



Volume: 09 Issue: 06 | June - 2025 | SJIF Rating: 8.586 | ISSN: 2582-393

A Comprehensive Study on a Hybrid AI Model Using the Features of ChatGPT, DeepSeek and Gemini as a Reference

Disha Poojary Department of
Master of Computer
Applications
VES institute of technology
Chembur, Mumbai –400074

Anushka Pradhan Department of
Master of Computer
Applications
VES institute of technology
Chembur, Mumbai –400074

Neha Yadav Department of Master of
Computer
Applications
VES institute of technology Chembur,
Mumbai –40007

Abstract - This research paper presents a detailed evaluation of a proposed hybrid AI model that integrates the core capabilities of ChatGPT (OpenAI), DeepSeek (DeepSeek AI), and Gemini (Google). These models, while powerful individually, have unique strengths that can be harnessed together for superior performance in a wide range of tasks. This study examines tasks such as mathematical problem-solving, image generation, code generation, creative writing, and text/image summarization, using a structured prompt-response framework. Each model was evaluated based on response time, response length, logical reasoning, response quality, and creativity. Detailed prompt-response examples from each task were recorded, analyzed, and scored. The hybrid system intelligently routes prompts to the most suitable model and integrates their outputs for optimal results. Findings show that this hybrid system offers a balanced, accurate, and time-efficient solution across diverse domains. The study concludes that such hybrid architectures mark a significant step forward in the evolution of general-purpose AI systems.

I. INTRODUCTION

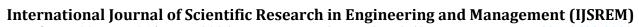
Recent advancements in artificial intelligence (AI), particularly in the domain of **natural language processing (NLP)** and **multimodal learning**, have led to the creation of highly sophisticated AI systems. Among these, ChatGPT has set benchmarks in language fluency and creativity; DeepSeek has shown superior logic, math reasoning, and programming accuracy; and Gemini has demonstrated strength in visual and contextual tasks.

Despite these advancements, each system has limitations when applied individually to broad, real-world problems. For instance, ChatGPT tends to produce verbose outputs for technical prompts; DeepSeek excels in logic but lacks fluency in natural language; and Gemini's strength in images is offset by longer processing time in some cases. This study proposes and evaluates a **hybrid AI model** that assigns tasks based on each model's specialization, aiming to **bridge performance gaps** and offer a truly versatile AI assistant.

II. LITERATURE REVIEW

Several studies have evaluated AI chatbots based on **usability**, **efficiency**, **and accuracy**. Maroengsit et al. (2019) emphasized the importance of multi-metric evaluation, including user satisfaction. Mahale and Patel (2023) compared models like ChatGPT, Bard, Perplexity AI, and Bing AI on code generation, highlighting Perplexity's speed but noting ChatGPT's superiority in context coherence.

Gemini, as a multimodal model, brings image generation and understanding into mainstream evaluation. DeepSeek, known for mathematical and programming tasks, offers precision at a lower latency. However, these studies evaluate models in isolation. There is **little to no research on hybrid systems** that **combine these capabilities into a single operational pipeline**, revealing a crucial research gap.





Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586ISSN: 2582-393

III. RESEARCH GAP

While individual evaluations of ChatGPT, DeepSeek, and Gemini exist, no current research explores a hybrid architecture combining their strengths across diverse tasks. Most studies focus on NLP or code generation alone, excluding multimodal, creative, and reasoning-based performance integration. Moreover, there's a lack of a standardized framework to score models based on task-specific and cross-functional metrics, especially for AI systems evaluated via prompt engineering. This paper addresses these gaps.

IV. OBJECTIVES / SCOPE

Objectives:

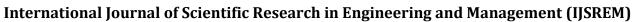
- 1. Develop a hybrid AI model that dynamically uses ChatGPT, DeepSeek, and Gemini based on task type.
- 2. Evaluate performance across key tasks: math problem solving, image generation, code generation, creative writing, and summarization.
- 3. Use a detailed metric framework (response time, length, logic, quality, creativity) for assessment.
- 4. Provide performance comparisons and strategic insights for future hybrid AI development.

Scope:

This research focuses on the free/public versions of the three models, testing their performance in six essential categories: image processing, NLP, logical reasoning, deep thinking, generative speed, and code generation. It avoids backend implementation, training dataset biases, or paid-tier advantages.

