Please treat it as confidential and share only for interview purposes.

Long-form blog post and LinkedIn post

Long-form blog post

How to Choose the Right AI: A Detailed Comparison of GPT-4o, Claude 3, and Llama 4

Since the late 1960s, when the first chatbot, ELIZA, was launched at MIT and laid the conceptual groundwork for human-computer dialogue systems (HCDS), we have witnessed stunning developments in natural language processing, leading to today's large language models (LLMs). Currently, there are dozens of significant LLMs, and hundreds that could be considered important for various reasons. So how to decide which model is right for your project?

In this blog post, we'll break down the key differences between three of the top LLMs with a detailed, feature-by-feature comparison to help you make an informed choice. Our focus will be on three industry leaders: OpenAI's GPT-4o, Anthropic's Claude 3, and Meta's LLaMA 4. We'll explore their architectures, strengths, limitations, and ideal use cases to give you a comprehensive overview.

OpenAI ChatGPT-4o

Let's start with one of the most prominent players in the field, that is, OpenAI's ChatGPT-4o.

ChatGPT-4o is a cutting-edge large language model designed to deliver advanced natural language understanding and generation, with a strong focus on versatility and safety. It uses reinforcement learning from human feedback (RLHF) combined with supervised learning to continuously refine its responses. This ensures that its answers are accurate, contextually relevant, and aligned with ethical guidelines.

The model excels at maintaining long, coherent conversations, understanding subtle and nuanced queries, and adapting its tone to fit different contexts and user preferences. One

of its standout features is its significantly extended context window—up to 128,000 tokens—allowing it to effectively process lengthy documents and complex interactions.

ChatGPT-4o supports multiple input and output formats, including text and images*, as well as audio input and output**—making interactions richer and more dynamic.

Accessible via the OpenAI API, ChatGPT-4o can be fine-tuned and customized for a wide range of applications, from content generation and coding assistance to tutoring and beyond.

Since ChatGPT-4o is highly versatile, it excels in a variety of tasks such as creative writing, coding help, and conversational agents. Its human-like text generation produces coherent, contextually relevant responses that make interactions feel natural and engaging. Additionally, the model is accessible through a robust and well-documented API, enabling easy integration into numerous applications.

However, these advanced capabilities come with some drawbacks. The model's computational demands can result in significant operational costs, which may be prohibitive for smaller organizations. Like many large language models, ChatGPT-4o can occasionally produce incorrect or misleading information if responses are not carefully monitored. Furthermore, it operates as a "black box," lacking transparency about how it arrives at specific conclusions, which can be a challenge for users needing explainability.

Anthropic Claude 3

Anthropic's Claude 3 (as all Anthropic models!) stands out for its strong emphasis on safety, ethical alignment, and minimizing harmful or biased outputs. It uses a proprietary approach called Constitutional AI, which shapes its behavior based on a clear set of ethical principles inspired by frameworks like the UN's Universal Declaration of Human Rights. By combining supervised learning and reinforcement learning with AI-generated feedback, Claude 3 continuously improves itself while maintaining consistent, reliable, and ethically grounded responses—all without heavy dependence on human labeling.

Claude 3 supports image inputs and boasts an exceptionally long context window of up to 200,000 tokens, allowing it to handle extensive documents and long conversations without losing track of context. While it primarily focuses on text, it also supports a variety of document formats such as PDF, DOCX, and CSV. This makes it well-suited for workflows involving both text and image data analysis, though it does not currently support audio or video inputs.

Available exclusively via API, Claude 3 can be integrated into a wide range of workflows, including automation platforms like Zapier. Unlike some competitors, it does not access

real-time internet data, instead relying solely on its extensive training dataset to generate responses.

Claude 3 is highly regarded for its strong focus on safety and ethical alignment, making it particularly suitable for sensitive applications where minimizing harmful or biased outputs is crucial. The model also excels at aligning closely with user intent, providing responses that meet user expectations effectively. Additionally, Anthropic is committed to transparency and works toward improving the explainability of Claude 3's outputs.

However, the strong emphasis on safety can sometimes lead to less creative or engaging text generation compared to other models. Claude 3's performance may also vary depending on the specific natural language processing task, meaning it may not always be the top performer across all applications. Finally, the model requires significant computational resources, which can make it less accessible for smaller organizations or those with limited infrastructure.

Meta AI LLaMA 4

Meta AI's LLaMA series, including the latest LLaMA 4, is a prominent player in the open-source large language model space. These models range from 7 billion to 65 billion parameters and are built on a transformer-based architecture that emphasizes both efficiency and strong performance. LLaMA models have demonstrated competitive results across a wide array of language tasks, sometimes even outperforming larger proprietary models like GPT-3 on key benchmarks.

