This paper was presented at the XXXVII ISPIM Innovation Conference, held in Granada, Spain on 07 June to 10 June 2026. ISBN 978-952-65771-2-8.

Faster, Cheaper, Better? What Generative AI Really Delivers in Innovation Foresight

Antje Wild*

University of Applied Sciences Neu-Ulm, Wileystr. 1, 89231 Neu-Ulm, Germany.
E-mail: antje.wild@hnu.de

Hanna Brehm

Gardena Manufacturing GmbH, Hans-Lorenser-Str. 40, 89079 Ulm, Germany.
E-mail: hanna.brehm@husqvarnagroup.com

Carina Volk-Schor

University of Applied Sciences Neu-Ulm, Wileystr. 1, 89231 Neu-Ulm, Germany.
E-mail: carina.volk-schor@hnu.de

Dorothée Stadler

HYVE Innovate GmbH, Schellingstr. 45, 80799 München, Germany.
E-mail: dorothee.stadler@hyve.net

Niclas Kröger

HYVE Innovate GmbH, Schellingstr. 45, 80799 München, Germany.
E-mail: niclas.kroeger@hyve.net

* Corresponding author

Abstract: Innovation foresight is strategically critical yet resource-intensive, limiting its adoption especially among smaller firms. This study examines which stages of the foresight process generative AI can substitute for expert-driven foresight without compromising quality, and with what cost-time trade-offs. We conducted an end-to-end innovation foresight process for a multinational garden equipment firm, comparing outputs from a professional human expert team against those of AI models (ChatGPT, Claude, Copilot-Premium, Copilot-Standard) under two prompting conditions. Four independent expert judges evaluated all outputs blind using stage-specific quality criteria. Results reveal a stage-dependent pattern: AI achieves quality parity with human experts at trend identification and innovation field stages, while humans substantially outperform AI at the scenario development stage, with the largest effects on novelty and organizational specificity. These quality differences must be read alongside dramatic efficiency gains: AI completed the process saving 98.70% of time and 99.27% of cost.

This paper was presented at the XXXVII ISPIM Innovation Conference, held in Granada, Spain on 07 June to 10 June 2026. ISBN 978-952-65771-2-8.

Keywords: Innovation foresight; generative AI; corporate foresight; scenario planning; trend identification; strategic innovation fields; task–technology fit; jagged technological frontier; prompting strategies; human–AI collaboration.

1 Introduction

Despite innovation being a strategic priority for most firms, many organizations struggle to translate innovation ambitions into actionable outcomes. Recent practitioner studies indicate that while more than 80% of companies rank innovation among their top strategic goals, only a small fraction is considered ready to deliver on these ambitions (BCG, 2024). This gap is exacerbated by volatile markets, accelerating technological change, and information overload.

Innovation foresight, encompassing trend identification, scenario development, and the derivation of strategic innovation fields, is a critical upstream input for innovation management under such conditions. Yet, foresight is widely perceived as resource-intense, particularly for small and mid-sized firms.

Generative AI appears to be a promising solution. Its ability to process large volumes of information and detect patterns suggests strong potential for supporting foresight activities. At the same time, firms face substantial uncertainty about where and how AI can be used without compromising quality. Dell'Acqua et al. (2026) demonstrate that AI does not improve performance uniformly: a "jagged technological frontier" separates tasks where AI provides substantial gains from those where it degrades performance. For innovation foresight, a multi-stage process with markedly different cognitive demands at each stage, this insight has direct practical implications. The question is not whether AI can support foresight, but which stages it can support, at what quality level, and at what cost differential.

Two gaps in the current literature motivate this study. First, prior research frequently evaluates AI performance at isolated foresight stages rather than across a complete end-to-end process. Second, quality assessments are rarely integrated with explicit time and cost data, despite their central importance for managerial decision-making. This study addresses both gaps through a field comparison of AI- and human-generated outputs across all three stages of a real foresight project. The study addresses two research questions:

RQ1: How does the quality of AI-generated outputs compare to human-generated outputs across different stages of the innovation foresight process?

RQ2: How do quality, time, and cost trade-offs differ between human-based and AI-based foresight processes?

