World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

Check for updates

Comparative Review of Modern AI Chatbots: Capabilities, design, and real-world applications

Kingsley Olunosen Osobase [1, *], Chijioke Cyriacus Ekechi [2], Aminat Oluwatimileyin Akinode [3], Adetola Elizabeth Adesanoye [4], Toluwanimi Williams Olatokun [5] and Chibuzo Lasbrey Opara [6]

[1] *Federal University of Technology, Akure, Meteorology, Akure, Ondo, Nigeria.*
[2] *Department of Electrical and Computer Engineering, Tennessee Technological University, Cookeville, Tennessee, USA.*
[3] *Olabisi Onabanjo University, Computer Engineering, Ago-Iwoye, Ogun, Nigeria.*
[4] *Department of Business Administration, Ahmadu Bello University, Zaria, Kaduna State Nigeria.*
[5] *Abiola Ajimobi Technical University, Mechanical and mechatronics engineering, ibadan, Oyo state, Nigeria.*
[6] *Federal University of Technology Owerri, Computer Science*

World Journal of Advanced Research and Reviews, 2025, 27(03), 097-112

Publication history: Received on 12 July 2025; revised on 23 August 2025; accepted on 25 August 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.27.3.2865

Abstract

Artificial intelligence chatbots are rapidly reshaping how individuals interact with information, tools, and workflows, spanning from code generation and document editing to policy drafting and academic search. This review explores the architecture, training philosophies, and deployment contexts of six prominent chatbot systems: ChatGPT, Gemini, Claude, Meta AI, GrammarlyGO, and Joules. At the heart of these systems are transformer-based or hybrid neural models trained on vast corpora and fine-tuned using reinforcement learning, instruction prompts, or constitutional principles. We compare these models across core features such as reasoning ability, multimodal capacity, user control, and professional integration. Beyond static comparison, we evaluate their real-world utility in coding, education, writing, compliance, and messaging environments, using a capability matrix and use-case mapping framework. The analysis also addresses deeper design tensions: autonomy versus oversight, generalization versus specialization, and transparency versus performance. Echoing lessons from our previous work on adaptive resilience in plant systems under dual stress, we argue that the future of chatbot intelligence lies not in scale alone but in functional alignment, collaboration with human judgment, and deployment sensitivity. The review concludes with key directions in unified multimodal agents, on-device reasoning, and ethical co-design, calling for a more plural, professional, and verifiable approach to next-generation chatbot development.

Keywords: Artificial intelligence chatbots; Large language models; Multimodal agents; Transformer architecture; Chatbot comparison; Retrieval-augmented generation; Human–AI collaboration; Constitutional AI; Enterprise deployment; Chatbot transparency

1. Introduction

Artificial Intelligence (AI) chatbots have shifted from novelty to necessity in the span of just a few years. What began as scripted, rule-based assistants, capable of only predefined replies, has now evolved into intelligent, highly responsive systems capable of natural language understanding, reasoning, and adaptive learning [1]. These chatbots, powered by increasingly sophisticated large language models (LLMs), are no longer limited to basic FAQs or customer support. They now support users in writing, coding, research, creative storytelling, mental health dialogue, and even decision-making in legal, medical, and enterprise environments [5,6].

* Corresponding author: Kingsley Olunosen Osobase

At the heart of this transformation is the rapid advancement of LLMs, neural architectures trained on massive volumes of diverse text and, in some cases, multimodal data such as images, audio, or code [7]. Models like OpenAI's GPT-4, Google's Gemini, Anthropic's Claude, and Meta's LLaMA have demonstrated the capacity to generalize across a wide range of tasks, generate coherent and context-aware responses, and engage in multi-turn conversations that closely mimic human discourse [8]. This ability to generate human-like output with minimal prompt engineering has accelerated their adoption across sectors and sparked intense interest in both academic and industrial communities [9].

The applicability of AI chatbots is especially evident in its integration into the daily set of tools and platforms. Most of them have been incorporated in search engines, office packages, messaging programs, operating systems, and cell phones [10]. They are used by businesses to automate both customer services and in house documentation. They are applied in learning by educators to have interactive learning [11]. Authors employ them to brainstorm or perfect the drafts [13]. Even programmers are now using chatbots to auto-code lines, debug, and simplify the words of complicated algorithms [12]. These interactions signal a further deeper paradigm shift in knowledge access, collaboration with machines, and human productivity power up.

Their growing influence also raises important questions about design philosophy, safety, bias, access, and purpose. In contrast to conventional software, LLM chatbots work probabilistically, that is, they predict the best next token, and thus can occasionally hallucinate or end up with biased, outdated, or otherwise wrong output [14]. This opens up other hazards, more so when users get overconfident about the fluency or confidence of the system. Meanwhile, such tools are relatively easy to use and designed in an intuitive manner, thus available to a population across the globe, frequently breaking technical boundaries and allowing access to AI that non-experts can use and participate in.

The aim of this review is to analyze the technological basis, developmental approaches and the growing functions of the contemporary AI chatbots. Instead of comparisons made right away between these systems, we will first examine the shared backbone of these models, their architecture, training methods, and alignment strategies. There, we trace how the systems have been implemented in professional and practical contexts and focus on applications that are not restricted to novelty and entertainment.The concluding section will draw direct comparisons between selected chatbots, offering critical insights into their strengths, weaknesses, and the emerging directions for future innovation.

By combining technical insight with practical analysis, this paper aims to serve as both a primer and a forward-looking roadmap. It is written with the understanding that AI chatbots are not just tools, they are becoming digital collaborators. Their ongoing development and integration raise important considerations for design ethics, user trust, policy, and long-term societal impact. As such, the perspectives in this review are informed by both academic research and professional expertise.

1.1. Statement of the Problem

Despite the widespread use and growing capabilities of AI chatbots, there remains a lack of structured, comparative understanding of how different systems perform across technical, functional, and professional dimensions. Most existing discussions focus either on general capabilities or are limited to promotional overviews from developers themselves. This establishes a gap among the users, researchers and professionals as they would need an objective, comprehensive evaluation of chatbot tools in the real-world. In addition, newer systems are being institutionalized in education, healthcare, legal services, and software developing; now, it is more pressing to recognize the limits, trade-offs in safety as well as the purpose behind the design of these systems. The lack of clear framework to compare two different chatbots can leave users unable to select the appropriate one to their needs; instead of this, they can overrate the potential of some solutions, which can be misused or inaccurately trusted.

1.2. Objectives of the Study

This review aims to:

1. To examine the core technologies and design principles behind modern AI chatbots.
2. To identify major application domains across professional and consumer contexts.
3. To analyze the functional distinctions among selected AI chatbots.
4. To highlight key challenges related to safety, ethics, and reliability.
5. To provide a structured comparative insight to guide informed use and development.

1.3. Overview of Chatbot Technologies

1.3.1. Core Architectures

The latest and most powerful AI chatbots rely on large language models (LLMs) deployed using transformer models. Transformer model essentially shifted paradigm in the field of natural language processing by the ability to parallelize attention along the token sequence instead of operating it sequentially as it had been done with previous models such as RNNs or LSTMs [15]. Such an architectural advance especially enhanced the capability of models to learn long-range dependencies and contextual relationships providing the foundation of the current systems of OpenAI ChatGPT, Anthropic Claude, Google Gemini, and Meta LLaMA [8].

