**INTERNATIONAL DESIGN CONFERENCE – DESIGN 2024**
https://doi.org/10.1017/pds.2024.198

# Generative large language models in engineering design: opportunities and challenges

Filippo Chiarello [1,✉], Simone Barandoni [1], Marija Majda Škec [2] and Gualtiero Fantoni [1]

[1] University of Pisa, Italy, [2] University of Zagreb Faculty of Mechanical Engineering and Naval Architecture, Croatia

✉ filippo.chiarello@unipi.it

**Abstract**

Despite the rapid advancement of generative Large Language Models (LLMs), there is still limited understanding of their potential impacts on engineering design (ED). This study fills this gap by collecting the tasks LLMs can perform within ED, using a Natural Language Processing analysis of 15,355 ED research papers. The results lead to a framework of LLM tasks in design, classifying them for different functions of LLMs and ED phases. Our findings illuminate the opportunities and risks of using LLMs for design, offering a foundation for future research and application in this domain.

*Keywords:* *artificial intelligence (AI), design process, generative AI, large language model (LLM), natural language processing (NLP)*

## 1. Introduction

Recent advances in the dynamic field of artificial intelligence (AI) render generative large language models (LLMs) a critical element in advancing engineering design (ED) processes. Language models are probabilistic models that facilitate the processing of natural language via algorithms. The term 'large' in LLMs signifies the extensive number of parameters involved in their training, while 'generative' denotes their capability to produce text. These models, particularly notable for text generation, have found numerous applications, from design idea generation to enhancing communication in design teams. These systems impact a wide range of human activities, and their user base has skyrocketed, surpassing 100 million as of January 2023 (Huang et al., 2023). ChatGPT, developed by OpenAI, exemplifies the rise of generative LLMs.

Extensive research has been conducted on Generative Models in the field of ED, with a focus on generative and parametric techniques to enhance the design process since the 1990s. Unfortunately, only a few works focus on text generation and even fewer on the use of LLMs. As the field of engineering design is knowledge-intensive, with much of its knowledge encapsulated in text, this presents a gap in the literature. Research needs to establish a direct link between the potential uses of Generative LLMs and the specific needs and stages of the design process, as well as with the potential practical applications in ED. Furthermore, studying the behaviour of these models and their user interactions presents challenges. Their 'black box' nature, owing to the immense size (e.g., ChatGPT's 175 billion parameters), raises questions in the domain of explainable artificial intelligence. Additionally, the proprietary nature of model prompts and responses limits a comprehensive understanding of LLM behaviour.

Considering these gaps and the challenging characteristics of LLMs, this paper aims to give a quantitative view of how generative LLMs can change the ED process. We conducted a study using a textual analysis of all the research papers published in journals identified as relevant for ED by the

Design Society, also adding the papers published in the two leading ED conferences (ICED and Design), resulting in 15,355 research papers. Then, we introduce a methodology that leverages Named Entity Recognition (NER), enabling us to map out the diverse uses of ChatGPT in ED. Starting from a series of actions that can be performed by ChatGPT (Barandoni, 2022), we identified these actions (e.g., create, assess, translate) in the scientific literature and extracted the object they are related to (e.g., idea, social media, environmental impact). We thus listed the potential activities that LLMs can impact in ED. Finally, we propose a framework of applications of LLMs in ED, focusing on the classes of actions performed by LLMs (generate, measure, and translate knowledge) and the phases of ED (Problem Definition, Conceptual Design, Embodiment Design and Detailed Design) (Chiarello et al., 2021). The results shed light on how generative LLMs shape the ED process and set the stage for future research challenges in the field.
The paper is structured as follows. Section 2 introduces generative AI and LLMs within the context of ED, highlighting the nascent exploration of LLMs despite the established history of generative methods. Section 3 details the method, with the development of a rule-based Named Entity Recognition (NER) approach to identify LLM-applicable tasks in ED. Section 4 discusses the emerged LLMs' roles in ED. Section 5 concludes by discussing the implications of integrating LLMs in ED.

