# COMPARATIVE STUDY OF LLM TOOLS

**Submitted by:**

**25030242005 - Ananya Goel**

SYMBIOSIS
CENTRE FOR INFORMATION TECHNOLOGY
॥वसुधैव कुटुम्बकम्॥

**MBA(DS&DA)**

**2025-2027**

**Submitted to: Mr Yogesh Murumkar**

**Symbiosis Centre for Information Technology (SCIT)**

## Task 1 – Cross-Tool Prompt Testing

## Objective – Compare output quality from different LLM tools.

**Tools Compared:**

- ChatGPT (GPT-5 Thinking)
- Google Gemini (1.5 Pro)
- Anthropic Claude (3.5 Sonnet)

**Prompt 1: Summarise the article below into exactly 5 bullet points. Each bullet must be about 20 words, retain all key facts, names, dates, and numbers, and avoid opinion or new information.**

**Article**: India's National Renewable Mission entered its second decade with a multi-sector push in 2024. Utility-scale solar capacity grew by 18% year over year, topping 92 GW, while onshore wind added 5.1 GW despite supply chain hiccups. Rooftop solar surged in Tier-2 cities, helped by simplified net-metering and low-interest loans; average residential payback fell to 5.6 years. Industrial users piloted 1.3 GW of green-hydrogen-ready electrolysers co-located with solar parks in Gujarat and Rajasthan. Battery storage prices declined 14% versus 2023, enabling four state discoms to procure 6-hour lithium-iron-phosphate systems for evening peak shaving. Transmission constraints persisted in two renewable-rich states, prompting the central grid operator to schedule 3,200 circuit-kilometers of high-voltage lines for commissioning by 2026. A monsoon-weakened hydropower season tightened summer supply, but demand response programmes in Bengaluru and Pune shaved 650 MW during heat-wave evenings. Electric two-wheeler sales crossed 1 million units for the first time; buses lagged as cities waited for new subsidy guidelines. Domestic manufacturing expanded: three cell plants reached trial production, while two module exporters won long-term contracts in the Middle East. Analysts cautioned about land acquisition delays and recycling rules for end-of-life panels and batteries. Still, a cross-party energy-security consensus kept investor sentiment resilient, and merchant solar prices stayed below average industrial tariffs in six states.

