Proceedings of the Design Society, Volume 5: ICED25
https://doi.org/10.1017/pds.2025.10172

Evaluating ChatGPT's role in collaborative CAD task completeness

Jelena Šklebar ,✉, **Tomislav Martinec** , **Stanko Škec** and **Mario Štorga**

University of Zagreb Faculty of Mechanical Engineering and Naval Architecture, Croatia

✉ jelena.sklebar@fsb.hr

ABSTRACT: This study explores the role of ChatGPT in the completeness of collaborative computer-aided design (CAD) tasks requiring varying types of engineering knowledge. In the experiment involving 22 pairs of mechanical engineering students, three different collaborative CAD tasks were undertaken with and without ChatGPT support. The findings indicate that ChatGPT support hinders completeness in collaborative CAD-specific tasks reliant on CAD knowledge but demonstrates limited potential in assisting open-ended tasks requiring domain-specific engineering expertise. While ChatGPT mitigates task-specific challenges by providing general engineering knowledge, it fails to improve overall task completeness. The results underscore the complementary role of AI and human knowledge.

KEYWORDS: design engineering, computer aided design (CAD), collaborative CAD task, artificial intelligence, ChatGPT

1. Introduction

Over 50 years ago, researchers began exploring and utilising artificial intelligence (AI) techniques to apply large amounts of data representing engineering knowledge to solve engineering problems (Mcdermott, 1982). Since then, AI has evolved dramatically, becoming one of the most discussed subjects in the modern digital era. With this progress, the focus has shifted from whether AI is feasible at all to whether it can adapt as a human collaborator or a team member (Korteling et al., 2021), contributing directly and proactively to solving a problem, or whether it merely responds reactively to human queries (McComb et al., 2023), functioning solely as a support tool. Among the latter, generative large language models (LLMs) have emerged as prominent tools in the field. In addition to efficiently acquiring, storing, and applying knowledge from a variety of domains, these models are based on a vast amount of general knowledge gained from large datasets. Thus, they can generate coherent and contextually credible text outputs based on text- or image-based inputs (Memmert et al., 2024). This capability enables users to interact with these models directly in a conversational manner. Such a rise of generative LLMs has been led by ChatGPT, the OpenAI's conversational tool built on the generative pre-trained transformer (GPT) model, launched in 2022. While earlier generative LLMs were often designed for specific tasks (Memmert & Tavanapour, 2023), ChatGPT is a versatile and broadly applicable tool. Due to the vast amount of general knowledge ChatGPT has been trained on, it can be utilised for various tasks (Bouschery et al., 2023).

Generative LLMs, such as ChatGPT, have demonstrated significant potential to be integrated into the engineering design process by acting as an AI expert to assist with various engineering tasks (Wang et al., 2023). This potential is already utilised across a variety of engineering design stages and is the subject of extensive research (Khanolkar et al., 2023). Such LLMs have also opened new possibilities for engineering design knowledge acquisition, holding the potential to transform the way engineering design problems are approached (Hu et al., 2023). In addition to domain-specific engineering knowledge, an essential part of today's design engineers' qualifications is Computer-Aided Design (CAD) knowledge

https://doi.org/10.1017/pds.2025.10172 Published online by Cambridge University Press

(Mandorli & Otto, 2013). Over time, CAD has evolved from standalone to cloud-based collaborative CAD, allowing multiple designers to work synchronously on the same CAD file regardless of their location (Cheng et al., 2023). In this context, there is a potential for ChatGPT to serve as a knowledge support tool in design activities that involve execution of collaborative CAD tasks. This raises an important question about the role of ChatGPT support on the execution of such tasks. More specifically, it is important to identify which collaborative CAD tasks could benefit from the support of ChatGPT and which should be approached without ChatGPT support. Therefore, this study aims to evaluate the ChatGPT's role in collaborative CAD by examining collaborative CAD task completeness as a standardised CAD output quality dimension (Company et al., 2015). The study aims to provide an answer to the following research question: *Does the completeness of collaborative CAD tasks vary when supported by ChatGPT, considering the requirement for different types engineering knowledge?*
The paper is structured as follows. In Section 2, an overview of existing research work is presented. Section 3 describes the research methodology. Furthermore, section 4 presents the results, followed by a discussion and conclusion in Sections 5 and 6, respectively.

2. Related work

Generative LLMs have gained increasing attention for their ability to support engineers in various engineering tasks. Therefore, this section presents the research work exploring the LLMs in engineering tasks (2.1.) and the current insights of the research work of LLM in CAD specifically (2.2.).

