Proceedings of the Design Society, Volume 6: DESIGN 2026
https://doi.org/10.1017/pds.2026.10597

Evaluating large language models for technology-oriented searches in engineering design

Vasileios Koutsouvelis[1], Filippo Silipigni[2] and Niccolò Becattini[1,✉]

[1] Politecnico di Milano, Italy, [2] Fondazione Politecnico di Milano, Italy

✉ niccolo.becattini@polimi.it

ABSTRACT: This study evaluates the efficacy of various freely available Large Language Models (LLMs) in conducting semi-automated purpose-oriented technology searches to support design activities as well as Technology Intelligence for innovation management, using a systematic manual search as a baseline for comparison. The case to run the comparison focuses on identifying water purification technologies suitable for mobile systems. The results show that LLMs can target more technologies than human-based searches, reducing time demands and providing wider entry points for additional technology analysis.

KEYWORDS: early design phase, artificial intelligence (AI), supportive technologies, information retrieval, design strategy

1. Introduction

In engineering design, integrating an innovative core technology into a functional system requires identifying complementary components whose specifications align with the constraints the novel device imposes. Candidate technologies must satisfy well-defined design requirements, but understanding their underlying operating principles is equally essential - it expands the design space and surfaces innovation opportunities that a specification-only search would miss. This paper presents such a scenario in humanitarian water treatment: an innovative photovoltaic-driven desalination device that demineralizes water while recovering waste heat imposes specific requirements on both inlet water quality (to ensure optimal device performance) and outlet water composition (to meet potability standards for emergency contexts). The design challenge thus extends beyond the desalination device, demanding systematic identification of pre-treatment technologies to prepare diverse source waters to inlet specifications, and post-treatment solutions to remineralize the output to potable standards. Mapping the operating principles of candidate technologies is equally critical, as it reveals adaptation and extension possibilities that a specification-only search would miss. This scenario exemplifies a recurring class of engineering design problems in which successful system integration depends on comprehensive, requirement-driven knowledge of supporting technologies, which directly conditions the breadth of the explorable design space and the innovation potential of the resulting solution. Conducting such technology landscape analyses presents significant challenges: relevant information is scattered across diverse sources (scientific literature, patents, technical standards, etc.); technologies are described using inconsistent terminology across domains; and the sheer volume of information makes manual review through traditional (academic indexing) engines time-consuming and potentially incomplete. The recent emergence of generative AI tools based on Large Language Models (LLMs) offers a potentially transformative alternative, promising faster information synthesis, cross-domain knowledge integration, and more natural interaction paradigms. However, the reliability, completeness, and accuracy of LLM-generated technology analyses in engineering design contexts remain largely unexplored. Using the water treatment system design as a case study, this paper presents a comparative investigation between human-conducted technology landscape analysis, performed through systematic literature review using Scopus, and results obtained through LLM-based generative AI tools. This comparison evaluates the potential and limitations of LLM-assisted approaches for supporting technology analysis activities in

engineering design, providing empirical evidence on how these emerging tools might complement or augment traditional methods in complex, multi-technology design scenarios.
The next section presents a brief overview of existing technology analysis/exploration approaches and the existing AI-based opportunities to support scholars and professionals to retrieve and analyse knowledge in a specific field/domain. Section 3 presents the method adopted to compare human vs AI performance in technology exploration processes. Section 4 describes how the method has been applied to the case study of water treatment, while Section 5 summarizes and discuss the results of both analysis and provide a comparison of the outcomes.

2. Relevant background

This section addresses two foundational pillars of this research. First, it examines technology intelligence as it relies on the analysis of technology-oriented searches and its diverse methodological approaches. These, despite varying terminologies, represent related activities focused on technology analysis for design and innovation decision-making. Second, it explores how Artificial Intelligence is being applied to support academic research activities, particularly systematic literature reviews. These two domains share significant commonalities: both rely on systematic search and information retrieval processes, and a systematic literature review can itself be conceptualized as a form of knowledge landscape analysis.

