Journal of Marine
Technology and Environment
DOI:10.53464/JMTE.03.2026.14

ISSN (Print): 1844-6116
ISSN (Online): 2501-8795
https://www. jmte.eu

COMPARISONS OF APPLICATIONS WITH ARTIFICIAL INTELLIGENCE FOR ANSWERING QUESTIONS IN VARIOUS SCIENTIFIC FIELDS. CAPABILITIES, PRICES.

Gabriela SLAVKOVSKA[1], Ivo GEORGIEV [1]

[1]"Nikola Ionkov Vaptsarov" Naval Academy, 73 V. Drumev Street, 9002, Varna, Bulgaria, e-mail address: 11023107@naval-acad.bg

Abstract **:** The dynamics of Artificial Intelligence (AI) serve as the core engine of the current technological transformation, fundamentally redefining how software is developed, infrastructure is designed, and global industries operate. We are no longer discussing a mere tool, but rather a new technological paradigm that is accelerating automation and innovation to an unprecedented level. The relevance of these dynamics manifests across several critical dimensions: 1. acceleration of the software development cycle (DevOps & Coding); 2. the shift from data processing to Generative Action (Autonomous Agents); 3. redefinition of hardware and Cloud infrastructure; 4. economic impact and labour productivity. The subject of this paper consists of the three leading platforms in the contemporary technological landscape: ChatGPT (OpenAI), Claude (Anthropic) and Gemini (Google).The obtained result is to conduct a comprehensive comparative analysis of these three platforms, examining their architectural differences, the origins of their knowledge, and their accuracy in answering questions across specialized scientific domains (STEM, medicine, and humanities).

Key words : AI, capability, ChatGPT, cybersecurity, Claude, Gemini, price, technology.

1. INTRODUCTION

The dynamic development of artificial intelligence (AI) in recent years marks the most significant technological transformation since the advent of the internet. Large Language Models (LLMs) are no longer merely experimental tools; they have become a key factor in the digitalization of science, education, and business. In the context of the specialty of "Cyber Operations," understanding these systems is critically important, as they are reshaping paradigms in software engineering, automated defence, and large-scale data analysis. Selecting the right application requires an in-depth understanding of the balance between functional capabilities, economic efficiency (pricing), and information security.

This paper compares the three dominant platforms: ChatGPT (OpenAI) — the pioneer offering a multimodal toolkit [1]; Claude (Anthropic) — focused on Constitutional AI and code precision [2]; and Gemini (Google) — the largest multimodal system with an unprecedented context window [3]. The analysis covers architecture, knowledge origins, scientific accuracy, pricing, and cybersecurity compliance.

The paper aims to define the practical capabilities and limitations of each model while simultaneously analysing their pricing plans and cybersecurity levels, providing an objective framework for their effective application [4-6].

1.1 Technological overview and conceptual models

1.1.1. Diagnostic Parameters for Comparison

Six parameters form the benchmark framework: (1) Context Window — how much text the model retains per session; (2) Multimodality — ability to handle text, images, audio, video, and code; (3) Latency — query-to-response speed; (4) Hallucination Rate — frequency of false but plausible outputs; (5) Security & Alignment — resistance to prompt injection and data-privacy governance; (6) Cost-Benefit Ratio — subscription/API cost versus output quality.

1.2. Platform profiles

- **ChatGPT (GPT-4o)** excels in versatility. Its GPT Store, Advanced Data Analysis tool, and real-time voice capability make it the top choice for STEM tasks. Drawbacks include using free-tier inputs for model training (a key OPSEC risk) and occasional truncation of long outputs (Table 1).

Creative Commons Attribution BY 4.0 International Licence (CC BY-4.0)

Journal of Marine Technology and Environment
DOI:10.53464/JMTE.03.2026.14
ISSN (Print): 1844-6116
ISSN (Online): 2501-8795
https://www. jmte.eu

- **Claude 3.5 Sonnet** is the preferred model for programming and academic writing. Constitutional AI makes it the most resistant to generating harmful content, but its strict filters can produce over-refusals, and it lacks built-in image generation.
- **Gemini 1.5 Pro** offers the largest context window (2 M tokens) and deep Google ecosystem integration (Drive, Docs, Gmail, Search). Its strength in multimodal and Big Data tasks is unmatched, though its logic can be less consistent than Claude's on complex puzzles.