2.1. Generative LLMs in engineering tasks

The impact of AI on the success of engineering tasks depends not only on the task itself but also on the type of AI tool being used and users' AI-related knowledge and skills, such as prompt engineering skills. Understanding this distinction is important, as task-specific AI systems are designed to be utilised in narrowly defined domains related to the specific task, thus leveraging task-specific knowledge. Significant research has shown that using task-specific or custom-made LLMs in engineering tasks can outperform traditional approaches for the same tasks. For example, studies related to the CAD domain show that LLMs utilising historical knowledge from multiple modalities (CAD models, text, and images) enhance the efficiency of the assembly process and overall manufacturing efficiency (Hu et al., 2023). They also outperform traditional approaches when searching design repositories and creating engineering knowledge bases (Meltzer et al., 2024).
In contrast, general-purpose LLMs, such as ChatGPT, are trained on broad datasets encompassing knowledge from various domains. That makes them suitable for general knowledge acquisition (Ritala et al., 2024), including engineering knowledge (Xu et al., 2024). Unlike task-specific AI tools that are limited to well-defined tasks, general-purpose LLMs can address both open-ended and well-defined tasks (Wang et al., 2023). For example, Urban et al. (2024) found that designers produced higher-quality solutions in open-ended tasks like concept generation with ChatGPT support. Moreover, participants who had prior experience with ChatGPT generated even higher-quality solutions. However, while ChatGPT, on average, outperforms humans in providing creative solutions to a problem, the best human solutions exceeded those of ChatGPT (Koivisto & Grassini, 2023). Research also shows that individuals with low creative abilities benefit from ChatGPT support in brainstorming (Memmert et al., 2024). However, when focusing only on the ideas generated independently by humans, excluding ChatGPT's, both low- and high-creativity individuals produced fewer ideas. When observing how individuals engage with the ChatGPT, low-creativity individuals request and accept more ChatGPT suggestions than high-creativity individuals. Furthermore, the output quality in open-ended engineering tasks, such as brainstorming, is highly sensitive to the prompts provided by ChatGPT (Memmert et al., 2024). Refining prompt engineering could enhance output quality (Memmert & Tavanapour, 2023). In prototyping tasks, design teams supported with ChatGPT yielded outcomes similar to human teams without ChatGPT support, but faced challenges such as ChatGPT's tendency to abandon concepts too early, unnecessary complexity, providing vague responses with forgetting previous information or prompts and design fixation (Ege et al., 2024). For well-defined engineering tasks, particularly in programming, ChatGPT yielded lower output quality compared to task-specific generative LLMs that receive code-based prompts (Mnguni et al., 2024). However, among general-purpose LLMs, ChatGPT demonstrated the highest output quality (Coello et al., 2024). Despite that, the role of user feedback in a manner that the models can understand, which is proven to be time- and resource-intensive, remains important. Furthermore, the

https://doi.org/10.1017/pds.2025.10172 Published online by Cambridge University Press

LLMs' programming style can affect the output quality, which may differ from that of the users. Engineering expertise is, therefore, essential for providing contextual understanding, particularly when dealing with ambiguities associated with problems that require domain-specific engineering knowledge (Wang et al., 2023), especially since ChatGPT may generate articulate, accurately sounding solutions that are nonetheless incorrect (Mnguni et al., 2024). That aligns with the notion that domain-specific engineering knowledge and hard skills are crucial in the interaction of user and ChatGPT (Giordano et al., 2024). Ultimately, this suggests a complementary relationship between human intelligence and AI in solving both open-ended and well-defined engineering tasks (Wang et al., 2023). Thus, ChatGPT also demonstrates potential in supporting CAD tasks since CAD encompasses both open-ended tasks like concept generation and well-defined detailed design engineering tasks. That highlights the need to explore which types of CAD tasks are best suited for ChatGPT support, particularly as these tasks require varying types engineering knowledge.

2.2. LLMs for Computer-Aided Design (CAD)

CAD has long been an essential tool used throughout various design phases. Designers utilise CAD tools to create and test digital representations of their ideas, such as virtual product concepts or prototypes, prior to manufacturing (Azemi et al., 2018). With the integration of LLMs like ChatGPT into the design process, researchers are also exploring how ChatGPT can support various CAD tasks. Those endeavours mainly focus on automating the generation of 3D objects or CAD models and variations of those models based on different ChatGPT input modalities, including text- and image-based prompts (Makatura et al., 2024). Text inputs being explored involve interpreting human language instructions, while others focus on utilising CAD programming languages such as CADQuery and OpenSCAD (Makatura et al., 2023). The findings suggest the potential benefits of using ChatGPT, such as generating CAD models representing standard mechanical engineering elements and their variations. However, it struggles with creating complex CAD models or those not representing standard mechanical engineering elements solely through text. Additionally, using ChatGPT in CAD workflows can be time-consuming and requires multiple iterations, as it generates flawed CAD models, requiring user interactions to correct them. While most studies have focused on CAD tasks that primarily require CAD knowledge, specifically the creation or modification of CAD models, there has been limited research into how LLMs like ChatGPT can assist with other types of CAD tasks. These tasks may require other types of expertise and knowledge, such as tasks that include optimisation of models based on various design requirements, including material or manufacturing technology selection.

This gap is even more pronounced in collaborative CAD, which has gained considerable research focus due to the shift from standalone CAD. This transition addresses common challenges associated with collaboration, such as limitations in cloud-based synchronous editing, issues with seamless file sharing, and visibility of design changes (Cheng et al., 2023). Researchers have focused on various collaborative CAD tasks that require different types of knowledge, including concept creation (domain-specific engineering knowledge) (Deng et al., 2022) or modification of existing CAD models (CAD knowledge) (Phadnis et al., 2021). The success of these tasks has been evaluated based on several criteria, particularly the outcome quality (Sadeghi et al., 2016), with task completeness being the important dimension (Company et al., 2015). Findings suggest that the quality of collaborative CAD outcomes depends on the context, including team size and the nature of tasks. However, there remains a research gap regarding the role of generative LLMs, such as ChatGPT, in solving collaborative CAD tasks with different types of knowledge, especially concerning the quality and completeness of such tasks.