2.1. Technology-oriented searches

Technology Intelligence (Veugelers et al, 2010) encompasses systematic methodologies for supporting strategic and operational decision-making through the acquisition, processing, and interpretation of information about technological developments. Within this field, technology-oriented searches constitute the foundational activities through which organizations identify, retrieve, and synthesize relevant technological knowledge from distributed sources. These searches are distinguished from general information retrieval by their specific focus on technical content, innovation trajectories, and capability assessments required for design and development processes. The literature identifies several established approaches to technology-oriented searches, each characterized by distinct objectives and scopes. Technology landscaping (or technology landscape analysis) involves comprehensive, domain-wide information gathering aimed at creating holistic overviews of technological fields, including identification of key technologies, actors, trends, maturity levels, and competitive positions within a specific domain (e.g. Zhou & Carbajales-Dale, 2018). Technology mapping focuses on establishing systematic relationships between technologies and their functional applications, performance characteristics, or design requirements, often involving targeted searches to link technical capabilities with specific use cases or product features (e.g. Ghaffari et al, 2023). Technology scouting represents proactive, forward-looking search activities aimed at early identification of emerging technologies, innovations, and scientific breakthroughs that may offer competitive advantages, typically involving continuous monitoring of research institutions, startups, patent filings, and specialized conferences (e.g. Rohrbeck, 2010). Technology forecasting and technology roadmapping similarly rely on technology-oriented searches to retrieve historical and current data that inform projections about future technological developments and strategic planning timelines. Patent analysis and competitive technology intelligence constitute more specialized search approaches focused on retrieving and analyzing intellectual property data and competitor technological activities to assess innovation patterns, technology trajectories, and potential threats or opportunities (e.g. Cascini et al, 2015).
Despite terminological and methodological variations, these approaches share common characteristics defining technology-oriented searches as distinct information retrieval activities. They require access to heterogeneous sources (academic literature, patents, technical standards, industry reports, manufacturer documentation); involve domain-specific terminology that varies across disciplines and contexts, challenging consistent retrieval and require technical expertise to formulate queries, evaluate source quality, and interpret information within specific design or innovation contexts. Critically, identifying what the relevant technologies are and which sources provide authoritative information becomes essential for directing subsequent in-depth investigations into their specific characteristics, performance parameters, and applicability to particular design challenges.

2.2. Artificial Intelligence for scholarly literature review

Manual technology reviews are a labour-intensive process; this has driven the interest in automation using Large Language Models (LLMs). These models, trained on a vast text corpora, can synthesize information and generate text with remarkable proficiency (Zhao et al., 2025).

Their application in academic research is growing, showing promise for tasks like literature screening and data extraction (Luomala et al., 2025; Scherbakov et al., 2025). This capability is, then, relevant to technology-oriented search too, as both are structured search-based information retrieval tasks.
However, the utility of LLMs is constrained by significant limitations, primarily "hallucination", the generation of plausible but incorrect information (Ji et al., 2023), and a dependence on prompt quality and source credibility (Bsharat et al., 2024). While LLMs excel at rapid information retrieval and synthesis, human expertise remains crucial for validation and critical insight. Consequently, while AI-assisted literature reviews are well-studied, a clear gap exists in rigorously evaluating LLMs for purpose-oriented technological search/exploration in engineering design. This study directly addresses that gap by benchmarking multiple LLMs against a human-generated baseline to assess their practical value.

3. Research methodology

This study employs a structured, multi-phase methodology to ensure a transparent, repeatable, and rigorous comparison between systematic manual searches and AI-assisted methods. The research method, illustrated in Figure 1, comprises four core stages, which are repeatable for other technologies:

1. Establishing a verified baseline through a systematic technology exploration;
2. Designing and executing AI-assisted searches with a standardized prompt across LLMs;
3. Apply both approaches to a focused case study (here water purification technologies); and
4. Quantitatively and qualitatively evaluating AI-generated vs human-generated outputs.

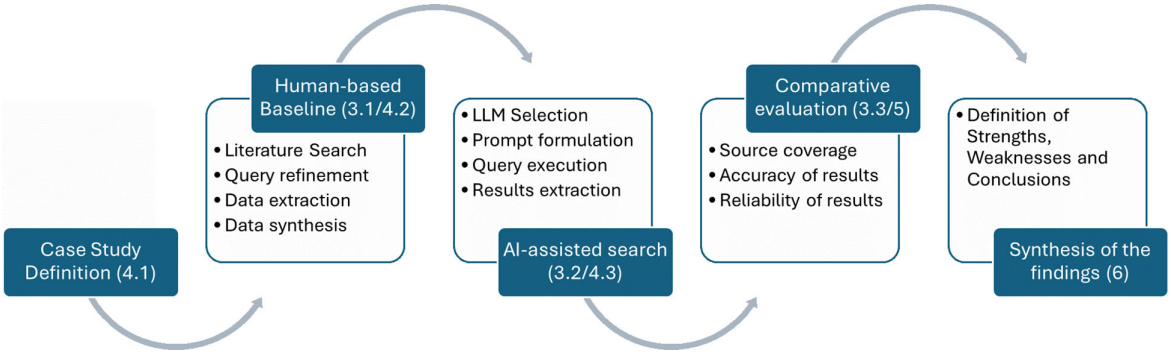

Figure 1. Manual vs AI-assisted technology exploration - research methodology overview

