Design Science

Artificial intelligence support in synchronous collaborative CAD activities: task-dependent effects and methodological implications

Jelena Šklebar[1], Tomislav Martinec[1], Stanko Škec[1] and Mario Štorga[1,2]

[1]*University of Zagreb Faculty of Mechanical Engineering and Naval Architecture, Zagreb, Croatia*
[2]*Luleå University of Technology, Luleå, Sweden*

Abstract

This study explores the task-dependent effects of ChatGPT support in synchronous collaborative Computer-Aided Design (CAD) activities. While large language models (LLMs) such as ChatGPT are increasingly being integrated into engineering design, evidence on their influence in collaborative CAD activities remains limited. Addressing this gap, a between-subjects experiment was conducted with 22 pairs of engineering students performing three synchronous collaborative CAD tasks requiring varying combinations of CAD and domain-specific knowledge. To assess the task-dependent effects, the study evaluated taskwork through the number of CAD user actions, productivity, efficiency and CAD dominance. Results show that ChatGPT's influence was highly task-dependent. In tasks relying primarily on CAD knowledge, ChatGPT-supported pairs exhibited lower productivity and efficiency. Conversely, in tasks requiring domain-specific engineering knowledge, ChatGPT support showed potential to improve efficiency. The study also offers methodological insights for designing controlled experiments involving ChatGPT support in collaborative CAD, underscoring that effective integration depends on task, tool and user-related factors. The findings further imply that integrating ChatGPT into engineering design education and practice requires developing competencies for critical and coordinated ChatGPT use within collaborative workflows, ensuring that ChatGPT complements rather than replaces designers. Future research should examine professional settings, teamwork and long-term adaptation to ChatGPT-supported synchronous collaborative CAD activities.

Keywords: Engineering design, Collaborative CAD, Artificial intelligence, Large language models

Received 22 October 2025 UTC
Revised 23 July 2026 UTC
Accepted 28 July 2026 UTC

Corresponding author
Jelena Šklebar
jelena.sklebar@fsb.unizg.hr

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Des. Sci., vol. 12, e40
journals.cambridge.org/dsj
DOI: 10.1017/dsj.2026.10073

1. Introduction

With the advancements of artificial intelligence (AI) and, specifically, generative AI, its utilisation is already fundamentally transforming various domains, such as business, education or healthcare (Fui-Hoon Nah *et al.* 2023). AI is conceptualised as "a system's ability to interpret external data, to learn from such data, and to use those learning to achieve specific goals and tasks through flexible adaptation" (Haenlein & Kaplan 2019). It can function in two ways (McComb, Boatwright, &

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

Cagan 2023): it can be proactive, anticipating needs and taking initiative, acting as a collaborator or as a member of a team, or it can be reactive, responding to user requests or inputs and functioning as a support tool.

Generative AI serves as a highly promising example of AI support tools that can generate human-like content, such as text, images or music. Within this category, large language models (LLMs) have gained significant attention with the launch of OpenAI's ChatGPT in 2022. Based on the generative pre-trained transformer (GPT) model, ChatGPT has been trained on vast datasets representing general knowledge. While generative AI encompasses a wide range of content generation tasks (Chiarello *et al.* 2024), ChatGPT is specifically designed to perform language-related tasks in a conversational manner, positioning ChatGPT as a broadly applicable support tool across many domains.

In the field of engineering design, LLMs such as ChatGPT have already demonstrated potential for integration in various stages (Khanolkar et al. 2023), supporting the execution of design tasks (Wang *et al.* 2024). The introduction of LLMs' support in engineering design may transform the way the associated engineering activities and tasks are approached and executed (Khanolkar *et al.* 2023). Among many, this transformation may also be present in computationally supported activities and tasks, such as those where computer-aided design (CAD) is utilised.

CAD is defined as using computers to assist in the creation, modification, simulation, analysis and optimisation of the design artefact, typically a digital three-dimensional information object or a CAD model, which captures both the structure and physical attributes of the designed artefact. CAD is utilised to create and test virtual product concepts or prototypes prior to manufacturing (Azemi, Mehmeti, & Maloku 2018).

In parallel with broader AI developments, including LLMs, CAD research has increasingly explored capabilities to integrate AI techniques into CAD tools and workflows (Zou *et al.* 2024). Recent work has examined how LLMs can take part in automating CAD tasks and activities by, for example, generating CAD models of standard mechanical engineering elements from text input, while also highlighting that such outputs in CAD require verification to meet precise geometric and engineering constraints. This requirement emphasises that human insight remains important since CAD includes the decision-making associated with the creation, modification and analysis of design representations, where both CAD and LLMs, despite the latter's broad knowledge base, remain limited. As a result, this positions LLMs as a suitable support tool within CAD activities, with the designer interpreting, validating and acting accordingly rather than just delegating the CAD task to the LLM tool completely. According to that, while recent research explores automation-oriented usage of LLMs in CAD activities, comparatively fewer studies have focused on how individual designers use LLMs as support in CAD activities and how this affects both the performance and the output in CAD (Zhang *et al.* 2026).

This is increasingly important because CAD is no longer limited to single-user, standalone CAD but is transformed into and is increasingly used in cloud-based collaborative versions of CAD, known as collaborative CAD (Cheng *et al.* 2023). Although CAD was initially developed to support individual work, CAD activities have long been inherently collaborative. Improvements to collaboration emerged when standalone CAD was integrated with data management systems, enabling

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

geographically distributed designers to collaborate asynchronously. More recently, the development of collaborative CAD has extended to support multiple designers to collaborate, enabling designers to view, edit and manipulate CAD models synchronously, with updates in real time to all designers, regardless of their location. This shift addresses common collaboration challenges in both standalone and distributed CAD, such as inability to edit synchronously, file sharing and limited visibility of design changes (Cheng *et al.* 2023). As a result, it has become possible to explore collaborative CAD activities, in which two or more collaborative CAD users actively and reciprocally engage in achieving a shared goal (Bedwell *et al.* 2012). On the activity level, it means that collaborative CAD users form a team working jointly to achieve a shared goal, integrating both social (team) and technical (engineering design and CAD) aspects. In this context, team performance, referring to the sum of teamwork and taskwork, presents an important consideration (Passmore 2017). Teamwork describes the interaction between team members that makes team functioning and the achievement of their goals possible, while taskwork refers to the performance of the task execution necessary to achieve the team's shared objectives and goals (Salas, Sims, & Shawn Burke 2005). In addition to tasks, it also involves team members' interactions with tools, machines or systems to carry out the team's work. Although both are important for the success of collaborative CAD activities, understanding taskwork is especially relevant when investigating how tools like ChatGPT might influence these activities.

Taskwork is known to change when CAD work shifts from single users to teams. Evidence from synchronous collaborative CAD activities suggests that pairs are less efficient than individuals working alone, due to the real-time communication and coordination overheads, and that their CAD output is also of lower quality than individuals working alone (Phadnis *et al.* 2021). Furthermore, different working collaborative modes also shape the taskwork within synchronous collaborative CAD activities. Working in parallel mode (all team members have control over the same CAD model they work on) led to a lower quality of resulting CAD models than working in shared mode (only one team member controls the CAD model, while the other/s still see the changes in real time and can take control of that CAD model anytime) (Arshad & Olechowski 2020). Taken together, both the shift from individual to synchronous collaborative CAD and the specific working collaboration mode influence taskwork performance. However, it remains unclear how the introduction of LLM support such as ChatGPT influences such activities. Introducing such support may introduce an additional interaction stream that teams need to coordinate around, including deciding when to consult ChatGPT, how to formulate and refine the input prompt and how to interpret, validate and integrate the suggestions. Such integration may introduce additional effort that can increase the time to solve a task (Feng *et al.* 2024). At the same time, ChatGPT and other LLM-based support also represent an additional knowledge source that may reduce teams' reliance on only teams' own expertise and support higher quality outputs. Accordingly, evidence from teams supported by ChatGPT suggests that performance benefits are task-dependent (Vaccaro, Almaatouq, & Malone 2024), being more advantageous for prediction and content-generation tasks while potentially disadvantageous on tasks requiring decision-making and human judgement. This is relevant for CAD taskwork, which combines the creation and manipulation of design representations in CAD with decision-making under geometric and engineering constraints. Thus, it would be naive to

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

claim that ChatGPT support influences all tasks in synchronous collaborative CAD activities in the same way, because its effect depends on the specific task type and the knowledge demands. Despite these indications, empirical evidence on ChatGPT support in synchronous collaborative CAD activities remains limited, and one study in this context indicates task-dependent effects when assessed using the quality taskwork variable, namely, task completeness (Šklebar *et al.* 2025b). This leaves open how ChatGPT support might influence other taskwork variables across different synchronous collaborative CAD task types.

To address this gap, this study proposes an exploratory evaluation of taskwork variables in synchronous collaborative CAD activities supported by ChatGPT. Given the lack of existing research in this context, this study not only explores the potential of ChatGPT on taskwork but also aims to contribute to the broader understanding of how to design and conduct experiments in the context of synchronous collaborative CAD activities. In other words, it reflects on lessons learned in designing such experiments. These insights are intended to inform future research and guide the development of more robust experimental designs (Robson & McCartan 2016). On a larger scale, this study aims to be one of the first steps or a proxy towards the theoretical description or a model of the influence of LLMs on taskwork in any collaborative design activities. More specifically, it aims to clarify when and why LLMs enhance or hinder the performance of such activities, thereby contributing to the broader understanding of collaborative design and even design cognition and human–AI interaction. Finally, it also aims to open questions about, for example, the competencies required for AI-supported engineering design, the potential redefinition of engineering design processes and the implications for future engineering education.

The paper is structured as follows. In Section 2, an overview of existing related research work is presented. Section 3 describes the research methodology. Furthermore, Section 4 presents the results, followed by discussion and conclusion in Sections 5 and 6, respectively.

2. Literature review

This section provides background work relevant to the study. Section 2.1 provides an overview of the LLMs in engineering design. Section 2.2 presents the usage and support of LLMs in CAD activities and tasks. Next, Section 2.3 describes collaborative CAD and approaches to studying collaborative CAD activities. Section 2.4 presents the factors influencing the effective usage of LLMs.

2.1. Large language models in engineering design

AI has already seen adoption across various stages of engineering design (Khanolkar *et al.* 2023). This aligns with the paradigm of human–AI collaboration because many design problems are knowledge-intensive and often exceed the capabilities of an individual designer or a specialised AI tool alone. Although there are many arguments about AI being a new member or being solely an automation technology, design research often emphasises its socio-technical nature (Song, Zhu, & Luo 2024). In other words, the value it brings to engineering design depends on how it is integrated into work practices. In addition to that, evidence from engineering design research is contrary. Some studies report that AI has been

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

associated with improvements in designer creativity and team performance (Song *et al.* 2022a) or communication in teams (Zhang *et al.* 2023), while others suggest that usage of AI may hinder high-performing teams (Zhang *et al.* 2021a) or negatively influence learning (Martin & Selva 2020). Complementing these, McComb *et al.* (2023) have proposed a roadmap to different ways in which AI tools support, work with and guide engineering design teams. Distinguishing these by focus (problem versus process support) and modality (reactive versus proactive), they highlight that different variations correspond to different AI roles in teams and provide distinct implications for when and how AI should be employed in engineering design. This is particularly relevant for LLMs such as ChatGPT, whose conversational interface, broad knowledge and generative context-aware capabilities make them suitable for supporting design work, while their limitations in analytical and spatial reasoning, self-verification mechanisms and inability to decompose tasks mean that human oversight and validation are still required (Makatura *et al.* 2024).

This is also reflected in the context of engineering design research, where LLMs such as ChatGPT demonstrate both benefits and shortcomings in the early design stages. In requirement engineering, ChatGPT has demonstrated potential to accelerate requirement elicitation (Ataei *et al.* 2025). However, most existing work evaluates the performance of LLMs, rather than analysing which requirement engineering tasks align with current LLMs' capabilities (Norheim *et al.* 2024). In concept generation and related creative tasks, ChatGPT has been found to produce novel, though not always useful, concepts when compared to conventional methods (Filippi 2023). Also, while engineering designers generated the highest-quality concepts, those produced with ChatGPT support were of higher average quality (Urban *et al.* 2024). Further insights into ChatGPT's role in conceptual design tasks reveal that it supports problem definition and idea generation, while idea selection and evaluation still depend on human judgement (Chen *et al.* 2025). Others also showed that while ChatGPT solutions were better in terms of stylistic quality, human designers outperformed in novelty, and human–AI collaboration solutions ranked lowest, suggesting that ChatGPT can support certain creative aspects but still relies on human contextual judgement and conceptual reasoning (Vo 2025). In brainstorming and design space exploration activities, ChatGPT has been proved to be beneficial for both low- and high-creative designers, with the former relying more on ChatGPT's suggestions (Memmert, Mies, & Bittner 2024b), while others reported that participants prioritised suggestions from other engineering designers over suggestions from ChatGPT (Han *et al.* 2024). They also reported that ChatGPT can introduce communication challenges between engineering designers (Han *et al.* 2024). Consequently, multiple studies also found that the effectiveness of ChatGPT support in such tasks is highly sensitive to the designers' prompt engineering skills (Memmert, Cvetkovic, & Bittner 2024a) and therefore requires human oversight and structured interaction practices, such as tailored prompts, templates and the implementation of iterative feedback loops (Ege *et al.* 2025) to avoid premature convergence and to ensure promising solutions are not discarded (Nygård Ege *et al.* 2024).

In the later stages of engineering design, LLMs have been explored in material selection and topology optimisation. While LLMs can suggest material families meeting general specifications and support conceptual understanding, they struggle with identifying materials matching specific numeric requirements, accurate

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

identification of load and boundary condition positions, and optimisation code generation (Picard *et al.* 2025). Given the central role of CAD in embodiment and detailed design (Pahl & Beitz 1996), LLM support becomes coupled with CAD work and associated activities. In line with the socio-technical perspective and the paradigm of collaboration of AI and designers, the relevant question is not only what LLMs can create but how their use is integrated in CAD activities and how such support influences task execution and performance.

2.2. LLMs in Computer-Aided Design

CAD has long been an essential tool used throughout various design phases. Regarding the integration of LLMs into CAD, recent studies show that LLMs can support a variety of CAD-related activities and tasks. A systematic literature review by Zhang *et al.* (2026) categorises key areas where LLMs have made considerable contributions, including data generation, CAD code and parametric model generation, image and text generation, and CAD model evaluation. Among these, CAD code generation and text generation have been the most explored, with studies showing that ChatGPT and other fine-tuned multimodal LLMs can generate CADQuery code from various types of input (Sun, Li, & Sha 2025) and support generation of complex geometries such as shafts and gears, depending on input modality (Li, Sun, & Sha 2025). In parallel, LLMs have also been applied to generate text prompts for CAD model variation (Kodnongbua *et al.* 2023), produce structured parametric data, such as JSON-formatted sequences that define geometric primitives, constraints and relationships (Li *et al.* 2024a), generate synthetic CAD datasets (Yuan, Shi, & Huang 2024) and support generation or restoration of CAD drawings (Tang *et al.* 2025) and CAD model evaluation based on visual content or attributes like size (Du *et al.* 2024).

