Unlocking Innovation: AWS and Anthropic Push the Boundaries of Generative AI Together

Amazon Bedrock is the best place to build and scale generative AI applications with large language models (LLM) and other foundation models (FMs). It enables customers to leverage a variety of high-performing FMs, such as the Claude family of models by Anthropic, to build custom generative AI applications. Looking back to 2021, when Anthropic first started building on AWS, no one could have envisioned how transformative the Claude family of models would be. We have been making state-of-the-art generative AI models accessible and usable for businesses of all sizes through Amazon Bedrock. In just a few short months since Amazon Bedrock became generally available on September 28, 2023, more than 10K customers have been using it to deliver, and many of them are using Claude. Customers such as **ADP, Broadridge, Cloudera, Dana-Farber Cancer Institute, Genesys, Genomics England, GoDaddy, Intuit, M1 Finance, Perplexity AI, Proto Hologram, Rocket Companies and more** are using Anthropic's Claude models on Amazon Bedrock to drive innovation in generative AI and to build transformative customer experiences. And today, we are announcing an exciting milestone with the next generation of Claude coming to Amazon Bedrock: Claude 3 Opus, Claude 3 Sonnet, and Claude 3 Haiku.

Introducing Anthropic's Claude 3 models

Anthropic is unveiling its next generation of Claude with three advanced models optimized for different use cases. Haiku is the fastest and most cost-effective model on the market. It is a fast compact model for near-instant responsiveness. For the vast majority of workloads, Sonnet is 2x faster than Claude 2 and Claude 2.1 with higher levels of intelligence. It excels at intelligent tasks demanding rapid responses, like knowledge retrieval or sales automation. And it strikes the ideal balance between intelligence and speed – qualities especially critical for enterprise use cases. Opus is the most advanced, capable, state-of-the-art FM with deep reasoning, advanced math, and coding abilities, with top-level performance on highly complex tasks. It can navigate open-ended prompts, and novel scenarios with remarkable fluency, including task automation, hypothesis generation, and analysis of charts, graphs, and forecasts. And Sonnet is first available on Amazon Bedrock today. Current evaluations from Anthropic suggest that the Claude 3 model family outperforms comparable models in math word problem solving (MATH) and multilingual math (MGSM) benchmarks, critical benchmarks used today for LLMs.

1. **Vision capabilities** – Claude 3 models have been trained to understand structured and unstructured data across different formats, not just language, but also images, charts, diagrams, and more. This lets businesses build generative AI applications integrating diverse multimedia sources and solving truly cross-domain problems. For instance, pharmaceutical companies can query drug research papers alongside protein structure diagrams to accelerate discovery. Media organizations can generate image captions or video scripts automatically.
2. **Best-in-class benchmarks** – Claude 3 exceeds existing models on standardized evaluations such as math problems, programming exercises, and scientific reasoning. Customers can optimize domain specific experimental procedures in manufacturing, or audit financial reports based on contextual data, in an automated way and with high accuracy using AI-driven responses. Specifically, Opus outperforms its peers on most of the common evaluation benchmarks for AI systems, including undergraduate level expert knowledge (MMLU), graduate level expert reasoning (GPQA), basic mathematics (GSM8K), and more. It exhibits high levels of comprehension and fluency on complex tasks, leading the frontier of general intelligence.
3. **Reduced hallucination** – Businesses require predictive, controllable outputs from AI systems directing automated processes or customer interactions. Claude 3 models mitigate hallucination through constitutional AI techniques that provide transparency into the model's reasoning, as well as improve accuracy. Claude 3 Opus shows an estimated 2x gain in accuracy over Claude 2.1 on difficult open-ended questions, reducing the likelihood of faulty responses. As enterprise customers rely on Claude across industries like healthcare, finance, and legal research, reducing hallucinations is essential for safety and performance. The Claude 3 family sets a new standard for reliable generative AI output.

Claude 3 Benchmark							
	Claude 3 Opus	Claude 3 Sonnet	Claude 3 Haiku	GPT-4	GPT-3.5	Gemini 1.0 Ultra	Gemini 1.0 Pro
Undergraduate level knowledge *MMLU*	86.8% 5-shot	79.0% 5-shot	75.2% 5-shot	86.4% 5-shot	70.0% 5-shot	83.7% 5-shot	71.8% 5-shot
Graduate level knowledge *GPQA, Diamond*	50.4% 0-shot CoT	40.4% 0-shot CoT	33.3% 0-shot CoT	35.7% 0-shot CoT	28.1% 0-shot CoT	-	-
Undergraduate level knowledge *GSM8K*	95.0% 0-shot CoT	92.3% 0-shot CoT	88.9% 0-shot CoT	92.0% 5-shot CoT	57.1% 5-shot CoT	94.4% Maj1@32	86.5% Maj1@32
Math problem-solving *MATH*	60.1% 0-shot CoT	43.1% 0-shot CoT	38.9% 0-shot CoT	52.9% 4-shot	34.1% 4-shot	53.2% 4-shot	32.6% 4-shot
Multilingual math *MGSM*	90.7% 0-shot	83.5% 0-shot	75.1% 0-shot	74.5% 8-shot	-	79.0% 8-shot	63.5% 8-shot
Code *HumanEval*	84.9% 0-shot	73.0% 0-shot	75.9% 0-shot	67.0% 0-shot	48.1% 0-shot	74.4% 0-shot	67.7% 0-shot
Reasoning over text *DROP, F1 score*	83.1 3-shot	78.9 3-shot	78.4 3-shot	80.9 3-shot	64.1 3-shot	82.4 Variable shots	74.1 Variable shots
Mixed Evaluations *BIG-Bench-Hard*	86.8% 3-shot CoT	82.9% 3-shot CoT	73.7% 3-shot CoT	83.1% 3-shot CoT	66.6% 3-shot CoT	83.6% 3-shot CoT	75.0% 3-shot CoT
Logical reasoning *ARC-Challenge*	96.4% 25-shot	93.2% 25-shot	89.2% 25-shot	96.3% 25-shot	85.2% 25-shot	-	-
Logical reasoning *HellaSwag*	95.4% 10-shot	89.0% 10-shot	85.9% 10-shot	95.3% 10-shot	85.5% 10-shot	87.8% 10-shot	84% 10-shot