3.1. Establishing the baseline: systematic human-based search

As stated, the manual exploration was conducted in Scopus with the goal of identifying as many relevant technologies as possible in the shortest amount of time. This was achieved by developing a query through an iterative process. The process began by defining the question that best described the targeted technologies. For example, here the question was: "What are the technologies used to purify water in a mobile scenario?". The resulting keywords need to be combined through appropriate boolean operators. The query strategy followed an iterative refinement process to manage the scope of the results:

1. **Initial Broad Search** The initial query can have a broad target spectrum by searching in the article title, abstract, and keywords, trying to maximize Recall.
2. **Refinement for Precision** This broad approach often leads to a high number of results, some/many of which are not relevant, affecting the screening time. Then, the query was refined to target the most pertinent articles.
3. **Preliminary Review for Guidance** This query refinement was guided by a preliminary review of the initial results. This helped identify how the target technologies are referred to in the literature and, crucially, what common characteristics were presented in irrelevant articles.
4. **Strategic Narrowing** Based on this analysis, the search spectrum was decreased. This was done by changing the search to a specific field, such as the article title only, when appropriate. Furthermore, Boolean operators like 'AND NOT' were used in the final stage to exclude non-relevant articles that shared common, identifiable characteristics.

The final optimized query that resulted from this process is presented and analysed in Section 4.2.
The articles retrieved from the final, optimized query underwent a systematic three-stage screening process to ensure relevance. The first stage involved a title screening, where articles were retained only if

their titles suggested the potential inclusion of the targeted technologies. Articles passing this initial filter advanced to the abstract screening stage, where their abstracts were assessed to determine if they contained substantially relevant information. Finally, articles deemed appropriate based on their abstract proceeded to the full-text review, initiating the formal data extraction process for the desired technological information. The data exported included the names of technologies, their key advantages, limitations, and the source link. This was done by identifying tables that some of this article already had with a similar format, or if they had the desired information in text form.

3.2. AI-assisted search and synthesis

To ensure a fair and consistent evaluation of the LLMs, a standardized prompt was designed to extract the necessary data for comparison against the human-generated baseline. The following subsection details the prompt engineering strategy and the subsequent data handling process.

3.2.1. Prompt engineering strategy

A critical factor dictating LLM performance is the quality of the instructions, or prompt. For comparability - which would be impossible to obtain with an LLM-based conversation (multi-prompt) - a single prompt was engineered, to give all models an identical task. The prompt was structured coherently with prompt engineering principles (Bsharat et al., 2024), incorporating these elements.

1. **Role Assignment** The LLM was instructed to "Act as an expert in <domain>" to frame its responses.
2. **Explicit Task Definition** Clear, imperative language ("Your task is . . . ", "You MUST") was used to state the objective: e.g. "generate a list of all possible technologies used to <function>"
3. **Structured Output Format** To facilitate automated parsing and direct comparison, the LLM was mandated to "Present the findings strictly as a concise table." The table's structure was explicitly defined with four columns: Technology (only the name), Pros (a bullet list of key advantages), Cons (a bullet list of key disadvantages), and Link (source URL).
4. **Context and Constraints** The prompt included guidelines to "Ensure the information is factual and focused on the core functionalities, advantages, and limitations of each technology from a technical perspective."
5. **Example Provision** An example row illustrating the expected format and detail for each column was included within the prompt to minimize ambiguity.

3.2.2. AI-driven results extraction and organization

The standardized prompt was submitted to each selected LLM in a new chat session to avoid context contamination from previous interactions. The resulting tables from each model were copied directly from the chat interface and pasted into a centralized Microsoft Excel workbook. This method allowed for the efficient aggregation and side-by-side review of all outputs. Care was taken to preserve the hyperlinks in the "Link" column. A single master sheet was compiled containing all AI-generated tables, which served as the direct input for the evaluation framework described in Section 3.3.

