

AN EXPERIMENTAL STUDY OF THE MECHANICAL ENGINEERING KNOWLEDGE OF CHATGPT AND GEMINI IN HIGHER EDUCATION

Michael Wendland

Dr., Department of Mechanical Engineering, Westphalian University of Applied Science,
Bocholt (Germany)

Abstract

This paper presents an experimental study benchmarking the mechanical engineering (ME) knowledge of advanced Large Language Models (LLMs), specifically ChatGPT 5o and Gemini 3.0, within the context of higher education. Utilizing a methodical seven-stage comprehension framework, this study evaluates the models' abilities to process skills ranging from basic dimensioning to complex spatial reasoning and mechanical design processes. The results indicate that while both models excel at reproducing standardized knowledge – scoring 80% (Gemini) and 78% (ChatGPT) overall – they exhibit significant limitations in spatial awareness and the interpretation of complex technical drawings. Specifically, both LLMs failed in tasks requiring the generation of projections and the estimation of dimensions for detailed manufacturing drawings. The study concludes that while LLMs are powerful tools for general ME inquiries, their tendency to act as "skilled deceivers" by presenting incorrect technical data as valid solutions necessitates that students still need strong foundational expertise to verify AI-generated content.

Keywords: Mechanical engineering, LLM, engineering, benchmarking, engineering education.

1. Introduction and testing methodology

Since the introduction of AI and the corresponding Large Language Models (LLMs) in 2022, students as well as researchers are exploring their limits and expanding their applications (Wang et al. 2025). The current models seem capable of developing an extensive understanding of the mechanical engineering (ME) domain (Yavartanoo et al. 2025). They seem to be able to analyze and understand complex technical drawings in 2D as well as simple 3D models (Wendland 2025). This recent development led to the idea of methodically testing their abilities and finding out which limits currently exist. Due to their current performance (Nelson et al. 2023), ChatGPT 5o and Gemini 3.0 were selected.

To thoroughly test the capabilities of a LLM, a methodical testing framework is developed, including seven stages (see figure 1, left side). All technical drawings can be found in Kurz et al. 2014.

Figure 1. Seven Stage Comprehension Model (left) and exemplary answers of stage 2 (right side), (drawings Kurz et al. 2014).

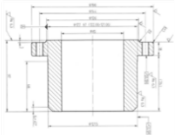
Seven Stages of ME Understanding

Rubric Levels 0–3

- 1 General Skills: Dimensions, Measurements, Tolerances
- 2 Rough Sketches & Drawing Types
- 3 Analyzing Projections of Existing 3D Parts
- 4 Generating Isometric Projections from a Given 3D Model
- 5 Interpreting Complex Technical Drawings
- 6 Functional Understanding of a Mechanical Design
- 7 Comprehension of the Mechanical Design Process

Rubric Levels 0–3

2.1 What kind of drawing type is this? Explain in detail what you see! (3/3)

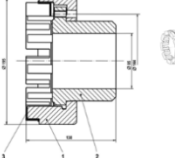


This image you uploaded is a Detailed Drawing (specifically a Cross-Section View). It focuses on a single component and provides the specific technical data needed to manufacture it. Here is a breakdown of the elements that identify this drawing type:

Key Features in the Image

- **Cross-Sectioning:** The diagonal "hatch" lines (at 45-degree angles) indicate that the part has been "cut in half" so we can see the internal geometry and wall thickness.
- **Dimensional Tolerances:** You can see specific size constraints, such as the diameter marked $\varnothing 122 \pm 0.1$, which includes a specific "IT" tolerance for a shaft or housing.

2.2 What kind of drawing is this? Explain in detail what you see! (3/3)



Based on the image you provided, the document on the left is an Assembly Drawing, containing a Section View, while the image on the right is an Exploded View.

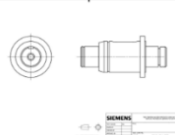
Together, these represent the two main ways engineers communicate how a complex mechanism fits together.