3. Research methodology

The study is designed as an experiment to explore the influence of ChatGPT support on completeness of collaborative CAD tasks. Participants in the experiment were pairs of individuals engaged in three different collaborative CAD tasks under two experimental conditions, categorising independent variable as: pairs supported by ChatGPT and those not supported by ChatGPT.

3.1. Experimental sample and tasks

The study involved 44 mechanical engineering students (11 females and 33 males) from the Faculty of Mechanical Engineering and Naval Architecture from the University of Zagreb, spanning undergraduate and graduate levels. Participation required students to have completed both basic (first year) and

https://doi.org/10.1017/pds.2025.10172 Published online by Cambridge University Press

advanced (third year) CAD courses at the university, ensuring they possessed both declarative and procedural CAD knowledge (Chester, 2007), along with a prior understanding of design for manufacturability principles. They were divided into 22 pairs. Each pair was assigned to one of two experimental conditions: with or without ChatGPT support.

The experiment consisted of three consecutive collaborative CAD tasks aimed at integrating different types of knowledge - CAD knowledge (declarative and procedural) and knowledge from the engineering design domain, such as design for manufacturability. The first task, referred to as Task 1, involved both the declarative and procedural CAD knowledge. Participants were tasked with creating a 3D CAD model of a crankshaft by measuring and replicating a provided model in STEP format (Figure 2). Task 2 required both CAD and domain-specific knowledge. Participants were tasked with creating a functional assembly from the five models provided previously by researchers that were not dimensionally adequate: the housing, along with crankshaft, upper and lower parts of connecting rod and a piston. Unlike creating a single part, assembly creation requires understanding how components fit together and interact in a functional assembly. Domain-specific knowledge was required to maintain appropriate clearances and ensure the parts' correct alignment and interaction. Participants also needed CAD knowledge on the declarative level to select the appropriate CAD assembly operations to complete the assembly successfully. Among the five provided parts, four had visible feature trees (a crankshaft, connecting rod parts and a piston), while housing was in STEP format, meaning its feature tree was not accessible and, consequently, not modifiable, requiring the other parts to be adjusted accordingly. The third task, Task 3, required more domain-specific knowledge than the previous tasks by involving redesigning a CAD model of a crankshaft based on design-for-manufacturability principles specific to forging technology. Participants needed to consider the shape and dimensions of the model as it would be forged before any machining occurred. The model to be modified represented the same crankshaft geometry as in Task 1 (Figure 2) and included a visible feature tree, enabling participants to modify the CAD model.

3.2. Experimental setup and procedure

The experimental setup for both conditions consisted of one room with two working places facing each other, as shown in Figure 1. Each working place was equipped with an office chair and table, a high-performance computer, two monitor screens (22" with a resolution of 1920x1080 pixels), a keyboard and a mouse. In both conditions, participants utilised Onshape, a cloud-based CAD software accessed via web browser on each monitor screen. Onshape enables synchronous collaboration on a single CAD model, allowing participants to simultaneously interact with the same CAD model, follow each other's work and share views. Like other CAD software, Onshape, in addition to creating or modifying a CAD model (single part or assembly), also provides basic viewing functionalities like measuring, sectioning, moving, and hiding parts. Each monitor screen had a different setup. The left screen displayed an Onshape working document. In Task 1, this document was blank; Task 2 included an empty assembly document alongside five CAD models of parts to be assembled; and Task 3 contained a CAD model of a crankshaft that needed modification. The right screen presented another Onshape document with detailed task explanations (imported as a PDF) and the crankshaft CAD model, which participants used to create their models. Task 3 also provided the original crankshaft model in case participants needed it while modifying the model on the left screen. Pairs with the ChatGPT support had an additional web browser window on the right monitor screen with the ChatGPT application (Figure 2). To facilitate synchronous collaboration during the execution of CAD tasks, the content of the right monitor of one team member was shared on the right monitor of the other one screen mirroring. This setup enabled team members to follow each other's work in real-time, whether it involved measuring the part in Task 1 or interacting with the ChatGPT. Participants utilizing ChatGPT received support from ChatGPT-4 model, from the customized GPT *Onshape Usage and Collaboration*, developed specifically for the experiment by the authors of this study. To prevent the information from being reused for other pairs' tasks, the training option for the ChatGPT model was disabled, and its memory was cleared after each pair's session by the researcher. Additionally, all screen content was recorded for the entire duration of the experiment using OBS Studio. Video and audio recordings of the tasks were captured using a conference camera, and small video cameras were placed on each monitor screen. The experimental procedure consisted of 5 steps: 1) Pre-experiment preparation, 2) Introductory collaborative CAD session, 3) First collaborative CAD session, 4) Second collaborative CAD session, and 5) Third collaborative CAD session.

https://doi.org/10.1017/pds.2025.10172 Published online by Cambridge University Press