While the technical capabilities and applications of LLMs in CAD are expanding, comparatively fewer studies examine how LLM support influences the execution and performance of CAD activities. This limitation becomes even more important given that CAD activities are no longer restricted to standalone or distributed individual activities but are shifting towards cloud-based collaborative CAD that enables multiple CAD users to manipulate the same CAD model synchronously in real time (Cheng *et al.* 2023). That being said that task execution in synchronous collaborative CAD activities is intertwined with team interaction, therefore, performance cannot be conceptualised as an aggregation of the performance of individual collaborative CAD users or team members (Mathieu *et al.* 2000). According to that, taskwork in synchronous collaborative CAD activities is influenced and changed when CAD work shifts from an individual user to a team of users working synchronously, which is empirically indicated in differences in efficiency and output quality (Stone *et al.* 2017b). The mention has been attributed to real-time coordination and communication overheads, something that is not present in either individuals working in CAD or groups of people working at different times, or asynchronously.

Introducing support of LLMs such as ChatGPT into this setting therefore adds an additional interaction stream that teams have to manage alongside CAD actions and communication. Teams must coordinate around deciding when to use the tool, who formulates the input prompt and how, as well as how to validate and integrate suggestions into CAD actions (Han *et al.* 2024). This can also introduce

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

overheads. It may also provide an additional knowledge resource depending on the task knowledge demands.

Despite these indications, empirical evidence in understanding how the use of LLM tools influences synchronous collaborative CAD activities where two or more users work synchronously on the same CAD task remains limited. To date, only initial work has examined task-dependent effects using the task completeness variable as a proxy for assessment of taskwork, more specifically of resulting CAD model quality (Šklebar *et al.* 2025b). This leaves open how ChatGPT support influences other taskwork variables across different CAD tasks. The following section, therefore, reviews collaborative CAD as a research context and clarifies how taskwork has been operationalised and evaluated in previous research.

2.3. Collaborative CAD

Although the adoption of collaborative CAD is in its early stages (Cheng *et al.* 2023), the development and release of commercial collaborative CAD tools enabled studying and exploring collaborative CAD activities (Andreadis, Fourtounis, & Bouzakis 2015). Unlike approaches reliant on post-task interviews or subjective assessments, collaborative CAD offers non-intrusive, real-time data collection capabilities through built-in action logging systems. This supports conceptualising CAD as a research instrument for engineering design, as it enables simultaneous observation of evolving artefacts, participants and actions performed during the design process in CAD, supporting the perspective of using CAD and collaborative CAD not only for post-hoc outcome assessment but also for studying CAD work execution as a process (Vrolijk & Olechowski 2025). At the operationalised level, CAD captures user actions (e.g., creating, editing, viewing, deleting) and their sequences via action logs, enabling researchers to study the workflow in situ (Gopsill *et al.* 2016). Such data can then be used to operationalise and analyse taskwork through various variables and to generalise about taskwork in CAD activities, with implications for understanding engineering design performance in the corresponding design activities and stages (Vrolijk & Olechowski 2025).

In CAD in general, taskwork pertains to the execution of tasks such as constructing parts, assembling components and making modifications. Taskwork is commonly assessed through a workflow lens (e.g., type, frequency and sequence of CAD user actions) and an outcome lens (e.g., CAD model completeness, consistency and accuracy, or the outcome relative to design objectives). The workflow lens operationalises taskwork through the analysis of CAD user actions, such as type, frequency and sequences, which enables analysing how teams progress through work and how distinct actions relate to performance. To support such analyses, Deng *et al.* (2022b) introduced a categorisation of CAD user actions according to action type (e.g., constructive versus organisational) and design space (e.g., part versus assembly level). Applying this approach, Celjak, Horvat, & Škec (2023) show that high-performing teams were more likely to engage in deletion and assembly-level editing actions, whereas low-performing teams engaged more frequently in repetitive creation or reversal actions. The performed CAD user actions have been found to be influenced by users' experience level, with novice teams performing more reversing actions and fewer viewing or constructive actions compared to experienced teams (Deng *et al.* 2022b). In parallel, the outcome lens evaluates the resulting CAD model relative to design objectives.

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

For example, Stone *et al.* (2017a) found that teams using collaborative CAD outperformed individuals in CAD model completeness, while Phadnis *et al.* (2021) reported that individuals outperformed pairs in synchronous collaborative CAD activities, possibly due to unclear role division or ineffective communication. Prior research has also begun to examine how taskwork is distributed among team members in synchronous collaborative CAD activities. Again, through the analysis of CAD user actions, Olechowski *et al.* (2022) showed that taskwork is not necessarily distributed evenly across team members. Moreover, Roy *et al.* (2024) further suggest that such asymmetries in contribution to taskwork may reflect a division of work within synchronous collaborative CAD activities. A complementary perspective is offered by Shockley *et al.* (2025), who, for teams using CAD representations of design, assessed contribution equality through verbal participation. Together, these studies point to an uneven contribution to taskwork that should not be assumed to arise from CAD representation itself but may instead emerge from team composition, task structure and the way work is coordinated within synchronous collaborative CAD activities (Roy *et al.* 2024; Shockley *et al.* 2025).

Accordingly, literature has also focused on defining appropriate metrics for evaluating taskwork in collaborative CAD activities that can be derived from the CAD action log or CAD user actions. A systematic literature review identified four core taskwork evaluation metrics, whose operationalisation is influenced by the specific collaborative CAD tasks performed: efficiency, productivity, quality and effectiveness (Šklebar *et al.* 2025a). Efficiency refers to the extent to which the gains are achieved relative to the resources used, often measured by time taken per task or per user (Deng *et al.* 2022a), influenced by the mode of collaboration, be it shared or parallel (Phadnis *et al.* 2021), or by the number of collaborative CAD users involved in the activities (Stone *et al.* 2017b). Furthermore, productivity is defined as the quantity of work produced within a given time frame in collaborative CAD activities, measured, for instance, by the number of CAD changes implemented in parts tasks (Phadnis *et al.* 2019) or the number of assembly mates added per unit of collaborative CAD activity time during assembly tasks (Cheng & Olechowski 2024, 2021). Quality is assessed using CAD model quality criteria, including completeness, conciseness, consistency and validity (Camba, Contero, & Company 2016), and is also influenced by the collaboration mode (Arshad & Olechowski 2020; Phadnis *et al.* 2021) and number of users involved in the activity (Stone *et al.* 2017b). Furthermore, differences in the quality of created CAD models have been shown to reflect the distinct approaches in utilising CAD, revealed by the analysis of CAD user actions performed by collaborative CAD users (Celjak *et al.* 2023). Finally, effectiveness evaluates how well the design outcomes align with predefined objectives, independent of resources used. According to the currently available research, this evaluation is highly dependent on the specific collaborative CAD task (Singh *et al.* 2022).

Despite the growing empirical base, current research on taskwork in collaborative CAD activities remains fragmented (Šklebar *et al.* 2025a). Taken together, this body of work suggests that collaborative CAD offers a valuable context for studying taskwork, enabled particularly by collecting real-time objective data. However, the literature has not yet converged on comprehensive models explaining how support of LLMs like ChatGPT influences taskwork in synchronous collaborative CAD activities. Building on the initial evidence of task-dependent

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

effects on task completeness (Šklebar *et al.* 2025b), this study extends the evaluation focus from completeness analysis to broader taskwork measures. Addressing this gap requires controlled experimental studies that isolate the effects of LLMs' support while accounting for task-, tool- and user-related factors that influence usage and can confound taskwork effects.

2.4. LLMs' usage influencing factors

The conversational interfaces, their broad knowledge base, the ability of knowledge synthesis and generative capabilities make LLMs such as ChatGPT attractive for a variety of domains. However, their effective utilisation is not uniform. The effectiveness of LLM use is shaped by an interplay of factors related to the task being performed, the characteristics of the LLM tool itself and the user or team of users interacting with them.

Task-related factors influence how LLMs are utilised (Hemmer *et al.* 2021). Empirical findings indicate that LLM support tends to improve performance in creative tasks, where users are expected to create open-ended content, but may lead to performance losses in decision tasks, where users decide between a finite set of options (Vaccaro *et al.* 2024). One explanation is that in open-ended tasks, LLMs provide the ability to generate additional content that complements human creativity and insight. Beyond task type, the nature of the task input (image, text, video, numeric data) and the required task output also influence the effectiveness of LLMs' utilisation (Vaccaro *et al.* 2024), which is particularly relevant in engineering design, where tasks range from open-ended tasks such as ideation or requirements elicitation to highly constrained and well-defined tasks such as finite element analysis. For instance, LLMs have demonstrated considerable promise in open-ended tasks like brainstorming (Memmert *et al.* 2024a) but may be less reliable in tasks requiring precise geometric or parametric outputs, where engineering expertise is essential for providing contextual understanding (Wang *et al.* 2024).

Tool-related factors further influence the effective utilisation of LLMs. In the first place, prompt engineering is repeatedly emphasised as central for realising the benefits of LLMs, such as ChatGPT. LLMs allow users to express problems and objectives without formal syntax or domain-specific language (Bubeck *et al.* 2023), yet the utility of interaction depends on the clarity and specificity of user prompts, including alignment with the LLM model's knowledge and formulations that reduce inaccuracy and bias (Robertson *et al.* 2024). In design applications, prompt formulation, prompt structure and user intent were found to influence generated design outputs (Chong *et al.* 2025), implying that effective interaction focuses on instructing LLMs to generate appropriate and useful responses, rather than mimicking human conversation (Tholander & Jonsson 2023). Therefore, prompt engineering practices such as delegating tasks (Jing *et al.* 2024) and breaking down large tasks into smaller subtasks (Makatura *et al.* 2024) can improve performance and efficiency (Jing *et al.* 2024). Moreover, the distinction between general-purpose LLMs, such as ChatGPT, and task-specific, fine-tuned models is often highlighted as a determinant of human performance in LLM-supported activities (Yang 2024), as general-purpose LLMs offer broad applicability and flexibility but struggle with domain precision. Additionally, factors of competence (Kaplan *et al.* 2023), robustness (Li *et al.* 2024b), accuracy and reliability (Thielsch, Meeßen, &

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

Hertel 2018) of LLMs remain persistent concerns, underscoring the importance of verifying ChatGPT output before integration into task execution (Al-Mughairi & Bhaskar 2025). Transparency and explainability are also important influencing factors, moderating user trust and reliance on LLMs. Users need to understand the reasoning for the LLMs' output (Li *et al.* 2024b), underlying AI algorithms (Glikson & Woolley 2020), their reliability (Song *et al.* 2022b), performance (Zhang *et al.* 2021b) and contextual information (Barda, Horvat, & Hochheiser 2020). Anthropomorphism, or the tendency to attribute human traits to non-human agents, also affects user perception and the effectiveness of LLMs' utilisation (Bach *et al.* 2024). ChatGPT's conversational fluency and contextual understanding can elicit human-like attributes, with trust increasing when responses appear helpful but decreasing when outputs are unsatisfactory or repetitive (Rapp, Di Lodovico, & Di Caro 2025). Finally, the support for multiple modalities (e.g., visual or CAD-specific terminology) further influences how well LLMs can be integrated into activities.

The user- and team-related factors complete this picture, as utilisation depends on who interacts with the tool and under what collaborative conditions (Kirkeby-Hinrup & Stenseke 2025). At the individual level, the literature identifies a set of inherent characteristics (e.g. age; gender (Shilpa & Menon 2024); personality traits (Hemmer *et al.* 2021); acquired characteristics (AI literacy, prompt engineering and domain-specific expertise) (Jing *et al.* 2024); attitudes (user trust in tools such as ChatGPT (Ulfert & Georganta 2020), performance expectancy and effort expectancy (Grassini, Aasen, & Møgelvang 2024; Strzelecki 2024); and external variables such as time spent interacting with a tool (Lee, Frank, & IJsselsteijn 2021). High AI literacy (AI awareness and AI usage literacy) and domain-specific expertise are particularly associated with the effective utilisation of LLMs, as they support appropriate prompt formulation and critical evaluation of LLM accuracy and reliability before applying them in a task-specific context (Jing *et al.* 2024). Also, user trust towards tools such as ChatGPT may lead to automation bias or overreliance on ChatGPT responses presented in confident language, regardless of their accuracy, thus even reducing users' thinking ability and, in the long term, leading to dependence on the LLMs for even minor tasks (Al-Mughairi & Bhaskar 2025). Furthermore, the more time users spend interacting with LLMs like ChatGPT, the more likely they are to develop trust, possibly due to a deeper understanding of its strengths and limitations (Lee *et al.* 2021). In collaborative settings, team-level factors also influence the effective utilisation of LLMs. Drawing from the Input–Mediator–Outcome (IMO) framework (Ilgen *et al.* 2005), effective use of LLMs is influenced by team-based input variables, such as team composition or role distribution (Zercher & Jussupow 2023), with emergent states, such as team trust, team cohesion, shared mental models and psychological safety (Georganta & Ulfert 2024), being equally important. These, in turn, mediate how LLM support is enacted through coordination (Dell'Acqua, Kogut, & Perkowski 2025) and communication (Musick *et al.* 2021) during task execution.

While these are influencing factors in general LLM and human–AI collaboration research, their combined implications for studying the LLM support in synchronous collaborative CAD activities remain under-specified. This creates a methodological challenge because effects attributed to LLM support can be confounded by task and tool characteristics and user- and team-level differences. Accordingly, experimental studies of LLM support in synchronous collaborative

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

CAD activities require explicit consideration and control of these factors to support interpretable claims about taskwork effects.

3. Research methodology

To initially explore the task-dependent effects of ChatGPT support in synchronous collaborative CAD activities, this study is designed as a between-subjects randomised controlled experiment (Robson & McCartan 2016). The independent variable was defined as the presence or absence of ChatGPT support during the synchronous collaborative CAD activity. Participants engaged in three consecutive collaborative CAD tasks, performed in a multi-user synchronous setting, with each task requiring a distinct type of engineering knowledge. The dependent variables included the number of CAD user actions, productivity, efficiency and CAD dominance.

Participants were randomly assigned to experimental conditions, enabling comparability between groups and helping to mitigate selection and population bias (Braga *et al.* 2025). Although the randomised controlled design maximises internal validity, full control of tool- and user-based variables was not feasible. Additionally, due to the nature of the independent variable, participant blinding could not be implemented. Nevertheless, key variables such as the collaborative CAD activity tasks, time constraints, participants' prior CAD and domain-specific knowledge and experimental setup were held constant across conditions to support a consistent and controlled comparison.

3.1. Participants

In total, 44 participants took part in the experiment. Participants were students (11 females and 33 males) at the Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb, spanning undergraduate and graduate levels (the average age was 22.5 years). Participation required students to have successfully completed both basic (first year) and advanced (third year) CAD courses at the university with at least grades 4 and 5 (according to the Croatian 1–5 grading system, where 5 is the highest grade and 1 is fail), ensuring they possessed both declarative and procedural CAD knowledge (Chester 2007), along with knowledge of design for manufacturability (DfM) principles. Participants were assigned to 22 pairs, with each pair randomly allocated to one of the two conditions: with or without ChatGPT support.

3.2. Tasks

The experiment consisted of three consecutive collaborative CAD tasks (Task 1, Task 2 and Task 3) designed to require varying combinations of CAD (declarative and procedural) and domain-specific engineering knowledge (DfM). In this context, declarative CAD knowledge refers to knowledge about the options within the tool, such as features or commands, whereas procedural CAD knowledge concerns the ability to apply the commands and features to achieve a goal, for example, to create CAD models that are easy to modify or reuse (Bhavnani, John, & Flemming 1999). Domain-specific engineering knowledge pertains to engineering principles that ensure functional designs. In this study, such knowledge was

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

required both for understanding how mechanical components interact within an assembly (e.g., alignment, clearances, or mechanical motion) and for applying DfM principles (Ulrich & Eppinger 2015).