Although the transformer is a common ground, various chatbot providers have come up with variations that suit their goals. GPT models used by OpenAI such as GPT-3.5 and GPT-4 [16] are constructed as dense stacks of transformer-based models, where the probabilities of generating successive tokens are computed in an autoregressive fashion, given a context provided by the previous tokens. The models are additionally refined by means of hybrid supervised fine-tuning and reinforcement learning based on human feedback (RLHF) to more closely match their output with human values, and instruction and task requirements [16].

Google's Gemini, particularly in its 1.5 release, presents a Mixture-of-Experts (MoE) Mixture-of-Experts (MoE) architecture with its own selective activation of a small subset of neural pathways [17]. This selective routing is effective in its efficiency and scalability as Gemini can process an input context of very high length of the order of 1 million tokens with high responsiveness and coherence [18]. This allows it to be quite good at long-document comprehension, analysing code and retrieval-intensive operations.

Anthropic's Claude retains the transformer core and adds a special alignment strategy called Constitutional AI [19]. Claude does not completely rely on human raters in a training process but operates within a set of rules that disclose a set of moral principles embedded in its training loop [20]. The principles are employed in their criticism, revision of the outputs promoting the safety, neutrality, non-manipulative conduct in responses [20]. This architectural philosophy emphasizes controllability and value alignment without excessive human intervention.

More open and community-based LLaMA models were built by Meta in their third version. LLaMA is modeled as a set of transformer-based foundation models and provides access to open-weighting to academic and commercial experimentation. LLaMA 3 has been optimized for scalability and multilingual performance, and serves as the core engine behind Meta's in-house AI assistant deployed across platforms like Instagram, WhatsApp, and Messenger [21].

These architectural innovations go beyond not only technical diversities but philosophical differences in the understanding, construction, and application of chatbot intelligence. Others have zoomed towards safety and ethical monitoring, and others have oriented toward openness, scalability, and circumstances over extended sequences. Figure 1 [2] shows that the chatbot systems are often developed on the modules framework based on natural language processing engines, validation layers, message routing logic, and API based integrations with external services. This multi-level technology allows its deployment to a variety of different platforms such as messaging apps and social media, as well as IoT systems and enterprise APIs with the support of dynamic language understanding and response generation. Conclusively, every architectural decision has measurable consequences for a chatbot's ability to interpret instructions, sustain coherent interactions, and operate reliably in real-world, multi-channel environments.

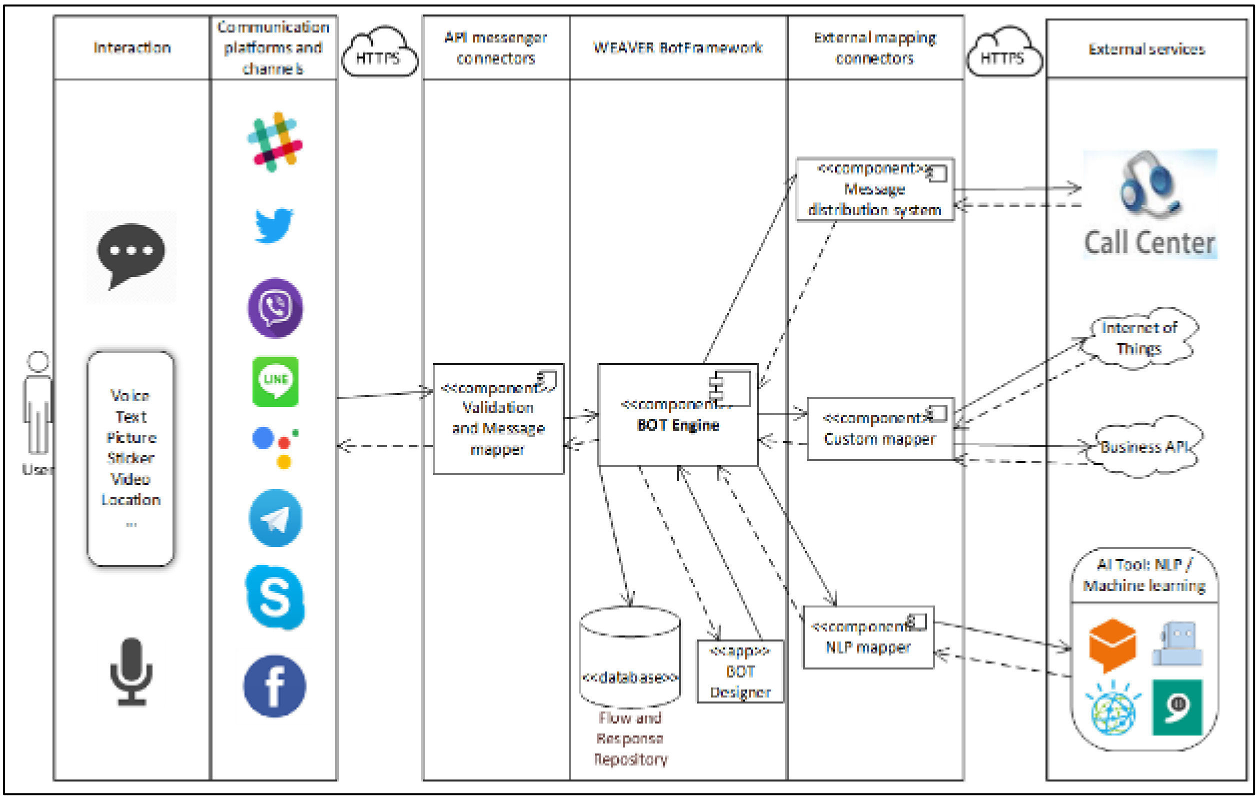

Figure 1 Layered Architecture of a Chatbot System with Natural Language Processing and API Integration [1]

This figure illustrates a modular chatbot architecture involving multiple communication platforms, a centralized bot engine, validation layers, and natural language processing (NLP) components. User interactions, ranging from voice and text to multimedia, are routed through messaging APIs and processed via the WEAVER Bot Framework. The bot engine communicates with external services such as call centers, IoT platforms, and business APIs, enabling both rule-based and AI-enhanced conversational capabilities. The architecture supports NLP mapping, flow design, and message distribution through extensible and decoupled components.

1.3.2. Training Data and Fine-Tuning Methods

The training data used to build large language models (LLMs) plays a critical role in shaping their generalization capabilities, domain knowledge, and the biases they may carry [22]. Most state-of-the-art chatbots are initially pre-trained on massive, diverse corpora that include web text, encyclopedic content, books, open-source code, forums, and multilingual datasets [23]. These data sources are often filtered to reduce noise, remove explicit content, and improve linguistic diversity, although the exact datasets used are typically not publicly disclosed due to proprietary concerns.

Pretraining alone, however, is insufficient to make these models useful, safe, or aligned with human intentions. Post-pretraining, most developers apply multiple rounds of fine-tuning to guide the model's behavior more deliberately [8,16]. One widely adopted approach is supervised fine-tuning (SFT), in which human-annotated examples are used to explicitly teach the model how to follow instructions, engage in multi-turn dialogue, or perform structured tasks [16]. This method has been integral to models like ChatGPT, which learned to mimic helpful, concise, and context-aware conversational styles [8].

Beyond SFT, many developers incorporate reinforcement learning from human feedback (RLHF) to further refine outputs [16]. In this setup, human evaluators rank the model's responses to prompts, and these preferences are then used to adjust the model's behavior through reinforcement learning algorithms. This approach has been credited with improving response quality, discouraging toxic or incoherent content, and aligning the model more closely with user expectations. However, it also introduces subjectivity and potential bias, depending on who provides the feedback and what guidelines are used [16].

