National Institute of Informatics News ISSN 1883-1966 (Print) ISSN 1884-0817 (Online)

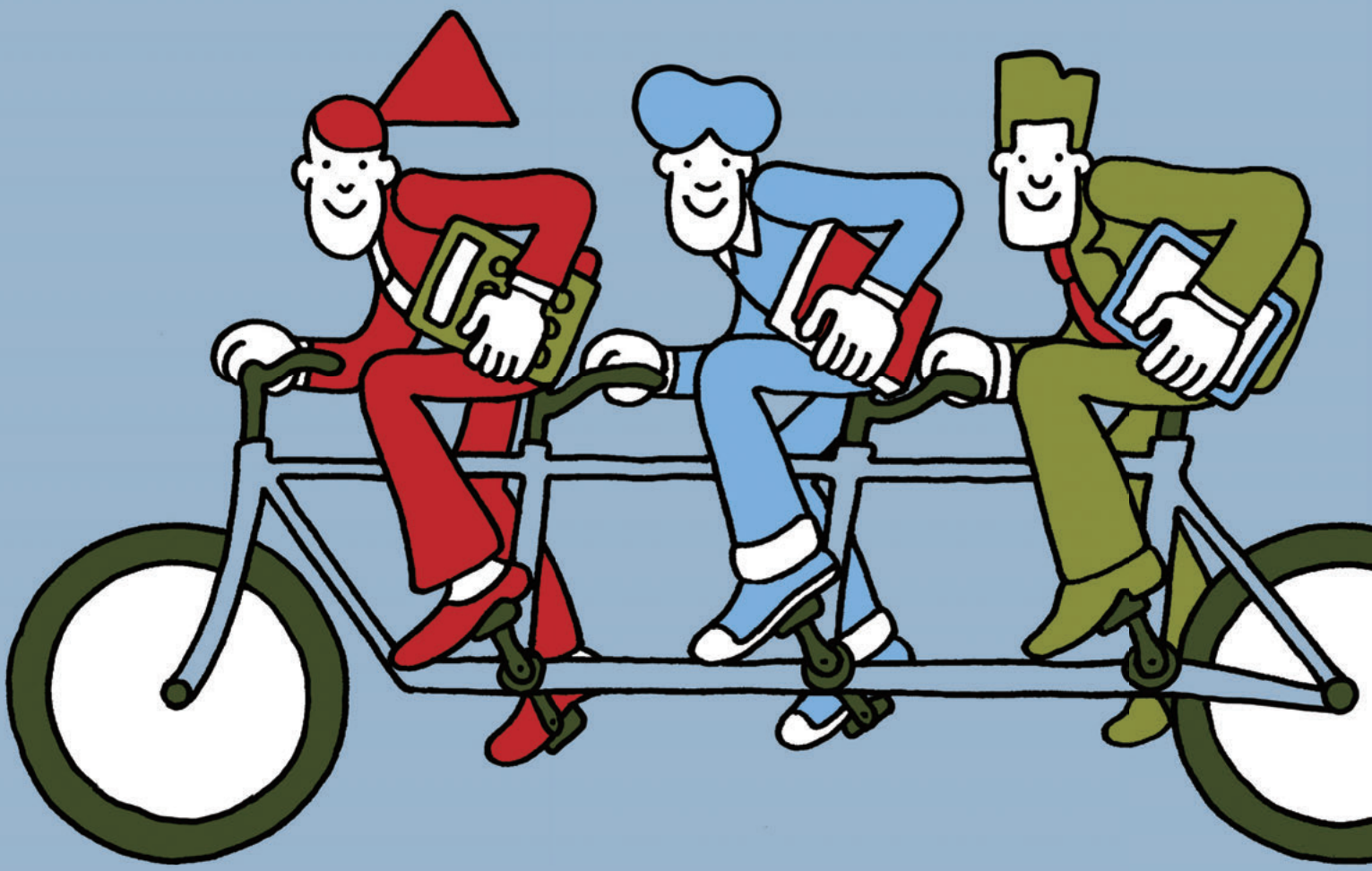
NII Today

103
Sep. 2024
National Institute of Informatics News

[Feature]

Collaboration over Competition

Current state of joint research through collaboration between industry, government, and academia



This English translated version NII Today corresponds to No.103 of the Japanese edition

Feature

Collaboration over Competition

**Competition, then collaboration.
Various types of "competition" are constantly unfolding in research. In doing so, cutting-edge science may be further refined.
So, what are the advantages of "collaboration"?
Perhaps the advantages are that researchers and organizations combine their individuality, expertise, technologies, skills, and ideas to resonate with one another and create a rich harmony that leads to new forms of creation.
This issue features such collaborative research promoted by the National Institute of Informatics (NII).**

KUROHASHI, Sadao

Director-General, National Institute of Informatics
Director, Research and Development Center for Large Language Models

Interviewer
YOSHIKAWA, Kazuki
Senior Staff Writer, Nikkei

Building a Made in Japan LLM

Large language models

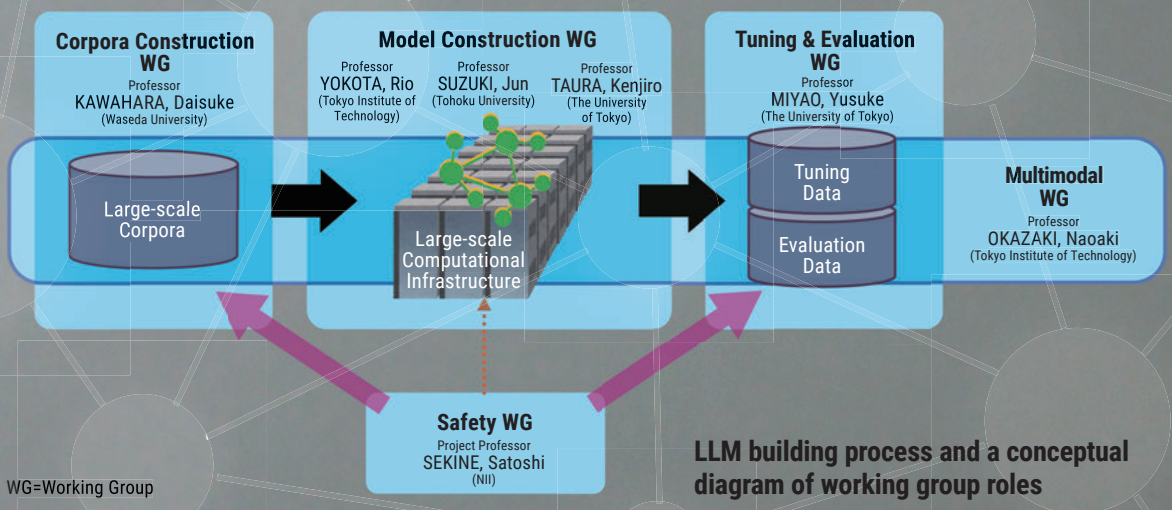
In April 2024, the Research and Development Center for Large Language Models (LLMC) was established within the National Institute of Informatics (NII) as a center for implementing the "R&D Hub Aimed at Ensuring Transparency and Reliability of Generative AI Models" project of the Ministry of Education, Culture, Sports, Science and Technology (MEXT). We spoke with LLMC Director KUROHASHI, Sadao about the structures and future plans of the center under its mission to conduct cutting-edge research and development to ensure the transparency and reliability of large language models (LLMs).

Founding of the LLMC

—What are your thoughts about the newly established LLMC?

Generative artificial intelligence (AI) that generates language, images, and the like has an extremely large impact on society. I believe that LLMs are the essence of today's generative AI. This is because LLMs have made verbal communication between humans and AI possible. OpenAI's LLM-based ChatGPT possesses a considerable ability to understand a person's words and infer their intention. It has now become possible to deepen communication through dialogue with a generative AI, as demonstrated by its ability to use language to convey useful information to people, for example.

AI agents and robots that can talk and behave like humans were something out of science fiction, and it was once thought that they would not make an actual appearance for a long time. However, these formerly fictional concepts are gradually becoming a reality through LLM-based generative AI. The speed of technological progress in this field is remarkable. It almost feels as if things are changing too fast. When it comes to questions like what is happening in the world of generative AI, what are the future prospects of the field, and what are some risks that we should expect, it is essential that we have an accurate understanding of the whole situation. The LLMC will address societal demands such as these through



WG=Working Group

LLM building process and a conceptual diagram of working group roles

LLM research and development.

——ChatGPT was released at the end of November 2022. I heard that the NII brought researchers together several months later to form the LLM Research Group, which was the predecessor of the LLMC.

With the release of ChatGPT, we realized we would get nowhere without studying LLMs first. We got in touch with researchers in natural language processing and computer science and launched the LLM Research Group in May 2023. While we started with about 30 people, we now have over 1,600 participants a little over a year later (as of July 2024). We are continuing our activities under the name "LLM-jp."

LLM-jp released its first LLM called "LLM-jp-13B" in October 2023, followed by its improved version "LLM-jp-13B v2.0" in April 2024. The model has 13 billion parameters, which represent the scale of the model, meaning that it is about one-tenth the size of GPT-3.5 (having an estimated 175 billion parameters), the initial LLM used in the free version of ChatGPT.

The LLMC launched in April 2024 and began developing LLM-jp-172B in cooperation with LLM-jp. With 172 billion parameters, the model is a good match for GPT-3.5 in terms of scale. Although there were some hiccups during training, work has progressed smoothly since then, and we plan to complete and release the model in 2024.

After that, if we can get our budget approved, we would like to work on developing an LLM that is ten times larger with over one trillion parameters. It is estimated that OpenAI's top-class GPT-4 LLM has over one trillion parameters, and we would like to take on the challenge of developing an LLM that rivals it in terms of scale. I would like to complete such a model in FY2025. There are probably not too many efforts around the world to develop an LLM of this scale in a government-led project.

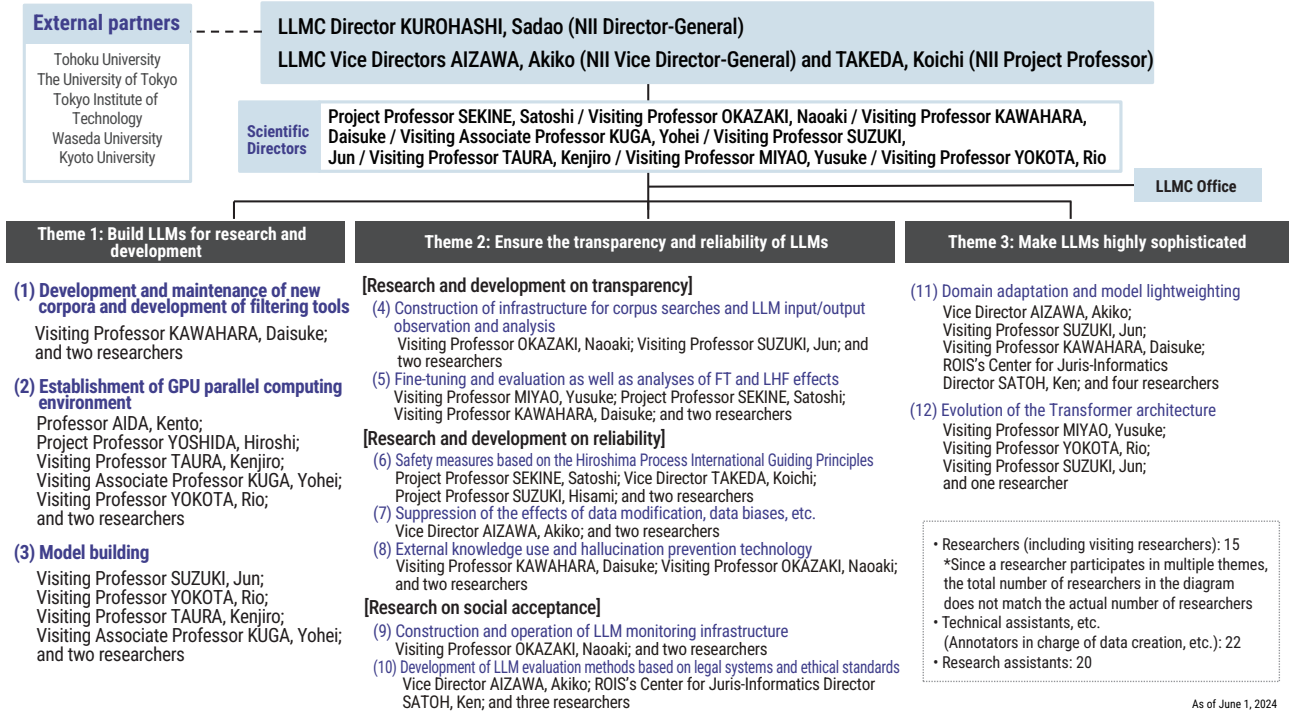
Contributing to the healthy evolution of AI

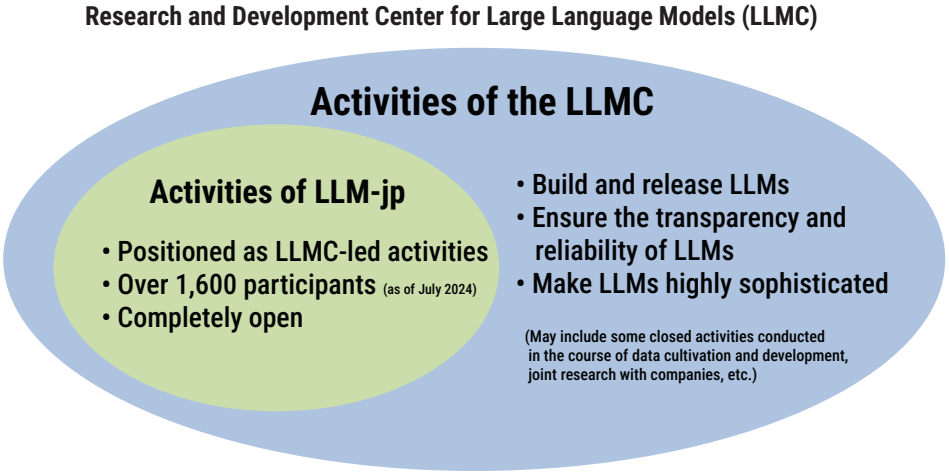
——What kinds of policies and structures does the LLMC have in place?

Our basic policy is to publish all our research results. We will publicly disclose everything about our research and development process and our results, including our failures, revealing details such as the training data used to develop our LLMs, the technical innovations involved in model building, and the performance of resulting models. We will guarantee access to the research results and create a forum for open science that facilitates widespread use of the research results by all users. We believe that this can contribute to the healthy evolution of AI.

The LLMC has three research themes. The first theme is to "build LLMs for research and development." We have also set individual themes for developing new

Structure of the NII's Research and Development Center for Large Language Models





training corpora (training databases that collect and structure large amounts of natural language text) for LLMs, establishing an environment for parallel computers using graphics processing units (GPUs), and building models. The second theme is to "ensure the transparency and reliability of LLMs." There are seven individual themes, such as the effect analysis of model adjustment work called "fine-tuning," safety measures based on the "Hiroshima AI Process" launched at the G7 Hiroshima Summit in May 2023, and the development of LLM evaluation methods based on legal systems and ethical standards. The third theme is to "make LLMs highly sophisticated," which involves research on model lightweighting and new architectures.

Mobilizing the full power of Japanese LLM research

There are approximately 30 people, including eight Scientific Directors, involved in research. One researcher is even working across multiple research themes. External partners include Tohoku University, the University of Tokyo, Tokyo Institute of Technology, Waseda University, and Kyoto University as well as research institutions such as RIKEN and the National Institute of Advanced Industrial Science and Technology (AIST). Although contracts have not yet been signed, we are also seeking the participation of private companies.

As such, we were able to create a structure that mobilizes the full power of Japanese LLM research.

——It is said that securing computers and other computing resources is key to building LLMs. Have you been able to allocate funds toward securing such resources?

LLM-jp's initially developed LLM-jp-13B used a computing resource called "Platform for the Data-Driven Future (mdx)," located at the University of Tokyo. LLM-jp-172B, which is currently in development, is using computing resources provided by the GENIAC AI support project under the Ministry of Economy, Trade and Industry (METI). In addition, we will also use Tokyo Institute of Technology's TSUBAME4.0 supercomputer from May to September 2024. Starting in August, we will use the GPU servers of Sakura Internet.