## 2. Generative AI and LLMs in Design

Generative AI tools have widely been studied in the context of Design (Thoring et al., 2023; Regenwetter et al., 2022). The development of generative and parametric methods to aid the design process began as early as the 1990s (Gunaratnam and Gero, 1994). More recently, Sarica et al. (2019) examined the use of text-mining methods for the symbolic representation of design knowledge and the automatic generation of ontologies from patent documents. Nobari et al. (2021) examined the use of generative adversarial networks for innovative bicycle designs. In digital manufacturing, Buonamici et al. (2020) described the use of generative AI in developing alternative, weight-efficient designs for a robot arm component.
As it is evident from this brief picture of the state of the art, textual generation with LLMs is still understudied in the context of ED. The fundamental methodology underpinning LLMs, Natural Language Processing (NLP), is different. These techniques have been pivotal in evolving design processes, fundamentally transforming the way design knowledge is encoded, evaluated, and enhanced (Siddharth et al., 2022). NLP's key strengths lie in its ability to tokenise text data, enabling the execution of tasks such as similarity measurement, topic extraction, and sentiment analysis. This has opened avenues for knowledge reuse or elicitation of users and their needs (Chiarello et al., 2020). The versatility of NLP is evident as it supports various design process applications, from brainstorming to detailed prototyping (Han et al., 2022). Particularly relevant is the building of ontologies with NLP. Researchers have demonstrated how ontologies can be leveraged to reduce ambiguity, improve coherence in sentence structures, and ultimately aid in the retrieval of design knowledge that forms the foundation of artefacts in the design process (Ahmed & Štorga, 2009). It is interesting to notice that LLMs go in the opposite direction. If ontologies are top-down approaches to knowledge modelling for machines led by engineers and designers, machine learning (ML) models are bottom-up, with models emerging from a large amount of data.

## 3. Method

### 3.1. Data collection and text preprocessing

We collected all the papers published in design-related journals indexed in Scopus. We took the list of the Design Society as a reference[1], also adding the "International Conference on Engineering Design (ICED)" and "DESIGN conference". In total, 15,355 scientific articles were gathered, from which we

[1] List of Journals: Artificial Intelligence for Engineering Design, Analysis and Manufacturing (AI EDAM); CoDesign - International Journal of CoCreation in Design and the Arts; Design Science Journal; Journal of Design Research; Journal of Engineering Design; Journal of Mechanical Design; Research in Engineering Design; The International Journal of Design Creativity and Innovation.



https://doi.org/10.1017/pds.2024.198 Published online by Cambridge University Press

extracted the title, the abstract, and the keywords. We merged the abstract, title, and keywords of each article in a single piece of text. Then, we applied a standard preprocessing procedure on the texts (Puccetti et al., 2023): lowercasing (conversion of the text to lowercase), lemmatisation (reduction of each word to their dictionary form, called lemma), and Part-of-Speech (POS) Tagging (assignment of grammatical categories to each word). These steps standardise the forms tasks can assume in the text (lowercasing and lemmatisation) and add information to the text (POS Tagging), consequently increasing the probability of task matching.

## 3.2. Named Entity Recognition (NER)

In this work, we developed a rule-based Named Entity Recognition (NER) methodology. For a panoramic view of different approaches for NER in ED, see Puccetti et al. (2023). Rule-based NER systems function by establishing specific rules that incorporate regular patterns and linguistic structures to detect entities in text. In this paper, the named entities we want to extract are tasks that ChatGPT can perform. To develop our NER methodology for task extraction, we focused on the identification of the three components of tasks: verb + (object + characteristic).

To collect the entity verb (i.e., actions that LLMs can perform), we utilised a dataset from Kaggle comprising 3.8 million tweets discussing LLMs (Barandoni et al., 2023). This dataset includes specific tasks performed by LLMs, as described by the users asking for these tasks. We began by isolating a collection of 503,000 tasks and extracted the primary verb from each task. We identified 13,531 unique verbs. We eliminated verbs that appeared in fewer than 0.01% of the tasks. Following this, we thoroughly reviewed the remaining 977 verbs, selecting those most pertinent to tasks within the ED domain. The verbs were screened against practical applications in engineering design. This involved the authors independently assessing whether a verb corresponds to an action that designers and engineers perform in real-world scenarios. After this personal revision, the authors find a consensus on the verbs that were not classified coherently by all the authors (2.3%). Consequently, we finalised a list of 37 verbs: analyse, answer, assess, brainstorm, browse, chat, clarify, code, convert, create, define, describe, detect, draft, estimate, evaluate, explain, extract, generate, guess, interpret, know, list, measure, predict, process, read, recommend, respond, review, rewrite, solve, suggest, summarise, translate, understand, write. The verbs were considered both in English and American spelling (e.g., analyse and analyse). As can be seen, these verbs are information-processing verbs, which are essential for Design Theory (Johannesson & Perjons, 2014).

To extract the entity object (related to the verb), we used the dependency trees (Honnibal et al., 2017). The dependency tree of a text represents its grammatical structure by illustrating the relationships between words of the sentences, showing connections such as subject, object, modifier, and other syntactic relationships through directed links or edges. To accurately identify tasks, we formulated a specific rule which follows the pattern: verb + object + characteristics, where the verb is one of the 37 verbs previously identified; the object is the word that has the syntactic relation "object" to the verb similarly to the Subject-Action-Object (SAO) methodology (Shankar Bhattacharjee et al., 2016); and the characteristics, ranging from one to four words (articles, conjunctions, and prepositions excluded) are dependent on the object. The characteristics allow matching tasks where the object is a compound term. Examples of the extracted objects are "ideas", "patents", and "procedures". Examples of objects + characteristics are "user need", "design concept", and "online review".

