

- March 4, 2025
- By James A. Larson program co-chair, SpeechTEK 2021
- Forward Thinking

ChatGPT: A Generative AI Revolution



The landscape of speech technologies underwent a monumental shift in November 2022 when Sam Altman and his team at OpenAI unveiled ChatGPT. This generative AI model leverages vast amounts of data to create humanlike text responses to a wide range of prompts and questions. In the November-December 2024 issue of Speech Technology, <u>I asked readers</u> who they would name to a hypothetical Hall of Fame for speech technologies, and you chose Altman and his team at OpenAI.

Generative AI: A New Era of Interaction

Generative AI marks a significant shift in how users interact with computers. Unlike traditional methods requiring complex commands, ChatGPT allows users to ask questions and receive answers in natural language. It can generate concise responses to simple

queries or produce creative text formats like poems, code, scripts, and more. [Ed note: A new competitor has arrived with a jolt: At press time, the genAI breakthroughs by Chinese company DeepSeek in terms of cost and efficiency had upended the AI world.]

Speech Technology Redefined by Generative AI

The speech technology landscape is experiencing a wave of innovation driven by generative AI. Large language models (LLM), which form the basis of generative AI, contain massive amounts of data never before available to applications. New applications use this data to provide users with new information and services. LLMs are trained to understand the jargon and nuances of specialized domains, such as medicine, nuclear physics, and environmental science, so applications can better respond to user requests. And using advances in AI technology, new classes of applications are appearing; here's a small sample:

Personalized agents: Al agents learn user preferences and assist with daily tasks, like shopping or scheduling appointments.

Content creation agents: Generate content like lectures, podcasts, and audiobooks, or even turn websites into interactive experiences.

Improved accessibility: Generate audio descriptions for videos, provide real-time closed captioning, and adjust conversation speeds for diverse needs.

How Generative AI Works

ChatGPT is powered by neural networks, which mimic the human brain's structure and function. These networks learn patterns and make decisions by processing information through interconnected layers. Here are the steps:

- **Training:** LLM developers train LLMs on massive amounts of text data to understand human language patterns, grammar, and context.
- **Fine-tuning:** To teach an AI model to understand complex tasks, like analyzing drug structures or health data, engineers provide it with thousands of examples. These examples include specific instructions (called prompts) and the correct responses, using the specialized language needed for the task. This way, the AI learns to answer questions accurately using the right concepts and terminology.
- **Generation:** Prompt engineers submit a prompt, possibly augmented with program code, to generate a conversational agent or application that uses the LLM.
- **Deployment:** Developers invoke the conversational agent or application.

Addressing Generative AI Challenges

One major challenge is integrating generative AI with other knowledge sources and applications. Here are some recent advancements:

- **2021:**Retrieval-augmented generation (RAG), developed by Patric Lewis, incorporates information from external sources.
- **2022:**LangChain, developed by Harrison Chase, provides a framework for building LLM-powered applications.
- **2023:**Program-aided language models (PAL), developed by Luyu Gao, leverage LLMs with programming languages for specific tasks.
- **2024:**DeepSeek 2, an open-source LLM developed by DeepSeek AI, and OpenAI's closed-source LLM o1 showcase reasoning capabilities.

And as for conversational agents built upon generative AI, further development is needed in three key areas:

- **Multimodal agents**that understand and respond using various modalities, like sight, sound, and touch.
- **Autonomous agents**(also called agentic agents) that can perform tasks, make decisions, and solve problems with minimal human intervention.
- **Collaborative agents**that each specialize in specific services while collaborating with others to leverage combined capabilities.

GenAI will also affect the role of programmers, who will likely transition into prompt engineer roles, analyzing systems and crafting prompts to leverage LLMs effectively. They will still write code but will also collaborate with AI to create new functionalities.

Generative AI, pioneered by Sam Altman and others, is transforming how we interact with computers and promises to revolutionize various industries. As technology evolves, we can expect even more innovative applications and a future where AI seamlessly integrates into our daily lives.

James A. Larson, PhD, is an independent voice technology expert and can be reached at jim42@larson-tech.com.