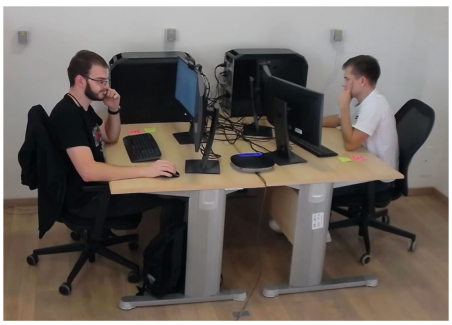


Figure 1. Experimental setup

In the first step, participants received an information package via email a few days before the experiment. The package included detailed tutorials on using Onshape, although all participants were already familiar with the software from their CAD courses at the university. Additional tutorials were provided on ChatGPT to participants in the experimental condition, focusing on the user interface, application usage and prompt engineering techniques. Consent forms were also signed in the first step. Furthermore, steps two through five took place on the same day as the experiment for each team, beginning with a 15-minute introductory session. Participants worked on a simple CAD task during this session, collaborating constantly and synchronously. Participants in the experimental condition were also required to consult ChatGPT to make every decision, the same as they would consult each other. After the introductory session, the first, second, and third collaborative CAD sessions were conducted, corresponding to the tasks mentioned earlier. Teams were allocated 40 minutes, 25 minutes, and 20 minutes for Task 1, Task 2, and Task 3, respectively. Before starting and upon finishing each of the four sessions, the researcher provided each team with details about the next steps. In addition to audio and video recordings, the researcher collected the final CAD outputs after each task session for data analysis.

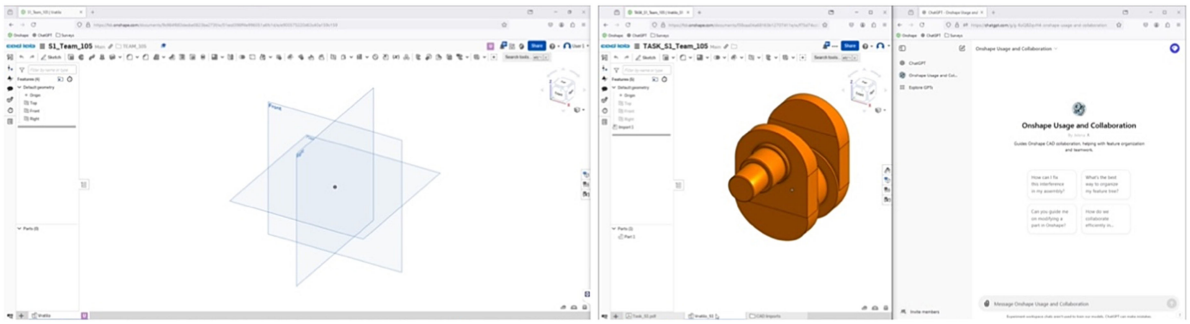


Figure 2. Left and right monitor screen set-up (Task 1 - experimental condition)

3.3. Data analysis

The collaborative CAD task completeness was measured via a CAD model completeness, one of the CAD model's quality dimensions (Company et al., 2015). According to that, a task is considered complete if it includes all the geometry aspects relevant to the design it represents, specifically regarding its sub-dimensions of shape and size. Therefore, the dependent variable in this study is collaborative CAD task completeness, measured by two specific sub-variables: shape replication and size replication (Company et al., 2015). Shape replication refers to the accuracy of the geometric features constituting the resulting design that the CAD model represents (e.g., a cylinder, a slot, a fillet, a constituting part of an assembly . . .). Size refers to the accuracy of geometric dimensions, which quantify the physical aspects of the design represented by a CAD model (e.g., length, width, height, radius, and angles). Even though the completeness variable is usually defined as a dichotomous variable, indicating that it is fully accomplished, herein we used intermediate scoring values for each sub-variable, namely both shape and size replication. This approach allowed us to capture partial and nuanced levels of task completeness. In Tasks 1 and 3, shape replication refers to the accuracy of the geometric features that make up the resulting CAD model. In contrast, size pertains to the accuracy of geometric dimensions that quantify the physical characteristics of the CAD model. In Task 1, the analysis focuses on evaluating whether all the geometric features present in the provided CAD model are also included in the CAD model created by

https://doi.org/10.1017/pds.2025.10172 Published online by Cambridge University Press

the pairs participating in the experiment (shape), and whether they are dimensionally accurate (size). Task 3 has a similar focus but relates specifically to design for manufacturability principles. That includes modifying geometric features (such as defining parting lines, adding drafted surfaces and rounded edges) (shape) and ensuring that the dimensions for these features are appropriate (size) (Bralla, 1998). Task 2 is inherently different since its output is a functional assembly. Therefore, the evaluation of shape and size follows the definition of assembly (Lupinetti et al., 2018), which considers the parts and their relationships while considering different levels of information (Chen, 2012): topological structure and relationships between assembly components (shape), and geometric information (size). In our case, the first two aspects involve verifying whether the constituent parts are included in the assembly, whether the parts are correctly placed in relation to each other, and whether the appropriate mates are applied (mates that mimic the intended movements). The geometric information concerns the dimensions between parts, as four of them are not dimensionally suitable for the provided housing. Additionally, functionality dictates certain dimensional relationships, such as those required for a simple engine mechanism represented in the assembly of Task 2.