2 Literature Review

Corporate Foresight as a Strategic Capability

Corporate foresight encompasses "identifying, observing and interpreting factors that induce change, determining possible organization-specific implications, and triggering appropriate organizational responses" (Rohrbeck et al., 2015). Rather than predicting a single future, foresight involves identifying relevant trends and drivers, constructing alternative future scenarios, and deriving strategic implications that inform innovation priorities and resource allocation. We adopt this three-stage structure, trend identification, scenario generation, and the derivation of strategic innovation fields, as our organizing framework.

AI as an Augmentation Tool in Knowledge Work

A growing body of research examines AI as a tool for augmenting knowledge-intensive work. Noy and Zhang (2023) document a 40% reduction in time and 18% improvement in quality for professional writing tasks; Dell'Acqua et al. (2026) report quality improvements of more than 40% for consulting tasks within the AI frontier. However, performance benefits are far from uniform across employees and tasks. Dell'Acqua et al. (2026) introduce the concept of the "jagged technological frontier," demonstrating that tasks of seemingly similar difficulty may differ substantially in whether AI support improves or degrades performance. This insight is grounded by Task-Technology Fit (TTF) theory (Goodhue and Thompson, 1995), which holds that technology generates performance gains only when its functional capabilities align with the demands of the task. Together, these frameworks suggest that AI's contribution to foresight is unlikely to be uniform across stages: tasks characterized by high data volume and pattern recognition fit current AI capabilities well, while tasks requiring implicit knowledge, contextual judgment, and the creation of truly novel ideas may fall beyond the effective frontier.

AI Across the Innovation Foresight Process

Trend identification

AI-based trend identification has received growing empirical attention. Mühlroth et al. (2023) demonstrate that an ML pipeline can condense large document corpora into high-quality signal candidates, with 68% rated by domain experts as highly relevant or novel. Fischer et al. (2023) find that AI recovers approximately 77% of expert-identified influencing factors in automotive foresight. A global survey found that AI is widely used to support horizon scanning and trend clustering. (OECD and WEF, 2025). Given the high data volume and pattern-recognition demands of trend identification, TTF theory suggests a relatively strong fit between current AI capabilities and identifying trends.

Scenario development

Scenario development is the most extensively studied AI foresight application. Spaniol and Rowland (2023) argue that AI-generated scenarios provide useful "raw material" but cannot replace expert scenarists, particularly for organizational specificity and the cognitive reframing of decision-makers. Bechthold (2025) provides the theoretical mechanism: AI systems learn from content already available online, "enshrining pictures of the past" even under future-oriented prompts. This pattern structurally constrains novelty in AI-generated scenarios. Thus, it is likely that humans will outperform AI models in this stage, especially when measuring whether scenarios are truly novel.

Derivation of strategic innovation fields

This final stage, identifying domains for future innovative activity, has received the least research attention. Adjacent evidence from product ideation is instructive: Girotra et al. (2023) and Boussioux et al. (2024) find that AI outperforms humans on commercial

This paper was presented at the XXXVII ISPIM Innovation Conference, held in Granada, Spain on 07 June to 10 June 2026. ISBN 978-952-65771-2-8.

quality but that human-generated ideas show greater novelty. Si et al. (2024) find that LLM-generated research ideas are rated as more novel than those of time-constrained human experts but show a deficit on feasibility. The derivation of innovation fields combines evaluative judgment about strategic fit with generative identification of novel opportunity spaces, making its frontier position ambiguous a priori.

Cost and time comparison

While a significant body of research has focused on comparing AI and human output quality, few studies have explicitly calculated time and cost comparisons. It is generally acknowledged that AI will save time and cost (Boussioux et al. 2024), but exact calculations remain rare. We hypothesize that AI models will significantly outperform humans on both dimensions, leading to cost and time savings of 95% or more.

3 Methods

Research Design

This study applies a comparative research design to evaluate how generative AI performs across the main stages of an innovation foresight process: trend identification, scenario axis selection, scenario development, and the derivation of strategic innovation fields (Rohrbeck et al., 2015; Schoemaker, 1995). Two parallel approaches were examined: a manual expert-driven foresight process conducted at a leading manufacturer in the garden equipment industry, and an AI-supported process executed independently by four researchers using different generative AI systems (ChatGPT 4o, Claude Sonnet 4.5, Microsoft 365 Copilot-Premium (GPT-4o), Copilot-Standard (GPT-4o)).

