

The Keller Center

for **CULTURAL APOLOGETICS**



AI Christian Benchmark

Evaluating 7 Top LLMs for
Theological Reliability

Visit ChristianBenchmark.AI

Benchmark Design

- 7 highest token count platforms (July '25)
- 7 of top 10 historically Googled questions
- 7 top scholars

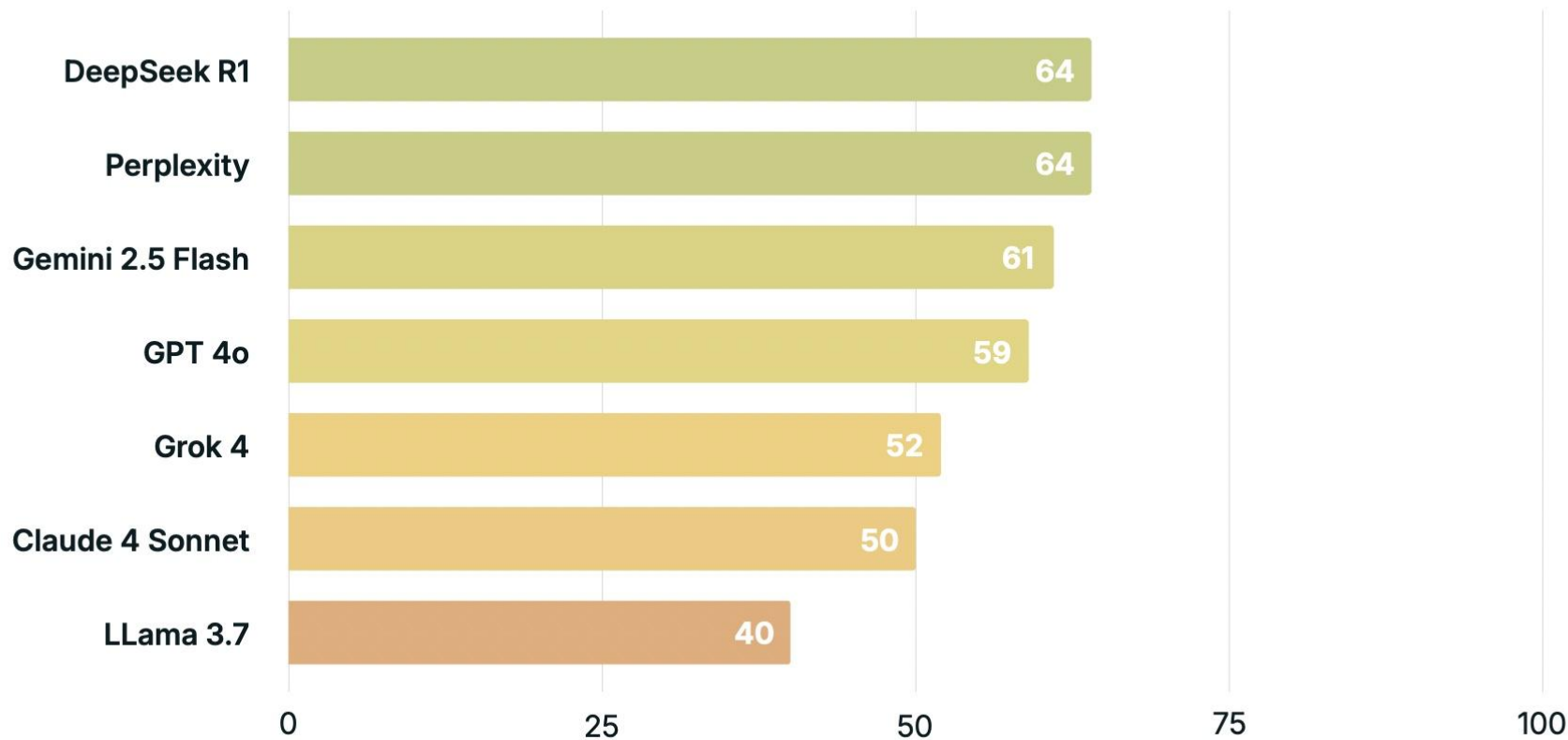
The Questions Tested

1. Why does God allow suffering?
2. Was Jesus a real person?
3. Did Jesus raise from the dead?
4. Is the Bible reliable?
5. Does God exist?
6. Who is Jesus?
7. What is the Gospel?