Some models, such as Claude and Gemini, also undergo instruction tuning, where they are exposed to a wide variety of prompt formats, tasks, and domains to improve generalization across unseen queries [18,19]. Instruction tuning helps reduce rigidity and improves the model's ability to follow zero-shot or few-shot instructions with minimal prompting.

A particularly novel strategy is employed by Anthropic's Claude through Constitutional AI [19, 20]. Instead of relying solely on human raters, this method embeds a set of ethical principles, such as avoiding harm, respecting privacy, or remaining neutral, directly into the training loop [19, 20]. The model is then taught to critique and revise its own outputs based on these principles, creating a self-supervised feedback mechanism that aims to reduce harmful responses without extensive human intervention.

Ultimately, the choice and balance of these fine-tuning methods significantly influence a chatbot's tone, reliability, ethical stance, and trustworthiness. As AI systems are increasingly deployed in sensitive or high-stakes environments, the alignment strategies behind them become just as important as their raw capabilities [16].

1.3.3. Retrieval-Augmented Generation (RAG), Agent Models, and Integration

Although large language models (LLMs) demonstrate remarkable capabilities, they are inherently limited by the static nature of their training data. Once trained, these models are not connected to the internet or real-time data sources, which means they cannot access new or evolving information. This often leads to hallucinations, confident but factually incorrect outputs, and outdated responses, particularly in knowledge-intensive or time-sensitive domains [14]. To respond to this, developers have moved to adopt a method known as Retrieval-Augmented Generation (RAG) where the language model is complemented with the so-called document retriever that can fetch relevant, current, or topic-focused material in outside sources. The information obtained is then applied to ground the answers provided by the model, and it is factual and in the context. As shown in Figure 2, The retrieval-augmented generation technique enables LLMs to have the capability of omitting or timely overusing static knowledge, by uploading external document call-ups at inference-time[1].

ChatGPT, among others, has features of browsing and plugins allowing the tool to interface with APIs, calculators, and external programs. This retrieval level allows the model to provide responses within and using live web-based information or data structure which expands the range of the models applicability in time-sensitive applications such as trip planning, live stocks information, or academic search [8]. Likewise, Google Gemini is linked to Google Search and Google Workspace Applications, which gives it the flexibility to access documents, spreadsheets and web results in a seamless manner [18]. This integration enhances its ability to summarize, extract, and reason across multiple data formats within the Google ecosystem. Models like Joules and Meta AI also rely on retrieval-based augmentation, pulling contextually relevant information from search results or internal knowledge graphs to improve accuracy during complex queries [21].

Indeed, scholars argue that generative AI should be viewed less as truth-generating systems and more as 'style engines' tools that simulate human-like fluency and creativity rather than traditional accuracy, thus reframing expectations around output veracity and interpretability [24].

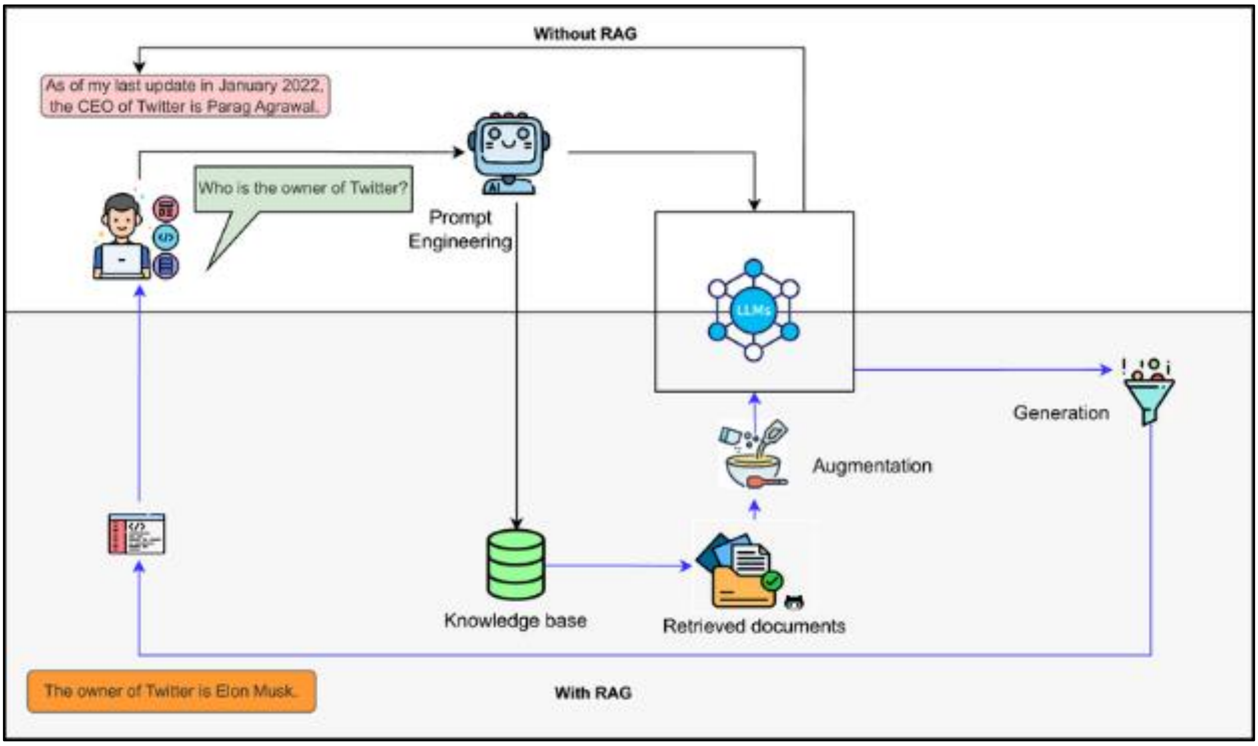

Figure 2 Illustration of Retrieval-Augmented Generation (RAG) in Large Language Models [2]

The diagram compares traditional language models without RAG, which rely solely on static pretraining, with RAG-based systems that dynamically fetch relevant, up-to-date information from external knowledge sources. The retrieved documents are used to augment the model's context, improving factual accuracy and temporal relevance during generation.

Beyond retrieval, there is growing interest in agent-based models, which move beyond simple prompt-response interactions and toward autonomous, goal-driven behavior. These models can break down user instructions into subtasks, plan multistep actions, and execute external commands through tools or APIs. OpenAI's experimental AutoGPT and Google's Gemini Advanced are early examples of this approach, where the chatbot behaves more like an intelligent agent capable of making decisions, iterating on strategies, or switching tools mid-task based on user intent [25]. While still in early stages, these models represent a major step toward interactive, task-oriented AI that can collaborate with users in more proactive and dynamic ways.

Integration also plays a pivotal role in shaping how chatbots are experienced and adopted. Today's AI assistants are no longer confined to isolated applications; they are embedded across the digital ecosystem. Browser-based systems like Gemini in Chrome allow users to query directly from search bars, while writing platforms such as GrammarlyGO offer in-line assistance for editing, summarizing, or generating content [17,26]. In developer environments, GitHub Copilot and ChatGPT's code interpreter assist with real-time programming, debugging, and logic analysis [27]. Messaging platforms, including WhatsApp, Instagram, and Facebook Messenger, now include Meta's AI assistant to provide instant responses, search suggestions, or content recommendations [21].