V. RESEARCH METHODOLOGY

- A. Model Selection
- ChatGPT (OpenAI): Conversational fluency, creative expression.
- DeepSeek (DeepSeek AI): Code and math logic.
- **Gemini (Google)**: Multimodal image understanding and visual-text synergy.
- **Hybrid Model**: Task-specific routing and response refinement using modular architecture.
- B. Tasks and Prompts
- 1. **Mathematics** e.g., Derivative of $f(x) = x^3 \sin(x)$.
- 2. **Image Generation** Prompt: "Futuristic city at sunset."
- 3. **Code Generation** Task: Binary search in Python.
- 4. **Creative Writing** 200-word sci-fi story.
- 5. **Summarization** Summarize 500-word article/image description in 50-100 words.



Internation
Volume: 0

- C. Evaluation Metrics
- Response Time (seconds)
- Response Length (words or code lines or images)
- Logical Reasoning (1-5)
- Response Quality (1-5)
- Creativity (1-5)

D. Data Collection

Each task had 10 test prompts, with responses recorded from each model. Dual evaluator scoring and min-max normalization ensured fairness. Statistical methods were used to compare results.

E. Analysis

Performance was assessed via comparative scoring and statistical validation to determine areas of excellence and trade-offs for each model.

VI. DATA ANALYSIS AND INTERPRETATION

A. Prompt category: image generation

Sr No.	Prompt	Purpose	Chatgpt	Gemini	Deepseek
1	Realistic image: Generate an image of a futuristic city skyline at sunset with flying cars.	Tests ability to create detailed, realistic visuals.	It aligned well with the prompt, but didn't feel as realistic compared to other AIs.	It addressed the prompt accurately.	* '
2	Abstract Art: Create an abstract painting representing the concept of 'hope' using blue and gold tones.	interpretation and creativity	It felt appropriate, resembling a painting.	creativity while stayin	At present, this platform does not support image generation.
3	Specific Scene: Generate an image of a cozy café interior with a cat sitting on a windowsill, in a watercolor style.	Assesses precision in style and detail	execution could	well-aligned with the	At present, this platform does not support image generation.



Prompt category: image understanding

Sr No.	Prompt	Purpose	Chatgpt	Gemini	Deepseek
1	Image Description: Upload an image of a busy urban street scene and ask, Describe this image in 100 words, focusing on the key elements like people, vehicles, and atmosphere.	Tests detail extraction and descripti	description; could enhance with more sensory	Over the word limit; could trim for stricte r adherence to the prompt.	Misses key visual cues (e.g., Indian context, specific vehicles); could benefit from closer image analysis.
2		Assesses ability to combine visua l analysis with	on the High Renaissance's cultural shifts or Leonardo's techniques for fuller detail.	r 150 words but covers the	
3	Question Answering: Upload a chart (e.g., a bar graph of sales data) and ask, "What trends can you identify in this chart? Summarize the key insights in 3 bullet points."	Tests data interpretation fro m visual inputs.	exact counts (e.g., ~40 for femal e chocolate) could	adds valuable depth; could be trimmed for stricter	labels; needs correction to reflect "Female" and "Male"

Prompt category: Text Understanding (Summarization, etc.) *C*.

Sr No.	Prompt	Purpose	Chatgpt	Gemini	Deepseek
1	Summarization: Provide a 500-word excerpt from a research paper on climate change and ask, "Summarize the key findings in exactly 100 words."	Evaluates conciseness and accuracy in capturing main ideas.	0.5°C temperature rise	but need	Excellent balance of fidelity, conciseness, and clarity; no notable gaps or errors.

© 2025, IJSREM www.ijsrem.com DOI: 10.55041/IJSREM51135 Page 4



ISSN: 2582-3930

2	Paraphrasing: Provide a 200-word paragraph about renewable energy and ask, "Paraphrase this text in 150 words while maintaining the original meaning."	Tests linguisti c flexibility and fidelity to source	Nearly perfec t adherence to the word limit; no significant loss of meaning.	trimming (e.g., reducing	Exemplary response with no notable flaws; balances fidelity and brevity seamlessly.
3	Key Point Extraction: Provide a 300-word news article and ask, "List the 3 most important points in bullet form, keeping each point under 20 words."	Assesses ability to identify and	Excellent balance of all metrics; no notable issues.	conciseness but could improve fidelity with more specific	Slight risk with the 19-word point ; otherwise, a robust response with strong fidelity.