Task 1 focused on 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 1a).

Task 2 combined declarative CAD and domain-specific knowledge. Participants assembled five CAD models of parts (the housing, along with the crankshaft, upper and lower parts of the connecting rod and a piston) provided by researchers into a functional assembly. Four parts (excluding the housing) were not dimensionally adequate and had accessible feature trees and could be modified as needed. The housing was provided in STEP format, making it unmodifiable, requiring the other parts to be adjusted accordingly (Figure 1b). 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.

Lastly, Task 3 required predominantly domain-specific knowledge, specifically related to design-for-manufacturability in forging. Participants were tasked with redesigning the crankshaft from Task 1 to ensure it would be suitable for forging, accounting for pre-machining shape and dimensional allowances (Figure 1c). More specifically, they were expected to define an appropriate parting line that separates the upper and lower die halves, add draft angles to relevant surfaces to enable part ejection from the die, and introduce rounded edges (fillets) to avoid sharp transitions and support more uniform material flow (Bralla 1998). In addition, they were expected to modify the geometry to achieve smoother changes along the shaft (i.e., shaft section transitions) and add material to surfaces that are intended for subsequent machining in order to achieve functional geometry after both forging and machining (Bralla 1998). The CAD model included a visible feature tree, allowing for full modification of the CAD model.

Since the three tasks were performed consecutively, potential learning and the carryover effects needed to be considered. To mitigate learning effects, the tasks were designed to require different combinations of engineering knowledge. This

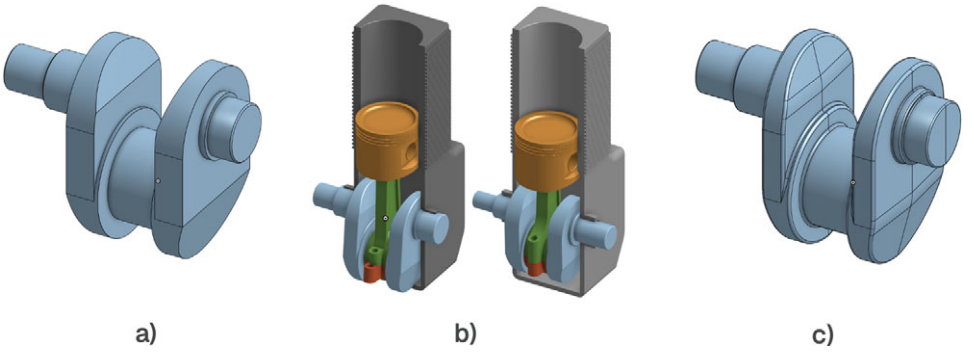


Figure 1. Task 1 (a), Task 2 (b) and Task 3 (c).

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

reduced the likelihood that knowledge gained in one task would simply transfer to the next. Regarding carryover effect, while Task 3 reused the crankshaft design from Task 1, the objective and required knowledge were different, moving from part replication to design-for-manufacturability for forging. Moreover, as all participants had comparable prior CAD and engineering knowledge gained through the same university courses, the risks of learning or carryover effects across participants were further minimised. To support this, all participants also received tutorials before the experiment to ensure a baseline of CAD and ChatGPT use. In addition, for participants with ChatGPT support, an introductory task was introduced to mitigate learning effects during the experimental tasks, while also limiting carryover effects, since the introductory task was independent of the experimental tasks and designed, among other reasons, to introduce interaction with ChatGPT.

Panels a–c show the three CAD tasks. Panel a is the original crankshaft used for measurement and replication in Task 1. Panel b shows two cutaway views of the Task 2 piston assembly, with an orange piston, green connecting rod, blue crankshaft and grey housing. Panel c is the Task 3 crankshaft redesign for forging, with added material, rounded edges and smoother transitions between the shaft and crank webs.

3.3. Experimental setup

The experimental setup consisted of one room with two working places facing each other, as shown in Figure 3. Each working place was equipped with an office chair and a table, a high-performance computer, two monitor screens (22″ with a resolution of 1920 × 1080 pixels), a keyboard and a mouse.

In both conditions, teams utilised Onshape on each monitor screen (Figure 2). Onshape is a cloud-based collaborative CAD software tool accessed via a web browser. Onshape allows synchronous collaboration on a single CAD model. Participants could simultaneously interact with the CAD model, follow each other's work and share views. Like other CAD tools, Onshape also provides orbital navigation features (panning, rotating and zooming) and basic viewing functionalities like taking screenshots, measuring, marking, sectioning, moving and hiding parts of a CAD model. The content displayed on the left screen varied by task. In Task 1, this document was blank; Task 2 included an empty assembly document and five CAD models provided as tabs within the working document (a common organisational approach in Onshape); and Task 3 contained a CAD model of the crankshaft to be modified. The right screen displayed task-specific content and varied by both experimental condition and task. All participants were shown a PDF with detailed task instructions via an Onshape document. For Task 1 and Task 3, the right screen was split: one half displayed the instructions, and the other half showed the reference crankshaft model in STEP format (used for replication in Task 1 and reference in Task 3).

For participants in the ChatGPT-supported condition, the right monitor additionally included a web browser window running ChatGPT with a customised GPT model, created for this experiment (Figure 2). The custom GPT titled "Onshape Usage and Collaboration" was configured to reflect the language and workflows of collaborative CAD in Onshape, including tailored prompt suggestions on the interface and an introductory message that helped participants

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

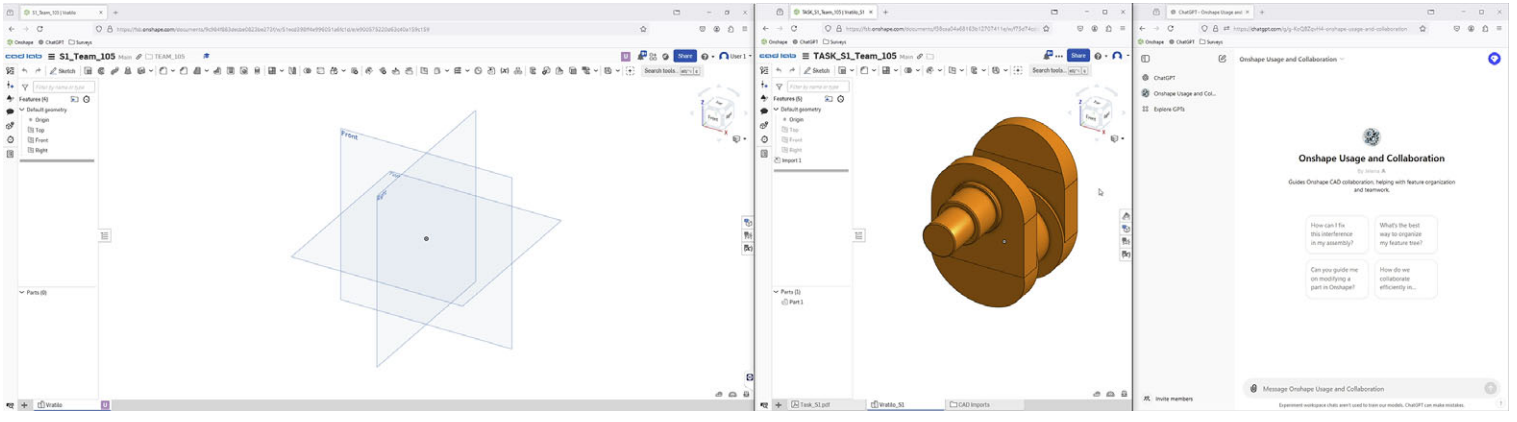


Figure 2. Left and right monitor screen setup for the experimental condition during Task 1.

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

formulate questions relevant to collaborative CAD usage, CAD modelling and collaboration in CAD. It was, therefore, implemented only at the user-interface level using the GPT builder functionality and did not involve any model-level adaptation (e.g., additional training data or tool integration). ChatGPT was accessed via the OpenAI ChatGPT web interface, using the GPT-4 model. The web interface does not expose a stable model snapshot identifier (build/version/date) or controllable sampling parameters (e.g., temperature) to end users, therefore, these settings could not be recorded or standardised in this study. Instead, the configuration was limited to the interface-level fields available during GPT creation, namely the GPT name and four predefined conversation starters ("Can you guide me on modifying a part in Onshape," "What is the best way to organise my feature tree?" "How can I fix the interference in the assembly?" and "How do we collaborate efficiency in Onshape during a shared modelling task?"), also shown on the landing page. The description communicated the intended scope of use (Onshape CAD feature organisation and collaboration in CAD), while the conversation starters provided participants with examples of questions to initiate interactions with GPT. Together with the prompt engineering tutorial, these interface elements served as standardisation to reduce variability in how participants interacted with ChatGPT across teams and sessions. Also, to maintain experimental integrity, ChatGPT's training memory was disabled, and its memory was cleared between experimental sessions. This ensured that no information persisted across experimental teams, thereby preventing answer accumulation or model learning that could bias results. To facilitate synchronous collaboration, the content of the right monitor of one team member was shared on the right monitor of the other team member by 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 ChatGPT.

As participants were co-located, they were required to communicate verbally during task execution. However, to keep the physical interaction conditions consistent across teams and to keep a parallel co-located workflow, participants were instructed to remain seated at their own workstation and were not allowed to stand up, approach their team member's workstation or use any gesticulations.

All screen content was recorded using OBS Studio, capturing both monitors per participant. Additionally, video recordings of the experimental sessions were collected using a conference camera positioned in the room, and audio was captured via small video cameras mounted on each monitor.

To control for artefact-related confounding variables, identical designs represented by CAD models were used across all tasks and conditions, standardising attributes such as dimensionality and complexity. Additionally, software functionality remained constant across both conditions, further eliminating potential tool-based variance between groups. The room was technically supported by the researcher, whose role was to guide participants through the experimental procedure and equipment setup, help them with software functionalities during the tutorial, and ensure that all data were being captured.

3.4. Experimental procedure

The experimental procedure consisted of five steps: (1) Pre-experiment preparation; (2) introductory synchronous collaborative CAD activity session; (3) first

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

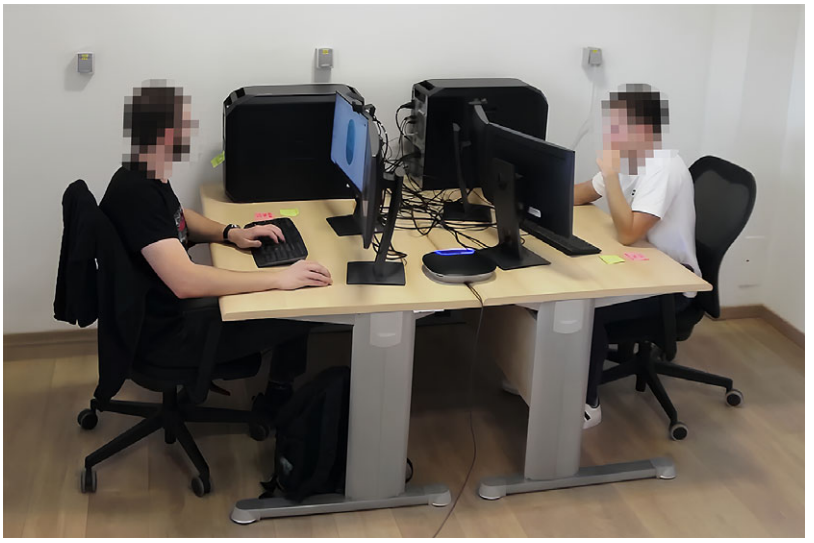


Figure 3. Experimental setup.

synchronous collaborative CAD activity session; (4) second synchronous collaborative CAD activity session; and (5) third synchronous collaborative CAD activity session.

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. Participants in the experimental condition also received additional tutorials on ChatGPT, which covered the user interface, application usage and prompt engineering techniques. This step aimed to familiarise participants with the CAD tool and, for those in the experimental condition, with ChatGPT. Consent forms were also signed during this phase. Steps two through five took place on the same day for each team. The experiment began with participants meeting in person, at which point a researcher guided them to randomly assigned seats in the physical room. The second step consisted of a 15-minute introductory session in which participants synchronously modelled a CAD model of a simple housing in Onshape. This session was designed to test Onshape and screen mirroring functionalities, allow team members to familiarise themselves with one another, and, for participants in the experimental condition, to test ChatGPT. After the introductory session and once participants were familiar with the setup, the researcher initiated screen recording, and the team proceeded to the first, second and third synchronous collaborative CAD activity sessions. These sessions corresponded to the tasks mentioned earlier. Teams were allocated 40 minutes, 25 minutes and 20 minutes for Task 1, Task 2 and Task 3, respectively. The duration of each task was defined based on testing to ensure that the allocated time was sufficient for completing the given task at the expected level of CAD and domain-specific knowledge.

Before starting and upon finishing each of the four sessions, the researcher provided each team with details about the next procedural steps. In addition to a document containing detailed task instructions, which remained visible throughout

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

the task duration (TD), participants in the experimental condition were instructed to consult ChatGPT at each decision point in solving the task, rather than deciding autonomously whether to use it. In other words, whenever the team reached a decision regarding how to proceed, whether determining the next modelling step or identifying necessary modifications to the CAD model, they were required to ask ChatGPT for input before taking action. This procedure was implemented to ensure consistent integration of ChatGPT across all teams, treating it as a team member consulted at every stage of solving the task.

Finally, in addition to audio and video recordings, the researcher collected the final CAD outputs after each task session for data analysis.

3.5. Data collection and analysis

As this study represents an initial exploration into the task-dependent effects of ChatGPT support in synchronous collaborative CAD activities, data collection and analysis were designed to capture a range of taskwork variables without presupposing specific outcomes.

Data were collected from two sources. The resulting CAD models (either parts or assemblies) created or modified by teams in each task served as the basis to evaluate the outcome. Furthermore, data were also extracted from Onshape Analytics, which enables non-intrusive tracking of CAD user actions or interactions with CAD during task execution. Recorded as an Onshape log file or a chronological audit trail, the dataset contained every CAD user action performed within CAD. Each data point included timestamps, document and tab details, the team member performing the action and the CAD user action's name. This dataset covered both constructing and modifying actions (e.g., creating a sketch and deleting a part) as well as organising actions (e.g., switching a tab, renaming a document and opening a part) performed while creating and modifying CAD models. To ensure accurate analysis, actions outside the study's focus were excluded from the dataset, such as those automatically logged by a CAD tool and the multiple records generated for a single CAD user action (e.g., when a user creates a sketch).

Although the objective of this study was exploratory, data analysis was based on descriptive and inferential statistics to examine trends and potential differences between experimental conditions and tasks. Descriptive statistics encompassed the average as a measure of central tendency and standard deviation as a measure of variability. Inferential analyses were used to examine overall effects across three tasks. A two-way mixed ANOVA was used to assess the effects of task and experimental condition, including interaction effects. It was conducted with Task (Tasks 1–3) as a within-subject factor and Experimental condition (with ChatGPT support versus without ChatGPT support) as a between-subject factor, including the interaction between task and experimental condition to assess whether the effect of ChatGPT support differed across tasks.

