Proceedings of the Design Society, Volume 6: DESIGN 2026
https://doi.org/10.1017/pds.2026.10525

AI-assisted leading sustainability criteria development: a multiple case study

Jennifer Woofter✉, Jesko Schulte and Matilda S. Watz

Blekinge Institute of Technology, Sweden

✉ jennifer.woofter@bth.se

ABSTRACT: This study examines how AI can support the development of Leading Sustainability Criteria in sustainable product development, comparing AI-generated outputs with human-facilitated workshop results from four Swedish companies. Results highlight AI's ability to accelerate and broaden sustainability framing, but emphasize that contextual relevance and legitimacy depend on participatory inputs. The findings suggest that AI is most effective when integrated into hybrid workflows that preserve human insight and stakeholder engagement—offering practical guidance for future implementation.

KEYWORDS: artificial intelligence (AI), sustainable design, sustainability criteria, sustainable product development, human-AI collaboration

1. Introduction

Sustainable Product Development (SPD) has gained increasing attention across research and industry as a means to integrate sustainability considerations proactively into design and innovation processes. Over the past two decades, numerous methods and tools have been developed to support this ambition (Faludi et al., 2020). However, the practical application of these tools often depends on extensive time commitments and skilled facilitation, representing a key bottleneck to the broader uptake of SPD methods and tools (Mallalieu et al., 2024). At the same time, advances in artificial intelligence (AI) are challenging traditional assumptions about how knowledge integration and decision support occur in design (Boncella, 2024). AI systems, and large language models (LLMs) in particular, show potential to perform or augment some of the cognitive and facilitative functions traditionally carried out by human experts, yet critical questions remain about the quality, reliability, and contextual relevance of AI-generated insights—especially in participatory sustainability design settings where tacit expertise and stakeholder interaction play central roles (Billi & Labraña, 2025; Corsini et al., 2025).

This study explores these questions by examining whether AI can substitute or complement human-generated outputs in the application of SPD methods and tools. The study focuses on one such method used to create company-specific leading sustainability criteria and compares AI-generated outputs with those produced by human groups of internal company representatives working together with external SPD experts. Specifically, the study asks, ***"To what extent can AI substitute or complement humans in the development of Leading Sustainability Criteria?"*** By investigating these questions, this paper contributes to a growing body of research on the intersection of AI, sustainability, and design practice. It offers empirical evidence on the strengths and limitations of AI in supporting knowledge-intensive, collaborative sustainability processes—helping clarify whether AI should be viewed primarily as a substitute for, or a complement to, human-only participatory processes in SPD.

2. Background

This section reviews how sustainability criteria are conceptualized in design literature and their importance for early-stage decision-making and examines emerging research on the extent to which AI can substitute for or complement humans in this context.

https://doi.org/10.1017/pds.2026.10525 Published online by Cambridge University Press

2.1. Leading sustainability criteria development in product development

Early-stage design decisions often determine the majority of a product's long-term sustainability performance (Ahmad et al., 2018). Once materials, suppliers, or production methods are locked in, the ability to make meaningful changes becomes limited (Hallstedt, 2017). To ensure the successful adoption of SPD methods, it is crucial to engage stakeholders through consensus building and participatory model building. These collaborative approaches foster shared understanding, align diverse priorities, and enhance commitment to sustainability goals throughout the design process (Mallalieu et al., 2024). To address this, sustainability researchers have developed tools and frameworks that help companies identify and prioritize Leading Sustainability Criteria (LSCs), i.e. forward-looking design requirements that translate broad sustainability principles into measurable, company-specific criteria that can guide decision-making in product and service development. By focusing on leading rather than lagging indicators, LSCs allow organizations to anticipate future sustainability challenges and opportunities, rather than reacting to them after impacts occur (Kravchenko et al., 2019; Saidani et al., 2022).

One approach for developing LSCs is the Leading Sustainability Criteria Development (LEASA) workshop method. LEASA builds on the Framework for Strategic Sustainable Development (FSSD) and uses a structured, participatory process to translate sustainability principles into actionable design guidance (Watz & Hallstedt, 2024). A typical LEASA process includes four main steps: 1) envision a sustainable future state or solution, 2) assess the current state, 3) identify the most material sustainability issues to develop leading criteria, and 4) develop sustainable design strategies using those criteria. This process is usually facilitated by sustainability experts and involves cross-functional workshops with company stakeholders. The output is a set of LSCs, each with associated indicators, lifecycle relevance, and sources of assessment data. These criteria can then be used to evaluate products, services, or innovation pathways—often supported by tools such as the Sustainability Compliance Index (SCI) and the Sustainability Fingerprint (SF), which help translate qualitative criteria into measurable performance profiles (Hallstedt, 2023). In this way, the SCI serves as a bridge between the defined LSCs and product-level assessment, translating qualitative sustainability criteria into measurable performance profiles that help define a "sustainability design space." This design space supports early-stage product development decisions by clarifying acceptable performance ranges and highlighting improvement opportunities.

