Proceedings of the Design Society, Volume 6: DESIGN 2026
https://doi.org/10.1017/pds.2026.10483

Large language models for identifying repurposing opportunities: a systematic evaluation

Adrian Dörnbach✉, Sebastian Sonntag, Abinash Selvarajah and Arun Nagarajah

University of Duisburg-Essen, Germany

✉ adrian.doernbach@uni-due.de

ABSTRACT: In a circular economy, repurposing extends product lifecycles and reduces resource use. However, identifying feasible repurposing opportunities remains challenging. This study therefore evaluates the capability of large language models to identify such repurposing scenarios and their relevant properties, using documented repurposing cases from peer-reviewed literature. Three models were tested, revealing potential in identifying repurposing scenarios, but also highlighting the need for structured methods and further research due to systematic limitations in property identification.

KEYWORDS: circular economy, sustainability, repurposing, large language model (LLM), multi-life products

1. Introduction

On July 24, 2025, mankind reached Earth Overshoot Day, marking the point at which all the resources that the Earth can sustainably regenerate within a year have been consumed (Wackernagel & Lin, 2023). Nevertheless, global material resource use is projected to more than double from 2020 to 2060, driven by economic growth models which are fundamentally dependent on resource consumption (Bruyninckx et al., 2024). In response to this challenge the European Commission has established policy frameworks targeting the decoupling of economic growth from resource consumption, positioning resource management and efficiency as central mechanisms for achieving these objectives (European Union, 2024). Consequently, the Circular Economy (CE) framework has emerged as a strategy to address resource limitation challenges. According to the Ellen MacArthur Foundation, the successful implementation of CE principles can reduce primary raw materials consumption by 53% (Ellen MacArthur Foundation, 2015). Within the spectrum of circular strategies, CE models reveal significant variation in resource preservation effectiveness. The EU Waste Hierarchy and the 6R Framework establishes a hierarchical order based on value retention, positioning repurposing as a high-value strategy. (Kurt et al., 2019). While recycling enables material recovery, it destroys the value added to products and requires substantial energy investment for reprocessing and new production (Reike et al., 2017). In contrast, repurposing preserves product structure and functionality by enabling continued use in alternative applications, thereby maintaining added value while significantly lowering overall energy and resource consumption (Ahmed et al., 2025). This positions repurposing as a particularly promising circular strategy. However, despite its potential, repurposing is not yet widely implemented in research or industrial practice. (Ali et al., 2020)

1.1. Related work

Existing literature addresses multiple barriers to repurposing implementation, which can be categorised into four main challenge areas: design, economic, ecological, and the identification and evaluation of repurposing applications (Psarommatis et al., 2025).
Pluhnau et al. identifies that products are often designed for a single life cycle, which makes repurposing difficult (Pluhnau et al., 2023). Current design-related approaches, such as design for modularity and

https://doi.org/10.1017/pds.2026.10483 Published online by Cambridge University Press

design for disassembly, are fundamentally limited as they must be considered in the early stages of product development, when potential scenarios are still unknown (Hewa Witharanage et al., 2025).
Although repurposing offers fundamental sustainability benefits, its environmental impact can vary depending on the product and process. Furthermore, rebound effects can cause resource shifts in later stages (Potting et al., 2017). Multi-Life-Cycle Assessments therefore quantify the environmental implications of repurposing by comparing material flows, energy consumption, and emissions across alternative scenarios (Bauer, 2018).
The repurposing of a car battery into stationary storage has been demonstrated to have significant resource preservation potential and economic benefits (Jeppe et al., 2023). However, this is only economic efficient if the costs of the repurposing scenario are lower than those of new production. Economic assessment frameworks evaluate the financial viability of projects by considering the costs of development, production, operation and end-of-life processing (Psarommatis & May, 2025).
Existing documented approaches are limited and do not demonstrate a systematic approach to identifying repurposing scenarios. Currently, there is no systematic approach to identifying potential applications for repurposing (Dörnbach et al., 2024).
Analysing the literature reveals that identifying repurposing scenarios poses a significant challenge for implementing repurposing practices. Since most approaches are based on predefined scenarios, addressing this identification gap is essential. Recent work has demonstrated that LLMs can generate novel repurposing ideas (Hewa Witharanage et al., 2024). However, it has not been established whether LLMs can reliably identify technically feasible repurposing scenarios.
Furthermore, identifying a feasible repurposing scenario is necessary but not sufficient for successful implementation. Critical success factors related to specific product properties must also be met. For example, when repurposing laptops as thin clients, 91% of devices proved unsuitable due to inadequate technical or functional specifications (Coughlan et al., 2018). This demonstrates that even fundamentally viable scenarios require product specific properties which must be met for repurposing to be successful. Therefore, it highlights the importance of property compatibility between components and target systems and the need for systematic methods to identify feasible repurposing scenarios.