1. The Assembly Drawing (Left Image)

This is a technical representation of multiple parts joined together to form a single functional unit.

- **Section View:** The drawing uses a "half-angle" perspective (indicated by the diagonal hatching break). This allows you to see the internal relationship between the different components, such as how the inner shaft (Part 2) fits inside the outer housing (Part 1).
- **Part Numbering (Callouts):** The numbers 1, 2, and 3 with leader lines are used to identify individual components. In a professional setting, these would correspond to a Bill of Materials (BOM) or Parts List located on the side of the drawing.

2.3 How many measurements are needed for this specific part? (1/3)



Feature Category	Estimated Count	Specific Examples from the Image
Dimensions (Size)	6 - 8	The outer flange, the central shaft, the inner bore, and the bolt hole diameters.
Lengths (Depth)	5 - 7	The total length of the part, the thickness of the flange, and the depth of the internal step.
Location (Patterns)	3 - 4	The Bolt Circle Diameter (BCD) which tells the CNC machine where to drill the ring of holes.
Detail Features	4 - 6	Small radius curves (fillets) at the corners and the "chamfers" (slanted edges) on the ends of the shaft.

Each stage is designed to represent a different skillset and comprehension level, similar to what is expected from undergraduate students. The prompt responses are evaluated in four levels on a fixed rubric, ranging from 0 = no knowledge; 1 = basic knowledge of 1st/2nd semester students with some mistakes; 2 = advanced knowledge with minor mistakes; 3 = expert knowledge with no mistakes. These criteria allow a final comparison of the state of readiness for application in the ME domain. Exemplary questions and test materials of stage 2 are shown on the right side of figure 1. For each stage, the framework questions will be discussed in detail and a summary of the LLMs ability will be given.

2. Testing framework: Seven stages of ME comprehension

Starting with the **first stage**, general skills like applying basic dimensions and analyzing measurements as well as checking mechanical tolerances are tested. The test questions are weighted equally and are worth 3 points each, the results are displayed in parenthesis (points achieved/maximum points available):

- 1.1 How many drawing types are there in general in mechanical engineering? (3/3)
- 1.2 Why does the design often start with a rough sketch instead of a drawing? (3/3)
- 1.3 How many dimensions are necessary to fully describe this design? (L-bracket) (3/3)
- 1.4 How many different types of tolerances are there for dimensions? (3/3)

The results: all drawing types (1.1) are correctly mentioned, e.g., a detail drawing for manufacturing, an exploding view drawing for assembly and an assembly drawing for the overall view of all components installed. Next, the answer why a design often starts with a sketch in engineering (1.2) is answered adequately, e.g. “in engineering and industrial design, starting with a rough sketch instead of a formal technical drawing is a deliberate strategy. While a technical drawing is meant to tell the truth (precision), a sketch is meant to explore a possibility (ideation)”. Moving on to a more complex question (1.3), an L-bracket is presented in an isometric perspective and the LLM is tasked to assess how many dimensions are needed to fully represent the design. Both models answer correctly by stating five necessary dimensions (height, length, width, vertical leg thickness and horizontal base thickness). Finally, the knowledge about tolerance for dimensions is tested (1.4). Both LLMs score maximum points by listing three tolerance types, e.g., size tolerances, geometric tolerances and fits + allowances. Stage 1 is successfully completed by both LLMs and shows that the questions type just query existing standard knowledge – something that modern LLMs excel at, thus both score 12/12 points.