2.2. AI in design and sustainability contexts

AI is increasingly integrated into design and sustainability domains, offering new possibilities for data analysis, creativity support, and decision facilitation. Most existing applications, however, remain concentrated in narrowly defined, data-driven tasks such as optimization, predictive maintenance, or sustainability reporting (e.g. Haddud, 2024). Studies in product development and circularity measurement show that AI can accelerate information processing, broaden the scope of explored options, and surface overlooked interconnections (Corsini et al., 2025). When combined with domain expertise, AI can strengthen creativity and efficiency, supporting what Marzi and Balzano (2025) describe as "adaptable, skill-differentiated" teams for sustainable product innovation. Yet, these benefits are contingent on hybrid collaboration: Nikolić and Bjelica (2025) caution that while AI may enhance efficiency, it can also disrupt the balance between human creativity and automated reasoning, underscoring the need for models in which AI complements—rather than replaces—human expertise. Scholars approaching AI from socio-technical systems and organizational perspectives further emphasize that expertise encompasses more than factual accuracy. It is socially and institutionally situated, shaped by trust, accountability, and interpretive capacity (e.g. Billi & Labraña, 2025; Trunk et al., 2020). AI excels at pattern recognition and rapid synthesis but lacks the tacit knowledge and legitimacy mechanisms that enable experts to contextualize uncertainty and build consensus. Human experts justify their reasoning, adapt to changing conditions, and mediate among conflicting stakeholder values—functions not easily automated. The opacity of algorithmic systems ("black-box" reasoning) further complicates their credibility in participatory settings where transparency and dialogue are essential (Pasquale, 2015). However, several studies suggest that well-designed human–AI hybrids can outperform either side alone. Gupta et al. (2024) and Ghosh et al. (2025) characterize AI as a potential "force multiplier" that enhances sensemaking and organizational learning when integrated into human decision loops. In sustainability contexts, AI's analytical reach can extend experts' cognitive span—rapidly scanning standards, databases, and public disclosures—while experts provide the interpretive framing that ensures relevance and ethical grounding (Corsini et al., 2025).

https://doi.org/10.1017/pds.2026.10525 Published online by Cambridge University Press

Within design for sustainability, these insights converge on a central tension: the promise of automation versus the indispensability of human facilitation (Brambila-Macias & Sakao, 2021). While AI systems can generate structured outputs that resemble expert work, their effectiveness depends on the type and richness of information they are given and the degree of human interpretation that follows. Consequently, current discourse calls for empirical investigation into how AI performs in knowledge-intensive, participatory design processes and whether its role should be conceived as substitution, augmentation, or collaboration (Billi & Labraña, 2025).

3. Research design and methodology

This study adopted a comparative, multi-case research design to investigate the extent to which AI can replicate or augment the role of humans in the LEASA method. The inquiry was structured as a series of controlled tests, in which an AI system (GPT-5, via Microsoft CoPilot) was prompted to generate LSCs for four Swedish companies operating in different sectors. These companies were purposefully selected: each was an active participant in ongoing sustainability research collaborations and had either already completed, or expressed willingness to conduct, a LEASA workshop and the Sustainability Fingerprint tool. This ensured both access to relevant workshop outputs and a baseline level of sustainability engagement. At the same time, this sampling strategy introduces a limitation: all cases are mature, sustainability-oriented Swedish firms, which may bias results toward comparatively structured or well-developed inputs. This study, therefore, explicitly treats the findings as context-specific and notes where conclusions may not generalize to companies with lower sustainability maturity or in other geographic contexts. Figure 1 provides an overview of the methodology.