## 3.3. Tasks classification

To guide towards a framework to explore where and how generative LLMs can impact the ED steps, we classified the tasks that LLMs can perform. A two-level classification was used. For both levels, the employed approach was qualitative and based on the expertise of the authors, who performed the tasks separately and then merged the results together. This involved the three authors independently assessing whether a verb belong to the class or not. In the first level, the tasks were grouped considering the verb. These verbs, as described before, are typical actions that LLMs can perform. After revising these verbs, we classified them into three classes, referring to the function that they can operate (revised from Johannesson & Perjons, 2014). *Generate* are tasks related to creating new knowledge; *Evaluate* comprises tasks of assessing, evaluating, and analysing knowledge; and *Describe* are tasks associated



https://doi.org/10.1017/pds.2024.198 Published online by Cambridge University Press

with converting knowledge from one form to another to make it more accessible or applicable to a different audience, context or purpose. The second classification level is linked to the phases of ED, as defined by Chiarello et al. (2021): *Problem Definition, Conceptual Design, Embodiment Design, and Detailed Design.* The authors evaluated each entity against the definitions of the classes to determine the most appropriate classification. We measured an inter-rater reliability between our independent decisions by calculating Fleiss' Kappa. The achieved scores were 81% for verb classification and 88% for ED phase classification.

## 4. Results

Out of 15,355 scientific papers related to ED, 7,034 (45.08%) contained at least one task. Thus, roughly 40% of our scientific works could be impacted by generative LLMs. In this section, we leverage the results of our analysis to speculate how. Our extraction process yielded a total of 12,130 tasks (9,588 unique) and 6,140 unique objects, demonstrating the varied potential applications of LLMs in ED.

Table 1 categorises ED papers according to the primary functions of generative LLMs: 'Generate', 'Evaluate', and 'Describe'. This classification provides the number and percentage of papers aligned with each category, further detailed by the top five most frequent verbs associated with these functions.

The 'Generate' category encompasses 17.88% of the papers, featuring verbs such as 'create', 'suggest', and 'answer' among the top five. Predominantly driven by the nature of the technology, the initial and most apparent application of generative LLMs is in supporting creativity. However, this may introduce a research bias, as ED encompasses a broader spectrum of tasks beyond generation, as seen from the following lines of the table.

'Evaluate', the most represented class with 42.77% of papers, includes verbs like 'solve' and 'understand'. This indicates that design research predominantly focuses on evaluation tasks, which LLMs can significantly augment. This presents both opportunities and challenges for researchers and practitioners, especially considering the importance of critically reviewing and validating LLM outputs in these contexts, where inaccuracies (i.e., "hallucinations") could lead to more significant repercussions, particularly for less experienced designers. This risk is exacerbated by the typically well-formatted and persuasive nature of LLM outputs.

Lastly, the 'Describe' tasks, accounting for 18.35% of papers, are almost on par with 'Generate'. This group includes verbs like 'define' and 'explain', highlighting LLMs' potential to transform how designers transfer knowledge within and across teams. Often undervalued in design practice (Moses et al., 2023), the ability to communicate effectively is vital, mainly as it influences the efficiency and success of transitioning between different design phases or in collaborations between marketing, design, and engineering teams. The observation that generative LLMs can impact this category of activity opens new avenues for guiding design practices in this direction, even if an over-reliance on automation might stifle creativity and reduce the opportunity for human designers to engage deeply with the problem-solving process.

**Table 1. Distribution of ED papers by classes and specific verbs (top 5) that LLMs can perform, alongside the percentage of scientific papers discussing them**

| LLM Verb Class | N. of Papers (%) | LLM Verbs (% of Papers) |
|---|---|---|
| Generate | 2,745 (17.88) | generate (7.67); create (7.30); suggest (1.71); answer (0.94); list (0.08) |
| Evaluate | 6,567 (42.77) | evaluate (6.72); solve (6.18); understand (5.77); assess (4.46); analyse (4.17) |
| Describe | 2,818 (18.35) | describe (9.71); define (3.94); explain (1.77); clarify (0.74); translate (0.59) |

Table 2 shifts the focus to the four phases of ED: for each one, it shows the three LLMs Verbs Classes and the top 2 to 5 corresponding objects of these verbs, alongside the percentage of scientific papers discussing the objects. The table offers insights into the potential roles of LLMs in each ED phase, which are discussed in the following sections, where the authors critically reviewed the results and discussed them in relation to previous relevant literature in the context of ED.



https://doi.org/10.1017/pds.2024.198 Published online by Cambridge University Press

