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

NII Today

100
Sep. 2023

National Institute of Informatics News

[Feature]

The Challenge to Generative AI

This English translated version NII Today corresponds to
No. 100 of the Japanese edition

INTERVIEW

How LLMs and Robots Will Shape the Future

ChatGPT, a large language model (LLM)-based interactive artificial intelligence (AI) technology, continues to cause shockwaves in Japan.
Not to be shocked into submission, the National Institute of Informatics (NII) launched the LLM-jp (group studying LLM) in May 2023, bringing together over 300 (as of August 2023) researchers and engineers from universities, research institutes, and companies to form a research team on the scale of a large-scale physics experiment.
The group aims to find out how LLMs work and construct an LLM for Japanese, with a view to collaboration with robotics – one of Japan's strengths.
Waseda University Professor OGATA, Tetsuya, who is engaged in humanoid robotics research, and NII Director General KUROHASHI, Sadao, who has spearheaded Japanese LLM research, discussed their dreams for the future.

OGATA, Tetsuya

Professor, Department of Intermedia Art and Science, Waseda University
Director, Institute for AI and Robotics, Waseda University
Joint Appointed Fellow, Artificial Intelligence Research Center (AIRC), National Institute of Advanced Industrial Science and Technology (AIST)

KUROHASHI, Sadao

NII Director General/Professor
Professor, Graduate School of Informatics, Kyoto University

Interviewer
TSUJIMURA, Tatsuya
Editorial Committee Member, Kyodo News

Constructing a Japanese LLM and applying it to robots

—How do you unravel the operating principle of an LLM?

KUROHASHI: The first thing we need to do is make the training data open. We don't know what corpora ChatGPT has been trained on, or what neural networks it uses; it's like a black box. If you ask ChatGPT a question and get an answer, you don't know what that answer is based on. It involves so many unknowns, there are growing concerns that copyrights are being infringed without it being realized.

That's why we want to create an environment where all the corpora are publicly available, so

OGATA, Tetsuya

Graduated from Waseda University. Previously associate professor at Kyoto University's Graduate School of Informatics before his current position. Involved in cognitive robotics research using neural circuit models and robot systems, especially predictive learning, imitation learning, multimodal integration, language learning, and communication research. Awards include the 2023 MEXT Award for Science and Technology.

we can say “these are the data it was trained on,” and all outputs can be analyzed against the input. That way, if a sentence is produced as an answer to a certain question, at least we will be able to observe how many sentences in the corpus are similar to that question, and how hallucinations occur (when the system outputs nonsense answers). Then, we want to work out what’s happening inside this neural network, which is a very challenging goal. And, unlike ChatGPT, we want to make a system that is good at Japanese.

Based on this, we will be able to take on new challenges in various fields of research. One that I find particularly interesting is the application to robots.

OGATA: Using deep learning to control robots is a fairly recent area of research. It was only about six years ago that it became a keyword at major international conferences. There’s still a lot we don’t know about how it should be used. Because of my background in neuroscience, I’m looking at deep learning from that perspective, and researching how it can be used to achieve stable movement of robots.

Until now, we have mainly used humans to teach robots how to move their hands to perform tasks like folding a towel or making scrambled eggs. A robot can move by itself, so even if the situation in front of it is slightly different from the situation when it learned the action, it can act to make the action more like what it knows, and perform the action as instructed. This kind of thinking is called active inference.

The next stage is to make words from actions, or actions from words. We want robots to understand instructions that are worded vaguely. However, we don’t have the huge database required for teaching actions. For example, what movement is meant by the words “stretch out your arm?” It could be anything from picking up a coffee cup to throwing a punch in the boxing ring. This requires wide-ranging teaching and learning, but there is no large-scale database for this purpose.

I want to see inside the LLM to find out whether it can be used to overcome this barrier. I have high hopes for the LLM Working Group (LLM-jp) as a forum to achieve this. Ultimately, I also wonder whether robots can be used at the LLM training stage.

LLM and robots: the international situation

KUROHASHI: There are limits to the research we can do in Japan alone, so once we have established a certain level of computational resources and so on, we are planning to work with at least Asia and Europe. With small projects, we don’t stand a chance against the big tech companies like GAFA (Google, Apple, Facebook and Amazon), either intellectually or financially. By working on a bigger scale, we hope to attract more interest and funding from companies to make it sustainable.

By the way, how is research linking LLMs and robotics being done in other countries?

OGATA: The people working on LLMs at GAFA don’t appear to be connecting it with robots. Robotics developers are not necessarily interested in LLMs, either. They are both treating models as separate modules, once a model has been trained. I wonder if there are people like us who are considering whether robots are required when making an LLM, or if it could be interesting to incorporate some aspects of robotics into an LLM.

OpenAI, the creator of ChatGPT, disbanded its robotics division in 2021, but has started investing in robotics companies this year (2023). Google’s RT-1 project is collecting data from 13 robots operating for 18 months. Apparently the researchers involved

Developmental robotics research is mainly being done in Japan and Europe

believe it is important to actively move robots, similar to the developmental process of living organisms.

There is an area of research called "developmental robotics," which attempts to understand the human cognitive development process through robots. When the world's first humanoid robot was developed fifty years ago at Waseda University, the motivation was as a tool to understand humans. Using LLMs, we might be able to research the question of how language is involved in development.

Research into developmental robotics is mainly happening in Japan and Europe, not in the USA, so this is an opportunity for us.

KUROHASHI: We can't do anything without first creating an LLM. Exploring the process of how humans perceive the world with a physical body, by training a system including robot motor information and visual information, could be the next major development.

Could robots learn more efficiently using LLMs?

—I had the impression that human understanding and robotics were unrelated areas of research.