3.3. Evaluation framework

The evaluation of the LLMs was performed by comparing their outputs against the baseline established by the systematic manual search. The quantitative evaluation method consisted of three steps that ensured repeatability and consistency. First, technology coverage was assessed by calculating the percentage of technologies, from the whole human-based set used as baseline, that were correctly identified by each LLM. Second, content accuracy was evaluated. For each technology, the provided "Pros" and "Cons" were scored on a scale from 0 to 5 based on their factual correctness and the precision of the numerical data stated, using the criteria detailed in the Table 1 according to the sources linked by the LLMs. The scores for a given LLM were then averaged to produce a single accuracy score for that model. For instance, a perfect score of 5 required all descriptions to be true and any numerical data to be accurate and non-misleading. Third, source validity was investigated. Every source link provided by the LLMs was checked for accessibility. Each accessible source was then categorized by type (e.g., academic paper, government website, commercial blog, newsletter, Wikipedia page, school website, company website, or non-existing/error). It's noted that when LLMs provided specific hyperlinks for individual data points, for consistency, this analysis focused only on the primary source link listed for each technology. This three-step method was applied consistently to the results from each LLM.

Table 1. Scoring criteria

Score	Description of score
5	All descriptions are true, and any description data is accurate and non-misleading
4	All descriptions are true, but numerical data are imprecise, OR statement is only true in a specific case.
3	One statement not true, OR numerical data misleading
2	One statement not true, AND numerical data misleading
1	Two statements not true, AND numerical data misleading
0	All or nearly all data are incorrect.

4. Comparative analysis

4.1. Case study application: mobile water treatment technologies

The methodology was applied to a focused case study: identifying technologies for a mobile water purification system. This domain presents specific constraints (e.g., portability, energy independence, ease of deployment) that make it a suitable testbed to evaluate the technology orientation of the searches. For this study, a "technology" was defined as a distinct unit process or a commonly combined series of processes used for water purification. This definition accommodates both broad technology clusters (e.g., Pressure-Driven Membrane Processes) and their specific sub-categories (e.g., Ultrafiltration (UF), Microfiltration (MF)). The core comparative analysis, therefore, assesses the overlap and divergence between the technologies identified in the systematic manual review (Section 3.1) and those generated by the LLMs (Section 3.2).

4.2. Human-based search strategy and query definition

The systematic manual search was conducted using the Scopus database. The objective was to identify water purification technologies suitable for mobile systems. An initial search query using core concepts ('Water Treatment' OR 'Water Purification') AND 'Technology' AND 'Mobile' AND 'System' in titles, abstracts, and keywords returned a high volume of irrelevant results. The query was refined iteratively to improve precision. The final strategy required the "water" to appear in the title and excluded documents containing the keywords "Epilepsy", "Superconductivity", and "Fire", which were found to be frequent sources of noise in the initial results. The search was also limited to English-language review articles. The final query used was:

> *(TITLE-ABS-KEY ("Water treatment") OR TITLE-ABS-KEY ("Water purification") AND TITLE-ABS-KEY (Technology) AND TITLE-ABS-KEY (Mobile) AND TITLE-ABS-KEY (System) AND NOT TITLE-ABS-KEY (Epilepsy) AND NOT TITLE-ABS-KEY (Superconductivity) AND NOT TITLE-ABS-KEY (fire) AND TITLE (Water)) AND (LIMIT-TO (DOCTYPE, 're')) AND (LIMIT-TO (LANGUAGE, 'English'))*

Executed on 15 October 2025, the search returned 33 documents. A title-based screening process was conducted, which identified 15 articles as potentially relevant, yielding a retrieval precision of 45%. These 15 papers were advanced to abstract and full-text review. The remaining 18 articles were excluded manually at the title stage because, despite containing the search keywords, their titles clearly indicated a research focus entirely unrelated to mobile water purification. A representative example is "An Overview of Characterisation, Utilisation, and Leachate Analysis of Clinical Waste Incineration Ash," which deals with ash waste management rather than water treatment. This manual exclusion was deemed more efficient than further query refinement, as the small result set and the non-systematic nature of the irrelevance made adding AND NOT operators impractical and risked the unintended exclusion of pertinent literature. After the full text review, the technologies were extracted and gathered in a table, together with their key advantages and disadvantages. These technologies were: Chlorination; Sand filtration; Adsorption; Solar disinfection; Microfiltration; Ultrafiltration; Nanofiltration; Reverse osmosis; Forward osmosis; Membrane distillation; Electrodialysis; Capacitive Deionization; UV treatment; Ozonation; Ion-Exchange; Electro-deionization; Membrane Bioreactors.