Furthermore, prior to individual task-level comparisons, assumptions were assessed using the Shapiro–Wilk test for normality ($p < 0.05$) and Levene's test for homogeneity of variances ($p < 0.05$). Depending on the nature of the comparison and the results of statistical assumptions, different tests were applied. When assumptions were met, an independent t-test was applied. If the assumptions were violated, the non-parametric Mann–Whitney U test was used instead. To

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

communicate the magnitude and uncertainty of the estimated between-condition differences, corresponding 95% confidence intervals were calculated. For comparisons conducted using the independent-samples *t*-test, confidence intervals were calculated for the mean difference between the ChatGPT-supported and unsupported conditions using the *t* distribution. For Mann–Whitney U comparisons, the Hodges–Lehmann estimator and its corresponding 95% confidence interval were used to estimate the between-condition location shift. To account for multiple comparisons across the tasks within each dependent variable, *p*-values were adjusted by using Bonferroni corrections and reported in the results (*p*-value adjusted). However, because different statistical tests were used for individual task comparisons, the statistical protocol is not fully consistent. In contrast, the repeated measures across tasks were conducted using a single statistical test. This inconsistency in statistical testing reflects a methodological limitation, and results should be interpreted cautiously. Additionally, effect sizes were calculated to complement significance testing and interpret the magnitude of observed differences: partial eta squared (η^2) for ANOVA results, Cohen's *d* for *t*-tests and *r* for non-parametric tests.

The data were also plotted using boxplots that compared conditions across tasks. For this purpose, values of all three variables (number of CAD actions, productivity and efficiency) were *z*-score normalised within each task to enable cross-task comparisons on a standardised scale, allowing all three tasks, despite differing in scale and absolute values, to be visualised in a comparable format. Nevertheless, all inferential statistics were conducted on raw, untransformed values to preserve the meaning of absolute differences. Data were analysed using the Python programming language.

Selected to reflect different aspects of taskwork, focusing on observable CAD output and resource utilisation, four taskwork variables were examined: the number of CAD user actions, productivity, efficiency and CAD dominance. These variables were selected to reflect different aspects of taskwork, with a focus on observable CAD output and resource utilisation.

3.5.1. Number of CAD actions

The number of CAD actions (CA_{Total}) captures the overall amount of single-user-initiated events recorded by Onshape Analytics. Before counting performed CAD actions, CAD actions were mapped to the action type dimension of the MUCAD-CLF framework (Deng *et al.* 2022b), which provides a structured approach for categorising CAD user actions. For the purpose of this study, only the action type dimension of the framework was applied, allowing each CAD user action to be assigned to a predefined action type, namely creating, editing, deleting, revising and other actions. However, these categories were not analysed separately. Instead, the categorisation was used to ensure consistency in identifying and counting actions, with all CAD user actions subsequently treated as equivalent when counting the total number of CAD user actions performed per team (CA_{Total}).

3.5.2. Productivity

Productivity (*P*) was conceptualised as the rate of interaction with CAD during task execution, operationalised as the number of CAD actions (CA_{Total}) performed per minute of TD. TD was operationalised as the total time from the official task start to

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

task completion under the assigned experimental condition, as determined from the session recordings.

$$P = \frac{CA_{Total}}{TD}.$$

Specifically, this metric provides insight into the team's number of performed CAD user actions relative to the time spent on the task. Therefore, when the number of CAD user actions remains constant, shorter task completion time results in higher productivity, reflecting faster task progression for the same amount of interaction with CAD. Conversely, when time is held constant, higher productivity corresponds to a higher number of CAD user actions.

Consequently, in the ChatGPT-supported condition, TD included time spent interacting with ChatGPT (e.g., composing prompts, reading responses and integrating or rejecting suggestions) as part of the task execution. Such interaction represents part of the overall workflow and associated time cost. Importantly, productivity should not be interpreted as a standalone performance metric since the higher rate of CAD actions per time does not reflect the quality of the resulting CAD model, nor does it reflect whether all performed CAD actions were added with respect to the task's requirements. For that reason, productivity is not an objective to be maximised and should be interpreted jointly with other CAD-based measures. The efficiency metric is therefore also reported as a more task-related measure that takes into account what has been gained in terms of the task outcome or, in the case of this study, the resulting CAD models.

3.5.3. Efficiency

Efficiency (E) was conceptualised as the relationship between what has been materially gained and the level of resource used (O'Donnell & Duffy 2005). In the context of the CAD activities examined within this study, material gain corresponded to the achieved task outcome, namely the resulting CAD models of the created or modified parts or assemblies. The resource used is time. Accordingly, efficiency was operationalised by evaluating the quality of resulting CAD models, utilising the completeness score (C) as the quality dimension of the resulting CAD model, relative to the TD.

$$E = \frac{C}{TD}$$

The completeness score was derived from the completeness dimension of CAD model quality (Company *et al.* 2015), evaluated along two sub-dimensions: shape replication and size replication. Shape replication refers to the accuracy of the geometric features constituting the resulting CAD model (e.g., a cylinder, a slot, a fillet, a constituting part of an assembly…). Size replication refers to the accuracy of geometric dimensions, quantifying the physical aspects of the CAD model (e.g., length, width, height, radius and angles). For each task, the shape and size completeness values were calculated as the proportion of correctly modelled or defined elements in accordance with the respective task requirements. The final completeness score for each team and each task was obtained as the average of the shape and size completeness values.

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

In Task 1, which required replicating a reference crankshaft model (provided as a STEP file), shape completeness reflected whether all geometric features were modelled, such as the three-level left shaft, the three-level right shaft, the central cylindrical body and the crank lobes with chamfers and semicylindrical features. Size completeness captured the accuracy of their corresponding dimensions (e.g., diameters of each shaft level, lengths of cylindrical parts and chamfer angles).

In Task 2, which required creating a functional assembly of a simplified engine mechanism, the shape and size evaluation followed the definition of assembly (Lupinetti *et al.* 2018). Shape completeness referred to the structural and kinematic correctness of the assembly; that is, whether all required parts (housing, crankshaft, connecting rod, rod cap and piston) were included, positioned correctly and properly constrained using mates that replicated intended mechanical movement. This included two revolute mates between the crankshaft journals and housing sleeves, a fastened mate between the connecting rod and rod cap, a revolute mate between the connecting rod and the piston, and a slider mate between the piston and the housing bore. Size completeness referred to the dimensional relationship between parts, four of which were intentionally incompatible with the provided housing.

In Task 3, which required redesigning an existing crankshaft model for manufacturability by forging, shape was assessed to determine whether all forging-related features were correctly implemented according to the design for forging requirements. This included defining an appropriate parting line that separates the upper and lower die halves, adding draft angles to facilitate part ejection from the die, introducing rounded edges to avoid sharp transitions and ensure uniform metal flow, and locally modifying geometry to achieve smoother cross-sectional changes along the shaft. In addition, the design required material additions to surfaces intended for subsequent machining operations. Size completeness referred to the correctness of the associated dimensions. For instance, whether the applied draft angles, fillet radii and material thicknesses were within manufacturing limits and consistent with general forging design guidelines (Bralla 1998).

3.5.4. CAD dominance

CAD dominance is conceptualised as the difference in contributions of team members to the CAD work during the collaboration within synchronous collaborative CAD activities. The metric was calculated from the same Onshape Analytics action log used to derive the total number of CAD actions (CA_{Total}). However, for CAD dominance, the actions were disaggregated by team member or CAD user involved in the activities. For each team and task, the number of CAD actions performed by the first CAD user (CU_1) and the second CAD user (CU_2) were divided by the total number of CAD actions performed.

$$S(CU_1) = \frac{CA_1}{CA_{Total}} \text{ and } S(CU_2) = \frac{CA_2}{CA_{Total}}$$

CAD dominance was defined as the absolute difference between these shares (Roy *et al.* 2024).

$$CD = |S(CU_1) - S(CU_2)|$$

The resulting values range from 0 to 1, where 0 indicates even contributions and 1 indicates complete dominance by one CAD user within synchronous collaborative

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

CAD activities (i.e., one CAD user performed all recorded CAD actions). Higher CAD dominance, therefore, reflects greater imbalance in contributions, whereas lower values reflect more even participation. Thus, CAD dominance is not interpreted as a metric that should be maximised or minimised. Instead, it serves as an indicator of how CAD execution is organised within the pair (i.e., whether contributions are balanced or concentrated in one member) and should be interpreted in relation to the task context and the other taskwork outcomes.

4. Results

This section presents the results of the experiment, focusing on the task-dependent effects of ChatGPT support in synchronous collaborative CAD activities. The results are structured around four taskwork variables: number of CAD user actions, productivity, efficiency and CAD dominance. For each variable, comparisons are made between teams with and without ChatGPT support across the three collaborative CAD tasks. Each subsection is organised as follows: it begins with descriptive statistics, followed by assumption checks, inferential statistics, effect size reporting and visualisations using boxplots.

Depending on whether assumptions for parametric testing were met, different statistical tests were used for individual task comparisons to ensure validity. However, this approach introduced an inconsistency in the statistical protocol which should be considered when interpreting results. As such, the reported results following in the next subsections offer preliminary and exploratory insights into the task-dependent effects of ChatGPT support in synchronous collaborative CAD activities.

4.1. Number of CAD user actions

The comparison of the number of CAD user actions across tasks and experimental conditions indicates that the distribution of actions varied between teams with and without ChatGPT support, as shown in Figure 4. In Task 1, teams supported by ChatGPT performed a higher mean number of CAD user actions (M = 59.45, SD = 33.92) compared to teams without ChatGPT support (M = 48.91, SD = 18.17), though the variability in the experimental condition group was notably larger. In Task 2, teams without ChatGPT support had a higher mean number of performed CAD user actions compared with teams with ChatGPT support (M = 205.55, SD = 42.60 versus M = 151.36, SD = 39.76), while in Task 3, teams without ChatGPT support had a slightly higher mean number of performed CAD user actions (M = 39.27, SD = 12.79) compared to teams with ChatGPT support (M = 33.82, SD = 10.05).

Furthermore, the Shapiro–Wilk test revealed a violation of normality for the experimental condition group in Task 1 (p = 0.003), while normality was met in all other groups. Levene's test confirmed homogeneity of variances for all tasks (Task 1: p = 0.569; Task 2: p = 0.713; Task 3: p = 0.785).

To assess the main and interaction effects of task and experimental condition, a two-way mixed ANOVA was conducted. The analysis, detailed in Table 1, revealed a significant main effect of task ($F(2, 40) = 193.99$, p = 0.007), indicating significant differences in the number of performed CAD user actions across the three tasks,

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

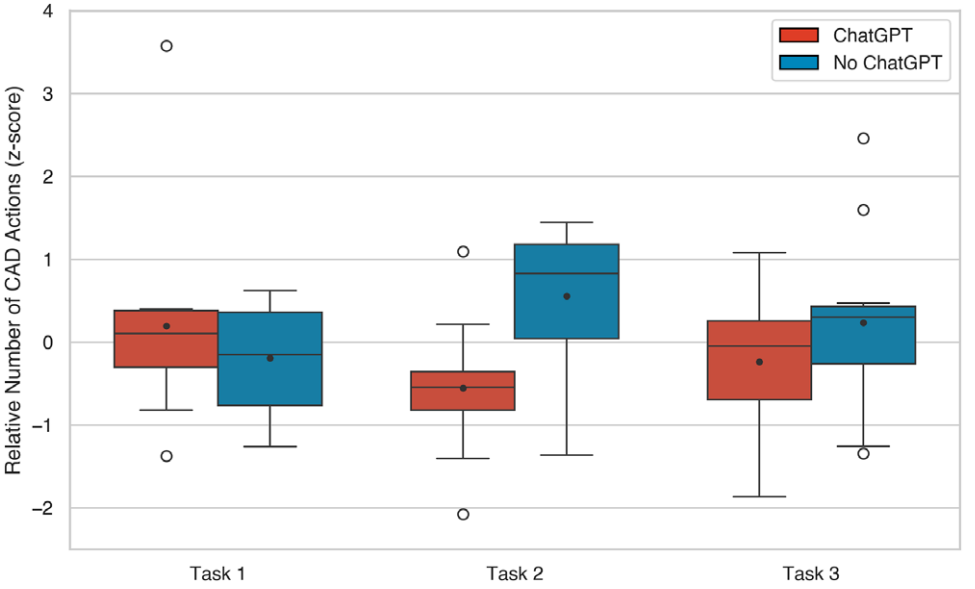


Figure 4. Number of performed CAD user actions across synchronous collaborative CAD activity tasks.

Table 1. Two-way mixed ANOVA results for number of CAD user actions across tasks and ChatGPT conditions

Effect	*F*	df	*p*_value	Effect size
Experimental condition	2.14	(1, 20)	0.159	$\eta^2 = 0.097$
Task	193.99	(2, 40)	0.007*	$\eta^2 = 0.907$
Task × Condition interaction	5.74	(2, 40)	0.006*	$\eta^2 = 0.223$

*Significant value ($p < 0.05$).

with a large effect size ($\eta^2 = 0.907$). The main effect of the experimental condition was not statistically significant ($F(1, 20) = 2.14$, $p = 0.159$), with a small-to-medium effect size ($\eta^2 = 0.097$). The interaction between task and condition was significant ($F(2, 40) = 5.74$, $p = 0.006$) with a large effect size ($\eta^2 = 0.223$), indicating that the effect of ChatGPT support varied depending on the task.

Post hoc comparisons (Table 2) revealed no significant difference in Task 1 (estimated difference = 6.00 actions, 95% CI [−10.00, 26.00], $U = 64.5$, $p = 0.818$, p_adj = 1.000), with a small effect size ($r = -0.07$), though ChatGPT teams had a higher mean with greater variability. In Task 2, teams without ChatGPT support performed a significantly higher number of CAD user actions (estimated difference = −54.18 actions, 95% CI [−90.84, −17.52], $t = -3.084$, $p = 0.006$, p_adj = 0.018) with a large negative effect size ($d = -1.31$). Task 3 showed no significant difference (estimated difference = −5.45 actions, 95% CI [−15.72, 4.81], $t = -1.112$, $p = 0.280$, p_adj =0.840), but a moderate negative effect size

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

Table 2. Comparison of number of performed CAD user actions in synchronous collaborative CAD activity tasks with and without ChatGPT support

Task	Shapiro p (ChatGPT/No ChatGPT)	Levene p	Between-group difference [95% CI]	Test	p_value	p_value adjusted	Effect size
Task 1	0.003*/0.313	0.569	6.00 [−10.00, 26.00]	Mann–Whitney U	0.818	1.000	$r = -0.07$
Task 2	0.626/0.115	0.713	−54.18 [−90.84, −17.52]	t-test	0.006*	0.018*	$d = -1.31$
Task 3	0.969/0.403	0.785	−5.45 [−15.72, 4.81]	t-test	0.280	0.840	$d = -0.47$

*Significant value ($p < 0.05$).

($d = -0.47$) indicated a tendency for a reduced number of CAD user actions with ChatGPT support.

