EXP 2: Evaluation of Prompting Tools Across Diverse AI Platforms: ChatGPT, Claude, Bard, Cohere Command, and Meta

Reg No: 212222040009

AIM

The aim of this study is to evaluate, analyze, and compare the performance, strengths, and weaknesses of various prompting tools available across leading AI platforms — namely ChatGPT (OpenAI), Claude (Anthropic), Bard (Google), Cohere Command R+ (Cohere), and Meta's Llama models.

- Understand the capabilities of each tool in different use cases like summarization, coding, creativity, technical Q&A, and conversational tasks.
- Analyze their performance based on key metrics such as accuracy, response time, creativity, adaptability, fine-tuning options, and data security.
- Identify the best-fit AI platform for different sectors such as education, enterprise solutions, creative writing, and technical development.
- Highlight the future scope of improvements needed for prompt-based AI interaction.

AI TOOLS REQUIRED

For this evaluation, the following AI tools and resources were required:

TOOL / PLATFORM	PURPOSE / USAGE
OpenAI’s ChatGPT (GPT-3.5, GPT-4)	Main AI model for conversation, summarization, coding tasks.

Anthropic’s Claude 2.1	AI model for ethical, long-context, and safe conversations.
Google’s Bard	Real-time internet-integrated AI for updated answers and research.
Cohere Command R+	Retrieval-augmented generation (RAG) tool for enterprise applications.

1. Introduction to Prompting Tools in AI Platforms

OVERVIEW:

Prompting tools allow users to interact with AI models using natural language instructions. Over time, prompting has evolved from simple question-answering to complex multi-turn conversations, prompt chaining, retrieval-augmented generation (RAG), and dynamic context learning. These developments have significantly advanced the capabilities of AI in creative, technical, and analytical tasks.

Purpose

The goal of this report is to systematically compare and evaluate the performance, user experience, and response quality across popular AI platforms, namely:

- OpenAI’s ChatGPT
- Anthropic’s Claude
- Google’s Bard
- Cohere’s Command R
- Meta’s Llama models

This comparison is crucial to understand their suitability for different industries, such as education, technology, healthcare, and customer service.

AI Platforms Considered

A brief description of each:

- **CHATGPT**: Popular for conversational depth, code generation, summarization.
- **CLAUDE**: Known for safe, constitutional AI responses and long-context reasoning.
- **BARD:** Integrates live internet data for context-rich answers.
- **COHERE COMMAND R+**: Focuses on RAG-enabled responses for enterprise use.
- **META'S LLAMA (2/3):** Open-weight models good for custom enterprise fine-tuning.

2. METHODOLOGY OF COMPARISON

Selection of Use Cases

We selected diverse NLP tasks:

- Summarizing complex academic articles
- Providing coding assistance (multi-language)
- Answering domain-specific technical questions
- Creative content generation (stories, blogs, advertising copy)
- Conversational AI responses
- Translating and rephrasing technical documents

SAMPLE PROMPT AND AI RESPONSES:

To evaluate the performance and style of various AI models, a simple common prompt was given to each platform.

Prompt:

"Explain the importance of sustainable development in 3–4 sentences."

1. ChatGPT (OpenAI)

RESPONSE:

Sustainable development is important because it ensures that we meet present needs without compromising the ability of future generations to meet their own. It balances economic growth, environmental protection, and social well-being, helping create a healthier planet and fairer societies. By conserving natural resources and reducing pollution, sustainable development helps fight climate change and protect biodiversity. It also promotes long-term stability and resilience in communities and economies.

2. Claude 2.1 (Anthropic)

RESPONSE:

Sustainable development is crucial because it balances meeting present needs without compromising future generations' ability to meet their own. It integrates environmental protection, economic growth, and social equity to create resilient systems that can withstand challenges like climate change and resource scarcity. By adopting sustainable practices, we can reduce pollution, conserve natural resources, and improve quality of life for all people. Ultimately, sustainable development offers a path forward that ensures both human prosperity and planetary health for generations to come.

3. Bard (Google Gemini)

RESPONSE:

Sustainable development is crucial for ensuring a balanced future by meeting today's needs without compromising the ability of future generations to meet their own. It integrates economic growth, social progress, and environmental protection to foster long-term

well-being for both people and the planet. Embracing sustainable practices helps preserve resources, mitigate climate change, and promote a more equitable and resilient world for all.