The following number of measurement elements for sub-variables of shape and size is established for each task: Task 1 has 18 elements for shape and 32 for size; Task 2 has 16 for shape and 10 for size, and Task 3 has 14 elements for both shape and size. The researcher evaluated each element in the resulting CAD models by determining if the element was correctly addressed in the task, receiving a score of 1 for correct and 0 for incorrect. Shape and size scores were calculated by summing the corresponding elements for each sub-variable. These scores were then normalized as percentages and averaged for each pair to provide an overall task completeness. Statistical analyses were conducted using Python.

4. Results

On average, teams without support of ChatGPT achieved higher completeness scores in Task 1 with a score of 100%, compared to 67.1% for teams with support of ChatGPT. The same is observed in Task 2, where teams without ChatGPT support averaged CAD task completeness of 83.1%, while those with ChatGPT support averaged 62.5%. In Task 3, however, teams with ChatGPT support had a higher average CAD task completeness of 45.1%. The difference between teams with and without ChatGPT support in this task is smaller than in Tasks 1 and 2; teams without ChatGPT support completed an average of 36.4% of Task 3. The results are visualised in Figure 3. Before statistically testing the means across tasks within each experimental condition and the differences in completeness of each task, the normality and homogeneity of variance were assessed. Since statistical tests, such as ANOVA or t-tests, assume that the data is normally distributed and the variances are equal, these assumptions were tested using the Shapiro-Wilk test for normality and Levene's test for homogeneity of variance.

Table 1. Comparison of completeness in collaborative CAD tasks with and without ChatGPT support

Task	Condition	Completeness Mean (%)	Std. deviation (%)	Normality (p-value)	Homogeneity (p-value)
Task 1	ChatGPT support	67.1	29.3	0.056	1.669×10^{-6}*
	Without ChatGPT support	100.0	0.0	1.000	
Task 2	ChatGPT support	62.5	24.4	0.0498*	0.405
	Without ChatGPT support	83.1	17.9	0.005	
Task 3	ChatGPT support	45.1	18.1	0.126	0.113
	Without ChatGPT support	36.4	30.8	0.158	

*Significant value ($p<0.05$)

The Shapiro-Wilk test indicated that the data were normally distributed for both experimental conditions in Task 1 and Task 3, but not in Task 2. Levene's test showed that the assumption of equal variances was not met in Task 1, while it was met in Task 2 and Task 3. The following sections present the statistical

https://doi.org/10.1017/pds.2025.10172 Published online by Cambridge University Press

differences in task completeness means across tasks and between experimental conditions, accounting violations of normality and homogeneity. The results are shown in Table 1.

4.1. Comparing completeness within each condition

To assess whether task completeness differs significantly across tasks within the same experimental condition, the Friedman test was conducted. This non-parametric test is suitable for analysing repeated measures data when the assumption of normality is violated. Additionally, pairwise comparisons were conducted using Wilcoxon signed-rank tests with Bonferroni correction to identify differences between tasks when a significance is found.

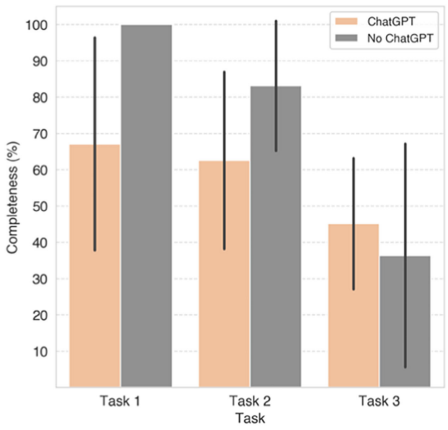


Figure 3. Task completeness across collaborative CAD tasks

The Friedman test revealed a statistically significant effect of task on task completeness in the condition without ChatGPT support (χ^2=17.610; p=0.002), with a large effect size (W=0.800). Pairwise comparisons using Wilcoxon signed-rank tests with Bonferroni correction, furthermore, indicated that completeness of Task 1 were significantly higher than Task 2 (W=0.0; p=0.020) and Task 3 (W=0.0; p=0.015). Additionally, completeness of Task 2 was significantly higher than Task 3 (W=1.0; p=0.006). In contrast, in the condition with ChatGPT support, the Friedman test did not reveal a significant difference (χ^2=3.818; p=0.148) in completeness across tasks. The corresponding effect size is small (W=0.174). The summary of results is shown in Table 2.