OGATA: Usually, they are unrelated. It is probably not necessary for industrial robots. But it's only natural that humanoid robot researchers should be interested in this. Why are humans able to move their bodies properly? There are a lot of aspects we don't fully understand. There is a research method known as the constructivist approach that involves actually making a robot or system that works well, and then finding out why it works.

Of course, LLMs are different from humans in many ways, but there are certainly similarities, too. Apparently, when a psychology researcher asked ChatGPT "What is this person imagining?" it correctly answered "I think he/she is imagining something like this." It can seem rather disturbing, but the fact that it can do such things means that it works in a similar way to us.

KUROHASHI: One of our goals is to find out whether robots could quickly become more intelligent with less training, using LLMs.

OGATA: That would be ideal. In our research, we have found that training with words or some kind of labels makes it easier to generate movement.

For example, when training a robot to fold a towel, adding inputs like "you are now holding the towel" and "you have started folding" clearly improves the accuracy of folding, even if the labels are somewhat vague.

When training a robot to perform a motion to hold some kind of object, usually it is just a case of "how close has it come to that action?" But if you add words, attention turns to, for example, the color of the object. In other words, introducing words to explain what the robot has done produces changes. I'm very interested in finding out at what stage this becomes effective.

When a human baby is learning to do something, I think words always act as an external bias, which somehow helps the child to acquire that action. I wonder if it would make it easier for a robot to learn if we could give it certain words within the constraints of the growth and development of its "body."

I'm also interested in the opposite direction: learning language based on a model trained using robot data.

The role of robots: Striking the right balance is vital

KUROHASHI: There is a debate about ethical issues like reliability and bias surrounding the use of LLMs. Robotics has a long history, and the "Three Laws of Robotics" were proposed in the early days. I get the impression that research is conducted with due consideration of ethical is-

sues. What is the reality?

OGATA: There are still many difficulties. For example, in the Japanese Cabinet Office's Moonshot Research and Development Program that I am involved in, a professor of social sciences told us "Please don't make all humanoid robots white." We need to consider race as well as gender.

In Japan, there is not much resistance to using robots for nursing care, but some people feel uncomfortable about having robots take care of their bodies. Using robots to serve food and drinks is another interesting area, but there is an aspect of entertainment in this service, so some people think it would be boring to be served by robots.

International surveys have found that in every country, people think it is OK to get robots to do jobs like cooking, washing, and cleaning. When it comes to cooking, if the main part was done by robots, the fun is also taken away, but robots are fine for tasks like chopping vegetables and stirring pans. The training needs to be just right, and there is strong resistance to using robots alone.

Thus, opinions differ, so we have to be very careful when people are involved, including how the application is presented.

I think the way we interact with ChatGPT is a good example. It's clearly a bad idea to empathize more than necessary with generative AI. We say we want robots to become butlers, silently and diligently doing as they're told. Starting from that point, if the social situation changes, the relationship might change, too. I think it is important to design AI considering how people will perceive it.

KUROHASHI: A language model is originally a technology for communicating with humans, so we must be careful.

OGATA: For robots, I think it's quite different if there is always a human by the robot's side. For nursing care, if the main job is done by a human, with robots doing anything that is dangerous or requires physical strength, it is less of a problem. It is not a good idea to leave it up to the robot to do the important parts involving interaction with people. It's important to strike the right balance.

The aim is to find out how LLMs work and construct a model that is tailored to Japanese

A Word from the Interviewer

It's been about forty years since the Fifth Generation Computer Systems research project astonished the world. Since then, I have not felt much enthusiasm for information science in Japan; it seemed we were one step behind foreign technologies like deep learning and generative AI. Listening to this conversation has made me feel more positive than I have in a long time. We need to keep looking ahead and taking action to maintain the momentum of the LLM-jp. In large-scale experiments in particle physics, the community unites across borders and continues to try new things. I hope this project will be a driving force for the development of information science.

TSUJIMURA, Tatsuya

Editorial Committee Member, Kyodo News

Graduated from 2nd Department of Physics, Faculty of Science, Tohoku University in 1984. Works in the Science Department at the head office (Tokyo), as well as in Osaka, Otsu, Kushiro, Sapporo, and Akita. Has written memorable articles such as "Takamatsu Crater not listed as a meteorite crater," "Ministry of Foreign Affairs refuses visa to Indian physicist," and "No need to reproduce or investigate STAP cells." Publications include "Nihon no chi, doko e?" (Co-author, Nippon Hyoron)

Photography by IMAMURA, Takuma

Press conference: “What generative AI can show us” Q&A session

July 28, 2023 (Friday)

Excerpts from the Q&A session with NII Director General KUROHASHI after the presentation on “What generative AI can show us” at a press conference held at the National Institute of Informatics.

Q1: Is there any prospect of unraveling the undisclosed part of GPT? Or will it never become clear without reaching the same scale of 175 billion parameters?

A Basically, there’s a lot that cannot be understood without making our own LLM. The original training corpus is not publicly available at present, so even if we observe that the model hallucinates (outputting content that does not fit the facts or context), we don’t know what is happening inside. That’s why we plan to make the training corpus clear, to create an environment where the LLM’s behavior can be analyzed by searching the training corpus against the input, which will be the first step towards mathematical model-based research.

Q2: Where will you get 175 billion parameters of training data? What about copyright?

