. Brown, G.D.; Largey, A.; McMullan, C. The impact of gender on risk perception: Implications for EU member states' national risk assessment processes. *Int. J. Disaster Risk Reduct.* **2021**, *63*, 102452.
54. European Commission. *Guide to Cost-Benefit Analysis of Investment Projects: Economic Appraisal Tool for Cohesion Policy 2014–2020*; Directorate-General for Regional and Urban Policy: Brussels, Belgium, 2014.
55. European Commission. *Better Regulation Guidelines*; Staff Working Document SWD(2021) 305 Final; European Commission: Brussels, Belgium, 2021.
56. Toolan, M.; Walpole, S.; Shah, K.; Kenny, J.; Jónsson, P.; Crabb, N.; Greaves, F. Environmental impact assessment in health technology assessment: Principles, approaches, and challenges. *Int. J. Technol. Assess. Health Care* **2023**, *39*, e13.
57. IAIA. Health Impact Assessment General Info. Available online: <https://iaia.org/health/health-impact-assessment-general-info/> (accessed on 1 December 2025).
58. European Union. *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions—Better Regulation for Better Results—An EU Agenda*; COM(2015) 215 Final; European Commission: Brussels, Belgium, 2015.
59. Vinuesa, R.; Azizpour, H.; Leite, I.; Balaam, M.; Dignum, V.; Domisch, S.; Fuso Nerini, F. The role of artificial intelligence in achieving the sustainable development goals. *Nat. Commun.* **2020**, *11*, 233.
60. Rolnick, D.; Donti, P.L.; Kaack, L.H.; Kochanski, K.; Lacoste, A.; Sankaran, K.; Bengio, Y. Tackling climate change with machine learning. *ACM Comput. Surv.* **2022**, *55*, 1–96.
61. Rahman, S.M.; Raihan, A.; Abudalfa, S. Generative artificial intelligence for environmental assessment: A new paradigm for sustainability analysis. *Environ. Manag.* **2026**, *76*, 93.
62. Nwagwu, C.C.; Ogorodnyk, O.; Sølvsberg, E.; Eleftheriadis, R.J.; Meskers, C. Integrating artificial intelligence into life cycle assessment: A framework for balancing automation and human expertise. *J. Sustain. Metall.* **2025**, *11*, 3590–3605.
63. Bouzari, P.; Fekete-Farkas, M.; Szalay, Z.G. Intelligent sustainability: Evaluating transformers for cryptocurrency environmental claims. *Information* **2025**, *16*, 1022.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.