Next up, the **second stage** focuses more on the interpretation of drawings and requires a deeper skillset. The three corresponding questions and their images are shown in figure 1 (right side), and reveal interesting results:

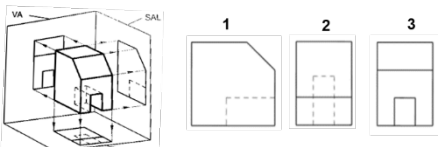
- 2.1 What kind of drawing type is this? Explain in detail what you see! (3/3)
- 2.2 What kind of drawing type is this? Explain in detail what you see! (3/3)
- 2.3 How many dimensions are needed for this specific part? (1/3)

In the first test, a cross-sectional view of a single component is uploaded (2.1), acting as a detail drawing for manufacturing. Both LLMs correctly identify it as such, and Gemini interprets the information from the drawing mostly correctly by stating the dimensions and tolerances depicted in the drawing. This means that current LLMs can read and understand images and extract as well as interpret information from simple detail drawings. Following this, a different type of drawing in the form of an assembly drawing is uploaded (2.2). The differences to the detail drawing are explained correctly and the ideas of the exploding view (showing the spatial relationship) and the assembly drawing (section views with assembly dimensions, orientation and part numbering for the bill of materials) are referenced. To increase the difficulty, a very complex manufacturing drawing is presented, which usually requires students 2-3 hours to complete. The LLMs are asked how many dimensions/measurements are thought to be needed. Here, their inadequate ability to process more detailed images step-by-step and apply/link their existing knowledge becomes apparent. Both LLMs fail and greatly underestimate the number of dimensions (ChatGPT: 20-22 / Gemini: 18-22; Solution: 35+ dimensions). It becomes obvious that the LLMs try to score with very general information instead of concise details related to the uploaded image, a method that seems very humanlike. Gemini scores 7/9, ChatGPT 6/9 points.

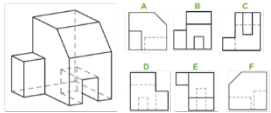
The **third and fourth stage** analyze the ability to read, understand and create projections from given 3D images. To distinguish between the skillsets, the third stage gives only quizzes (3.1-3.3, figure 2, left side), which require the LLM to recognize and sort the views according to a given projection method, while the fourth stage demands creating these projections (figure 2, right side).

Figure 2. Stage 3 (left side) and Stage 4 (right side) of projection quizzes (drawings Kurz et al. 2014).

3.1 The following part and the projections 1, 2 and 3 are given. According to projection method 1, which views correspond to 1, 2 and 3? (3/3)

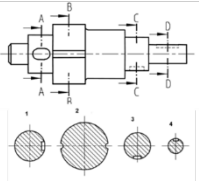


3.2 Which views (A-F) are necessary for projection method 1? (1/3)



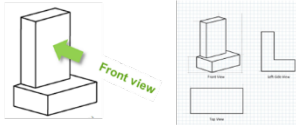
View Direction	Position on Drawing	Corresponding Letter
Front	Center	E
Top	Below Front View	B
Left Side	Right of Front View	A

3.3 Which cross-sections (A to D) are corresponding to the drawings 1-4? (3/3)

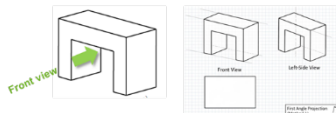


Cross-Section	Drawing Number	Key Identifying Feature
A-A	1	Central oval slot
B-B	2	Largest diameter with outer grooves
C-C	3	Bottom rectangular pocket
D-D	4	Smallest diameter with top keyway

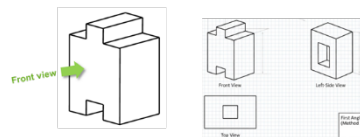
4.1 Generate the corresponding views in a drawing according to projection method 1 and the given isometric perspective of the part! (1/3)



4.2 Generate the corresponding views in a drawing according to projection method 1 and the given isometric perspective of the part! (1/3)



4.3 Generate the corresponding views in a drawing according to projection method 1 and the given isometric perspective of the part! (0/3)