A A non-profit organization called Common Crawl has accumulated data collected from websites around the world using a crawl program, and provides datasets free of charge. This includes data extracted from Japanese web pages and the data have been prepared to some extent, so we will work with those data first. But unless the original training corpus is enlarged as the parameters grow, data accuracy will become biased due to overfitting. So, by the time we create a model with 175 billion parameters, we will basically have to filter the entire Common Crawl dataset as training data. Recent research has found that the better the quality of the data, the better the model. Copyright is a sensitive issue, but we want to negotiate with various sources in Japan to get more data for this project. For example, I think some newspaper articles published online have already been crawled and are being used. I would like to experiment using this model by putting something in as data from a newspaper and seeing whether it really comes out as a verbatim copy or as digested text. I hope we can rely on the cooperation of the newspaper companies.

Q3: The performance of the GPT language model suddenly took off at a certain point. Does this mean quantitative performance improvement turned into something qualitative?

A There are various types of tasks measuring the performance of a language model (summarizing text, answering questions, and so on). If you increase the number of parameters of a language model by a factor of 10 or 100, at some point there is a sudden leap in performance in these tasks. Performance in a considerable number of tasks has been found to improve when it reaches 175 billion parameters. In a sense, we could say that quantity is turning into quality.

Q4: If models become even larger in future, will something unintended emerge?

A This is a very difficult question, and there is an ongoing debate about whether further training will cause language models to gain consciousness and harm humans. In terms of future predictions, we cannot say anything has zero possibility, but it is particularly difficult to discuss whether consciousness will emerge. If a language model is given a prompt telling it to harm humanity, it could give that kind of response, so it could do bad things without the language model itself being conscious. Just like the debate around knives, which can be used to prepare delicious dishes but can also hurt people, we need to talk about a social framework to use it in a way that is safe and healthy. To provide a safety net, I think it is vital to have a proper debate about how AI language models are used.

Q5: Would the LLM-jp consider collaboration with industry in future, such as leveraging LLM in manufacturing?

A That’s something we would not rule out. This project is completely open, so more people are gradually getting involved from various sectors of industry. We decided to allow anyone to participate, both researchers and company representatives, as long as they understand that all our discussions will be open. We have already come this far as a result, so we will carry on with this approach for now. If we are going to specialize in various fields in the future, it is possible that we may end up talking to companies individually, but we want to keep the LLM Study Group as a hub for information exchange in its current form.

Q6: As the language model learns, will it reach a limit to the amount of information, where it cannot produce new answers without new content created by humans?

A If it is given all kinds of knowledge, it will be able to answer within the scope of that knowledge. As for whether anything would happen beyond that, I don’t think it will come up with some kind of sudden insight. But solutions can often be found through collaboration between experts in different fields. For example, researchers in medicine are struggling with this issue, but this technology exists in engineering, and if this legal issue can be resolved, then this treatment would be possible… A language model with huge amounts of knowledge in every field might be able to produce new suggestions like this. It’s unlikely that a language model could ever come up with creative insights like Einstein, but I think it could certainly produce innovation by linking knowledge from different fields.

(Summary by NII PR Team)

The Newly Launched LLM-jp

Professor, Digital Content and Media Sciences Research Division, NII
Vice Director-General, NII
AIZAWA, Akiko

Associate Professor, Digital Content and Media Sciences Research Division, NII
KANAZAWA, Teruhito

Assistant Professor, Digital Content and Media Sciences Research Division, NII
SUGAWARA, Saku

Interviewer
TAKI, Junichi,
Senior Writer, NIKKEI INC.

It is important to lay the groundwork for research by putting the computational platform and language model building platform into place

The National Institute of Informatics (NII) has launched the LLM-jp, the group studying LLM, bringing together natural language processing researchers from Japanese universities and companies. In a spirit of "cooperation, not competition," it aims to build an open-source large language model that is tailored to Japanese, promote related research and development, build a network of researchers, and develop new human resources for the "AI native era." We talked to three of the main members, NII Vice Director-General AIZAWA, Akiko, Associate Professor KANAZAWA, Teruhito, and Assistant Professor SUGAWARA, Saku, about the background and aims of the LLM-jp and its ripple effects.

—First of all, I would like to ask about the aims of LLM-jp.

SUGAWARA: LLM-jp is made up of natural language processing researchers from all over Japan, invited by NII Director General KUROHASHI. Its purpose is to create a large language model (LLM) that is tailored to Japanese, and to promote research into understanding the principles and utilizing such a model. We held our first meeting in May and our third meeting in July. As of August 2023, over 300 people have registered to take part.

—I believe there are some teams within LLM-jp.


AIZAWA, Akiko

SUGAWARA: We have four teams: a team to prepare the data required to train the LLM; a team to build and train the model; a team to tune and evaluate the model for specific purposes; and a team to set up the large-scale computational infrastructure required to create the model.

AIZAWA: There is a large-scale platform called mdx (platform for building a data-empowered society) that is jointly operated by several academic organizations, including information technology centers in national universities and NII. We have launched a project to build a large language model using this platform. We plan to release a model with 13 billion parameters by the end of 2023, which is relatively small for an LLM. LLM-jp has set a goal of a model with 175 billion parameters.

As well as creating the model, later phases of the project will look at how to adapt it for use in various domains, whether it can be used for natural language processing research, whether it can be leveraged to address social issues, and so on.

A large language model is also called a foundation model, which means it can be used in all kinds of situations. I think this will become a large project involving the scientific community using computers in image processing, robotics, medicine, biotechnology, materials, and so on, in future. For example, in social sciences, analyzing society through language is an important area of research. Another mission of the LLM-jp is to answer the scientific question: what is human language? From the perspective of being tailored to the Japanese language, the language model will reflect social norms and rules and various human social activities, so it could also be used as a tool to observe and analyze our society.

Advanced information search and retrieval with the LLM

KANAZAWA: As part of the development team for CiNii Research, the academic information search platform provided by NII, I am making improvements by applying my own research results. I am taking part in LLM-jp with research, development, and practical application in mind, with a view to applying the LLM to CiNii Research. How can we make academic information easily accessible across domains without being research-field-specific or even domain-specific? I hope the LLM will help.

