

# Abstract

The results of the project “Testing AI as a Rational Theologian” carried out from 2023 to 2024 show that leading chatbots powered by large language models, such as ChatGPT based on GPT-4o and Gemini based on Gemini 1.5, possess the core skills of a rational theologian. Testing, based on Christian theological materials, has revealed that these chatbots accurately interpret biblical passages—providing context, conveying the correct meaning (including according to the theory of the four senses of Scripture), and solving logical puzzles related to difficult passages. The chatbots can also clarify theological problems, critically assess theological reasoning and formulate such reasoning, as well as logically link biblical utterances to theological statements. The chatbots are capable of discussing natural theology and can provide theological guidance. Finally, they have proven that they can apply the concept of a rational theologian by creating their own criteria for such a figure, and that they are able to evaluate the theological skills of other language models. Some limitations have been detected, such as hallucination when quoting theological sources and mistakes in logical inference. Moreover, the chatbots gave different answers when asked a practical question within the framework of the Islamic tradition, according to the Hanafi school. A simplified benchmark for theological skills assessment enabled us to measure their performance: ChatGPT achieved an overall score of 86%, while Gemini scored 84%. This report presents the study’s methodology and tools (tasks and assessment criteria), which can also be used to evaluate other chatbots. It includes a discussion of the results and an assessment of the theological skills of both chatbots. An appendix contains the full transcripts of conversations with ChatGPT and Gemini analysed in the study.

**Keywords:** chatbots, large language models, artificial intelligence, theology, religion, rationality, benchmarks, testing, performance