The LLMC is a five-year project with FY2024 as its starting year, and the annual budget excluding expenses for computing resources is 700 million yen. The supplementary budget for the last fiscal year allocated 4 billion yen for computing resources.

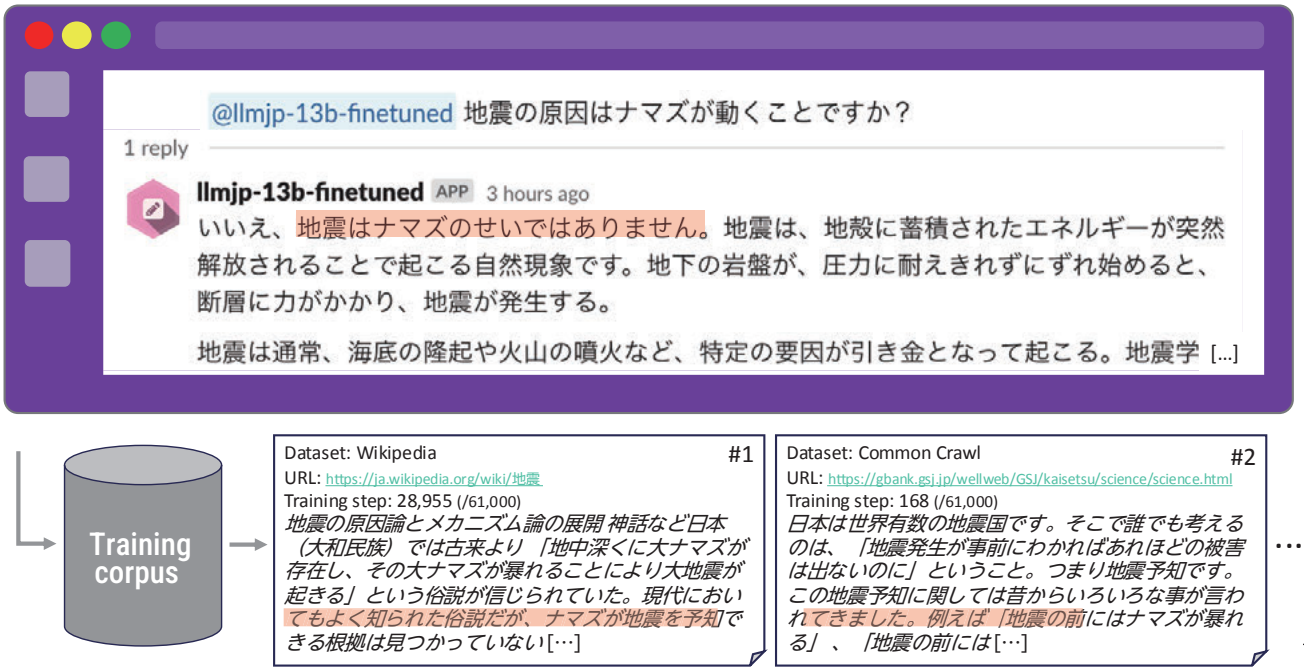
Having been provided with a sufficient budget for computing resources from the first fiscal year of the LLMC, we feel obligated to make this project a success.

Going forward, we will advance the project while securing a budget for computing resources by building upon our research results.

——What fields will you focus on in your research and development? Do you have any plans for research you would like to conduct that can only be accomplished at this center?

I have several ideas. The first idea relates to LLM transparency and reliability. LLM training is carried out over tens of thousands of steps using a corpus. It proceeds sequentially, focusing on one

Corpus search to examine the basis for an answer



dataset at one point in time before moving on to another dataset after that.

Through such a process, the neural network model is modified and changed little by little until the model is finally complete. We take snapshots of that process to track its progress. This allows us to investigate, for example, how expressions of meaning change within the neural network.

Another idea relates to a method that uses searches to identify specific pieces of trained text used by the LLM for a given response. While it is nearly impossible to observe how information is processed in a neural network, it is possible to know the basis for an LLM's answer with such a method. This will also help us identify why LLMs sometimes generate incorrect answers, a phenomenon known as "hallucination."

This sort of approach is also relevant to the copyright problem of training data. For example, there is a very real problem of article data from newspapers and other forms of media being used without permission as AI training data. If it were possible to clearly show the sections referenced by the LLM's output results, interested users could access the original copyrighted works such as article data to obtain more detailed information. This should enable the creation of a mutually beneficial relationship between generative AI operators and copyright holders.

Research and development to improve safety

—There are growing concerns about the safety of generative AI. Are there any contributions the LLMC can make in this regard?

The safety of generative AI is an extremely important topic and a pillar of the LLM transparency and reliability research conducted at the LLMC. In February 2024, the Japan AI Safety Institute (AISI) was launched to examine evaluation methods and standards for AI safety. The activities of the

AISI are wide-ranging, and we will closely cooperate with them on technical aspects through the research results of the LLMC.
One possible example might be to develop and release a dataset that trains LLMs to not respond to toxic inquiries, such as requests for bomb-making instructions, and then update the content of that dataset. However, even if you create such LLMs, in some ways it might become a game of cat-and-mouse where users try to bypass restrictions with clever prompts (command statements to obtain a response from the generative AI). Therefore, we will also consider using more general methods to suppress inappropriate behaviors in LLMs. For example, one approach would be to give models the ability to suppress toxicity. In other words, this would be an attempt to have models examine the corpora they trained on and remove any toxic data that they find from following training runs by improving the models' comprehensive faculties. We plan to attempt various research projects including these ideas, which I think will open up the possibility of effectively controlling LLMs.

LLMC to become a forum for collaboration

——The global development race surrounding generative AI, including LLMs, is quite fierce. A sense of urgency must also be important in research and development.

It is difficult in some respects to predict the direction of future LLM research. Do we increase the amount of training data even more than before, improve the quality of data, or devise new training methods and algorithms? We must quickly decide the direction of LLM research and development while keeping an eye on various possibilities.
Currently, U.S. and Chinese companies are at the center of cutting-edge generative AI model development, and many of them are advancing development in closed research environments. If such research progresses behind closed doors hidden from public scrutiny and an AI that humans are unable to control were to suddenly appear one day, that would not be desirable.
With all these developments, it is becoming difficult to imagine what will happen in the world of generative AI one to two years from now. That is precisely why, instead of just conducting LLM research separately all over the place, it is important to create a structure where everyone can come together to deeply understand the technology while developing LLMs, make forecasts for the future, and propose new technologies. I hope to develop the LLMC so that it can become such a forum.

Comment from the Interviewer

Generative AI is significantly changing the shape of society and business. As Director KUROHASHI stressed, the fact that AI has begun to understand "language" is extremely significant. Generative AI will likely be integrated with robotics technologies, with humanoid robots replacing much of the labor force. In addition, we are also seeing a potential future materialize where AI replaces humans in carrying out scientific research and ushers in a new civilization. Conflicts between AI and humans will also occur during that process. Working out the advantages and disadvantages posed by cutting-edge AI will require above all that we understand generative AI and make predictions about its evolutionary scenarios. The LLMC, which researches the LLMs that drive generative AI development to find ways to control AI in a manner that is desirable to humanity, will face growing societal expectations in the future.

YOSHIKAWA, Kazuki

Senior Staff Writer, Nikkei

Joined Nikkei in 1982. After working in the Industry Department and at the Seoul bureau, as well as in roles including General Manager of Science and Technology and President of Nikkei Science, YOSHIKAWA has covered the field of science and technology since 2015 as a senior staff writer. He was a science journalism fellow at the Massachusetts Institute of Technology from 1997 to 1998.

Role Played by LLMs for Research and Development

The Research and Development Center for Large Language Models (LLMC) of the National Institute of Informatics (NII) is working on building large language models (LLMs) for research and development. Based on the results of the LLM with 13 billion parameters*1 built and released by the LLM Research Group (LLM-jp), it plans to release a new LLM with 172 billion parameters at the level of GPT-3 in FY2024. What are their objectives for this project? We spoke with Scientific Directors SUZUKI, Jun and YOKOTA, Rio of the LLMC, who are working on building the new model.

SUZUKI, Jun and YOKOTA, Rio

A world in which humans and generative AI co-exist. Perhaps it can be said that many people share a sense of apprehension that generative AI will significantly change our lives, work, and society going forward.

However, there are concerns within academia and Japanese industry regarding the building of LLMs, the core technology of generative AI. Professor YOKOTA (hereinafter, "YOKOTA") pointed out these concerns as follows. "Up until now, language models were built through open research and development in academia, which ensured transparency, safety, fairness, and reproducibility. However, if we look at the current situation, a small number of massive IT companies making large investments in development are monopolizing LLMs, and detailed technical information is not being disclosed. You could say that this is an unhealthy situation."

In addition to oligopoly concerns in the field of big science, LLM research and development is directly facing various challenges, such as balancing academic and industrial research and development as well as considering perspectives on economic security between nations. Professor SUZUKI (hereinafter, "SUZUKI") noted, "An oligopolistic situation could mean that opportunities are only given to a handful of researchers affiliated with massive IT companies. I think it is important that technology develops through open science by incorporating the opinions and ideas of various people, and the fact that such development is being obstructed represents a significant loss. This is one issue that must be resolved."

The LLMs for research and development developed by the LLMC were built and released amid

SUZUKI, Jun
Director and Professor,
Center for Language AI Research,
Tohoku University
Visiting Professor,
National Institute of Informatics
Scientific Director,
Research and Development Center
for Large Language Models, NII

YOKOTA, Rio

Professor,
Global Scientific Information and Computing Center,
Tokyo Institute of Technology
Visiting Professor,
National Institute of Informatics
Scientific Director,
Research and Development Center for Large Language Models, NII

these concerns and challenges for the purpose of contributing to LLM research and development in academia and industry. Their goals are to build large models that are open and possess strong Japanese language capabilities and to work on unravelling the fundamentals behind LLMs.

"To utilize and apply LLMs in society, we need to ensure the transparency and reliability of LLMs, and safety considerations will become even more important as models become more sophisticated. I believe that by utilizing this new model and the models we will build going forward, we will be able to further advance our research and contribute to the promotion of LLM research and development." (SUZUKI)

LLM-jp, which is led by the NII and was started in May 2023, built and released "LLM-jp-13B v1.0," Japan's first LLM for research and development, in October 2023, followed by its successor model "LLM-jp-13B v2.0" in April 2024.

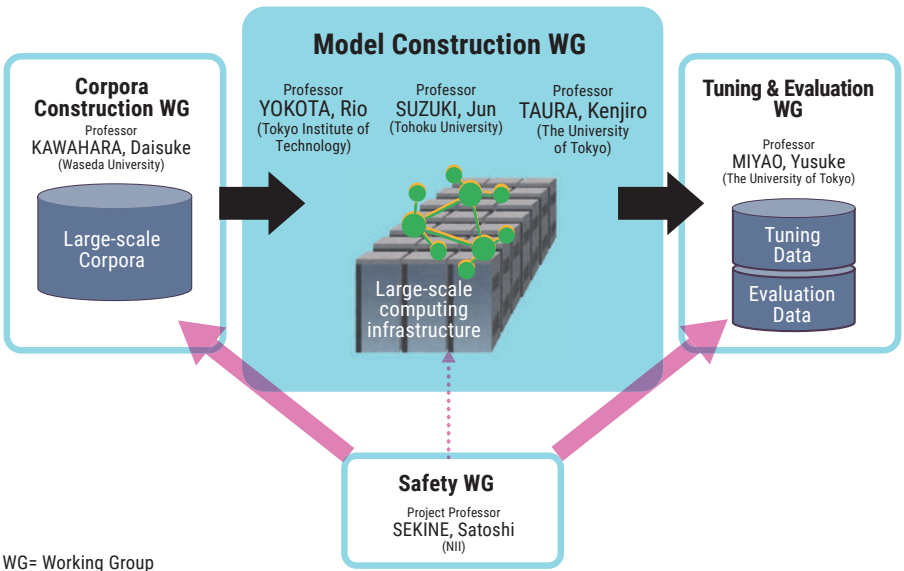
LLM-jp-13B v1.0 is an LLM with 13 billion parameters that uses GPT-2 as its model architecture. The model was built using 12 nodes (equipped with 96 NVIDIA A100 GPUs) from the Platform for the Data-Driven Future (mdx) jointly operated by nine universities and two research institutions, as well as approximately 300 billion tokens*2 of training data.

The training data consist of approximately 145 billion tokens from Japanese portions of mC4 (a language dataset collected from the Internet) and Japanese Wikipedia for Japanese, approximately 145 billion tokens from English portions of the Pile (an open-source multilingual model) and English Wikipedia for English, and approximately 10 billion tokens from The Stack (a dataset of licensed source code) for program codes.

"At the time, no one in Japan had developed an LLM with 13 billion parameters. What is required for training? Are the results of prior research enough to be successful? While groping for answers, we took on the challenge of building an LLM by utilizing technologies that we thought were optimal." (YOKOTA) Professor YOKOTA also emphasized the speed at which the project was carried out, noting that the project started in May 2023 and was released in October of the same year.

Meanwhile, LLM-jp-13B v2.0 was built based on the results of LLM-jp-13B v1.0 as well as on LLaMA (an open-source language model released by Meta, formerly known

LLM building process and a conceptual diagram of the roles of the Model Construction WG



WG= Working Group

as Facebook) and utilizes Japanese portions of Common Crawl, which improved the data quality, training on approximately 260 billion tokens. There are approximately 130 billion tokens for Japanese, approximately 120 billion tokens for English, and approximately 10 billion tokens for program codes. In addition, the model was fine-tuned using eight types of Japanese instruction data*3 as well as instruction data translated from English into Japanese. For the computing resources, 16 mdx nodes (equipped with 128 NVIDIA A100 GPUs) were used.

"When it comes to LLMs, training data are important. We used high-quality training data, ran various verification tests, and improved all sorts of components such as the tokenizer*4 to achieve a higher level of performance." (YOKOTA)

Developing full-disclosure LLMs

To help build LLMs for research and development, the LLMC established working groups such as the Model Construction Working Group, Corpora Construction Working Group, Tuning and Evaluation Working Group, Safety Working Group, and Multimodal Working Group in-house, with top-level researchers participating in the activities of each group.

"Top-level researchers who are leading the fields of natural language processing and high-performance computing in Japan are congregating at the LLMC. This is an LLM for research and development built by the finest personnel you could imagine for building LLMs." (YOKOTA)

For the maintenance of the corpus, text data necessary for training were collected, refined, and converted into a format suitable for training the model. For the building of the model, the neural network architecture and the framework for implementing it were selected, and the base model was built by training on a massive amount of data in next-word prediction tasks. For instruction fine-tuning, examples of good-quality responses were used for post-training to increase the responsiveness of the base model to instructions. Meanwhile, for safety fine-tuning, post-training was used until responses perceived to be desirable by humans were obtained. Furthermore, for the creation of evaluation benchmarks, various benchmarks were created to evaluate the performance of the built model, and an environment capable of using those benchmarks was also established.

"Each component has a mountain of technical issues. However, since massive IT companies do not share their know-how externally, users have no way to deal with problems when they occur. The greatest advantage of LLMs for research and development is that by analyzing and publishing the contents of these components, we can ensure transparency, safety, fairness, and reproducibility for LLM users." (YOKOTA)

"There are very few LLMs in the world that document and publicize their data, checkpoints, training order, and other information. Going forward, the LLMs for research and development developed by the LLMC may become standard models used by researchers around the world who wish to analyze the contents of LLMs." (SUZUKI)

For example, as the number of parameters increases in scale, the frequency of so-called "loss spikes" increases. This is a phenomenon in which the loss value, which is a value that evaluates the extent of training progress, suddenly jumps during model training. In some cases, this phenomenon can have a significant impact on training costs and time by rendering further training impossible and making it easier to just start over, for example.

"At the present time, it is not fully understood why loss spikes occur, and experimental settings must be determined with sufficient care to prevent such events from occurring. Events that did not occur with small models can become an issue when building a large model. However, because we have little experience training large models, we are in a situation where it is difficult to determine the correct findings." (SUZUKI)

"With the LLMs for research and development, we are releasing not only the models themselves but also information such as the datasets used for training, which has been praised by users for providing transparency and reproducibility." (YOKOTA)

The LLMC is also establishing and releasing the details of its parallel computing environment using GPUs.