4.2. Productivity

The comparison of productivity across tasks and experimental conditions shows that teams without ChatGPT support consistently achieved higher productivity than teams with ChatGPT support, as illustrated in Figure 5. Descriptive statistics showed that teams without ChatGPT achieved higher productivity across all three tasks. In Task 1, teams without ChatGPT support outperformed teams with

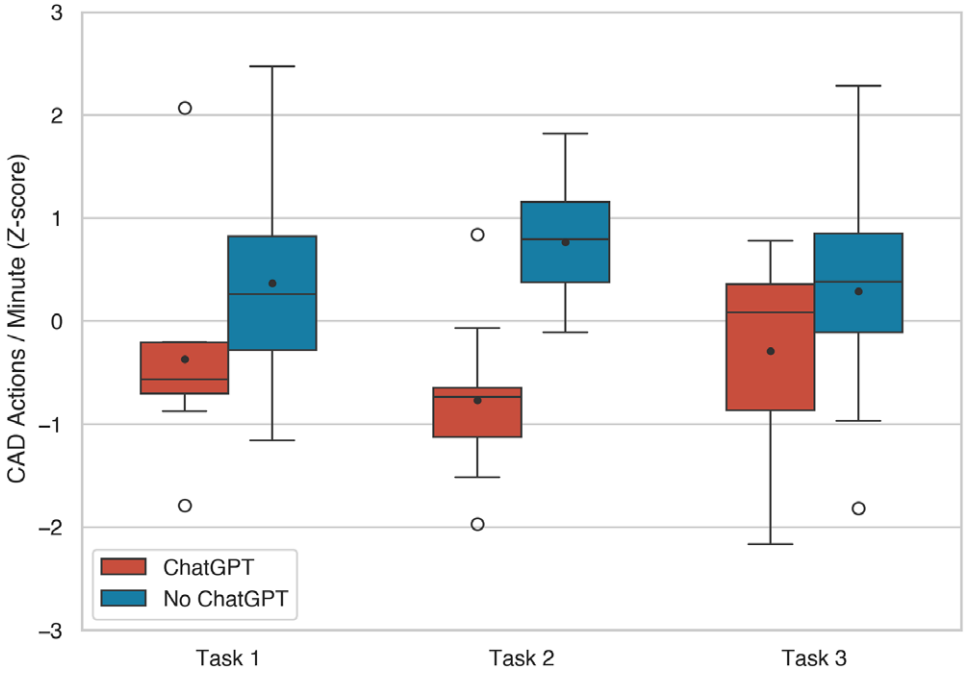


Figure 5. Productivity across synchronous collaborative CAD activity tasks.

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

ChatGPT support ($M = 4.80$, SD = 1.83 versus $M = 3.48$, SD = 1.67). In Task 2, the difference was more pronounced (No ChatGPT: $M = 24.22$, SD = 3.41; ChatGPT: $M = 15.63$, SD = 4.13), with the ChatGPT-supported teams showing lower productivity. In Task 3, the productivity gap between experimental conditions narrowed ($M = 4.75$, SD = 1.08 versus $M = 4.18$, SD = 0.93), but teams without ChatGPT support still had a higher average.

Furthermore, Shapiro–Wilk tests showed that the assumption of normality was met for all subgroups ($p > 0.05$) except for teams with ChatGPT support in Task 1 ($p = 0.0053$). Levene's test confirmed homogeneity of variances across conditions for each task (Task 1: $p = 0.4408$; Task 2: $p = 0.9831$; Task 3: $p = 0.8791$).

A two-way mixed ANOVA (Table 3) revealed a significant main effect of experimental condition ($F(1, 20) = 24.77$, $p = 7.241 \times 10^{-5}$, $\eta^2 = 0.553$), with a large effect size, suggesting a significant influence of ChatGPT support on productivity. A significant main effect of task was also observed ($F(2, 40) = 349.05$, $p = 4.773 \times 10^{-23}$, $\eta^2 = 0.946$), indicating that productivity varied significantly between tasks. The interaction between task and experimental condition was also significant ($F(2, 40) = 20.99$, $p = 5.823 \times 10^{-7}$, $\eta^2 = 0.512$), with a large effect size, showing that the influence of ChatGPT support differed by task.

Task-level comparisons further clarified this interaction, as detailed in Table 4. In Task 1, the estimated difference was −1.39 CAD user actions per minute, 95% CI [−2.65, −0.05]. The Mann–Whitney U test showed a significant difference ($p = 0.042$) that did not remain significant after the Bonferroni correction (p_adj = 0.125), with a medium-to-large effect size ($r = 0.521$). In Task 2, an

Table 3. Two-way mixed ANOVA results for the productivity variable across tasks and ChatGPT conditions

Effect	F	df	p_value	Effect size
Experimental condition	24.77	(1, 20)	7.241×10^{-5}*	$\eta^2 = 0.553$
Task	349.05	(2, 40)	4.773×10^{-23}*	$\eta^2 = 0.946$
Task × Condition interaction	20.99	(2, 40)	5.823×10^{-7}*	$\eta^2 = 0.512$

*Significant value ($p < 0.05$).

Table 4. Comparison of productivity variables in synchronous collaborative CAD activity tasks with and without ChatGPT support

Task	Shapiro p (ChatGPT/No ChatGPT)	Between-group difference [95% CI]	Levene p	Test	p_value	p_value adjusted	Effect size
Task 1	0.005*/0.865	−1.39 [−2.65, −0.05]	0.441	Mann–Whitney U	0.042*	0.125	$r = 0.521$
Task 2	0.436/0.914	−8.59 [−11.96, −5.21]	0.983	t-test	<0.001*	<<0.0001*	$d = -2.269$
Task 3	0.210/0.821	−0.58 [−1.47, 0.32]	0.879	t-test	0.194	0.581	$d = -0.574$

*Significant value ($p < 0.05$).

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

independent *t*-test revealed a significant difference (estimated difference = −8.59 CAD user actions per minute, 95% CI [−11.96, −5.21], $p < 0.001$, p_adj < 0.001) and a large effect size ($d = -2.27$), indicating a strong negative influence of ChatGPT support. In Task 3, the estimated difference was −0.58 CAD user actions per minute, 95% CI [−1.47, 0.32], the difference was not statistically significant ($p = 0.194$, p_adj = 0.581), but a moderate effect size ($d = -0.574$) suggests a trend towards reduced productivity of teams with ChatGPT support. The results indicate that ChatGPT support was associated with lower productivity and that these differences were not only statistically significant but also practically meaningful.

4.3. Efficiency

On average, teams without ChatGPT support were more efficient in the first two tasks, whereas in the third task the difference was reversed but less pronounced, as illustrated in Figure 6. In Task 1, teams without ChatGPT support had a higher mean efficiency ($M = 0.044$, SD = 0.025) than teams with ChatGPT support ($M = 0.017$, SD = 0.008). A similar pattern was observed in Task 2, where teams without ChatGPT support had higher mean efficiency than teams with ChatGPT support ($M = 0.040$, SD = 0.015 versus $M = 0.025$, SD = 0.010). In contrast, in Task 3, teams with ChatGPT support had a slightly higher average efficiency ($M = 0.023$, SD = 0.009) than teams without ChatGPT support ($M = 0.019$, SD = 0.016).

Furthermore, Shapiro–Wilk tests revealed a violation of the normality assumption for the control group in Task 1 ($p = 0.0029$) and in the experimental group in

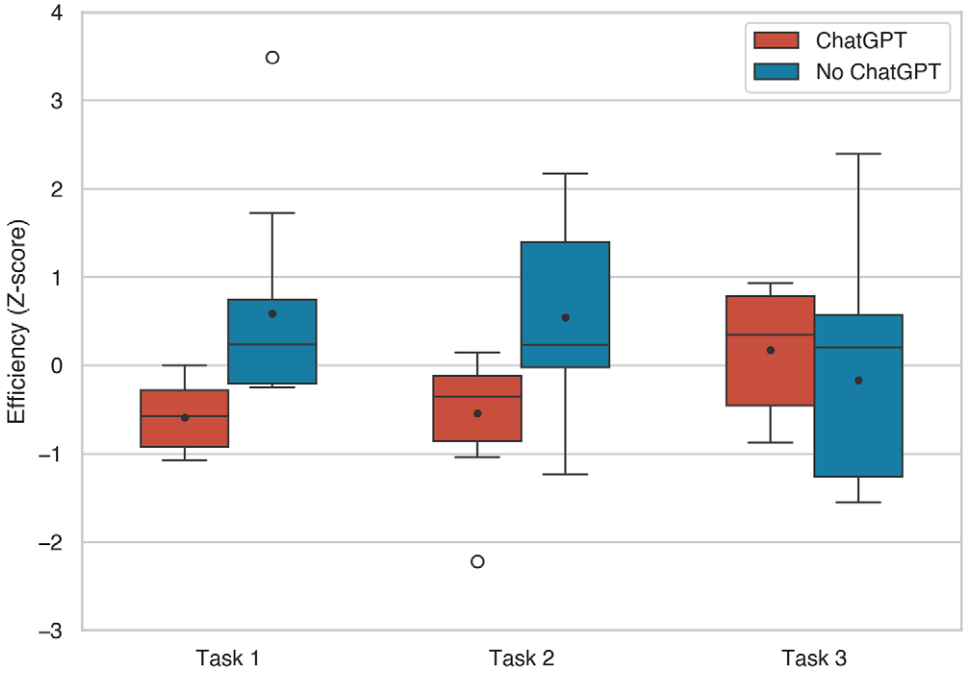


Figure 6. Efficiency across synchronous collaborative CAD activity tasks.

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

Table 5. Two-way mixed ANOVA results for efficiency variable across tasks and ChatGPT conditions

Effect	F	df	p_value	Effect size
Experimental condition	7.36	(1, 20)	0.013*	$\eta^2 = 0.269$
Task	5.58	(2, 40)	0.007*	$\eta^2 = 0.218$
Task × Condition interaction	8.20	(2, 40)	0.001*	$\eta^2 = 0.291$

*Significant value ($p < 0.05$).

Task 2 ($p = 0.0498$). Levene's test confirmed that homogeneity of variances was met for all three tasks (all $p > 0.05$).

A two-way mixed ANOVA (Table 5) revealed a significant main effect of experimental condition ($F(1, 20) = 7.36$, $p = 0.013$) with a medium to large effect size ($\eta^2 = 0.269$), indicating that teams with ChatGPT support were less efficient than teams without ChatGPT support. A significant main effect of task was also found ($F(2, 40) = 5.58$, $p = 0.007$), with a medium effect size ($\eta^2 = 0.218$), suggesting that efficiency varied across tasks. Furthermore, the interaction between task and experimental condition was significant ($F(2, 40) = 8.20$, $p = 0.001$), with a large effect size ($\eta^2 = 0.291$), indicating that the influence of ChatGPT support differed depending on the task.

Task-level comparisons provided additional insights, as detailed in Table 6. In Task 1, efficiency was significantly lower in teams supported by ChatGPT than in teams without ChatGPT support (estimated difference = −0.0183, 95% CI [−0.0336, −0.0109], $p = 0.0005$, p_adj = 0.002), with a large effect size ($r = 0.884$). In Task 2, efficiency was also significantly lower in teams supported by ChatGPT (estimated difference = −0.0140, 95% CI [−0.0273, −0.0040], $p = 0.011$, p_adj = 0.034), with a large effect size ($r = 0.645$). In contrast, Task 3 showed no significant difference (estimated difference = 0.0042, 95% CI [−0.0073, 0.0156], $p = 0.451$, p_adj = 1.000), and the small-to-moderate effect size ($d = 0.329$) suggests limited practical impact. Overall, these results indicate that

Table 6. Comparison of efficiency in synchronous collaborative CAD activity tasks with and without ChatGPT support

Task	Shapiro p (ChatGPT/No ChatGPT)	Between-group difference [95% CI]	Levene p	Test	p_value	p_value adjusted	Effect size
Task 1	0.429/0.003*	−0.0183 [−0.0336, −0.0109]	0.169	Mann–Whitney U	0.001*	0.002*	$r = 0.884$
Task 2	0.0498*/0.808	−0.0140 [−0.0273, −0.0040]	0.220	Mann–Whitney U	0.011*	0.034*	$r = 0.645$
Task 3	0.076/0.143	0.0042 [−0.0073, 0.0156]	0.087	t-test	0.451	1.000	$d = 0.329$

*Significant value ($p < 0.05$).

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

ChatGPT support was associated with reduced efficiency in the first two tasks but had no clear effect in the third task.

4.4. CAD dominance

The comparison of CAD dominance across tasks and experimental conditions is shown in Figure 7. In Task 1, teams supported by ChatGPT had a higher mean CAD dominance ($M = 0.590$) than teams without ChatGPT support ($M = 0.376$). In Task 2, teams without ChatGPT support showed a higher mean CAD dominance than ChatGPT-supported teams ($M = 0.258$ versus $M = 0.136$). In Task 3, teams supported by ChatGPT again showed a higher mean CAD dominance ($M = 0.440$) than teams without ChatGPT support ($M = 0.307$). Assumption checks indicated that normality and homogeneity of variances were met for both groups in Task 1 and Task 2 (Shapiro $p > 0.05$; Levene $p > 0.10$), and independent t-tests were therefore applied. For Task 3, normality was violated in both conditions (Shapiro $p < 0.05$), while homogeneity of variances was met, and a Mann–Whitney U test was used. A two-way mixed ANOVA (Table 7) showed a significant main effect of experimental condition ($F(2, 40) = 7.122$, $p = 0.002$), with a large effect size ($\eta^2 = 0.176$), while the main effect of task was not significant ($F(1, 20) = 1.447$, $p = 0.234$), with a small effect size ($\eta^2 = 0.018$), and the interaction between task and experimental condition did not reach significance ($F(2, 40) = 2.631$, $p = 0.0803$), with a medium effect size ($\eta^2 = 0.065$).

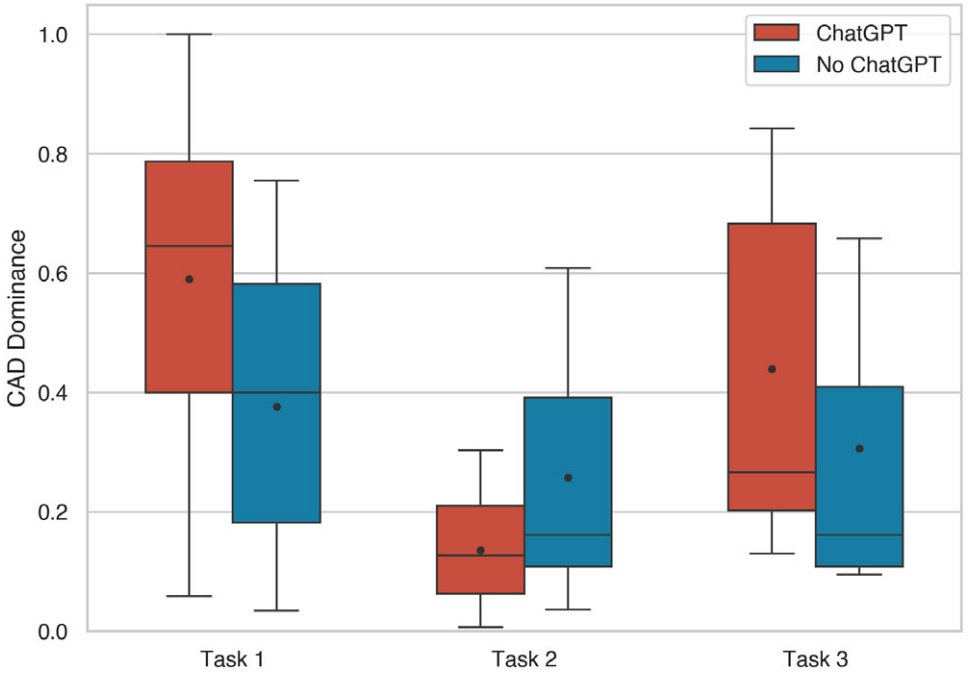


Figure 7. CAD dominance across synchronous collaborative CAD activity tasks.