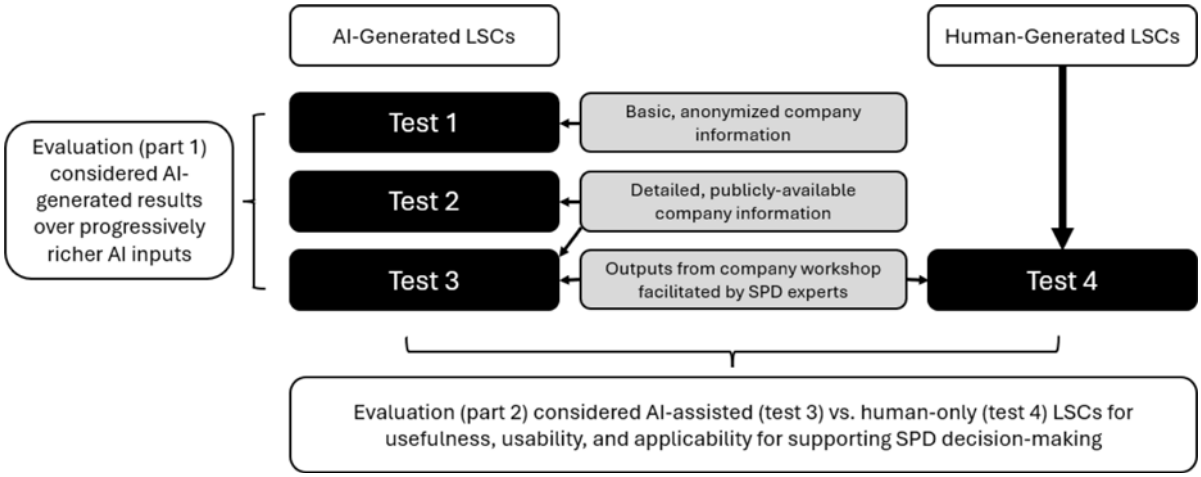


Figure 1. Research design and methodology, showing AI-generated LSCs (tests 1-3) and human-generated LSCs (test 4)

3.1. Test protocol

Each case proceeded through three sequential tests designed to simulate different levels of information availability. In the first test, the AI was given only a short, generic description of the company's industry, size, and basic characteristics, together with a high-level explanation of the LEASA method. In the second test, the AI was first instructed to generate a comprehensive company profile based on publicly available materials such as sustainability reports, websites, press releases, and news articles, covering the company's business model, operations and workforce, supply chain, and products and services, and then use this profile to generate a new set of LSCs and SCI rating. The third test built on this foundation by supplementing the company profile with outputs from human-facilitated LEASA workshops conducted in two half-day sessions at each company. Workshops involved mixed groups of 3-6 internal company representatives and 2-4 external SPD experts acting as guides and facilitators, while generating unstructured facilitator notes, digital whiteboard records, and cluster analysis of emergent themes. These participatory inputs offered stakeholder-generated insights and priorities, which challenged the AI to synthesize qualitative knowledge into structured sustainability criteria. Workshop materials used in Test 3 were shared with explicit company permission, contained no personal identifiers, and were processed exclusively through a university-approved AI platform. To ensure the independence of results, the AI

https://doi.org/10.1017/pds.2026.10525 Published online by Cambridge University Press

was instructed to "forget" previous inputs at the start of each test iteration, thereby avoiding contamination across stages.

3.2. Outputs

Each AI test produced a table containing 10 LSCs. Following the structure outlined by Hallstedt (2017) and Watz and Hallstedt (2024), each LSC was assigned an identifier (for example, LSC1 or LSC2) and accompanied by a brief description. To ensure measurability, one to three performance indicators were included for each criterion, along with an indication of the relevant life cycle phases. The criteria were further classified according to the sustainability dimension they addressed—ecological, social, or economic. Finally, each LSC specified potential sources of assessment data, ranging from utility bills and supplier audits to company reporting systems. Following this, the AI tool also generated an indicator-specific Sustainability Compliance Index (SCI), a four-level qualitative scale ranging from 1 (non-compliant/high risk) to 9 (strategic alignment/best practice), following guidance by Watz and Hallstedt (2024). The SCI ratings provide users with a methodical way to apply the LSCs to product concepts, enabling structured evaluation and comparison of sustainability performance. To mitigate the risk of false precision in its numerical levels, the analysis interprets the SCI scores as qualitative ranges and focuses on relative differences (e.g., low–medium–high alignment) rather than fine-grained scoring.

3.3. Evaluation and analysis

Four versions of the LSC outputs were evaluated and compared: the three AI-generated versions and one human-generated version. The first three tests reflected progressively richer informational inputs for AI-assisted LSC development, while the fourth represented the human-generated results produced through expert-facilitated LEASA/SF workshops. To compare the outputs across tests and against the expert-led LEASA workshops, qualitative interpretation was combined with a structured comparative framework. The analysis examined the extent of overlap and divergence, identified which LSCs appeared consistently and which emerged only under particular conditions, and considered the specificity of indicators. The analysis examined areas of overlap and divergence, the specificity of indicators, lifecycle coverage, the balance of sustainability dimensions, and the orientation of data sources (external benchmarks, public reports, or internal insights).

To assess the relative merits of the most advanced AI-assisted (Test 3) and human-only (Test 4) outputs, a three-part evaluation was conducted, structured along the three evaluation criteria: usefulness, usability, and applicability:

1. ***Company perspective:*** One of the workshop participants from each company reviewed both outputs, selecting the more useful version and commenting on its practical value.
2. ***Expert perspective:*** Two external SPD researchers familiar with the LEASA, SCI, and SF methods, independently compared the outputs, evaluating methodological rigor, contextual relevance, and potential for future integration.
3. ***AI self-assessment:*** The model was prompted to identify the most relevant version for early-stage SPD decisions and reflect on each version's strengths and limitations.

Risk of bias due to the qualitative evaluation approach with a limited participant sample was considered in the results evaluation.