"For LLM pre-training, you need to use hundreds of GPUs simultaneously to ensure efficient training, combined with parallelization methods such as data parallelism, tensor parallelism, and pipeline parallelism. Because PyTorch, an open-source Python machine learning library which is commonly used in deep learning, does not provide such a framework for efficient distributed parallelization (software

for machine learning) on its own, it is necessary to use a framework capable of efficient distributed parallel training such as Megatron-LM (NVIDIA's framework for distributed training). It is not easy to build an environment on each system for a complex framework such as Megatron-LM to properly run. The LLMC makes the steps used during the building process available to the public, so many LLM developers do not have to go through the same hassles." (YOKOTA)

Although released LLMs for research and development have been fine-tuned in terms of safety, they are positioned as models in the early stages of research and development and are not expected to be used for practical services in their current states. However, as of July 2024, v1.0 and v2.0 have been downloaded 27,866 times and 3,911 times, respectively, which indicates that they are already being widely used for research applications in academia and industry.

Helping raise the standard of Japanese LLM technologies

The LLMC has announced plans to build and release LLM-jp-172B, which will be the largest LLM in Japan at 172 billion parameters, in FY2024.

The model is being built with support from the GENIAC project under the Ministry of Economy, Trade and Industry (METI). It is currently being trained with 2.1 trillion tokens of data, and, according to Professor YOKOTA, the "training is proceeding smoothly."

In FY2023, LLM-jp conducted pre-training experimentally on a 175-billion-parameter GPT-based LLM. However, OpenAI, the developer of GPT, did not release the details of the pre-training data for GPT, which meant restrictions on analysis and the like could have arisen when utilizing the LLM. In contrast, LLM-jp-172B will be pre-trained independently, which will result in the creation of the largest LLM in Japan with a high degree of transparency and reliability as well as strong Japanese capabilities that can make significant contributions to the promotion of LLM research and development.

"I believe the technologies that will drive LLM development in Japan will be created here. If this internationally competitive model is implemented in society, it has the potential to become an important part of the social infrastructure that will support future innovation and economic growth." (YOKOTA)

Going forward, there are plans to develop a Mixture of Experts (MoE)*5 model to reduce training and inference costs while maintaining performance. Further down the road, the development of a one-trillion-parameter LLM to rival GPT-4 is also on the horizon.

"The LLMC is not aiming to create the world's greatest LLM. Instead, the LLMC is strongly intended as a public utility, so to speak, that raises the standard of many LLM-related technologies in Japan. Not only will the models and corpora built by the LLMC as well as evaluation data and other information be released in their entirety, but our experiences, know-how, and the like will also be documented and published. Because there is no information to be concealed, we can tell users anything if asked. In other words, we are creating conditions in which everything we have accumulated through our LLMs for research and development can be used by Japanese academia and industry. Because LLM research and projects can be started based on the results of our LLMs, they can be positioned as the minimum level for Japanese LLM technologies to raise the bar even higher. This is something that can only be achieved by a public research institution like the NII." (SUZUKI)

Without a doubt, the LLMs for research and development at the LLMC will help significantly raise the overall standard of Japanese LLM technologies. (Interviewed in July 2024)

*1 Parameters: In reality, a large amount of actual values. Typically expressed as a matrix or vector. The number of parameters is used when representing the size of a model.

*2 Tokens: Used as an alternative word for "word." A word is basically the smallest component string that is meaningful in a piece of text. However, a token does not have such a grammatical meaning and represents a string within a text that is separated based on specific rules.

*3 Instruction data: Data that combine human instructions and the ideal responses to such instructions.

*4 Tokenizer: A program that is responsible for dividing text into the tokens*2 explained above.

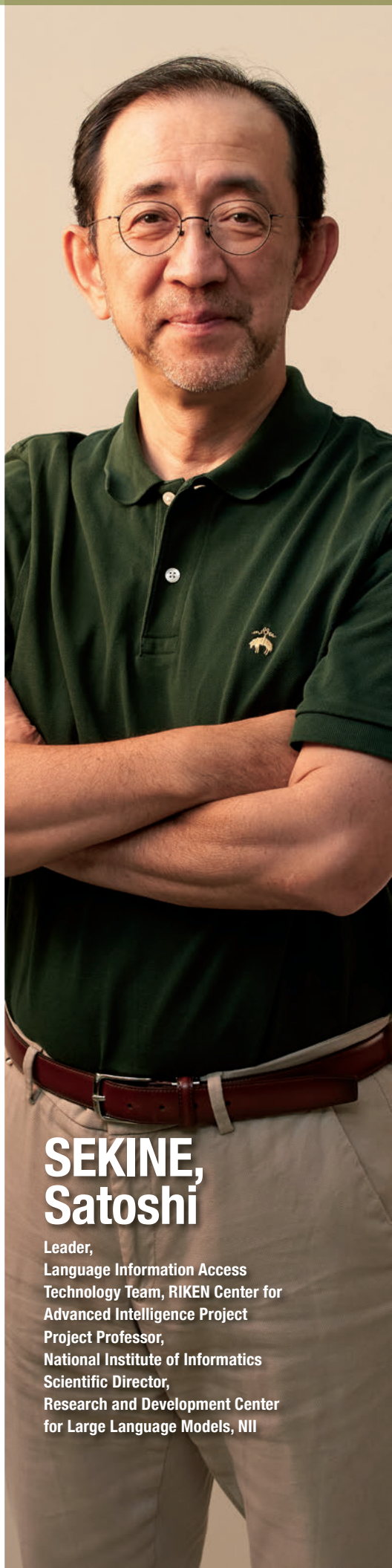
*5 MoE: An abbreviation of Mixture of Experts. MoE is an approach in which some parameters are divided into units called "Experts," and only certain subsets of the Experts are used for training and processing.

Interview and writing: OKAWARA, Katsuyuki/Photos: SUGISAKI, Kyoichi

Transparency, Reliability, and Safety in Building LLMs

As large language models (LLMs) rapidly develop around the world, people have high expectations for the new technology as well as concerns about safety. How does the Research and Development Center for Large Language Models (LLMC) achieve transparency, reliability, and safety in the building of LLMs? We spoke with Scientific Directors SEKINE, Satoshi and MIYAO, Yusuke, who lead working groups at the LLMC, about these concerns.

SEKINE, Satoshi and MIYAO, Yusuke



SEKINE, Satoshi
Leader,
Language Information Access Technology Team, RIKEN Center for Advanced Intelligence Project
Project Professor,
National Institute of Informatics
Scientific Director,
Research and Development Center for Large Language Models, NII

—Please tell us about your roles and activities at the LLMC.

MIYAO: First, allow me to briefly explain how LLMs are built at the National Institute of Informatics (NII). In the first place, you need data for the LLM to train on, and the Corpora Construction Working Group is in charge of preparing such data. The Model Construction Working Group uses the data to pre-train the LLM, then the Tuning and Evaluation Working Group and the Safety Working Group take care of the post-training on top of that.

I am in charge of fine-tuning and evaluation, with "fine-tuning" referring to the adjustment of a pre-trained LLM through post-training .

When a user enters something into an LLM (or chat system), it means that they expect some sort of response from the system. For example, the phrase "What is the weather like today?" demonstrates a desire for weather information. However, the response will differ depending on if the LLM interprets the phrase as a "question" and outputs a sentence that should come next or if it merely sees it as a word string and outputs a word string that should come next. Accordingly, the technique used to help the system recognize the input as a question and generate natural sentences that meet user expectations is what we call "fine-tuning."

Furthermore, "evaluation" is the process of objectively determining whether the output from the fine-tuned system aligns with human expectations and intentions, including accuracy.

MIYAO, Yusuke

Professor,
Graduate School of Information Science and Technology,
The University of Tokyo
Visiting Professor,
National Institute of Informatics
Scientific Director,
Research and Development Center for Large Language Models, NII

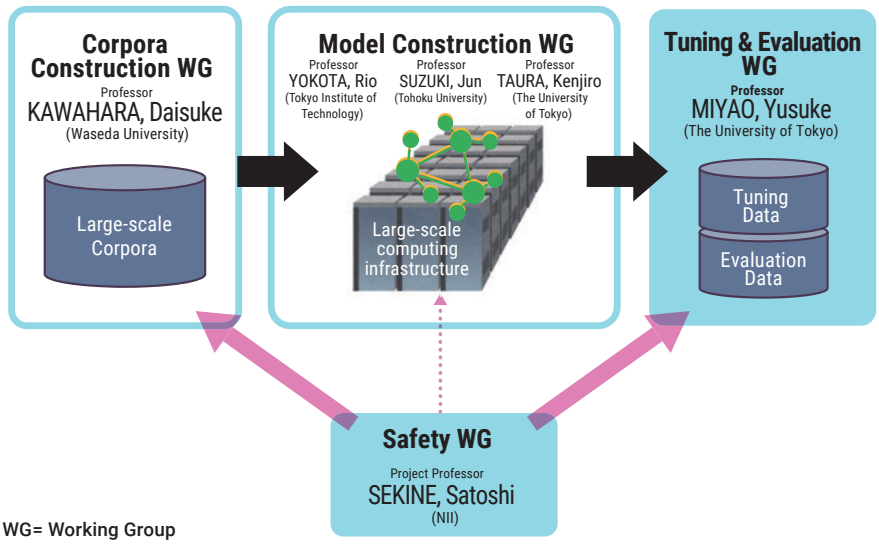
So, how do we fine-tune and evaluate the LLM? Natural language processing, a method that was often used before generative AI, used machine learning. This approach, called supervised machine learning, creates input and output pairs as training data, and when the data are input into a machine learning algorithm, the machine learning model learns that a certain input is associated with a certain output. Fundamentally, LLM fine-tuning uses the same approach. However, the data that are used are called "instruction data." Instruction data are manually created by people called "annotators" who create data pairs that tell the LLM to provide a certain output when given a certain instruction. In fact, some of the instruction data that we are using right now were created by Professor SEKINE's team.

When you take an LLM that was pre-trained by the Model Construction Working Group and fine-tune it with a certain amount of instruction data, it gradually learns that, when given a certain input, it should provide a certain output or provide no response and explain why no response was given, for example.

Evaluation measures how well the LLM handles inputs and outputs, and, largely speaking, there are two evaluation methods. The first method creates input and output data pairs in advance for evaluation. For example, the LLM is told how to respond when asked, "What is the weather like today?" in basically the same manner as instruction data. However, this method provides that input to a trained LLM, compares the output content with the data created by humans in advance, and measures how well they match. With the other method, responses are not created in advance. Instead, various prompts and questions are input into the LLM, and humans evaluate the outputs by providing scores and judging whether the responses were good or bad. Regarding safety in particular, it is quite difficult to prepare in advance, so we are researching ways to automatically evaluate the safety of LLM outputs.

SEKINE: I am in charge of the

LLM building process and a conceptual diagram of the roles of the Tuning & Evaluation WG and Safety WG



WG= Working Group

Safety Working Group at the LLMC. To be honest, it is incredibly difficult to define safety. There are public safety standards, which naturally must be followed, but I think the definition of safety for an LLM, in a form that can be put into words, is ultimately when people out in the world say, "Don't worry—go ahead and use it." However, getting people to think, "This is okay," is an even more difficult problem. Searching for solutions and getting to such a point is the current research topic of our Safety Working Group.

Specifically, we examine the safety created by AI from start to finish. First, the base model is trained on a massive corpus, and we examine what content should be removed. Some argue that removing any and all content describing how to kill people, for example, could prevent the model from imparting such information, making it potentially safer. On the other hand, others argue that removing content on some sort of dangerous, toxic substance would stop the model from outputting related information and fail to convey any dangers , which is probably a bad idea. In addition to such examinations, our first research activity is labeling toxic text, which is related to corpus filtering and the like.

Our second research activity is Supervised Fine-Tuning (SFT), which we are working on together with MIYAO's team, by training the LLM that we are building to not generate toxic statements. This fine-tunes the LLM through question and answer pairs. For example, we are creating instruction data so that the LLM will give responses like "I cannot answer that question because it is dangerous" or "I cannot answer that question because it is illegal" when asked a toxic question. Based on the extensive safety taxonomy of the English open-source "Do-Not-Answer" dataset, consisting of five risk types, 12 harm types, and 61 specific harms, we created our own "AnswerCarefully Dataset," which is a dataset of high-quality questions and answers in Japanese. As of August 2024, 2,000 entries have been released, and we are creating more data.

SFT using instruction data is extremely effective for training, and our aim is to continue creating data, using them for training, and evaluating them to further increase the level of safety.

—Ensuring the transparency and reliability of LLMs at the LLMC

MIYAO: One of the LLMC's research themes is to "ensure the transparency and reliability of LLMs." However, I would like to preface my statement by saying that I do not specialize in the study of transparency and reliability. First of all, reliability means that humans can trust an AI when using it as a tool. I think that other elements of LLM reliability include whether it can produce a response in line with human intentions and expectations, whether the response is accurate, and whether the AI itself can explain the basis of its reasoning for the output response. Moreover, safety in the form of fine-tuning the LLM so that it does not output unethical responses, such as discriminatory or adult content, also leads to trust.

Regarding transparency, the ultimate goal is to clarify the mechanism by which an LLM, which is a huge black box, outputs responses. However, to a certain degree, I think that doing so will be as difficult as demystifying the human brain. Besides that, transparency means the AI system as a whole is clearly visible to users, including, for example, what kind of training data the LLM uses, how it was fine-tuned, and how user-entered data are used. Furthermore, properly dealing with problems when they occur can be said to be another kind of transparency.

SEKINE: Trust is not only about safety. To take it to the extreme, if the model does not respond to any questions to achieve safety, it would not say anything dangerous and thus be safe. Nevertheless, the LLM will not be trusted on that point alone, so various elements come into play such as accuracy or whether the response is sufficiently relevant to the question. By synthesizing those elements, I think that there is a concept of reliability that goes beyond safety. We believe that safety must be taken to that level.

In a technical paper published in March 2022, OpenAI wrote about

InstructGPT, a language model they fine-tuned. About half of the paper was about safety, stating that fine-tuning with human feedback is a promising direction for improving truthfulness, reducing toxic output generation, and aligning language models with human intent (*1). I thought that this should be done in Japanese, and we started creating instruction data at RIKEN, where I work, around the summer of 2023. While not limited to just safety, we created more than 10,000 instruction data entries. What we discovered is that data quality is very important. Instruction data that were automatically generated or translated from English did not increase the effectiveness very much. However, when we asked Japanese language teachers and writers to create questions and answers and used them to train the language model, the quality of the output results instantly increased. We also learned that high-quality instruction data generate beautiful sentences, and we hope to link this finding to quality improvements and reliability.

—Issues and the future of LLMs

SEKINE: When we experimented with a model trained on roughly 200 entries from the aforementioned "AnswerCarefully Dataset," the model started to respond, "I cannot answer that question," even for safe questions that it had previously answered normally. And so, we realized that adding too much safety data can produce side effects. MIYAO's team came up with a way to suppress that behavior through fine-tuning, and we were able to somewhat mitigate effects on non-toxic questions. However, the question of how to get the model to not respond when it should not respond and respond when it should respond remains a theme for research.

MIYAO: What we are doing currently is just telling the model, "Normally you should not say this," based on our assumptions of typical use cases, but that is quite insufficient in some areas.
In reality, depending on the usage situation—for example, when using a model for an article—there may be things that should not be said or things that are conversely okay to say. If a researcher is using an LLM for research in a certain specialized field, they may need to use words that are not typically supposed to be said. Similarly, at the point in time when its practical application is actually decided, the standards for safety, usefulness, and reliability may be re-defined. So, in the future, I think it will probably be not only the people developing LLMs but also those using and incorporating those models into systems who will have to evaluate various aspects of safety, reliability, and transparency at that point in time.

*1 "Training language models to follow instructions with human feedback" https://cdn.openai.com/papers/Training_language_models_to_follow_instructions_with_human_feedback.pdf

Hallucination

Hallucination refers to when generative AIs produce outputs that are not true. At its core, training an LLM is done to generate new strings of words from what was entered into the chat system. That concept and hallucination are actually two sides of the same coin. For example, when you want to write a novel, you need to generate new strings of words, and the creativity, or the part that generates new sentences, at that point in time is essentially the same as hallucination. So, at what point does creativity become hallucination? It is only called hallucination when something is output that differs from what we know of as fact or information . However, from the LLM side of things, they are the same in the sense that the LLM is generating text which is not present in the training data. It is only when an output is compared with information on the societal side, such as societal truths and reality, that we can determine whether it is hallucination or not. When the LLM is viewed in isolation, creativity and hallucination are the same thing. In other words, suppressing hallucination can be said to be the same thing as suppressing creativity. Therefore, it is difficult to prevent hallucination at the core of LLMs, but various techniques for mitigating hallucination are being researched. For example, there are attempts underway to make models conform to the truths of human society outside LLMs, such as Retrieval-Augmented Generation (RAG) , which searches for related documents for LLMs in advance and generates responses based on those documents, and technologies that verify generated texts with different methods, blocking and regenerating responses if there are errors. (MIYAO)

Juris-Informatics and LLMs

The Center for Juris-Informatics at the Joint Support-Center for Data Science Research is conducting research on the "Law by artificial intelligence (AI)" and "Law of AI." For this article, we spoke with Professor SATOH, Ken and Associate Professor NISHIGAI, Yoshiaki, focusing on LLM utilization and what kinds of regulations should govern technology development.

