NRI

The Evolution of Conversational GenAI

Ryota Ogawa
21 Feb 2025

Nomura Research Institute, Ltd.

Executive Summary

Ryota Ogawa

Senior Associate

IT Platform Technology Strategy
Department

The advent of generative AI, particularly ChatGPT, has spurred corporate adoption of large language models (LLMs). Recent overseas advancements in conversational AI include pairing of LLMs with reinforcement learning and specialized LLMs developed for specific applications.

Large language models' promise

The advent of ChatGPT has catalyzed a global GenAI boom. LLMs in particular, a core building block of GenAI, have elicited much excitement in the corporate sector. Some companies, mostly large corporations, have started to use LLMs in applications such as chatbots and in-house information retrieval systems. Before GenAI, chatbots were programmed to select, based on keywords in users' input, the optimal response to inquiries from a fixed set of responses. As such, they could be flummoxed by minor differences in how inquiries were worded. LLMs, by contrast, are able to respond as if they understand questions' intent even when the question is posed in colloquial language.

LLM applications are now starting to broaden from textual to voice input/output. For example, Softbank's humanoid robot Pepper can converse vocally. Pepper for Care, a Pepper variant deployed in hospitals and nursing homes since February 2024, uses ChatGPT as its conversation engine. Softbank claims that Pepper for Care can engage patients in lively, personalized conversations on a wide variety of topics.

Advancements in conversational AI

While LLM-powered voice conversations have recently become reality, use cases in which conversational AI can take the place of humans are still limited. In a face-to-face sales scenario, for example, a conversational AI agent would need some technique to variably choose its words in response to feedback from the customer, such as the customer's concerns about the future, plans in terms of navigating life events or reactions to what the agent has just said. Developing a general-purpose conversational AI endowed with such flexibility and impromptu responsiveness is a formidable challenge. Conversational AI agents are now

emerging that can converse as effectively as a human, if not even more so, by personalizing conversations to their users' circumstances or using LLMs trained to possess specialized skills within a limited subject-matter domain.

One example is ElliQ, an empathetic companion care robot for the elderly sold by Intuition Robotics, an Israeli startup founded 2016. ElliQ, equipped with a camera and microphone for daily user monitoring, undergoes continuous reinforcement learning on observational data. This iterative training enables it to develop a decision-making model that optimally determines when to engage vocally, tailoring interactions to the user's personality, preferences, and daily behavioral patterns. A 2023 study conducted by the New York State Office for the Aging with over 800 elderly participants found that ElliQ reduced loneliness by 95% and substantially improved well-being among its users[1].

ElliQ3, released in January 2024, integrates a proprietary LLM[2]. The new LLM reportedly enables ElliQ3 to engage in conversations that more effectively reduce loneliness by encouraging social connections and addressing topics of interest to the elderly.

Hippocratic AI, a US startup founded in 2023, is developing conversational AI agents specifically for the healthcare sector to mitigate labor shortages at healthcare facilities[3]. Hippocratic AI is developing LLMs in collaboration with physician and nurse advisory councils, placing utmost priority on ensuring patient safety. Its LLMs have been evaluated by some 4,500 nurses and over 270 physicians. Unlike ChatGTP and other LLMs built from public information scraped from the Internet, Hippocratic AI's LLMs are a type of specialized LLM trained on datasets comprising ideal conversations with patients. In December 2024, Hippocratic AI formed a strategic partnership with the Nurses on Boards Coalition in the aim of educating nurses about its conversational AI agents and promoting their adoption at healthcare facilities.

Non-diagnostic tasks such as inquiring about patients' general state of health or medication compliance generally require an approach that induces the patient to speak candidly about their health concerns or symptoms. In a survey that asked respondents to evaluate how empathetic Hippocratic AI's agents are in comparison to human nurses, the agents performed on a par with the nurses overall and even scored a bit better than the nurses in terms of bedside manner[4].

NOTE

1) New York State Office for the Aging, *NYSOFA's Rollout of AI Companion Robot ElliQ Shows 95% Reduction in Loneliness*, August 1, 2023 (https://aging.ny.gov/news/nysofas-rollout-ai-companion-robot-elliq-shows-95-reduction-loneliness).

2) https://voicebot.ai/2024/01/09/intuition-robotics-releases-elliq-3-robot-with-upgraded-hardware-augmented-by-generative-ai/

3) https://www.hippocraticai.com/sarah

4) The AI agents received an 86.32% rating vs. 82.67% for the human nurses. (https://www.hippocraticai.com/foundationmodel).

Toward widespread utilization of conversational AI

5) Developed using NTT's Tuzumi LLM and Microsoft's Azure OpenAI Service, the AI agent was demoed at Docomo Open House '24.

Use of specialized LLMs built by subject-matter experts and integration of an LLM with a decision-making model that utilizes reinforcement learning are examples of advancements in conversational AI. Another example is a conversational AI agent developed by NTT Docomo for customer-service settings[5]. While conversing with a customer, the agent infers emotions from the words used by the customer and modulates its responses accordingly. This innovation is another milestone in Japan's long history of researching emotion recognition through audio analysis.

6) Moffatt v. Air Canada, 2024 BCCRT 149 (https://decisions.civilresolutionbc. ca/crt/crtd/en/item/525448/index.do).

Meanwhile, AI hallucination (factually false output) remains a risk. In one well-known example of this risk, Air Canada lost a lawsuit filed by a passenger who was misinformed about airfare refund policies by a chatbot. The court ordered Air Canada to pay US$483 in damages plus $27 interest and court costs of $93[6]. Although the suit was filed in small claims court, it raises the question of whether conversational AI agents can be trusted to process important matters like, e.g., life insurance claims.

Some companies have implemented guardrails to keep LLMs from causing harm. One example is Hippocratic AI. In addition to developing specialized LLMs, Hippocratic AI enhances the safety of its AI agents by deploying an ensemble of AI models to double-check conversation content and correct errors. Such guardrails are essential to safe LLM usage. As conversational AIs that utilize LLMs in concert with peripheral technologies become increasingly sophisticated, safety and reliability upgrades cannot be neglected.

about NRI

Founded in 1965, Nomura Research Institute (NRI) is a leading global provider of system solutions and consulting services with annual sales above $4.9 billion. NRI offers clients holistic support of all aspects of operations from back- to front-office, with NRI's research expertise and innovative solutions as well as understanding of operational challenges faced by financial services firms. The clients include broker-dealers, asset managers, banks and insurance providers. NRI has its offices globally including New York, London, Tokyo, Hong Kong and Singapore, and over 17,400 employees.

For more information, visit https://www.nri.com/en

https://www.nri.com/en/knowledge/publication/fis/lakyara/