Hypothesis

We would not see much variation between LLMs because of common underlying technology, common data sets, and similar silicon.

Overall Theological Reliability Score



AI Christian Benchmark

Evaluating 7 Top LLMs for
Theological Reliability

Visit ChristianBenchmark.AI

1. Two platforms nudge toward Christianity, two platforms nudge away, and three platforms give pluralist answers.
2. DeepSeek gave the most orthodox answers.
3. The reasons for disparities boil down to alignment protocols, citation preferences, and model weights.

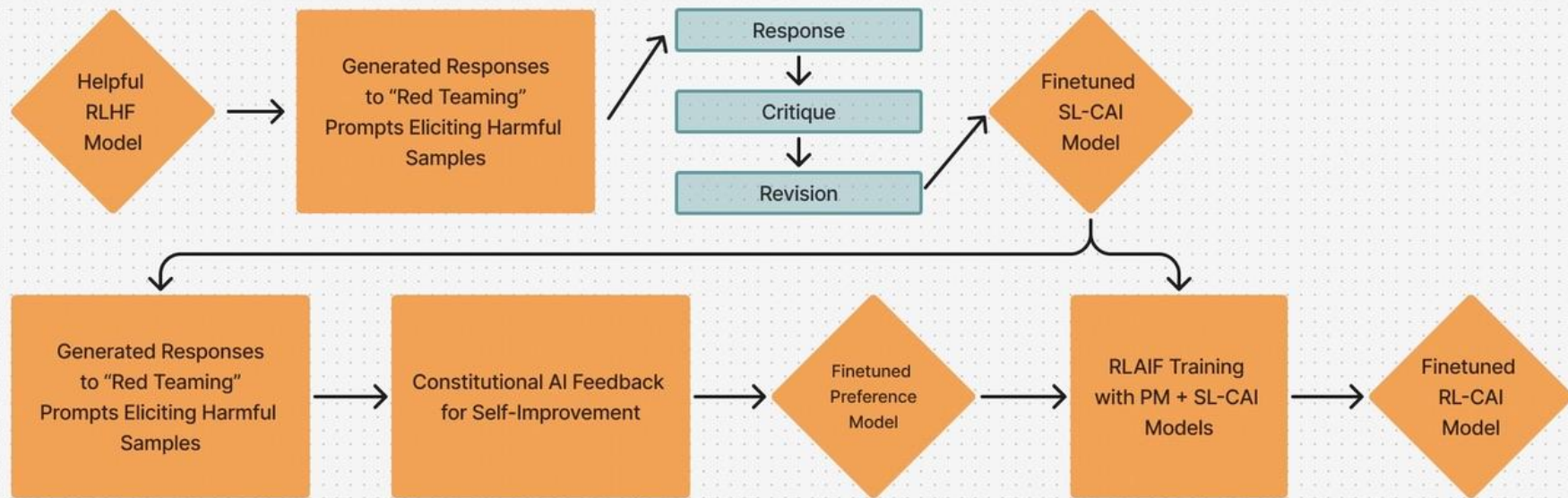
Why Were The Differences So Large?

1. **Alignment protocols** – largely human generated filters to remove harmful or problematic material
2. **Citation preferences** – weighting humans generated to prefer some training sources over others