Table 2. Distribution of Language Model Verbs Class in Engineering Design Phases and the top 2 to 5 corresponding objects of these verbs, alongside the percentage of scientific papers

| ED Phase | LLM Verb Class | Object (% of Papers) |
|---|---|---|
| 1- Problem Definition | Generate | research question (0.22); need (0.08); user insight (0.07); questionnaire (0.03); persona (0.02) |
| | Evaluate | customer profile (0.05); ill structure problem (0.03); customer preference (0.05); market (0.02); analysis technique (0.01); social media (0.01) |
| | Describe | case study (0.05); experience (0.05); narrative text (0.03) |
| 2-Conceptual design | Generate | creative idea (0.15); new design (0.09); printable variant (0.08); design concept (0.06) |
| | Evaluate | need (0.14); social/environmental impact (0.14); idea (0.07); function (0.07); novelty (0.05); fixation effect (0.05) |
| | Describe | concept (0.14); problem (0.08); need (0.05); design specification (0.04); design concept (0.03) |
| 3-Embodiment design | Generate | solution (0.18); geometry model (0.08); alternative solution (0.05); functional model (0.05); functional part (0.02) |
| | Evaluate | performance (0.59); effectiveness (0.47); optimization problem (0.24); reliability (0.12); engineering problem (0.11) |
| | Describe | function (0.1); requirement (0.07); model knowledge (0.01); simulation model (0.01) |
| 4-Detailed design | Generate | design solution (0.04); design image (0.03); feasible solution (0.02); patent (0.02); optimal solution (0.02) |
| | Evaluate | cost (0.14); isomorphism (0.11); configuration problem (0.05); assembly variation (0.03); requirement change (0.03); prototype (0.02) |
| | Describe | procedure (0.07); prototype system (0.03) |

## 4.1. Problem definition

**Generate:** LLMs may facilitate a systematic approach to defining *research questions* by leveraging extensive datasets and historical examples (including papers, reports, and policy documents), also mitigating human biases, which are crucial for objective question formulation (Agyemang et al., 2023). However, it's advisable to complement this with qualitative, expert-driven searches to counteract biases in the uploaded documents. Also, LLMs may enable rapid exploration of *user needs and insights*, analysing users' language more deeply than legacy NLP systems in the ED context (Chiarello et al., 2023). It will be interesting to see if LLMs can identify latent needs from large textual datasets (Yuan et al., 2023). LLMs may resolve language issues and handle datasets in multiple languages, considering users with different backgrounds simultaneously. However, there is a risk that ideas generated may only sometimes be practically feasible and might inadvertently steer designers towards standard solutions, reducing true innovation.
Furthermore, LMMs may support the creation of well-designed *questionnaires*, a crucial task in accurately capturing the voice of the customer (Mugge et al., 2023). LLMs can analyse existing data to suggest pertinent areas of inquiry and identify patterns in user responses that humans might overlook. Moreover, LLMs can tailor questionnaires to different user segments, ensuring comprehensive and representative data collection. Finally, LLMs can analyse large volumes of user data to identify common characteristics, behaviours, and preferences, aiding in segmenting the user base into distinct *personas*. They may assist in updating personas over time, keeping them relevant as user needs and market dynamics evolve. By providing insights potentially missed in manual analysis, LLMs enable the creation of more accurate and representative personas (Stevenson & Mattson,2019). The challenge lies in ensuring that the generated content remains relevant and closely aligned with the real-world context of the users' needs.
**Evaluate:** Evaluating *customer profiles* is crucial during the problem definition phase. LLMs can enhance the quality of these analyses by scanning benchmarks of older profiles, assisting designers in



https://doi.org/10.1017/pds.2024.198 Published online by Cambridge University Press

creating new ones more efficiently and effectively. This process can be further enriched by evaluating *social media content* (Chiarello et al., 2020). LLMs facilitate code-free analysis of this data source, empowering designers without coding experience. However, a potential drawback is the overemphasis on quantitative data, which overshadows important qualitative aspects in the problem definition. The data show that LLMs may support assessing the structure of *ill-structured problems*. It is intriguing to consider whether generative LLMs can aid in this phase by helping designers assess the quality of problem definitions and guiding them on the time allocation for this phase. Finally, LLMs' support in evaluating *customer preferences* and the *market* can lead to a deeper, data-driven understanding of the problem space, aiding in assessing the value of delving into specific design problems. Here, too, the collaboration between designers, engineers, and machines is critical in assessing this space. It is essential to determine what measures to use and how to integrate the human perspective in evaluating customer preferences, considering the qualitative aspects of the problem.

**Describe:** Effective communication of *case studies* and *design experiences* is crucial for knowledge transfer, allowing non-technical stakeholders to learn from real-world applications, successes, and failures, aiding in understanding the complexities of design problems and applying theoretical concepts in practice. However, the clear and comprehensive communication of these experiences is often challenging due to their intricacies and contextual specifics. LLMs can assist in synthesising vast data from case studies into structured and understandable formats. Despite these benefits, there is a risk that LLMs might need to be more concise or more context-specific subtleties, potentially leading to a loss of critical insights. LLMS can also be helpful in translating technical texts into *narrative texts.* Cummings & Teal (2023) highlight the effectiveness of design fiction and dialogic methodologies, especially in the initial phases of design. However, there is a risk that the subtleties and nuances of the design problem may be overlooked if descriptive tasks rely solely on LLMS, underscoring the importance of adequate human interpretation and judgment.

