

April 1, 2024

## Research and Development Center for Large Language Models Established at National Institute of Informatics

— Accelerating R&D to Develop Domestic LLMs and Ensure Transparency and Reliability of Generative AI Models —

The National Institute of Informatics (NII, Director-General: KUROHASHI Sadao, Chiyoda-ku, Tokyo, Japan) has established the Research and Development Center for Large Language Models (LLMC, Director: NII Director-General KUROHASHI Sadao) on April 1, 2024. The LLMC conducts research and development of large language models (LLMs).

In May 2023, NII set up the LLM Research Group (LLM-jp) which includes a wide range of people from domestic research institutes, private enterprises, and other organizations. Since then, NII keeps advancing the research and development of open generative AI. Recently, a new center has been established within NII for implementing the "R&D Hub Aimed at Ensuring Transparency and Reliability of Generative AI Models" project of the Ministry of Education, Culture, Sports, Science and Technology. Now NII is structurally ready to help up-and-coming AI researchers concentrate on the research and development of LLMs.

LLMC is first developing an LLM with 175 billion parameters which is equivalent to GPT-3 level with the goal of completing around the summer of 2024. We are also promoting research activities related to the development of large language models that are open and Japanese-Proficient, including advanced R&D activities to ensure the transparency and reliability of LLMs. And through these activities, we will accumulate a series of knowledge and experience that will contribute to the evolution of AI and, ultimately, to the creation of revolutionary innovations for the future.

Large language models (LLMs) have been increasingly used in all industries. They have the potential to change existing industrial bases drastically as foundation models and are also expected to function as a knowledge base that is indispensable for extensive science and technology research. Now, however, corpus data of major LLMs are kept private and not opened to the public. The models and the behavior of the models are a black box. Consequently, there are still a lot of problems such as hallucination and biases. The data used in the training of the major LLMs are focused on English and their ability to understand and generate Japanese language content is relatively low. In Japan, there are few development cases of LLM with the size of 100 billion parameters. And the lack of research examples leads to the delay of the acquisition of knowledge in the area of LLM development.

**National Institute of Informatics**

大学共同利用機関法人 情報・システム研究機構  
国立情報学研究所  
総務部企画課 広報チーム  
〒101-8430 東京都千代田区一ツ橋 2-1-2  
直通 : 03-4212-2164 E-Mail : media@nii.ac.jp

Web: <https://www.nii.ac.jp>  
X(旧 Twitter): @jouhouken  
facebook: <https://www.facebook.com/jouhouken>

The LLM Research Group (LLM-jp) having been led by NII developed and released its first model of LLM with 13 billion parameters in October 2023 and the release has contributed to LLM development in Japan. Now, NII has established LLMC as a base for implementing the " R&D Hub Aimed at Ensuring Transparency and Reliability of Generative AI Models" project of the Ministry of Education, Culture, Sports, Science and Technology, to set up a system to promote R&D to ensure transparency and reliability of generative AI models (Fig.1). Based on the knowledge about LLM development obtained through LLM-jp, we are building a knowledge hub where researchers and engineers can cooperate and are also creating an environment to nurture R&D capabilities related to generative AI models.