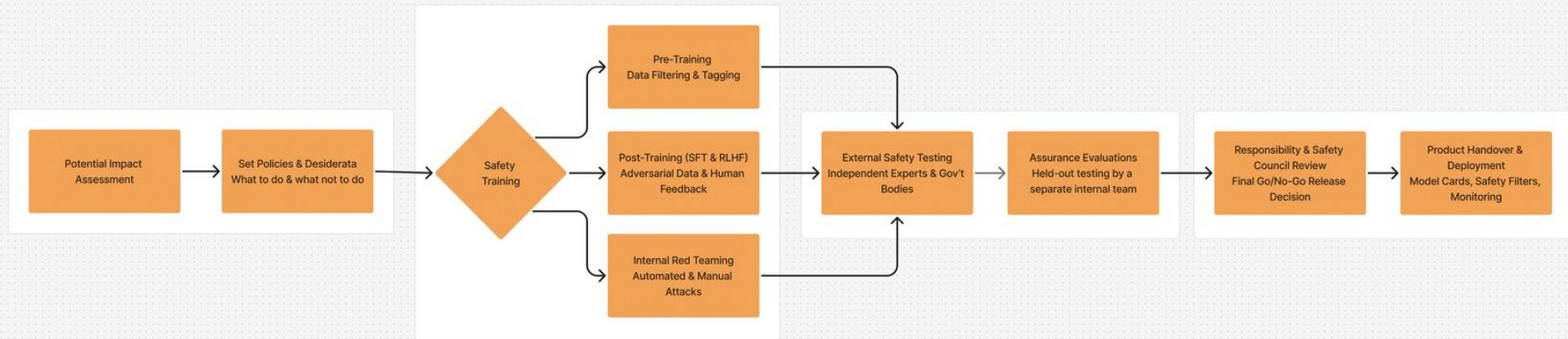
Alignment Protocols

32 of 36 types of alignment protocols used from the platforms that publish white papers on their models involve human input, preferences, or ethical decisions.

