Title:
Advanced Frameworks for Large Language Models in CAD Education: Evaluating and Enhancing Academic Problem Solving

Authors:
Xiang Li, xli@berkeley.edu, Research Institute of Advanced Materials, CISRI Group; China Iron and Steel Research Institute Group; UC Berkeley
Ling Chen, hust_chenling@outlook.com, China Railway Siyuan Survey and Design Group
Guang Feng, hb0102fengguang@163.com, Research Institute of Advanced Materials, CISRI Group
Sara McMains, mcmains@berkeley.edu, UC Berkeley

Keywords:
Large Language Models (LLMs), Computer-Aided Design (CAD), CAD Education

DOI: 10.14733/cadconfP.2025.126-131

Introduction:

The rapid advancement of Large Language Models (LLMs) has significantly accelerated innovation in scientific research, technology development, and data analysis, while also disrupting educational paradigms across disciplines. These models excel in solving academic problems by leveraging their sophisticated reasoning capabilities and extensive knowledge base. As a result, they have become valuable tools in education, providing insights, solutions, and enabling personalized learning pathways in subjects ranging from humanities to complex scientific domains [1]. However, while LLMs demonstrate promising abilities in many areas, their application to highly specialized and interdisciplinary subjects remains underexplored.

Computer-Aided Design (CAD) stands out as a particularly challenging domain due to its integration of diverse knowledge areas such as mathematics, algorithms, data structures, and manufacturing processes. Academic problems in CAD often push the boundaries of conventional learning and test problem-solving skills, making them a rigorous benchmark for both students and LLMs. Despite the growing importance of LLMs, to our knowledge, there has been no evaluating their proficiency in addressing CAD academic problems.

This paper makes a two-fold contribution: first, it provides a systematic evaluation of LLMs on CAD academic problems across four categories: mathematical theory, coding/algorithmic skills, conceptual understanding, and comprehensive application-based challenges. Second, upon identifying weaknesses in handling comprehensive application-based challenges, we introduce simple yet powerful two-phase frameworks tailored for LLMs, aimed at enhancing their performance on these challenging tasks. The findings from this study provide a deeper understanding of the capabilities and limitations of LLMs in this domain and offer valuable insights for future CAD education strategies.

Questions and LLM Models for Evaluation:

The test questions were collected from real-world academic CAD problems aimed at graduate students,

Proceedings of CAD'25, Shenzhen, China, June 23-25, 2025, 126-131

encompassing homework and exam challenges from institutions such as UC Berkeley, University of Cincinnati, Tianjin University, and the China Iron and Steel Research Institute. These questions span a broad range of topics and necessitate advanced problem-solving skills across multiple domains. They are organized into four distinct categories, each designed to evaluate different facets of CAD knowledge and reasoning:

1. Conceptual Understanding: This section consists of 30 multiple-choice questions and 30 Q&A problems, designed to evaluate how well students comprehend representations, foundational geometric theories, and computational geometry in CAD contexts. It assesses how well students grasp core principles and concepts before applying them to problem-solving.

2. Math Problems: This section consists of 30 multiple-choice questions and 30 Q&A problems, designed to assess analytical skills and quantitative reasoning. It focuses on the mathematical aspects of geometric concepts, algorithmic complexity, and constraint solving within CAD.

3. Coding/Algorithm: This section consists of 30 multiple-choice questions and 30 Q&A problems, designed to assess hands-on coding proficiency and algorithmic thinking. It encompasses mesh generation, constraint resolution, CAD software scripting, and related data structures, testing the ability to design, implement, and optimize algorithms relevant to CAD tasks.

4. Comprehensive Application-Based Problems: This section consists of 20 Q&A problems (no multiple-choice questions), designed to evaluate a comprehensive integration of mathematical theory, programming, conceptual reasoning, and applied knowledge in complex, real-world CAD scenarios. It further examines multi-step problem-solving, paper-reading skills, and scenario interpretation.

Examples of comprehensive application-based problems include reading and understanding the C-space (configuration-space) approach to tool-path generation for computer numerical control machining (as described in [2]), replicating the algorithm, and calculating the inverse tool-offset surface from real-world mechanical parts. Another example involves analyzing a recent computational geometry-based algorithm for metallographic analysis introduced in [3], implementing the algorithm, and proposing potential improvements for computational efficiency.