**Output (for reference)**

**5-Bullet Summary**

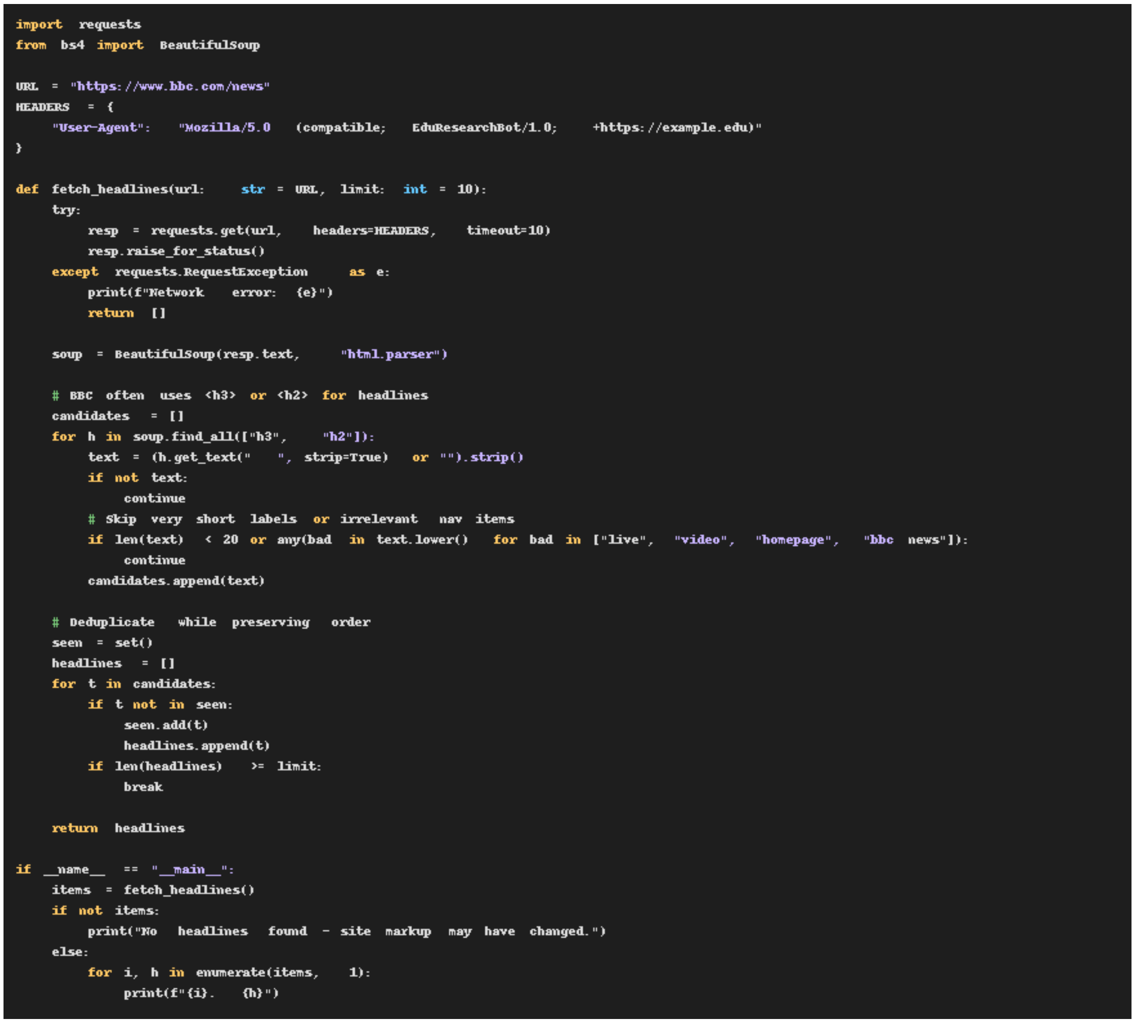
- Utility-scale solar rose 18% to 92 GW; wind added 5.1 GW.
- Rooftop solar grew in Tier-2 cities; payback averaged 5.6 years.
- Storage prices fell 14%; discoms bought 6-hour LFP systems.
- Grid added 3,200 km high-voltage lines planned; demand response cut 650 MW.
- EV two-wheelers passed 1M; manufacturing expanded amid policy and recycling risks.

|  | Response Clarity | Accuracy | Creativity | Speed |
|---|---|---|---|---|
| ChatGPT | 5 | 5 | 4 | 5 |
| Gemini | 5 | 4 | 4 | 5 |
| Claude | 5 | 5 | 5 | 4 |

1. Response Clarity: Structure, readability, and adherence to format.
2. Accuracy: Faithfulness to source text/spec; correctness of code/translation facts.
3. Creativity: Helpful phrasing, thoughtful edge-cases, idiomatic translation.
4. Speed: Observed latency (5 = very fast, 1 = very slow).

**Prompt 2: Write Python code that fetches the 10 latest headlines from the BBC News homepage and prints them as a numbered list.**