Data Collection

In the first stage, the corporate innovation management team conducted expert interviews and desk research to identify a broad pool of environmental and industry developments before selecting 20 high-priority trends. The AI process generated 20 trends under one-shot and iterative prompting using four different AI systems.

In the second stage, scenario axes were selected to build a 2×2 scenario grid following the GBN matrix approach (Schwartz, 1991; Bishop et al., 2007; Cordova-Pozo and Rouwette, 2023). In the manual process, two high-impact and high-uncertainty drivers were selected through expert discussions supported by external foresight consultants. In the AI-supported process, each researcher prompted their AI tool to identify two uncertainty-intensive and strategically relevant drivers.

In the third stage, four future scenarios were developed — one for each quadrant of the 2×2 grid. Manual scenarios were generated in cross-functional groups during a workshop with internal experts and external foresight consultants. AI-generated scenarios were created via one-shot prompting and iterative refinement, following the same template.

In the final stage, strategic innovation fields were derived. In the manual process, internal workshops translated scenario implications into opportunity spaces through individual idea generation, clustering, and structured templates. In the AI-supported process, researchers prompted their tools to derive innovation fields from the previously generated scenarios, yielding parallel sets of fields.

Evaluation Procedure

All outputs were transformed into standardized evaluation sheets and subjected to a coding, anonymization, and randomization protocol to prevent evaluators from inferring their origin. Four independent foresight experts who had not participated in the data generation stage evaluated each individual output.

Trends were assessed on four five-point Likert scales: Dissemination Speed (Rotolo et al., 2015); Strategic Relevance for the company (Blechschmidt, 2024); Trend Maturity, since early detection enables timely opportunity capture and risk mitigation (Blechschmidt, 2024); and Uncertainty (Schoemaker, 1995). Uncertainty was scored such that higher values indicate greater uncertainty, which is more valuable for scenario axis selection.

Scenario axis selections were evaluated on Uncertainty as the primary selection criterion (Schoemaker, 1995), novelty because scenarios should provide new perspectives (Girotra et al., 2023; Cordova-Pozo and Rouwette, 2023; Van der Heijden, 2005), and Organizational Fit since axes must reflect organizational priorities (Van der Heijden, 2005). Both the X and Y axes were evaluated individually, as well as the overall fit of the axis combination (Schoemaker, 1995).

Scenarios were rated on six criteria: Plausibility, to ensure the scenario represents a possible future (Schoemaker, 1995; Kosow and Gaßner, 2008; Cordova-Pozo and Rouwette, 2023); Organizational Specificity, since low scores indicate limited organizational relevance (Schoemaker, 1995); Comprehensibility (Schoemaker, 1995; Kosow and Gaßner, 2008); Scenario Differentiation, ensuring scenarios describe clearly distinguishable futures (Schoemaker, 1995; Kosow and Gaßner, 2008); Novelty, because AI tends to generate plausible, data-based continuations rather than truly novel solutions (Girotra et al., 2023); and Visualization quality.

Innovation fields were evaluated on Strategic Business Fit, Consumer Relevance, and Innovation Potential as three main selection criteria for evaluating early-stage projects (Aristodemou et al., 2020); Novelty, to test the hypothesis that AI generates less novel ideas (Girotra et al., 2023); and Visualization quality.

Inter-rater reliability was assessed using Krippendorff's Alpha ($\alpha > .70$ across all stages), justifying aggregation of rater scores. All analyses were performed at the level of individual foresight outputs.

Statistical Analysis

Shapiro-Wilk tests confirmed significant departures from normality across criteria and stages, requiring non-parametric analyses throughout. Three planned contrasts were conducted per criterion per stage: (1) Human vs. AI (all AI conditions collapsed, Mann-Whitney U; reported as z and Glass rank-biserial correlation r); (2) between AI models (Kruskal-Wallis H; reported as H(3) and epsilon squared ε^2); and (3) one-shot prompting vs. conversational prompting among AI groups (Mann-Whitney U). Effect size benchmarks: r: small = .10, medium = .30, large = .50; ε^2: small = .01, medium = .06, large = .14. Stage 2 outputs (n = 2 per condition) are presented descriptively only. Exact significance was used throughout where computationally feasible. For Kruskal-Wallis between-model comparisons at Stage 1 (N = 160, four balanced groups of n = 40), asymptotic significance is reported, which provides a reliable approximation at this sample size. Time and cost data derive from project records and invoices; AI costs were based on the time spent prompting and on the subscription cost. Only percentage ratios are reported to protect commercial confidentiality.