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

Table 7. Two-way mixed ANOVA results for CAD dominance across tasks and ChatGPT conditions

Effect	*F*	df	*p*_value	Effect size
Experimental Condition	7.12	(1, 20)	0.002*	$\eta^2 = 0.176$
Task	1.45	(2, 40)	0.234	$\eta^2 = 0.018$
Task × Condition Interaction	2.63	(2, 40)	0.081	$\eta^2 = 0.065$

*Significant value ($p < 0.05$).

Table 8. Comparison of CAD dominance in synchronous collaborative CAD activity tasks

Task	Shapiro *p* (ChatGPT/No ChatGPT)	Between-group difference [95% CI]	Levene *p*	Test	*p*_value	*p*_value adjusted	Effect size
Task 1	0.765/0.239	0.2146 [−0.0398, 0.4690]	0.731	*t*-test	0.0933	0.280	$d = 0.75$
Task 2	0.657/0.198	−0.1221 [−0.2660, 0.0218]	0.101	*t*-test	0.085	0.254	$d = -0.77$
Task 3	0.041*/0.001*	0.1054 [−0.0148, 0.4676]	0.681	Mann–Whitney U	0.076	0.228	$r = -0.49$

*Significant value ($p < 0.05$).

Task-level comparisons (Table 8) yielded non-significant differences for Task 1 (estimated difference = 0.2146, 95% CI [−0.0398, 0.4690], $p = 0.0933$, p_adj = 0.280), with a medium-to-large effect size ($d = 0.75$), for Task 2 (estimated difference = −0.1221, 95% CI [−0.2660, 0.0218], $p = 0.085$, p_adj = 0.254), with a medium-to-large effect size ($d = -0.77$) and for Task 3 (estimated difference = 0.1054, 95% CI [−0.0148, 0.4676], $p = 0.076$, p_adj = 0.228), with a large effect size ($r = -0.49$).

5. Discussion

The results show that the influence of ChatGPT support on synchronous collaborative CAD activities is task-dependent, with its effects varying across the examined taskwork variables. This section discusses the task-dependent effects of ChatGPT on the number of CAD user actions, productivity, efficiency and CAD dominance (Section 5.1). It then reflects on the methodological lessons learned from the experimental design (Section 5.2) before outlining study implications for practice and education (Section 5.3), as well as the study limitations (Section 5.4).

5.1. Task-dependent effects of ChatGPT support

This exploratory study examined the task-dependent effects of ChatGPT support in synchronous collaborative CAD activities. By analysing four taskwork variables, namely, number of CAD user actions, productivity, efficiency and CAD dominance, across three distinct collaborative CAD tasks, the study offers early evidence

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

that ChatGPT may play an important role depending on task characteristics and the type of knowledge required.

Rather than revealing a universally positive or negative effect, the results suggest that the impact of ChatGPT support is task-sensitive. In tasks where procedural and declarative CAD knowledge dominated (Tasks 1 and 2), ChatGPT did not appear to offer substantial benefit. In fact, teams with ChatGPT support often exhibited lower productivity and efficiency, indicating that the tool may introduce unnecessary overhead in contexts where CAD-based task execution is the primary requirement. These findings resonate with limitations of LLMs in structured, geometry-focused domains where precision and accuracy are required (Makatura *et al.* 2024), the ChatGPT support is more constrained by the limitations of natural language processing. Conversely, in Task 3, where domain-specific knowledge about DfM principles was essential, no statistically significant between-condition differences were found. Nevertheless, the small positive point estimate for efficiency, together with lower estimated numbers of CAD user actions and lower productivity in the ChatGPT-supported teams, suggests a different pattern from that observed in the first two tasks. These findings do not demonstrate that ChatGPT improved taskwork in Task 3, but they leave open the possibility that its use as an additional knowledge source may be more appropriate for tasks requiring general, conceptual or rule-based knowledge and reasoning, such as interpretation of design principles or the integration of abstract rules (e.g., DfM, as in the case of forging in this study). This aligns with ChatGPT's strength as a general-purpose language model that can support exploratory reasoning and knowledge supplementation, especially when users know what to ask and how to interpret and use the suggestions, recommendations or responses from LLMs like ChatGPT (Li *et al.* 2024b).

The addition of CAD dominance extends the interpretation of taskwork by clarifying that ChatGPT may influence not only how much interaction with CAD occurs (CAD user actions) and how effectively it translates into task outcomes (efficiency), but also, given the synchronous nature of collaborative CAD activities, whether the CAD work is concentrated in one team member or distributed more evenly across the pair. Rather than indicating imbalance, the observed CAD dominance is more consistent with asymmetries in the distribution of taskwork within the team, potentially reflecting a division of work rather than mere inequality (Roy *et al.* 2024). In the ChatGPT-supported condition, this measure is especially informative because the tool may introduce an additional interaction stream that must be managed in real time (Han *et al.* 2024). Even though task-level comparisons did not reach statistical significance, higher dominance by one team member in Task 1 and Task 3 may explain the emergence of a division of work due to the new coordination streamline, where one member works in CAD while the other is focused more on the work in ChatGPT (prompting, reading, interpreting and briefing the team member executing the CAD work). In a synchronous activity with a limited time, this work division can be a pragmatic adaptation to maintain continuity of CAD work while still leveraging an external knowledge source in the form of ChatGPT. However, whether it benefits taskwork depends on task demands. The fact that higher dominance appears in both Task 1 and Task 3 suggests that ChatGPT support may have encouraged role concentration, even if the nature of tasks differed. By contrast, the lower dominance in the ChatGPT-supported condition compared to teams without ChatGPT support may indicate that the integration

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

of ChatGPT has disrupted this work division in assembly tasks, possibly because assembly requires more balanced and joint work to manage mates and constraints, and immediate feedback, or alternate more frequently, both of which would reduce dominance relative to a division that is concentrated in one member according to that member role. Nevertheless, the present results cannot establish that these mechanisms occurred. Taken together, the observed pattern provides no clear evidence that ChatGPT support reconfigured CAD work within the team. Rather, it raises the possibility that any such reconfiguration may depend on task characteristics. This possibility should, therefore, be examined in future research.

While this task-dependent variation shows the promise of LLMs as support tools in synchronous collaborative CAD activities, especially in tasks needing wider engineering knowledge than purely knowledge about how to utilise the CAD functionalities, the findings should not be generalised. The influence of ChatGPT is mediated by multiple factors, including the nature of the task, the tool itself, the way it is integrated into the workflow, and the characteristics of the users themselves. In particular, the observed differences across tasks underscore that effective utilisation of ChatGPT is not automatic or guaranteed; it requires deliberate alignment between tool capabilities and task demands.

Moreover, as an exploratory study, this study was not designed to isolate all the variables that shape ChatGPT's effective use and influence on the synchronous collaborative CAD activities. The next section, therefore, discusses the lessons learned from the experimental design, highlighting the experimental control, participant preparation and tool integration that constrain the generalisability of the current findings. These reflections form the basis for designing more rigorous future studies to better understand when and how ChatGPT can enhance taskwork in collaborative CAD activities.

Importantly, these findings should also be considered in light of the synchronous nature of the activities. Similar to the use of collaborative CAD itself, ChatGPT was not consulted in isolation but in real time, embedded in joint and simultaneous work on the tasks. This integration means that each query to ChatGPT required coordinating and joint interpretation of ChatGPT's suggestions before continuing the CAD work. In Tasks 1 and 2, such synchronous work may disrupt the flow of taskwork, possibly introducing coordination overhead (Phadnis *et al.* 2021), whereas in Task 3 they supported shared reasoning around manufacturability principles for forging, where clarification and alignment within teams were valuable. This highlights that ChatGPT's role in synchronous collaborative CAD may extend beyond simply providing information and knowledge, as it also may reshape the teamwork processes. While our current analysis focused on taskwork measures, further qualitative investigation of interaction logs and communication is needed to better understand the overall influence of ChatGPT support on synchronous collaborative CAD activities and teamwork.

Finally, leveraging collaborative CAD as a research instrument for engineering design, the Onshape action logs enabled in situ observation of work execution and objective operationalisation of taskwork, including not only outcomes and time but also how work is distributed within the pairs (Vrolijk & Olechowski 2025). These action-based insights show that ChatGPT's influence is not uniform but varies with task demands, which supports the human–AI collaboration perspective in engineering design (Song *et al.* 2024). AI effects in engineering design are socio-technical and contingent on integration into work practices rather than being inherently

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

beneficial. More broadly, the results imply that introducing LLM support into the collaborative, tool-mediated engineering design activities can change how taskwork is executed.

5.2. Methodological lessons for designing experiments on ChatGPT-supported synchronous collaborative CAD activities

While this study provides initial exploratory evidence on how ChatGPT may influence taskwork in synchronous collaborative CAD activities, the primary value lies in what it reveals about the complexity of conducting controlled experiments in this domain. Given the lack of existing research in this context, it contributes to the broader understanding of how to design and conduct experiments in the context of synchronous collaborative CAD activities. The present study adopts a more comprehensive approach, incorporating additional taskwork variables to better capture the dynamics of ChatGPT-supported synchronous collaborative CAD activities. The exploratory nature limited experimental control over variables known to affect the effective use of LLMs like ChatGPT. The experimental design did not fully isolate the independent variable, ChatGPT support during synchronous collaborative CAD activities, which constrains the robustness and generalisability of the findings (Robson & McCartan 2016). Differences in taskwork across tasks and conditions underscore the need to systematically account for and control variables influencing the effective utilisation of ChatGPT. This further emphasises the importance of considering the intertwined effects of task-, tool- and user-based factors in experimental design. Although the study controlled for variables such as the AI tool, collaborative CAD task, artefact characteristics and both CAD and domain-specific engineering knowledge due to the randomisation, other variables remained unaccounted for.

The contrast between tasks in this study indicates that ChatGPT support offers limited benefit in tasks primarily focused on procedural CAD knowledge, whereas its potential was shown in Task 3, where domain-specific knowledge and rule-based reasoning were required. First, the results demonstrate that tasks need to be deliberately designed to require consultation with external knowledge sources, where ChatGPT can serve a meaningful role (Ray 2023). Future studies should, therefore, aim for domain-specific knowledge-intensive engineering design tasks, particularly in embodiment design (Veisz *et al.* 2012). Unlike early phases such as conceptual design, which can be hindered by using CAD due to premature fixation or bounded ideation (Brown 2009), embodiment design is not merely about creating design representations in terms of geometry and topology in CAD. It involves the iterative elaboration of solution principles into physical forms, integrating multifaceted, often conflicting, requirements regarding production, assembly, cost, maintenance, safety and lifecycle through the application of Design for X (DfX) rules and principles (Favi *et al.* 2022). In other words, embodiment design demands domain-specific knowledge, including, among many, mechanics or manufacturing technologies (Pahl & Beitz 1996). Therefore, CAD knowledge alone does not encompass the full breadth of knowledge and skills required for embodiment design. The present results provide empirical motivation for this choice, or when the task required manufacturability knowledge (Task 3), the disadvantages

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

observed in the CAD-procedural tasks were not clearly reproduced, although no statistically supported benefit from integrating an additional knowledge source was demonstrated. In other words, teams may have been more likely to benefit from integrating an additional knowledge source, whereas in CAD-procedural tasks the same integration tended to introduce overhead under the limited and fixed time limits. The variation in CAD-dominance estimates also raises the possibility that teams organised their work differently depending on whether consulting an external knowledge source could be advantageous. However, the present analysis did not demonstrate that such adaptation occurred. Together, these findings reinforce the need to examine tasks in which external knowledge consultation is essential.

Second, the study task-dependent outcomes also underline that ChatGPT support cannot be treated as a uniform independent variable unless the prompting and ChatGPT usage protocol are fully controlled (Robertson *et al.* 2024). ChatGPT often struggles to provide answers that are both on point and at an appropriate level of detail, an issue particularly relevant in engineering contexts (Schulze Balhorn *et al.* 2024). Although participants in the study received tutorials on ChatGPT and prompt engineering, the effective prompting remains dependent on both prior experience or time spent interacting with a tool (Grassini *et al.* 2024) and domain-specific knowledge (Jing *et al.* 2024). Furthermore, the interaction with ChatGPT was not analysed in this study; therefore, it is not controlled to a degree that would allow generalisation beyond this experimental setting. Future experimental designs should therefore reduce variability by providing predefined prompt templates and explicit prompting methods (Ege *et al.* 2025), as well as by controlling input and requested output modalities within those templates (Vaccaro *et al.* 2024). Additionally, because ChatGPT may exhibit knowledge gaps and limited critical reflection (Ege *et al.* 2025), protocols should include explicit verification steps aligned with engineering design requirements for factual accuracy. This would better isolate the effect of the ChatGPT support treatment by reducing variability in ChatGPT responses, given that the tool is known to produce different answers to identical prompts (Schulze Balhorn *et al.* 2024), and by limiting the influence of user-related factors such as AI awareness and AI usage literacy (Jing *et al.* 2024). Additionally, ChatGPT lacks critical reflection in its responses and may have knowledge gaps (Ege *et al.* 2025), which is particularly important in domains where factual accuracy is crucial, such as engineering design.

Third, findings also imply that user-related factors may not be sufficiently controlled. While randomisation in this study helped balance CAD and domain knowledge (Robson & McCartan 2016), the taskwork performance and allocation of CAD work execution captured by CAD dominance may also be interpreted as indicating that pairs differ in how they organise roles and whether they use ChatGPT meaningfully or merely to comply with the experimental requirement. Future experiments should therefore either assess relevant competencies prior to recruitment or provide targeted instruction on the knowledge required for the specific tasks, alongside structured prompting templates, so that participants can articulate requirements accurately and evaluate outputs critically (Jing *et al.* 2024). The present study therefore indicates the need to make the integration procedure realistic: if the goal is to study the influence of ChatGPT as a functional knowledge source, tasks and instructions should mirror authentic practices of consulting domain-specific knowledge sources (here, with ChatGPT), and the

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

workflow integration protocol should be explicit to distinguish meaningful use from compliance.

5.3. Implications for practice and education

In addition to implications for researchers through synthesis of influencing factors and methodological insights for designing experiments in the domain of ChatGPT influence on synchronous collaborative CAD activities, the exploratory study's findings also offer several implications for practitioners and educators. For practitioners, especially those integrating ChatGPT into collaborative CAD activities, the study reveals both the potential and limitations of such tools, particularly in tasks requiring domain-specific knowledge. Practitioners should be aware of the uneven utility of LLMs like ChatGPT across different CAD and collaborative CAD task types and should consider integrating such tools as support that complements human expertise rather than as proactive or autonomous agents or even team members. For educators, the results indicate a need to develop student competencies that encompass not only domain-specific and CAD knowledge but also AI literacy. This also means preparing students to use ChatGPT in ways that complement, rather than replace, their own engineering judgement. For example, students need to learn how to frame CAD-related problems, such as debugging assembly constraints, adapting and modifying CAD models, or interpreting DfX rules, into effective prompts that ChatGPT can process and how to critically evaluate the textual responses before applying them to the CAD model. In synchronous collaborative CAD activities, this competency extends further. Both practitioners and students might also be educated in how to coordinate the use of ChatGPT within the team, including collaboratively deciding when and how to utilise tools like ChatGPT, interpreting its responses and balancing the time spent interacting with ChatGPT against the time they have for the work at hand. Educating students to implement tools like ChatGPT into synchronous collaborative CAD activities in this way would ensure that they develop technical proficiency in CAD and the skills to utilise AI tools effectively and responsibly, ensuring they maintain accountability for their decisions.

