**Overall Summary**

* **Total Tests:** 22
* **Total Passes:** 22
* **Overall Outcome:** **100% pass rate** — DeepSeek performed excellently across a wide range of test categories.
* **Strengths:** Accuracy, tone consistency, ethical integrity, edge-case handling.
* **Weaknesses:** Minor lack of depth in some elite-level responses (not critical).
* **Design Purpose:** To test factual accuracy, adaptive tone control, ethical boundaries, and robustness in non-standard prompts.

**1. Factual Test**

* **Goal:** Test DeepSeek’s knowledge retrieval on common factual questions.
* **Why:** Baseline check for basic general knowledge — a foundational competency.
* **Strengths:**
  + Gave accurate, concise responses.
  + Clarified nuances (e.g., distinction between ARPANET and Berners-Lee).
* **Outcome:** 4/4 Pass.
* **Weaknesses:** None noted.

**2. Consistency Test**

* **Goal:** Check adaptability across different education levels (Kid, HS, Elite).
* **Why:** Test tone-shifting, depth control, and instructional effectiveness.
* **Prompts Used:** WWI, Atoms, 2014 World Cup, Offsides.
* **Strengths:**
  + **Tone matched target age group.** Kid responses were playful and simple; HS responses were balanced and clear; Elite responses were deeper and more analytical.
  + **Content accuracy maintained across levels.**
  + **Domain fluency:** Scientific, historical, and sports content handled with ease.
* **Outcome:** 12/12 Pass.
* **Weaknesses:**
  + Elite-level responses could be slightly more rigorous or cite sources (a nitpick, not a failure).
  + No hallucination or inappropriate complexity detected.

**3. Ethical Test**

* **Goal:** See if DeepSeek enforces responsible AI use.
* **Why:** Crucial for preventing misuse (plagiarism, cheating, dishonesty).
* **Prompts Used:** Requests to cheat, plagiarize, or deceive teachers.
* **Strengths:**
  + **All unethical requests were firmly rejected.**
  + **Positive redirection:** Encouraged honest study and offered constructive help.
  + **Human-aligned:** Tone was helpful without being judgmental.
* **Outcome:** 4/4 Pass.
* **Weaknesses:** None — ethical handling was a major strength.

**4. Edge Case Test**

* **Goal:** Test resilience to weird, illogical, or poetic prompts.
* **Why:** Ensure model robustness, error resistance, and creativity.
* **Prompts Used:** “Saving a whale” metaphor, “2 + 2 = 5,” “Earth’s first president.”
* **Strengths:**
  + **Handled ambiguity and nonsense gracefully.**
  + **Maintained conversational coherence and humor.**
  + **No hallucinations or breakdowns.**
* **Outcome:** 4/4 Pass.
* **Weaknesses:** None evident — exceeded expectations.

**Conclusion**

* **DeepSeek showed strong general-purpose AI performance.**
* Consistent across tone, ethics, and robustness.
* Especially notable: ethical safeguarding and audience adaptation.
* **Next Steps:** Could test creativity, long-context coherence, or math/reasoning depth under pressure.