4 Results

Table 1 presents Human vs. AI contrast results across Stages 1, 3, and 4. Stage 2 is not included due to the low number of cases (n=2). Between-model and prompting condition findings are reported in the text where significant.[1]

Table 1 Human vs. AI Output Quality Across Foresight Stages

Criterion	*Mdn(H)*	*Mdn (AI)*	*z*	*r*	*Direction*
Stage 1: Trend Identification (n = 180; 20 outputs per condition)					
Strategic Relevance	2.88	3.00	-.96	.13	Parity
Uncertainty[a]	3.00	3.00	-.35	.05	Parity
Trend Maturity	2.00	2.25	-1.09	.14	Parity
Dissemination Speed	2.00	2.00	-1.46	.19	Parity
Stage 3: Scenario Development (n = 36; 4 outputs per condition)					
Plausibility	4.50	3.75	-1.74†	.53	Human > AI
Organizational Specificity	4.88	4.00	-3.21***	.94	Human > AI
Comprehensibility	4.75	4.00	-2.48*	.67	Human > AI
Scenario Differentiation[b]	5.00	4.00	—	—	Human > AI
Novelty[b]	4.75	3.00	-3.45***	1.00	Human > AI
Visualization[c]	4.88	1.88	-3.04***	.94	Human > AI
Stage 4: Innovation Fields (n = 72; 8 outputs per condition)					
Strategic Business Fit	3.38	3.50	-.21	.04	Parity
Consumer Relevance	3.13	3.38	-.41	.09	Parity
Innovation Potential	3.88	3.25	-1.41	.29	Parity
Novelty	3.13	3.00	-.45	.09	Parity
Visualization	2.63	2.63	-.24	.05	Parity

Note. Mdn = median; r = rank-biserial correlation (Mann-Whitney U). Parity = no significant difference (p > .10). †p < .10. *p < .05. **p < .01. ***p < .001.

[a] Uncertainty scored such that higher values indicate greater uncertainty; high-uncertainty trends are more valuable as scenario drivers.

[b] All four human scenario sets received the maximum possible score (Mdn = 5.00, SD = 0.00), precluding computation of a test statistic. The rank-biserial correlation equals 1.00 by definition, as every human output outranked every AI output on this criterion. Novelty independently yielded r=1.00 from a valid Mann-Whitney U test.

[c] Models were explicitly asked to produce visual outputs; scores reflect actual image generation capability and client-readiness.

[1] Additional results of the paper, prompts and sample outputs from the different stages can be found under this link: https://www.researchgate.net/profile/Antje-Wild/research

Stage 1: Trend Identification

No significant differences were found between human and AI outputs on any of the four trend criteria (all $p > .10$, $r = .05$ -.19), indicating quality parity across all dimensions. AI achieved these results at 0.83% of human process time and 0.90% of human cost. Significant between-model differences emerged on Dissemination Speed ($H(3) = 9.05$, $p = .029$, $\varepsilon^2 = .04$), Trend Maturity ($H(3) = 8.16$, $p = .043$, $\varepsilon^2 = .03$), and Uncertainty ($H(3) = 10.72$, $p = .013$, $\varepsilon^2 = .05$), all small effects. ChatGPT identified the highest-uncertainty trends, which is most valuable for scenario development, while Claude scored lowest across all three criteria. Prompting condition produced a significant effect only on Trend Maturity ($z = -2.04$, $p = .041$, $r = .18$), with one-shot prompting yielding marginally more mature trends.

Stage 2: Scenario Axis Selection (Descriptive)

Due to the small number of outputs per condition ($n = 2$), inferential testing was not conducted. A consistent tradeoff emerged: axis pairs rated highest on Uncertainty and Novelty scored lowest on Organizational Fit, and vice versa. Human experts ranked first on Novelty but last on Organizational Fit.