1.2. Research gap

The analysis reveals two research gaps. First, no established methods exist for systematically identifying feasible repurposing scenarios. Second, even when promising scenarios are known, most specific product options prove incompatible due to property incompatibility. Analysis shows that even for feasible repurposing scenarios, not all individual products meet the property requirements for successful implementation. This reveals that repurposing-relevant properties determine technical feasibility in given repurposing scenarios, yet these properties are not systematically operationalised. This prevents a structured assessment of whether a specific component meets the requirements of a known target application.
Recent work has demonstrated that LLMs can generate novel repurposing ideas but, it has not been established whether LLMs can reliably identify technically feasible repurposing scenarios. Therefore, this study investigates whether LLMs can provide systematic support for identifying repurposing opportunities: *To what extent can Large Language Models support the identification of technically feasible repurposing scenarios?*
Two sub-questions guide the analysis:

> **RQ1:** How effectively can LLMs identify technically feasible repurposing scenarios for a given source system?
> **RQ2:** How effectively can LLMs identify the properties necessary for assessing technical feasibility for a given repurposing transition?

To address these research questions, this paper is structured as follows: Section 2 describes the study design, with the results presented in Section 3 and discussed in Section 4. Finally, the research concludes with a summary and an outlook on future work in Section 5.

2. Study design

To investigate the capability of LLMs to identify technically feasible repurposing scenarios, this study is based on a four-step process. In the first step, a dataset of documented repurposing cases is created to provide a transparent basis for evaluation (Section 2.1). In the second step, structured prompts were

https://doi.org/10.1017/pds.2026.10483 Published online by Cambridge University Press

designed to translate the documented repurposing cases into standardised input tasks for the LLMs (Section 2.2). In the third step, evaluation metrics are defined to assess the quality of the LLM outputs (Section 2.3). In the fourth step, the developed prompts are applied to different state-of-the-art LLMs, and the generated outputs are systematically analysed to identify patterns, limitations and dependencies on system complexity (Section 2.4). The following sections describe each step in detail. To ensure reproducibility the study materials are available on GitHub[1] including: (1) detailed documentation of the systematic literature review process, (2) the validated dataset of 72 repurposing cases and (3) all prompts used for both research questions.

2.1. Dataset creation

To evaluate the performance of LLMs, a validated dataset is required that includes documented cases of repurposed products and properties that determine technical feasibility, to provide an objective benchmark. Therefore, the objective was to identify a comprehensive set of documented repurposing cases. The dataset was established following a structured and transparent procedure based on the PRISMA 2020 methodology to ensure reproducibility and methodological rigour (Page et al., 2021).

Definition of Search Strategy and Databases – The search aimed to capture documented cases of repurposing, defined as the repurpose of products or components from an original system in a different technical environment with altered requirements. An initial unsystematic literature search for repurposing examples was conducted to identify relevant publication types and keywords. The resulting sample was then used to refine the search query into four main concept groups (Repurposing; Products or Components; Examples or Applications; and Circular Economy or Lifecycle) and define inclusion and exclusion criteria for the systematic review. The systematic literature search was performed in the Scopus and Web of Science databases, which were selected for their broad and complementary coverage of engineering and design research (Singh et al., 2021). The final search query combined four main concept groups using Boolean operators: terms related to repurposing, products or components, examples and applications and the circular economy or lifecycle.

Definition of Inclusion and Exclusion Criteria – To ensure the publications were both relevant and of a high quality, they were filtered according to defined inclusion and exclusion criteria. Only peer-reviewed journal and conference papers published between 2015 and 2025 in English or German were considered. The chosen time frame reflects the increasing relevance of the circular economy in recent years and the comparatively young scientific treatment of repurposing. The selected studies had to belong to the fields of engineering or environmental science and be accessible in full text. Publications that did not meet one or more of these criteria were excluded.

Data Extraction and Synthesis – After screening and eligibility checks, 50 publications were included in the final review. From these, 105 repurposing cases were extracted. The product or component properties considered relevant for repurposing by the original authors were manually extracted for each case.

Dataset Structuring – The dataset was refined through three analytical steps to establish a solid basis for evaluating LLM capabilities in terms of scenario identification and property analysis.

First, duplicate or similar scenarios were removed, resulting in 72 unique repurposing scenarios that provide a clearer and more consistent basis for analysis.