Each multiple-choice question is weighted as one-third of a Q&A problem to balance the grading scale. All questions are framed to yield quantitative or qualitative answers that can be objectively graded. They are presented exclusively in text form without figures or charts, ensuring compatibility with language-only models.

For evaluation, five state-of-the-art LLMs were employed: **OpenAI o1, OpenAI 4o, Anthropic Claude 3.5 Sonnet, Deepseek R1-Lite, and Alibaba Cloud Qwen2.5**. These models, selected for their diverse architectures and performance characteristics, provide a robust assessment of current LLM capabilities in solving academic CAD problems across the four categories.

Frameworks for Effective LLM Problem-Solving:

To address the challenges in solving academic problems in CAD and enhance the problem-solving capabilities of LLMs, we employed three distinct approaches: the naive approach, the knowledge refinement framework, and the Chain-of-Thought (CoT) framework (Fig. 1). While the naive approach serves as a baseline, the two advanced frameworks were to improve LLM performance, particularly for comprehensive, multi-domain problems.

In the ***naive approach***, the raw question and associated reference materials, such as academic papers or book chapters, are directly provided to the LLM. The model processes these inputs using its internal reasoning capabilities to generate an answer. However, this method often struggles with CAD problems due to the dense and complex nature of reference materials. CAD-related sources frequently contain intricate mathematical derivations, domain-specific terminology, and theoretical formulations that are difficult for LLMs to parse and understand effectively. Consequently, the naive approach often leads to suboptimal performance, especially on problems that require interdisciplinary integration and precise

Proceedings of CAD'25, Shenzhen, China, June 23-25, 2025, 126-131

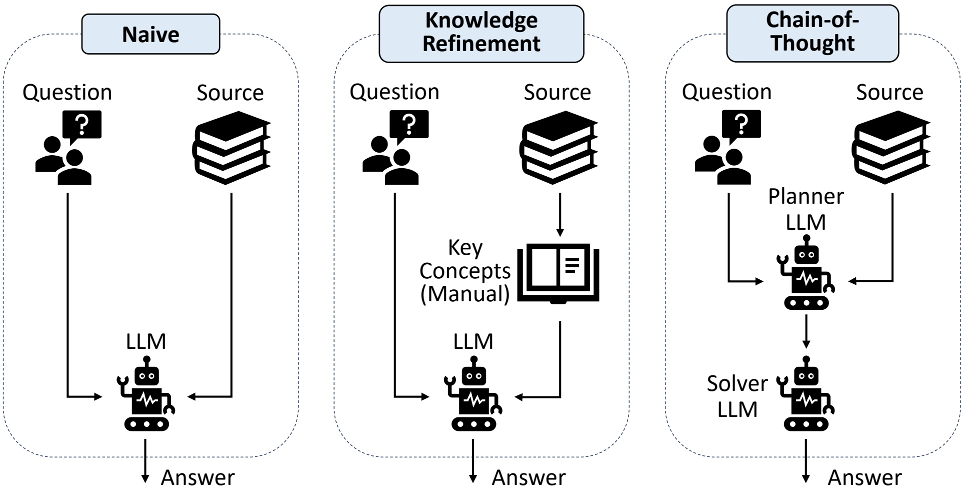

Fig. 1: Framework structures for solving CAD academic problems with LLMs: naive, knowledge refinement, and Chain-of-Thought.

comprehension of technical details.

The ***knowledge refinement framework*** mitigates the limitations of the naive approach by preprocessing the input to make it more focused and digestible for the LLM. Instead of directly feeding raw references, we manually extract and condense the most relevant definitions and theoretical insights from the source materials. These refined inputs are then provided alongside the original question, enabling the LLM to focus on essential knowledge without being overwhelmed by extraneous details. By simplifying verbose and technical reference materials, the knowledge refinement framework improves the model's ability to reason through complex problems and deliver accurate solutions, even for challenging CAD-related tasks.

The ***Chain-of-Thought (CoT) framework*** adopts the concept of CoT prompting, which emphasizes breaking down complex problems into smaller, logically connected reasoning steps to enhance problem-solving performance [4]. Building on this concept, our framework introduces a structured, multi-phase process specifically tailored for solving comprehensive problems. This framework consists of two distinct phases: **(1) Phase 1: Planning**, where a "Planner LLM" analyzes the raw question and references to design a logical, step-by-step plan for solving the problem. This phase decomposes the problem into manageable components, ensuring clarity and direction in the problem-solving process. The effectiveness of Chain-of-Thought reasoning has been widely demonstrated in improving LLM performance, particularly for tasks requiring multi-step reasoning and logical decomposition. **(2) Phase 2: Solving**, where the plan generated by the Planner LLM is passed to a "Solver LLM," which executes the outlined steps to generate the final solution. By separating the planning and execution processes, this framework leverages the complementary strengths of both models, with strategic reasoning in the planner and precise implementation in the solver.