5.4. Limitations

Several limitations constrain the interpretation of this exploratory study on ChatGPT's influence in synchronous collaborative CAD activities. First, due to the exploratory nature, not all task-, tool- and user-based influencing factors were controlled. The experiment did not control for user characteristics such as AI usage literacy, personality traits, or trust in ChatGPT, all of which are known to influence the effective utilisation of LLM tools. Second, the integration of ChatGPT into the workflow was based on general instructions rather than standardised prompting protocols or templates. This introduces variability in how participants interpreted and used ChatGPT, thus influencing taskwork. Consequently, the reported analysis focuses on the CAD outcome in terms of the resulting CAD model and user actions, therefore, it does not allow for interpretation of ChatGPT use as an interaction process. Prompting and response utilisation are also integrated into the synchronous collaborative CAD activities supported by ChatGPT. Prompts are formulated, and the ChatGPT suggestions are discussed among team members

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

before being translated into CAD user actions. Without examining both the human–AI interaction and the communication between team members, the study provides exploratory evidence at the CAD level, but it does not account for how differences in interactions with ChatGPT may have contributed to the observed task-dependent effects of ChatGPT support. The experimental protocol required teams in the ChatGPT-supported condition to consult ChatGPT at each decision point. Although this requirement ensured that teams were consistently exposed to the AI support, it may also have increased interaction and coordination overhead, particularly when participants would not otherwise have considered AI assistance necessary. This prescribed pattern of use differs from use in practice, where teams can decide whether and when consulting an AI tool is likely to be useful. Consequently, the findings reflect the integration of ChatGPT under a forced-use protocol and should not be generalised directly to situations in which AI use is optional. Future studies should compare prescribed and discretionary use protocols. These issues overlap with the methodological lessons outlined in Section 5.2. In addition, while the selected taskwork metrics (number of CAD user actions, productivity, efficiency and CAD dominance) provided valuable insights into taskwork, they are evaluated separately and are not synthesised into a holistic assessment of overall taskwork. As a result, the findings offer a fragmented rather than fully integrated account of taskwork in synchronous collaborative CAD activities. More specifically, the study does not provide a basis for determining how the observed variables should be interpreted in relation to one another. This limits broader conclusions about the overall influence of ChatGPT support in synchronous collaborative CAD activities. Moreover, the selected metrics do not fully capture the complexity of collaborative CAD activities, including teamwork in addition to taskwork. Finally, because the study was conducted with engineering students performing CAD tasks within a controlled laboratory setting, the findings may not directly translate to professional engineering contexts.

6. Conclusion

This study presents an exploratory investigation examining how ChatGPT influences taskwork in synchronous collaborative CAD activities. By employing a between-subjects experimental design across three collaborative CAD tasks, the study investigated how ChatGPT support influences four taskwork variables: number of CAD user actions, productivity, efficiency and CAD dominance. The results showed that the influence of ChatGPT on synchronous collaborative CAD activities is task-dependent and cannot be generalised across all collaborative CAD tasks. ChatGPT support was associated with lower productivity and efficiency in tasks primarily requiring CAD knowledge, whereas these disadvantages were not clearly reproduced when domain-specific engineering knowledge was needed. These conclusions reflect a prescribed-use protocol in which teams were required to consult ChatGPT at each decision point and may not extend directly to discretionary ChatGPT use in synchronous collaborative CAD activities. The study also demonstrates that ChatGPT support in synchronous collaborative CAD activities is not automatically beneficial. Rather, it requires consideration of when and how the tool should be integrated into the synchronous collaborative activities. Beyond the empirical findings, the study also provides methodological insights for designing future experiments that aim to explore the influence of LLMs, such as

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

ChatGPT, on synchronous collaborative CAD activities. Such methodological considerations are essential to ensure experimental validity when investigating AI-supported collaboration. These insights include the importance of isolating the independent variable of ChatGPT support through task selection, participant recruitment and preparation that develop both domain-specific engineering knowledge and AI knowledge to reduce user-based variability, as well as standardising prompt engineering and expanding taskwork evaluation metrics beyond using CAD user actions. As such, the study contributes both empirical and methodological knowledge to the growing body of work at the intersection of AI and engineering design; more specifically, in contexts where collaborative CAD is utilised. On a broader level, it represents a first step towards understanding when and why LLMs enhance or hinder the performance of activities in collaborative design. It also highlights emerging questions about the competencies required for AI-supported engineering design and implications for future engineering education.

Acknowledgements

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

Financial support

Croatian Science Foundation project IP-2022-10-7775: Data-driven Methods and Tools for Design Innovation – DATA-MATION.

Availability of data and material

Data are available upon request.

Competing interest

The authors declare no conflict of interest.

Ethics approval

Ethical approval was obtained for the Croatian Science Foundation projects IP-2022-10-7775.

Code availability

Not applicable.

Consent to participate

The authors obtained consent to participate from the participants.

Consent for publication

The authors obtained consent to participate from the participants.

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

References

Al-Mughairi, H. & **Bhaskar, P.** 2025 Exploring the factors affecting the adoption AI techniques in higher education: insights from teachers' perspectives on ChatGPT. *Journal of Research in Innovative Teaching & Learning* **18** (2), 232–247; doi:10.1108/jrit-09-2023-0129.

Andreadis, G., **Fourtounis, G.** & **Bouzakis, K.-D.** 2015 Collaborative design in the era of cloud computing. *Advances in Engineering Software* **81**, 66–72; doi:10.1016/j.advengsoft.2014.11.002.

Arshad, H. & **Olechowski, A.** 2020 *Paired computer-aided design: the effect of collaboration mode on differences in model quality. In Proceedings of the ASME 2020 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*; doi:10.1115/DETC2020-22730.

Ataei, M., **Cheong, H.**, **Grandi, D.**, **Wang, Y.**, **Morris, N.** & **Tessier, A.** 2025 Elicitron: a large language model agent-based simulation framework for design requirements elicitation. *Journal of Computing and Information Science in Engineering* **25** (2); doi:10.1115/1.4067388.

Azemi, F., **Mehmeti, X.** & **Maloku, B.** 2018 *The Importance of CAD/CAE Systems in Development of Product Design and Process of Optimization*. University for Business and Technology.

Bach, T. A., **Khan, A.**, **Hallock, H.**, **Beltrão, G.** & **Sousa, S.** 2024 A systematic literature review of user trust in AI-enabled systems: an HCI perspective. *International Journal of Human–Computer Interaction* **40** (5), 1251–1266; doi:10.1080/10447318.2022.2138826.

Barda, A. J., **Horvat, C. M.** & **Hochheiser, H.** 2020 A qualitative research framework for the design of user-centered displays of explanations for machine learning model predictions in healthcare. *BMC Medical Informatics and Decision Making* **20** (1), 257; doi:10.1186/s12911-020-01276-x.

Bedwell, W. L., **Wildman, J. L.**, **DiazGranados, D.**, **Salazar, M.**, **Kramer, W. S.** & **Salas, E.** 2012 Collaboration at work: an integrative multilevel conceptualization. *Human Resource Management Review* **22** (2), 128–145; doi:10.1016/j.hrmr.2011.11.007.

Bhavnani, S. K., **John, B. E.** & **Flemming, U.** 1999 The strategic use of CAD: an empirically inspired, theory-based course. In *CHI '99: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 183–190; doi:10.1145/302979.30303.

Braga, L. H., **Farrokhyar, F.**, **Dönmez, M. İ.**, **Nelson, C. P.**, **Haid, B.**, **Herbst, K.**, **Garriboli, M.**, **Cascio, S.**, **Nieuwhof-Leppink, A.**, **Kaefer, M.**, **Bägli, D. J.**, **Kalfa, N.**, **Ching, C.**, **Fossum, M.** & **Harper, L.** 2025 Randomized controlled trials – The what, when, how and why. *Journal of Pediatric Urology* **21** (2), 397–404; doi:10.1016/j.jpurol.2024.11.021.

Bralla, J. G. 1998 *Design for Manufacturability Handbook*. McGraw Hill.

Brown, P. 2009 CAD: do computers aid the design process after all? *Intersect: The Stanford Journal of Science, Technology and Society* **2** (1).

Bubeck, S. *et al.* 2023 Sparks of artificial general intelligence: early experiments with GPT-4. https://arxiv.org/abs/2303.12712.

Camba, J. D., **Contero, M.** & **Company, P.** 2016 Parametric CAD modeling: an analysis of strategies for design reusability. *CAD Computer Aided Design* **74**, 18–31; doi:10.1016/j.cad.2016.01.003.

Celjak, R., **Horvat, N.** & **Škec, S.** 2023 Comparing collaborative CAD modelling patterns of high-performing and low-performing teams. In *Proceedings of the Design Society (Vol. 3)*, pp. 1007–1016. Cambridge University Press. doi:10.1017/pds.2023.101

Chen, L., **Song, Y.**, **Guo, J.**, **Sun, L.**, **Childs, P.** & **Yin, Y.** 2025 How generative AI supports human in conceptual design. *Design Science* **11**, e9; doi:10.1017/dsj.2025.2.

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

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. In *Proceedings of the ACM on Human-Computer Interaction 7 (CSCW1)*, pp. 1–29; Association for Computing Machinery (ACM). doi: 10.1145/3579613.

Cheng, K. & **Olechowski, A.** 2024 Analysis of collaborative assembly in multi-user computer-aided design. *Journal of Mechanical Design* **146** (3); doi:10.1115/1.4063759.

Cheng, K. & **Olechowski, A.** 2021 *Some (team) assembly required: an analysis of collaborative computer-aided design assembly. In Volume 6: 33rd International Conference on Design Theory and Methodology (DTM).* American Society of Mechanical Engineers; doi:10.1115/DETC2021-68507.

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

Chiarello, F., **Giordano, V.**, **Spada, I.**, **Barandoni, S.** & **Fantoni, G.** 2024 Future applications of generative large language models: a data-driven case study on ChatGPT. *Technovation* **133**, 103002; doi:10.1016/j.technovation.2024.103002.

Chong, L., **Lo, I.-P.**, **Rayan, J.**, **Dow, S.**, **Ahmed, F.** & **Lykourentzou, I.** 2025 Prompting for products: investigating design space exploration strategies for text-to-image generative models. *Design Science* **11**, e2; doi:10.1017/dsj.2024.51.

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; doi:10.1016/j.cad.2014.10.001.

Dell'Acqua, F., **Kogut, B.** & **Perkowski, P.** 2025 Super Mario meets AI: experimental effects of automation and skills on team performance and coordination. *Review of Economics and Statistics* **107** (4), 951–966; doi:10.1162/rest_a_01328.

Deng, Y., **Marion, T.** & **Olechowski, A.** 2022a Does synchronous collaboration improve collaborative computer-aided design output: results from a large-scale competition. In *Proceedings of the ASME 2022 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*; American Society of Mechanical Engineers (ASME). doi:10.1115/DETC2022-89731.

Deng, Y., **Mueller, M.**, **Rogers, C.** & **Olechowski, A.** 2022b The multi-user computer-aided design collaborative learning framework. *Advanced Engineering Informatics* **51**, 101446; doi:10.1016/j.aei.2021.101446.

Du, Y. *et al.* 2024 BlenderLLM: training large language models for computer-aided design with self-improvement. https://arxiv.org/abs/2412.14203.

Ege, D. N., **Øvrebø, H. H.**, **Stubberud, V.**, **Berg, M. F.**, **Elverum, C.**, **Steinert, M.** & **Vestad, H.** 2025 ChatGPT as an inventor: eliciting the strengths and weaknesses of current large language models against humans in engineering design. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* **39**; doi:10.1017/s0890060425000010.

Favi, C., **Campi, F.**, **Germani, M.** & **Mandolini, M.** 2022 Engineering knowledge formalization and proposition for informatics development towards a CAD-integrated DfX system for product design. *Advanced Engineering Informatics* **51**, 101537; doi:10.1016/j.aei.2022.101537.

Feng, L., **Yen, R.**, **You, Y.**, **Fan, M.**, **Zhao, J.** & **Lu, Z.** 2024 CoPrompt: supporting prompt sharing and referring in collaborative natural language programming. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–21. ACM; doi:10.1145/3613904.3642212.

Filippi, S. 2023 Measuring the impact of ChatGPT on fostering concept generation in innovative product design. *Electronics* **12** (16), 3535; doi:10.3390/electronics12163535.

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

Fui-Hoon Nah, F., **Zheng, R.**, **Cai, J.**, **Siau, K.** & **Chen, L.** 2023 Generative AI and ChatGPT: applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research* **25** (3), 277–304; doi:10.1080/15228053.2023.2233814.

Georganta, E. & **Ulfert, A.** 2024 Would you trust an AI team member? Team trust in human–AI teams. *Journal of Occupational and Organizational Psychology* **97** (3), 1212–1241; doi:10.1111/joop.12504.

Glikson, E. & **Woolley, A. W.** 2020 Human trust in artificial intelligence: review of empirical research. *Academy of Management Annals* **14** (2), 627–660; doi:10.5465/annals.2018.0057.

Gopsill, J., **Snider, C.**, **Hicks, B.**, **Gopsill, J.**, **Snider, C.**, **Shi, L.** & **Hicks, B.** 2016 Computer-aided design user interaction as a sensor for monitoring engineers and the engineering design process. In *DS 84: Proceedings of the DESIGN 2016 14th International Design Conference*. Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb, pp. 1707–1718.

Grassini, S., **Aasen, M. L.** & **Møgelvang, A.** 2024 Understanding university students' acceptance of ChatGPT: insights from the UTAUT2 model. *Applied Artificial Intelligence* **38** (1); doi:10.1080/08839514.2024.2371168.

Haenlein, M. & **Kaplan, A.** 2019 A brief history of artificial intelligence: on the past, present, and future of artificial intelligence. *California Management Review* **61** (4), 5–14; doi:10.1177/0008125619864925.

Han, Y., **Qiu, Z.**, **Cheng, J.** & **Lc, R.** 2024 When teams embrace AI: human collaboration strategies in generative prompting in a creative design task. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–14. ACM; doi:10.1145/3613904.3642133.

Hemmer, P., **Schemmer, M.**, **Vössing, M.** & **Kühl, N.** 2021 Human-AI complementarity in hybrid intelligence systems: a structured literature review. In *PACIS 2021 Proceedings*. Association for Information Systems (AIS).

Ilgen, D. R., **Hollenbeck, J. R.**, **Johnson, M.** & **Jundt, D.** 2005 Teams in organizations: from input-process-output models to IMOI models. *Annual Review of Psychology* **56** (1), 517–543; doi:10.1146/annurev.psych.56.091103.070250.

Jing, Y., **Wang, H.**, **Chen, X.** & **Wang, C.** 2024 What factors will affect the effectiveness of using ChatGPT to solve programming problems? A quasi-experimental study. *Humanities and Social Sciences Communications* **11** (1); doi:10.1057/s41599-024-02751-w.