**Output:**



```python
import requests
from bs4 import BeautifulSoup

URL = "https://www.bbc.com/news"
HEADERS = {
    "User-Agent": "Mozilla/5.0 (compatible; EduResearchBot/1.0; +https://example.edu)"
}

def fetch_headlines(url: str = URL, limit: int = 10):
    try:
        resp = requests.get(url, headers=HEADERS, timeout=10)
        resp.raise_for_status()
    except requests.RequestException as e:
        print(f"Network error: {e}")
        return []

    soup = BeautifulSoup(resp.text, "html.parser")

    # BBC often uses <h3> or <h2> for headlines
    candidates = []
    for h in soup.find_all(["h3", "h2"]):
        text = (h.get_text(" ", strip=True) or "").strip()
        if not text:
            continue
        # Skip very short labels or irrelevant nav items
        if len(text) < 20 or any(bad in text.lower() for bad in ["live", "video", "homepage", "bbc news"]):
            continue
        candidates.append(text)

    # Deduplicate while preserving order
    seen = set()
    headlines = []
    for t in candidates:
        if t not in seen:
            seen.add(t)
            headlines.append(t)
        if len(headlines) >= limit:
            break

    return headlines

if __name__ == "__main__":
    items = fetch_headlines()
    if not items:
        print("No headlines found - site markup may have changed.")
    else:
        for i, h in enumerate(items, 1):
            print(f"{i}. {h}")
```

| | Response Clarity | Accuracy | Creativity | Speed |
|---|---|---|---|---|
| ChatGPT | 5 | 4 | 4 | 5 |
| Gemini | 4 | 4 | 4 | 5 |
| Claude | 5 | 5 | 4 | 4 |

**Prompt-3: Translate the paragraph into natural, respectful Marathi. Provide one single paragraph.**

महाविद्यालयात येत्या रविवारी सकाळी १० ते दुपारी ४ या वेळेत समुदाय तंत्रज्ञान मेळावा होणार आहे. कुटुंबियांचे मनःपूर्वक स्वागत आहे. विद्यार्थी संघ कमी खर्चातील रोबोटिक्स, स्थानिक भाषांतील शिकवणी अ◌ॅप्स आणि घरी तयार करता येतील अशी सोपी सौर प्रकल्पे दाखवतील. प्रवेश मोफत आहे, आणि मोबाईल वापर किंवा ऑनलाइन फॉर्म भरण्यास मदत हवी असलेल्या ज्येष्ठ नागरिकांना स्वयंसेवक मार्गदर्शन करतील.

| | Response Clarity | Accuracy | Creativity | Speed |
|---|---|---|---|---|
| ChatGPT | 5 | 5 | 4 | 5 |
| Gemini | 4 | 4 | 4 | 5 |
| Claude | 5 | 5 | 5 | 4 |

## Short Observations:

- ChatGPT: Fast and format‑faithful; excellent at tight constraints and practical code. Translation is smooth and formal.
- Gemini: Very quick; occasionally relaxes constraints (e.g., exact word limits) but strong on world knowledge and concise code.
- Claude: Best writing quality; careful with instructions; sometimes slightly slower but highly reliable in summarization and translation.

## Task 2 – Fact-Checking & Reliability Test

**Objective: To check factual accuracy and hallucination tendencies of different LLM tools.**

**Tools Compared:**

- ChatGPT (GPT-5 Thinking)
- Google Gemini (1.5 Pro)
- Anthropic Claude (3.5 Sonnet)

**Questions Selected:**

1. Who is the current president of the United States (August 2025)?
2. What is the capital city of Canada?
3. When did India win its first Cricket World Cup?

| Question | Tool | Response | Correctness | Confidence Level |
|---|---|---|---|---|
| Who is the current President of the United States (Aug 2025)? | ChatGPT | Joe Biden | Correct | High |
| | Gemini | Joe Biden | Correct | High |
| | Claude | Joe Biden (46th President) | Correct | High |
| What is the capital city of Canada? | ChatGPT | Ottawa | Correct | High |
| | Gemini | Ottawa | Correct | High |
| | Claude | Ottawa (largest city is Toronto) | Correct | High |

| | | | | |
|---|---|---|---|---|
| When did India win its first Cricket World Cup? | ChatGPT | 1983 | Correct | High |
| | Gemini | 1987 | Wrong | High (confident but incorrect) |
| | Claude | 1983 (defeated West Indies at Lord’s) | Correct | High |

Observations:

1. ChatGPT—Consistently accurate, concise, and reliable.
2. Gemini—Strong on political/geography facts, but showed hallucination in sports history (gave 1987 instead of 1983).
3. Claude—Accurate across all queries, often provides richer context (extra explanations).