Table 2. Analysis of collaborative CAD task completeness across experimental conditions

Condition	Test	Task comparison	Statistic	p-value	Effect size
Without ChatGPT support	Friedman	-	χ^2=17.610	0.002*	W=0.800
	Wilcoxon signed-rank	Task 1 vs Task 2	W=0.0	0.020*	-
	Wilcoxon signed-rank	Task 1 vs Task 3	W=0.0	0.015*	-
	Wilcoxon signed-rank	Task 2 vs Task 3	W=1.0	0.006*	-
ChatGPT support	Friedman	-	χ^2=3.818	0.148	W=0.174

*Significant value (p<0.05)

4.2. Comparison of completeness within each task

To compare task completeness between conditions with ChatGPT and without ChatGPT support for each task, different statistical tests were used based on the characteristics of the data. For Task 1, Welch's t-test was used due to violations of the equal variances. For Task 2, the Mann-Whitney U test was applied due to violations of the normality assumption, as it is a non-parametric alternative suitable for comparing medians between two independent groups. For Task 3, the standard t-test was appropriate as both normality and homogeneity of variance assumptions were met.

https://doi.org/10.1017/pds.2025.10172 Published online by Cambridge University Press

For Task 1, Welch's t-test indicated a statistically significant difference between conditions (t=-3.728, p=0.004), with a large effect size (d=1.589). In Task 2, the Mann-Whitney U test also indicated a statistically significant difference (U=21.5, p=0.011), with a large effect size (r=0.645). In Task 3, however, the standard t-test did not reveal a statistically significant difference (t=0.815, p=0.425) of completeness scores between pair with ChatGPT and without ChatGPT support. The effect size for Task 3 is small to medium (d=0.347). The results are summarised in Table 3.

Table 3. Analysis of difference in task completeness across collaborative CAD tasks

Task	Test	Statistic	p-value	Effect size
Task 1	Welch's t-test	t=-3.728	0.004*	d=1.589
Task 2	Mann-Whitney U test	U=21.5	0.011*	r=0.645
Task 3	Standard t-test	t=0.815	0.425	d=0.347

*Significant value (p<0.05)

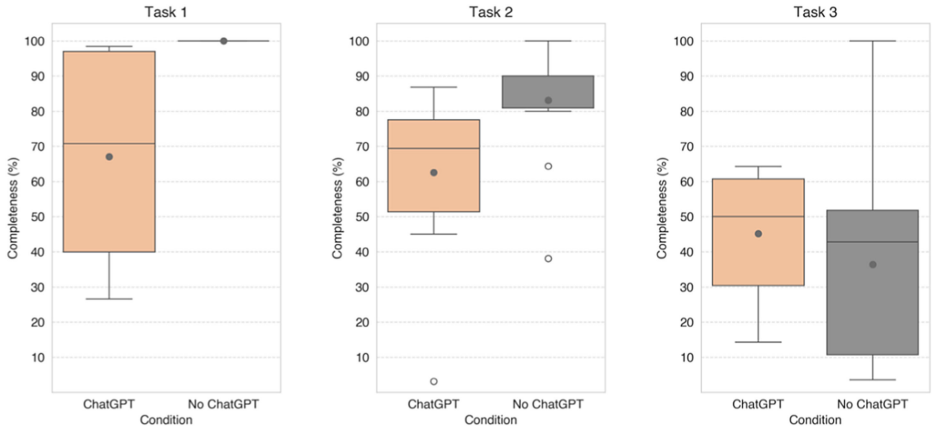


Figure 4. Comparison of completeness across tasks

5. Discussion

The study reveals insights into the role of ChatGPT support on the completeness of collaborative CAD tasks. Overall, the findings suggest that ChatGPT support, with its broad general knowledge, is not suitable for CAD-specific tasks requiring minimal general engineering knowledge. Task-specific LLMs, which focus on more CAD-specific inputs, may be more suitable. That is similar to findings in the programming domain, where the users with support of task-specific LLMs that process lines of code produce higher quality task output than ChatGPT (Mnguni et al., 2024). Additionally, the lower task completeness observed, despite the ChatGPT support, implies that participants may lack proficiency in prompt engineering (Memmert et al., 2024) since the research has shown that effective prompt engineering enhances the quality of task outputs. Furthermore, user feedback and the need for additional prompting when ChatGPT provides vague responses or factually inaccurate information (Ege et al., 2024) can be time-consuming (Coello et al., 2024), potentially contributing to lower task completeness. Additionally, ChatGPT's approach to solving tasks may differ from that of the participants, necessitating additional prompting and spending time that could have been dedicated to the tasks. That underscores the importance of human intervention in refining ChatGPT responses (Coello et al., 2024) through additional prompting, implying the complementary relationship between human intelligence and AI (Wang et al., 2023). Humans have to remain the decision-makers in CAD tasks. Therefore, ChatGPT should be seen as a tool to augment, not replace, the engineer's knowledge.

The study also revealed a significant effect of task type on the participants' performance who completed tasks without ChatGPT support. The effect size was large, indicating practical significance. Task completeness declines as the tasks become more open-ended and require greater domain-specific engineering knowledge. In contrast, pairs with ChatGPT support did not show significant differences in task completeness across the tasks, suggesting that ChatGPT support mitigates task-specific challenges,

https://doi.org/10.1017/pds.2025.10172 Published online by Cambridge University Press

such as the knowledge required to solve a task; however, it did not substantially improve overall task completeness. All of that indicates a complementarity between humans and AI, highlighting the importance of engineering knowledge in interactions with ChatGPT (Giordano et al., 2024). Additionally, collaborative CAD tasks require coordination and communication among team members, which can further slow the process when handling ChatGPT's suggestions (Phadnis et al., 2021).