These integrations signal a broader shift in how AI is deployed, not as a separate tool, but as a background collaborator woven into daily workflows. By embedding intelligence into familiar interfaces, chatbot systems gain both immediacy and invisibility, increasing user reliance and expanding the range of tasks they can support. This fusion of retrieval, autonomy, and integration marks a turning point in chatbot evolution, where the distinction between AI tool and digital co-worker begins to blur.

Table 1 Summary of Chatbot Architectures and Design Characteristics

Model	Developer	Architecture	Context Length	Alignment Method	Notable Feature
ChatGPT (GPT-4o)	OpenAI	Dense Transformer (autoregressive)	Up to 128K tokens	RLHF + SFT	Multimodal input (text, image, audio), Code Interpreter
Gemini 1.5	Google DeepMind	Mixture-of-Experts (MoE Transformer)	Up to 1M tokens	RLHF + Instruction Tuning	Native integration with Google Search, Workspace
Claude 3 Opus	Anthropic	Transformer with Constitutional AI	~200K tokens	Constitutional AI	Safety-aligned self-critique mechanism
LLaMA 3	Meta AI	Open-Weight Transformer (dense)	~65K–128K tokens	Community fine-tuning	Open-source availability, high multilingual capability
Joules AI	You.com	Transformer-based (retrieval-augmented)	~32K tokens (est.)	SFT + RAG	Open web integration, search-driven answers
GrammarlyGO	Grammarly	Proprietary fine-tuned transformer	~4K–8K tokens (est.)	SFT	Writing-focused, context-aware tone adaptation

1.4. Chatbot Profiles and Comparative Matrix

As AI chatbots become increasingly embedded into consumer tools and professional workflows, understanding their practical characteristics becomes essential. While all modern systems leverage transformer-based large language models, they differ significantly in terms of input modalities, integration design, target user base, and underlying optimization strategies. This section offers a comparative overview of six widely used chatbot systems, summarizing their origins, capabilities, and trade-offs based on publicly available documentation, technical benchmarks, and observed user experience.

OpenAI's **ChatGPT** (GPT-4o) stands out for its broad multimodal support, including text, images, and audio, and its seamless plugin ecosystem, enabling access to code execution, file uploads, and web browsing [28]. It is known for strong reasoning ability and instructional fluency but remains costly for high-volume use and is prone to hallucination, especially when generating long-form content or citations.

Gemini 1.5, developed by Google DeepMind, features a Mixture-of-Experts transformer architecture that supports text, images, and code. It integrates tightly with Google Search and Workspace, offering real-time document access and retrieval-augmented answers [18]. Despite its rich features, availability remains limited in certain regions and enterprise contexts.

Claude 3, by Anthropic, is designed around a safety-first principle using Constitutional AI. It supports long-context reasoning with a strong emphasis on neutrality and non-manipulative responses [29], although users have reported slower interface speeds and restricted availability compared to its competitors.

Meta's **LLaMA 3**, integrated into **Meta AI** across platforms like WhatsApp and Instagram, is optimized for lightweight inference and multilingual support [21]. It accepts text and images but has limited capabilities in code or abstract reasoning, reflecting Meta's social-first design philosophy.

GrammarlyGO, a writing assistant rather than a general-purpose chatbot, focuses on grammar correction, tone adjustment, and editorial rephrasing [26]. It handles only short-form text and lacks broader reasoning capabilities, making it ideal for narrow, writing-focused domains.

Joules, developed by You.com, takes a retrieval-augmented approach that allows fast response generation, particularly for coding and development tasks [30]. It blends open-source models with in-browser tools but suffers from occasional UI inconsistency and lacks the polish of more mature platforms.

These variations reflect distinct design intentions, from generalist intelligence and code generation to safety, writing precision, and speed. Table 2 summarizes these chatbot profiles for quick reference.

Table 2 Chatbot Profiles and Comparative Matrix

Feature	ChatGPT	Gemini	Claude	Meta AI	GrammarlyGO	Joules
Developer	OpenAI	Google DeepMind	Anthropic	Meta	Grammarly	You.com
Model	GPT-4o	Gemini 1.5	Claude 3	LLaMA 3	Proprietary	Mix of open-source
Input Modes	Text, Image, Audio	Text, Image, Code	Text	Text, Image	Text only	Text, Code
Strength	Reasoning, plugins	Web-integrated, multimodal	Safe, constitutional AI	Messenger-first, social	Grammar/style aid	Fast, coding
Limitation	Cost, hallucinations	Limited in some regions	Slower UI, availability	Limited reasoning	Narrow domain	UI inconsistency

1.5. Evaluation Criteria

As AI chatbots become more widely integrated into both professional and consumer applications, the need for robust and multidimensional evaluation has become increasingly important. Traditional benchmarks focused on language modeling performance alone are no longer sufficient to assess real-world utility. Instead, evaluation must incorporate a blend of technical, functional, and ethical dimensions to reflect the complex roles these systems now occupy. This section presents six core criteria used to evaluate contemporary chatbot systems.

Accuracy and reliability remain foundational for any AI assistant. While most large language models demonstrate fluency, their ability to generate factually correct and verifiable responses continues to vary. Chatbots relying solely on pre-trained data are particularly vulnerable to hallucinations, confidently stating incorrect information [31]. Reliability also includes consistency across prompts, deterministic behavior when required, and tolerance to ambiguous input.

Multimodal capacity is an increasingly critical differentiator among advanced systems. The ability to process and respond to not just text, but also images, audio, and code inputs, allows for broader functionality and cross-domain interactions. Systems like ChatGPT and Gemini exemplify this direction, enabling use cases in design review, visual reasoning, and audio transcription, whereas text-only systems may be limited in their range of interaction [32].

Context retention refers to a chatbot's ability to remember and utilize prior information across multiple turns or within long prompts. Models with extended context windows (e.g., Gemini 1.5 and ChatGPT) are better suited for complex tasks like contract review, research synthesis, and long-form writing [33]. Effective context handling reduces the need for repeated input and enables more human-like conversation flow, particularly in professional domains.

Professional applications challenge a model to be precise and consistent in carrying out tasks that are deemed within the domain. This incorporates software development, scholarly writing, legal, and clinical information retrieval, spheres that require meticulous formatting, structuring of outputs and content with respect to the expert-level expectations [34].

Another significant requirement is user customization especially in cases where the models are applied in enterprise or content development context. More precise outputs can be made possible by the possibility to adjust tone, style, verbosity or even persona. Other platforms provide user-specified instructions or memory capabilities, or can be used with fine tuning or plugin extensions which add to the behavior of chatbots [35].

Lastly, safety and ethics are vital parts of any cognizant assessment. This includes the capacity of a chatbot not to produce dangerous or discriminatory information, defend the privacy of users, and be transparent in ambiguous situations or when it comes to sensitive data. Some systems such as Claude have safety mechanisms built in during training, whereas some others have moderation layers or users flagging systems to deal with problematic output [36].

These criteria form a comprehensive foundation for analyzing not just what chatbots can do, but how well they meet the expectations of accuracy, accessibility, adaptability, and responsible deployment in modern digital environments.

1.6. Real-World Use Cases

Beyond their technical capabilities, the true measure of AI chatbots lies in how effectively they support real-world tasks. Each system is designed with specific user goals and interaction contexts in mind, leading to divergent strengths across domains such as education, development, writing, compliance, and communication. This section examines how major chatbots are currently deployed in professional and consumer environments, focusing on the alignment between model design and task execution. As shown in Figure 3 [3], the adoption and academic investigation of AI chatbots remain geographically uneven, with most empirical studies emerging from Asia and very few from African or South American contexts. This disparity highlights a need for more globally inclusive evaluations of chatbot use, especially in underrepresented educational systems.