While the first question (3.1) is answered correctly, the more complex situation and choices of the second question (3.2) seem to disrupt both LLMs, resulting in low scores. The last question of stage 3 involves assigning given cross-sections to the corresponding drawings (3.3), and the LLMs both have no problem in solving the task correctly. Next, complexity is increased by switching from analyzing to creating projections in stage 4. **This stage produced by far the weakest results from both LLMs.** Although the given part designs were geometrically simple, they were not able to generate the correct corresponding views for projection method 1 and failed in all three questions (see figure 2, right side). One hypothesis to explain this behavior is that the LLMs need to switch to their image generation module, and while text analysis and generation is excellent in current models, there seems to be a loss of information, especially regarding the spatial arrangement (tested in January 2026 with Nano Banana for Gemini 3.0 and DALL-E for ChatGPT 5o). Stage 3 quizzes were completed by the LLMs with 7/9 points, while stage 4 only awarded 2/9 points. It becomes apparent that spatial awareness and spatial arrangement is a missing skill in current LLMs.

Stage 5 and 6 ask to interpret technical drawings, a skill where most students seem to have difficulty at first. The questions get progressively more complex (see figure 3):

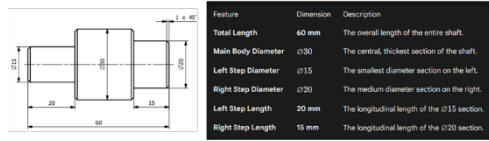
- 5.1 Interpret this drawing! What mechanical part do you see? (3/3)
- 5.2 Interpret this manufacturing drawing? What kind of part is shown here? (3/3)
- 5.3 What is shown in this assembly drawing? (1/3)
- 6.1 Which parts do you see in this mechanical drawing? (3/3)
- 6.2 Which mechanical parts are shown in this exploded view? (3/3)
- 6.3 Which mechanical design element is shown here and what is its purpose? (3/3)

Results of **stage 5** indicate that simple mechanical parts and detail drawings can be easily understood by the LLMs, with regards to their dimensions (5.1) and their corresponding function (5.2). Once more parts are added and a complex assembly drawing with shafts, gears and keys is introduced, the interpretations get inaccurate and lead to guessing. One possible reason could be that the missing spatial awareness is hindering the LLMs in making the right conclusions. Stage 5 resulted in 7/9 points for both LLMs.

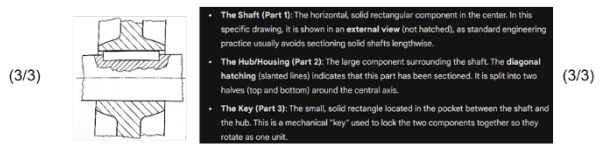
In contrast to this, **stage 6** showed their full potential by scoring the maximum 9/9 points. For question 6.1, both LLMs correctly identify the shaft, the hub/housing and a feather key which connects both parts. Since this is a commonly used connection method in mechanical engineering, the widespread information in textbooks and other sources allow the LLMs to successfully interpret this drawing. Next, an exploding view of multiple parts joint together is presented (6.2). Again, the information given is accurate and results in full scores. The last question shows a double direction thrust ball bearing (6.3), which is designed to support axial loads from two directions. Both LLMs correctly identify it as such and therefore the overall stage 6 score results in 9/9 points.

Figure 3. Stage 5 (left side) and Stage 6 (right side) for drawing knowledge (drawings Kurz et al. 2014).

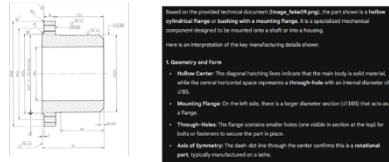
5.1 Interpret this drawing! What mechanical part do you see?



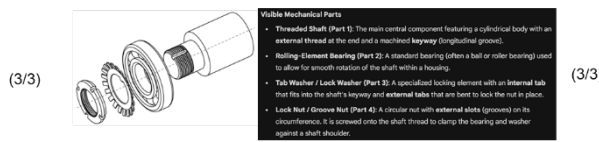
6.1 Which parts do you see in this mechanical drawing?