4. Results

The results section is organized in two parts. First, it examines how AI-generated outputs evolved across three tests (Tests 1–3) as progressively richer information was introduced—from generic industry profiles to company-specific data and participatory workshop inputs. Second, it compares the most advanced AI-assisted outputs (Test 3) with human-generated results (Test 4), drawing on evaluations from company representatives, SPD experts and AI self-reflection to identify patterns of convergence and divergence, usability considerations, and time efficiency.

4.1. Company A – food industry

Company A is a family-owned ice cream manufacturer with in-house production, packaging, and cold-chain distribution. It sources dairy and berries locally but relies on global supply chains for ingredients like cocoa and vanilla. Sustainability is central to its brand, reflected in climate and packaging goals. In

https://doi.org/10.1017/pds.2026.10525 Published online by Cambridge University Press

Test 1, AI outputs focused on familiar sector themes—sustainable sourcing, packaging recyclability, water use, labor practices, and cold-chain efficiency—supported by generic indicators based on external benchmarks such as certification ratings and carbon footprints. Test 2 produced more tailored criteria, introducing renewable packaging transitions, factory ergonomics, climate targets, and product-level GHG tracking. Test 3 incorporated participatory workshop insights, adding context-specific priorities such as taste integrity, market compatibility, supplier collaboration, animal welfare, and transparency. Indicators became more operational (e.g., shelf-life measures, percentage of rain-fed crops, supplier proximity). A summary of results is presented in Table 1.

Table 1. Evolution of LSCs for Company A, showing evolution across increasingly detailed AI inputs

Dimension	Test 1	Test 2	Test 3
LSC Overlap	~6 of 10 LSCs aligned with later tests	~8 of 10 LSCs aligned with Test 3	Introduces new themes (local identity, consumer trust, and stand-alone animal welfare)
Indicator Specificity	Generic (e.g. % recyclable)	Semi-specific (e.g. CO_2e/kg)	Highly specific (e.g. rain-fed crops, shelf life in years)

4.2. Company B – cleantech industry

Company B develops water-treatment technologies and is part of a Scandinavian industrial group with roots in shipping and engineering. Its systems are assembled in-house from globally sourced components and sold internationally across manufacturing, aquaculture, and process-industry sectors. Sustainability is central to its value proposition, particularly the elimination of harmful biocides and improvements to worker health and safety. In Test 1, AI criteria reflected generic industrial concerns—energy efficiency, packaging waste, recyclability, and supplier labor rights—supported by broad indicators and a simplified upstream–operations–downstream lifecycle model. Test 2 began to capture the company's distinct value proposition, introducing biocide-free treatment, health and safety, and customer cost savings, with more specific indicators (e.g., % additive-free systems, fluid-disposal reductions) and expanded lifecycle coverage including design and assembly. Test 3 produced the most operationally grounded criteria, adding biological control, bacterial risk management, toxic-emission concerns, and hazardous-component oversight. Social and governance dimensions deepened through customer empowerment and supply-chain resilience. Indicators reached high specificity (e.g., mean time between failures, traceability scores, downtime costs), and lifecycle coverage extended to installation, use, and upgradeability. A summary of results is presented in Table 2.

Table 2. Evolution of LSCs for Company B, showing evolution across increasingly detailed AI inputs

Dimension	Test 1	Test 2	Test 3
LSC Overlap	~5 shared themes with later tests	~7 shared themes with Test 3, with added specificity	Most refined and prioritized set
Indicator Specificity	Moderate	Improved (e.g., % biocide-free)	Highly specific (e.g., Mean Time Between Failures, traceability index)

4.3. Company C – industrial machinery industry

Company C manufactures construction equipment with global production and supply chains and is part of a multinational industrial group. Its products are sold worldwide through dealerships and direct contracts, and the company is investing in sustainability-related innovation, including electrification and fuel-efficiency technologies. In Test 1, AI generated broad manufacturer-oriented criteria—emissions reduction, energy efficiency, responsible sourcing, labor rights, hazardous substances, and waste management—supported by abstract, compliance-oriented indicators and a generic upstream–operations–downstream lifecycle model. Test 2 aligned more closely with Company C's portfolio and innovation agenda, introducing criteria related to electrification, fuel efficiency, and global supply-chain management. Indicators became more concrete (e.g., % electric models sold, Scope 1–2 emissions per unit, supplier code-of-conduct coverage), and lifecycle mapping reflected company control over

https://doi.org/10.1017/pds.2026.10525 Published online by Cambridge University Press

design, production, and distribution. Test 3 marked the biggest shift, moving from external compliance to internal priorities. New criteria included refurbishment tracking, diversity targets, and end-of-life traceability, with highly specific indicators such as risk-observation frequency, refurbishment economics, and logistics optimization. Lifecycle coverage expanded to include end-of-life and sub-supplier tiers, and circularity emerged as a central theme supported by take-back and reuse programs. A summary of results is presented in Table 3.