Secondly, the scenarios were classified by complexity. The 72 scenarios were divided into two complexity clusters: Cluster 1 ($n = 50$) consists of scenarios characterised by physical reuse without system integration or technical interfaces representing low repurposing complexity. Cluster 2 ($n = 22$) consists of scenarios requiring system integration, including electronics, control interfaces and functional coupling between components representing high repurposing complexity.

Third, the repurposing-relevant properties were analysed. To ensure statistical reliability, cases with fewer than three documented properties were excluded, yielding 32 scenarios suitable for property analysis. All identified properties were ranked in descending order according to how frequently they were documented across the reference dataset. As shown in Figure 1, the distribution varied considerably: some properties appeared in over 50% of cases, while others were case-specific and only appeared in a single scenario.

[1] https://github.com/IPE-PEP/DESIGN-2026

https://doi.org/10.1017/pds.2026.10483 Published online by Cambridge University Press

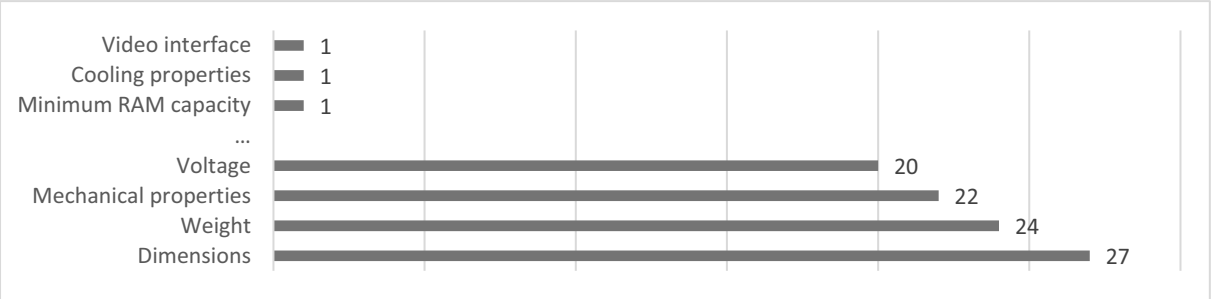


Figure 1. Distribution of repurposing properties across scenarios

This ranking was then used to divide the properties into three distinct categories. The top 20% represented the most frequently documented properties, the middle 60% of documented properties and the bottom 20% represented properties that were documented least frequently.

2.2. Prompt design

The step prompt design should translate the dataset information into standardized input tasks for LLM evaluation. The objective in this study is to have a consistent, reproducible and model independent comparison of LLM capabilities in identifying technically feasible repurposing scenarios and related relevant repurposing properties. To ensure comparability, prompts followed a structured format of three elements: *problem description, task description* and *format description* (Sahoo et al., 2025). Therefore, two prompts were designed, one for each research question, to systematically evaluate LLM performance. Figure 2 shows an example prompt for RQ1, which asks for scenarios for repurposing a high-voltage battery from an electric car.