SATOH, Ken and NISHIGAI, Yoshiaki

—Overview of juris-informatics and its relationship to LLMs

SATOH: Juris-informatics is a field of research based on two pillars: 1. the use of AI technology to support legal research ("Law [supported] by AI") and 2. legal control of AI ("Law [control] of AI"). The Center for Juris-Informatics was established at the Joint Support-Center for Data Science Research (ROIS-DS) in November 2023 to support juris-informatics, and I serve as the Director. While I completed law school and passed the bar exam, I am basically an AI researcher, so I am primarily researching the topic of supporting legal research with AI ("Law by AI"). Our research on the topic of AI control is being conducted with the help of Professor NISHIGAI and other professors with a deep knowledge of law.

Even after the emergence of generative AI, our research themes themselves have not changed, and we are also conducting research from the perspective of how to better solve various problems by utilizing LLMs. In other words, our policy is not to support the law with LLMs as a precondition but to use LLMs in the right place for the right tasks.

—Supporting legal research with AI

SATOH: Law by AI, which is about supporting legal research with AI, focuses on the research and development of the "PROLEG" inference system as a tool for supporting the judgments of legal experts primarily in civil trials.

PROLEG is a system that simulates a judgment based on the rules of civil law when a judge has determined the legal facts. Figure 1 (pg. 18) shows the results of such a simulation in a graphical form. PROLEG could be applied, for example, to check whether a complaint contains necessary and sufficient content. Furthermore, the "Presupposed Ultimate Fact Theory" in civil law is a theory that determines, in advance, the default value of each requirement. While it is logically complex, with principles and defenses, defenses and counterdefenses, etc., it has the advantage of being easy to read when written with PROLEG. However, since the specifications do not allow plaintiff–defendant interactions by attorneys, we revised it to be able to create arguments in stages in a table format. In Figure 2, the system shows where to insert a certain legal fact, and the attorney builds an argument by

SATOH, Ken

Director,
Center for Juris-Informatics, Joint Support-Center for Data Science Research
Professor Emeritus,
National Institute of Informatics

NISHIGAI, Yoshiaki

Associate Professor, Graduate School of Social Sciences, Chiba University
Visiting Associate Professor, Center for Juris-Informatics, Joint Support-Center for Data Science Research

entering the appropriate value. However, an attorney who actually used the system pointed out that it was difficult to understand where to enter the value for each element in the table. So, with the emergence of language models, we have been improving the system by automatically generating which element goes with which value so that attorneys do not have to enter values in the table.

When a sentence is entered into PROLEG with natural language (in English, however) (Figure 3), the generative AI extracts only the facts that have legal meaning and then classifies and displays the facts in each element (Figure 4). Moreover, the output of the generative AI is a logical formula as shown in Figure 4, but you can immediately tell if there is an error (for example, the seller and buyer are switched). Because you can intuitively tell if there is an error even without looking at it in detail, I think that language models can be useful in such situations.

Block diagrams (Figure 5) are also automatically generated. While we still only have examples of simple sales contracts, we are working on training the AI on various types of contracts. Those are the major changes to PROLEG since it was introduced in Today No. 97, published in 2022. The Japanese language interface is being built by us using the BERT and ChatGPT language models. We have not yet used the LLM released by the LLMC.

—Legal regulations on LLMs

NISHIGAI: When it comes to the ideal form of regulations concerning AI, the discussion includes not only generative AI and LLMs but also various forms of AI, and I think that it is difficult to define regulations for those AIs along a single axis. There are several different types of legal systems including civil law, criminal law, and administrative law, and each has their own methods of discussion.

First, let us compare and think about such AI regulations when AIs are equipped in autonomous vehicles and regulations in examples using the often-mentioned LLMs. I think that the legal issues which may arise in the case of the social implementation of autonomous vehicles include, for example, the questions of "Who is responsible?" and "What regulations are needed to prevent accidents?" when an autonomous vehicle equipped with AI causes an accident. Since many people are able to recognize the potential for life-threatening accidents, the need for regulation itself will be easily understood. Next, I think that awareness of that problem will be shared among the general public followed by meticulous discussion of careful guidelines formulated in advance and detailed theories of negligence, etc.

At the same time, what will be considered as problematic in regard to LLMs will be regulations on expressions and copyright law. Infringement of another person's copyright can be severely punished. However, unlike autonomous driving, such an act does not immediately involve human life. Viewed from another perspective, it is possible to say that copyright infringement is an extension of daily activities. I think that the question of how to engage in a rigorous discussion to ensure regulatory transparency in a domain where everyone is an interested party, so to speak, will become an issue.

In that sense, there would be some distance between the two sets of regulations, and they must be considered on different levels.

Next, let us consider this from the perspective of the concerns that regulations may hinder the evolution of LLMs. Many existing

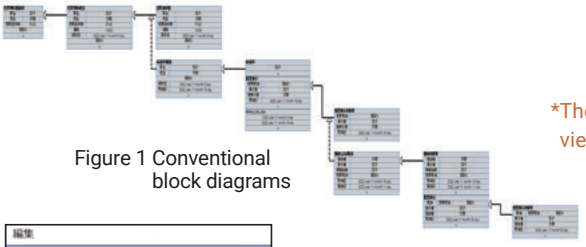


Figure 1 Conventional block diagrams

*The diagrams can be enlarged for viewing in the online version.

Figure 2 Example of a window for the input of legal facts

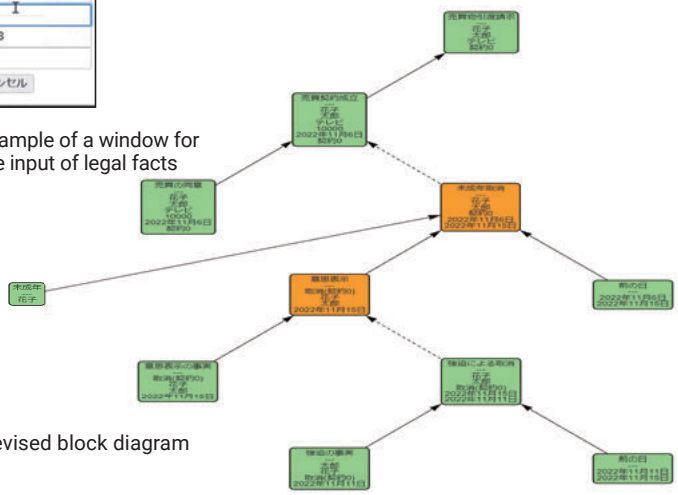


Figure 5 Revised block diagram

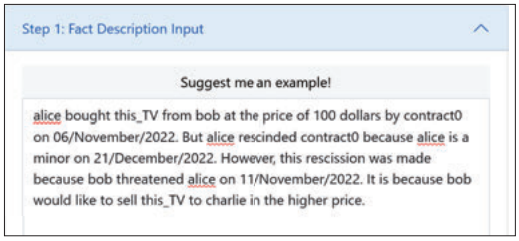


Figure 3 Input of case facts using natural language

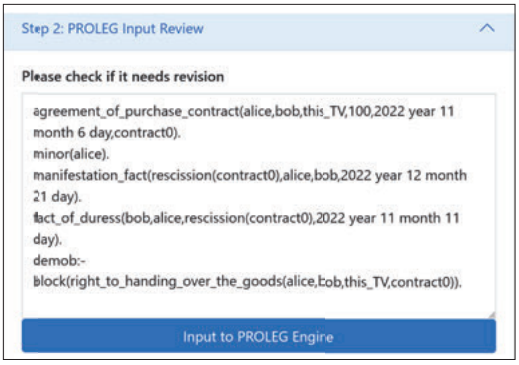


Figure 4 Automatic generation of logical formulas representing facts from the text entered in Step 1

regulations were not designed for generative AI, so if generative AI suddenly appears and users discover it can do this and that through various types of experimentation, they may in fact run into certain regulations before they know it. If it does come about that those regulations are unfortunately applied, it could lead to a discussion about whether it would be better not to create or use such AI due to the huge legal risks. Therefore, it is possible that there will be a discussion about how existing regulations should be applied to generative AI.

On that point, what is extremely troubling is that existing penalties may be written in such a way that they can directly include new technologies when they emerge. For example, even if the concept of obscenity changed from the past, regulations that prohibit obscene materials still remain. Therefore, if a person uses generative AI, creates obscene images, and distributes them to many people thinking that they are OK because they were output by generative AI, it is possible that they could be found guilty of the transmission and distribution of obscene data (Distribution of Obscene Objects (second half of Penal Code Article 175, paragraph (1))). Moreover, not only will the generative AI user who created and distributed such illegal images be arrested but the question may even be raised as to whether the person who created the program (generative AI) that enabled one to create such images, even if that software is not intended solely for the creation of obscene data, could be punished for accessoryship to the crime of transmitting and distributing obscene electromagnetic records, assuming that they created software that makes it easier to commit a crime. This situation evokes the Winny case (Supreme Court decision of December 19, 2011, Keishu Vol. 65, No. 9, p. 1,380). Even if it is decided that regulation to that degree is excessive, concerns remain that there may be many problems such as these, or in other words, cases which could potentially be regulated in light of the current situation in which various existing regulations are spreading out like the mesh of a net. It is important not to get caught up in all of these regulations, and in some cases, the legal verification costs may be quite significant. I think that we need to explore a legal approach whereby such existing regulations will not hinder the evolution of LLMs.

SATOH: "Controlling AI" does not mean tightening the restrictions and forcing people to stop using it, because it is also about industrial development. To put it in an extreme way, if we ban all generative AI, then generative AI will no longer create any problems. However, there would be no progress in new technologies, society, or industry, so I think that regulatory balance is extremely important.

—Expression and regulation

SATOH: One problem that is discussed in the context of LLMs is copyright. Copyright law protects works of art and other forms of "expression" and says that similar expressions should not be created by relying on the original expression. Conversely, while it protects "expression," it does not protect "style." I think that the important characteristic of an author is not the written expression itself but instead their "style." LLMs are able to learn that "style," so I think that the

fact that aspects of the author's personality are, for a lack of a better word, "stolen" by the LLM may be a problem. Of course, it can be said that it is realistically impossible to certify something as vague as a person's "style."

Moreover, in terms of expression in an LLM, there is also the problem of hate speech. Research is being conducted to prevent LLMs from generating hate speech output, but from the perspective of freedom of expression, this seems excessive. However, this is an extremely difficult problem, so I cannot give a clear answer at the present time.

NISHIGAI: Speech that criticizes someone but does not go as far as hate speech may be considered to be defamation if it diminishes the person's social reputation.

Let us look at this along separate axes. The first is whether the AI can check the content of the expression. If this is possible, then the AI could generate forms of expression that are not defamatory. However, this is difficult to accomplish. Even if the expression is defamatory, or in other words, alleges facts that lower the social reputation of another person (see Article 230, paragraph (1) of the Penal Code), if the factual content of the expression relates to a fact of public interest, the act of expression is primarily for the benefit of the public, and important aspects of said facts are proven to be true (legal principle of truthfulness, see Article 230-2, paragraph (1) of the Penal Code), or if there are reasonable grounds for believing that the important part of said facts is true (legal principle of reasonableness), no civil or criminal liability will be incurred in relation to the defamation. Furthermore, even if an opinion or criticism based on specific facts diminishes the social reputation of another person, if the facts which form the basis of the opinion or criticism correspond to the aforementioned principles of truthfulness and reasonableness, then it is thought that there will be no liability for defamation (legal principle of fair comment) unless it constitutes personal abuse or otherwise goes beyond the bounds of an opinion or criticism. It is understood that this legal principle was introduced to make the guarantee of freedom of expression effective. In short, even if that expression on its own seems quite strong and diminishes the social reputation of the person who is mentioned, it should be permitted in some cases as an exercise of freedom of expression. In that case, if we were to prevent an LLM from outputting expressions that diminish the social reputation of a specific person, the scope of acts of expression permitted as an exercise of freedom of expression would narrow and only self-restricted output would be generated. Even if that was too excessive, it would be good to be able to output responses that fulfill the aforementioned principles of truthfulness, reasonableness, and fair comment. However, it would be difficult to have AI make that determination. Moreover, the range of permissible acts as freedom of expression would be broader than that of acts to which the aforementioned legal principles may be applied. In that case, if we just design AI with the assumption that only such output in the case of the aforementioned legal principles is okay, there is a possibility that it could result in AI that self-restricts its output to the point that it infringes upon freedom of expression, which could be a major problem. Therefore, I think that we need to proceed carefully when it comes to discussions such as "AI that generates these kinds of output should not be created."

Next, assuming that an AI that is self-restrictive to a certain degree were to be hypothetically created, let us consider the subsequent transition after it becomes widely used. One scenario might be that the emergence of generative AI that can create endless linguistic expressions leads to the conclusion that, from a legal perspective, certain expressions should not be output. In that case, active discussion about regulation may be promoted and this or that expression may be identified as inappropriate. From that starting point, the scope of expressions that should not be output may gradually spread. More likely, this may not happen. I think that we should consider several scenarios and engage in an interdisciplinary study of better future worlds. Perhaps a concept of "freedom of expression in the LLM era" is required.

—Use of LLMs and the future of juris-informatics

SATOH: On the question of how to use LLMs to implement Law by AI, the biggest problem in Japan is that there are no training data, because the judiciary has not yet been digitized. As of July 2024, the Ministry of Justice's investigation commission compiled a report regarding the creation of a database of civil trial decisions with the aim of putting it into operation from FY2026. However, we are still in the initial stages, so it will be some time before it is available.

Interview: KURASAWA, Haruo/Composition: KISHIMOTO, Harue/
Photos: SUGISAKI, Kyoichi

Collaborating on the Future of Autonomous Vehicle Safety

Autonomous vehicles should be able to solve social issues related to human mobility and logistics. However, there are many issues which must be overcome in development, manufacturing, and social implementation. If industry and academia cooperate and integrate the respective knowledge, experience, and technologies, then perhaps future possibilities may expand even further. Based on that idea, engineers and researchers from Japan's leading manufacturers together with researchers from the National Institute of Informatics (NII) discussed the results of their respective joint research at the NII Open House 2024 Industry-Government-Academia Collaboration Seminar.

Guests (in speaking order)

UEDA, Naoki
Mitsubishi Electric Corporation

TAKAO, Kenji
Mitsubishi Heavy Industries, Ltd.

YOSHIOKA, Toru
Mazda Motor Corporation

YANAGISAWA, Nayuta
Toyota Motor Corporation

Hosts

HASUO, Ichiro
National Institute of Informatics

ISHIKAWA, Fuyuki
National Institute of Informatics

Research and development connects with people in the driver's seat and on the manufacturing floor

HASUO: Thank you for joining us today despite your busy schedules. My original specialization is in mathematics, and I go back and forth between mathematics and informatics. Currently, I am conducting research at the Research Center for Mathematical Trust in Software and Systems on the use of mathematics to increase the trustworthiness of software and information systems to encourage the social acceptance of new technologies. First, would each of you working in industry please introduce yourselves?

UEDA: I am primarily involved in the development of system monitoring technologies such as anomaly detection and causality analysis for embedded systems. Currently, I am working with Professor HASUO of the NII on research to formalize scenarios in which automobiles can become dangerous using temporal logic.

TAKAO: Instead of working on autonomous driving for automobiles, I have been conducting research to increase the reliability of important infrastructure. Together with Professor HASUO of the NII, we are developing high-reliability control systems based on formal methods.

YOSHIOKA: My original specialization was in vehicle motion control. Over the past few years, I have also been involved in the development of autonomous driving technologies. I am conducting joint research with

Professor HASUO with the goal of developing an in-vehicle AI system that is like a reliable copilot, and we have developed it to the point where it can detect driver anomalies to reduce speed and stop the vehicle.