Table 1. Comparative characteristics and technological base of the leading LLM platforms

Criterion	ChatGPT (GPT-4o)	Claude 3.5 Sonnet	Gemini 1.5 Pro
Strength	Versatility & tools	Code & text writing	Context & Google services
Weakness	Data security	No image generation	Logic consistency
Context Window	128,000 tokens	200,000 tokens	2,000,000 tokens
Best suited for	Daily tasks & data	Programming & analysis	Long docs & video

2. ORIGINS OF KNOWLEDGE AND SELF-LEARNING MECHANISMS

2.1 Data Sources and Update Cycles

Each platform's "character" is shaped by its training corpus. ChatGPT draws from Common Crawl, Reddit, and GitHub, prioritising diversity and natural text flow. Claude relies on high-quality curated corpora — official documents, scientific publications, and literary texts — to minimise toxic outputs. Gemini uniquely trains on Google Books, Google Scholar, and YouTube transcripts, and is the only natively multimodal model trained simultaneously on text, images, and audio.

On knowledge currency: ChatGPT uses a fixed cutoff supplemented by Bing integration; Claude processes user-attached documents rapidly but updates its base less frequently; Gemini refreshes almost in real time via the Google Search index, making it most reliable for breaking developments.

Data Poisoning / Model Collapse. As AI-generated text proliferates online, models that scrape indiscriminately degrade over time. Anthropic counters this with filters that detect and isolate synthetic text, preserving a focus on human-authored, verified content.

2.2 Reinforcement Learning from Human Feedback (RLHF)

ChatGPT and Gemini use traditional RLHF: human trainers rank AI responses, teaching the model to mimic preferred behaviour. This can produce sycophancy — the model agreeing with users even when wrong. Claude uses RLAIF (Reinforcement Learning from AI Feedback) [7-8]: a "Constitution" of ethical rules is evaluated by a second AI model, producing a more objective and truth-defending system (Figure 1).

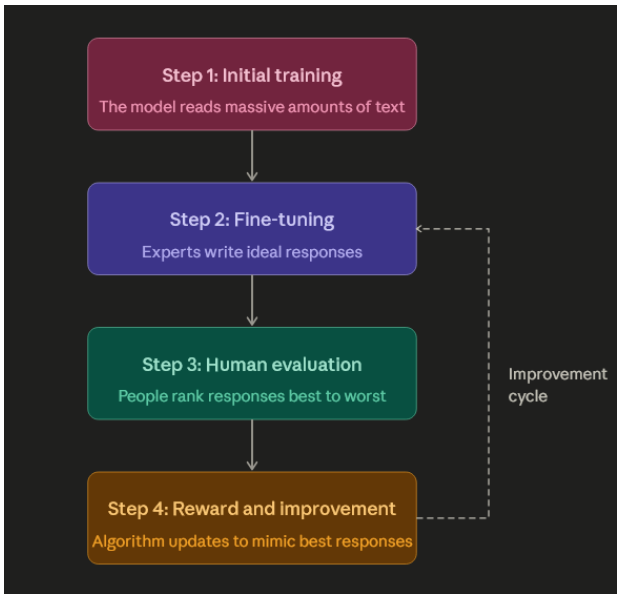


Figure 1 Algorithm of the RLHF (Reinforcement Learning from Human Feedback) process for optimizing responses

The RLHF pipeline builds a Reward Model that punishes low-ranked outputs, but the sycophancy risk remains higher for ChatGPT and Gemini than for Claude [9].