4.3. AI-generated results and prompting

The LLMs selected for this study (including ChatGPT, Gemini, Claude, Perplexity, and DeepSeek) were chosen based on their status as the most popular and freely accessible tools at the time of research. A key difference among them is their context window: the amount of text (measured in tokens) they can process in a single interaction. This capacity ranges from standard sizes (e.g., 128k tokens) to the massively enlarged window of Gemini 2.5 Pro experimental 1M-token context (Gemini 2.5 Pro). This selection also enables design students, besides scholars, of benefiting from the findings of the research. The AI-assisted search was executed using the standardized prompt presented in Section 3.2.1. The prompt was intentionally designed to identify general water purification technologies, omitting the specific constraint of "mobile" systems to evaluate the LLMs' inherent ability to generate a broad technology set, allowing for a subsequent analysis on advantages and limitations. The prompt was:

> *Act as an expert in water treatment and environmental engineering. Your task is to generate a list of all possible technologies used for purifying water to make it drinkable within a treatment system. Present the findings strictly as a concise table. The table must have only the following four columns:*
>
> *- Technology: The name of the purification method.*
>
> *- Pros: A bulleted list of key advantages.*
>
> *- Cons: A bulleted list of key disadvantages.*
>
> *- Link: The link of the source you got the information from*
>
> *Ensure the information is factual and focused on the core functionalities, advantages, and limitations of each technology from a technical perspective.*

The novel technologies funded by the LLM tools are: Coagulation; Flocculation; Sedimentation; Biological Treatment [Perplexity only]; Aeration [Perplexity only]; Advanced Oxidation Processes; Packed Tower Aeration [Claude only]; Multi-Stage Flash (MSF) Distillation [Claude only]; Multi-Effect Distillation [Claude only]; Electrodialysis Reversal [Claude Only]; Capacitive Deionization [Claude Only]; Membrane Capacitive Deionization [Claude Only]; Electrocoagulation [Claude only]; Air Stripping [Claude Only]; Biological Filtration.

5. Evaluation of the results

The findings of the comparative analysis between the technologies found by the systematic manual baseline and the LLM conceptualized technologies are presented in this section. The results are organized to evaluate the models based on technology coverage, novelty, the credibility of their sources, and the factual accuracy of their outputs. The systematic manual search established a baseline of 17 distinct mobile water purification technologies (Section 4.2). A preliminary finding was that all technologies

Table 2. Example of the AI-generated table, with technologies, pros, cons, and sources (excerpt)

Claude Sonnet 4.5				
#	**Technology**	**Pros**	**Cons**	**Link**
1	Coagulation-Flocculation	Removes 27-87% of bacteria and viruses; Effective for suspended particles and turbidity; Removes natural organic matter (NOM); Cost-effective when chemicals are available; Can remove heavy metals like iron	Requires chemical addition (aluminum/iron salts); Generates large amounts of sludge; Chemical dosage must be adapted to water composition; Does not completely remove all pathogens; Requires trained operators	https://www.safewater.org/fact-sheets-1/2017/1/23/conventional-water-treatment
2	Sedimentation	Simple gravity-based process; Low energy consumption; Effective for removing heavy particles and floc; Well-established technology; Can handle large volumes	Requires large basin area; Time-consuming process; Not effective for small/light particles; Requires regular sludge removal; Cannot remove dissolved contaminants	https://www.cdc.gov/drinking-water/about/how-water-treatment-works.html

generated by the LLMs (Section 4.3) were relevant to water purification. Table 2 presents an excerpt from the results generated by Claude Sonnet 4.5 when it receives the prompt presented in Section 4.3.

5.1. Technology coverage and novelty

The technology coverage measures each model's ability to recall the technologies identified in the manual baseline. As detailed in Figure 2a,b,d and Table 3, performance varied from model to model. Claude Sonnet 4.5 achieved the highest coverage of technologies, correctly identifying all 17 baseline technologies and adding 14 more to them. It was interesting, though, that instead of stating the general category of the technology cluster, it listed the subcategories and these increasing significantly increased the number of technologies. An example is the "Membrane Filtration", where Claude listed all four filter types (Microfiltration (MF), Ultrafiltration (UF), and Nanofiltration (NF)) and the Sand filter, which again included both subcategories of Slow and Rapid sand filter. This was also done by other models, which complicated a direct one-to-one comparison. Perplexity and Gemini 2.5 Pro tied for the next highest coverage, each identifying 9 technologies (53%). GPT-5 and DeepSeek DeepThink both covered 8 technologies (47%). DeepSeek identified 7 technologies representing 41% of the baseline. DeepSeek DeepThink & Search had the lowest coverage, identifying only 5 technologies, representing only 29% of the baseline. Regarding novelty, all models except one introduced technologies not found in the baseline: Claude Sonnet 4.5 generated the highest number of novel technologies (14), all of which were verified as actual water purification technologies. In contrast, DeepSeek DeepThink & Search introduced zero novel technologies.