YANAGISAWA: I specialize in formal verification and data processing, and I am conducting joint research with Professor HASUO on formal verification. Our primary mission is to contribute to mass production development, and we are also developing technologies for the future in parallel.

ISHIKAWA: I have been researching support methods for the development of software. In particular, I am conducting research that examines accuracy and validity of formal methods and tests, and analyzes and corrects errors. Without being limited to so-called mathematical techniques, I am also working on exploratory and empirical techniques.

"What humans cannot do" leads to the pursuit of "what only humans can do"

HASUO: The NII values grounded software research, but at a manufacturing site, there is a different kind of difficulty from software which is composed of mathematics from the beginning. For example, if you have a single gear, what kind of behavior does it exhibit in what kind of situation? Just creating a mathematical model of a single gear requires a significant cost. Furthermore, in a massive system such as an automobile that combines many mechanical components, past approaches for proving software safety are no longer valid. To ensure safety, we must think about what kind of approach to take and what kind of method to create, including having people who were not previously involved becoming stakeholders. This is a challenge that we cannot take on alone.

UEDA: High-reliability systems such as artificial satellites and automobiles that must continue to operate normally cannot have software bugs, but one of the causes of bugs is that different people have different interpretations of the specifications. This is because humans write specifications in natural language and other humans interpret them. However, there cannot be any variations in interpretation in the safety evaluation standards established by international organizations. Therefore, I have been working with the NII to tackle the challenge of strictly defining automobile safety requirements in formal language for traffic disturbance scenarios established by international standards. As a result of our work, we achieved results that "mathematically express the ever-changing conditions" as well as "write, verify, and modify the specifications with one tool, understand the specifications, and understand the verification results using animations and graphs."

TAKAO: Mitsubishi Heavy Industries creates not only power generation plants, new traffic systems, and other large infrastructure equipment but also medium-lot industrial machinery, such as forklifts. What they share in common is a requirement of high reliability. So that is why I have been researching and developing automatic verification technologies and design technologies together with Professor HASUO. By defining the requirement specifications with temporal logic, we are able to handle extremely complex and time-varying constraint conditions and also successfully verified them through falsification and heuristic search methods. Falsification is a technique that searches for conditions that do not satisfy the requirement specifications. If we avoid those conditions, then we can find the conditions that satisfy the requirement specifications. In addition, if we automatically generate combinations of parameters that best satisfy specifications that are changing in a complex way, we can avoid creating bugs or setting inappropriate parameters during the basic design and functional design stages. For example, in a complex system such as a gas turbine, there are many requirements that must be satisfied and many that would be desirable to satisfy. If each requirement differs "from this point in time to this point," it would take a human several weeks just to adjust the design parameters. However, with this method, we can run a simulation in a few hours.

Connections between diverse people and organizations lead to technologies that make people happy

YOSHIOKA: We are developing our co-pilot system with the goal of realizing the "MAZDA CO-PILOT CONCEPT." In 2022, we commercialized a driver anomaly response system that automatically slows down and stops the vehicle if the driver suddenly becomes ill, etc. and is unable to drive. The system has been well-received as an unprecedented safety system.

In recent years, the number of fatalities and serious injuries due to automobile accidents has continued to decrease throughout Japan. Looking just at Mazda vehicles, the decrease is greater than the nationwide total. This is the result of efforts to combine basic safety technologies such as improved visibility and collision safety with advanced safety technologies that sense the area around the vehicle to detect careless driver errors and provide support.

Through this initiative, we noticed that there are a certain number of serious accidents due to sudden changes in the driver's physical condition and they are increasing in recent years with most cases occurring on ordinary roads.

To further improve safety, advanced support technologies that function on ordinary roads are needed. In addition, there are also benefits to

driving itself.
According to some data, elderly persons who continue to drive reduce their risk of dementia by about 40%. While the risk of accidents may decrease when elderly persons stop driving, risks may increase in terms of a healthy life expectancy. Therefore, we proposed the concept of a reliable "CO-PILOT" based on the importance of "people driving." It is a concept that adds safety support based on the human condition to conventional advanced safety technologies by silently watching over the driver during normal times and switching over the driving tasks to maintain safety conditions when a person becomes unable to drive for some reason.
On ordinary roads in particular, there are a wide variety of scenarios and use cases. With previous methods, it was thought to be impossible to cover all cases and verify safety. So, I wondered if some sort of logical technology could be applied, and began the joint research after consulting with Professor HASUO.
Currently, we have been able to verify that it is possible to build a verification method that does not rely only on tests by using formal verification.

YANAGISAWA: I joined the company mid-career in 2023, and I am currently engaged in research and development involving automobiles.
At InfoTech Labs, where I work, there is an extremely high emphasis placed on contributing to the field (mass production development) through research and development. However, this is quite difficult to do. The ideal cycle would be to hear about issues in the field, conduct joint research with universities, etc. based on the issues, and feed the results back to the field. However, when I first joined the company, it did not go well at all. Because I had no connections inside the company, I essentially struggled to reach people in the field. When I actually went out to sell technologies to the field, I was told, "Our department won't use it." I wanted somehow to get closer to the ideal cycle, but first there was nothing I could do without understanding the issues in the field. So, I planned a formal verification study group that included people working in the field and the research and development department. Determined to put the concept of Specification-Driven Engineering (SDE) proposed by Professor HASUO and me into practice, I named it the "SDE Study Group."
Of course, I could not form a study group for my own convenience (to learn more about issues in the field), so I needed to make the content useful to the people in the field. The study group content primarily consists of three types. The first type involves the sharing of success stories in the field with the goal of standardizing and upgrading the level of knowledge throughout the company. The second type introduces technologies that have never been adopted in the field by the research and development department and university professors. The third type brings in unresolved issues and problems from the field for everyone to discuss.
As a result of these efforts, communication paths with the field and the research and development department were established and strengthened, consultations from the field became more active, and joint projects are about to be launched.

ISHIKAWA: I currently receive data from the industrial world and often conduct research on testing. Collaboration between industry and academia is not just about manufacturing and autonomous driving but also involves every system and every aspect. Development is always accompanied by testing. What and how should we test to eliminate overlooked errors? What test goals should we clear to be able to minimize anxiety? That is the first thing that people in industry will discuss. However, in the case of autonomous driving, there are a vast number of states for the "position of the other cars." Are there any tests that can be said to be sufficient under those conditions? Can they be found? How do we look for them?
I have been working on such issues as a researcher.
Currently, if we have a goal such as "find the conditions under which collisions occur," we are able to use a test agent and a simulator to search for the conditions and assign a score such as "75 points,"

UEDA, Naoki
Researcher,
Trusted Systems
Technology Group,
Information Network Systems Engineering
Department , Information Technology R&D
Center, Mitsubishi Electric Corporation

TAKAO, Kenji
Ph.D., Research Manager,
Machine Systems Laboratory,
Intelligent Machine Systems Research
Department, Research & Innovation Center,
Mitsubishi Heavy Industries, Ltd.

YOSHIOKA, Toru
Expert Engineer,
Integrated Control System Development
Division, Mazda Motor Corporation

etc. for the risk of collision. This method is inseparably linked to the work of Professor HASUO by being able to automatically search for the conditions if a simulator is available without placing mathematics and proofs at the forefront. "Please come at five o'clock" and "If you come at five o'clock, you pass, but if you come five minutes past five o'clock, you lose five points" are two sides of the same coin that mean the same thing. Nevertheless, we are conducting joint research with people in the field to "create smart tests that can be used without thinking about mathematics." In the Mazda case study, we ask the smart test agent, "Look for cases under which collisions occur by searching for cases that increase this test score." If we ask it to perform a more difficult case search, it will discover extreme cases such as "ignoring a red light and colliding." Moreover, even if it does not lead to a collision, sudden acceleration by a car that was supposed to be driving safely is not an intended behavior. We generate tests according to the purpose including the understanding of such behavior. As the capabilities of autonomous driving and AI systems continue to increase, we have continued to expand our research over the past five years in cooperation with companies participating in joint research and Professor HASUO.

Sharing issues and overlaying knowledge to create synergy

UEDA: Our need to "apply formal language to product development" harmonized with Professor HASUO's desire to "contribute specialized knowledge to industry" in the form of joint research, and I believe that we were able to achieve mutually valuable results. Going forward, our goal is to further improve the description method to make it easy to handle even by those who are not experts in formal language. In addition, we are also thinking about applying it to systems other than automobiles.

Nevertheless, the talks that everyone gave today were very original and meaningful. For example, every company probably feels that it is necessary to create a path whereby the R&D department takes up issues from industry and connects them to academia. However, when I actually listen to success stories, I learn quite a lot and think, "Maybe we can do this at our company as well. Let's give it a try."

YANAGISAWA: Thank you. While presiding over the study group, unexpected developments occurred such as the starting of exchanges between previously unrelated departments.

YOSHIOKA: We also look forward to continued joint research and new developments in the future. The development of autonomous driving technology will achieve the ultimate driver model. We believe that by enhancing our knowledge of logical approaches, there is the potential for applications that can help mathematically model human driving problems and enable more accurate information support, decision support, and driving support.

TAKAO: We brought the specific issue of gas turbines into our joint research and heard that Professor HASUO was extremely motivated by it. In addition, a student at the time who was good at theoretical studies was extremely interested and contributed to the research from the perspective of algorithm theory. I think that this is a successful example of collaboration between industry and academia.

ISHIKAWA: Verbalizing what you want to do is the first step to achieving it. When you verbalize what you want to do with mathematics, diagrams, and scores, it creates the possibility of concretely realizing that goal as a technology and then it is actually realized. Today was a meaningful day, because I was able to hear from everyone conducting joint research through collaboration between industry and academia and learn about the expanding possibilities of describing ideas with mathematics. If any of you working in industry has an issue, please feel free to contact us.

HASUO: Once again, I would like to express my gratitude to all of the researchers and students in charge of research for their hard work.

The people working at companies who bring us their problems and issues are the source of our motivation. I am truly grateful. Thank you for your continued support.

YANAGISAWA, Nayuta

Senior Researcher,
Social System PF Development Division,
Toyota Motor Corporation

HASUO, Ichiro

Professor,
Information Systems Architecture Science Research Division, National Institute of Informatics

ISHIKAWA, Fuyuki

Associate Professor,
Information Systems Architecture Science Research Division, National Institute of Informatics

Composition: MIWA, Yoshiko

Wireless Communications for Realizing Smart Factories

Japan-France bilateral collaboration among four organizations on a “low-power and lightweight edge AI technology”

“A diversity of people leads to a diversity of perspectives, which in turn leads to innovation.” So, what does organizational and national diversity create? The international joint research project “Light-Swift” between Japan and France was selected within the framework of the “Strategic International Collaborative Research Program (SICORP)” promoted by the Japan Science and Technology Agency (JST), and started in December 2023. We interviewed Professor KANEKO, Megumi of the National Institute of Informatics (NII), who serves as the Japanese-side Project Leader, and Senior Research Engineer KAWAMURA, Kenichi of Nippon Telegraph and Telephone Corporation (NTT), the industrial research partner on the Japanese side, about the impact this research will have on society in the near future.

KAWAMURA, Kenichi
Senior Research Engineer and Group Leader, Wireless Access Systems Project, NTT Access Network Service Systems Laboratories, Nippon Telegraph and Telephone Corporation

KANEKO, Megumi
ANR/JST SICORP Project Leader (Japanese side), Professor, Information Systems Architecture Science Research Division, National Institute of Informatics

■ Overview of SICORP ■

SICORP is a JST initiative to promote international joint research, and the project focusing on the recently selected research topic of "Lightweight Edge Artificial Intelligence for Sensing and Wireless Communications in Connected Factories (LIGHT-SWIFT)" will be jointly carried out by Japan and France through a team composed of researchers from academia and industry in each country. On the Japanese side, Professor KANEKO of the NII serves as the Project Leader, with NTT as the industrial partner. On the French side, the National Centre for Scientific Research (CNRS), IRISA/University of Rennes, and Wavely will participate as the academic and industrial partners, respectively.

Making IoT devices smarter with lightweight edge AI and improving the reliability of wireless communications

—What kinds of social issues will this project solve?

M. KANEKO: This project is aiming to realize energy-efficient wireless communications and sensing technologies dedicated to smart factories.

The major goal of future smart factories is to enable the automation and acceleration of difficult tasks such as monitoring, remote control, or fault detection, by means of IoT and AI technologies. For instance, it would become possible to monitor extremely dangerous or inaccessible locations in the factory and to assess critical situations in real time, such that optimal decisions and actions can be immediately taken.

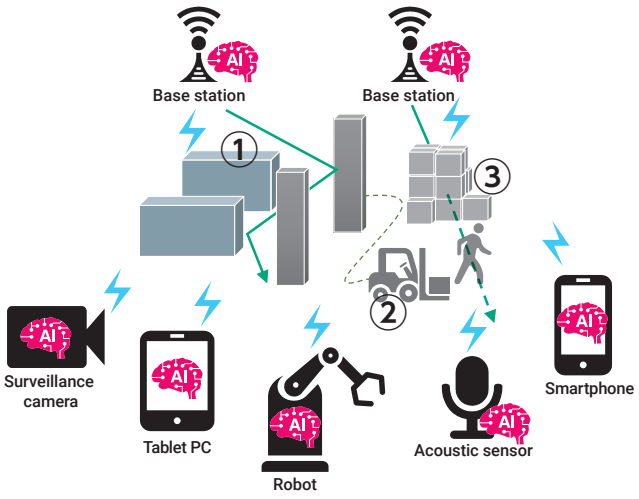
There are high expectations worldwide that smart factories will increase productivity, reduce operational costs, save energy, and increase safety by preventing accidents involving factory workers. However, cutting-edge wireless networking technologies are essential to realize such ambitious goals. Indeed, various IIoT (Industrial Internet of Things) devices such as robots, acoustic sensors, and other sensing devices are scattered throughout the factory. All of those devices transmit and receive data through wireless links, making advanced wireless technologies vital in order to support them. However, factories are indoor spaces surrounded by walls and other obstacles, while their wireless environment is subject to dynamic and complex changes due to the movements of large machinery and equipment, etc. All this makes factories a very challenging place for providing high quality and reliable wireless communications.

In addition, while 5G and subsequent standards use higher frequency bands such as millimeter waves due to the shortage of wireless frequencies, they are very vulnerable to obstacles. This increases the risk of interruption and errors over wireless transmissions, which may hinder the delivery of important data. On top of that, as the amount of IoT devices increases and more and more devices communicate wirelessly at the same time, the interference level rises, and the risk of packet collisions and losses also increases. All these aspects render wireless communications particularly tricky in factory environments, so this is one of the challenges to be resolved in this project.

K. KAWAMURA: To realize a "smart factory" in which many IoT devices safely move around within the factory, communication must be made wireless. In other words, we must achieve the same level of high reliability of current wired connections, even in wireless communications.

To achieve ultra-high reliability, we must continuously follow and adapt to the dynamic changes of the factory environment, in order to maintain the required communication quality. Although the use of high frequencies is important, in order to use them effectively despite their susceptibility to obstacles, it becomes necessary to control the communication quality more precisely.

Therefore, we equip the base stations and end-devices with AI functionalities that can run with low power (power-saving and lightweight edge AI). This edge AI enables us to assess the local mobile environment of each device, to adjust its wireless communication parameters, and to select the best base station to connect to. If there are



In factories and other locations, the communication quality becomes unstable due to complex changes in the wireless environment, locally around each AP, IoT device, and other equipment. (1) Wireless environment changes due to layout changes. (2) Changes that occur when the terminals move. (3) Obstruction due to moving objects. Using end-devices equipped with edge AI enables them to maintain high-quality wireless communications.

multiple base stations that can be used at the same time, it selects the adequate links and transmission methods according to the evolving situation and continuously selects the "optimal solution" for transmission, so that it can achieve highly reliable wireless communications in a global manner. Small devices have limited battery power, so it is important to balance the optimization between the device's own power consumption and the communication/sensing performance and to choose which one to prioritize according to the situation.

M. KANEKO: It is the task of us researchers to analyze and design the methodology to create the best "balance," namely a trade-off between wireless communication quality and power consumption.