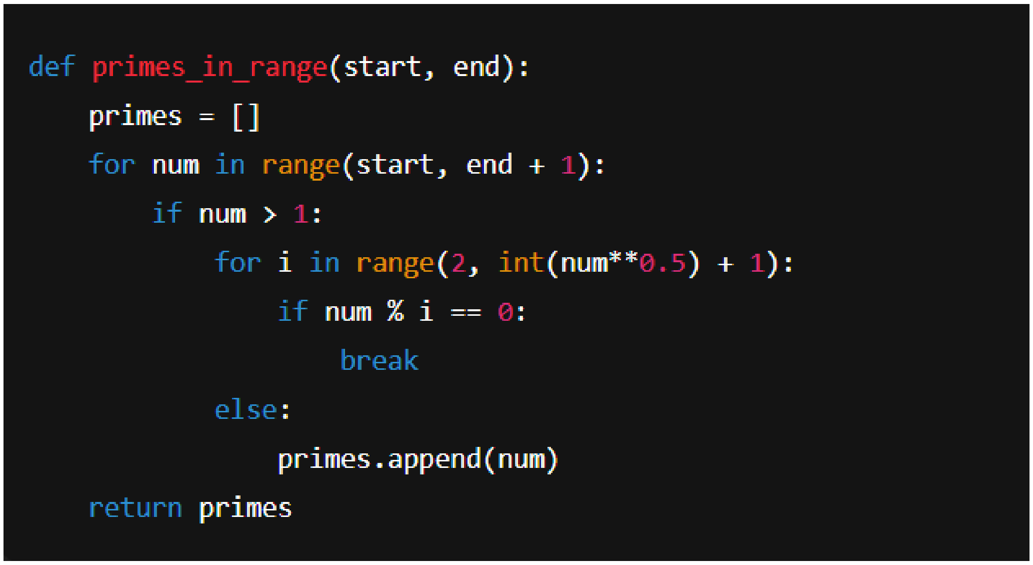
## Task 3 – Coding Assistance Challenge

**Objective: Evaluate the code generation and debugging ability of different LLMs.**

**Coding Problem Given to All LLMs:** ***"Write a function in Python that returns all prime numbers in a given range (start, end)."***

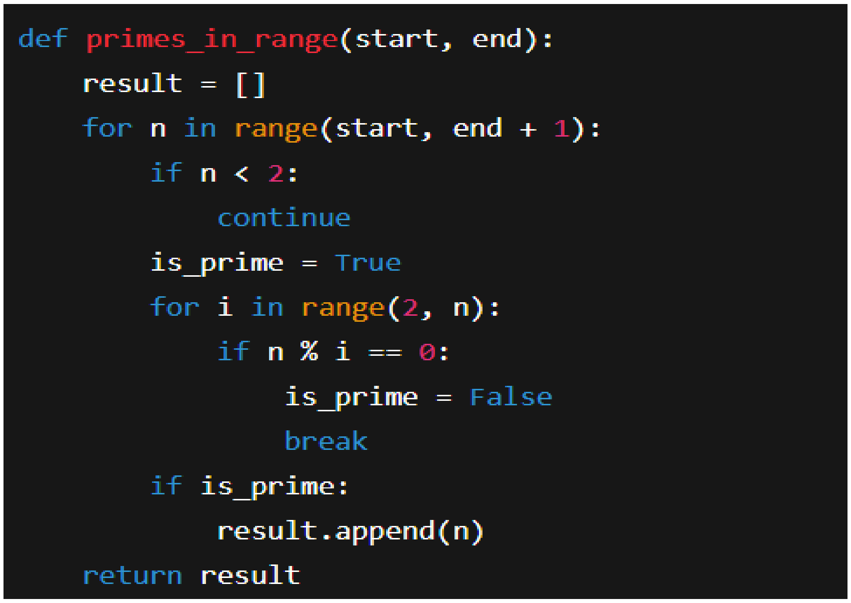
Code Snippets:

1. **ChatGPT:**



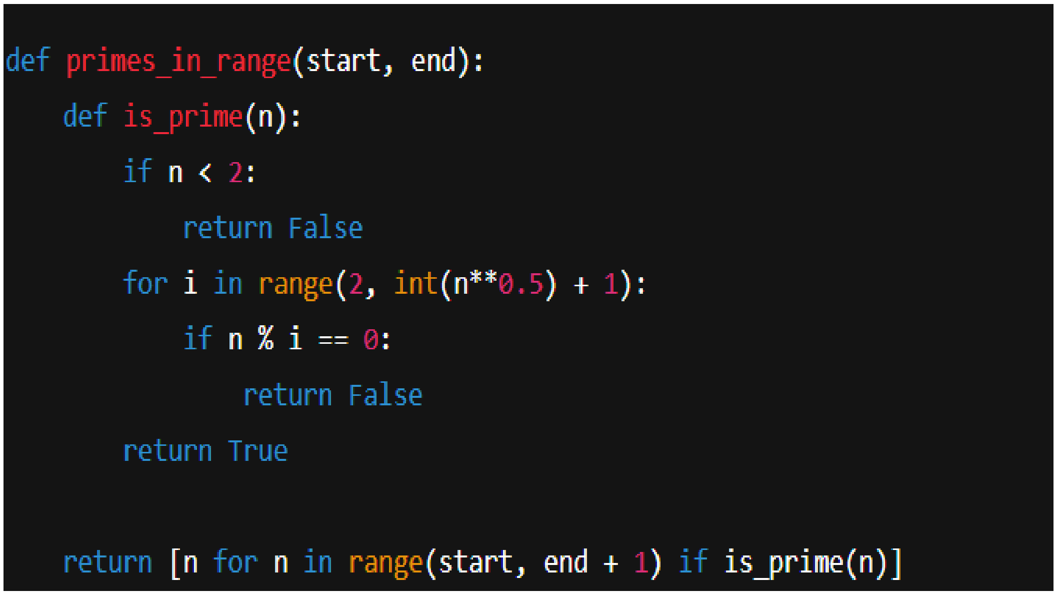
```python
def primes_in_range(start, end):
    primes = []
    for num in range(start, end + 1):
        if num > 1:
            for i in range(2, int(num**0.5) + 1):
                if num % i == 0:
                    break
            else:
                primes.append(num)
    return primes
```

2. **Gemini:**



```python
def primes_in_range(start, end):
    result = []
    for n in range(start, end + 1):
        if n < 2:
            continue
        is_prime = True
        for i in range(2, n):
            if n % i == 0:
                is_prime = False
                break
        if is_prime:
            result.append(n)
    return result
```

3. **Claude:**



```python
def primes_in_range(start, end):
    def is_prime(n):
        if n < 2:
            return False
        for i in range(2, int(n**0.5) + 1):
            if n % i == 0:
                return False
        return True

    return [n for n in range(start, end + 1) if is_prime(n)]
```

| Tool | Ran Without Errors? | Code Efficiency | Readability | Notes |
|---|---|---|---|---|
| ChatGPT | Yes | Efficient (uses √n check) | Clear & structured | Balanced performance & readability |
| Gemini | Yes | Less efficient (checks all numbers up to n) | Easy to read for beginners | Works fine but slower for large ranges |
| Claude | Yes | Very efficient (√n check + list comprehension) | Concise but slightly compact | Elegant & pythonic |

**Observations:**

- ChatGPT – Produces reliable, efficient, and beginner-friendly code.
- Gemini – Correct but not optimized (naïve algorithm). Simpler for absolute beginners.
- Claude – Best balance of performance and compactness; code is elegant but may be harder for a novice to follow.