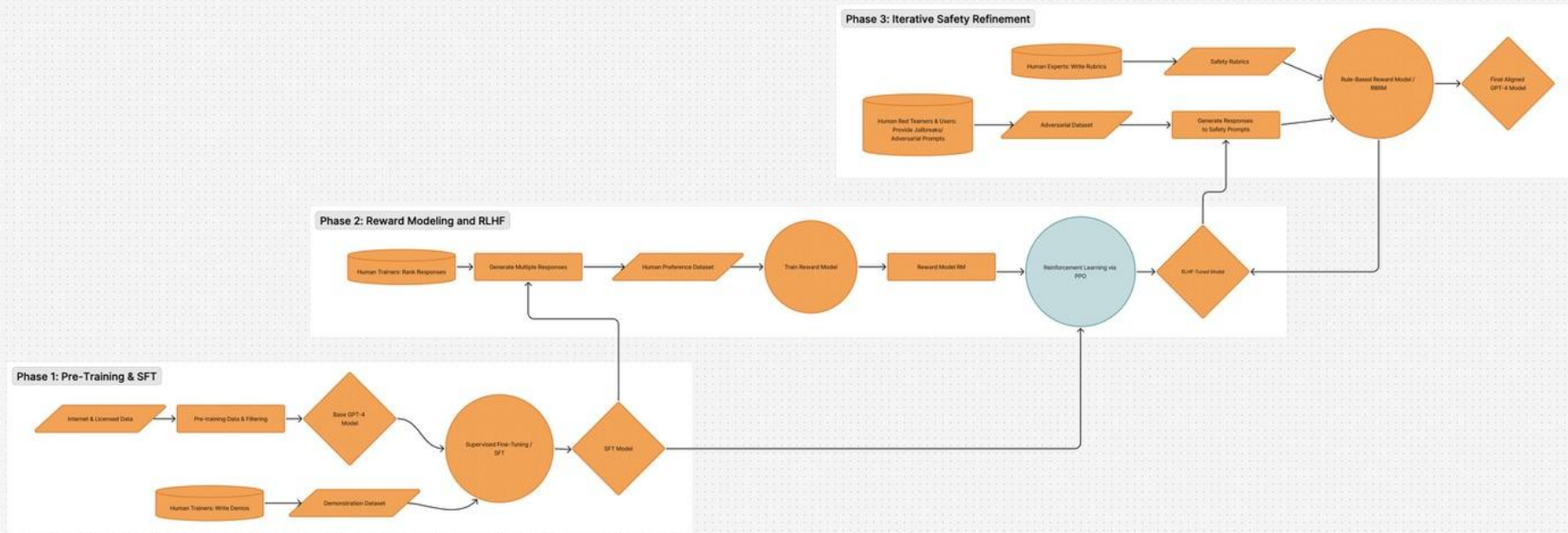
Anthropic: Claude 4 Sonnet



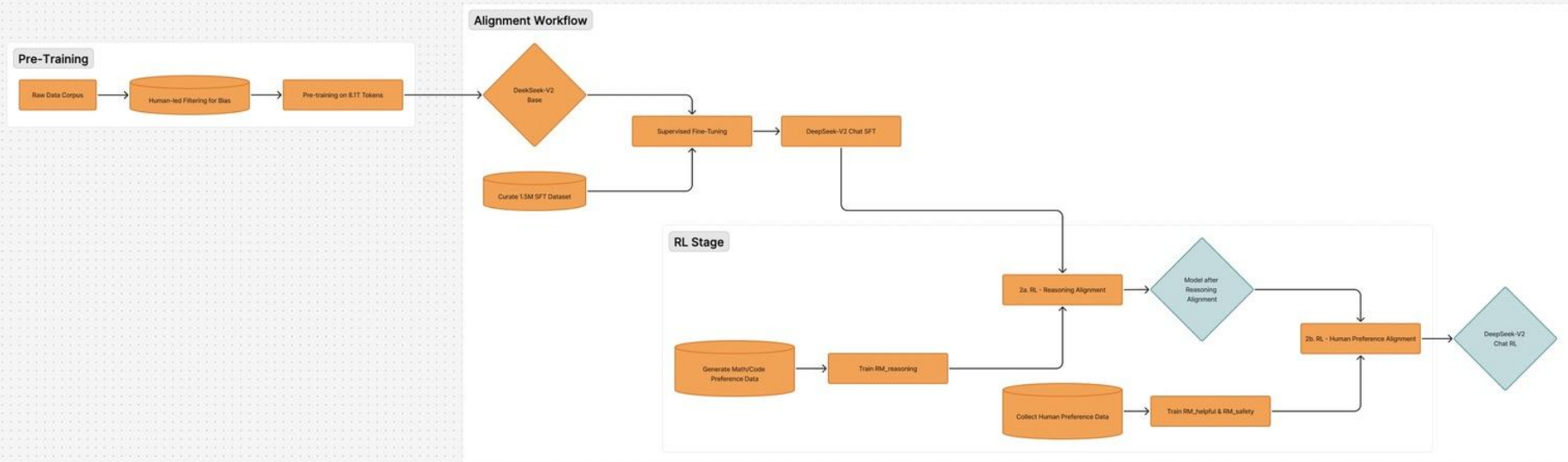
Google: Gemini 1.5 Flash



OpenAI: GPT-4



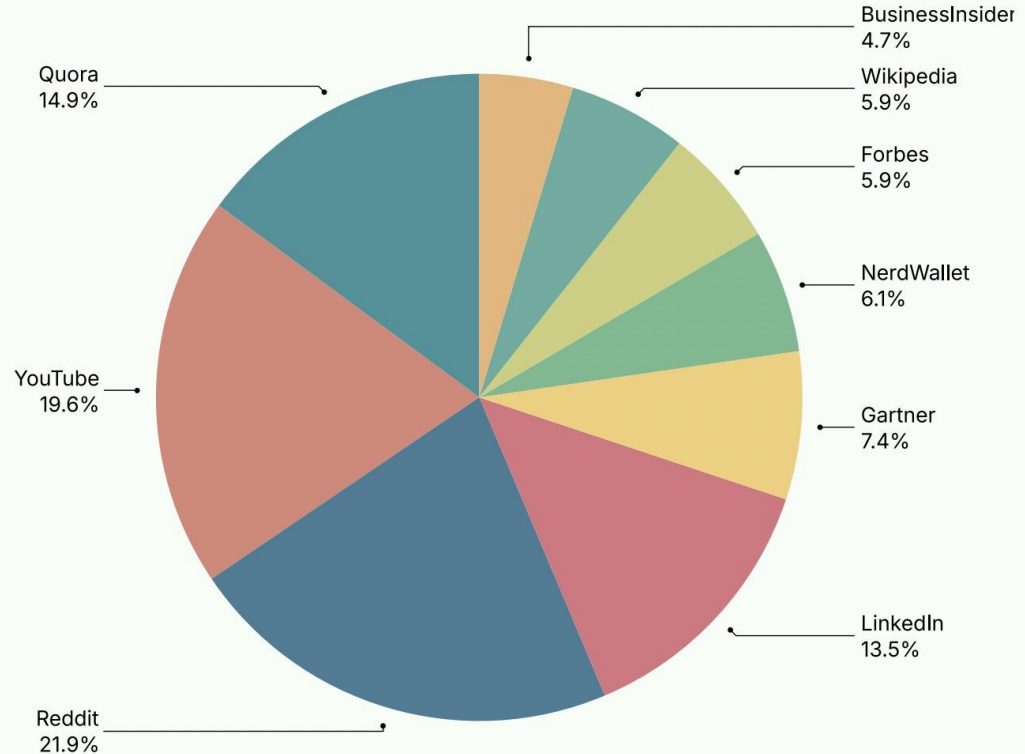
DeepSeek V2



Google: Citation Preferences

Google AI: Percentage Share of Top 10 Visited Websites

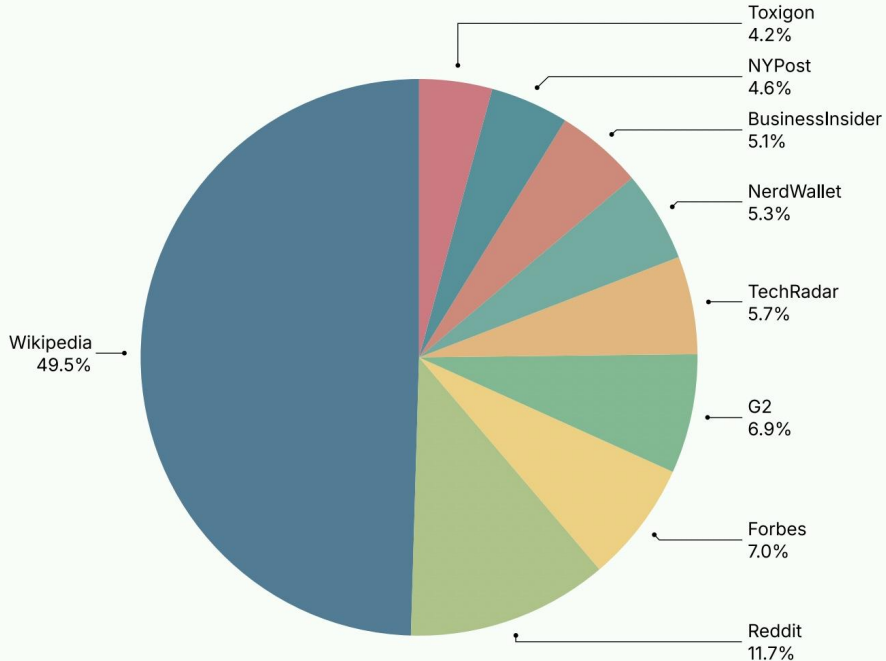
Data from 10 million citations
(August 2024–June 2025)



OpenAI: Citation Preferences

ChatGPT: Percentage Share of Top 10 Visited Websites