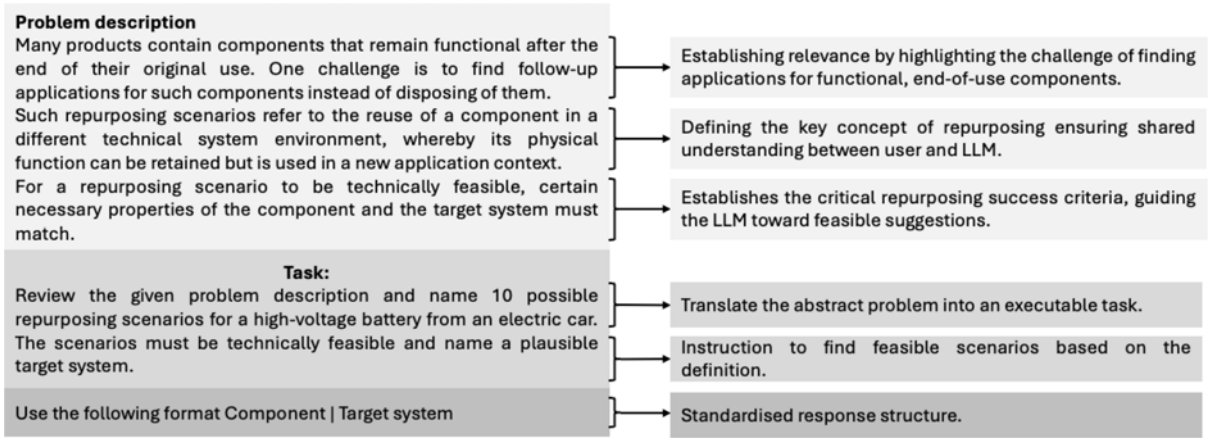


Figure 2. Prompt design for RQ1 based on the given structure

Problem description – Both prompts share the same description, outlining the challenge of identifying technically feasible repurposing scenarios for systems that remain functional at end-of-use.
Task description – The first prompt targets the identification of possible follow-up applications, requesting technically feasible repurposing scenarios for a given component. The second prompt extends this by asking for all the technical properties required to realize a given repurposing scenario.
Format description – To ensure consistency, output formats were standardized. Scenario identification followed the format Component | Target System, while for property identification (RQ2) only the respective property was reported, without a predefined structure. This uniform formatting ensured clear, machine-readable outputs that could be quantitatively evaluated.
Output constraints – The research approach aims to assess the consistency and reliability of LLMs in identifying technically feasible repurposing scenarios. To evaluate this, the models were instructed to generate ten repurposing scenarios for RQ1 across multiple runs. Scenarios that consistently appear in the top ten across runs indicate reliable identification by the LLM. For RQ2, no output limit was applied, as comprehensive property identification was required to assess completeness. The prompt explicitly requests all relevant properties, consequently, any missing properties could be classified as either irrelevant or not recognized by the LLM.

https://doi.org/10.1017/pds.2026.10483 Published online by Cambridge University Press

Testing – The aim of the study was to ensure that outputs reflect actual model performance, rather than suboptimal prompt formulation. Therefore, approximately 10% of the dataset was used for iterative prompt optimisation to minimise variability in task interpretation and ensure consistent, reproducible results.

2.3. Evaluation metrics

This study evaluates the ability of LLMs to identify technically feasible repurposing scenarios and their relevant properties. Therefore, LLM performance is evaluated using quantitative metrics that assess the alignment between LLM-generated outputs and the validated dataset (Section 2.1). This approach ensures methodological transparency, reproducibility and comparability across models, providing an objective basis for addressing the research questions (Section 1.2).

Methodological approach – The evaluation focuses exclusively on correctly identified documented cases, providing a conservative assessment of LLM capabilities. Two task specific metrics enable this evaluation: Scenario Identification Rate (SI) measures the coverage of documented repurposing scenarios, while Property Identification Rate (PI) measures the identification of relevant properties by the LLM. Documented cases and LLM outputs were compared semantically for each run, using LLMs to detect equivalent expressions, followed by manual verification to ensure accuracy.

(SF = identified scenarios, NS = documented scenarios, PF = identified properties, NP = documented properties).

$$SI = \frac{SF}{NS} \tag{1}$$

$$PI = \frac{\mathrm{PF}}{\mathrm{NP}} \tag{2}$$

Evaluation procedure – The evaluation was conducted at two levels of detail: per individual component (micro level) and aggregated (macro level). For SI, an instance is considered successfully identified if the corresponding documented scenario appears among the 10 scenarios suggested by the LLM. For PI, the evaluation measures the proportion of correctly identified properties relative to all properties documented in the dataset for the respective scenario. Model outputs were normalised before analysis to ensure consistent terminology and format.

2.4. Study execution

This study evaluates the fundamental ability of LLMs to identify technically feasible repurposing scenarios. To establish the upper bound of LLM capabilities for this task, three state-of-the-art models representing the current performance threshold were selected. Testing these models with identical prompts on the same dataset ensures comparability, while also revealing any fundamental limitations of current LLM architectures if the performance is insufficient. Following models were included: GPT-5 (OpenAI, 2025), Claude Sonnet 4.5 (Anthropic, 2025) and Gemini 2.5 Pro (Google, 2025). All models were accessed via official APIs. Experiments were conducted between July and September 2025 using the latest publicly available model versions. All models were tested under identical zero-shot conditions with a fixed moderate temperature and consistent evaluation metrics. Basic settings were selected to balance consistency and variation in model outputs, as the study aims to assess fundamental model performance. Each model run represented one independent execution of a prompt. To ensure robustness, each query was executed ten times per model, resulting in 3,120 runs in total.

3. Results

The results summarise the evaluation of LLM performance in identifying technically feasible repurposing scenarios (RQ1) and the corresponding relevant properties required for their realisation (RQ2).

3.1. Performance findings

The quantitative analysis is based on the metrics defined in Section 2.3 and is followed by a cluster specific comparison to examine the influence of system complexity on model performance.

SI and PI were calculated for each individual run and then averaged across runs to receive the reported metrics. Table 1 presents an overview of the achieved metrics across both research questions

https://doi.org/10.1017/pds.2026.10483 Published online by Cambridge University Press

Table 1. Comparison of the achieved metrics of the LLMs

Evaluation Metric	GPT-5	Gemini 2.5 Pro	Claude 4.5 Sonnet
SI	84%	78%	81%
PI	45%	42%	46%

Table 1 shows that LLMs perform significantly better at scenario identification than at property recognition, with minimal performance differences between models. However, the performance gap between tasks is more pronounced than between models. All three LLMs struggle considerably more with identifying repurposing-relevant properties than with recognising scenarios.