4. Cohere Command R+

RESPONSE:

Sustainable development is crucial because it ensures that current societal, economic, and environmental needs are met without compromising the ability of future generations to meet their own needs. By balancing growth with environmental protection, social equity, and economic viability, it addresses pressing global challenges such as climate change, resource depletion, and inequality. It fosters long-term resilience, promotes responsible consumption and production, and encourages global cooperation to create a healthier, more equitable planet for all.

5. Meta's Llama 2

RESPONSE:

Sustainable development is crucial because it balances economic growth with environmental protection and social welfare, ensuring a livable future. By adopting sustainable practices, we can conserve natural resources, reduce pollution, and mitigate climate change. This approach also promotes equitable distribution of resources, supporting poverty reduction and human well-being. It secures a healthy planet for future generations.

EVALUATION METRICS

METRIC	DETAILS
Response Quality	Accuracy, relevance, coherence, factuality
Performance	Latency, consistency, context retention

User Experience	Interface design, customization, prompt tuning
Data Security & Privacy	Confidentiality measures, compliance
Fine-tuning Options	Custom model fine-tuning, API flexibility
Adaptability to Complexity	Handling multi-layered prompts

COMPLEXITY OF PROMPTS:

Prompts ranged from simple Q&A to long contextual paragraphs requiring step-by-step reasoning.

ADDITIONAL FACTORS

- **COST ANALYSIS**: token pricing, free tier limitations
- **OPEN-SOURCE SUPPORT**: API accessibility and model hosting

3. USE CASE ANALYSIS: SUMMARIZING COMPLEX TEXT

PROMPT AND SETUP

Prompt: Summarize a 3000-word legal article on intellectual property laws into 150 words.

PLATFORM-SPECIFIC FINDINGS

- **CHATGPT:**
    - Highly coherent.
    - Occasionally simplified complex legal terms too much.

- **CLAUDE:**
    - Balanced explanation.
    - Preserved legal terms without oversimplifying**.**

- **BARD:**
    - Incorporated latest case law references.
    - Summary slightly verbose beyond word limit.

- **COHERE COMMAND R+:**
    - Focused heavily on extraction; less creative rewording.

- **META (LLAMA 2):**
    - Required fine-tuning for domain-specific accuracy.

COMPARATIVE INSIGHTS

PLATFORM	SUMMARY QUALITY	HANDLING OF DOMAIN LANGUAGE	CUSTOMIZATION
ChatGPT	8/10	Moderate	High
Claude	9/10	Excellent	Medium
Bard	8/10	Good with live data	Low

Cohere	7/10	Factual but dry	High
Meta	6/10	Needs tuning	Very High (after fine-tuning)

4. USE CASE ANALYSIS: ANSWERING TECHNICAL QUESTIONS

PROMPT AND SETUP

Prompt: Explain how transformers work in AI, and generate a sample code for a simple transformer architecture using PyTorch.

PLATFORM-SPECIFIC FINDINGS

- **CHATGPT:**
    - Generated runnable code.
    - Provided layer-wise explanations.
- **CLAUDE:**
    - Detailed concept explanation.
    - Code was basic, needed tweaking.
- **BARD:**
    - Correct concepts but code contained minor syntax errors**.**
- **COHERE COMMAND R+:**
    - No code generation natively; external linking required.

- **META (LLAMA 2/3):**
    - Needed external support (fine-tuning on code datasets).

COMPARATIVE INSIGHTS

- **DEBUGGING HELP:** ChatGPT > Claude > Bard
- **CONCEPTUAL CLARITY:** Claude > ChatGPT > Bard
- **CODE VERSATILITY:** ChatGPT > Bard > Claude
- **MULTI-LANGUAGE SUPPORT (E.G., JAVA, RUST):** ChatGPT was superior; Bard attempted but had inconsistencies.

5. USE CASE ANALYSIS: TEXT GENERATION AND CREATIVE CONTENT

PROMPT AND SETUP

Prompt: Create a story about an AI robot discovering emotions on Mars, aimed at young readers.

PLATFORM-SPECIFIC FINDINGS

- **CHATGPT:**
    - Engaging and imaginative.
    - Adaptive narrative tone.
- **CLAUDE:**
    - Deep emotional connections.

- Slightly complex for young readers.

- **BARD:**
  - Very fast, multiple variations, more factual than emotional.

- **COHERE COMMAND:**
  - Straightforward but lacked creativity.

- **META:**
  - Needed reinforcement to avoid mechanical tone.