## 4.2. Conceptual Design

**Generate:** Augmenting the *creative idea* process with a vast array of possibilities, LLMS can lead to a proliferation of new design alternatives. However, the generation of these elements must be carefully curated to ensure their relevance and feasibility. Excessive generation in the conceptual design phase can result in too many choices, potentially causing analysis paralysis in the design process. Also, Data indicates that LLMS can aid in generating *printable variants* of products. Investigating how LLMS can optimise design parameters for 3D printing and additive manufacturing processes is intriguing. These models have the potential to enhance both the efficiency and quality of final products by analysing extensive datasets to identify optimal design configurations, material choices, and printing strategies. This approach can significantly reduce trial-and-error during the prototyping phase.

**Evaluate:** *Needs* evaluation is a critical step in ED, particularly in understanding customer profiles more deeply. If the problem definition phase is successful, designers may find themselves evaluating various needs. LLMS can provide a holistic view of needs evaluation. By inputting needs and evaluation criteria into LLMS, designers can get an initial assessment of these needs, which they can then refine. However, a potential drawback of using LLMS for evaluation is their current limitations in understanding complex interdependencies and long-term consequences, areas where human experts typically excel. Also, LLMs can assist in evaluating the *novelty* of an *idea* (Lee et al., 2023). LLMS, with their access to large textual datasets, can streamline this process. Future research could compare legacy machine learning systems against LLMS to aid in this phase. LLMS are likely to be more accurate and require less training data, enabling one-shot learning. This could lower the cost of using these systems in ED, democratising the tools. LLMS might also help overcome fixation effects by providing objective critiques to challenge existing ideas. Also, LLMs may assist in analysing the potential *social and environmental impacts* of design concepts, thereby aligning them with sustainability goals. This capability could enhance the sustainability performance of designers, shifting their focus from merely the technical aspects of a solution to encompassing the entire design process (Stevenson et al., 2023). LLMs can be fine-tuned using sustainability frameworks, guidelines, or even sustainability reports from other companies, aiding in the evaluation of the sustainability impact of proposed solutions. Finally, LLMS can assist by ensuring the proper form and content of *functions* and functional analysis. Here, experts are crucial in providing guidelines to LLMS for appropriate function assessment, and here, a fine-tuning of the LLMs can be fundamental. Past literature on functional design will be important in guiding the correct use of LLMS (She et al., 2022), which can execute measurements well but could be more adept at designing metrics, especially when dealing with complex concepts like engineering functions.

**Describe:** In the conceptual design phase, accurately describing the design *concept* and identifying the design *problem* or *need* is crucial. As this phase often involves collaboration with other functions, such as marketing and engineering, effective communication of outputs is essential, a step that is frequently overlooked by technical personnel. Generative LLMs can assist in tailoring the description of design concepts, problems, and needs to suit different audiences. Also, LLMs may ensure a comprehensive and clear presentation of *design specifications*, adhering to regulatory standards. They can analyse vast repositories of design documents, learning from existing examples to generate appropriate specifications. Leveraging their extensive training on diverse datasets, LLMs can suggest specifications that align with industry standards and best practices, ensuring both innovation and compliance. Furthermore, LLMs' ability to process and interpret natural language enables them to refine complex technical jargon into more accessible language. Anyway, the nuances of context-specific requirements, unique innovation elements, and the designer's intent might need to be fully captured by a model, risking the loss of nuanced understanding crucial for high-quality design specifications.

## 4.3. Embodiment Design

**Generate:** LLMs can enable rapid exploration of multiple embodiment design *solutions*. They could play a crucial role in optimising design parameters through the analysis of patterns and correlations within extensive datasets, aligning designs with both technical specifications and practical constraints. Again, it is essential to manage the generation of solutions to avoid an excess of non-viable options. Also, integrating language models with 3D models could enable the generation of *geometry models*. A critical challenge in this integration is aligning generated models with the practical constraints of materials and manufacturing processes, which an LLM may need to correctly model (Nie et al., 2021). Finally, the impact of LLMs in supporting *functional model* creation is noteworthy, as functional modelling is a language-intensive task. Fine-tuning LLMs specifically for this task, with examples of previous functional models, could enhance their performance.