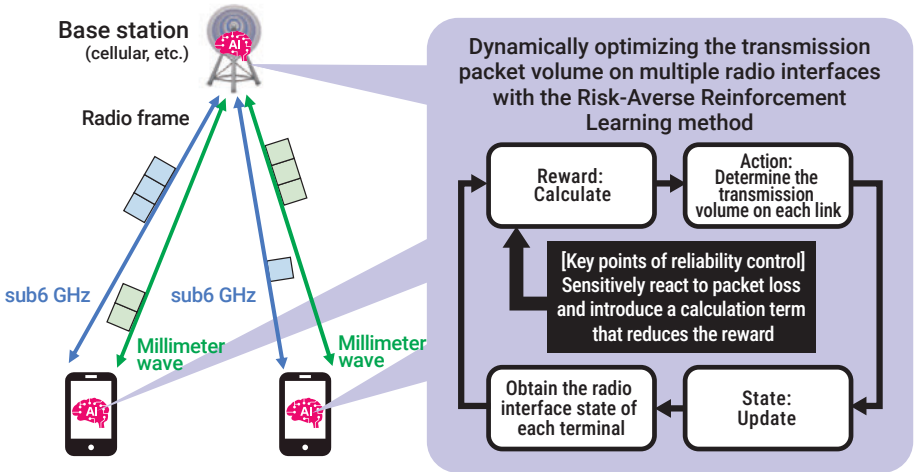
For example, when it comes to IoT devices, if the wireless environment suddenly changes, a significant amount of power is required to process that information. However, if the environment remains stable, we may limit the devices to a normal monitoring state, thereby reducing both their power consumption and processing functions. We design algorithms and protocols providing such a trade-off, by means of AI that exploits risk-averse reinforcement learning.

For example, in conventional IoT networks with automatic control of air conditioners, refrigerators, and other electronic devices, sensing data are only transferred, and intelligent processing is carried out by edge nodes on the network side, or by cloud equipment in the central part of the network. In contrast, this project equips IoT devices that are the terminal entities of the network (edge) with a small "lightweight edge AI" brain. This enables them to effectively utilize their limited processing and battery power, to assess their surrounding wireless environment and acoustics, and to process their sensing data to make their own decisions, using centralized control only when deemed necessary. Hence, making each individual device just slightly intelligent enables the entire system to benefit from substantial performance improvements.

——What is the story behind the incorporation of risk-averse reinforcement learning?

M. KANEKO: Previous technologies were insufficient for implementing optimized access control in a complex wireless environment where many IoT devices are sharing the same wireless resources. In recent years, reinforcement learning approaches for improving wireless communications have sparked many research interests both in Japan and overseas. In the first place, the fields of AI and machine learning are not within my field of research, which is wireless communications. However, I got interested in the tool of risk-averse reinforcement learning, so I designed wireless access methods by incorporating multi-agents (multiple terminals) and jointly developed them with NTT as the "Multi-radio Access Technology Using Risk-averse Reinforcement Learning."

In reinforcement learning, the agent (terminal that implements AI) observes the state of the surrounding environment and learns to select actions with the goal of maximizing the accumulation of "rewards" (cumulative reward function), by making use of signals fed back to the agent by the environment. Part of that research focused on the design of risk-averse rewards and their implementation method. By considering the errors that result due to poor wireless transmissions and failures to deliver desired information as "risk," the observation state, actions, and rewards were set so as to avoid errors that are specific to wireless communications. In addition, our core technical idea is to make each individual IoT device take such risk-avoiding actions and to enable many agents to cooperate at minimal cost, so that they select more risk-avoiding actions for enhancing wireless communications.



Overview of the multi-radio access control technology using risk-averse reinforcement learning

—What is the future of communications and devices in factories?

K. KAWAMURA: Various wireless communication standards are used in IoT devices. For example, while Wi-Fi has the advantage of being easy for anyone to use, there is a lot of interference, and the stability of the communication quality may be low in some cases. While cellular communications suffer little from interference, their base stations may be unable to provide sufficient radio coverage inside the factory, so there may be no cellular signal inside the factory at all. In addition, since the cellular frequency band is also used by general users in the surrounding area, the communication performance cannot be guaranteed. Local 5G enables the installation of your own base station, and the quality is good. However, its high cost is a disadvantage compared to Wi-Fi. The current situation in which diverse forms of communication using various radio waves and communication technologies co-exist is expected to continue going forward.

M. KANEKO: Under 6G, energy utilization efficiency will become even more important, as it constitutes an indicator of the balance between communication performance and its required power consumption and cost, and is an important target of this research. IoT devices in particular have limited batteries, so the question of how to guarantee a high level of quality in advanced wireless communications and sensing inside factories with minimum power usage is also related to energy efficiency.

Combining each partner's strengths in joint research to reach new heights

—What kind of collaboration will take place in the SICORP research project?

M. KANEKO: As the academic partner on the Japanese side, the NII will conceive and evaluate the core algorithms, while NTT will participate as our industrial partner. We'll elaborate on the power-saving, lightweight edge AI technology, examine applicable architectures, conduct the test evaluation, etc. On the French side, CNRS IRISA/University of Rennes, as our academic partner, will be in charge of developing the power-saving AI compression technology and designing the edge AI software and hardware. Wavely, an acoustic equipment company, will develop the AI technology for IIoT acoustic sensing, environmental variation extraction, and anomaly detection.

—So each of the four organizations, with their own features and expertise, will innovate together through this project.

K. KAWAMURA: I truly believe that we can realize SICORP's policy of "international joint research based on equal partnership," solve common international issues that cannot be solved by one country, and help strengthen Japan's scientific and technological capabilities.

M. KANEKO: I conduct research in wireless systems including interference management, performance analysis, and protocol design, which have a significant impact on the overall communication quality. In particular, I had mainly focused on research using mathematical models, so it is a more basic type of research within my field. I have been engaged in joint research with NTT since 2018, and it has become possible to conduct research and development for applications toward various use cases. As a result, we have made a fruitful collaboration through outcomes that include not only top international conferences and journal publications, but also 12 patent applications, among which four have been registered.

K. KAWAMURA: Since NTT is a company, their principal objective is to consider how the technology will be implemented, practically applied, and tied to services. The kind of basic research that Professor KANEKO conducts on component technologies is something that is beyond our reach. We have continued to conduct joint research focusing on two topics: "Machine Learning-Based Methods for Distributed Wireless Environment Prediction" from 2018 to 2021 and "Optimized Multi-Connectivity and Resource Utilization for High Reliability Wireless Communications" from 2021 to 2024. Our current international project builds upon those past projects.

We also hope to further enhance wireless communications technologies through our research.

M. KANEKO: Currently, wireless communication and sensing systems equipped with a low-power edge AI that can run even in simple IoT devices are yet to be deployed in the world, but I look forward to seeing them several years from now not only in factories, but also in other fields of society. It is also my aim to make that development a reality, in cooperation with NTT and our two French partner organizations within SICORP Light-Swift.

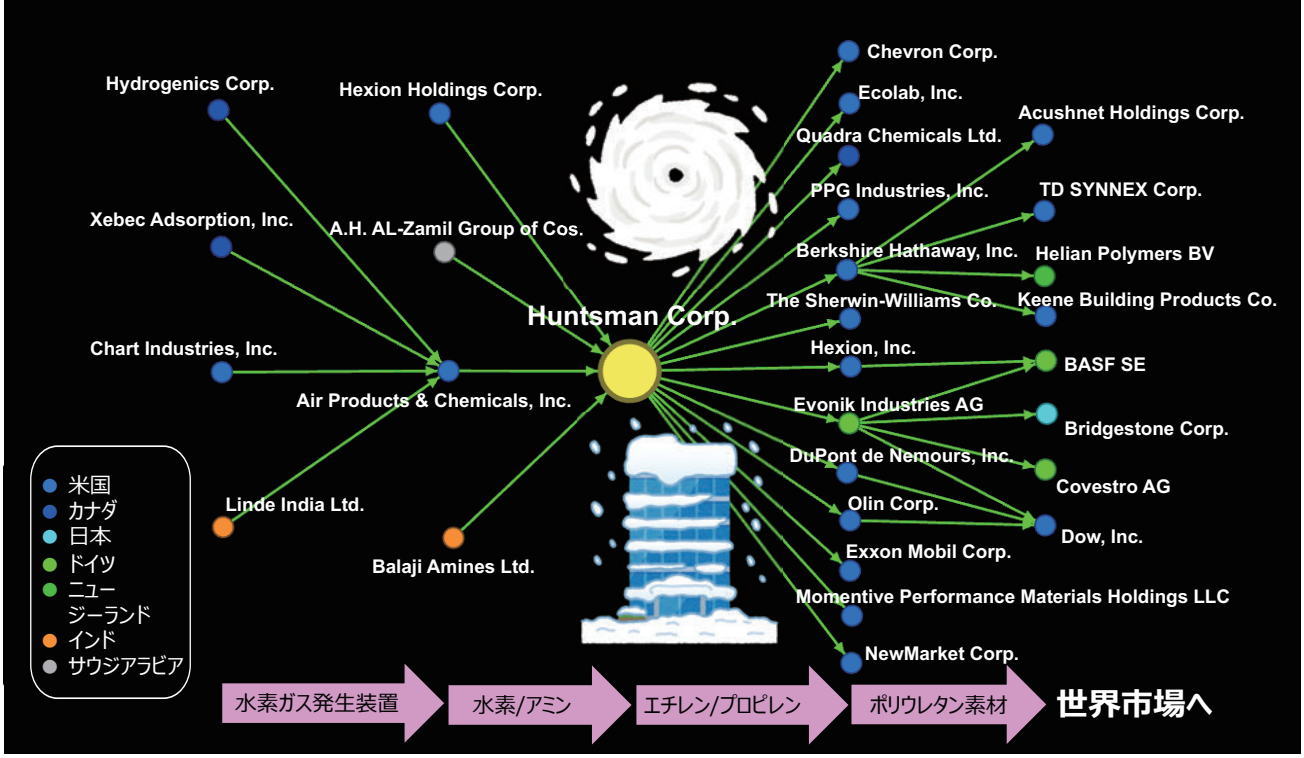
Interview: MIWA, Yoshiko/Composition: KISHIMOTO, Harue/Photos: KURASHIMA, Kazuki

Visualizing Supply Chain Risks with Predictive AI

Supply chain risks are becoming more complex and increasing due to frequent disasters, conflicts, pandemics, etc. Tokio Marine dR Co., Ltd. and the National Institute of Informatics (NII) are using such problems as a starting point to conduct joint research. This research uses predictive AI technology to quantitatively assess how a shock occurring at a business partner will spread to a target company. We spoke with the two researchers in charge of this project about the value that their results will provide to society.

SATO, Ryoji
Senior Risk Analyst, Fusion-Oriented Data Business Unit , Property Risk Engineering Department
Tokio Marine dR Co., Ltd .

MIZUNO, Takayuki
Associate Professor, Information and Society Research Division, National Institute of Information



Estimated from FactSet Supply Chain Relationships 2021

——How did you decide to conduct this joint research project?

MIZUNO: I became involved in researching global supply chains in 2016. The frequency of extreme weather patterns around the world and the growing globalization of corporate trading networks have made it easier for shocks that occur somewhere in the global chain to spread out to other regions and industries. That is exactly what happened when Hurricane Katrina struck the U.S. in 2005 and when the Great East Japan Earthquake hit in 2011. The impact of shortages of procured goods and the shutdown of transportation infrastructure, etc. affected companies in locations far from the disaster area.

In order to visualize such supply chain effects, we started joint research with Tokio Marine dR in 2018. In our field of informatics, we have researched how information propagates over the Internet. The proposal behind this latest initiative is to replace the Internet with the supply chain and use informatics methods to trace how the negative impact of disasters, geopolitical risks, pandemics, etc. propagate.

SATO: The Property Risk Engineering Department of Tokio Marine dR provides quantitative evaluations and consulting on fire and natural disaster risks to our client companies. We combine national and local government data on natural disasters, etc. with information from our client companies to evaluate risks, enabling our clients to utilize those results in disaster prevention measures, appropriate insurance design, and other forms of risk management.

In contrast, when it comes to the supply chains of our client companies, they involve complex, highly confidential information, which makes it difficult to understand them in detail. In many cases, even our client companies are only able to manage the "primary business partners" that they conduct direct business with. Therefore, we are forced to make huge assumptions to evaluate risks such as supply chain disruptions due to disasters, etc. I felt that was a major issue in our business.

Around that time, Professor MIZUNO was using big data that could be externally obtained and researching the extraction of insights from that data with AI. I thought that perhaps that technology could be combined with our company's knowledge of accidents and disasters to conduct a risk evaluation of client companies and proposed that we engage in joint research.

Approximately 17% of business partner shocks spread to target companies

——What kind of research did you specifically carry out?

SATO: One of our initiatives is the quantitative assessment of "shock ripples" that occur when the negative impact of a disaster, etc. spreads through business relationships. The existence of shock ripples has been demonstrated in prior research. However, conventional approaches are unable to quantitatively evaluate their magnitude. Accordingly, we built a predictive model using AI machine learning as a new approach.

MIZUNO: Specifically, we adopted the following method.

First, we extract the primary business partners of the target companies including both the suppliers and customers from a global supply chain dataset which contains approximately 380,000 business relationship entries. Next, we add the "sales growth rate," "country," "industry type," and other information to the extracted target companies and primary business partners to build a model that predicts the sales growth rate of the target companies from the other added information. We then analyze the behavior of the constructed pre-

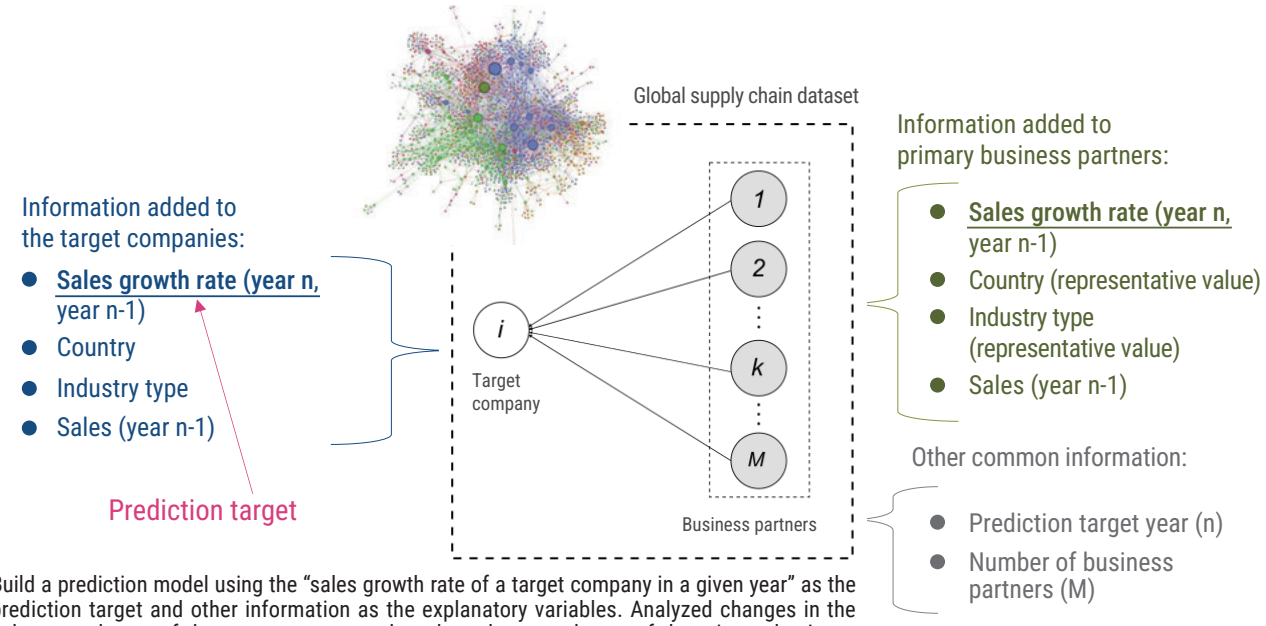
diction model to investigate how a decline in the sales growth rate of a business partner impacts the target company. Furthermore, we adopted "CatBoost (evolutionary algorithm for decision tree)" as the AI training method. The AI model built using this method is visualized using the partial dependency plot method which shows "how the predicted target value (sales growth rate of the target company) changes when only the information of interest (sales growth rate of the business partner) changes."
According to the results, when the sales growth rate of a business partner declines, the sales growth rate of the target companies also tends to fall, and when the shock to a business partner is set to 100, on average it spreads to approximately 17% of the target companies. Furthermore, there are no significant differences based on whether the business partner is a supplier or customer or whether it was in a manufacturing or non-manufacturing industry. The magnitude of the shock ripple was generally the same.