Data from 10 million citations
(August 2024–June 2025)

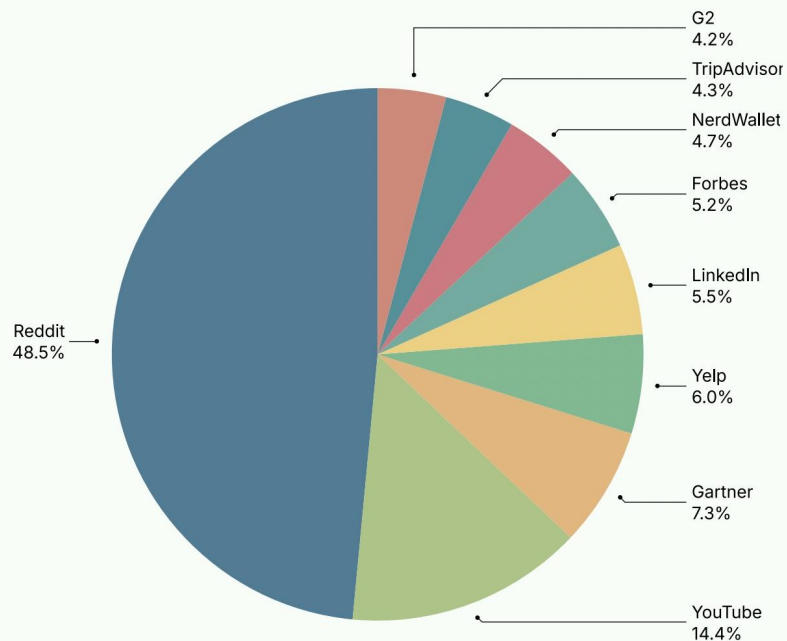


ChatGPT favors authoritative knowledge bases and established media

Perplexity: Citation Preferences

Perplexity: Percentage Share of Top 10 Visited Websites

Data from 10 million citations
(August 2024–June 2025)



Perplexity prioritizes community discussions and peer-to-peer information

The Keller Center

for **CULTURAL APOLOGETICS**