Meta's LlaMA 4 is available through various platforms and interfaces, including direct downloads, web interfaces, and integration with existing platforms. It is also accessible through AI assistants like Meta AI, and cloud platforms such as AWS, Snowflake, and Google Cloud.

One of LLaMA's main strengths lies in its open-source nature, making it highly attractive for researchers, academic institutions, and developers who want to customize and fine-tune models to fit specific needs without incurring the high operational costs typical of commercial LLMs. This openness promotes community-driven innovation and continuous improvement, enabling users to adapt LLaMA for specialized applications ranging from scientific research to multilingual processing.

However, the reliance on community contributions for updates and enhancements can lead to variable consistency and slower development compared to commercially backed models. Additionally, commercial support is more limited, which may pose challenges for enterprises requiring dedicated resources. Despite these limitations, LLaMA's accessibility,

scalability, and solid performance make it a valuable tool for experimentation and deployment in research and niche commercial contexts.

How about the use cases?

By now, you might still be wondering: “So, what are the actual use cases? And which model should I pick if they’re all such strong competitors?” Don’t worry—we’re coming to an aid! Here are some practical examples to help guide your decision:

OpenAI’s GPT-4o shines in creative writing, chatbots, and coding assistance. If you’re developing a virtual assistant that requires natural, engaging conversations or generating detailed, coherent blog content, GPT-4o is a great choice. It’s also highly valued by developers for code generation, debugging, and programming help due to its strong understanding of code syntax and logic.

Meta AI’s LLaMA 4 is ideal for research projects and academic work thanks to its open-source nature. For example, a university team looking to fine-tune a model on specialized scientific papers or multilingual datasets will benefit from LLaMA’s flexibility and cost-effectiveness. It’s also perfect for organizations needing scalable, customizable solutions while maintaining data privacy.

Anthropic’s Claude 3 excels where safety and ethical considerations are paramount. Healthcare providers or financial institutions managing sensitive data can trust Claude’s robust safeguards against biased or harmful outputs. It’s also well-suited for customer support in regulated industries, where responsible and trustworthy AI interactions are critical.

What the future will bring?

As we move deeper into 2025, large language models like GPT-4o, Claude 3, and LLaMA 4 are not just evolving—they’re shaping the very direction of AI development. One major trend is the rise of multimodal capabilities, where models can handle not just text but also images, audio, and video in a single pipeline, allowing for more intuitive and context-rich interactions. We’re also seeing greater integration of GenAI into everyday tools—from writing assistants in Microsoft Word to AI copilots in enterprise workflows—bringing AI closer to the average user. At the same time, open-source models like LLaMA are enabling a new wave of academic and industry experimentation, especially in fields like healthcare and scientific research. On the ethical side, models like Claude are setting new standards for safe, transparent AI, a crucial development as governments worldwide introduce more stringent regulations. In short, the future of LLMs is not just about who’s smartest—but about who’s safest, most adaptable, and most usable at scale.

Choosing the right large language model ultimately comes down to your specific needs—whether that's creative flexibility, ethical safety, cost-effective customization, or research-driven innovation. OpenAI's GPT-4o offers unparalleled versatility and conversational fluency, Anthropic's Claude 3 prioritizes safe and principled interactions, and Meta's LLaMA 4 stands out for its openness and adaptability. Each has its strengths, and there is no one-size-fits-all solution. By understanding the unique advantages of each model and matching them to your use case, you'll be better equipped to select the AI that will power your next big idea or project with confidence.

LinkedIn post

The complexities of AI implementation in large enterprises

In 2024, over 41% of large EU enterprises are using AI, according to Eurostat. Yet, implementing AI at scale still remains challenging. Find out the biggest hurdles organizations encounter and the best ways to overcome them!

Consider infrastructure hurdles

Many organizations still rely on outdated infrastructure and face system incompatibilities. Legacy IT environments often lack the computing power, storage, and flexibility required for modern AI tools. To overcome this, investing in cloud solutions and high-performance computing is essential.

Upskill your team

AI adoption lies at the intersection of engineering, domain expertise, and analytics. Therefore, carefully assessing your team's skill set is essential for success. To ensure successful AI implementation, I strongly recommend investing in ongoing education and reskilling.

Prepare financial justification

Many organizations hesitate to invest in AI because the financial benefits aren't always clear. Building a compelling business case means focusing on tangible wins—whether it's cutting costs, managing risks more effectively, or creating new revenue streams. Pinpointing practical use cases by starting small with pilot projects can deliver early proof points to justify broader adoption.

Don’t overlook ethical implementation

AI systems must be designed with fairness, transparency, and accountability in mind. Bias in data or models can lead to unintended harm and reputational risk. Establishing clear ethical guidelines, conducting regular audits, and involving diverse stakeholders in AI development are essential to building trustworthy and responsible AI systems.

AI adoption is a journey requiring strategic planning, investment, and continuous learning. Yet, the gains are worth it! Challenge accepted?