Table 3. Evolution of LSCs for Company C showing evolution across increasingly detailed AI inputs

Dimension	Test 1	Test 2	Test 3
LSC Overlap	~6 of 10 LSCs aligned with later tests	~8 of 10 LSCs aligned with Test 3; added specificity	Full alignment with stakeholder priorities;
Indicator Specificity	Generic indicators (e.g., "% of spend covered by conflict minerals policy")	Company-specific indicators (e.g., "% electric models sold")	Operationally grounded indicators (e.g., "# of risk observations")

4.4. Company D – aerospace industry

Company D designs and manufactures aerospace components, often using advanced processes such as additive manufacturing (AM), and is part of a multinational engineering group. Its operations span multiple facilities and global supply chains, serving both civil and defense aviation markets. Sustainability is increasingly central in this sector, driven by regulatory pressures and competitiveness, with priorities including responsible sourcing, waste reduction, lightweighting, and repairability. In Test 1, AI generated criteria anchored in external standards (e.g., double-materiality frameworks), covering sectoral "must-haves" such as responsible sourcing, hazardous substances, worker safety, energy efficiency, waste reduction, and reliability. Indicators were abstract, and lifecycle mapping followed a generic upstream–operations–downstream model. Test 2 aligned more closely with Company D's technologies and strategy, referencing AM, lightweighting, and supplier code-of-conduct requirements. Indicators reflected operational realities such as AM energy use and supplier compliance, and lifecycle coverage expanded across supply chains, manufacturing, and product use. Test 3 produced the most context-specific outputs, incorporating participatory insights to introduce criteria on buy-to-fly ratios, AM powder safety, repair compatibility, and waste quality for reuse. Indicators became highly granular (e.g., refurbishment economics, risk reporting, recycled waste revenues), and lifecycle coverage became comprehensive, with strong emphasis on repair, reuse, and circularity. Social criteria also became more operational—such as conflict-free sourcing and workplace safety—while others were deprioritized. A summary of results is presented in Table 4.

Table 4. Evolution of LSCs for Company D, showing evolution across increasingly detailed AI inputs

Dimension	Test 1	Test 2	Test 3
LSC Overlap	~6 of 10 LSCs aligned with later tests	~8 of 10 LSCs aligned with Test 3	Overlaps on core themes, but introduces new themes like diversity, refurbishment, and traceability
Indicator Specificity	Generic sector indicators (e.g., % recycled content)	More tailored (e.g., buy-to-fly, AM powder sourcing, lightweighting R&D)	Highly specific, measurable (e.g., % repairs using AM, waste reuse rates)

4.5. Comparative analysis

To understand how AI-generated sustainability criteria evolve and compare to human-led approaches, this section synthesizes patterns across four company cases. It highlights both the progression of AI outputs with richer inputs and the trade-offs between automation and participatory design.

https://doi.org/10.1017/pds.2026.10525 Published online by Cambridge University Press

4.5.1. Evoluton of AI-assisted LSCs

Across all cases, a consistent trajectory emerged as information richness increased: generic, standards-driven LSCs in Test 1; company-specific criteria in Test 2; and stakeholder-informed, operationally grounded priorities in Test 3. Within this shared progression, several cross-case dynamics were evident.

Overlap and divergence. Core sustainability themes—emissions, energy efficiency, materials, responsible sourcing, and worker safety—appeared across all companies. Divergence emerged only in Test 3, when participatory data introduced sector- and context-specific nuances. For example, animal welfare and taste integrity surfaced only for the food manufacturer; biological control and toxic-water risks became prominent in the cleantech case; refurbishment economics and diversity targets emerged for the machinery company; and AM powder safety and repairability appeared only in aerospace. These variations show that AI alone does not surface latent priorities without stakeholder inputs.

Shifts in data orientation. Across cases, AI moved from externally anchored criteria (Test 1, grounded in standards and certifications) to company-specific disclosures (Test 2) and finally to internal operational knowledge and stakeholder judgments (Test 3). This indicates that the type of information matters as much as the amount: qualitative, participatory inputs consistently redirected AI's focus toward more contextually relevant criteria.

Evolution of focus. AI's framing became more strategic and operational as inputs deepened. Early outputs emphasized compliance; mid-stage outputs aligned with company strategies and market positioning; and final outputs incorporated nuances such as supplier dynamics, organizational culture, customer empowerment, and circularity. Social aspects shifted from baseline labor rights to deeper themes of inclusion, workforce well-being, and community trust, depending on the case.

Cross-sector contrasts and exceptions. Consumer-facing sectors emphasized trust, quality, and supply-chain transparency, whereas industrial sectors focused on system performance, operational risk, and circularity. Aerospace remained an exception: despite strong alignment with general manufacturing patterns, it showed uniquely high emphasis on process-level safety, repair loops, and regulatory alignment, reinforced by participatory data. Despite these contrasts, all cases exhibited the same macro-pattern: AI became most valuable when guided by participatory insights, not when operating on generic or publicly available information alone.