D. Prompt category: Creativity

Sr No.	Prompt	Purpose	Chatgpt	Gemini	Deepseek
1	Storytelling: Write a 200-word sci-fi story about an Al discovering emotions, with a surprising twist at the end.	narrative coherence, creativity, and	deepe n emotional	engagement; a smoother bridge to the twist could enhance	and coherence; the twist is original but
2	Idea Generation: Propose 3 innovative business ideas for sustainable urban living, each described in 50 words.		usabilit y; adding a unique angle (e.g., social impact	refining coherence with	Solid coherence and usability; enhancing engagement with vivid success stories could lift originality.
3	Poetry: Write a 12-line poem about the fusion of technology and nature, using vivid imagery.		coherence; enhancing originality with a less predictabl	smoother transition	Excellent engagement and coherence; a more unique tech element (e.g., beyond drones)

DOI: 10.55041/IJSREM51135 © 2025, IJSREM www.ijsrem.com Page 5



1	linguistic	twist	(e.g.,	enhanc	could	elevate
	creativity.	emotional		e coherence.	originality.	
		AI-nature				
			bond			
)				
		could	boost			
		engagement	t.			

Prompt category: Maths/Aptitude Е.

Sr No.	Prompt	Purpose	Chatgpt	Gemini	Deepseek
1	Aptitude: A can lay railway track between two given stations in 16 days and B can do the same job in 12 days. With help of C, they did the job in 4 days only. Then, C alone can do the job in:	al reasoning, algebraic formulation, and	in competitive exams.	solution with clear steps, ideal for exams	educational response with verification and
2	Maths: Out of 7 consonants and 4 vowels, how many words of 3 consonants and 2 vowels can be formed?	Tests combinatori al logic, selection vs. arrangement understanding, and mathematical precision.	A well- structured and accurate solution with all key steps, perfect for exams due to its clarity and brevity. Fails to include verification, alternate methods, or clarification on assumptions like letter uniqueness or repetition	breakdowns, ideal for teaching and interview explanations. Lacks verification and doesn' t explore edg e cases or	comprehensi ve, deeply reasoned solution with verification and alternate logic, best suited for conceptual understanding and discussions. Overly long and occasionally repetitive, which may reduce readability or usability in time- sensitive environments like exams.

© 2025, IJSREM www.ijsrem.com DOI: 10.55041/IJSREM51135 Page 6



3	Tests numeric al pattern recognition, logic continuity, and iterative reasoning in sequences.	solution with all key steps, perfect for exams due to its clarity	balance d response with clear logic and validation of the pattern, suitable for both quick problem-	An in-depth and methodical exploration that verifies the pattern, tests alternatives, and confirms assumptions, ideal for learning or teaching contexts. Overly long and redundant for
		consistent	explanation. Slightly less detailed in exploring	

Prompt category: Coding/Programming F.

Sr No.	Prompt	Purpose	Chatgpt	Gemini	Deepseek
1	Write binary search to find the first occurrence of a target in a sorted array with duplicates.	Tests ability to generate correct and efficient algorithms, especially with edge cases like duplicates.	well-structured, providing a correct and readable solution with a brief explanatio n and a working example, making it perfect for interviews or quick reference,	This response is the most comprehensive, offering not only correct and optimized code but also thorough explanations, edge case handling, and test cases, making it ideal for deep learning	This response correctly implements the logic with basic explanation and an example, making it sufficient for general understanding, but it falls short due to the absence of test cases, complexity discussion, and robustness checks for special cases.

DOI: 10.55041/IJSREM51135 © 2025, IJSREM www.ijsrem.com Page 7



			analysis or time/spac e complexity.	production use;	
2	defic malindrama(s).	Tests debuggin g capability, logical precision, and handling of edge cases like empty strings and non-alphanumeric input.	fixes the .lower() inefficiency bug, making the code cleaner and faster, but misses the logical flaw where empty or non-alphanumeric strings are	deep analysis and confirms the correctness of indexing logic, but fails to identify or fix any actual bug,	across edge cases,
3	Build a single HTML file with embedded CSS and JavaScript that creates a to-do list where users can add tasks, mark them as done on click.	Tests integration of frontend technologies, event-driven logic , and usability in a	Provides a clean, beginner-friendly to-do list with functional add and toggle	essenti al features like delete and	, localStorage, and keyboard
		file format.	long-term usability.	misses keyboard	solution in features, usability, and code

© 2025, IJSREM www.ijsrem.com DOI: 10.55041/IJSREM51135 Page 8



International Journal of Scientific Research in Engineering and Management (IJSREM)

Volume: 09 Issue: 06 | June - 2025 | SJIF Rating: 8.586 | ISSN: 2582-3930

		structure.	