**Evaluate:** The results show that the evaluation tasks show a more significant opportunity for application in the Embodiment design phase with respect to generating and describing related tasks. Proper judgment of design *performance* in this phase is crucial (Andersson, 2020). LLMs can scour existing scientific literature or patents to suggest existing similar tests, potentially increasing the quality of the evaluation. LLMs may also contribute significantly to risk assessment by identifying potential failure points through the analysis of historical data and similar projects. This enables the development of pre-emptive mitigation strategies, enhancing the *reliability* and safety of the design. Their proficiency in analysing extensive datasets allows them to identify underlying patterns and trends indicative of potential reliability issues or areas for improvement. The predictive modelling capabilities of LLMs provide foresight into potential design failures, helping designers simulate a range of operational conditions and stress scenarios to evaluate the resilience of designs under various what-if scenarios.

**Describe:** *Functional* analysis is a unique language spoken by a subset of designers and has been extensively studied in the context of AI-aided design (Fantoni et al., 2013). LLMs can assist in translating functional models into natural language and provide different levels of abstraction for functional modelling, depending on how the model should be used. For example, experiments can be made to test if a low-level functional model, which incorporates many technical details of how the design systems work, can be synthesised in a more abstract functional model, which can be more beneficial for communication with non-technical figures in the company. However, there is a risk of losing some information during translation, which is why designers should view this as a support tool for communication rather than a replacement for human expertise. Similarly, the translation of users' needs into technical *requirements*, a critical step of the ED phase, can be helped by LLMs. Design methods like Quality Function Deployment (QFD) can be improved and utilised more effectively by simplifying the comparison of large datasets of technical/engineering design requirements (Franceschini & Rossetto,



https://doi.org/10.1017/pds.2024.198 Published online by Cambridge University Press

1995). LLMs can also help designers write more readable requirements, contributing to requirement elicitation (Cheligeer et al., 2022).

### 4.4. Detailed Design

**Generate:** Using LLMs to generate feasible and optimal final *design solutions*, building on the analysis and work done in previous phases, is an intriguing concept, as old as the use of AI in design (Schön, 1992). While we are still far from achieving this, LLMs can assist designers in considering all the documentation created in previous phases to help generate preliminary design solution alternatives within certain constraints. However, this approach might limit the designers' creativity in developing out-of-the-box solutions. Similarly, the use of multimodal systems that manage both text and *design images* can streamline the creation of figures from existing document repositories. This can also aid designers in rapidly searching databases of existing solutions by visualising the design, a known method for quickly searching large information sets. As stated in the introduction, the literature in this area is more extensive than that focused on language generation and LLMs. It will be intriguing to see how multimodal systems will be studied, leveraging existing literature on the topic (Han et al., 2018). Finally, with the advancement of LLMs, the topic of AI-aided *patent* generation may attract interest from both theory and practice. If an invention is worth patenting, the vast amount of knowledge generated in the previous design phases can be utilised to feed LLMs and support the generation of a patent. However, there are many risks, ranging from designers' decreased ability to read and write patents to intellectual property damages.

**Evaluate:** The results indicate that in the Detailed Design phase, the focus on evaluation is higher than generation, as supported by previous research (Shafqat et al., 2019). This opens more possibilities for ED research and practice in using LLMs for Evaluating and not just for generating knowledge. Automating these tasks with LLMs could optimise design decisions against *design costs* and manage the large quantity of knowledge previously developed. However, if not carefully balanced, there is a risk of overemphasising cost at the expense of other critical factors like user experience or long-term sustainability. Also, in the context of *configuration problems* (Brown & Hwang, 1993), LLMs cannot still manage spatial problems, requiring integration with other tools. Multimodal generative systems, mixing text and 3D models, can be game changers in this context. This integration may allow for the consideration of previous design information in solving configuration problems. A similar discourse applies to *assembly variation*. In this final phase, LLMs may also quickly analyse and categorise incoming *requirements*, assess their impact on the current project scope, and suggest necessary adjustments. They can identify inconsistencies and potential conflicts between new and existing requirements. Additionally, LLMs can automate the documentation process, maintaining a clear and up-to-date record of changes. However, LLMs may need a more nuanced understanding of context-specific details, leading to potential oversights in complex or specialised projects. By analysing design specifications, customer feedback, and market trends, LLMs can also generate insightful suggestions for evaluating *prototype* development. However, LLMs may need to capture the complexities and physical constraints inherent in material and design choices, leading to impractical suggestions. Moreover, the lack of tactile and real-world testing experience means that LLMs can only partially replicate the nuanced understanding that comes from physical prototyping.

**Describe:** LLMs can synthesise and articulate complex design concepts into clear, accessible language. This skill is crucial in bridging the gap between technical experts and diverse stakeholders in sharing design *procedures*, ensuring a coherent and consistent understanding across multidisciplinary teams, particularly in projects with extensive and intricate design processes. However, a primary concern with using LLMs is the potential for data privacy and security issues when handling sensitive design information as procedures. This challenge highlights the need for robust security measures and privacy protocols when integrating LLMs into the design process to protect confidential and proprietary information.