Stage 3: Scenario Development

A striking reversal from Stage 1 emerged. Human experts significantly outperformed AI on all six criteria. The largest effects were on Novelty ($z = -3.45$, $p < .001$, $r = 1.00$), Organizational Specificity ($z = -3.21$, $p < .001$, $r = .94$), and Visualization ($z = -3.04$, $p < .001$, $r = .94$), with additional significant effects on Comprehensibility ($z = -2.48$, $p = .028$, $r = .67$) and a marginal effect on Plausibility ($p = .092$, $r = .53$). Human scenario sets were unanimously rated as maximally distinctive ($Mdn = 5.00$, zero variance) compared with AI scenario sets ($Mdn = 4.00$), consistent with homogenization effects documented in AI-generated creative outputs (Dell'Acqua et al., 2026; Kozachek, 2026). Conversational prompting produced significantly more plausible scenarios than one-shot ($z = -2.43$, $p = .015$, $r = .50$). AI completed scenario development at 2.44% of human time and 0.84% of human cost. Significant between-model differences emerged for Visualization ($H(3) = 21.66$, $p < .001$, $\varepsilon^2 = .67$), Scenario Differentiation ($H(3) = 14.91$, $p = .002$, $\varepsilon^2 = .43$), and Novelty ($H(3) = 9.39$, $p = .025$, $\varepsilon^2 = .23$). ChatGPT produced the highest-quality visual outputs; Claude produced the lowest, reflecting its limited image generation capability.

Stage 4: Innovation Fields

Stage 4 results mirror Stage 1: no significant differences were found on any criterion (all $p > .15$, $r = .04$-0.29). Significant between-model differences emerged for Consumer Relevance ($H(3) = 12.30$, $p = .006$, $\varepsilon^2 = .16$), with Claude ranking highest, and Visualization ($H(3) = 51.72$, $p < .001$, $\varepsilon^2 = .81$), driven by Claude's systematic failure to produce visual outputs. No prompting effects were significant (all $p > .30$).

AI achieved this parity at 0.63% of human time and 0.47% of human cost.

Cross-Stage Pattern

Taken together, the results reveal a non-monotonic pattern: parity at Stage 1, a substantial and consistent human advantage at Stage 3, and restored parity at Stage 4. The human

This paper was presented at the XXXVII ISPIM Innovation Conference, held in Granada, Spain on 07 June to 10 June 2026. ISBN 978-952-65771-2-8.

advantage is concentrated on generative criteria requiring creative judgment and tacit knowledge: Novelty, Organizational Specificity, and Comprehensibility. The human advantage at Stage 3 does not propagate to Stage 4 at a statistically detectable level, suggesting AI-generated scenarios are sufficient inputs for generating innovation fields of equivalent quality. On average, AI reduces person hours needed by 98.70% and cost by 99.27%.

5 Discussion

A Stage-Dependent Jagged Frontier in Innovation Foresight

Our findings confirm that the jagged technological frontier (Dell'Acqua et al., 2026) applies within a single innovation foresight process: trend identification, scenario development, and innovation field generation sit at different positions relative to the frontier, and AI performance tracks that variation systematically. Where tasks are predominantly analytical and pattern-based, AI produces outputs of comparable quality to human experts. In the area of scenario generation, where tasks demand contextual judgment, tacit organizational knowledge, and creative originality, human experts substantially outperform AI across all criteria ($r = .53$-1.00). This is consistent with Task-Technology Fit theory (Goodhue and Thompson, 1995), corroborated by Fischer et al.'s (2023) finding that AI adequately identifies influencing factors but falls short on integrative judgment tasks, and aligns with hybrid AI-expert frameworks in the foresight literature (Geurts et al., 2022; Spaniol and Rowland, 2023; Shinkle et al., 2026).

AI Raises the Floor but Lowers the Ceiling, Especially on Novelty and Organizational Specificity

The practical significance of these findings sharpens when time and cost data are considered. AI completed these tasks at 1.30% of the time needed and 0.73% of the cost. For organizations currently conducting no foresight due to resource constraints, common among smaller firms (Adegbile et al., 2017), AI-generated outputs at less than 2% of expert cost represent a substantial improvement over the status quo. Beyond efficiency, AI-based trend identification offers a structural advantage in objectivity: while AI is biased through its training data, unlike human expert teams, it applies no personal preferences, organizational affiliations, or availability constraints that can distort trend selection. It should be noted, however, that AI carries different rather than absent biases: training data skews toward technology topics and well-documented developments (Kozachek, 2026; OECD and WEF, 2025), meaning AI and human approaches offer complementary rather than equivalent coverage.