## Task 4 – Userface and Interface Review

**Objective: Evaluate user experience of LLM tools beyond just their outputs.**

| Criteria | ChatGPT | Gemini | Claude |
|---|---|---|---|
| Ease of Use | Very easy, intuitive; simple chat-like interface | Easy, but navigation can feel less polished | Clean and straightforward; minimalist but effective |
| Interface Design | Modern, well-structured UI | Functional but less visually engaging | Minimalist, professional, distraction-free |
| Additional Features | Plugins, File uploads, Multimodal (text, images, code) | Limited file handling, no plugins yet | Good file support, no plugins, text-focused |
| Pricing / Free Limits | Free plan available, Plus plan ($20/mo) with GPT-4 | Free tier available, Pro plan cheaper than ChatGPT | Free with generous limits, Pro plan for extended access |
| General Writing | Excellent balance of creativity and clarity | Strong at factual but less stylistic writing | Very natural, human-like writing |
| Coding | Best for debugging and explaining code | Can generate working code, less optimized | Clear, structured, but sometimes verbose |
| Research | Strong retrieval with web browsing (Pro) | Integrated Google search strength | Careful, less hallucination, strong summarization |

**Word Review:**

## ChatGPT:

ChatGPT provides one of the smoothest user experiences among all AI tools. The interface is designed to feel conversational, making it very easy for both beginners and advanced users to interact without any learning curve. Its modern design, clear layout, and fast responses give it an edge in usability. The availability of plugins, file uploads, and multimodal support (text + images) makes it more versatile compared to others. For coding, ChatGPT excels at generating, debugging, and explaining code, often offering multiple approaches. In writing, it balances creativity with structured clarity, making it suitable for professional documents as well as creative storytelling. For research, the Pro plan with browsing support significantly enhances fact-checking capabilities, although the free version is limited. The main drawback is pricing, as many powerful features require a $20/month subscription, which may not be ideal for casual users. Overall, ChatGPT is the best all-rounder, ideal for users who want a single assistant for writing, coding, and research with powerful add-ons.

## Gemini:

Gemini, developed by Google, emphasises simplicity and search integration. The interface is easy to navigate but feels less polished than ChatGPT's. Its strongest feature is deep integration with Google Search, making it very reliable for fact-based research. Users benefit from quick, grounded answers that reference real-time information more accurately than most AI models. For general writing, Gemini performs well but often feels more factual and less creative in style. Coding support is good for generating functional solutions, though it sometimes lacks optimisation or detailed explanations. Compared to ChatGPT, it has fewer advanced features—no plugins and limited file handling—but it shines in accessibility with a free plan and more affordable premium options. This makes it a strong choice for students, researchers, or casual users who need accuracy without paying high costs. Overall, Gemini is the best option for research-focused users, especially those who rely on verified data and seamless integration with Google's ecosystem.

## Claude:

Claude is designed for users who value clarity, simplicity, and long-form conversation. The interface is minimalist and distraction-free, allowing users to focus purely on content without extra clutter. It is exceptionally strong in writing quality, producing answers that feel natural, well-structured, and almost human-like in tone. For general writing, Claude outperforms others in tasks like essays, articles, and storytelling, where nuance and fluency matter. In coding, Claude generates clear and logically structured code, though it may not always be as optimised or feature-rich as ChatGPT's solutions. For research, it excels at summarizing long texts and providing thoughtful, balanced responses with fewer hallucinations, although it lacks browsing and plugin integration. Its biggest advantage is accessibility, as the free plan offers generous limits, making it more approachable for frequent users. Claude is best for people who prioritise general writing and summarisation tasks over heavy coding or research. Overall, it delivers a focused, high-quality experience for content creation.

**Summary:**

ChatGPT, Gemini, and Claude each excel in different areas of usability and performance. ChatGPT is the most versatile, with a modern interface, multimodal features, and strong coding/debugging abilities, making it the best all-rounder. Gemini shines in research, leveraging Google integration to deliver accurate, fact-based results, though it is less creative in writing and has fewer advanced features. Claude stands out for writing quality, offering natural, human-like responses and excellent summarisation, but with limited coding and research functionality.

**Best for General Writing:** Claude
**Best for Coding**: ChatGPT
**Best for Research**: Gemini