

# Approaches to Generative AI Technology: From Foundational Technologies to Application Development and Guideline Creation

As expectations for generative AI continue to rise, NEC is engaging in advanced applications such as supporting expert tasks with generative AI and automating video analytic tasks by integrating face and object recognition technologies. In support of these initiatives, NEC is not only improving the core generative AI technology but also advancing the development of related technologies that facilitate its application in various tasks and ensure its safe use. Furthermore, NEC has constructed an AI supercomputer capable of processing these technologies at high speeds. Additionally, to ensure societal acceptance of generative AI, we are developing guidelines and risk management methods aligned with global standards. In this paper, we will introduce NEC's initiatives in generative AI, showcasing them with real-world examples.

## SAKAI Junji

Senior Principal Engineer  
Global Innovation Business Unit

---

## 1. Introduction

The year 2023 saw generative AI becoming widely recognized and discussed among the general public. The introduction of ChatGPT, capable of conducting intelligent conversations, has led to its extensive trial across various applications such as document creation and summarization. Furthermore, companies like Microsoft and GitHub have integrated large language models (LLMs) into their office applications and software development environments, aiming to enhance user productivity. Additionally, the competition to develop proprietary high-performance LLMs is intensifying among tech companies and research institutions.

In this paper, we will introduce NEC's various initiatives in generative AI, including LLMs, covering everything from foundational technologies and application development to aspects of societal utilization.

## 2. The Rapid Expansion of Generative AI in the Marketplace

While much recent attention has focused on the inter-

active use of ChatGPT for text generation, the practical applications of generative AI span text, video, and even software code. We will examine these in turn.

Starting with text, NEC has developed technology that combines general LLMs with domain expertise to enable AI assistance for tasks that previously required human experts. For example, an LLM trained on medical terminology introduced in a hospital setting can summarize doctor-patient conversations and automatically generate medical records. In cybersecurity, LLMs can take over the analysis of the huge volume of daily vulnerability reports to readily assess risks for an organization's IT system. LLMs can also dramatically accelerate materials and drug development by screening libraries of chemistry and pharmacology papers to identify promising new compound candidates.

Combining video data with LLMs is an area expected to see major growth and advancement. NEC is developing various applied technologies that fuse our expertise in face and object recognition technologies with generative AI capabilities. For instance, by having LLMs generate descriptive text based on recognition results, solutions become possible that identify people in video

analytics and verbalize their actions and situational context. Furthermore, we have successfully developed an application technology that can extract traffic accident scenes from dashcam records and automatically generate accident investigation reports by contextualizing and verbalizing the significance of the recognized results relevant to each scene. Additionally, with the vast number of photos shared on social media, LLMs can be used in an interactive way to sort through these images when a disaster strikes. They help pinpoint the photos showing the affected areas, speeding up the assessment of damage and aiding early response efforts.

In addition to these cutting-edge initiatives, NEC has also begun trialing the use of generative AI to boost productivity in day-to-day business operations. In May 2023, NEC launched the NEC Generative AI Service (NGS) as an in-house LLM service for use by our employees in their daily tasks, including creating progress reports and setting work objectives, driven by their own ideas. Additionally, we have established mechanisms and internal guidelines to ensure the safe use of LLMs in tasks involving confidential information.

Beyond text, generative AI can produce software code and is expected to greatly enhance developer productivity. This paper also highlights NEC Group's efforts to incorporate generative AI into software and system development workflows.

### **3. Foundational Technologies Driving Generative AI Potential**

In July 2023, NEC completed the development of its proprietary LLM and has already started offering it as the generative AI "cotomi." Cotomi is characterized by its compact size relative to its performance, making it easily customizable for client-specific needs. It is also equipped to handle lengthy texts, making it highly effective for handling large volumes of business documents.

Looking towards future business applications, we see four main directions for the development of LLM technology.

The first is the ability to handle data beyond text. As mentioned earlier, solutions using video data are expected to expand further in the future. NEC is currently developing a foundational model for generative AI (Foundational Vision LLM) that can integrate and process both video and text data.

The second direction involves improvements in fine-tuning. Generally, LLMs are initially created as versatile foundational models and later customized for specific tasks through fine-tuning, which involves further training with data from the targeted application. How-

ever, in reality, there are often situations where there is not enough data available for fine-tuning, or there is a reluctance to share confidential data for additional training. To address these challenges, NEC is advancing research and development in technologies that can fine-tune effectively with limited data and perform fine-tuning without disclosing sensitive information.

The third direction focuses on improving prompt instructions. When using LLMs, instructions are given in the form of prompts, which are textual cues. There are limitations to the size of prompts that can be given to an LLM, and effectively communicating within these constraints requires specialized know-how. Therefore, NEC is also working on technologies that facilitate prompt instructions, such as efficiently extracting information from compact prompts.

The fourth direction is about developing technologies to use LLMs with confidence. Since LLMs analyze training data stochastically and do not actually understand the content, they can sometimes produce answers that are not based on facts or are socially undesirable. This poses a significant concern when using LLMs in critical decision-making processes. To mitigate these issues, NEC is developing technologies that provide justifications for answers and methods to evaluate whether the responses have any societal biases.