In the past, information retrieval involved metadata precisely attached to information. Users could refine a search by using appropriate search terms, but it was not easy to master the system. With CiNii Research, you just have to enter a keyword and it will come up with suggestions and give you options to further refine the results. Complex queries (processing requests) can now be refined without any thought on the part of the human. But if the user interface is too simple, it is sometimes difficult to express a query in an appropriate way.

I think if we get the LLM to write complex search expressions, appropriate instructions can be given with input in natural language. It is hard to put such requests into program code, but it is now possible to write high-quality code in a short time using an LLM. In the world of programming, there is talk of “no code/low code” methods allowing users to give instructions without any programming knowledge, with an LLM expected to handle complex operations. We


KANAZAWA, Teruhito

want to apply a similar "no search/low search" approach, allowing users to access the information they want simply by asking for what they want to know.

Another of my research themes is developing technology to analyze and strengthen research capabilities to promote and support research work. We are investigating a system to match up researchers for interdisciplinary research, but it is difficult and time-consuming for researchers to explain their research topics in a way that can be easily understood by those working in different areas. We are looking at whether we could apply an LLM to create a system to help researchers understand each other in a shorter time. As Professor AIZAWA mentioned, we want to encourage joint research between researchers in every area, from informatics to medicine and social sciences.

Laying the groundwork for research for the future

SUGAWARA: Speaking from my perspective as an early-career researcher, large American companies are taking the lead in LLMs, and it would be very expensive to reproduce an openly available model for research. Just to run the largest open-source, free-to-use model requires a server costing around 10 million yen, which is a significant barrier for early-career researchers and students.

Individual research labs in Japan often don't have a lot of money. By getting together and sharing a platform, data, codes, and programs to create the model, it means students and post docs can also get involved. This is one of the major benefits.

Personally, I am interested in evaluating the model. Even with the latest models, there are aspects where it is doubtful whether they can read text properly. If you ask the model to explain what its answer is based on, it might not be able to explain properly; this appears to be an issue.


SUGAWARA, Saku

AIZAWA: I think there are two key aspects: this LLM is being developed by academia, and we are focusing on the Japanese language. ChatGPT has brought a sharp increase in the scale of language models. With these rapid changes, natural language processing, or perhaps artificial intelligence, has suddenly joined the realm of "big science." Although we don't have the same budget scale, it feels like we have suddenly entered the arena of big science, just like the large particle accelerator project in particle physics, in the sense that progress is only possible if researchers join forces.

Director General KUROHASHI led an emergency panel at the annual conference of the Association for Natural Language Processing in March 2023, which spurred researchers to join forces. Senior researchers seemed to be united in realizing this needed to be done for their junior colleagues, looking five years ahead. While the aim of LLM-jp is to create an open model to be widely used, as a senior researcher, I am aware of creating a large scientific arena for the next generation.

Meanwhile, I have been saying for several years that in the era of data-led science, it will be difficult to protect Japan's identity if we remain passive. I believe it is important to invest in the information we have in Japan, but this does not produce research achievements. While researchers compete to get their papers published in journals with a high impact factor, there is no motivation to focus on Japan. I think many researchers were in agreement that it is important to work on collecting language and information as part of Japan's identity.

—How do you see the impact of LLMs on scientific research?

KANAZAWA: It will make it easier and cheaper to access academic information, which is very significant, whatever domain of research you work in. I cannot give specific examples, but the model created by LLM-jp will be lever-

LLM-jp
llm-jp.nii.ac.jp

Purposes of LLM-jp

- Build an open-source large-scale model that is tailored to Japanese
- Promote regular exchange of information among researchers and cross-organizational collaboration between researchers
- Make everything open, including results, discussion process, and failures

Issues related to the LLM

- Technical issues: Mathematical explanation of learning principle, efficiency
  Social issues: Explainability & interpretability, fairness, safety, reliability
- Expansion to multiple fields: Expansion to medicine, law, education, etc. Link-up with multi-modal information, robot control, etc.

Future policy (as of August 2023)

- Build a model with at least 175 billion parameters in Japan and work out the principle. Build a 13 billion parameter model by the end of 2023.

aged in many different areas, and it will be evaluated in future. As a researcher of information retrieval and access technology, I am waiting in anticipation.

SUGAWARA: In the field of AI, it is usual these days for top researchers to get together to conduct research and write papers in teams. Strengthening connections between researchers is important for producing globally impactful results. We should form teams driven by academia, and I think LLM-jp will play a significant role as a hub for collaboration.

AIZAWA: I think the effect of the LLM on science will be huge, a really critical impact, with the potential to accelerate research in every field. When I asked biotechnology researchers how they would use it, they anticipate searching databases of genetic information – an advanced usage that researchers of natural language processing like myself would never have thought of.

Will LLM research lead to cognitive and language research?

—Will research into the mechanism of how an LLM generates fluent sentences lead to answers to questions about human language and how the brain processes language?

SUGAWARA: I think it's quite difficult to say whether what we find out from observing a language model would necessarily apply to the human brain. But we will be able to see if the model behaves like a human in certain conditions but not in other conditions. So, as we narrow down the conditions in which it behaves the same as a human, and as we devise ways to train the model, by observing that "if we train it on this kind of information, then it can understand this," then we might find things that can also be said about the human brain.

Nobody can give clear answers yet, but for now, perhaps the most important thing is more active communication with researchers in fields like linguistics, cognitive science, and neuroscience.

AIZAWA: I think it will lead to cognitive research. The mechanism must be completely different, but I hear that a language model has been successfully used to decode brain waves in cognitive research. The model seems to capture some characteristics of how the brain works to understand language. This could be a future topic for LLM-jp.