Notably, the probability of retrieving all documented properties for a component within a single run was 0% for all models. This means that no model ever identified the complete set of documented properties for a single repurposing scenario. Each run missed at least one relevant property, indicating a systematic gap in comprehensive property retrieval. This suggests that LLMs are effective at identifying scenarios, but inadequate at recognising relevant properties for repurposing.

Two other analyses were conducted to examine performance patterns systematically. First, the complexity patterns and second, the property identification rate, were statistically analysed. Both the complexity clusters and the property distribution were introduced in Section 2.1.

Table 2. Complexity-based performance of LLMs for scenario identification

	GPT-5		Gemini 2.5 Pro		Claude 4.5 Sonnet	
Evaluation Metric	Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 1	Cluster 2
SI	92%	65%	88%	55%	90%	61%

Table 2 shows scenario identification performance across low-complexity repurposing cases (Cluster 1) and high-complexity repurposing cases (Cluster 2). The results show that performance varies more between the two clusters than between individual models. Across all three LLMs, accuracy remains high for low-complexity scenarios but is noticeably low for high-complexity ones. Therefore, all models perform consistently well when the underlying structural and functional relationships are simple. However, performance is low when scenarios involve a greater number of interacting properties, functional dependencies or technical constraints. The decline across all three models indicates a complexity-dependent pattern in scenario identification, higher system complexity is consistently associated with lower accuracy.

Figure 3 shows property retrieval performance across the three documentation frequency categories defined in Section 2.1.

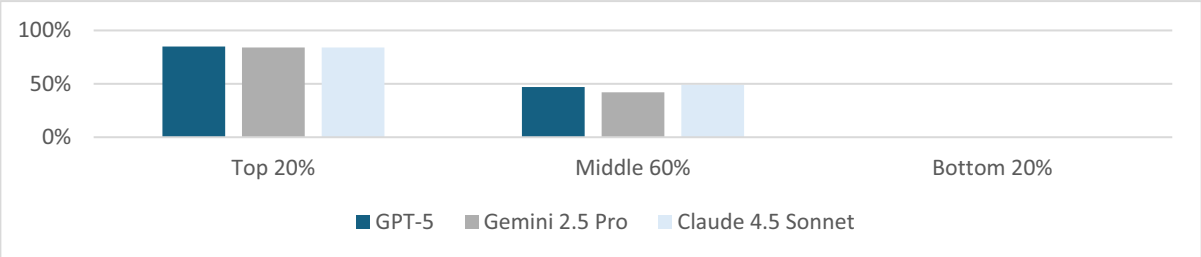


Figure 3. Property identification performance by dataset frequency

The results show substantial differences in retrieval rates between the categories, while differences in performance between models remain small. Property retrieval in the top 20% was high for all three LLMs, indicating stable performance for frequently documented properties. Retrieval accuracy decreases substantially in the middle 60%. For the bottom 20%, all models show zero retrieval rates, suggesting that LLMs struggle to identify properties that are rarely documented or cases specific. This consistent pattern across all three LLMs suggests a frequency dependent retrieval effect: properties that appear frequently in the dataset are more likely to be retrieved. However, properties appearing infrequently remain unrecognised, regardless of model architecture.

https://doi.org/10.1017/pds.2026.10483 Published online by Cambridge University Press

The results shows that LLMs reliably identify repurposing scenarios with low complexity, but their performance is low in high complexity cases, indicating a dependency on the complexity of the scenario. Properties that frequently occurred in the dataset were recognised consistently, while those that occurred less frequently were identified inconsistently. No single run across all models returned the complete set of relevant properties, indicating that property recognition remains incomplete regardless of model choice. Overall, the performance differences between the three LLMs are minimal, suggesting that task complexity and property frequency have a much stronger impact on performance than architectural differences between the models.

3.2. Results analysis

The results demonstrate both practical application potential and systematic limitations of current LLMs. The underlying influencing factors are explained below in order to analytically classify the observed patterns. All models showed consistent performance in identifying repurposing scenarios (RQ1) and recognizing technical properties (RQ2), with comparable limitation patterns and no significant performance differences.

For RQ1, the evaluation shows that LLMs can identify repurposing scenarios in low-complexity cases. In these scenarios, all models achieved consistently high performance and reliably generated valid repurposing concepts. This demonstrates their practical applicability to simple technical systems. However, performance declined substantially as system complexity increased, with identification accuracy dropping from approximately 90% to 60%. This pattern indicates that there are systematic challenges in processing complex technical relationships, suggesting that current models struggle to generalise beyond simple structural relationships.