Table 3. Summary of AI-generated technology novelty

	GPT-5	Perplexity	Gemini 2.5 Pro	DeepSeek	DeepSeek DeepThink	DeepSeek DeepThink & Search	Claude Sonnet 4.5
Results	10	11	11	10	10	5	29

5.2. Source credibility

A critical aspect of any literature review is the reliance on credible sources. This section analyses the types of sources the LLMs provided for their information. The results are summarized in Table 4, revealing stark differences in sourcing strategy and credibility. GPT-5, DeepSeek, and DeepSeek DeepThink each had a high number of source errors (e.g., broken or expired links). In comparison, the other models provided links that correctly led to the source. Figure 1c shows the percentage of sources that were Authoritative, Non-Authoritative, or resulted in an error. Authoritative sources were assumed to be the sources coming from school websites, academic papers, government websites, and books. The rest of the sources, like blog posts and Wikipedia, were classified as Non-Authoritative. In some cases, the specific data points in the "Pros" and "Cons" were not found in the cited source, though the general information about the technology was present elsewhere.

Table 4. Source credibility: breakdown of source types by AI model

	GPT-5	Perplexity	Gemini 2.5 Pro	DeepSeek	DeepSeek DeepThink	DeepSeek DeepThink & Search	Claude Sonnet 4.5
Error on Link	5			9	8		
Blog						3	3
Company							1
News		8					4
Wiki							7
School							1
Paper				1	1		2
Government	7	3	11		1	2	7
Book							3