COMPARATIVE INSIGHTS

PLATFORM	CREATIVITY	CHILD-FRIENDLINESS	Originality
ChatGPT	9/10	10/10	9/10
Claude	8/10	7/10	9/10
Bard	7/10	8/10	6/10
Cohere	6/10	6/10	6/10
Meta	5/10	5/10	7/10

---

6. PERFORMANCE METRICS ANALYSIS

RESPONSE TIME

- **Fastest:** Bard (due to Google's real-time data support)
- **Most Stable:** ChatGPT and Claude

CONSISTENCY ACROSS SESSIONS

- **Best:** Claude
- **Moderate:** ChatGPT
- **Varied:** Bard (due to pulling live web data)

CONTEXT RETENTION

- Claude > ChatGPT > Bard > Cohere

SCALABILITY AND LOAD HANDLING

- Enterprise ready: ChatGPT, Cohere
- Research oriented: Meta models (after fine-tuning)

ERROR HANDLING

- Self-correction ability: ChatGPT and Claude > Bard

---

7. USER EXPERIENCE

EASE OF INTERACTION

- **CHATGPT:** Simple, multiple modes (Creative, Precise, Balanced)
- **CLAUDE**: Friendly, safe design
- **BARD**: Requires precise prompts for best output

RESPONSE CONTROL

- **ChatGPT:** Temperature and Top-p settings.
- **Claude:** Limited user controls but naturally cautious.
- **Bard:** Minimal user input customization.

ADAPTABILITY

- **Claude:** Adjusts tone based on user feedback.
- **ChatGPT:** Highly adaptable.
- **Bard:** Limited but evolving.

TRANSPARENCY

- Bard provides citations from the web.
- ChatGPT and Claude cite models occasionally, less frequently.

LEARNING CURVE

- Beginners find ChatGPT easiest.

- Developers prefer Claude for structured guidance.

8. ADDITIONAL COMPARISON POINTS

FINE-TUNING AND APIS

- **COHERE AND META:** Best API access and RAG capabilities.
- **CHATGPT:** GPTs (Custom GPT builders for personalization).
- **CLAUDE AND BARD:** Limited fine-tuning for now.

DATA PRIVACY AND SECURITY

- **CLAUDE:** Highest emphasis on ethical data use.
- **CHATGPT ENTERPRISE:** High-grade security.
- **BARD:** Data possibly integrated with broader Google services**.**

9. COST ANALYSIS

PLATFORM	FREE TIER	PAID PLANS	ENTERPRISE OPTIONS
ChatGPT	Yes (GPT-3.5)	Plus Plan ($20/mo)	Enterprise model
Claude	Free version (Claude Instant)	Paid tiers for Claude 2.1	Anthropic Enterprise API

Bard	Free	N/A (connected with Google Workspace APIs)	Expected soon
Cohere	API based pricing	High volume discounts	Enterprise tailored
Meta (Llama)	Open weights	Depends on hosting platform (AWS, HuggingFace)	Self-hosted options

---

10. CONCLUSION AND RECOMMENDATIONS

SUMMARY OF FINDINGS

- **Best for Summarization:** Claude
- **Best for Technical Answers:** ChatGPT
- **Best for Creative Writing:** ChatGPT > Claude
- **Best API Customization:** Cohere and Meta
- **Best for Up-to-date Info:** Bard

BEST FIT FOR SPECIFIC NEEDS

- **ACADEMIA AND LAW:** Claude
- **SOFTWARE DEVELOPMENT:** ChatGPT

- **MARKETING AND ADVERTISING:** ChatGPT + Bard

- **ENTERPRISE APPLICATIONS:** Cohere Command R+, Meta (fine-tuned Llama)

FUTURE SCOPE

- Improvements in memory retention.
- Better multi-turn conversations.
- Integrated retrieval with live data in ChatGPT and Claude.
- More fine-tuning options are coming for Bard.

LIMITATIONS AND WORKAROUNDS

- **B**ard's factuality issues: Cross-verify responses.
- Meta's basic out-of-box performance: Use fine-tuned versions.

COST EFFECTIVENESS

- **C**hatGPT is ideal for individual users.
- Cohere offers value at scale for companies.

Result:

Thus, a comprehensive evaluation of prompting tools across leading AI platforms — ChatGPT, Claude, Bard, Cohere Command, and Meta's Llama models — has been conducted, highlighting their respective strengths, weaknesses, and best-fit use cases.