Kaplan, A. D., **Kessler, T. T.**, **Brill, J. C.** & **Hancock, P. A.** 2023 Trust in artificial intelligence: meta-analytic findings. *Human Factors: The Journal of the Human Factors and Ergonomics Society* **65** (2), 337–359; doi:10.1177/00187208211013988.

Khanolkar, P. M., **Vrolijk, A.** & **Olechowski, A.** 2023 Mapping artificial intelligence-based methods to engineering design stages: a focused literature review. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM* **37**; doi:10.1017/S0890060423000203.

Kirkeby-Hinrup, A. & **Stenseke, J.** 2025 The psychology of LLM interactions: the uncanny valley and other minds. *Journal of Psychology and AI* **1** (1); doi:10.1080/29974100.2025.2457627.

Kodnongbua, M., **Jones, B.**, **Ahmad, M. B. S.**, **Kim, V.** & **Schulz, A.** 2023 ReparamCAD: zero-shot CAD re-parameterization for interactive manipulation. In *SIGGRAPH Asia 2023 Conference Papers*, pp. 1–12. ACM; doi: 10.1145/3610548.3618219.

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

Lee, M., **Frank, L.** & **IJsselsteijn, W.** 2021 Brokerbot: a cryptocurrency chatbot in the social-technical gap of trust. *Computer Supported Cooperative Work (CSCW)* **30** (1), 79–117; doi:10.1007/s10606-021-09392-6.

Li, X., **Song, Y.**, **Lou, Y.** & **Zhou, X.** 2024a CAD translator: an effective drive for text to 3D parametric computer-aided design generative modeling. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 8461–8470. ACM; doi: 10.1145/3664647.3681549.

Li, X., **Sun, Y.** & **Sha, Z.** 2025 LLM4CAD: multimodal large language models for three-dimensional computer-aided design generation. *Journal of Computing and Information Science in Engineering* **25** (2); doi:10.1115/1.4067085.

Li, Y., **Wu, B.**, **Huang, Y.** & **Luan, S.** 2024b Developing trustworthy artificial intelligence: insights from research on interpersonal, human-automation, and human-AI trust. *Frontiers in Psychology* **15**; doi:10.3389/fpsyg.2024.1382693.

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; doi: 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.** 2024 Large language models for design and manufacturing. *An MIT Exploration of Generative AI*; doi:10.21428/e4baedd9.745b62fa.

Martin, A. V. I. & **Selva, D.** 2020 Daphne: a virtual assistant for designing earth observation distributed spacecraft missions. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **13**, 30–48; doi:10.1109/JSTARS.2019.2948921.

Mathieu, J. E., **Heffner, T. S.**, **Goodwin, G. F.**, **Salas, E.** & **Cannon-Bowers, J. A.** 2000 The influence of shared mental models on team process and performance. *Journal of Applied Psychology* **85** (2), 273–283; doi:10.1037/0021-9010.85.2.273.

McComb, C., **Boatwright, P.** & **Cagan, J.** 2023 Focus and modality: defining a roadmap to future ai-human teaming in design. In *Proceedings of the Design Society* (Vol. **3**), pp. 1905–1914. Cambridge University Press; doi:10.1017/pds.2023.191.

Memmert, L., **Cvetkovic, I.** & **Bittner, E.** 2024a The more is not the Merrier: effects of prompt engineering on the quality of ideas generated by GPT-3. In *Proceedings of the 57th Hawaii International Conference on System Sciences*. University of Hawai'i at Manoa.

Memmert, L., **Mies, J.** & **Bittner, E.** 2024b Brainstorming with a generative language model: the role of creative ability and tool-support for brainstorming performance. In *Conference: 45th International Conference on Information Systems (ICIS).*

Musick, G., **O'Neill, T. A.**, **Schelble, B. G.**, **McNeese, N. J.** & **Henke, J. B.** 2021 What happens when humans believe their teammate is an AI? An investigation into humans teaming with autonomy. *Computers in Human Behavior* **122**, 106852; doi:10.1016/j.chb.2021.106852.

Norheim, J. J., **Rebentisch, E.**, **Xiao, D.**, **Draeger, L.**, **Kerbrat, A.** & **De Weck, O. L.** 2024 Challenges in applying large language models to requirements engineering tasks. *Design Science* **10**; doi:10.1017/dsj.2024.8.

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. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* **39**, e6.

O'Donnell, F. J. & **Duffy, A. H.** 2005 *Design performance*. Springer. doi:10.1007/1-84628-147-4

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

Olechowski, A., **Deng, Y.**, **DaMaren, E.**, **Verner, I.**, **Rosen, U.** & **Mueller, M.** 2022 All's not fair in CAD: an investigation of equity of contributions to collaborative cloud-based design projects. *Computer-Aided Design and Applications*, 574–583; doi:10.14733/cadaps.2023.574-583.

Pahl, G. & **Beitz, W.** 1996 *Engineering Design: A Systematic Approach*. Springer.

Passmore, J. 2017 *Wiley Blackwell Handbooks in Organizational Psychology*. Wiley.

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); doi:10.1115/1.4050734.

Phadnis, V., **Leonardo, K.A.**, **Wallace, D.R.** & **Olechowski, A.L.** 2019 An exploratory study comparing CAD tools and working styles for implementing design changes. In *Proceedings of the International Conference on Engineering Design, ICED*, pp. 1383–1392. Cambridge University Press; doi:10.1017/dsi.2019.144.

Picard, C., **Edwards, K. M.**, **Doris, A. C.**, **Man, B.**, **Giannone, G.**, **Alam, M. F.** & **Ahmed, F.** 2025 From concept to manufacturing: evaluating vision-language models for engineering design. *Artificial Intelligence Review* **58** (9), 288; doi:10.1007/s10462-025-11290-y.

Rapp, A., **Di Lodovico, C.** & **Di Caro, L.** 2025 How do people react to ChatGPT's unpredictable behavior? Anthropomorphism, uncanniness, and fear of AI: a qualitative study on individuals' perceptions and understandings of LLMs' nonsensical hallucinations. *International Journal of Human-Computer Studies* **198**, 103471; doi:10.1016/j.ijhcs.2025.103471.

Ray, P. P. 2023 ChatGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems* **3**, 121–154; doi:10.1016/j.iotcps.2023.04.003.

Robertson, J., **Ferreira, C.**, **Botha, E.** & **Oosthuizen, K.** 2024 Game changers: a generative AI prompt protocol to enhance human-AI knowledge co-construction. *Business Horizons* **67** (5), 499–510; doi:10.1016/j.bushor.2024.04.008.

Robson, C. & **McCartan, K.** 2016 *Real World Research*. Wiley.

Roy, D., **Calpin, N.**, **Cheng, K.**, **Olechowski, A.**, **Argüelles, A. P.**, **Soria Zurita, N. F.** & **Menold, J.** 2024 Designing together: exploring collaborative dynamics of multi-objective design problems in virtual environments. *Journal of Mechanical Design* **146** (3), 031702; doi:10.1115/1.4063658.

Salas, E., **Sims, D. E.** & **Shawn Burke, C.** 2005 Is there A 'big five' in teamwork? *Small Group Research* **36** (5), 555–599; doi:10.1177/1046496405277134.

Schulze Balhorn, L., **Weber, J. M.**, **Buijsman, S.**, **Hildebrandt, J. R.**, **Ziefle, M.** & **Schweidtmann, A. M.** 2024 Empirical assessment of ChatGPT's answering capabilities in natural science and engineering. *Scientific Reports* **14** (1), 4998; doi:10.1038/s41598-024-54936-7.

Shilpa, K. & **Menon, D.** 2024 Consumer's interaction with ChatGPT – a UTAUT perspective. *Science Talks* **9**, 100281; doi:10.1016/j.sctalk.2023.100281.

Shockley, R., **Herchenroether, K.**, **Xu, Z.**, **Martelaro, N.** & **McComb, C.** 2025 The impact of design representation on equal contribution in engineering design teams. In *Proceedings of the ASME 2025 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*; doi:10.1115/DETC2025-168883.

Singh, H., **Horvat, N.**, **Škec, S.** & **Becattini, N.** 2022 A longitudinal study of teamwork quality in design teams. In *Proceedings of the Design Society (Vol. 2)*, pp. 2383–2392. Cambridge University Press; doi: 10.1017/pds.2022.241.

Šklebar, J., **Martinec, T.**, **Škec, S.** & **Štorga, M.** 2025a Collaborative CAD in engineering design: a systematic literature review and a taxonomy of collaborative CAD activities. *Advanced Engineering Informatics* **74**, 104805; doi:10.2139/ssrn.5222432.

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

Šklebar, J., **Martinec, T.**, **Škec, S.** & **Štorga, M.** 2025b Evaluating ChatGPT's role in collaborative CAD task completeness. *Proceedings of the Design Society* **5**, 1585–1594; doi:10.1017/pds.2025.10172.

Song, B., **Gyory, J. T.**, **Zhang, G.**, **Soria Zurita, N. F.**, **Stump, G.**, **Martin, J.**, **Miller, S.**, **Balon, C.**, **Yukish, M.**, **McComb, C.** & **Cagan, J.** 2022a Decoding the agility of artificial intelligence-assisted human design teams. *Design Studies* **79**, 101094; doi:10.1016/j.destud.2022.101094.

Song, B., **Soria Zurita, N. F.**, **Nolte, H.**, **Singh, H.**, **Cagan, J.** & **McComb, C.** 2022b When faced with increasing complexity: the effectiveness of artificial intelligence assistance for drone design. *Journal of Mechanical Design* **144** (2); doi:10.1115/1.4051871.

Song, B., **Zhu, Q.** & **Luo, J.** 2024 Human-AI collaboration by design. In *Proceedings of the Design Society*, pp. 2247–2256; doi:10.1017/pds.2024.227.

Stone, B., **Salmon, J.**, **Eves, K.**, **Killian, M.**, **Wright, L.**, **Oldroyd, J.**, **Gorrell, S.** & **Richey, M. C.** 2017a A multi-user computer-aided design competition: experimental findings and analysis of team-member dynamics. *Journal of Computing and Information Science in Engineering* **17** (3); doi:10.1115/1.4035674.

Stone, B., **Salmon, J. L.**, **Hepworth, A. I.**, **Red, E.**, **Killian, M.**, **la, A.**, **Pedersen, A.** & **Jones, T.** 2017b Methods for determining the optimal number of simultaneous contributors for multi-user CAD parts. *Computer-Aided Design and Applications* **14** (5), 610–621; doi:10.1080/16864360.2016.1273578.

Strzelecki, A. 2024 To use or not to use ChatGPT in higher education? A study of students' acceptance and use of technology. *Interactive Learning Environments* **32** (9), 5142–5155; doi:10.1080/10494820.2023.2209881.

Sun, Y., **Li, X.** & **Sha, Z.** 2025 Large language models for computer-aided design fine tuned: dataset and experiments. *Journal of Mechanical Design* **147** (4); doi:10.1115/1.4067713.

Tang, J., **Xiao, H.**, **Li, X.**, **Wang, W.** & **Gong, Z.** 2025 ChatCAD: an MLLM-guided framework for zero-shot CAD drawing restoration. In *ICASSP 2025–2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. IEEE; doi:10.1109/icassp49660.2025.10890248.

Thielsch, M. T., **Meeßen, S. M.** & **Hertel, G.** 2018 Trust and distrust in information systems at the workplace. *PeerJ* **6**, e5483; doi:10.7717/peerj.5483.

Tholander, J. & **Jonsson, M.** 2023 Design ideation with AI – sketching, thinking and talking with generative machine learning models. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, pp. 1930–1940. ACM; doi:10.1145/3563657.3596014.

Ulfert, A.-S. & **Georganta, E.** 2020 A model of team trust in human-agent teams. In *Companion Publication of the 2020 International Conference on Multimodal Interaction*, pp. 171–176. ACM; doi:10.1145/3395035.3425959.

Ulrich, K. T. & **Eppinger, S. D.** 2015 *Product Design and Development*. McGraw-Hill Higher Education.

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**, 105031; doi:10.1016/j.compedu.2024.105031.

Vaccaro, M., **Almaatouq, A.** & **Malone, T.** 2024 When combinations of humans and AI are useful: a systematic review and meta-analysis. *Nature Human Behaviour* **8** (12), 2293–2303; doi:10.1038/s41562-024-02024-1.

Veisz, D., **Namouz, E. Z.**, **Joshi, S.** & **Summers, J. D.** 2012 Computer-aided design versus sketching: an exploratory case study. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* **26** (3), 317–335; doi:10.1017/s0890060412000170.

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press

Vo, K. H. T. 2025 "Who" designs better? A competition among human, artificial intelligence and human–AI collaboration. *Design Science* **11**, e37; doi:10.1017/dsj.2025.10029.

Vrolijk, A.-P. & **Olechowski, A.** 2025 Computer-aided design as a research instrument for engineering design: supporting multiple methods by observing artifacts, actors, and actions. *Journal of Engineering Design* **36** (12), 2151–2182; doi:10.1080/09544828.2024.2432835.

Wang, K. D., **Burkholder, E.**, **Wieman, C.**, **Salehi, S.** & **Haber, N.** 2024 Examining the potential and pitfalls of ChatGPT in science and engineering problem-solving. *Frontiers in Education* **8**; doi:10.3389/feduc.2023.1330486.

Yang, D. 2024 Human-AI interaction in the age of large language models. *Proceedings of the AAAI Symposium Series* **3** (1), 66–67; doi:10.1609/aaaiss.v3i1.31183.

Yuan, Z., **Shi, J.** & **Huang, Y.** 2024 OpenECAD: an efficient visual language model for editable 3D-CAD design. *Computers & Graphics* **124**, 104048; doi:10.1016/j.cag.2024.104048.

Zercher, D. & **Jussupow, E.** 2023 When AI joins the team: a literature review on intragroup processes and their effect on team performance in team-AI collaboration. In *CIS 2023 Research Papers*.

Zhang, G., **Raina, A.**, **Cagan, J.** & **McComb, C.** 2021a A cautionary tale about the impact of AI on human design teams. *Design Studies* **72**, 100990; doi:10.1016/j.destud.2021.100990.

Zhang, L., **Le, B.**, **Akhtar, N.**, **Lam, S.-K.** & **Ngo, T.** 2026 Large language models for computer-aided design: a survey. *ACM Computing Surveys* **58** (9), 1–39; doi:10.1145/3787499.

Zhang, R., **Duan, W.**, **Flathmann, C.**, **McNeese, N.**, **Freeman, G.** & **Williams, A.** 2023 Investigating AI teammate communication strategies and their impact in human-AI teams for effective teamwork. *Proceedings of the ACM on Human-Computer Interaction 7(CSCW2)* **7** (CSCW2), 1–31; doi:10.1145/3610072.

Zhang, Z., **Genc, Y.**, **Wang, D.**, **Ahsen, M. E.** & **Fan, X.** 2021b Effect of AI explanations on human perceptions of patient-facing AI-powered healthcare systems. *Journal of Medical Systems* **45** (6), 64; doi:10.1007/s10916-021-01743-6.

Zou, Q., **Wu, Y.**, **Liu, Z.**, **Xu, W.** & **Gao, S.** 2024 Intelligent CAD 2.0. *Visual Informatics* **8** (4), 1–12; doi:10.1016/j.visinf.2024.10.001.

https://doi.org/10.1017/dsj.2026.10073 Published online by Cambridge University Press