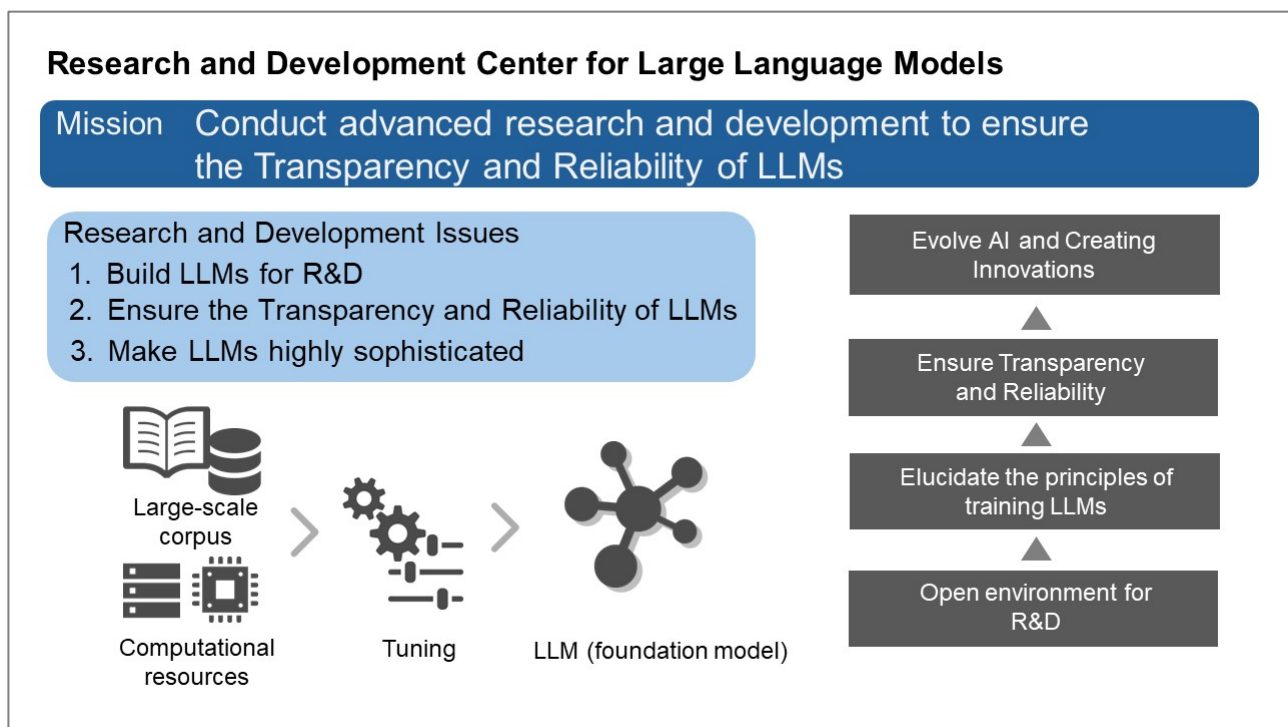


Fig. 1 Outline of LLMC

LLMC conducts the following R&D activities.

1. Build LLMs for R&D

LLMC makes LLMs fully open to researchers as well as prepares corpus data, computing environments, and assessment benchmarks for R&D.

2. Ensure the transparency and reliability of LLMs

LLMC ensures the transparency and reliability of generative AI by elucidating generative AI behavioral principles and developing technologies to control the impact of data alteration, data bias, etc.

3. Make LLMs highly sophisticated

LLMC ensures that R&D activities, such as domain adaptation and making models lighter, aid the development of generative AI models.

We keep advancing the R&D activities of LLM-jp. We are planning to release of 175 billion parameters scale LLM (equivalent to GPT-3) around this summer. By utilizing our research results in experimental LLM models, we seek to establish a method for creating transparent and reliable generative AI models. At the same time, we accumulate knowledge and experience as it contributes to the evolution of AI and innovation toward the future.

### Overview of the Center

Name: Research and Development Center for Large Language Models

Director: KUROHASHI Sadao (Director-General, National Institute of Informatics / Program-Specific Professor, Kyoto University)

Vice Director: AIZAWA Akiko (Vice Director-General, National Institute of Informatics / Professor, Digital Content and Media Sciences Research Division)  
TAKEDA Koichi (Project Professor, National Institute of Informatics)

Many researchers from universities, private companies and other institutions will join in our R&D activities through LLM-jp Research Group. We will continue to enhance our research organization. If you want to know more about us, please visit our website.

Link: <https://www.nii.ac.jp/research/centers/llmc/>

### Comment from KUROHASHI Sadao, LLMC Director

“In May 2023, the National Institute of Informatics established the LLM Research Group (LLM-jp) in which anyone who agrees with our philosophy can join. We’ve also disclosed all of our model’s mechanisms, development data, tools, technical documents and other materials, including the development processes, discussions and even failures. Thanks to the participation from various universities and companies, the number of participants in the LLM-jp has exceeded 1000 people. The recognition of our activities has led to the establishment of LLMC. We will prepare the necessary computational resources and strive to elucidate the principles of generative AI and establish methods of developing LLM. We hope our new research center will become the place where talented and energetic young researchers can join and the hub of research and development of LLMs in Japan. We also want to build an international cooperation system of researching open LLMs.”

<Media Contact>

**National Institute for Informatics Research Organization of Information and Systems**

Publicity Team, Planning Division, General Affairs Department

TEL : 03-4212-2164 E-mail : [media@nii.ac.jp](mailto:media@nii.ac.jp)

**National Institute of Informatics**