The following analysis is based on documented platform features, public user interfaces, and the author's hands-on usage of the respective systems.

OpenAI's ChatGPT, particularly in its GPT-4o release, is among the most versatile and widely adopted chatbots, supporting a broad spectrum of applications. In software development, it offers real-time assistance with code generation, debugging, and documentation. The integration of tools such as the code interpreter, Python environment, and file analysis plugins extends its utility into areas like data science, mathematics, and file parsing. In customer service, ChatGPT is increasingly being used for handling support tickets, live chat automation, and knowledge base generation, often enhanced by plugin access to external APIs or enterprise systems.

Gemini, a creation of Google DeepMind is effective in academic-based and knowledge-intensive settings. Its seamless interconnection with Google Search and Workspace allows live augmentation of web content, text summarization, and live lookups of any data, which is why it is especially useful to students, researchers, and analysts. Gemini can also be used in the creation of educational content, fact-checking jobs, particularly those in which a longer-context comprehension and faithful quotation are needed.

Claude, a product of Anthropic, is specially placed to be used in policy, legal, and compliance-related purposes. Being trained according to the principles of Constitutional AI, it is unable to be easily manipulated, biased, or deal with sensitive data carelessly. The above characteristics render Claude worthwhile in regulatory writing, content moderation, risk assessment, and other fields that require an ethics-first approach to response tone and alignment.

Based on the LLaMA architecture, Meta AI was implemented to run on its own platforms, including WhatsApp, Instagram, and Facebook Messenger which are owned by Meta. It is adjusted towards conversational perceptiveness in everyday, societal or amusement-oriented communications. The common applications are responding to trivia, content recommendations, social post support, and simple, low-stakes customer engagement with messaging bots, thus making it best suited in high-volume low-risk interactions of social ecosystems.

GrammarlyGO focuses on improving the quality and tone of written communication. It is embedded into word processors, email clients, and web-based editors, offering real-time suggestions for grammar, clarity, tone, and sentence structure. While narrow in scope, it excels at editorial refinement, especially for professionals seeking to improve communication effectiveness without relying on full generative AI solutions.

Lastly, Joules, a chatbot developed by You.com, targets developers, researchers, and power users who require fast, retrieval-augmented assistance. By drawing from the open web and code repositories, Joules provides real-time answers for coding queries, documentation lookup, and snippet generation. Its integration with browser tools and preference for transparency over closed-source fluency make it a lightweight but practical assistant for technical users.

These varied use cases demonstrate that while all modern chatbots are built on language models, their real-world value is tightly linked to how well they are optimized and integrated for domain-specific demands.

Table 3 Real-World Use Cases of Leading AI Chatbots

Chatbot	Primary Domains of Application	Key Strengths in Practice
ChatGPT	Software development, customer support, file analysis, and education	Code writing, debugging, math tools, and plugins for external functions
Gemini	Academic research, web search, and document summarization	Live Google Search integration, Workspace access, long-context tasks
Claude	Legal writing, compliance, ethics-focused communication	Safe alignment, neutrality, policy drafting, and handling sensitive input
Meta AI	Social media messaging, content replies, and informal engagement	Fast responses in WhatsApp/Instagram, lightweight assistant design
GrammarlyGO	Business communication, editorial workflow, emails	Tone rephrasing, grammar correction, and clarity enhancement
Joules	Coding queries, developer tools, web-augmented answers	Fast retrieval, code-focused outputs, browser-based productivity

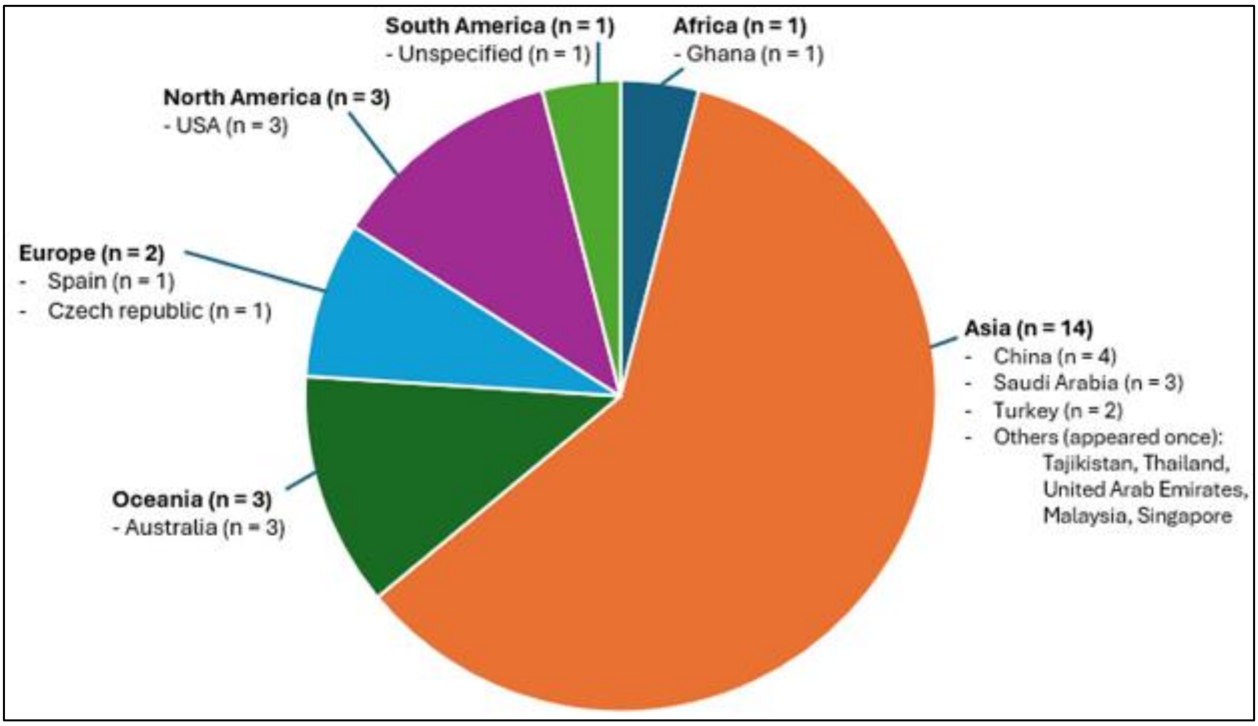

Figure 3 Geographical distribution of studies on AI chatbot usage in higher education (2024) [3]

The figure presents the global spread of empirical research on chatbot use among students, revealing a concentration of studies in Asia (n = 14), followed by North America and Oceania (n = 3 each). Limited representation is seen from Africa, South America, and parts of Europe. This suggests a regional imbalance in research output and may reflect differences in infrastructure, policy, or institutional adoption of AI tools.

1.6.1. Strengths and Gaps Across Systems

The following comparative assessment is based on publicly documented features, platform demonstrations, and the author's direct usage and testing of each system.

In real-world use, the strengths of today's leading chatbot systems become immediately obvious the moment you move beyond simple prompting. Chat Generative Pre-trained Transformer version 4 omni (ChatGPT-4o), developed by OpenAI, remains one of the most powerful for general-purpose reasoning, code generation, mathematical problemsolving, and file analysis. Its support for image and audio inputs in the same chat session makes it the most capable in terms of multimodal interaction. The ability to switch between reasoning, document parsing, creative writing, and code execution in a single session is unmatched. The major limitation, however, is that high-quality access (that is, Chat Generative Pre-trained Transformer version 4 omni-level intelligence) is gated behind a paid plan, and even then, hallucinations can still occur with citations or niche factual queries.