## 5. Conclusions

The integration of LLMs in engineering design tasks holds the potential to significantly enhance human capabilities by automating various aspects such as generation, evaluation, and description of design-related knowledge. This brings practical benefits like increased efficiency, a more comprehensive exploration of options, and improved decision-making. Theoretically, this integration could lead to a paradigm shift in the understanding and execution of design tasks. However, it is essential to ensure that this automation supports rather than replaces human expertise, maintaining a balance between computational efficiency and human-centric design values.

This paper serves as a roadmap for future research to investigate the impact of LLMs on ED. It emphasises that ED scholars should not only focus on the generative capacity of LLMs but also explore their evaluative and descriptive abilities. Moreover, the paper calls for measuring the performance of these systems in the specified application domains and studying collaboration between designers and LLMs in all the design phases. By doing so, we will guide designers in effectively utilising and studying this powerful technology to create better products and to guide future design practices.

## Acknowledgement

This work was partly funded by the *DETAILLs* Project (DEsign Tools of Artificial Intelligence in Sustainability Living LabS) - European Union. Erasmus + KA2 - Cooperation partnership in higher education (Project Number: 2023-1-IT02-KA220-HED-000158755).

## References

Agyemang, M., Andreae, D. A., & McComb, C. (2023), "Uncovering potential bias in engineering design research", *Design Science*, *9*. https://doi.org/10.1017/dsj.2023.17

Ahmed, S., & Štorga, M. (2009). Merged ontology for engineering design: Contrasting empirical and theoretical approaches to develop engineering ontologies. AI EDAM, 23(4), 391-407. https://doi.org/10.1017/S0890060409000146

Andersson, T. (2020), "Design judgement processes in mature Swedish manufacturing companies", *Journal of Design Research*, Vol. 18 No. 5-6, pp. 410-433. https://doi.org/10.1504/JDR.2020.118668

Barandoni, S., Giordano, V., Fantoni, G., & Chiarello, F. (2023), "The ChatGPT Tweets Dataset [Data set]", *Kaggle*, https://doi.org/10.34740/KAGGLE/DSV/6241027

Buonamici, F., Carfagni, M., Furferi, R., Volpe, Y., Governi, L., 2020, "Generative Design: An Explorative Study", *Computer-Aided Design and Applications*, Vol. 18, pp. 144–155. https://doi.org/10.14733/cadaps.2021.144-155

Brown, D. R., & Hwang, K. Y. (1993), "Solving fixed configuration problems with genetic search", *Research in Engineering Design*, Vol. 5, pp. 80-87. https://doi.org/10.1007/BF02032577

Cheligeer, C., Huang, J., Wu, G., Bhuiyan, N., Xu, Y., & Zeng, Y. (2022), "Machine learning in requirements elicitation: A literature review", *AI EDAM*, Vol. 36. https://doi.org/10.1017/S0890060422000166

Chiarello, F., Belingheri, P., & Fantoni, G. (2021), "Data science for engineering design: State of the art and future directions", *Computers in Industry*, Vol. 129, 103447. https://doi.org/10.1016/j.compind.2021.103447

Chiarello, F., Bonaccorsi, A., & Fantoni, G. (2020), "Technical sentiment analysis, Measuring advantages and drawbacks of new products using social media", *Computers in Industry*, Vol. 123, 103299. https://doi.org/10.1016/j.compind.2020.103299

Cummings, M., & Teal, G. (2023), "Healing fabulations: a dialogic methodology for digital codesign in health research", *CoDesign*, Vol. 19 No. 1, pp. 74-90, https://doi.org/10.1080/15710882.2022.2157837

Fantoni, G., Apreda, R., Dell'Orletta, Felice., & Monge, Maurizio (2013), "Automatic extraction of function–behaviour–state information from patents", *Advanced Engineering Informatics*, Vol. 27 No. 3, pp. 317-334. https://doi.org/10.1016/j.aei.2013.04.004

Franceschini, F., & Rossetto, S. (1995), "QFD: the problem of comparing technical/engineering design requirements", *Research in Engineering design*, Vol. 7, pp. 270-278. https://doi.org/10.1007/BF01580463

Gunaratnam, D. J., & Gero, J. S. (1994), "Effect of representation on the performance of neural networks in structural engineering applications". *Computer-Aided Civil and Infrastructure Engineering*, Vol. 9 No. 2, pp. 97-108.

Han, J., Sarica, S., Shi, F., & Luo, J. (2022). Semantic networks for engineering design: state of the art and future directions. *Journal of Mechanical Design*, 144(2), 020802. https://doi.org/10.1115/1.4052148.

Han, J., Shi, F., Chen, L., & Childs, P. R. (2018), "The Combinator–a computer-based tool for creative idea generation based on a simulation approach", *Design Science*, Vol. 4, No. 11. https://doi.org/10.1017/dsj.2018.7

Honnibal, M., and Montani, I. (2017), "spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing".