The presented insights have implications for both industry and education. This study highlights the complementary relationship between human intelligence and AI, emphasizing that knowledge of CAD and engineering remains irreplaceable, even with the support of LLMs like ChatGPT. Experienced practitioners have to maintain and continuously improve their engineering expertise to effectively validate ChatGPT's outputs while addressing complex, open-ended CAD or design challenges. Education should focus on developing strong CAD skills and domain-specific engineering knowledge, ensuring novice engineers can critically evaluate suggestions from ChatGPT. Additionally, effective communication with tools like ChatGPT through prompt engineering techniques is becoming an increasingly valuable skill for practitioners, both novices and experts.

6. Conclusion

This study explored the role of ChatGPT on the completeness of collaborative CAD tasks solved by pairs of CAD users requiring varying types of engineering knowledge - CAD and domain-specific engineering knowledge. The findings highlight the context- and task-dependency of ChatGPT support in collaborative CAD. While ChatGPT support hinders task completeness in tasks requiring more CAD-specific knowledge (Tasks 1 and 2), it showed limited potential in assisting open-ended tasks (Task 3) that demand more domain-specific engineering knowledge. However, this study has several limitations. The evaluation was limited to a specific set of tasks, which may not fully represent the diversity of collaborative CAD tasks. Also, the quality and consistency of prompts used by participants were not standardised, which could influence the ChatGPT's support and pairs' interaction with ChatGPT. Lastly, the study focused only on novice users and did not explore how individual differences, such as prior experience with CAD or ChatGPT, might have affected task completeness. Therefore, future research work should explore the influence of ChatGPT in the similar setups that include professional engineers and CAD users. It should also examine the potential reasons behind the lower task completeness observed in pairs supported by ChatGPT. This includes pairs' interaction with ChatGPT, impact of used prompt engineering techniques, prior experience with both CAD and ChatGPT, and the dynamics of collaboration and communication.

Acknowledgement

This work was supported by/supported in part by the Croatian Science Foundation under the project DATA-MATION number [IP-2022-10-7775].

References

Azemi, F., Mehmeti, X. & Maloku, B. (2018, October 27). The Importance of CAD/CAE systems in development of Product Design and Process of Optimization. https://doi.org/10.33107/ubt-ic.2018.344

Bouschery, S. G., Blazevic, V. & Piller, F. T. (2023). Augmenting human innovation teams with artificial intelligence: Exploring transformer-based language models. *Journal of Product Innovation Management*, 40(2), 139–153. https://doi.org/10.1111/jpim.12656

Bralla, J. G. (1998). *Design for Manufacturability Handbook* (2nd ed.). McGraw Hill.

Cheng, K., Davis, M. K., Zhang, X., Zhou, S. & Olechowski, A. (2023). In the Age of Collaboration, the Computer-Aided Design Ecosystem is Behind: An Interview Study of Distributed CAD Practice. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1), 1–29. https://doi.org/10.1145/3579613

Chester, I. (2007). Teaching for CAD expertise. *International Journal of Technology and Design Education*, 17(1), 23–35. https://doi.org/10.1007/s10798-006-9015-z

Coello, C. E. A., Alimam, M. N. & Kouatly, R. (2024). Effectiveness of ChatGPT in Coding: A Comparative Analysis of Popular Large Language Models. *Digital*, 4(1), 114–125. https://doi.org/10.3390/digital4010005

Company, P., Contero, M., Otey, J. & Plumed, R. (2015). Approach for developing coordinated rubrics to convey quality criteria in MCAD training. *CAD Computer Aided Design*, 63, 101–117. https://doi.org/10.1016/j.cad.2014.10.001

Deng, Y., Mueller, M., Rogers, C. & Olechowski, A. (2022). The multi-user computer-aided design collaborative learning framework. *Advanced Engineering Informatics*, 51. https://doi.org/10.1016/j.aei.2021.101446

Giordano, V., Spada, I., Chiarello, F. & Fantoni, G. (2024). The impact of ChatGPT on human skills: A quantitative study on twitter data. *Technological Forecasting and Social Change*, 203. https://doi.org/10.1016/j.techfore.2024.123389

https://doi.org/10.1017/pds.2025.10172 Published online by Cambridge University Press