The two advanced frameworks (knowledge refinement and Chain-of-Thought) were applied to address the shortcomings of the naive approach when solving comprehensive application-based problems, which require a seamless integration of mathematical reasoning, programming, conceptual understanding, and practical application. The refined inputs provided by the knowledge refinement framework and the structured reasoning in the CoT framework offered significant enhancements to LLMs' ability to effectively tackle challenges that were otherwise difficult to solve, improving their problem-solving capabilities across complex academic tasks.

Results:

Fig. 2 illustrates the evaluation results of the naive approach applied to academic problems across four categories: conceptual understanding, math, coding/algorithm, and comprehensive application-based problems. These results were obtained from tests conducted on both LLMs and a baseline of 5 graduate students majoring in CAD-related disciplines. To simulate an authentic test situation, these graduate students were provided with all the required background knowledge one week in advance and given one week to study before the assessment. The findings indicate that, even with the naive approach, LLMs outperform graduate students in conceptual understanding, math, and coding/algorithm problems, achieving an average score of 96.5 compared to 73.7 points for students. This demonstrates the strong foundational and technical capabilities of LLMs in processing and reasoning within these domains.

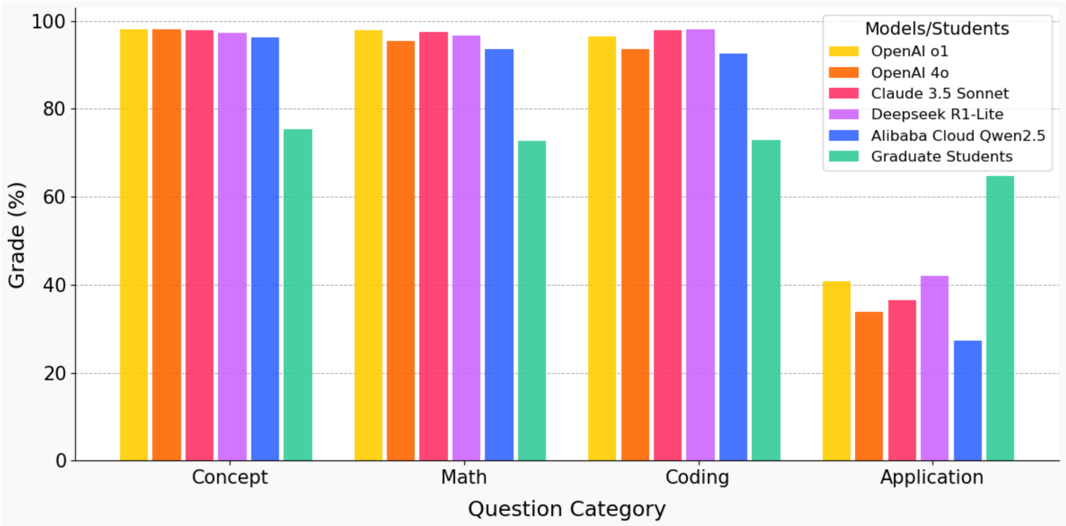

Fig. 2: Performance on CAD academic problems across LLMs and graduate students.

However, for comprehensive application-based problems, LLMs perform poorly when using the naive approach, with an average score of 36.1 points, compared to 64.8 points achieved by graduate students. These results suggest that the naive approach is insufficient for solving problems that require a seamless integration of reasoning, programming, and real-world application. In contrast, the use of our frameworks, namely the knowledge refinement framework and the Chain-of-Thought (CoT) framework, results in significantly improved performance (as shown in Table 1). The top three LLMs (OpenAI o1, Anthropic Claude 3.5 Sonnet, and Deepseek R1-Lite) achieve average scores of 94.1 and 96.5 points using the knowledge refinement and CoT frameworks, respectively. These results highlight the ability of LLMs to logically understand and solve complex problems when guided by refined inputs or structured reasoning processes.