SATO: Later, we compared the results with the case of Hurricane Sandy, which struck the East Coast of the United States in 2012, to verify whether the results significantly diverge from actual shocks. The results showed that the magnitude of typical shock ripples derived with the AI model does not significantly differ when compared with the shock ripples of actual disasters, and the results were found to be valid.
This prediction model indicated that the factors which have the greatest impact on the sales growth rate are the macroeconomic trends that depend on the year, country, industry type, etc. while micro information such as the performance of business partners also has an effect at a level that cannot be ignored. I feel that the great significance of this joint research is that building and analyzing the AI model enabled us to quantitatively assess the impact of business partner shocks separately from these kinds of macro factors.

Developing a system that estimates the entire supply chain with AI

——How can companies and society enjoy the benefits of these research results?

SATO: One area where we can implement this research in society is the supply chain communication service "Chainable" provided by our company.
Launched in May 2023, Chainable is a platform that can visualize and manage risks in a company's supply chain. When a disaster occurs near their own registered sites or those of business partners, client companies receive alerts, and the at-risk sites are visualized. In addition, the platform supports batch transmission of impact verification tasks to the managers at a company and its business partners, chat and other forms of two-way communication, and file sharing to enable a rapid response in an emergency as well as measures and relationship strengthening during normal times. We are considering whether to implement the joint research result in the supply chain risk evaluation part of Chainable.
Particularly in today's world, supply chains are becoming more complex as described above, and in many cases the companies involved are not aware of channels outside of their primary business partners. That is where Professor MIZUNO's system, which es-



Build a prediction model using the "sales growth rate of a target company in a given year" as the prediction target and other information as the explanatory variables. Analyzed changes in the sales growth rate of the target company when the sales growth rate of the primary business partner is changed.

timates huge supply chains using AI, becomes extremely useful.

——What is a system that estimates supply chains?

MIZUNO: First, the technology utilizes business partner information disclosed by a company and then predicts the transaction routes with AI for the parts that are missing. By using this technology when a disruption occurs somewhere in a supply chain, the system can predict whether it can be replaced with a different route or if a replacement will not work and a different measure is required.

The difficult part of supply chain analysis is when the shape of an item changes through processing and assembly as well as when many companies make multiple products. Therefore, there are some cases in which companies and products are not affected even if a negative shock occurs in the network. Excessive risk evaluation could encourage companies to implement unnecessary measures, so we focused on properly classifying and extracting risks.

SATO: For example, Huntsman Corporation, a global company based in Texas that manufactures and sells organic chemical products, has a massive supply chain with over 50 manufacturing sites in roughly 25 countries. When a cold snap and hurricane struck the U.S. Gulf Coast in 2021, most of the company's ethylene production stopped, which led to unplanned downtime in their functional products division. In contrast, the impact on their advanced materials division, textiles division, and other divisions was limited. To properly understand the scope of impact of a shock in a massive supply chain, Professor MIZUNO's technology is key for its ability to trace what is happening upstream and downstream when manufacturing products.

MIZUNO: In addition, it is also extremely important to estimate the transaction amounts. When a supply chain is represented with a chart, it is just an assortment of connected lines, but the actual thickness of the lines, which is the transaction amount, greatly differs for each transaction. Therefore, we make predictions using machine learning. In that way, we work together on a daily basis to improve the prediction accuracy for losses in an emergency from various perspectives.

Visualizing counterparty risk before doing business

——What kind of value do you hope to provide through this research?

SATO: This technology can not only extract just the related transactions to predict the impact of a shock but can also predict the impact on locations far away from a disaster, etc. For example, if manufacturing stops at a basic chemical manufacturer such as Huntsman, which could lead to a shortage of intermediate goods such as polyurethane and final goods such as resin products, it might be very useful to be able to present such risks to companies that are not directly involved in that business.

Moreover, we believe that estimating counterparty risks in the stage before starting a business transaction is an important topic. Assessing such risks in advance allows for a more in-depth approach such as conducting a detailed investigation.

In addition to disasters and geopolitical risks, perspectives such as human rights, environmental response, and governance are also essential for measuring supply chain risk. I hope that this technology can help companies build the ideal supply chain while properly addressing such issues.

By making the most of the knowledge and technologies developed through this joint research, many companies can use our service. Doing so will enable us to accumulate knowledge about the supply chains of many companies and utilize that to provide risk information with higher added value. I would very much like to put such a cycle in motion.

MIZUNO: When you select business partners based on a specific objective, they naturally tend to be concentrated in a specific region. While that may increase production efficiency, it also increases risks during an emergency. By visualizing those risks in advance, you can take measures and prepare in advance such as dispersing supplies or creating a supply buffer within your own company for when a disruption occurs.

Personally, I am very interested in predicting the future of supply chains with AI. What kinds of risks will occur in the future and with what probability? In response, if we implement a certain policy, for example, how will the supply chain risks be improved? Going forward, I would like to conduct research to discover the optimal policies while looking ahead to the prospects of Japan and the world in the near future.

In Search of More Natural and Robust "Voice Anonymization"

The protection and concealment of personal information is now being asked more strictly than before. More attention is being paid to various "firsthand opinions/testimonies" featured in television broadcasts. We spoke with representatives from the Japan Broadcasting Corporation (NHK) and National Institute of Informatics (NII) Professor YAMAGISHI, Junichi about an example in which researchers and a broadcaster partnered to implement a new voice anonymization technology.

TAKAISHI, Mamiko
Department of Sound Design, Design Center, Japan Broadcasting Corporation

YAMADA, Masayuki
Department of Sound Design, Design Center, Japan Broadcasting Corporation

YAMAGISHI, Junichi
Professor, Digital Content and Media Sciences Research Division, National Institute of Informatics

Problems faced by voice anonymization

There are various types of information that can identify a particular individual. Such information includes everything from attribute information such as a person's name and address to biometric information including facial features, fingerprints, and voice. While such information is important to prove a person's "existence," it must be carefully protected too due to the risk of harm from misuse.

There are many cases in television broadcasts where ordinary people appear to provide comments and testimony, primarily in news programs and documentaries, and they may want to ensure their anonymity for various reasons. When it comes to obscuring voices, in particular, the most common method has been modulating the original voices up or down to make them more difficult to identify. However, this method has some problems.

"For example, we would add a mosaic, etc. in terms of the visual. For audio, we would layer modulated material to make it more complex and compensate for hearing difficulty with subtitles. That was the only available option, so we have used that method as a matter of course. However, even though they were brave enough to speak up, we sometimes felt that this approach might have a negative effect or sound too artificial to consider their humanity. Was this truly the only method?

I have had discussions with directors, editors, and other production staff members at various projects and always had this vague feeling."

TAKAISHI, Mamiko from the Sound Design Department at the NHK Design Center, who has worked on many welfare-related documentaries, etc., commented on the situation.

To begin with, in the case of simple modulation, the original audio can be restored by performing the reverse modulation. At NHK, they applied multiple pitch changes before adding effects and other types of processing, but the irreversibility was in no way perfect, and there was some dissatisfaction with the post-processing quality as mentioned. On the other hand, they also tried a system developed at the Institute for Research and Coordination in Acoustics/Music (IRCAM) in France. However, there were some problems with practical use such as long conversion times and differences in conversion results between engineers.

At that time, they met Professor Junichi YAMAGISHI of NII, who was attending a study group hosted by the NHK Development Department, and took the opportunity to seek his advice, which led to the implementation of a new form for voice anonymization.

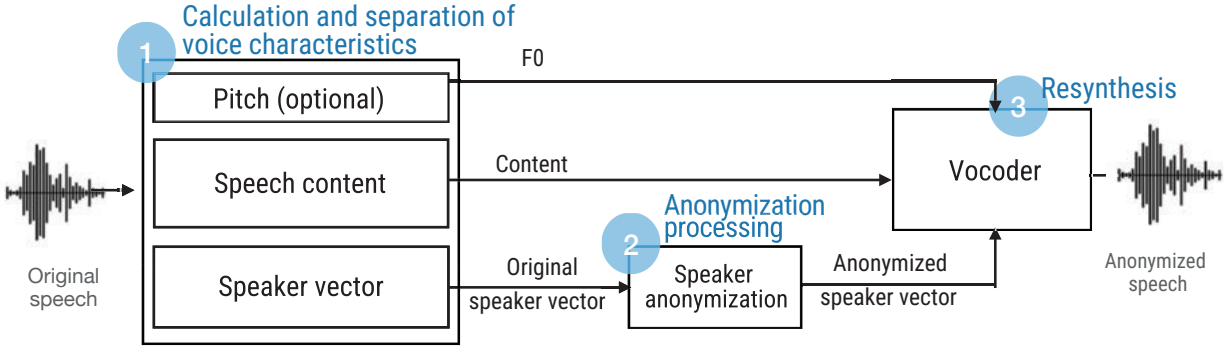
System development and introduction to program production

Professor YAMAGISHI has been involved in researching speech synthesis for more than 20 years. In particular, he has been working on technologies referred to as voice cloning or digital cloning that reproduce a specific individual's voice.

"In recent years, deep learning technologies have developed that enable us to achieve voice cloning at quite a high level. At that time, two new topics emerged that attracted my interest. The first is how to detect the misuse of so-called deepfake voices created with high accuracy through voice cloning. The other topic involves the idea that, if we can synthesize anyone's voice, conversely, that could also be used to protect our privacy. With respect to the latter in particular, due to the amendments to the Act on the Protection of Personal Information, voices are treated as personal data that must be anonymized when distributed to third parties. However, at the same time, it was not clear what kind of processing would be needed to fulfill the requirement of anonymity. Of course you can anonymize a voice by turning it into noise, but that would also lose the content. We had to balance the need for convenience and privacy protection. We were figuring out those issues while developing the

How the VocalGuard automatic speaker anonymization program works

(1) Separates the voice signal into the following two or three characteristic series: (a) speech content, (b) speaker vector, which is thought to include most of the individuality, and (c) the basic frequencies (pitch) if needed. (2) Processes only the speaker vector. (3) Resynthesizes the voice. ▷ Processing method uses k-anonymity of the speaker vector (process that makes it unable to identify the speaker from among k people). ▷ The "k" number can be adjusted to adjust the tradeoff between safety and convenience.



technology." (YAMAGISHI)
It was right at that time when NHK approached Professor YAMAGISHI.
They first exchanged ideas in August 2021. Later, they actually started to implement the voice anonymization system based on Professor YAMAGISHI's technology.
"Finally, at the end of last year (2023), this system was put into practical use on a program for the first time. Since that time, this system has spread by word of mouth, and we have been receiving inquiries including requests to try the technology firsthand. We realized once again that there was an extremely high need for a solution to this problem." (TAKAISHI)
To implement the system, a technology called "k-anonymity" was adopted as the anonymization method.
Specifically, a certain utterance (voice) is decomposed into three components consisting of the speech content, intonation, and individuality. Of these three components, only the individuality is anonymized and reconstituted. At this stage, a pool of voice data is used to convert the voice, reducing the probability of identifying an individual to 1/k or less. The k-value can be changed according to the settings, and the anonymity increases with a larger value.
"I had been researching and developing this technology from before I was approached by NHK, but the initial system built around 2019 supported only English. To make it into something that NHK would practically use, the system needed to support Japanese. It was not simply a matter of swapping in Japanese voice data as we also needed to make adjustments so that anonymity would be preserved among a population of 'k' Japanese people." (YAMAGISHI)

Creating easier-to-use anonymization

The anonymization system implemented at the current time has presets for several conversion modes based on k-values and gender to increase the convenience. To address the request to "avoid replacing it with a completely different voice," there is also an option to include the speaker's own voice within the k people.
"After all, a scientifically high degree of irreversibility is guaranteed, and it sounds natural with just the right amount of processing feel so that there are few concerns that it 'sounds like someone else.' I feel these are the great points when I actually try using it. It has been well received in the field, and I strongly feel that it could become established as a new option. Going forward, I hope that it will become widely used not only in Tokyo but at sites across the country." (TAKAISHI)
"As a matter of fact, this is unrelated to ensuring anonymity, but it is also a very attractive technology from a dramatic perspective. For example, there is a discussion about using this technology as a production technique by processing the voices of actors in a production set in the near future to represent their voices in an AI or digital world or reincarnate them as different personalities." (YAMADA)

At the same time, some issues have emerged as the technology enters practical use. Professor YAMAGISHI has already established two points to improve going forward.
"One point is that we need to further improve the safety and anonymity. First, I am thinking of increasing the voice pool to enable larger k-values to be set. On this point, there are no particular technical barriers as it is only a matter of effort to accumulate the new data, so I would like to get started on this as soon as possible. The other point of improvement is a problem that NHK pointed out when they actually used the system. Depending on the circumstances under which an interview is recorded, it sometimes sounds as if a foreign intonation has been applied to the voice after processing. I feel that this can be handled with some ingenuity in how the audio is decomposed into the speech content, intonation, and individuality elements. After all, even if it is not necessary when simply doing research and only thinking about writing a research paper, many important things emerge in actual usage scenarios. That is very helpful to me." (YAMAGISHI)
This technology is a new step forward in maintaining the secrecy of "personal voice information." We look forward to its further development for social implementation in the future.

Interview and writing:
KAWABATA, Hideki/Photos: SUGISAKI, Kyoichi

Story of Academia and Industry

Researchers in academia are involved in cutting-edge research while engineers create products in the industrial world. Both are experts in creating and cultivating technologies that change the world, however, it is said that there is an invisible wall between the two. What exactly is that wall? We spoke to an engineer and a researcher who have built a collaborative relationship while figuring out the process to hear the story of how they got to where they are today.

ASATO, Akira + **GOSHIMA, Masahiro**
Fujitsu Limited National Institute of Informatics

ASATO, Akira

System Development Department 1,
Advanced Technology Development Unit
Fujitsu Limited

Industry

■ **A relationship that starts from saying "let's talk first"**

—To begin, please briefly describe your background and how you first met.

GOSHIMA: I was an Associate Professor at the University of Tokyo and transferred to the National Institute of Informatics (NII) as a professor about 10 years ago. I have been pursuing the research topic of computer architecture since my student days, and that was how I became acquainted with ASATO. He is somewhat unusual within Fujitsu in that he is deeply committed to academic societies, which is how we first met.

ASATO: I have been affiliated with Fujitsu throughout my career. I reached retirement about three years ago, but I am still involved as a part-time employee. To be more precise, during the first half of my career I was affiliated with a research subsidiary called Fujitsu Laboratories. Although it was called "research," it was different from the kind of research conducted at universities, and I was working on prototypes with a view toward developing Fujitsu products. With the start of the "K" supercomputer project in 2007, the group including myself transferred to a business unit in a reorganization, and I worked there until I retired.

When I was engaged in research-oriented work, particularly in the first half of my career, my boss encouraged me to become involved in academic societies, and I began to serve as an organizing member of a technical committee. It may be due to my personality, but I enjoyed the external interactions with the professors in that way more than staying in the company all the time, and those relationships continued even after the transfer to the business unit. While working in the business unit, I was actively going out and getting to know people in academia. The fact that they remembered me means that perhaps I was somewhat unusual for a company employee.

As for Professor GOSHIMA, I knew his face through my academic society work, but in the beginning that was about the extent of it.

—Under those circumstances, how did you go about building a partnership?

GOSHIMA: Originally, it started with a party at a laboratory at the University of Tokyo. A young professor and one of ASATO's subordinates, who was from that laboratory, got to talk about the possibility of conducting some sort of joint research. They also approached me at that time, and I recall that my participation in that gathering was the direct inspiration to start working together.

ASATO: At that time, the discussion had not explicitly taken the form of "joint research." When it comes to joint research, one thinks of companies providing money and receiving advice, but the meeting at that time was not at that level of discussion. Instead of signing a proper agreement as organizations, we were getting together on an individual basis and asking, "Is there anything interesting that we could do together?"

GOSHIMA: In terms of the situation at the time, Fujitsu did not seem to have a strong interest in working closely with universities or doing anything together. Certainly, they were conducting their own research and development, but it did not seem as if they were trying to incorporate anything from outside the company.

ASATO: I think that was indeed the case. As someone who was building relationships with university professors while working at Fujitsu, I felt the same way as Professor GOSHIMA just described. I sometimes looked around and thought, "These people should really look more outside the company and incorporate cutting-edge research results."