For RQ2, a similar trend emerged in identifying repurposing-relevant properties. Frequently documented properties were retrieved reliably, whereas rarely documented or case specific properties were often missed. This selective retrieval pattern suggests a reliance on statistical associations in the training data rather than genuine domain reasoning. Furthermore, no model successfully identified the full set of relevant properties in a single run, highlighting a fundamental limitation in systematic property retrieval.

Collectively, the results suggest that while LLMs can recognise familiar patterns, they lack the deeper technical understanding required to consistently interpret complex systems and systematically identify all relevant properties for repurposing assessments.

4. Discussion

This study evaluated the capability of LLMs to systematically identify technically feasible repurposing scenarios (RQ1) and the recognition of relevant technical properties (RQ2). Results demonstrate that state-of-the-art models achieve substantial scenario identification rates (78-84%), confirming fundamental capability for this domain. Property identification (RQ2) revealed more pronounced challenges with moderate property identification rates (42-46%). The fact that none of the models retrieved all documented properties for any component within a single run indicates systematic limitations in capturing complete property-level information.

4.1. Root causes

LLM failures can be classified into two fundamental causes: errors in describing the problem (insufficient prompt formulation) and errors in understanding the problem (model comprehension limitations) (Xu et al., 2024). The two hypotheses were put to systematic testing in this study.

Problem Description – Errors in the problem description occur when task specifications are inaccurate or incomplete. This leads to model confusion, regardless of the model's underlying capability. If this error were applicable, performance would be inconsistent across cases due to inconsistent prompt quality. To isolate this factor, all models received identical zero-shot prompts across all test cases to ensure a uniform problem description. However, the results did not support this hypothesis because performance was consistently high (90%) for low-complexity repurposing scenarios, but systematically low (60%) for high complexity scenarios across all models. This pattern indicates that prompt formulation is not the limiting factor. The consistent performance differential based on scenario complexity suggests that the models adequately understood the task structure but struggled with domain specific complexity.

https://doi.org/10.1017/pds.2026.10483 Published online by Cambridge University Press

Problem Understanding – Errors in the problem understanding occur when the model fails to comprehend the underlying technical relationships, despite an adequate task specification. If this error were applicable, performance should decrease systematically with increasing problem complexity, while the prompt structure remains constant. The experimental design controlled for variability in the prompts by maintaining identical prompt structures across all runs. Under these constant conditions, three observations provide convergent evidence for limitations in problem understanding:

All models showed significantly lower performance (60%) for high complexity repurposing scenarios compared to low-complexity scenarios (90%), indicating complexity-dependent performance degradation. Inconsistent property identification across runs suggests that LLMs may identify properties based on statistical pattern matching in the training data rather than genuine technical understanding. Properties that are frequently mentioned in training data are more likely to be retrieved, while case-specific technical requirements that are different from common patterns are systematically missed.

No model achieved 100% property identification in any single run (complete identification rate of 0%). This universal failure across all models and scenarios indicates a fundamental limitation in systematic property identification, rather than isolated errors.

These three observations: complexity-dependent degradation, pattern-based identification rather than reasoning-based identification, and incomplete property retrieval, collectively point to limited technical problem comprehension as the primary failure mode. This suggests that the models demonstrate pattern recognition capabilities but lack the structured domain knowledge and reasoning mechanisms required to assess complex repurposing scenarios and systematically identify repurposing-relevant properties.

4.2. Limitation and future work

The primary goal of this study was to evaluate the fundamental capabilities of LLMs in identifying feasible repurposing scenarios and the corresponding repurposing-relevant properties. The study provided baseline insights into model performance under controlled conditions but also revealed several methodological and analytical limitations that should be addressed in future work.

Methodological Improvements – The study revealed strong indications that large language models struggle with missing contextual or domain-specific knowledge. This leads them to rely on statistical pattern matching rather than genuine technical reasoning. These findings suggest a structural limitation in problem understanding. Further targeted experimentation is needed to determine the root causes of the observed errors more precisely.

In terms of the study design, the reference dataset consists of scenarios and properties extracted from peer-reviewed publications and therefore could not be considered exhaustive. Rather, it represents a lower-bound benchmark, where failure to meet this baseline indicates a fundamental limitation in LLM capabilities. Future research should expand this dataset to allow for more detailed investigations.