Although the Chain-of-Thought framework achieves marginally higher scores in our evaluations, each approach offers distinct advantages and disadvantages. The knowledge refinement framework requires researchers to manually extract and define critical terminologies from technical literature, a process accessible even to those with limited expertise in LLMs. In contrast, the CoT framework employs a specialized planner LLM to perform this task automatically, but it demands a higher level of understanding in selecting high-performance LLMs and in prompt engineering to guide multi-step reasoning effectively.

Framework	o1	4o	Claude3.5	R1-Lite	Qwen2.5
Naive Approach	40.7	33.8	36.6	42.1	27.2
Knowledge Refinement	92.5	70.7	93.8	95.9	53.5
Chain-of-Thought	95.1	84.6	97.3	97.1	59.3

Table 1: Scores of different LLM models on comprehensive application-based CAD problems across the naive approach, knowledge refinement framework, and CoT framework. The CoT framework results were obtained by using OpenAI o1 as the planner and these LLMs as solvers.

Moreover, for researchers with sufficient LLM background knowledge, the CoT framework is expected to provide superior performance because it not only identifies critical terminologies but also supports deeper analysis and synthesis of the remaining background material.

Our results indicate that all tested LLMs provide better results when guided by advanced frameworks compared to the naive approach on comprehensive application-based CAD problems. However, since these models are developed by different organizations, not every LLM demonstrates uniformly good overall performance (for example, Qwen2.5 demonstrates relatively lower performance). Therefore, the careful selection of a high-performance LLM is crucial for achieving optimal performance across all tasks. Given the rapid evolution of LLM technology, continuous evaluation and appropriate model selection remain essential for maintaining and enhancing performance across diverse applications.

These findings suggest that the naive approach is effective for solving most CAD academic problems, including conceptual, mathematical, and coding challenges. However, it is not suitable for handling complex application-based problems. The two advanced frameworks, by contrast, demonstrate that with moderate effort in knowledge extraction or structured reasoning, LLMs can overcome these limitations and deliver robust solutions for comprehensive challenges.

Implications for CAD Education:

The integration of LLMs into CAD education has significant implications. By leveraging frameworks such as knowledge refinement and CoT, students and educators can solve complex academic problems more efficiently, transforming how CAD is taught and learned. However, the ease with which students can rely on LLMs to solve academic problems raises critical questions about the role of education. It becomes essential to reconsider teaching methods to ensure that students develop a deep understanding of CAD concepts, rather than relying solely on LLM outputs for completing problem sets. Balancing the use of LLMs as tools for enhancing learning while maintaining rigorous academic engagement is a challenge that educators must address to avoid superficial learning and encourage meaningful intellectual growth.

Conclusions:

Our systematic evaluation reveals that LLMs excel at foundational CAD tasks, encompassing conceptual, mathematical, and coding components, yet encounter notable difficulties when confronted with comprehensive application-based problems requiring an integration of theory, implementation, and context. We address these challenges by introducing the knowledge refinement and Chain-of-Thought (CoT) frameworks, which significantly elevate LLM performance through targeted content preparation and multi-phase reasoning. These findings not only highlight the unrealized potential of LLMs as powerful tools in CAD education but also underscore the importance of curricular designs and pedagogical policies that balance automated problem-solving with rigorous conceptual understanding.

Xiang Li, https://orcid.org/0000-0001-5936-8457

Ling Chen, https://orcid.org/0009-0004-1086-8222
Guang Feng, https://orcid.org/0000-0002-1500-2282
Sara McMains, https://orcid.org/0000-0002-7152-9409

References:

[1] Kasneci, E.; Sessler, K.; Küchemann, S.; et al.: ChatGPT for good? On opportunities and challenges of large language models for education, Learning and Individual Differences, 103, 2023, 102274. https://doi.org/10.1016/j.lindif.2023.102274

[2] Choi, B.K.; Kim, D.H.; Jerard, R.B.: C-space approach to tool-path generation for die and mould machining, Computer-Aided Design, 29(9), 1997, 657–669. https://doi.org/10.1016/S0010-4485(97)00012-2

[3] Li, X.; Cui, L.; Li, J.; et al.: Automation of intercept method for grain size measurement: A topological skeleton approach, Materials & Design, 224, 2022, 111358. https://doi.org/10.1016/j.matdes.2022.111358

[4] Wei, J.; Wang, X.; Schuurmans, D.; et al.: Chain-of-thought prompting elicits reasoning in large language models, Advances in Neural Information Processing Systems, 35, 2022, 24824–24837.