A Word from the Interviewer

A newspaper article is the combined work of many reporters, and is the property of the newspaper company. The Japan Newspaper Publishers & Editors Association has expressed concerns about the unauthorized and unregulated use of news content by generative AI. But my personal opinion is that news articles are in the public interest, so it would be better to find a way to allow them to be used with the appropriate protection of rights.

TAKI, Jun-ichi
Senior Writer, NIKKEI INC.

After graduating from the School of Political Science and Economics at Waseda University, joined Nikkei, Inc. After working in branch offices and covering corporate news, began covering science and technology, as well as environmental fields, starting from the mid1980s. Authored "*Eco-Uma ni Nore!*" , among others.

Photography by KOYAMA, Kosuke

Interview

"SYNTHETIQ VISION" to Automatically Detect Fake Media

Global Research Center for Synthetic Media

SynMedia Center

In addition to generative AI for natural language (text), rapid progress is also being made in generative AI for images, video, and audio. As it becomes easier to create AI-generated photos and voices that can be mistaken for the real thing, there are growing concerns about this technology being misused for purposes like fake news or scams. We asked ECHIZEN, Isao and YAMAGISHI, Junichi, leaders of the Global Research Center for Synthetic Media, what the Center is doing to protect against deepfakes, and about the "SYNTHETIQ VISION" system they have developed to automatically detect AI-generated fake facial images.

ECHIZEN, Isao
Professor, Information and Society Research Division, NII/ Director, Global Research Center for Synthetic Media

YAMAGISHI, Junichi
Professor, Digital Content and Media Sciences Research Division, NII/ Deputy Director, Global Research Center for Synthetic Media

Interviewer
YAMADA, Tetsuro
Editorial Writer, The Yomiuri Shimbun

—What are the research goals of the Global Research Center for Synthetic Media?

ECHIZEN: It is now possible for AI to learn from human faces and voices and generate "synthetic media" – synthetic images and voices that can be mistaken for the real thing. This can be used in various ways: you can communicate with a virtual avatar, or create videos and music for entertainment. But it also has a negative side. There are growing fears about synthetic media being misused for scams or to manipulate public opinion, or fake videos being shared by criminals for their own amusement, which is becoming a major social issue.

The Center is an international research center working on ways to detect fake media generated for dishonest purposes, to ensure the credibility of the media.

—Please tell us how it came about.

ECHIZEN: The Center was launched in July 2021 with funding from the JST Strategic Basic Research Program (CREST), a large-scale national research fund. Dr. YAMAGISHI started research on speech synthesis and voice conversion as a joint project between Japan and France. I joined in, researching the theme of how to deal with fake media. As our fields are closely related, working together produces synergy, allowing for diverse joint research.

YAMAGISHI: Even before the CREST project began, we have actually had many opportunities to work together. We were doing pioneering research into deepfakes with Professor by Special Appointment BABAGUCHI, Noboru, from Osaka University, another member of the Center, so we were already heading in this direction. For example, in 2014 we presented a

liveness detection technique to determine whether a voice is a real person or a generated voice.

This area is now called "generative AI" or "information security," but at the time it was known as "digital cloning" or "media cloning." It's the same concept: looking at images, videos, audio or text created by AI technology, identifying social issues and problems for biometric authentication or security systems, and finding ways to tackle these issues in advance.

—Automatic translation, for example, improved so rapidly from a certain stage that it now eclipses the language ability of the average person. How about image and speech synthesis?

YAMAGISHI: Deep learning technology first entered the field of speech information processing in 2013, which dramatically improved the accuracy of speech recognition. What's more, it quickly reached the same level as human recognition ability. Since 2013, there have been remarkable changes in speech synthesis, too. The technology behind the apps we have now was already developed by around 2018. But it took time to catch on due to issues with processing time and the cost of using it as a product. Now that computer performance has improved and the cost issue has been solved, it has started to be widely used. The same is true for speech generation AI. It's now so easy to get a computer to accurately imitate somebody's voice that anybody can use this technology. Well-known image generation AI systems like Stable Diffusion have appeared, too.

ECHIZEN: When it comes to images, issues with deepfakes, like replacing a face in a photo with a different person's face, began to appear around 2017. That's when fake photos started to look just like the real thing.

Visualizing detection points

—The SYNTHETIQ VISION fake facial image detection program that the Center has developed is now drawing attention.

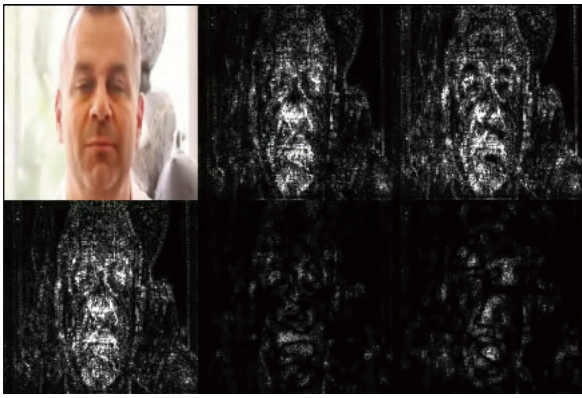
YAMAGISHI: Of course it's a good thing that everyone can now benefit from AI technology, but on the flipside, this means that unscrupulous people can easily create fake videos with the faces replaced. Ordinary people can create deepfake videos using an app. For example, you could generate a video of an imaginary person giving a "customer review" and use it to advertise your product, or insert a celebrity's face in place of someone else's face and upload it to social media.