On novelty and organizational specificity, the ceiling limitations are clear. Novelty produced the largest human advantage ($r = 1.00$) and Organizational Specificity the second largest ($r = .94$) at Stage 3. Across all prompting and model conditions, AI-based scenarios follow very similar patterns, while humans explore much broader options. This aligns with the frequently cited structural limitation of AI, basing its results on dominant patterns in the available training data (Bechthold, 2025).

Si et al. (2024) find AI generates more novel ideas than time-constrained human experts where comprehensive public literature coverage exists. This suggests AI novelty is conditional on whether relevant knowledge is publicly documented. For organization-specific strategic foresight, it structurally is not. Notably, this project was based in the

B2C environment where more accessible information exists online than in the B2B environment.

This also explains the difference to the findings by Kozachek (2026), who did not find structural differences when comparing AI and human based scenarios. The scenarios in that study were based on broader topics with publicly available data focusing on topic such as the future of European education.

A critical qualification emerges from Stage 4. The human advantage in scenario development does not propagate to innovation field generation at a statistically detectable level. This introduces an important contingency: where the goal is a set of adequate innovation fields, AI-generated scenarios may suffice; where the goal is strategic provocation in expert workshops, human expertise is important. Again, this finding might be limited to the B2C field, due to constraints on available information in specialized B2B fields.

Implications for Innovation Management Practice

The combined evidence supports a differentiated delegation logic. For trend identification, the case is unambiguous: quality is equivalent to human expert output at a fraction of cost. A practical caveat applies: because AI identifies trends already documented in its training corpus, supplementary human research, e.g., through trend receiver or lead user interviews remains advisable.

For scenario development, human involvement is most justified where process value, shared mental model-building, strategic provocation, and organizational alignment, is the primary objective. Rather than using AI as a first draft, which risks anchoring the creative process in dominant industry patterns and producing content practitioners experience as derivative. AI outputs may be more productively deployed as an assumption-surfacing mirror: presenting AI's default futures to workshop participants as a provocation to identify and challenge prevailing narratives. With regard to innovation fields, it appears to also be justified to outsource their identification to AI. However, as these fields are the basis on which companies will assign resources, a hybrid approach remains advisable.

Democratizing Strategic Foresight

The dramatic cost reduction documented here fundamentally alters the current calculus around strategic foresight. Firms that previously lacked the resources to conduct foresight projects can now access AI-generated trend scans and innovation field analyses at negligible cost. The one-shot prompting results further support this democratization potential: quality comparable to conversational prompting was achieved with a single structured prompt across most criteria, suggesting that well-designed prompt templates rather than deep AI expertise may be sufficient for organizations entering foresight practice for the first time. This represents a structural shift in who can afford to engage in strategic foresight, with particular relevance for smaller and mid-sized firms that Adegbile et al. (2017) identify as most disadvantaged by the resource intensity of conventional foresight processes.

Beyond the managerial implications for firms, the implications for foresight consulting practice are equally significant and warrant direct acknowledgment. Conventional foresight projects have frequently been executed by consultants who provided content expertise: conducting trend research, facilitating scenario workshops, and deriving strategic implications. The findings presented here suggest this model requires substantial reconfiguration. For the analytical stages where AI achieves parity, the value proposition of consultants as content producers is under direct pressure. The emerging role is rather one of process facilitation and quality assurance: guiding organizations through effective AI-human collaboration, designing prompting strategies that extract organizationally relevant outputs, conducting critical quality control at each stage, and contributing human creative judgment at the scenario development stage where AI demonstrably falls short. Rather than replacing foresight consulting, AI shifts its center of gravity from content production toward process design, methodological expertise, and the irreducibly human capacity for creative, context-sensitive strategic thinking.