Huang, F., Kwak, H., An, J., (2023), "Is ChatGPT Better than Human Annotators? Potential and Limitations of ChatGPT in Explaining Implicit Hate Speech", *arXiv* Preprint ArXiv:2302.07736. https://doi.org/10.48550/arXiv.2302.07736

Hwang, A. H. C. (2022, April), "Too late to be creative? AI-empowered tools in creative processes", *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pp. 1-9. https://doi.org/10.1145/3491101.3503549

Johannesson, P., & Perjons, E. (2014), "An introduction to design science", Vol. 10, pp. 978-3, Cham: Springer. https://doi.org/10.1007/978-3-319-10632-8

Kicinger, R., Arciszewski, T., & DeJong, K. (2005), "Evolutionary design of steel structures in tall buildings", *Journal of Computing in Civil Engineering*, Vol. 19 No. 3, pp. 223-238.

Lee, J. W., Daly, S. R., Vadakumcherry, V., & Rodrigues, D. (2023), "Idea generation, development and selection: a study of mechanical engineering students' natural approaches and the impact of hybrid learning blocks", *Design Science*, Vol. 9. https://doi.org/10.1017/dsj.2023.26

Moses, N. D., Wojciechowski, L. R., Daly, S. R., & Sienko, K. H. (2023), "Front-end design prototyping strategies during remote stakeholder engagement", *Design Science*, Vol. 9, e24. https://doi.org/10.1017/dsj.2023.23

Mugge, R., Schoormans, J. P., & Schifferstein, H. N. (2009), "Emotional bonding with personalised products", *Journal of Engineering Design*, Vol. 20 No. 5, pp. 467-476. https://doi.org/10.1080/09544820802698550

Nie, Z., Lin, T., Jiang, H., & Kara, L. B. (2021), "Topologygan: Topology optimisation using generative adversarial networks based on physical fields over the initial domain", *Journal of Mechanical Design*, Vol. 143 No. 3, 031715. https://doi.org/10.1115/1.4049533

Nobari, A. H., Rashad, M. F., & Ahmed, F. (2021), "Creativegan: Editing generative adversarial networks for creative design synthesis". *arXiv* preprint arXiv:2103.06242.

Puccetti, G., Giordano, V., Spada, I., Chiarello, F., & Fantoni, G. (2023), "Technology identification from patent texts: A novel named entity recognition method", *Technological Forecasting and Social Change*, Vol. 186, 122160. https://doi.org/10.1016/j.techfore.2022.122160

Regenwetter, L., Nobari, A. H., & Ahmed, F. (2022), "Deep generative models in engineering design: A review", *Journal of Mechanical Design*, Vol. 144 No. 7, 071704. https://doi.org/10.1115/1.4053859

Sarica, S., Song, B., Low, E., Luo, J., (2019), "Engineering knowledge graph for keyword discovery in patent search", *Proceedings of the International Conference on Engineering Design*, Cambridge University Press, pp. 2249–2258. https://doi.org/10.1017/dsi.2019.231

Shafqat, A., Oehmen, J., Welo, T., & Willumsen, P. (2019), "The cost of learning from failures and mistakes in new product development projects: An exploratory framework", *Proceedings of the International Conference on Engineering Design, ICED*, Vol. 1, pp. 1791-1800. https://doi.org/10.1017/dsi.2019.171

Shankar Bhattacharjee, K., Kumar Singh, H., & Ray, T. (2016), "Multi-objective optimisation with multiple spatially distributed surrogates", *Journal of Mechanical Design*, Vol. 138 No. 9, 091401. https://doi.org/10.1115/1.4034035

She, J., Belanger, E., Bartels, C., & Reeling, H. (2022), "Improve Syntax Correctness and Breadth of Design Space Exploration in Functional Analysis", *Journal of Mechanical Design*, Vol. 144 No. 11, https://doi.org/10.1115/1.4054875

Siddharth, L., Blessing, L., & Luo, J. (2022), "Natural language processing in-and-for design research", *Design Science*, Vol. 8, e21. https://doi.org/10.1017/dsj.2022.16

Stevenson, P. D., & Mattson, C. A. (2019), "The personification of big data", *Proceedings of the International Conference on Engineering Design (ICED)*, pp. 4019-4028. https://doi.org/10.1017/dsi.2019.409

Stevenson, P. D., Mattson, C. A., Dahlin, E. C., & Salmon, J. L. (2023), "Creating predictive social impact models of engineered products using synthetic populations", Research in Engineering Design, Vol. 34 No. 4, pp. 461-476. https://doi.org/10.1007/s00163-023-00424-4

Thoring, K., Huettemann, S., & Mueller, R. M. (2023), "The Augmented Designer: a Research Agenda for Generative AI-Enable Design", *Proceedings of the Design Society*, Vol. 3, pp. 3345-3354. https://doi.org/10.1017/pds.2023.335