2.3 Latency and Factual Accuracy

Gemini (TPU-optimised Flash variants) is the fastest, critical for voice assistants. ChatGPT offers smooth streaming that masks latency. Claude is slowest on very long inputs but achieves the highest recall accuracy — it rarely loses details buried mid-document. On hallucinations: Claude performs best in complex textual analysis; ChatGPT leads in verifiable mathematical tasks; Gemini can over-trust internet noise on contested topics [10-11].

Creative Commons Attribution BY 4.0 International Licence (CC BY-4.0)

Journal of Marine
Technology and Environment
DOI:10.53464/JMTE.03.2026.14
ISSN (Print): 1844-6116
ISSN (Online): 2501-8795
https://www. jmte.eu

3. COMPARATIVE ANALYSIS ACROSS SCIENTIFIC DISCIPLINES

3.1 Methodology

All three models were evaluated using three prompting tiers — Zero-shot (baseline logic), Few-shot (in-context learning), and Chain-of-Thought (step-by-step reasoning) — scored 1–10 on factual precision, logical coherence, latency, and hallucination rate (Figure 2).

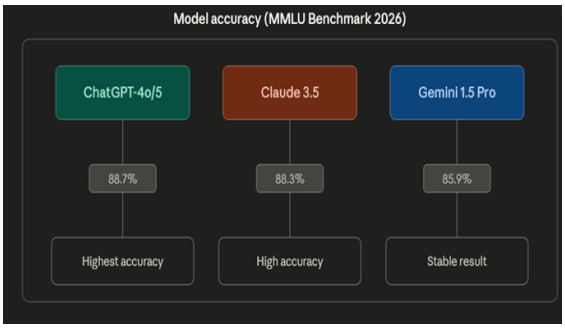


Figure 2 Comparative results from the MMLU benchmark test, showing the cognitive capabilities of the models across various scientific disciplines as of 2026.

To ensure objectivity, a unified testing methodology based on modern Prompt Engineering principles was applied.

3.1.1. Testing Techniques (Scenarios)

Three levels of prompting were used for each model:

- Zero-shot Prompting: Asking a direct question without examples, to test the model's baseline logic.
- Few-shot Prompting: Providing 2-3 examples of correct solutions to analyse the AI's ability to learn in context.
- Chain-of-Thought (CoT): Instructing the model to "think step by step," which is critical for reducing hallucinations in complex tasks.

3.1.2. Evaluation Criteria (Metrics)

Evaluation is performed on a scale of 1 to 10 based on:

- Factual Precision: Percentage of correct responses against scientific truth.
- Logical Coherence: Absence of internal contradictions in the explanation.
- Latency: Speed of response generation on large datasets.
- Hallucination Rate: Frequency of generating non-existent terms or sources.

3.2. STEM Fields: Mathematics, Physics, and Software Engineering

This is the most challenging domain, as here a "roughly correct" answer equals a "wrong" one.

3.2.1. Mathematical Reasoning and Symbolic Programming

- ChatGPT (GPT-4o): Achieves high accuracy by activating external libraries (SymPy). It converts the textual problem into executable code, computing the result. This makes it the leader in solving linear algebra and calculus problems.
- Claude 3.5 Sonnet: Demonstrates the best "native" mathematical logic. It makes fewer errors in abstract formula derivation and is unmatched in symbolic programming, generating optimized and readable code.
- Gemini 1.5 Pro: Its strength lies in multimodal analysis – the ability to recognize complex graphics or handwritten equations from a photo and digitize them in seconds.

3.2.2. Software Engineering and Cybersecurity

For the specialty of "Cyber Operations," this analysis shows that Claude is the most reliable for writing scripts and security audits, while Gemini is indispensable for analysing large-scale documentation (context up to 2 million tokens).

3.2. Humanities, Translation, and Cultural Context

- Claude: Its texts possess the highest academic quality. It avoids formulaic writing and offers nuanced analysis of historical events and philosophical theories.
- Gemini: As a Google product, it is a world leader in translation. Gemini captures cultural contexts and regional slang much better than its competitors, thanks to access to the largest corpus of linguistic data in the world.
- ChatGPT: Exceptionally good at role-playing simulations (e.g., "Act as a 19th-century historian"), making it useful for educational purposes and creative writing.