4.5.2. Comparing AI-assisted outputs to human-only results

Evaluations by company representatives, SPD researchers, and the AI model (Table 5) reveal differences between AI-assisted outputs (Test 3) and human-generated outputs (Test 4).

Interpretation of "strategic." SPD researchers were split in their preferences. One SPD researcher consistently favored the AI-assisted version, citing its holistic coverage and structured phrasing as advantageous for strategic decision-making. In contrast, the second SPD researcher preferred the human version, emphasizing its principle-based orientation and forward-looking vision. This divergence underscored that "strategic" was interpreted differently: breadth and comprehensiveness for one researcher, versus normative alignment and creative tension for the other. In the same vein, Company C viewed strategic value as a balance between ambition and feasibility, favoring a hybrid model that combined AI's structured depth with the human version's contextual flexibility, cautioning that overly "crisp" AI-generated SCI ratings could arbitrarily limit criteria applicability and flatten nuance.

Trade-offs between breadth and contextual relevance. AI-assisted outputs consistently introduced a more holistic range of sustainability themes—such as supplier traceability, governance, and lifecycle waste reduction—while human-generated versions concentrated on priorities most critical for near-term decisions. For example, for Company B, AI introduced governance and supplier traceability as systemic priorities, while the human version focused on biocide-free treatment and worker safety—issues tied directly to the company's core technology and immediate customer needs. For Company D, AI emphasized additive manufacturing powder safety, refurbishment economics, and waste quality for reuse, whereas the human version prioritized climate impact and regulatory compliance, reflecting sector-specific imperatives.

Time efficiency a clear advantage of AI-assisted outputs. While expert-led workshops required multiple hours of preparation and facilitation—often spanning one or two full days—AI generated structured criteria sets in under an hour. For instance, Test 2, which relied solely on publicly available company data, produced a near-complete output in approximately 30 minutes. This speed offers a compelling benefit for organizations facing resource constraints, though it comes with trade-offs in contextual relevance and stakeholder engagement.

https://doi.org/10.1017/pds.2026.10525 Published online by Cambridge University Press

Usability considerations. Representatives valued clarity and simplicity for team engagement, even at the cost of strategic depth. For example, Company A warned that overly complex criteria risk "analysis paralysis," while Company D noted that legal compliance thresholds, though imperfect, provide actionable starting points. One SPD researcher highlighted that AI-generated outputs, while more comprehensive, introduced greater cognitive load due to the breadth of sustainability aspects covered. In contrast, human-generated outputs were more streamlined, with clearer patterns and fewer dimensions—potentially increasing user-friendliness but compromising a more holistic view of product sustainability. Taken together, these findings indicate that neither approach fully satisfied both strategic ambition and operational usability. While AI-assisted outputs offered breadth and structure, human-generated versions provided contextual relevance and ease of application.

Table 5. Preferences between AI-generated (Test 3) vs human-generated (Test 4) LSCs and SCI ratings when prompted, "Which version do you think is best for navigating SPD decision-making?"

Company	Company Representative	SPD Researchers	AI
A	Human	Split	Human
B	AI-Assisted	Split	AI-Assisted
C	Human	Split	AI-Assisted
D	Human	Split	AI-Assisted

Across cases, preferences reflected predictable biases: company representatives tended to favor the human-generated LSCs they had co-developed, while the AI model preferred its own output. Yet these biases also reveal a broader dynamic: human evaluators anchored their judgments in contextual relevance and legitimacy, whereas the AI optimized for structural completeness and internal consistency. Taken together, the pattern suggests not a competition between approaches but a complementary pairing—AI excels at widening the solution space, while humans determine which solutions are meaningful, credible, and actionable in practice.

5. Concluding discussion

This study found that while AI-generated criteria were technically coherent, they remained generic until enriched with participatory inputs—highlighting a key limitation in AI's ability to interpret context. Expertise in sustainability is not just cognitive but socially embedded, requiring interpretive framing and consensus-building. This supports Billi and Labraña's (2025) argument that AI lacks contextual adaptation, strategic foresight, and social legitimacy. At the same time, the study extends their work by illustrating "hybrid expertise," where AI contributes structure and speed, but human input ensures relevance and legitimacy. Once company-specific and stakeholder-derived data were introduced, AI outputs became strategically aligned and operationally grounded. These findings also reinforce Corsini et al. (2025), who emphasize AI's potential to support collaborative KPI definition—while adding that the quality of such collaboration depends on human framing. These findings directly address the research question by showing that while AI can accelerate criteria development, it cannot replace the participatory processes essential for effective sustainability decision-making. These dynamics extend beyond LEASA: most sustainability design tools require structured exploration (which AI accelerates) and socially mediated prioritization (which AI cannot replace). The findings therefore inform a broader class of AI-supported workflows, indicating that hybrid configurations—AI for scope and synthesis, humans for strategic framing and legitimacy—are likely to be transferable across SPD methods. Four additional insights deepen this conclusion:

First, ***the value of LEASA lies not only in its outputs but in the process itself***—bringing diverse actors together to build shared understanding. Replacing workshops entirely with AI risks losing this learning dimension, which Watz & Hallstedt (2024) identify as essential for building trust, contextual relevance, and stakeholder legitimacy in sustainability decision-making. Across both company and expert evaluations, participants consistently emphasized that the LEASA workshop process fosters critical reflection and contextual prioritization—benefits that AI alone cannot replicate. This ties directly into the themes of learning, trust-building, and stakeholder legitimacy, while reinforcing the limits of AI in completely replacing human-led processes.

https://doi.org/10.1017/pds.2026.10525 Published online by Cambridge University Press

Second, ***AI can offer targeted support in several areas***: (1) jumpstarting the process by presenting companies with a broad, holistic set of sustainability issues to ensure they do not overlook critical topics or narrow their focus too early; (2) quickly generating a starting list of SCI thresholds to promote consistency and reduce ambiguity; and (3) closing the "translation gap" between workshop notes and actionable outputs, accelerating the move from discussion to implementation. This translation gap is not merely technical—it reflects the broader "alignment problem" described by Boncella (2024), where AI systems struggle to interpret human values, ethics, and nuanced preferences. Without human oversight, AI-assisted outputs risk misinterpretation or contextual drift, potentially leading to poor strategic decisions. Moreover, the opaque nature of deep learning models compounds this issue, making it difficult to explain or justify AI-generated recommendations in stakeholder settings.

Third, ***customizations do matter***: company preferences for concise versus descriptive language—and clarity on whether the tool is intended for visionary transformation or short-term operational guidance—should be explicitly integrated into AI prompting. These adjustments could significantly improve usability and help surface the inherent tension between strategic ambition and practical feasibility, key challenges with AI-assisted outputs identified by Billi & Labraña (2025) and Corsini et al. (2025).

Finally, ***time efficiency emerged as a compelling advantage***. AI produced structured outputs in under an hour compared to time intensive workshops, addressing a key bottleneck for SPD methods (Watz & Hallstedt, 2024). Yet this speed comes with trade-offs in contextual relevance and stakeholder engagement—underscoring that hybridization, rather than substitution, offers the most viable path forward.

This study is bounded by its four-company sample within a single national context and by the use of one AI model, limiting generalizability. The evaluation relied on qualitative expert and practitioner judgments rather than standardized performance metrics, and some bias was observed: company representatives favored the human-generated criteria they had co-developed, while the AI tended to prefer its own outputs. Taken together, these limitations suggest caution in interpreting the comparative results but do not alter the insight that AI strengthens—rather than replaces—participatory approaches to developing sustainability criteria.

The challenge ahead is operationalizing this hybridization: designing workflows that preserve the learning value of workshops while leveraging AI for speed, consistency, and scalability. Doing so will require explicit prompting strategies, usability customization, and capability development within organizations. Equally important is addressing AI's "black-box" opacity—the limited visibility into how models interpret context or justify recommendations—which constrains trust and accountability (Schmid, 2023). Future research should move beyond proof-of-concept toward implementation studies—testing hybrid models in diverse sectors, evaluating decision quality, and exploring how AI can support both visionary transformation and short-term operational goals.

Acknowledgement

This research was funded partly by the Knowledge Foundation, Sweden, through the Sustainable Product Innovation for Rewarding Transformation (SPIRIT) Synergy Project (contract 20240015). Sincere thanks to the company participants and to Josefin Lövdahl and Adam Mallalieu for contributing to this research.

References

Ahmad, S., Wong, K. Y., Tseng, M. L., & Wong, W. P. (2018). Sustainable product design and development: A review of tools, applications and research prospects. *Resources, Conservation and Recycling*, 132, 49–61. https://doi.org/10.1016/j.resconrec.2018.01.020

Billi, M., & Labraña, J. (2025). Is AI a functional equivalent to expertise in organizations and decision-making? Opportunities and pitfalls for AI in the context of just transitions. *Frontiers in Artificial Intelligence*, 8, 1571698. https://doi.org/10.3389/frai.2025.1571698

Boncella, R. (2024). AI and management: Navigating the alignment problem for ethical and effective decision-making. *Issues in Information Systems*, 25(4), 194–204. https://doi.org/10.48009/4_iis_2024_116

Brambila-Macias, S. A., & Sakao, T. (2021). Effective ecodesign implementation with the support of a lifecycle engineer. *Journal of Cleaner Production*, 279, 123520. https://doi.org/10.1016/j.jclepro.2020.123520