Source: Anthropic

Benefits of Anthropic Claude 3 FMs on Amazon Bedrock

Through Amazon Bedrock, customers will get easy access to build with Anthropic's newest models. This includes not only natural language models but also their expanded range of multimodal AI models capable of advanced reasoning across text, images, charts, and more. Our collaboration has already helped customers accelerate generative AI adoption and delivered business value to them. Here are a few ways customers have been using Anthropic's Claude models on Amazon Bedrock:

"We are developing a generative AI solution on AWS to help customers plan epic trips and create life-changing experiences with personalized travel itineraries. By building with Claude on Amazon Bedrock, we reduced itinerary generation costs by nearly 80% percent when we quickly created a scalable, secure AI platform that can organize our book content in minutes to deliver cohesive, highly accurate travel recommendations. Now we can repackage and personalize our content in various ways on our digital platforms, based on customer preference, all while highlighting trusted local voices–just like Lonely Planet has done for 50 years."

— Chris Whyde, Senior VP of Engineering and Data Science, Lonely Planet

"We are working with AWS and Anthropic to host our custom, fine-tuned Anthropic Claude model on Amazon Bedrock to support our strategy of rapidly delivering generative AI solutions at scale and with cutting-edge encryption, data privacy, and safe AI technology embedded in everything we do. Our new Lexis+ AI platform technology features conversational search, insightful summarization, and intelligent legal drafting capabilities, which enable lawyers to increase their efficiency, effectiveness, and productivity."

— Jeff Reihl, Executive VP and CTO, LexisNexis Legal & Professional

"At Broadridge, we have been working to automate the understanding of regulatory reporting requirements to create greater transparency and increase efficiency for our customers operating in domestic and global financial markets. With use of Claude on Amazon Bedrock, we're thrilled to get even higher accuracy in our experiments with processing and summarizing capabilities. With Amazon Bedrock, we have choice in our use of LLMs, and we value the performance and integration capabilities it offers."

— Saumin Patel, VP Engineering generative AI, Broadridge

The Claude 3 model family caters to various needs, allowing customers to choose the model best suited for their specific use case, which is key to developing a successful prototype and later production systems that can deliver real impact—whether for a new product, feature or process that boosts the bottom line. Keeping customer needs top of mind, Anthropic and AWS are delivering where it matters most to organizations of all sizes:

1. **Improved performance** – Claude 3 models are significantly faster for real-time interactions thanks to optimizations across hardware and software.
2. **Increased accuracy and reliability** – Through massive scaling as well as new self-supervision techniques, expected gains of 2x in accuracy for complex questions over long contexts mean AI that's even more helpful, safe, and honest.
3. **Simpler and secure customization** – Customization capabilities, like retrieval-augmented generation (RAG), simplify training models on proprietary data and building applications backed by diverse data sources, so customers get AI tuned for their unique needs. In addition, proprietary data is never exposed to the public internet, never leaves the AWS network, is securely transferred through VPC, and is encrypted in transit and at rest.

And AWS and Anthropic are continuously reaffirming our commitment to advancing generative AI in a responsible manner. By constantly improving model capabilities committing to frameworks like Constitutional AI or the White House voluntary commitments on AI, we can accelerate the safe, ethical development, and deployment of this transformative technology.

The future of generative AI

Looking ahead, customers will build entirely new categories of generative AI-powered applications and experiences with the latest generation of models. We've only begun to tap generative AI's potential to automate complex processes, augment human expertise, and reshape digital experiences. We expect to see unprecedented levels of innovation as customers choose Anthropic's models augmented with multimodal skills leveraging all the tools they need to build and scale generative AI applications on Amazon Bedrock. Imagine sophisticated conversational assistants providing fast and highly-contextual responses, picture personalized recommendation engines that seamlessly blend in relevant images, diagrams and associated knowledge to intuitively guide decisions. Envision scientific research turbocharged by generative AI able to read experiments, synthesize hypotheses, and even propose novel areas for exploration. There are so many possibilities that will be realized by taking full advantage of all generative AI has to offer through Amazon Bedrock. Our collaboration ensures enterprises and innovators worldwide will have the tools to reach the next frontier of generative AI-powered innovation responsibly, and for the benefit of all.

Conclusion

It's still early days for generative AI, but strong collaboration and a focus on innovation are ushering in a new era of generative AI on AWS. We can't wait to see what customers build next.

About the author

Swami Sivasubramanian is Vice President of Data and Machine Learning at AWS. In this role, Swami oversees all AWS Database, Analytics, and AI & Machine Learning services. His team's mission is to help organizations put their data to work with a complete, end-to-end data solution to store, access, analyze, and visualize, and predict.

Source & Image Credit: AWS

Published on : Mon, 4 Mar 2024