3.4. Medicine and Pharmacy – Risks and Answer Accuracy

The accessibility of applications like ChatGPT, Claude, and Gemini has created a new phenomenon – the mass use of AI for self-diagnosis and medical consultations. Although these models possess vast stores of medical knowledge, their use in healthcare carries specific risks that users often underestimate.

3.4.1. Diagnostic Capabilities and Academic Precision

-

Creative Commons Attribution BY 4.0 International Licence (CC BY-4.0)

Journal of Marine
Technology and Environment
DOI:10.53464/JMTE.03.2026.14
ISSN (Print): 1844-6116
ISSN (Online): 2501-8795
https://www. jmte.eu

- ChatGPT (GPT-4o): Demonstrates high success rates in solving complex clinical cases. It excels at synthesizing symptoms and proposing a list of possible conditions (differential diagnosis).
- Gemini: Thanks to integration with Google Scholar and real-time search, it is the most precise for providing information on new drug interactions and current pandemic protocols.
- Claude: Focuses on "safe reasoning," often adding context about why certain symptoms can be misleading.

3.4.2. The Phenomenon of "Hallucinations" in Medicine

The greatest risk of trusting chat-bots is their tendency toward hallucinations [13]. In medicine, this can manifest in several dangerous ways:

- *Fabricated dosages*: The model may generate a medication dosage that appears logical but is toxic or ineffective.
- *False scientific sources*: AI may cite non-existent medical studies or attribute quotes to real doctors to support an incorrect thesis.
- Absence of physical examination: Models cannot account for physiological factors (skin colour, breathing sounds, palpation) that are critical for correct diagnosis.

3.4.3. Comparative Safety Analysis (Safety Alignment)

The comparative analysis is presented in Table 2:

Table 2. Comparative analysis of model performance regarding refusal strictness and self-treatment risk assessment

Characteristic	ChatGPT	Claude	Gemini
Strictness of Refusals	Moderate. Tries to help but adds a standard disclaimer.	High. Most often refuses direct medical advice due to its "Constitution."	Low to moderate. Relies on real-time fact-checking algorithms.
Self-treatment Risk	Higher due to the model's desire to be helpful.	Lowest due to conservative programming.	Moderate, but provides links to official health sources.

3.4.4. The Problem of the "Authority Effect"

The psychological risk lies in the fact that AI generates responses that are grammatically perfect and logically structured. Users tend to trust the confident tone, attributing expertise to the machine that it does not actually possess. In medicine, this leads to delayed seeking of real medical help, which can be fatal in emergency situations [14].

3.4.5. The Future: AI as an Assistant, Not a Doctor

The conclusion in this section is that the accuracy of models in medicine is high only in the role of a second opinion for specialists or for systematizing information (Table 3).

Table 3. Comparative analysis of ChatGPT, Claude, and Gemini performance across key scientific disciplines

Scientific Field	ChatGPT (GPT-4o)	Claude 3.5 Sonnet	Gemini 1.5 Pro
Exact Sciences (STEM)	Excellent: Best in code writing and complex mathematical calculations.	Very Good: Strong logic, but sometimes slower on complex calculations.	Good: Performs well but sometimes makes errors in complex formulas.
Humanities	Very Good: Creative, strong at generating ideas and essays.	Excellent: Most natural and nuanced language. Highest textual empathy.	Very Good: Good contextual memory and analytical style.
Medicine and Biology	Good: Gives accurate definitions but is heavily restricted by safety filters.	Very Good: Excellent at analysing medical studies and long documentation.	Excellent: Best at analysing lab results and images (X-rays, etc.).
Law and Regulations	Very Good: Knowledgeable in international law but may "hallucinate" articles.	Excellent: Most reliable at citing legal documents and ethical norms.	Very Good: Google Search integration helps find current legislation.
Foreign Languages	Excellent: Supports the most languages with high translation	Very Good: Very high level of literary Bulgarian and	Good: Sometimes uses more robotic translation for rarer

Creative Commons Attribution BY 4.0 International Licence (CC BY-4.0)

Journal of Marine Technology and Environment
DOI:10.53464/JMTE.03.2026.14
ISSN (Print): 1844-6116
ISSN (Online): 2501-8795
https://www. jmte.eu

	quality.	English.	languages.