Corsini, F., Iannuzzi, T., Fundoni, M., & Frey, M. (2025). Can (A)I help you measure your circularity? An AI-driven and collaborative framework for strategic KPI definition. *Journal of Cleaner Production*, 524, 146488. https://doi.org/10.1016/j.jclepro.2025.146488

https://doi.org/10.1017/pds.2026.10525 Published online by Cambridge University Press

Faludi, J., Hoffenson, S., Kwok, S. Y., Saidani, M., Hallstedt, S. I., Telenko, C., & Martinez, V. (2020). A Research Roadmap for Sustainable Design Methods and Tools. *Sustainability*, 12(19), 8174. https://doi.org/10.3390/su12198174

Ghosh, V., Chotia, V., Alghafes, R., Basahel, S., & Nazrul, A. (2025). Bridging AI minds and corporate sustainability—Exploring the role of business model innovation and organizational sensemaking. *Technological Forecasting and Social Change*, 218, 124229. https://doi.org/10.1016/j.techfore.2025.124229

Gupta, A., Tyagi, L., Sisodia, V. S., & Verma, A. (2024). Strategic Framework for Implementing Artificial Intelligence in Organizations from Conceptualization to Practical Execution and Performance Evaluation. *2024 1st International Conference on Sustainable Computing and Integrated Communication in Changing Landscape of AI (ICSCAI)* (pp. 1-8). https://doi.org/10.1109/ICSCAI61790.2024.10866620

Haddud, A. (2024). ChatGPT in supply chains: Exploring potential applications, benefits and challenges. *Journal of Manufacturing Technology Management*, 35(7), 1293–1312. https://doi.org/10.1108/JMTM-02-2024-0075

Hallstedt, S. I. (2017). Sustainability criteria and sustainability compliance index for decision support in product development. *Journal of Cleaner Production*, 140, 251–266. https://doi.org/10.1016/j.jclepro.2015.06.068

Hallstedt, S. I., Villamil, C., Lövdahl, J., Nylander, J.W. (2023). Sustainability Fingerprint – guiding companies in anticipating the sustainability direction in early design. *Sustainable Production and Consumption*, 37, 424–442. https://doi.org/10.1016/j.spc.2023.03.015

Kravchenko, M., Pigosso, D. CA., & McAloone, T. C. (2019). Towards the ex-ante sustainability screening of circular economy initiatives in manufacturing companies: Consolidation of leading sustainability-related performance indicators. *Journal of Cleaner Production*, 241, 118318. https://doi.org/10.1016/j.jclepro.2019.118318

Mallalieu, A., Isaksson Hallstedt, S., Isaksson, O., Watz, M., & Almefelt, L. (2024). Barriers and enablers for the adoption of sustainable design practices using new design methods – Accelerating the sustainability transformation in the manufacturing industry. *Sustainable Production and Consumption*, 51. https://doi.org/10.1016/j.spc.2024.08.023

Marzi, G., & Balzano, M. (2025). Artificial intelligence and the reconfiguration of NPD Teams: Adaptability and skill differentiation in sustainable product innovation. *Technovation*, 145, 103254. https://doi.org/10.1016/j.technovation.2025.103254

Nikolic, M, Bjelica, D. (2025) Digital product success under the microscope: When artificial intelligence in projects helps – and when it hurts. *PLoS One* 20(8), e0331229. https://doi.org/10.1371/journal.pone.0331229

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press. https://www.jstor.org/stable/j.ctt13x0hch

Saidani, M., Cluzel, F., Leroy, Y., Pigosso, D., Kravchenko, M., & Kim, H. (2022). Nexus Between Life Cycle Assessment, Circularity and Sustainability Indicators—Part II: Experimentations. *Circular Economy and Sustainability*, 2(4), 1399–1424. https://doi.org/10.1007/s43615-022-00160-2

Schmid, U. (2023). Trustworthy artificial intelligence: Comprehensible, transparent and correctable. In *Introduction to Digital Humanism* (pp. 151–164). Springer. https://doi.org/10.1007/978-3-031-45304-5_10

Trunk, A., Birkel, H., & Hartmann, E. (2020). On the current state of combining human and artificial intelligence for strategic organizational decision making. *Business Research*, 13(3), 875–919. https://doi.org/10.1007/s40685-020-00133-x

Watz, M., & Hallstedt, S. I. (2024). Depth and Detail or Quick and Easy? Benefits and Drawbacks of Two Approaches to Define Sustainability Criteria in Product Development. In S. Fukushige, H. Kobayashi, E. Yamasue, & K. Hara (Eds.), *EcoDesign for Sustainable Products, Services and Social Systems I*. Springer Nature Singapore (pp. 413–427). https://doi.org/10.1007/978-981-99-3818-6_28

https://doi.org/10.1017/pds.2026.10525 Published online by Cambridge University Press