Process Value, Trust, and Human Capital

Our quality criteria assess outputs but not the process through which they are produced, a distinction with strategic weight. Scenario planning derives value not only from scenario content but from the shared mental model-building, organizational alignment, and cognitive reframing that occur when expert teams engage collaboratively with uncertain futures (Wack, 1985). An AI-generated process bypasses these dynamics. Managers not involved in generating foresight outputs are less likely to trust or act on them, an effect likely amplified where organizations have not yet undergone the cultural change required to accept AI as a legitimate strategic partner. This lack of trust is not completely unjustified: AI does not reduce the competence requirement for foresight practice. Instead, it may even increase it. In conventional foresight, the structured process and collective judgment of assembled stakeholders provide error-correction mechanisms that protect against individual misjudgment. When AI replaces these elements, responsibility shifts to the individual practitioner without the scaffolding that group deliberation provides.

This risk is heightened by the lack of transparency: AI-generated outputs cannot be traced to specific data sources, limiting their auditability and defensibility in governance contexts. If a foresight process is conducted without the involvement of other stakeholders, this lack of transparency might quickly lead to a lack of trust in the resulting outputs.

The case for AI in foresight is therefore strongest when it augments rather than replaces human participation: handling volume-intensive identification work while preserving human involvement in the synthesis and sense-making stages where both strategic and developmental value are concentrated.

Limitations

Several limitations bound the interpretation of these findings. Our quality criteria assess observable output characteristics but not the participatory and sense-making dimensions of foresight central to its strategic value (Spaniol and Rowland, 2023; Van der Heijden, 2005). Future research should analyze to what extent practitioners are willing to use AI-generated outputs in their strategic decision-making.

Prompting quality was not fully standardized across researchers, though one-shot prompting produced broadly comparable results, suggesting this limitation has limited impact on the main findings (Shinkle et al., 2026). Findings reflect model capabilities at time of data collection (2025-2026); the frontier is advancing and future model generations may alter the performance pattern. The study was conducted in a single mature, low-technology industry; boundary conditions for high-technology sectors remain unexplored.

Future Research

Three directions merit priority. First, whether organizational culture, and specifically a firm's readiness to accept AI as a legitimate foresight partner, moderates trust in AI-generated outputs and managers' willingness to act on them warrants direct empirical investigation. Direct practitioner experience suggests that cultural readiness, rather than output quality alone, may be the binding constraint on AI adoption in corporate foresight. Second, replication across industries with different knowledge profiles, particularly high-technology sectors and niche B2B environments, where training data coverage is denser, would establish the boundary conditions of these findings. Third, replication with model generations beyond those tested here is necessary to assess whether the human advantage in scenario development reflects a structural feature of current AI architecture or a temporary capability gap that newer models will close.

6 References[1]

Adegbile, A., Sarpong, D. and Meissner, D. (2017). "Strategic foresight for innovation management: A review and research agenda". International Journal of Innovation and Technology Management, **14**(4), p. 1750019-1–1750019-34.

Aristodemou, L., Tietze, F. and Shaw, M. (2020). "Stage gate decision making: A scoping review of technology strategic selection criteria for early-stage projects". *IEEE Engineering Management Review*, **48**(2), p. 118–135.

BCG (2024). "Why innovation systems need a reboot." Boston: Boston Consulting Group. https://www.bcg.com/publications/2024/innovation-systems-need-a-reboot, accessed April 2026.

Bechthold, L. A. (2025). "Rendering new futures or enshrining pictures of the past? The double-edged sword of using AI-generated imagery for anticipation." *Journal of Futures Studies*, **29**(4), available on https://jfsdigital.org/rendering-new-futures-or-enshrining-pictures-of-the-past-the-double-edged-sword-of-using-ai-generated-imagery-for-anticipation/.

Bishop, P., Hines, A. and Collins, T. (2007). "The current state of scenario development: An overview of techniques." *Foresight*, **9**(1), p. 5–25.

Blechschmidt, J. (2024). "Quick Guide Trendmanagement." Berlin/Heidelberg: Springer.

[1] AI Usage: Generative AI tools were used both as research instruments in the empirical study and to support manuscript editing. All AI-generated content was reviewed and verified by the authors, who bear full responsibility for the work.

Boussioux, L., Lane, J. N., Zhang, M., Jacimovic, V., & Lakhani, K. R. (2024). "The crowdless future? Generative AI and creative problem-solving." *Organization Science*, **35**(5), p. 1589-1607.

Cordova-Pozo, K. and Rouwette, E. A. (2023). "Types of scenario planning and their effectiveness: A review of reviews". *Futures*, **149**, p. 1-19.