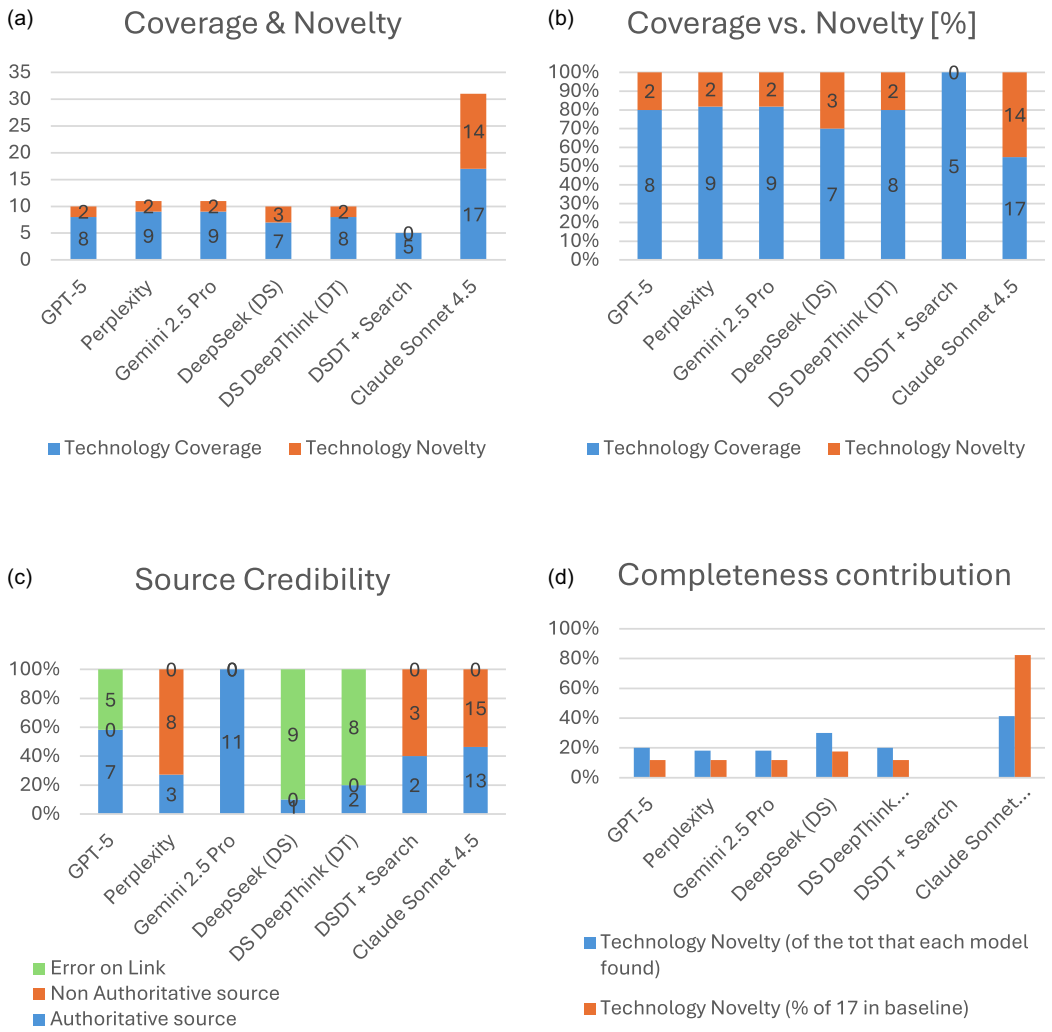

Figure 2. a. LLM technology coverage and its novelty; b. percentage of covered and novel technologies of the LLMs; c. performance of each LLM according to the authoritative and the error of the sources; d. performance of the LLMs according to the novelty and cove

5.3. Content accuracy

The factual correctness of the "Pros" and "Cons" generated by each model was scored on a 0-5 scale, as explained in the Section 3.3. The average scores, shown in Table 5, were consistently high, with small variations. Most of the models performed extremely well with average scores above 4.9, and no individual description score fell below 4. This means that when a numerical value was included, it was imprecise but within the range found in the manual search, or a statement was only true in specific cases and lacked context, potentially misleading the reader.

Table 5. Average score of "pros" and "cons" of the technologies of every LLM

	GPT-5	Perplexity	Gemini 2.5 Pro	DeepSeek	DeepSeek DeepThink	DeepSeek DeepThink & Search	Claude Sonnet 4.5
Accuracy - Pros	5	5	5	5	5	5	4.7
Accuracy - Cons	4.9	5	5	5	4.9	4.6	4.8

The slightly lower values recorded for Claude are to be balanced with the higher number of technologies it provides, which inevitably increases the chances to hit a potential lack of accuracy (which is in any case sufficiently rare).

6. Discussion and conclusion

This paper addressed the challenge of conducting technology-oriented searches in engineering design contexts by investigating whether Large Language Models can effectively support traditional technology-oriented searches approaches. Using a case study in mobile water purification technologies, we compared human-conducted technology landscape analysis through Scopus with single prompt AI-assisted approaches using multiple freely accessible LLMs. This research introduces a repeatable, structured methodology that can be systematically applied to different technology domains and evaluated with evolving sets of LLMs as these tools continue to develop, despite search refinements, via prompt adjustment and the expanding memory capabilities of LLM models can radically change the relationship between a single and robustly engineered prompt and technology-relevant results

The comparative analysis revealed substantial variation in LLM performance. Claude Sonnet 4.5 demonstrated superior performance, achieving 100% coverage of the 17 baseline technologies while introducing 14 additional relevant ones, though with slightly lower accuracy scores compared to other models. Perplexity and Gemini 2.5 Pro tied for second-highest coverage at 53%, while GPT-5 and DeepSeek DeepThink both achieved 47%. DeepSeek DeepThink & Search exhibited the weakest performance, covering only 29% of baseline technologies and introducing no novel ones. Source credibility varied considerably: GPT-5 and DeepSeek variants produced numerous broken or expired links, while content accuracy remained consistently high across models, with most averaging above 4.9 on a 0–5 scale. A notable pattern across models—most prominently in Claude—was the tendency to decompose technology clusters into subcategories rather than reporting them at the general level (e.g., listing Microfiltration, Ultrafiltration, and Nanofiltration separately rather than "Membrane Filtration"). While this granularity increases practical utility for design purposes, it complicates direct coverage comparisons and should be accounted for in future evaluation frameworks. These findings have significant implications for Technology Intelligence activities and design processes. LLMs can dramatically accelerate the initial exploration phase of technology landscape analysis, rapidly generating comprehensive technology inventories that would require substantially more time through manual searches. For complex, multi-technology system integration challenges (e.g. water treatment scenario examined here) this rapid synthesis capability supports more informed decision-making regarding technology compatibility and system architecture. The free accessibility of the selected LLMs makes this approach particularly valuable for engineering design education, enabling students to find and explore technologies they were previously unaware of, using the structured methodology presented here as a reference framework. However, this preliminary investigation has important limitations. First, the set of evaluated LLMs was limited and arbitrary, reflecting popular freely accessible tools rather than a comprehensive survey. Second, the human-generated baseline was derived exclusively from academic literature through Scopus, while comprehensive technology intelligence should incorporate patents, technical standards, manufacturer specifications, and industry reports. Extending the search to patent databases would substantially increase both coverage and effort: a comparative mapping against the Cooperative Patent Classification system confirmed that manual literature review, even when conducted over approximately 24 hours of focused work, covers only a portion of the technology categories captured by structured patent hierarchies—and navigating those hierarchies to identify relevant subcategories, exclude non-pertinent classifications, and link patent classes to actual deployable technologies represents a significant additional undertaking. This limitation, however, potentially affects coverage assessments, as technologies primarily documented in non-academic sources may have been overlooked. Third, the results reflect the current state of LLM development and their training data cutoffs, which may not include recently emerging technologies, though models with web search capabilities may partially address this. Fourth, despite high accuracy scores, the persistent risk of hallucinations necessitates careful validation of LLM outputs until these tools reach full maturity. Critically, the purpose of LLM-assisted technology exploration is not to replace human judgment but to efficiently bring to the user's attention technologies that may not have been previously known or considered. The approach serves as an initial discovery mechanism, expanding the designer's awareness of the technology landscape and potentially revealing unexpected options. The subsequent steps of any technology intelligence-related activity—including detailed technical assessment, performance verification, feasibility analysis, compatibility evaluation, and strategic selection—still require substantial human expertise and control. Designers must critically evaluate AI-generated suggestions, validate technical claims against primary sources, assess applicability to specific design contexts, and