We created SYNTHETIQ VISION to deal with these issues. With this system, the technology we have developed to distinguish between real and fake images can be used in the real world. It works for various generation methods, and can accurately detect whether an image is fake, even when a human cannot tell the difference at first glance.

—How does it detect fakes?

YAMAGISHI: Basically, a neural network searches for reference points to tell whether an image is real or fake. Looking only at the results, it is a "black box" technology where we don't know what is happening inside, but there are ways to visualize where the AI is looking to determine whether an image is real or fake. Of course, the focus points depend on the individual fake image and the algorithms used to generate it. So, it's not like you can look at it with the human eye and say "this part looks odd" or "this doesn't look right," but it shows what the neural network is referring to.

Figure 1: Visualizing the parts that AI focuses on



Taking a certain facial image (Figure 1) as an example, AI is reacting to the areas around the eyes and mouth. In the end, perhaps it was the area around the nose that was useful for prediction. With a different photo, it might be the chin. This only applies to this method of processing for this particular photo, so the same cannot always be said. In any case, the points that a neural network looks at will not necessarily be useful to a human, so it is always difficult to explain. Unfortunately, this is common to all deep learning systems.

The mission to detect fakes

—SYNTHETIQ VISION was implemented very quickly, wasn't it?

ECHIZEN: A lot of companies are concerned about fake me-

dia, or see it as a business opportunity, but this is not something that a company could develop quickly. Firstly, as with any deep learning technology, you need an enormous amount of image data and fake data to train the system. Next, you need a huge neural network. Finally, you need a GPU (graphics processing unit) server to efficiently train that huge neural network, which means the barrier to entry is quite high. I think this is why, besides us, there are still very few companies offering this kind of service in Japan.

We designed all the modules that we envisaged would be needed for a company to be able to use the system for business applications straight away. We have created all the components required for the whole process: a user registration module, a module to receive videos, a module to visualize the real/fake detection results, and so on. If we didn't go to these lengths, even if we developed the core technology, nobody would end up using it.

—I think it's rare for researchers to be involved up until the stage of providing a service.

ECHIZEN: It's because the field of generative AI is moving so fast. It would not work if the researchers just focused on the core inference part and left the rest to someone else. We presented the world's first fake detection model for facial images in 2018. We were the first to see the need for such a system. We decided it would be quicker to make the whole thing ourselves from start to finish.

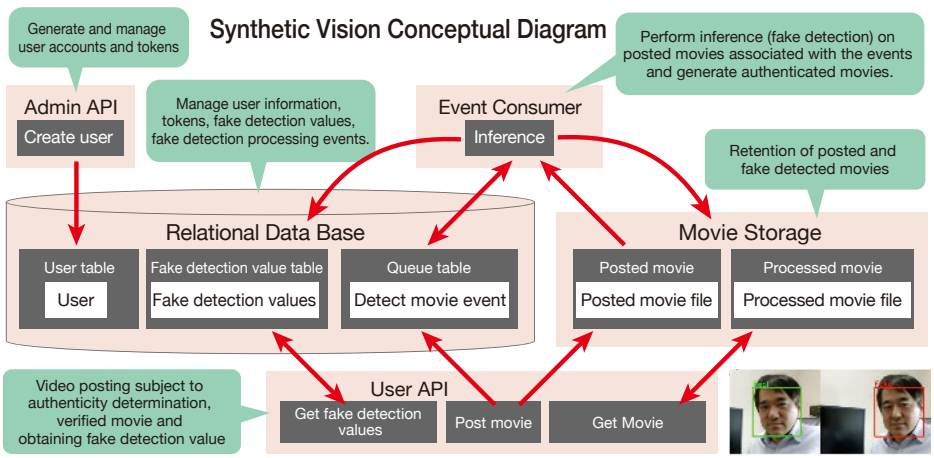
—Do you get a lot of inquiries from companies?

ECHIZEN: It has already been adopted and put into use by major digital advertising company CyberAgent Inc., and we have received many other inquiries. Since its launch in September 2021, the most common inquiry has been for deepfake detection in eKYC (Electronic Know Your Customer) for online identity verification systems, rather than identifying celebrity deepfakes. It seems that many users want to detect fraudulent applications in online verification.

YAMAGISHI: This service is not being provided by many companies yet. We are currently looking for a partner company to deliver our product to as many users as possible.

—What is the significance of the all-Japanese development team, with NII at the heart?

ECHIZEN: Now that ChatGPT is in the spotlight, there are growing calls for a Japanese large language model (LLM). The key point is that it must be made in Japan. If a foreign technology becomes the de facto standard, it will be controlled from outside Japan. The other party could decide to stop providing the service or narrow down the functions at their own convenience. Similarly, I think developing our own Japanese model for fake detection and strengthening our defense against fake information has huge benefits for the whole country. In our own small way, we will continue with our research and development with this as our mission.



A Word from the Interviewer

If misused, generative AI could undermine democracy and our electoral systems. It is vital to develop and maintain our own technical infrastructure to protect Japan from the threat of deepfakes. I hope the government will continue to promote research and development from a strategic perspective to ensure security and uphold democracy.

YAMADA, Tetsuro

Editorial Writer, The Yomiuri Shimbun

Joined the Yomiuri Shimbun in 1990 after graduating from The University of Tokyo. In 2006, he studied at the Massachusetts Institute of Technology (MIT) as a Knight Science Journalism research fellow. After working in the economics and science departments of the newspaper, and as a special correspondent at its Washington bureau, in 2018, he was appointed Science Department Chief. Since 2019, he has served as an editorial writer (on science and technology).