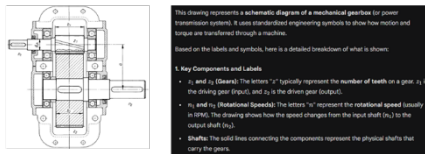
5.2 Interpret this manufacturing drawing? What kind of part is shown here?



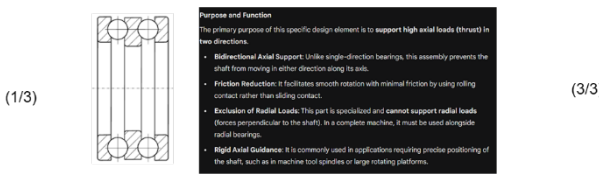
6.2 Which mechanical parts are shown in this exploded view?



5.3 What is shown in this assembly drawing?



6.3 Which mechanical design element is shown here and what is its purpose?



The last **stage 7** tests the comprehension of the mechanical design process with three questions:

7.1 What are major steps in a mechanical design process? List all steps and give examples!

7.2 Which methods do you recommend ensuring a successful planning phase? And what is the main difficulty in the planning phase?

7.3 Which product is described in this functional structure plan?

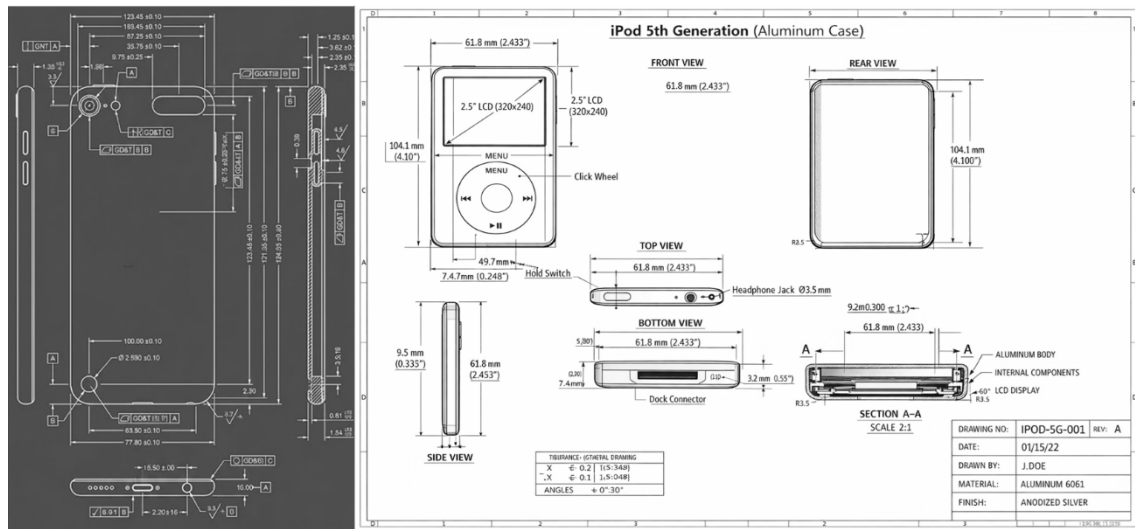
As expected, no problems arise for the LLMs due to their foundational knowledge of mechanical engineering and design processes: they offer the right steps (7.1 – from problem definition via specification to concept design, prototyping, testing and final documentation), the fitting methods (7.2 – Quality Function Deployment, Requirement specifications, Design Failure Mode and Effects Analysis, etc.) and more interestingly, they also list the main difficulties while designing (7.2 - Low Info vs. Cost Paradox, Scope Creep and Discipline Silos). Even when prompted with an unknown functional structure (7.3), both LLMs correctly identified the shown product as a Drip Coffee Maker. This shows a very broad ability to interpret and provide information in context of a mechanical design process. Therefore, the final result for stage 7 is 9/9 points.