In addition to developing the LLM itself and related technologies, we are also focusing on creating the computational environments necessary for running LLMs at high speeds. LLMs consist of complex and vast data structures, and to fully leverage their capabilities, it is crucial to optimally adjust the 13 billion or more parameters that these data structures contain. This adjustment requires extensive computational resources, which is often considered a significant challenge when developing LLMs in-house. NEC has long recognized this challenge and has designed and constructed its own supercomputer, known as an AI supercomputer, with an architecture specifically tailored for the research and development of AI technologies, including LLMs. This AI supercomputer has been in full operation since March 2023, serving as a crucial behind-the-scenes force that significantly accelerates the development of cotomi and related technologies.

### **4. For AI Technology to Permeate Society**

While AI technologies, including LLMs, are expected to boost efficiency across society and enrich lives, concerns have also been raised about their potential negative aspects. These include decisions of significant matters being made automatically without human involvement and

the possibility of their use in criminal activities.

Looking beyond NEC, discussions among countries, such as those at the G7 Hiroshima Summit, have revolved around establishing global guidelines for the use and regulation of AI. NEC has a long history of involvement in international standard-setting activities, particularly in telecommunications, and is now participating in efforts to develop standard specifications for the broader adoption of AI through organizations like the Institute of Electrical and Electronics Engineers (IEEE), the International Organization for Standardization/International Electrotechnical Commission (ISO/IEC), and the European Telecommunications Standards Institute (ETSI).

Internally, to ensure that NEC's business activities utilizing AI respect human rights, the NEC Group AI and Human Rights Principles has been established, guiding the formation of internal structures and related regulations. Through the operation of the Digital Trust Advisory Council, we also integrate diverse external expert opinions into the design of our AI projects.

Given that it is not possible to completely eliminate the risks associated with AI, it is important to conduct a preliminary assessment of potential social risks when developing AI-driven solutions. In collaboration with the University of Tokyo, NEC has pilot-tested an AI risk management tool and confirmed its effectiveness. We are considering offering this system as an AI risk management service in the future.

## 5. Conclusion

Generative AI is anticipated to profoundly impact existing IT and AI solutions, with technological development and real-world integration advancing simultaneously at an unparalleled pace. In this paper, we have primarily concentrated on the technical aspects of generative AI. Moving forward, NEC will continue to showcase its wide-ranging initiatives in generative AI through detailed articles and exhibitions. Stay tuned for further updates.

---

\* Microsoft is a registered trademark or a trademark of Microsoft Corporation in the U.S. and other countries.

\* ChatGPT is a trademark of OpenAI.

\* All other company names and product names that appear in this paper are trademarks or registered trademarks of their respective companies.

# Information about the NEC Technical Journal

Thank you for reading the paper.

If you are interested in the NEC Technical Journal, you can also read other papers on our website.

Link to NEC Technical Journal website

Japanese

English

## Vol.17 No.2 Special Issue on Revolutionizing Business Practices with Generative AI

– Advancing the Societal Adoption of AI with the Support of Generative AI Technologies

Remarks for Special Issue on Revolutionizing Business Practices with Generative AI  
Approaches to Generative AI Technology: From Foundational Technologies to Application Development and Guideline Creation

### Papers for Special Issue

#### Market Application of Rapidly Spreading Generative AI

NEC Innovation Day 2023: NEC's Generative AI Initiatives  
Streamlining Doctors' Work by Assisting with Medical Recording and Documentation  
Using Video Recognition AI x LLM to Automate the Creation of Reports  
Understanding of Behaviors in Real World through Video Analysis and Generative AI  
Automated Generation of Cyber Threat Intelligence  
NEC Generative AI Service (NGS) Promoting Internal Use of Generative AI  
Utilization of Generative AI for Software and System Development  
LLMs and MI Bring Innovation to Material Development Platforms  
Disaster Damage Assessment Using LLMs and Image Analysis

#### Fundamental Technologies that Enhance the Potential of Generative AI

NEC's LLM with Superior Japanese Language Proficiency  
NEC's AI Supercomputer: One of the Largest in Japan to Support Generative AI  
Towards Safer Large Language Models (LLMs)  
Federated Learning Technology that Enables Collaboration While Keeping Data Confidential and its Applicability to LLMs  
Large Language Models (LLMs) Enable Few-Shot Clustering  
Knowledge-enhanced Prompt Learning for Open-domain Commonsense Reasoning  
Foundational Vision-LLM for AI Linkage and Orchestration  
Optimizing LLM API usage costs with novel query-aware reduction of relevant enterprise data

#### For AI Technology to Penetrate Society

Movements in AI Standardization and Rule Making and NEC Initiatives  
NEC's Initiatives on AI Governance toward Respecting Human Rights  
Case Study of Human Resources Development for AI Risk Management Using RCModel

### NEC Information

2023 C&C Prize Ceremony



Vol.17 No.2

June 2024

Special Issue TOP