Hu, X., Tian, Y., Nagato, K., Nakao, M. & Liu, A. (2023). Opportunities and challenges of ChatGPT for design knowledge management. *Procedia CIRP*, 119, 21–28. https://doi.org/10.1016/j.procir.2023.05.001
Hu, Z., Li, X., Pan, X., Wen, S. & Bao, J. (2023). A question answering system for assembly process of wind turbines based on multi-modal knowledge graph and large language model. *Journal of Engineering Design*. https://doi.org/10.1080/09544828.2023.2272555
Khanolkar, P. M., Vrolijk, A. & Olechowski, A. (2023). Mapping artificial intelligence-based methods to engineering design stages: a focused literature review. *In Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM* (Vol. 37). Cambridge University Press. https://doi.org/10.1017/S0890060423000203
Koivisto, M. & Grassini, S. (2023). Best humans still outperform artificial intelligence in a creative divergent thinking task. *Scientific Reports*, 13(1). https://doi.org/10.1038/s41598-023-40858-3
Korteling, J. E. (Hans), van de Boer-Visschedijk, G. C., Blankendaal, R. A. M., Boonekamp, R. C. & Eikelboom, A. R. (2021). Human- versus Artificial Intelligence. *Frontiers in Artificial Intelligence*, 4. https://doi.org/10.3389/frai.2021.622364
Lupinetti, K., Giannini, F., Monti, M. & Pernot, J. P. (2018). Multi-criteria retrieval of CAD assembly models. *Journal of Computational Design and Engineering*, 5(1), 41–53. https://doi.org/10.1016/j.jcde.2017.11.003
Makatura, L., Foshey, M., Wang, B., HähnLein, F., Ma, P., Deng, B., Tjandrasuwita, M., Spielberg, A., Owens, C. E., Chen, P. Y., Zhao, A., Zhu, A., Norton, W. J., Gu, E., Jacob, J., Li, Y., Schulz, A. & Matusik, W. (2023). How Can Large Language Models Help Humans in Design and Manufacturing? http://arxiv.org/abs/2307.14377
Makatura, L., Foshey, M., Wang, B., Hähnlein, F., Ma, P., Deng, B., Tjandrasuwita, M., Spielberg, A., Owens, C. E., Chen, P. Y., Zhao, A., Zhu, A., Norton, W. J., Gu, E., Jacob, J., Li, Y., Schulz, A. & Matusik, W. (2024). Large Language Models for Design and Manufacturing. *An MIT Exploration of Generative AI*. https://doi.org/10.21428/e4baedd9.745b62fa
Mandorli, F. & Otto, H. E. (2013). Negative knowledge and a novel approach to support MCAD education. *Computer-Aided Design and Applications*, 10(6), 1007–1020. https://doi.org/10.3722/cadaps.2013.1007-102
McComb, C., Boatwright, P. & Cagan, J. (2023). Focus and Modality: Defining a Roadmap to Future AI-Human Teaming in Design. *Proceedings of the Design Society*, 3, 1905–1913. https://doi.org/10.1017/pds.2023.191
Mcdermott, J. (1982). *Domain knowledge and the design process*.
Meltzer, P., Lambourne, J. G. & Grandi, D. (2024). What's in a Name? Evaluating Assembly-Part Semantic Knowledge in Language Models Through User-Provided Names in Computer Aided Design Files. *Journal of Computing and Information Science in Engineering*, 24(1). https://doi.org/10.1115/1.4062454
Memmert, L., Cvetkovic, I. & Bittner, E. (2024). The More Is Not the Merrier: Effects of Prompt Engineering on the Quality of Ideas Generated By GPT-3. *Proceedings of the 57th Hawaii International Conference on System Sciences*. https://www.researchgate.net/publication/374942989
Memmert, L., Mies, J. & Bittner, E. (2024). Brainstorming with a Generative Language Model: The Role of Creative Ability and Tool-Support for Brainstorming Performance. https://aisel.aisnet.org/icis2024
Memmert, L. & Tavanapour, N. (2023). Towards Human-AI-Collaboration in Brainstorming: Empirical Insights into the Perception of Working with a Generative AI. https://www.researchgate.net/publication/370801376
Mnguni, N. M., Nkomo, N., Maguraushe, K. & Mutanga, M. B. (2024). An Experimental Study of The Efficacy of Prompting Strategies In Guiding ChatGPT for A Computer Programming Task. *Journal of Information Systems and Informatics*, 6(3), 1346–1359. https://doi.org/10.51519/journalisi.v6i3.783
Nygård Ege, D., Øvrebø, H. H., Stubberud, V., Francis Berg, M., Elverum, C., Steinert, M. & Vestad Affiliations, H. (2024). ChatGPT as an inventor: Eliciting the strengths and weaknesses of current large language models against humans in engineering design.
Phadnis, V., Arshad, H., Wallace, D. & Olechowski, A. (2021). Are two heads better than one for computer-aided design? *Journal of Mechanical Design*, 143(7). https://doi.org/10.1115/1.4050734
Ritala, P., Ruokonen, M. & Ramaul, L. (2024). Transforming boundaries: how does ChatGPT change knowledge work? *Journal of Business Strategy*, 45(3), 214–220. https://doi.org/10.1108/JBS-05-2023-0094
Sadeghi, S., Dargon, T., Rivest, L. & Pernot, J.-P. (2016). Capturing and analysing how designers use CAD software. *Tools and Methods for Competitive Engineering (TMCE'16)*, 447–458.
Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C. & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. *Computers and Education*, 215. https://doi.org/10.1016/j.compedu.2024.105031
Wang, K. D., Burkholder, E., Wieman, C., Salehi, S. & Haber, N. (2023). Examining the potential and pitfalls of ChatGPT in science and engineering problem-solving. *Frontiers in Education*, 8. https://doi.org/10.3389/feduc.2023.1330486
Wang, X., Anwer, N., Dai, Y. & Liu, A. (2023). ChatGPT for design, manufacturing, and education. *Procedia CIRP*, 119, 7–14. https://doi.org/10.1016/j.procir.2023.04.001
Xu, W., Kotecha, M. C. & McAdams, D. A. (2024). How good is ChatGPT? An exploratory study on ChatGPT's performance in engineering design tasks and subjective decision-making. *Proceedings of the Design Society*, 4, 2307–2316. https://doi.org/10.1017/pds.2024.23

https://doi.org/10.1017/pds.2025.10172 Published online by Cambridge University Press