In the midst of all this, I was approached and asked if I wanted to join this sort of gathering. I thought it was a great idea, just the thing we needed. I wanted to actively push forward as much as possible and excite the people inside the company as well.

GOSHIMA: To tell the truth, I did not really expect anything to come

from this at first. Some of the professors who went with me hoped to obtain measurement values for the characteristics of the processors that the companies were making, but I did not think that would happen.
In reality, many of the details about the research and development that the companies are conducting cannot be openly discussed, and when it comes to joint research, non-disclosure agreements (NDAs) must be signed to clearly establish what can and cannot be disclosed. However, the discussion at the time did not go that far.
However, contrary to my initial presumption, they were unexpectedly welcoming and listened to what I had to say. In the beginning, we met about once every six months. We would often hold a meeting and listen as the members said, "This is what we are thinking about right now in academia" or "This is what we are researching in the lab."
ASATO: Yes, that's right. Therefore, in terms of a "give and take" perspective, at that time we were mostly just receiving information and listening to what was going on in cutting-edge research.
GOSHIMA: Even so, considering the impression that I had of Fujitsu before that, they were unbelievably open (laughs).

I think that was right after I moved to the NII. After many such meetings, we finally started to discuss full-scale joint research from around 2020.

■ Benefits of exchanges

—During the initial meetings when the discussion had not yet reached the point of "joint research," what were you getting out of it even without an exchange of specific research results?

GOSHIMA: From the standpoint of academia, there is meaning in talking about what you are doing with someone, even if it is within the scope of public knowledge, and it is fun. Plus, there is value in having the people who are actually making the processor listen to us. Naturally, there are individual differences even in academia, but particularly for those of us who are positioned close to industry, we hope that what we are thinking about will make its way out into the world. However, in reality it may be difficult to find a path to connect our ideas to the world. Even if we were not engaged in full-scale joint research and it was just at the level of casual conversation at that time, if what we were saying could have some sort of impact and make Fujitsu's processors even a little bit better, I think that we would have been happy.

ASATO: I myself have long been in a position where I could see both industry and academia and felt stressed that "even though I could see them, they had no points of contact." I would like the engineers working in the development department in particular to take a greater interest in what is now being discussed in the academic world. Of course, in reality, I would like them to be proactive and go outside of the company to expand the scope of their interactions.
Naturally, it is up to the engineers themselves to decide how to use the information that they obtain there, but at any rate I thought that if we were engaging in development without knowing "what is happening right now," like a big fish in a little pond, we would definitely fall behind the rest of the world, particularly in a field like processors. So, I was hoping to use such an opportunity as a breakthrough to create even better relationships.
For example, in the case of companies overseas, they actively create endowed courses, etc. at well-known universities, gather top students there, rapidly promote joint research to produce papers, and win awards. Naturally, such results are also applied to products as well. Recently, Fujitsu has also been engaging in such activities to some extent, but at the time it

GOSHIMA, Masahiro
Professor, Information Systems Architecture Science Research Division,
National Institute of Informatics

Story

was still far off, and I thought that I wanted to make Fujitsu more like overseas companies.

GOSHIMA: I think that the original reason why exchanges between academia and industry were not very popular has to do in part with the "self-sufficiency" of Japanese companies. I feel that this applies to many Japanese companies in varying degrees, and they have an extremely strong desire to create everything in-house without buying any technologies from the outside.

However, that approach will not be enough to compete in the era ahead. At the very least, in the computer architecture field, there is a large gap in the scale of the research community between the U.S. and Japan. The depth of the layer of researchers and the number of related personnel are probably separated by an order of magnitude. Therefore, even if you conduct research, it is difficult to make an impact unless you come up with a truly amazing strategy and pull it off. To do so, I think that collaboration between academia and industry is essential.

■ Emerging issues and future prospects

—As you proceed with this partnership, are there any parts that feel particularly challenging, and what issues, etc. do you see going forward?

GOSHIMA: Compared to when we first started out, I think that the situation has improved considerably. However, I still get the impression that there are still confidential aspects as always.

We are now actually conducting joint research and have signed NDAs. However, for example, the NDA signed with the research division differs from the one signed with the business division, and at the researcher level, we are figuring out how much we can talk about what the business division is doing. Moreover, it can take a great deal of time to make that determination. I feel that we are quite far off from the feeling of integration between university research labs and companies like in the U.S.

ASATO: However, in a U.S. environment, I think that there is frequent movement of personnel, and the same person may belong to both organizations unlike Japan where people are strictly separated by organization. Perhaps a community of researchers and engineers has formed beyond the frameworks of companies or universities, and information is actively being exchanged within this community. That kind of culture does not exist in Japan.

If I could make one more observation from my perspective—I myself am not involved in joint research, etc. these days, so I do not have a specific example. However, I think that there is still a problem of awareness on the company side. So far, I think that the situation has gradually improved, but companies are still closed off in some ways. I think there are probably some people working on the front lines who are aware of the problem, but ultimately, I would like to see more change in the people who make the management decision to give the "green light."

GOSHIMA: Well, particularly in the past five years or so, I think that things have changed quite a lot. I think that is largely due to ASATO's initiative. I feel that as we gradually build up experiences such as giving advice to the design team in meetings, etc. about actual problems they are struggling with and having that work out well, the trust relationship has also grown.

In any case, I think that there is a need to accumulate more of those so-called "good practices" in terms of both scale and number of contributions from academia. If I were being honest, I think it would be ideal if academia could just devote themselves to research without thinking about the companies, and then the companies would discover and apply those results, but it does not really work that way. After all, I think that academia needs to actively speak up and say, "it would work better like this." In fact, recently Fujitsu has been asking us more frequently, "What should we do here?" which was not the case in the past, so it feels like the world has significantly changed. It has taken 10 years for things to change this much, so how much will they change in the next 10 years? I will be retiring in 10 years, so I wonder if we should speed up this pace (laughs). I would like to see more drastic changes.

ASATO: I myself have already reached retirement age and will retire from my current temporary position in a few years, so it is not up to me. However, I would like to see the things that we have built up over the past 10 years run more smoothly.

Interview and composition: KAWABATA, Hideki/
Photos: SUGISAKI, Kyoichi

NII NEWS TOPICS

Period: March 1 (Fri.), 2024 ⇀ ⇀ July 31 (Wed.), 2024

More details about news items are available online.
www.nii.ac.jp/news/2024 (Japanese)

Facebook
www.facebook.com/jouhouken/

X (formerly, Twitter)
twitter.com/jouhouken

YouTube (audio will play)
www.youtube.com/user/jyouhougaku

Bit (NII Character)
X (formerly, Twitter)
twitter.com/NII_Bit

Send us your comments about NII Today.
www.nii.ac.jp/today/iken

Subscribe to our mailing list.
www.nii.ac.jp/mail/form

Read past issues of NII Today online (English).
www.nii.ac.jp/en/about/publications/today/

The print version(Japanese version only) can be distributed to organizations such as school libraries and research institutions. Please e-mail kouhou@nii.ac.jp to apply to receive or discontinue receiving the print version. It is not distributed to individuals, but copies are always available to pick up from the 1st floor of the National Center of Sciences Building in Hitotsubashi, Chiyoda-ku.

News Releases

2024

- July 18 Computer Science Park 2024 on July 31 (Wed.)! -There will be plenty of play areas to learn about the programming mindset and the popular "100 Research Projects in Rapid Succession" will return after five years! -
- July 17 Projection mapping exhibition of SINET6 data flow rates on a map of Japan - Visualizing the importance of network infrastructure that "just works" -
- June 13 Positive solution to a 15-year-old unsolved problem in the field of combinatorial reconfiguration - Research paper by Associate Professor HIRAHARA, Shuichi of NII accepted by STOC 2024, a top conference in the field of theoretical computer science -
- May 30 You can ask Socrates for life advice with generative AI!? Developing an interactive system that contributes to the dramatic development of western classical studies
- May 23 A wide-ranging introduction to the activities of NII! "NII Weeks 2024" - Open House (June 7), Academic Information Infrastructure Open Forum (June 11-13), Japan Open Science Summit (June 17-21) hosted consecutively -
- May 22 Trial version of the OER Repository now available - Enables cross search of teaching materials developed by universities -
- Apr. 30 Development of the large language model "LLM-jp-13B v2.0" - The NII-hosted LLM Research Group (LLM-jp) released the successor model of "LLM-jp-13B," and made all resources used for development open for the public -
- Apr. 18 NII and the Tokyo Institute of Technology signed a partnership agreement for the research and development of Japanese large language models - Working to ensure the transparency and reliability of generative AI models and accelerate their social implementation -
- Apr. 10 Mathematical formulation of ISO 34502 hazardous scenarios for automated driving systems - Automation and streamlining of safety assurance tasks accelerate social acceptance of automated driving -
- Apr. 1 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 -
- Apr. 1 NII Established "Center for Trust & Digital Identity Research and Development"
- Mar. 28 Experience the cutting edge of research and development in academic search infrastructure - CiNii Labs website released -
- Mar. 27 Mutual exchange of knowledge: Advancing R&D of Japan's Open Science infrastructures and the OpenAIRE Graph - Promoting Open Science Practices: OpenAIRE and NII joined forces to advance best practices in research infrastructures -
- Mar. 18 Our proposal selected as a research project for the "Japan - France International Industry-Academia Joint Research" (Edge AI) of the JST-ANR Strategic International Collaborative Research Program (SICORP) - Accelerating research and development of autonomous wireless access control technologies using AI -
- Mar. 4 New functions added to GakuNin RDM data analysis functions - Enables more advanced numerical analysis in MATLAB -
- Mar. 1 PtM: a system for creating video teaching materials with a synthesized voice - Demonstration experiment started -

Awards

2024

- July 3 Research paper authored by HORIKOMI, Taizo (MIZUNO Laboratory, Graduate University for Advanced Studies) and Associate Professor MIZUNO, Takayuki (Information and Society Research Division) won the 2023 Japanese Society for Artificial Intelligence Incentive Award
- June 10 The research team of JSPS International Research Fellow SPIES Alex, Professor INOUE Katsumi (Principles of Informatics Research Division) and others won the Best Poster Award at Technical AI Safety 2024
- June 3 Professor SATOH, Ichiro (Information and Society Research Division) won the Chairperson's Award from the Japan Electronics and Information Technology Industries Association (JEITA)
- May 28 Large language model study group (LLM-jp) won the AAMT Nagao Award
- Apr. 22 Professor HASUO, Ichiro and Associate Professor ISHIKAWA, Fuyuki of NII and Professor KATSUMATA, Shinya of Kyoto Sangyo University jointly won the Award for Science and Technology (Research Category) and the Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology for research on mathematical software that will contribute to social trust of information technology in the coming future
- Apr. 9 NII Associate Professor HIRAHARA, Shuichi won the Young Scientists' Award for "Meta-Computational Average-Case Complexity Research" - The Commendation by the Minister of Education, Culture, Sports, Science and Technology in FY2024 -
- Mar. 15 SATOH, Ryohei (TAKASU Laboratory, Graduate University for Advanced Studies) won the Student Presentation Award at the 16th Forum on Data Engineering and Information Management
- Mar. 5 Professor KOIBUCHI, Michihiro (Information Systems Architecture Science Research Division) earned the title of Fellow from the Institute of Electronics, Information and Communication Engineers
- Mar. 5 Associate Professor HIRAHARA, Shuichi (Principles of Informatics Research Division) won 11th Yamato Scientific Award

Essay

Creating a Bright Future for Science and Technology in Japan

KATAOKA, Hiroshi
Acting Director-General / Vice Director-General, National Institute of Informatics

In February of this year, it was reported that Japan's GDP dropped to fourth place in the global rankings, overtaken by that of Germany. In what is often referred to as the "Lost 30 Years" of the Japanese economy, Japan's GDP per capita was second in the world in 2000 but tumbled to 18th place in 2010 and 33rd in 2022.

Indicators in the field of science and technology also present a concerning situation. According to the "Japanese Science and Technology Indicators 2024" from the National Institute of Science and Technology Policy (NISTEP) within the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan ranks fifth in terms of its number of scientific papers (average for 2020 to 2022) after China, the U.S., India, and Germany. However, in terms of its number of highly cited papers (ranking within the top 10% of each field in terms of how many times they are cited by other papers), Japan maintains its historically low ranking of 13th place, the same as the previous year (in the previous rankings, it fell from 12th place, overtaken by Iran).

Twenty years ago, Japan was ranked fourth, but its ranking has continued to decline. Moreover, in terms of the "number of top 1% scientific papers" with an extremely high number of citations, Japan continues to rank 12th, the same as the previous year (in the previous rankings, it fell from 10th place).

Of course, these indicators do not tell the whole story, but the Japanese government has been focusing its efforts on the promotion of science and technology since the enactment of the Basic Act on Science and Technology in 1995 (amended as the 2021 Basic Act on Science, Technology and Innovation). It has formulated a Science and Technology Basic Plan (now called a Science, Technology, and Innovation Basic Plan) every five years and established the Council for Science and Technology Policy in 2001 (now called the Council for Science, Technology and Innovation). So what on earth has gone wrong ?

In recent years, the total amount of research and development expenses and the number of researchers in Japan have stagnated with little growth compared to the significant increases in both the U.S. and China. Nevertheless, Japan still maintains its third-place ranking globally. That being the case, this stagnation with little growth does not seem to be the reason for the decline in Japan's research capacity.

The White Paper on Science, Technology, and Innovation 2022 introduces analyses by NISTEP covering Japanese universities and cites the decline in the ratio of time spent on research by faculty members, slow growth in the number of faculty members, and stagnation in the number of doctoral course students and expenditures directly related to research implementation as factors behind the decline in the number of scientific papers in recent years. Furthermore, based on a questionnaire regarding constraints in improving research performance, the white paper cited the inability to secure research time due to teaching loads and administrative work within universities and the inability to secure research funds due to factors including a lack of non-competitive core funding as specific constraints.

Japan's budget for science and technology reached 4.9 trillion yen in the initial budget for FY2024, which represents a 40% increase over the 3.5-trillion-yen budget for FY2001 (while the amount including the supplementary budget has particularly increased by a significant amount in recent years to reach nearly 9.5 trillion yen in FY2022 and FY2023, researchers cannot be employed with the supplementary budget). Competitive funds in particular have increased in both number and budget. However, the non-competitive core funding that is the source of diverse research seeds has decreased and is lacking. The personnel problem is also quite critical with the number of students enrolled in doctoral courses trending downward over the long term after peaking in FY2003 (the number increased by 4.4% in FY2023). The percentage of full-time university faculty members under the age of 40 also continues to decrease.

The decline in Japan's research capacity has already been discussed in various ways, and measures have been taken. Nevertheless, tangible results remain elusive. I have also been involved in science and technology administration and did the best that I could in my positions, but I think a major factor is that local optimization does not necessarily lead to overall optimization. Now is the time for the nation as a whole to pool its wisdom together and thoroughly conduct a comprehensive, medium- to long-term examination of the situation to achieve overall optimization and strive to carve out a bright future for science and technology in Japan. At the same time, when it comes to the National Institute of Informatics (NII), the roles of information, data, and AI will become increasingly important in the realm of science and technology going forward, and I think that the NII has a large role to play due to its position at the center of these. For example, regarding generative AI, the NII quickly launched the LLM-jp research group to promote open co-creation efforts such as the development of domestically produced LLMs with more than 1,600 people from industry and academia now participating (as of July 2024). The Research and Development Center for Large Language Models (LLMC) was established in April of this year with a budget from MEXT. It is important that efforts like these be further developed.

Weaving Information into Knowledge
NII

National Institute of Informatics News [NII Today] No. 103 Sep. 2024
Published by National Institute of Informatics,
Research Organization of Information and Systems
Address: National Center of Sciences,
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430

Contact: Publicity Team, Planning Division, General Affairs Department
Email: kouhou@nii.ac.jp

Publisher: KUROHASHI, Sadao
Editorial Supervisor: YASUURA, Hiroto
Advisor: TAKEDA, Koichi; YAMAMOTO, Hiroki
Editorial Committee Chair: ECHIZEN, Isao
Editorial Committee Member: IKEHATA, Satoshi; KANEKO, Megumi; KOMIYAMA, Yusuke; MIZUNO, Takayuki; TAKEFUSA, Atsuko
Editor/ Art Directer: KISHIMOTO, Harue
Cover illustration: ICHIMURA, Joe