Gemini version 1.5, by Google, is similarly strong, especially in long-context comprehension. It handles large Portable Document Format (PDF) files, spreadsheets, and research papers without breaking context, something Chat Generative Pre-trained Transformer still struggles with at times. Gemini also shines in tasks that require up-to-date information or direct interaction with Google Docs and Gmail. However, its interface is less stable under load, and its output is sometimes overly verbose or vague in technical tasks like coding or data structure debugging. Availability also remains limited in some countries, and application programming interface (API)-level access is still restricted compared to OpenAI.

Claude version 3 is the most "human-sounding" chatbot in many cases, and it excels in tasks that require careful tone, neutrality, or ethical balance. It is a go-to for policy writing, rewriting sensitive text, or explaining complex issues without bias. Its context window (up to two hundred thousand tokens) is a huge advantage for reviewing legal documents or structured content. The downside? Claude is slower, lacks true plugin or tool integration, and does not offer audio or image input, making it limited for broader applications.

Meta Artificial Intelligence (Meta AI), built into platforms like WhatsApp and Instagram, is more of a lightweight, always-on assistant than a deep reasoning model. It handles casual queries, simple summaries, and conversational replies well. But it does not offer code-level intelligence or the ability to deeply reason through multi-step logic. You cannot upload files or parse documents. It is fast and good for social media interaction, but not suitable for research, coding, or professional writing.

Grammarly Generative Output (GrammarlyGO) is hyper-focused: it improves email tone, rewrites text with better clarity, and fits seamlessly into everyday writing platforms. It does not hallucinate because it does not invent content, it just adjusts yours. But you cannot ask it to explain a topic, compare ideas, or generate structured documents. It is helpful for communication polish, but that is where its usefulness ends.

Joules, on You.com, feels like a coder's search companion. It gives quick answers, grabs snippets from websites, and can combine multiple sources into one response. It is fast and clean, but prone to factual inconsistencies if the source links are not vetted. It does not reason well with complex instructions and lacks memory or deeper contextual understanding. Also, its web-first interface lacks the maturity of Chat Generative Pre-trained Transformer or Claude in structured, multi-turn tasks.

In summary, Chat Generative Pre-trained Transformer is best for power users, Gemini for document-heavy or Google-integrated workflows, Claude for ethics and long text review, Meta Artificial Intelligence for casual use, Grammarly Generative Output for editing, and Joules for quick code and search tasks. The gap lies in how few chatbots currently balance depth, accuracy, and multi-functionality. None are perfect, but the strengths are clear when matched to the right task.

1.6.2. Deployment Environments: Browser, Mobile, Enterprise

The way artificial intelligence chatbots are deployed heavily influences both how people use them and what they are realistically useful for. Chat Generative Pre-trained Transformer, Gemini, and Joules are primarily browser-based, making them easily accessible without any installation or setup. Chat Generative Pre-trained Transformer's web application is clean and responsive. With the professional subscription plan, users gain access to Chat Generative Pre-trained Transformer version 4 omni, along with advanced tools such as the code interpreter, file uploads, and long-term memory support. Gemini's browser interface is smooth and particularly effective for document-heavy workflows, especially when integrated with Google Drive or Gmail. However, it may experience lag during extended sessions or when handling large contextual prompts. Joules operates entirely within the browser and functions as a fast, retrieval-augmented assistant for coding and web search. Its lightweight design offers speed but lacks persistent session memory and deeper analytical capacity.

Meta Artificial Intelligence is integrated directly into mobile messaging platforms such as WhatsApp, Facebook Messenger, and Instagram. This makes it extremely convenient for casual users, it is always "on" and feels like a native part of the chat experience. You can ask it to summarize a message, answer a question mid-conversation, or suggest a caption. However, its limitations become obvious quickly: it cannot handle structured documents, it has no programming support tools, and it does not accept multimodal input such as Portable Document Format files or other uploaded materials. It is built for lightweight interactions, not deep or technical workflows.

Grammarly Generative Output is not a standalone chatbot but rather a feature embedded within the Grammarly desktop application and web browser extension. It appears contextually when writing emails, LinkedIn posts, or blog articles, and offers suggestions for tone shifts, sentence rephrasing, and grammar corrections in real time. You never have an ongoing dialogue with it, the way you do with Chat Generative Pre-trained Transformer, it is invisible unless you need to rewrite or polish something. This embedded design makes it powerful for improving writing quality, but completely unsuitable for tasks beyond communication editing.

In enterprise environments, deployment becomes even more critical. Claude, developed by Anthropic, is becoming a preferred choice in compliance-heavy sectors due to its safety-first design and alignment with ethical language modeling principles. Its limited application programming interface access and transparent data handling policies make it easier to integrate into sensitive institutional workflows without raising privacy concerns. Gemini, by contrast, is natively integrated into Google Workspace, enabling enterprise users to summarize content in Google Docs, generate replies in Gmail, or analyze spreadsheets in Google Sheets. This gives it a major advantage in collaborative work environments where seamless integration with existing productivity tools is essential.

Chat Generative Pre-trained Transformer is also evolving into a highly flexible enterprise solution. With support for custom instructions, modular plugin integrations, specialized artificial intelligence agents known as Generative Pre-trained Transformers, and file upload capabilities, it offers deep compatibility with automation platforms like Zapier, custom application programming interfaces, and protected organizational datasets. The business-tier plans from Open Artificial Intelligence also prioritize privacy, providing enterprise contracts that guarantee no customer data is used to train future language models.

In practice, deployment is not only about where the chatbot is accessed, it is about how well it fits into existing daily workflows, whether those involve communication, messaging, software development, or formal enterprise documentation. The most effective chatbot is not always the most powerful; it is often the one that appears precisely where the user needs support.

1.6.3. Human–Artificial Intelligence Collaboration in Practice

Modern artificial intelligence chatbots are not replacing users; they are becoming collaborators. In practice, the most effective systems are those that support shared cognitive work, where the human sets the goals, applies judgment, and provides oversight, while the artificial intelligence handles language generation, data processing, or structural planning. This model, human-in-the-loop by design, is what makes these tools usable in real-world workflows.

Chat Generative Pre-trained Transformer, especially in its version 4 omni release, is a prime example of this collaborative logic. It is most effective when used iteratively, writing a chunk of computer code, pausing to get feedback, refining it, and prompting again. The same applies to data analysis using the embedded code interpreter or structuring long documents with user-defined formatting. It rarely gets everything right the first time, but when guided, it delivers high-quality, co-authored results.

Claude, developed by Anthropic, takes a different approach. Its training via Constitutional Artificial Intelligence allows for outputs that reflect neutrality, restraint, and ethical tone even without being explicitly told what to avoid. In professional environments such as human resources policy drafting or ethical reviews, this allows users to work with the artificial intelligence as a filter, not just a generator. Users do not need to manually prompt it for fairness or balance; the model is pre-aligned toward those goals, making it easier to trust in regulated settings.

Grammarly Generative Output collaborates in an even more embedded way. It does not generate content from scratch but offers tone adaptation, sentence restructuring, and clarity improvements in real time, often without interrupting the user's workflow. Instead of replacing writing, it quietly co-authors it, letting the human lead and refining only when invited. This kind of passive collaboration is subtle but effective, especially in professional communication.