Photography by KOYAMA, Kosuke

"Cyber Vaccine" to Combat Infodemics

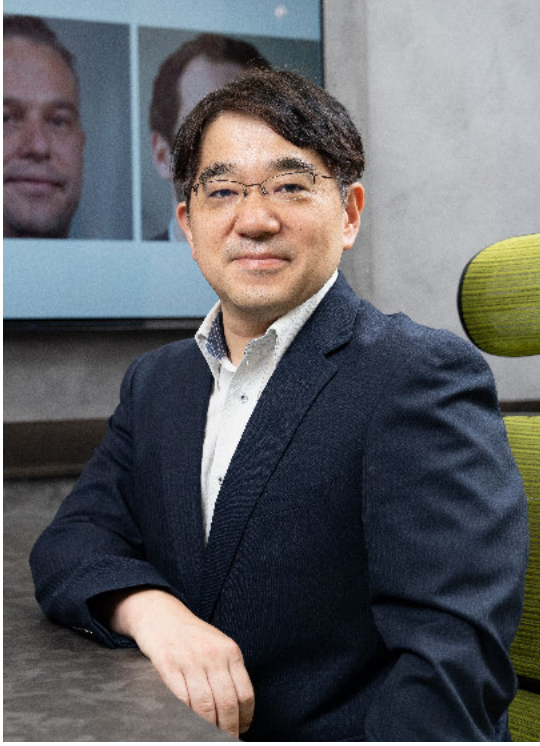
Professor, Information and Society Research Division, NII/ Director, Global Research Center for Synthetic Media

ECHIZEN, Isao

Reporting and writing

YAMADA, Tetsuro

Editorial Writer, The Yomiuri Shimbun



Dr. ECHIZEN, head of the Global Research Center for Synthetic Media, has been a pioneer in new areas of research into fake media since before the Center was established. Sometimes, thinking about methods of attack in an imaginative way can help to strengthen our defenses.

Rich imagination: an unusual talent in security

Tremendous amounts of information exist on the internet. Countless photographs and videos are shared on social media every day. In future, this information will be used – and misused – in ways we cannot even imagine.

In 2019, a man was arrested in Tokyo for stalking a female celebrity. He had compared scenes reflected in the victim's eyes in selfies that she had posted online with Google Street View to identify her local train station, and then followed her home from the station. This shows that images are being exploited in new ways, now that photos are so detailed that it is possible to enlarge a section to identify what is reflected in the eyes.

Dr. ECHIZEN realized a long time ago that as image quality improves, more information can be extracted. In 2018, he successfully used image processing to extract fingerprint patterns from photos of people holding up two fingers in a peace sign taken at a distance of three meters. Fingerprints are often used for biometric identification because they are unique to each individual, but this research demonstrated that, in principle, it could be possible to overcome such systems.

"Rather than collecting fingerprints left at a crime scene, I wanted to consider the threat of fingerprints being stolen remotely," says Dr. ECHIZEN. "It's something nobody had thought of before, so you could say it's a bit weird, but I am intrigued by such things."

Another interesting example of this incredible imagination is glasses to protect privacy. If your face happened to appear in a photo that someone else posted on social media, it could be compared against information online to identify you. Looking for ways to protect people against such inadvertent invasions of privacy, he developed in 2012 a pair of glasses equipped with near-infrared LEDs to prevent automatic facial recognition. An improved version that prevents facial recognition using reflection on the lens surface is now sold by a manufacturer in Sabae (Fukui) as "Privacy Visor."

"I think the ability to think of new threats that have not appeared yet is vital in the field of information security. People sometimes call me an unusual talent in security," laughs Dr. ECHIZEN.

"Cyber Vaccine" against fake facial images

The culmination of this pioneering research is SYNTHETIQ VISION, Japan's first fake facial image detection program, developed by the Global Research Center for Synthetic Media. The system analyzes facial images and indicates real faces with a green frame and AI-generated (fake) faces with a red frame, making it possible to tell at a glance whether an image is real or not.

The focus of research has now moved on from detecting fake images to developing a "Cyber Vaccine" against deepfakes. For example, information about facial features can be embedded into the edges of a video before it is published to "vaccinate" the image. It looks almost the same as a normal, unvaccinated video, but if a deepfake attacker attempts to replace the image with a different face, the vaccine is activated using the reconstruction model, so it will return to the original face, based on the facial information embedded in the periphery.

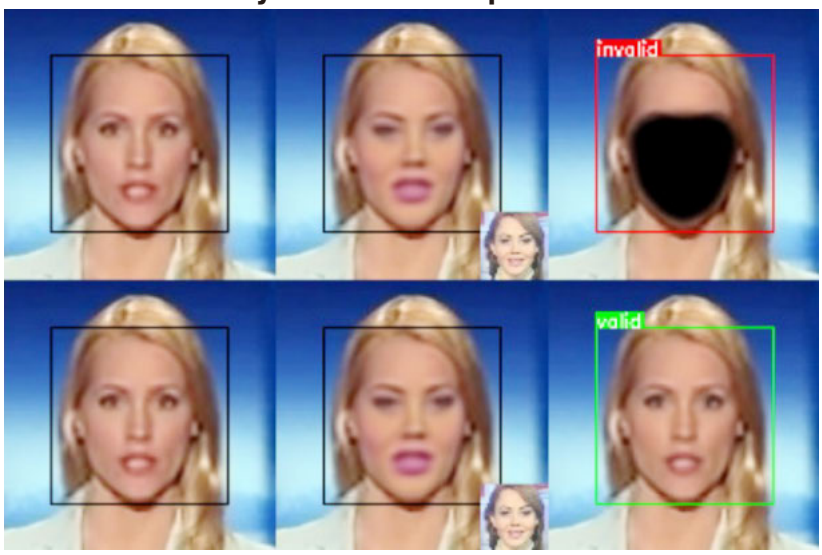
Dr. ECHIZEN and his team are also developing another type of vaccine to prevent online photos of people from being automatically collected without the subject's knowledge. By embedding invisible noise into people in photos, it makes it impossible to extract people from photos using "person segmentation" technology. "The Privacy Visor" is a real product to protect your face from facial recognition. This is a virtual version," explains Dr. ECHIZEN.