Furthermore, default hyperparameters were set for all models to ensure comparability. However, this approach limited the ability to gain insight into the impact of temperature or sampling settings on stability and diversity. Future work should systematically vary these parameters to identify configurations that could improve the reliability of reasoning. The evaluation only considered true-positive matches, excluding novel but potentially feasible model-generated scenarios and preventing an assessment of creativity and technical plausibility. Incorporating expert review in future studies would enable the validation of new cases and improve interpretability. Finally, the study did not examine small-scale and open-source models, systematic testing could reveal differences in performance compared to large-scale systems. If these models were found to be competitive, they could provide cost-efficient, privacy-preserving and locally deployable alternatives.

Model Evaluation Improvements – The study indicates that LLMs rely more on statistical pattern matching than on genuine engineering reasoning due to their lack of contextual and domain-specific knowledge. To overcome these limitations, future work should focus on strengthening contextual grounding and integrating domain knowledge directly into the reasoning process. The study applied a zero-shot setup, which isolated the model's intrinsic capabilities but withheld contextual cues that are particularly relevant for complex repurposing scenarios. More informative prompting strategies, such as context-rich, one- or few-shot prompts with explicit boundary conditions and property lists, could provide the necessary structure for more reliable technical reasoning. Additionally, exclusively using general-purpose models without access to external knowledge prevented grounding in verified engineering data. Retrieval-augmented generation approaches and fine-tuning with curated domain datasets offer promising ways to compensate for missing context and enhance domain-specific problem understanding.

https://doi.org/10.1017/pds.2026.10483 Published online by Cambridge University Press

5. Conclusion

Identifying technically feasible repurposing opportunities requires linking the characteristics of components with alternative system contexts through expert reasoning about functions, interfaces and boundary conditions. This study examined to which extent LLMs can support this process by identifying repurposing scenarios and the relevant properties required to assess their feasibility. Three state-of-the-art models (GPT-5, Claude 4.5 Sonnet and Gemini 2.5 Pro) were evaluated under controlled zero-shot conditions using a validated dataset of 72 documented repurposing cases from peer-reviewed literature. The results show that LLMs can reliably identify low-complexity scenarios. However, performance was low for high complexity repurposing scenarios, indicating that current models struggle to identify more complex repurposing cases. Property identification revealed greater limitations, while frequently documented properties were recognised consistently, properties occurring rarely in the dataset were identified only sporadically, and no model returned a complete property set in any single run. There were small performance differences between the investigated models, suggesting that task characteristics such as complexity and property frequency are the dominant influencing factors. Overall, the findings indicate that LLMs can provide support in the identification of possible repurposing scenarios but currently lack the domain-specific reasoning capabilities needed for reliable feasibility assessment. These results establish a methodological baseline and highlight key research directions for advancing the use of LLMs in repurposing.

- Analysing hyperparameter sensitivity and prompt-structuring techniques, including few-shot approaches
- Integrating external engineering knowledge via retrieval-augmented generation or domain-specific fine-tuning, particularly for smaller models
- Incorporating expert validation of generated scenarios to enhance evaluation reliability.

Together, these directions outline pathways towards developing LLM systems that are reliable for generating and assessing the feasibility of repurposing scenarios.

References

Ahmed, W., Bäckstrand, J., Siva, V., Sarius, N., & Sundberg, H.-Å. (2025). Extending lifespan: Wear and non-wear component analysis in end-of-life products. *MethodsX*, 14, 103423. https://doi.org/10.1016/j.mex.2025.103423

Ali, M. I. M., Shaukat, M. M., Merah, N., & Pashah, S. (2020). Product repurposing: typology and design considerations. *Journal on Innovation and Sustainability RISUS*, 11(2), 18–32. https://doi.org/10.23925/2179-3565.2020v11i2p18-32

Anthropic. (2025, September 29). Introducing Claude Sonnet 4.5. https://www.anthropic.com/news/claude-sonnet-4-5

Bauer, T. (2018). Concevoir un produit pour plusieurs vies: Propositions pour la conception et l'évaluation environnementale de solutions en cascade.

Bruyninckx, H., Hatfield-Dodds, S., Hellweg, S., Schandl, H., Vidal, B., Razian, H., Nohl, R., Marcos Martinez, R., West, J., Lu, Y., Miatto, A., Lutter, F. S., Giljum, S., Lenzen, M., Li, M., Cabernard, L., Fischer-Kowalski, M., Kulionis, V., Oberschelp, C., & Silva, D. (2024). UN Global Resources Outlook 2024: Bend the trend – Pathways to a liveable planet as resource use spikes (A Report of the International Resource Panel).

