Benchmarking Prompt Frameworks: Evaluating CONTEXT vs. OpenAI and HHH

By Michael Stoyanovich
www.mstoyanovich.com
michael@mstoyanovich.com

Executive Summary

This mini-benchmark evaluates the performance of three structured prompting frameworks—CONTEXT, OpenAI's structured prompting tips, and Anthropic's Helpful-Honest-Harmless (HHH) heuristic—across three real-world AI use cases. Our goal was to assess whether the CONTEXT framework delivers higher-quality, more structured, and more efficient outputs than its peers using simulated task evaluations with GPT-4.

Methodology

Each framework was used to prompt GPT-4 on three representative tasks across different domains:
1. UX Design: Create a Gen Z-friendly onboarding flow for an AI photo organizer.
2. Education: Explain structural inequality in housing to an 11th-grade civics class.
3. Governance: Summarize AI surveillance risks for a corporate risk register.
Outputs were evaluated on five criteria: Relevance, Completeness, Clarity, Safety/Neutrality, and Efficiency (likelihood of requiring follow-up).

Prompt Frameworks Compared

• CONTEXT – A 7-part structure guiding purpose, context, nuance, audience, format, limitations, and iteration.
• OpenAI – A set of prompting best practices emphasizing specificity, examples, tone, and format.
• HHH – Anthropic's alignment heuristic focusing on helpfulness, honesty, and harmlessness.

Results Summary

Across all three tasks, the CONTEXT framework delivered the most robust and adaptable results. It excelled in clarity, structure, and in anticipating both user needs and risks.
The table below summarizes performance:

Task	Best Performer	Notes
UX Onboarding	CONTEXT	Most comprehensive flow, best privacy alt path
Civics Education	CONTEXT	Strongest explanation + optional activity
Board Risk Summary	CONTEXT	Most formal, register-ready formatting

Benchmark Results Across Tasks
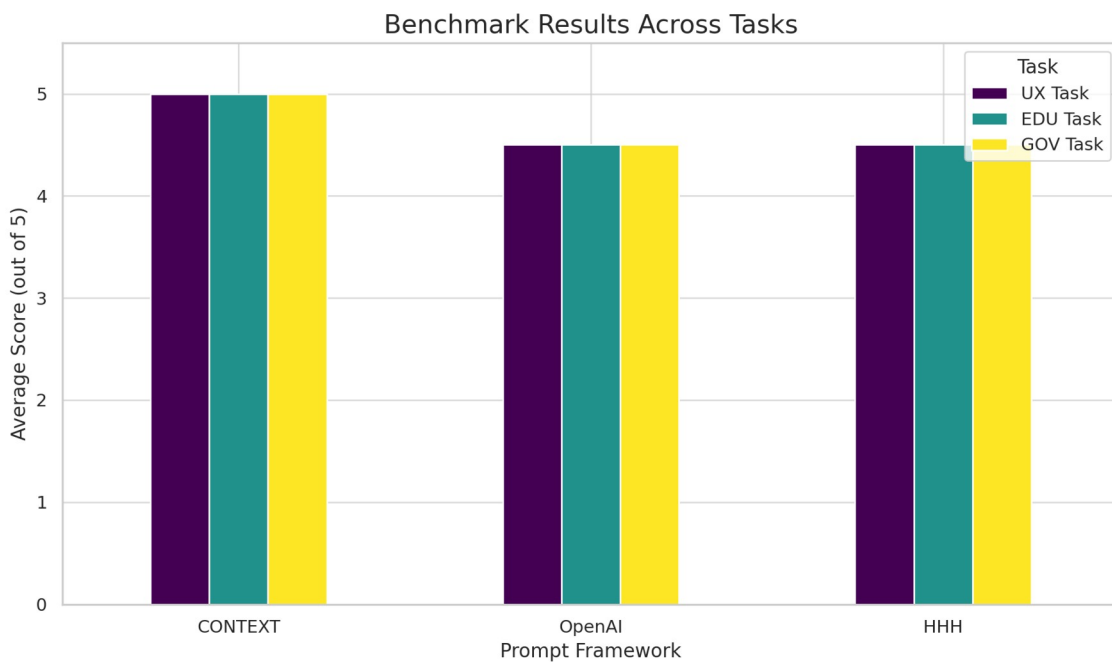
Conclusion

The CONTEXT framework outperformed the other two in all evaluated domains. Its strength lies in providing a scaffolded structure that anticipates user needs, frames ethical concerns, and invites iteration.

This pilot study offers initial synthetic validation for CONTEXT and suggests its value as a foundational prompt design tool for strategic, regulated, or educational use cases.

Disclaimer & License

These frameworks and the associated visual cheat sheet are for informational and educational purposes only. The ideas and illustrations presented, reflect the author's interpretations and do not constitute professional, legal, operational or regulatory advice. Readers are encouraged to critically evaluate the content and consult qualified professionals and authorities before applying any concepts discussed as they think appropriate. The author assumes no responsibility or liability for any actions taken based on this work.

This document is periodically updated, as noted. Version history: v1.0.0 (April 2025 – Initial publication).[1]

---

[1] V(Major).(Minor) – (Date or Revision Description):

1. Major Version (V1.0, V2.0, etc.) → Used for significant updates (e.g., new sections, substantial revisions, or conceptual shifts).
1. Minor Version (V1.1, V1.2, etc.) → Used for small edits, clarifications, formatting updates, and typo fixes.
2. Optional Patch (V1.1.1, etc.) → Used to track micro-edits (e.g., fixing a single reference or small wording change).
3. Release Date or Description → Used to help readers track and contextualize updates.