The most successful human and artificial intelligence collaboration happens when the model respects boundaries, knowing when to follow and when to wait for guidance. When users retain control over direction, tone, and judgment, artificial intelligence outputs become trusted inputs, not unquestioned answers. These systems are not decision-makers; they are accelerators of human decisions.

Going forward, collaboration will deepen through features such as persistent memory, where chatbots remember user preferences and style across sessions; goal tracking, where they assist over time rather than just per prompt; and tool-based autonomy, where the artificial intelligence can act on behalf of users with explicit permissions. But none of these advancements work without the user at the center. The future of artificial intelligence chatbots is not fully autonomous, it is deeply collaborative, with the human setting the agenda and the machine executing, suggesting, and refining in real time.

1.6.4. Evolving Expectations of Chatbot Intelligence

As artificial intelligence chatbots have grown more capable, user expectations have shifted dramatically. Early chatbot systems were seen as helpful if they could respond fluently to simple queries or complete a sentence correctly. Today, the expectations are far more demanding. Users now look for chatbots that can sustain long conversations, follow multi-step instructions, interpret tone, offer emotionally appropriate responses, and even reason through technical or regulatory prompts. Whether asked to draft a legal disclaimer, solve a recursive programming problem, or compose a sensitive condolence message, these systems are increasingly expected to act with insight and discretion, not just linguistic fluency.

This rising bar has driven developers to focus not only on computational performance but also on trust, transparency, and adaptability. Modern systems like Chat Generative Pre-trained Transformer have introduced multimodal capabilities that allow for interpreting images, audio, and text together. Claude's safety-focused alignment framework guides its responses using clearly defined ethical principles. Gemini's ability to process large Portable Document Format files, summarize, and cross-reference web results within a single interaction reflects a growing emphasis on intelligent reasoning. These evolutions are not accidental, they are responses to a user base that now sees chatbots not merely as tools but as collaborative assistants.

At the same time, the expectation of transparency has become central. Users increasingly ask: Where is the chatbot getting its information? How current is the response? What data was the system trained on? When responses contain factual errors or vague disclaimers, user trust is quickly compromised. This has placed pressure on developers to create systems that are not only effective but also interpretable, auditable, and clear about their data sources and limitations.

Ethical alignment is another dimension of this shift. Users want to know that the chatbot will not reinforce harmful stereotypes, expose private data, or deliver manipulative or inappropriate content. In academic, legal, or professional writing contexts, there is a growing expectation that artificial intelligence systems will handle sensitive content with neutrality, fairness, and awareness of social responsibility. This is why platforms like Claude are gaining traction in settings where emotional tone and ethical restraint are essential, while enterprise deployments of systems like Chat Generative Pre-trained Transformer have introduced features to control plugin access, memory retention, and system instructions to ensure alignment with organizational values.

What emerges is a clear trend: people are no longer just evaluating whether a chatbot works, they are evaluating whether it understands, adapts, and can be trusted. As artificial intelligence continues to evolve, future systems must be designed not simply for scale or speed, but for ethical reliability, transparency, and long-term human collaboration. The measure of intelligence is no longer just what the chatbot knows, but how responsibly and contextually it behaves.

1.7. Future Directions

The next generation of artificial intelligence chatbots will likely be defined by five emerging trajectories. First is the rise of unified multimodal agents. As users increasingly expect systems to process images, audio, and text in a single interaction, chatbots are being re-engineered to act across modalities seamlessly. This shift expands their relevance in fields like education, design, and diagnostics, where inputs are not purely textual.

Second is the growing challenge of balancing autonomy with controllability. As chatbots are given more responsibility, whether planning tasks, interacting with external tools, or calling functions, they must remain auditable, interruptible, and explicitly permissioned. Systems that act without transparency or override user intent will struggle to earn trust, especially in regulated environments.

Third, advances in edge computing are enabling hybrid artificial intelligence models that combine on-device reasoning with cloud-based retrieval. These systems reduce latency, improve privacy, and can operate even in bandwidth-constrained environments. On-device artificial intelligence also opens the door for more secure applications in healthcare, finance, and legal contexts where sensitive data must remain local.

Fourth, the role of professionals in artificial intelligence design and oversight will become more central. Educators, legal practitioners, engineers, and health professionals must be involved not only in dataset curation but also in defining system behavior, validating outputs, and guiding ethical frameworks. Artificial intelligence models trained without input from domain experts risk reinforcing bias, producing inappropriate content, or failing to meet legal or institutional standards.

Fifth, to guide future adoption and comparative assessment, a final matrix of the leading chatbot systems has been presented. This matrix evaluates six major platforms across five key dimensions: reasoning depth, modality support, enterprise readiness, ethical alignment, and accessibility. It highlights the functional strengths and limitations of each system and reinforces the importance of aligning model selection with task-specific needs.

Artificial intelligence chatbots are moving from experimental tools to collaborative digital agents. Their continued success will depend on the degree to which they support human goals, operate transparently, adapt to context, and behave in a way that reflects both technical precision and ethical reliability. As these systems evolve, their value will no longer be measured by how much they can say, but by how well they understand and serve the people who use them.

2. Conclusion

The evolution of artificial intelligence chatbots has demonstrated significant advancement in reasoning ability, natural language fluency, multimodal processing, and domain-specific applications. From general-purpose systems like ChatGPT and Gemini to more specialized tools such as GrammarlyGO and Joules, these models reflect a spectrum of design philosophies, training approaches, and integration strategies. ChatGPT stands out for its broad reasoning capacity and plugin ecosystem; Gemini excels in long-context document handling and search-linked intelligence; Claude focuses on safety, neutrality, and extended input comprehension; Meta AI serves lightweight messaging tasks with speed and accessibility; GrammarlyGO is embedded for real-time writing enhancement; and Joules performs well for coding tasks and retrieval-based queries.

Despite these strengths, no single system meets all user needs. Some models are gated by subscription costs or regional restrictions, while others lack the depth or flexibility for cross-domain applications. Tools like Claude prioritize safety and ethical alignment but fall behind in multimodal capability. Meta AI is accessible and responsive, but not suitable for complex professional use. These trade-offs underscore the importance of choosing chatbots based on context, task complexity, and the required level of autonomy or user control. As chatbots increasingly become part of academic, business, and personal workflows, their effectiveness will hinge on more than just performance, it will depend on ethical deployment, reliability, and how well they align with user expectations.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Matic, R., Kabiljo, M., Zivkovic, M., & Cabarkapa, M. (2021). Extensible Chatbot Architecture Using Metamodels of Natural Language Understanding. *Electronics*, *10*(18), 2300. https://doi.org/10.3390/electronics10182300

[2] Rocco, J. D., Ruscio, D. D., Di Sipio, C., Nguyen, P. T., & Rubei, R. (2025). On the use of large language models in model-driven engineering. *Software & Systems Modeling*, *24*, 923–948. https://doi.org/10.1007/s10270-025-01263-8

[3] Schei, O. M., Møgelvang, A., & Ludvigsen, K. (2024). Perceptions and Use of AI Chatbots among Students in Higher Education: A Scoping Review of Empirical Studies. Education Sciences, 14(8), 922. https://doi.org/10.3390/educsci14080922

[4] Caldarini, G., Jaf, S., & McGarry, K. (2022). A literature survey of recent advances in chatbots. Information, 13(1), 41. mdpi. https://doi.org/10.3390/info13010041

