

Evaluation of AI Tutor Tools for Question-Answering and Error Handling

Based on Output Quality, Usability, and Output Trustworthiness

1. Introduction

This report evaluates two AI-based tutor tools, DeepSeek and OpenAI, in the context of question-answering (QA) and error-handling support during software development and learning processes. The evaluation is conducted using three quality areas: Output Quality, Usability, and Output Trustworthiness. These criteria are essential for determining how effectively AI tutors assist users in understanding concepts, resolving errors, and improving problem-solving skills.

2. DeepSeek Evaluation

2.1 Output Quality

DeepSeek provides technically detailed and logically structured answers, especially for programming-related questions.

- Explanations are often precise and focused on implementation details.
- Error-handling responses usually identify the root cause of the problem.
- However, explanations may sometimes lack clarity for beginners.

Overall, DeepSeek delivers high-quality technical outputs but assumes a certain level of prior knowledge.

2.2 Usability

DeepSeek's usability is moderate.

- The interface is functional but minimal.
- Responses are direct but sometimes dense.
- Less guidance is provided for step-by-step learning.

Usability is acceptable for experienced users but less friendly for beginners.

2.3 Output Trustworthiness

DeepSeek generally provides reliable and technically accurate answers.

- Solutions are logically consistent.

- Some responses require validation due to limited contextual awareness.
- Best used with developer verification.

Output trustworthiness is moderate to high.

3. OpenAI Evaluation

3.1 Output Quality

OpenAI excels in producing clear, structured, and well-explained answers.

- Responses are detailed and easy to understand.
- Error-handling explanations include reasoning and alternative solutions.
- Suitable for both beginners and advanced users.

Output quality is consistently high across different question types.

3.2 Usability

OpenAI offers high usability.

- Natural language interaction is intuitive.
- Step-by-step explanations improve comprehension.
- Adaptive tone and examples enhance learning.

Usability is very high, especially for tutoring and learning scenarios.

3.3 Output Trustworthiness

OpenAI generally provides trustworthy answers.

- Explanations are logically sound.
- Edge cases are often mentioned.
- Some outputs still require validation for critical applications.

Overall trustworthiness is high when combined with user review.

4. Comparison of DeepSeek and OpenAI

Output Quality:

- DeepSeek is strong in technical precision.
- OpenAI balances precision with clarity and explanation.

Usability:

- OpenAI is more user-friendly and suitable for tutoring.
- DeepSeek is better suited for experienced users.

Output Trustworthiness:

- Both tools provide reliable answers.
- OpenAI offers more contextual validation and guidance.

5. Conclusion

Both DeepSeek and OpenAI are effective AI tutor tools for question-answering and error-handling, but they serve different user needs.

DeepSeek is more suitable for advanced users seeking concise and technical explanations.

OpenAI is better suited for educational purposes, error explanation, and guided problem-solving.

When evaluated across output quality, usability, and output trustworthiness, OpenAI demonstrates stronger overall performance as an AI tutor, while DeepSeek excels in technical depth.