Coughlan, D., Fitzpatrick, C., & McMahon, M. (2018). Repurposing end of life notebook computers from consumer WEEE as thin client computers – A hybrid end of life strategy for the Circular Economy in electronics. *Journal of Cleaner Production*, 192, 809–820. https://doi.org/10.1016/j.jclepro.2018.05.029

Dörnbach, A., Luttmer, J., & Nagarajah, A. (2024, August 25). Design Guidelines for Repurposing: A Comprehensive Methodological Literature Review. Volume 5: 29th Design for Manufacturing and the Life Cycle Conference (DFMLC). ASME 2024 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Washington, DC, USA. https://doi.org/10.1115/detc2024-143700

Ellen MacArthur Foundation. (2015, Dezember 1). Towards a circular economy: Business rationale for an accelerated transition. https://www.ellenmacarthurfoundation.org/towards-a-circular-economy-business-rationale-for-an-accelerated-transition

European Union. (2024). Regulation (EU) 2024/1781 of the European Parliament and of the Council of 13 June 2024 establishing a framework for the setting of ecodesign requirements for sustainable products, amending Directive (EU) 2020/1828 and Regulation (EU) 2023/1542 and repealing Directive 2009/125/ECText with EEA relevance.

https://doi.org/10.1017/pds.2026.10483 Published online by Cambridge University Press

Google. (2025, März 25). Gemini 2.5: Our most intelligent AI model. Google. https://blog.google/technology/google-deepmind/gemini-model-thinking-updates-march-2025/
Hewa Witharanage, S. D., Li, W., Otto, K., & Holtta-Otto, K. (2024). Identifying Opportunities to Repurpose Decommissioned Products Using Large Language Models (SSRN Scholarly Paper No. 4985604). *Social Science Research Network*. https://doi.org/10.2139/ssrn.4985604
Hewa Witharanage, S. D., Otto, K., Li, W., & Holtta-Otto, K. (2025). Repurposing as a Decommissioning Strategy for Complex Systems: A Systematic Review. *Journal of Mechanical Design*, 147(5). https://doi.org/10.1115/1.4067090
Jeppe, A., Proff, H., & Eickhoff, M. (2023). Economic Potentials of Ecologically Attractive Multi-Life Products—The Example of Lithium-Ion Batteries. *Sustainability*, 15(14), 11184. https://doi.org/10.3390/su151411184
Kurt, A., Dat, V., Mangione, F., Cortes Cornax, M., & Front, A. (2019). An Extended Circular Supply Chain Model Including Repurposing Activities (S. 6). https://doi.org/10.1109/ICCAD46983.2019.9037929
OpenAI. (2025, August 7). Entdecke GPT-5. https://openai.com/de-DE/index/introducing-gpt-5/
Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ*, 372, n160. https://doi.org/10.1136/bmj.n160
Pluhnau, R., Lübke, J., & Nagarajah, A. (2023). Multi-Life-Products – towards a new paradigm of product development aligning on module's multiple usage-scenarios. *Procedia CIRP*, 119, 873–878. https://doi.org/10.1016/j.procir.2023.03.132
Potting, J., Hekkert, M. P., Worrell, E., & Hanemaaijer, A. (2017). Circular Economy: Measuring innovation in the product chain.
Psarommatis, F., & May, G. (2025). A Cost–Benefit Model for Sustainable Product Reuse and Repurposing in Circular Remanufacturing. *Sustainability*, 17(1), 245. https://doi.org/10.3390/su17010245
Psarommatis, F., May, G., & Azamfirei, V. (2025). Product reuse and repurpose in circular manufacturing: A critical review of key challenges, shortcomings and future directions. *Journal of Remanufacturing*, 1–38. https://doi.org/10.1007/s13243-025-00153-y
Reike, D., Vermeulen, W., & Witjes, S. (2017). The circular economy: New or Refurbished as CE 3.0? — Exploring Controversies in the Conceptualization of the Circular Economy through a Focus on History and Resource Value Retention Options. *Resources, Conservation and Recycling*, 135. https://doi.org/10.1016/j.resconrec.2017.08.027
Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., & Chadha, A. (2025). A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications (No. arXiv:2402.07927). arXiv. https://doi.org/10.48550/arXiv.2402.07927
Singh, V., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, 126. https://doi.org/10.1007/s11192-021-03948-5
Wackernagel, M., & Lin, D. (2023). Earth Overshoot Day. In N. Wallenhorst & C. Wulf (Hrsg.), *Handbook of the Anthropocene* (S. 569–572). Springer International Publishing. https://doi.org/10.1007/978-3-031-25910-4_90
Xu, F., Lin, Q., Han, J., Zhao, T., Liu, J., & Cambria, E. (2024). Are Large Language Models Really Good Logical Reasoners? A Comprehensive Evaluation and Beyond (No. arXiv:2306.09841). arXiv. https://doi.org/10.48550/arXiv.2306.09841

https://doi.org/10.1017/pds.2026.10483 Published online by Cambridge University Press