[5] Han, Q., & Zhao, C. (2025). Unleashing the potential of chatbots in mental health: Bibliometric analysis. Frontiers in Psychiatry, 16. https://doi.org/10.3389/fpsyt.2025.1494355

[6] Aldoseri, A., Khalifa, K. N. A., & Hamouda, A. M. (2024). AI-Powered innovation in digital transformation: Key pillars and industry impact. Sustainability, 16(5), 1790. mdpi. https://doi.org/10.3390/su16051790

[7] Moroto, Y., Maeda, K., Togo, R., Ogawa, T., & Haseyama, M. (2024). Multimodal transformer model using time-series data to classify winter road surface conditions. Sensors, 24(11), 3440. https://doi.org/10.3390/s24113440

[8] McIntosh, T. R., Susnjak, T., Liu, T., Watters, P., Xu, D., Liu, D., & Halgamuge, M. N. (2025). From Google Gemini to OpenAI Q (q-star): A survey on reshaping the generative artificial intelligence (AI) research landscape*. Technologies, 13(2), 51–51. https://doi.org/10.3390/technologies13020051

[9] Yang, L., & Su, R. (2025). From machine learning-based to LLM-enhanced: An application-focused analysis of how social IoT benefits from LLMs. IoT, 6(2), 26. https://doi.org/10.3390/iot6020026

[10] Davar, N. F., Dewan, M. A. A., & Zhang, X. (2025). AI chatbots in education: Challenges and opportunities. Information, 16(3), 235. https://doi.org/10.3390/info16030235

[11] Zviel-Girshin, R. (2024). The good and bad of AI tools in novice programming education. Education Sciences, 14(10), 1089–1089. https://doi.org/10.3390/educsci14101089

[12] Grant, E. S., Shao, S., Shi, Q., & Arinaitwe, M. (2025). Applying artificial intelligence in software development education. 2024 IEEE 6th Eurasia Conference on IoT, Communication and Engineering, 92(1), 26. https://doi.org/10.3390/engproc2025092026

[13] Park, J.-H., Kim, S.-J., & Lee, S.-T. (2025). AI and creativity in entrepreneurship education: A systematic review of LLM applications. AI, 6(5), 100–100. https://doi.org/10.3390/ai6050100

[14] Farquhar, S., Kossen, J., Kuhn, L., & Gal, Y. (2024). Detecting hallucinations in large language models using semantic entropy. Nature, 630(8017), 625–630. https://doi.org/10.1038/s41586-024-07421-0

[15] Lin, T., Wang, Y., Liu, X., & Qiu, X. (2022). A survey of transformers. AI Open, 3, 111–132. https://doi.org/10.1016/j.aiopen.2022.10.001

[16] González Barman, K., Lohse, S., & de Regt, H. W. (2025). Reinforcement learning from human feedback in LLMs: Whose culture, whose values, whose perspectives? Philosophy & Technology, 38(2). https://doi.org/10.1007/s13347-025-00861-0

[17] Imran, M., & Almusharraf, N. (2024). Google Gemini as a next-generation AI educational tool: A review of emerging educational technology. *Smart Learning Environments*, *11*(1). https://doi.org/10.1186/s40561-024-00310-z

[18] Pichai, S., & Hassabis, D. (2024, February 15). *Our next-generation model: Gemini 1.5*. Google. https://blog.google/technology/ai/google-gemini-next-generation-model-february-2024/#sundar-note

[19] Anthropic. (2023, September 9). *Claude's constitution*. Www.anthropic.com. https://www.anthropic.com/news/claudes-constitution

[20] Campbell, S. (2025). Ethics by design: Understanding Claude AI's constitutional framework. SimpleMindedRoBot. https://simpleminded.bot/claude-ethical-ai-framework/

[21] Meta AI. (2024, July 23). Introducing Llama 3.1: Our most capable models to date. Meta AI. https://ai.meta.com/blog/meta-llama-3-1/

[22] Han, S., Wang, M., Zhang, J., Li, D., & Duan, J. (2024). A review of large language models: Fundamental architectures, key technological evolutions, interdisciplinary technologies integration, optimization and compression techniques, applications, and challenges. Electronics, 13(24), 5040–5040. https://doi.org/10.3390/electronics13245040

[23] Ilieva, G., Yankova, T., Klisarova-Belcheva, S., Dimitrov, A., Bratkov, M., & Angelov, D. (2023). Effects of generative chatbots in higher education. Information, 14(9), 492. https://doi.org/10.3390/info14090492

[24] Riemer, K., & Peter, S. (2024). Conceptualizing generative AI as style engines: Application archetypes and implications. International Journal of Information Management, 79, 102824–102824. https://doi.org/10.1016/j.ijinfomgt.2024.102824

[25] Piccialli, F., Chiaro, D., Sarwar, S., Cerciello, D., Qi, P., & Mele, V. (2025). AgentAI: A comprehensive survey on autonomous agents in distributed AI for Industry 4.0. Expert Systems with Applications, 291, 128404. https://doi.org/10.1016/j.eswa.2025.128404

[26] Grammarly. (2024). Introducing generative AI assistance. Grammarly Support. Retrieved from https://support.grammarly.com/hc/en-us/articles/14528857014285-Introducing-generative-AI-assistance

[27] Nebiu Hyseni, L., & Dermaku, A. (2025). Comparative analysis of GitHub Copilot and ChatGPT in web application development: An experimental study. International Journal of Computational and Experimental Science and Engineering, 11(2). https://doi.org/10.22399/ijcesen.1846

[28] Fard, A. (2025). How to use ChatGPT-4: A comprehensive guide. Adam Fard, UX Agency. https://adamfard.com/blog/how-to-use-chatgpt-4

[29] Caswell, A. (2025, May 22). What is Claude? Everything you need to know about Anthropic's AI powerhouse. Tom's Guide. https://www.tomsguide.com/ai/what-is-claude-everything-you-need-to-know-about-anthropics-ai-powerhouse

[30] James, A., Trovati, M., & Bolton, S. (2025). Retrieval-Augmented generation to generate knowledge assets and creation of action drivers. Applied Sciences, 15(11), 6247–6247. https://doi.org/10.3390/app15116247

[31] Williamson, S. M., & Prybutok, V. (2024). The era of artificial intelligence deception: Unraveling the complexities of false realities and emerging threats of misinformation. Information, 15(6), 299. https://doi.org/10.3390/info15060299

[32] Rane, N., Choudhary, S., & Rane, J. (2024). Gemini versus ChatGPT: Applications, performance, architecture, capabilities, and implementation. Journal of Applied Artificial Intelligence, 5(1), 69–93. https://doi.org/10.48185/jaai.v5i1.1052

[33] Dwivedi, Y. K. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management, 71(0268-4012), 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

[34] Anisuzzaman, D. M., Malins, J. G., Friedman, P. A., & Attia, Z. I. (2024). Fine-Tuning LLMs for specialized use cases. Mayo Clinic Proceedings: Digital Health, 3(1). https://doi.org/10.1016/j.mcpdig.2024.11.005

[35] Garrido-Merchán, E. C., Arroyo-Barrigüete, J. L., Borrás-Pala, F., Escobar-Torres, L., Martínez de Ibarreta, C., Ortíz-Lozano, J. M., & Rua-Vieites, A. (2024). Real customization or just marketing: Are customized versions of generative AI useful? F1000Research, 13, 791. https://doi.org/10.12688/f1000research.153129.3

[36] Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-Physical Systems, 3(1), 121–154. ScienceDirect. https://doi.org/10.1016/j.iotcps.2023.04.003.