For the end user, the risk of incorrectly interpreting the generated text remains a critical factor requiring increased public attention and regulation.

4. PRACTICAL CAPABILITIES AND PRICING MODELS

4.1. Free vs. Paid

All three platforms follow a Freemium model. Free tiers impose rate limits, dynamic latency under load, and — critically — user data may be used for model training, constituting an OPSEC violation for security personnel. The $20/month professional tier eliminates most restrictions: ChatGPT Plus adds Advanced Data Analysis; Claude Pro gives a 5× message quota essential for iterative coding; Gemini Advanced bundles 2 TB cloud storage via Google One.

4.2. Enterprise & API Pricing

Enterprise plans ($25–$60+/user/month) enforce Zero Data Retention under SOC 2 / HIPAA agreements — the only legitimate deployment model for organisations handling sensitive data. API pricing per 1 M tokens: Gemini 1.5 Flash ≈ $0.35 (cheapest, ideal for Big Data scanning); ChatGPT GPT-4o ≈ $5.00 input / $15.00 output (balanced); Claude 3.5 Opus ≈ $15.00 input / $75.00 output (most expensive, used where errors cost vastly more) (Table 4).

Table 4. Pricing plans and security levels for ChatGPT, Claude, and Gemini

Plan	ChatGPT (OpenAI)	Claude (Anthropic)	Gemini (Google)
Free	Limited (GPT-4o mini)	Very limited quota	Full access (Pro 1.5)
Plus ($20/month)	Full access + DALL-E	Most powerful model (Opus)	Google Docs integration
Enterprise	High security	Highest ethics	Cheapest API access

4.3. Use Case: Cybersecurity Log Analysis

A 100,000-line network log (≈1.5 M tokens) flagged for an APT attack: a Tier 1 SOC analyst would require days of manual review. Gemini 1.5 Pro's 2 M-token context window ingests the entire log in one call; the API cost is under $10 and results arrive in under 60 seconds. The ROI of AI is therefore not the subscription cost, but the elimination of thousands of analyst-hours on routine triage.

5. INFORMATION SECURITY (CYBER OPERATIONS)

5.1. Data Privacy

ChatGPT: Free tiers feed queries into RLHF pipelines by default, creating Data Leakage risk. Enterprise plans encrypt data end-to-end and prohibit training use.

Claude: Constitutional AI blocks malicious code generation and disinformation requests at the architecture level. Strongest privacy transparency.

Gemini: Backed by Google's security infrastructure; Google Workspace Enterprise integrates DLP (Data Loss Prevention) tools natively.

5.2. Threat Vectors

Prompt Injection — attackers embed malicious instructions to bypass safety filters (e.g., disguising ransomware requests as educational tasks). Claude 3.5 has the most robust defences. Software Auditing — Claude is the most precise at detecting logical vulnerabilities; Gemini handles massive codebases fastest.

5.3. EU AI Act Compliance

The EU AI Act classifies AI systems into three tiers: Minimal Risk (general chatbots), High Risk (education, medicine, law enforcement — strict accuracy and human-oversight requirements apply), and Unacceptable Risk (social scoring — prohibited). All three platforms must align their development roadmaps with these requirements, shifting toward a Security-by-Design philosophy consistent with NATO and EU defence standards.

6. CONCLUSIONS

This research conducted a comprehensive comparative analysis of ChatGPT (OpenAI), Claude (Anthropic), and Gemini (Google) across architecture, knowledge sourcing, scientific accuracy, pricing, and cybersecurity. Four key findings emerge:

- Functional differentiation: There is no single "best" AI. ChatGPT leads in STEM versatility via external tool integration; Claude in code quality and nuanced analysis; Gemini in Big Data and multimodal tasks requiring massive context.