3. Summary and discussion

Both LLMs score evenly across the 7-stage testing framework and achieve 53/66 points (Gemini, =80%) or a close 52/66 points (ChatGPT, =78%). **In general, the more precise a design question or task is, the better the results of a LLM are.** The more knowledge in form of texts, graphs, images, textbooks, blogs, forums is available, the better they reproduce this existing knowledge. **This means that common ME problems, machine elements and other standardized design challenges are well within the capability realm of modern LLMs.** If it has been discussed somewhere before, the LLMs will most likely present potential solutions.

On the other hand, the less information is available for a certain ME item, the more the LLM diverts from the available data points and the more vague the results become. **Challenged to understand and interpret complex assembly and manufacturing drawings, both LLMs failed**, but presented their wrong answers as true. **This is the “dark side” in current LLMs, they seem to act by design under the concept of the “skilled deceiver”.** Being excellent at presenting made-up information as valid solutions, they appear to offer good knowledge at first glance. It is that often only with appropriate expert knowledge that the user can notice the incoherences. This poses the biggest problems for students, because in complex design situations, the given LLM solutions often conceive the novice users, while being unusable in reality.

Figure 4. Results of Gemini (left) and ChatGPT (right) for the prompt: “Create a mechanical manufacturing drawing for an iPod 5. Gen with all the necessary dimensions, include multiple projection”.



One example of this can be seen in figure 4, where Gemini and ChatGPT were asked to produce a manufacturing drawing for an iPod Gen. 5 case made of aluminum. It becomes obvious that the LLMs try to score with very general information instead of concise details related to the complex question or an uploaded image. **The results shown reflect in no way a feasible manufacturing drawing and would result in a complete fail in an exam.** Furthermore, once spatial awareness or arrangement is necessary to solve a task, current LLMs fail completely. One hypothesis of the author that could explain this behavior is that the LLMs need to switch to their image generation module, and while text generation is excellent in current models, there seems to be a loss of information, especially regarding the spatial arrangement (tested in January 2026 with Nano Banana for Gemini 3.0 and DALL-E for ChatGPT 5o). **So, whenever spatial reasoning is required, the current LLMs do not generate adequate results and have to be used with utmost care.**

As with regards to **teaching in Higher Education**, it becomes apparent that one major skill for the students must be to be able to verify the results of LLMs when using them. This is even more applicable when faced with complex ME design situations, but fact-checking is only possible with a minimum of design knowledge and know-by-heart expertise about the ground-truths of engineering: materials, manufacturing, design rules. Although the impressive performance of both Gemini and ChatGPT in this testing framework indicate that they will play a large role in modern teaching and ME, the **students must still learn the basic skillsets and ground truths before applying AI.** A common question, which arises at this point is “how to motivate the students to still learn the basic knowledge?”, even though that part is easily covered by LLMs. The uncomfortable answer might be that it is the only way to decipher the “Skilled Deceivers” in the future.

References

- Kurz, U., Wittel, H., & Jannasch, D. (2014). *Böttcher/Forberg Technisches Zeichnen*. Wiesbaden: Springer Vieweg.
- Nelson, M. D., Goenner, B. L., & Gale, B. K. (2023). Utilizing ChatGPT to Assist CAD Design for Microfluidic Devices. *Lab on a Chip*, 17, 3778-3784.
- Wang, R., Yuan, Y., Sun, S., & Bian, J. (2025). *Text-to-CAD Generation Through Infusing Visual Feedback in Large Language Models*. arXiv. Retrieved April 7, 2025, from <https://arxiv.org/abs/2501.19054>
- Wendland, M. (2025). Exploring the potentials and limitations of AI-based CAD design for Mechanical Engineering. *International Conference on Education and New Developments*, Budapest, Hungary.
- Yavartanoo, M., Hong, S., Neshatavar, R., & Lee, K. M. (2024). *Text2CAD: Text to 3D CAD Generation via Technical Drawings*. arXiv. Retrieved April 7, 2025, from <https://arxiv.org/abs/2411.06206>