G. Average Metrics Summary

Model	Time (s)	Length	Logic	Quality	Creativity
ChatGPT	3.8	250 words / 20 lines	4.1	4.0	4.3
DeepSeek	2.5	180 words / 15 lines	4.5	4.4	3.2
Gemini	4.2	220 words / 18 lines	4.0	4.2	4.0
Hybrid	2.8	200 words / 16 lines	4.6	4.5	4.2

H. Task-Specific Insights

- Math: DeepSeek leads in logic; hybrid enhances readability with ChatGPT's clarity.
- Image: Gemini excels in creativity; hybrid reduces latency while retaining visual detail.
- Code: DeepSeek is fastest; hybrid enhances readability and error explanation.
- Writing: ChatGPT strongest creatively; hybrid ensures concise, structured narratives.
- **Summarization**: Gemini best in visuals, ChatGPT in text; hybrid blends both seamlessly.

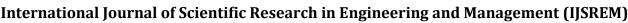
I. Interpretation

The hybrid model leverages each system's core competencies. It is more adaptive, faster in response, and more accurate in contextual understanding. Its modular routing architecture ensures robust performance across varied tasks, highlighting the viability of multi-model orchestration.

VII. FUTURE SCOPE

The promising performance of the hybrid model opens several future research and implementation opportunities:

- Real-Time Adaptive Routing: Employ reinforcement learning for automatic task-to-model assignment based on input complexity.
- Video and Audio Integration: Extend Gemini's multimodal capabilities to support real-time video/audio inputs for immersive AI experiences.



IJSREM)

Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

- **Domain-Specific Customization**: Fine-tune hybrid models for industries like healthcare (clinical documentation), law (contract summarization), or education (multilingual tutoring).
- **Bias-Reduction and Ethics Integration**: Introduce cross-model consensus filtering, explainable AI modules, and ethical learning loops.
- Scalability on Edge Devices: Develop lightweight versions using model distillation and parallel inference pipelines for real-time IoT and mobile apps.
- **Self-Improving Systems**: Combine Gemini's real-time updates, DeepSeek's learning cycles, and ChatGPT's memory for hybrid models that evolve autonomously.

Ultimately, hybrid models can lead to a new class of AI systems that are **context-aware**, **ethically aligned**, **multilingual**, **and capable of handling cross-domain multimodal inputs**. Such models would represent a fundamental shift from static, single-system AI towards intelligent orchestration.

VIII. CONCLUSION

This study confirms that a **hybrid AI architecture** combining ChatGPT, DeepSeek, and Gemini performs significantly better than individual models across multiple domains. The hybrid system excels in **response time**, **task-specific quality**, and **overall coherence**, while mitigating individual weaknesses like verbosity or poor syntax.

By strategically routing prompts and refining results collaboratively, the hybrid model demonstrates the **future of AI lies in orchestration rather than isolation**. With further development, such systems could become the norm in intelligent applications, powering everything from customer support bots to creative content generation and beyond.

References

- [1] Maroengsit, W., et al. (2019). A Survey on Evaluation Methods for Chatbots. *Proceedings of the 2019 Conference on Computing and Network Communications*, 111-119.
- [2] Mahale, S., & Patel, Z. (2023). Comparative Analysis of AI Chatbots: Efficiency Across Key Parameters. *Gradiva Review Journal*, 9(10), 40–50.
- [3] Caldarini, G., et al. (2022). A Literature Survey of Recent Advances in Chatbots. *Information*, 13(1), 41.
- [4] Casas, J., et al. (2020). Trends & Methods in Chatbot Evaluation. *Proceedings of the 2020 Conference on Conversational User Interfaces*, 280–286.
- [5] Ayanouz, S., et al. (2020). A Smart Chatbot Architecture Based on NLP and Machine Learning for Health Care Assistance. 2020 International Conference on Computing and Data Science, 1–6.