Creative Commons Attribution BY 4.0 International Licence (CC BY-4.0)

Journal of Marine
Technology and Environment
DOI:10.53464/JMTE.03.2026.14
ISSN (Print): 1844-6116
ISSN (Online): 2501-8795
https://www. jmte.eu

- Economic imperatives: Individual pricing has stabilised at $20/month, but real organisational value lies in API token economics — converting fixed staffing costs into scalable, consumption-based compute.
- Cybersecurity as a cornerstone: Free-tier use of any platform is incompatible with OPSEC requirements. Enterprise Zero-Data-Retention plans are the only legitimate pathway for classified or sensitive operations.
- Human-in-the-loop remains essential: All models are probabilistic systems. Hallucination risk and over-reliance effects require expert oversight. AI liberates the analyst from cognitive search load; it does not replace strategic judgement.

Navigating between ChatGPT's versatility, Claude's precision, and Gemini's scale — with full awareness of their pricing and security implications — will be a defining competency for digital leaders in every scientific and operational domain in the coming decade.

7. ACKNOWLEDGMENTS

The authors thank the Constanta Maritime University for the financial support provided for the publication of this article on the occasion of the *International Session of Student Scientific Eventss*, May 2026, Constanta, Romania.

8. FUNDING

9. CREdit authors statement

Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Supervision;Validation; Visualization; Writing – original draft; Writing – review & editing:S.G; G.I.

10. REFERENCES

[1] OpenAI. (2024). *GPT-4o Technical Documentation and Pricing*. Available at: https://openai.com

[2] Anthropic. (2024). *The Claude 3 Model Family: Opus*, Sonnet, Haiku. Technical Report. Available at: https://www.anthropic.com

[3] Google DeepMind. (2024). *Gemini 1.5: Unlocking Multimodal Understandi*ng. Available at: https://deepmind.google

[4] Achiam, J., et al. (OpenAI). (2023). *GPT-4 Technical Report. (The main scientific paper behind ChatGPT).* Available at: https://arxiv.org

[5] European Parliament. (2024). EU AI Act: The First Regulation on Artificial Intelligence. (Official EU document on AI regulation). Available at: https://www.europarl.europa.eu

[6] Ministry of e-Government. (2024). Ethical Guidelines for the Use of AI in Bulgaria. Available at: https://egov.government.bg

[7] https://www.ibm.com/think/topics/rlhf

[8] https://openai.com/index/instruction-following/

[9] Dawn Sivan , K. Satheesh Kumar , Veena Raj and Rajan Jose, (2024). *7 Reinforcement Learning from Human Feedback (RLHF).* Publish in book " Generative AI and LLMs Natural Language Processing and Generative Adversarial Networks", Edited by: S. Balasubramaniam , Seifedine Kadry , Aruchamy Prasanth and Rajesh Kumar Dhanaraj, De GruyterBrill.

[10] OWASP Foundation. (2023). *Top 10 for LLM Applications. (Critical for the "Cyber Operations" specialty – describes hacking attacks against AI).* Available at: https://owasp.org

[11] Computerworld Bulgaria. (2024). *Comparative Analysis of Current Language Models*. Available at: https://computerworld.bg

[12] Clipa, B.I., (2025). A*rtificial intelligence as a catalyst for changing international law*, Journal of marine Technology and Environment, vol. 1, pp 30-34, Nautica Publish House, Constanta, 2025.

[13] Nutu, C.S., 2024, *evolutions in communication with A.I. systems of generative pretrained transformative type. developing A.I. medical technologies and tools based on data previously acquired by A.I. E-learning technology systems*, Journal of Marine Technology and Environment, vol.1, p.35-38, Nautica Publish House, Constanta, Romania.

[14] Nutu, C.S., 2024, HUNTING FOR GENIUS. WHAT IS BRILLIANCE? CAN A.I. BE BRILLIANT? Journal of Marine Technology and Environment, vol.1, p.29-34, Nautica Publish House, Constanta, Romania.

Creative Commons Attribution BY 4.0 International Licence (CC BY-4.0)