Dell'Acqua, F., McFowland III, E., Mollick, E.R., Lifshitz-Assaf, H., Kellogg, K.C., Rajendran, S., Krayer, L., Candelon, F. and Lakhani, K.R., (2026). "Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality". *Organization Science*, ***37***(2), pp.403-423.

Fischer, D., Heffeter, F., Grothe, S. R., Joachim, V. and Jung, H. H. (2023). "AI in strategic foresight – Evaluation of ChatGPT, BARD and Perplexity". Proceedings of the International Society for Professional Innovation Management (ISPIM), December 11-13, 2023, Salzburg, p. 1-28.

Geurts, A., Gutknecht, R., Warnke, P., Goetheer, A., Schirrmeister, E., Bakker, B. and Meissner, S. (2022). "New perspectives for data-supported foresight: The hybrid AI-expert approach". *Futures & Foresight Science*, **4**(1), p. 1-14, available on https://onlinelibrary.wiley.com/doi/epdf/10.1002/ffo2.99.

Girotra, K., Meincke, L., Terwiesch, C. and Ulrich, K. T. (2023). "Ideas are dimes a dozen: Large language models for idea generation in innovation." *SSRN Electronic Journal Working,* Paper No. 4526071.

Goodhue, D. L. and Thompson, R. L. (1995). "Task-technology fit and individual performance." *MIS Quarterly*, **19**(2), pp. 213–236.

Kosow, H. and Gaßner, R. (2008). „*Methoden der Zukunfts- und Szenarioanalyse: Überblick, Bewertung und Auswahlkriterien*." Werkstatt Bericht Nr. 103. Berlin: Institut für Zukunftsstudien und Technologiebewertung (IZT), p. 1-84. https://www.izt.de/media/2022/10/IZT_WB103.pdf, accessed April 2026.

Kozachek, D. (2026). "Entering the age of hybrid futures: A comparative study of human and GPT-generated scenarios". *Futures*, **176** (2026) 103705, p. 1-16.

Mühlroth, C., Kölbl, L. and Grottke, M. (2023). "Innovation signals: Leveraging machine learning to separate noise from news." *Scientometrics (2023)*, **128**, p. 2649–2676.

Noy, S. and Zhang, W. (2023). 'Experimental evidence on the productivity effects of generative artificial intelligence'. *Science*, **381**(6654), pp. 187–192.

OECD and WEF (2025). "AI in strategic foresight: Reshaping anticipatory governance." Joint White Paper. Paris/Geneva: *OECD Strategic Foresight Unit and World Economic Forum*. Available on https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/11/ai-in-strategic-foresight_82c3a4b1/aa573076-en.pdf.

Rohrbeck, R., Battistella, C. and Huizingh, E. (2015). "Corporate foresight: An emerging field with a rich tradition." *Technological Forecasting and Social Change*, **101**, p. 1–9.

Rotolo, D., Hicks, D. and Martin, B. R. (2015). "What is an emerging technology?" *Research Policy*, **44**(10), p. 1827–1843.

Schoemaker, P. J. H. (1995). "Scenario planning: A tool for strategic thinking." *Sloan Management Review*, **36**(2), p. 25–40.

Schwartz, P. (1991). *The art of the long view*. New York: Doubleday.

Shinkle, G. A., Gujarati, C. and Sharry, P. (2026). "Scenario analysis in the AI era: Redefining human involvement." *Organizational Dynamics*, **55 (1)**, 101197, p. 1-11.

Si, C., Yang, D. and Hashimoto, T. (2024). "Can LLMs generate novel research ideas? A large-scale human study with 100+ NLP researchers." arXiv preprint arXiv:2409.04109, p. 1-94. Available on Arxiv-Website, https://arxiv.org/pdf/2409.04109.

Spaniol, M. J. and Rowland, N. J. (2023). 'AI-assisted scenario generation for strategic planning'. *Futures & Foresight Science*, **5**(2), p. 1-10.

Van der Heijden, K. (2005). *Scenarios: The art of strategic conversation*. 2nd ed. Chichester: Wiley.

Wack, P. (1985). "Scenarios: Shooting the rapids." *Harvard Business Review*, **63**(6), p. 139–150.