synthesize information across multiple sources. The methodology presented here should thus be viewed as a complementary front-end tool that enhances rather than replaces the comprehensive, multi-stage process of technology intelligence in engineering design.

Acknowledgement

This study was developed in the context of the MISSION4WATER Project—Interreg IPA-ADRION 0219—Multidisciplinary strategic partnership providing innovative Solutions to reduce pollutants dispersion in WATER.

References

Bsharat, S. M., Myrzakhan, A., & Shen, Z. (2024). Principled Instructions Are All You Need for Questioning LLaMA-1/2, GPT-3.5/4 (No. arXiv:2312.16171). arXiv. https://doi.org/10.48550/arXiv.2312.16171

Cascini, G., Becattini, N., Kaikov, I., Koziolek, S., Kucharavy, D., Nikulin, C., … & Vanherck, K. (2015). FORMAT–building an original methodology for technology forecasting through researchers exchanges between industry and academia. *Procedia Engineering*, 131, 1084–1093.

Ghaffari, M., Aliahmadi, A., Khalkhali, A., Zakery, A., Daim, T. U., & Yalcin, H. (2023). Topic-based technology mapping using patent data analysis: A case study of vehicle tires. *Technological Forecasting and Social Change*, 193, 122576.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Chen, D., Dai, W., Chan, H. S., Madotto, A., & Fung, P. (2023). Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys*, 55(12), 1–38. https://doi.org/10.1145/3571730

Loo, S.-L., Fane, A. G., Krantz, W. B., & Lim, T.-T. (2012). Emergency water supply: A review of potential technologies and selection criteria. *Water Research*, 46(10), 3125–3151. https://doi.org/10.1016/j.watres.2012.03.030

Luomala, M., Naarmala, J., & Tuomi, V. (2025). Technology-Assisted Literature Reviews with Technology of Artificial Intelligence: Ethical and Credibility Challenges. *Procedia Computer Science*, 256, 378–387. https://doi.org/10.1016/j.procs.2025.02.133

Rohrbeck, R. (2010). Harnessing a network of experts for competitive advantage: technology scouting in the ICT industry. *R&d Management*, 40(2), 169–180.

Rubel, J., Buysschaert, F., & Vandeginste, V. (2025). Review and selection methodology for water treatment systems in mobile encampments for military applications. *Cleaner Water*, 3, 100065. https://doi.org/10.1016/j.clwat.2025.100065

Scherbakov, D., Hubig, N., Jansari, V., Bakumenko, A., & Lenert, L. A. (2025). The emergence of large language models as tools in literature reviews: A large language model-assisted systematic review. *Journal of the American Medical Informatics Association*, 32(6), 1071–1086. https://doi.org/10.1093/jamia/ocaf063

Veugelers, M., Bury, J., & Viaene, S. (2010). Linking technology intelligence to open innovation. *Technological forecasting and social change*, 77(2), 335–343.

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., … Wen, J.-R. (2025).

Zhou, Z., & Carbajales-Dale, M. (2018). Assessing the photovoltaic technology landscape: efficiency and energy return on investment (EROI). *Energy & Environmental Science*, 11(3), 603–608.