After the Russian invasion of Ukraine, a deepfake video was created which appeared to show Ukrainian President Zelenskyy telling his troops to surrender. There have been many cases of deepfakes such as celebrities' faces being inserted into pornographic videos.

The face is a piece of information that has a special meaning for humans. If a face is misused for purposes of political propaganda, manipulating public opinion, or slander, it can have very damaging effects. Research into vaccine technology has only just begun, but there are high hopes that vaccines could be developed to combat infodemics (global spread of misinformation), as well as pandemics.

Anticipating and guarding against unknown methods of attack

Cyber Vaccine: comparison



[Top] With an unvaccinated facial image (left), if the face is replaced by a deepfake attack (center), it cannot be restored to the original face using a reconstruction model (right).
[Bottom] With a vaccinated facial image (left), if the face is replaced (center), it can be restored to the original face using a reconstruction model (right).

The basis of a deterrent is to mitigate the enemy's attack in advance. Just like developing glasses to protect privacy, we always need to stay one step ahead of emerging threats. Dr. ECHIZEN believes that "the time is coming when we will need to be able to tell what is real and what is fake." This is what has motivated his research.

To a non-expert, it is difficult to conceive of unknown methods of attack, but Dr. ECHIZEN has plenty of ideas. The manager of a building can open all the doors as long as they have a master key. If a thief were to get hold of the master key, it would be a disaster. "Master Face" is exactly the kind of technology that Dr. ECHIZEN believes attackers could use. An artificial face with common features matching the features of many faces could be used as a "Master Face" to bypass facial recognition systems. No Master Face attacks have yet been reported, but this research is sure to prove useful when this threat becomes a reality.

These days, we are seeing more and more cases of spies hacking into the computers of rival companies or countries to steal information, and cyberattacks to disable IT systems. Having full knowledge of how the attackers work helps you to fix any holes in your defense network. Often, using technology in good ways and bad ways are two sides of the same coin.

Dr. ECHIZEN looks for unknown methods of attack and ways to guard against them. Attackers looking to use deepfake images certainly would not want him as an enemy.

Speech Synthesis Creates a Voice That Can Be Heard Over a Crowd

Professor, Digital Content and Media Sciences Research Division, NII/ Deputy Director, Global Research Center for Synthetic Media

YAMAGISHI, Junichi

Reporting and writing

YAMADA, Tetsuro

Editorial Writer, The Yomiuri Shimbun

Advances in speech synthesis technology have made it possible to produce natural synthetic voices that sound just like human voices, as well as voices with added value, such as voices resistant to noise. Dr. YAMAGISHI, an expert in speech synthesis and speech recognition using deep learning, tells us about the latest trends.

Speech synthesis is another area that has undergone a dramatic transformation since 2013 with the emergence of deep learning. "Vocorder" technology treats speech as data, allowing speech to be encoded and then reproduced from the encoded data. "Neural vocorders" use neural networks instead of signal processing, which has dramatically improved the audio quality.

Another innovation is a technique called "speaker vectors" to mathematically represent the characteristics of a speaker. It is very good at identifying and summarizing the characteristics that make you sound like you, meaning that AI can now accurately capture the characteristics of an individual's voice. "This was originally developed as an authentication technology, but by telling the system to synthesize a voice using these speaker vectors, it can accurately copy a person's voice," explains Dr. YAMAGISHI.

Speaker vectors also have the benefit of not requiring a lot of training data. Voice cloning technology, which allows a voice to be replicated in just a few minutes after recording and uploading a small sample, relies on speaker vectors.

It is no longer possible to tell the difference between an actual human voice and a synthesized imitation of that voice. It sounds as if the same person is repeating the same words. "Speech synthesis technology has basically been completed by around 2018," recalls Dr. YAMAGISHI.

Although speech synthesis has reached the point where it can produce natural speech that sounds exactly like the original voice, research has not stopped there. Realizing that "rather than just improving quality, we should think from a different perspective," Dr. YAMAGISHI started to think differently, aiming to improve clarity.

The result is a clear voice that can be heard above noise, which has been used for announcements on the Tokaido Shinkansen line since May 2023. This speech synthesis system allows content to be input as text, so there is no need for an announcer to record each announcement in a studio. The "clarity enhancement" effect enables passengers to hear announcements clearly over the noise of train vibrations, wind and rain, and crowds on a noisy station platform.

"AI storytellers" and speaker vectors

An intriguing area of research is the development of speech synthesis to perform the traditional Japanese art of rakugo comic storytelling. As well as producing natural speech, this involves developing a voice that can express subtle emotions and comic effects. With speech synthesis, the performance itself sounds natural enough, but it is rather hard to follow the story. This is because different characters of different genders appear in rakugo, and speech synthesis is not good at making a distinct voice for each role. At present, AI lacks the skill of a warm-up act, let alone a star performer.

Interestingly, it appears that a skilled star rakugo performer does not distinguish between characters in the story simply by changing his voice. “We have analyzed performances and we still don’t know, but the different roles might be expressed by subtle adjustments in rhythm or speed,” speculates Dr. YAMAGISHI. One day, I would love to see a scientific explanation of this mysterious skill that cannot be unlocked with speaker vector technology.

Music is actively created using the MIDI standard. Speech and music have a lot in common, so the same research can be applied. When the sound of a piano is synthesized by inputting MIDI data instead of text, the piano music plays naturally, but it occasionally skips or goes out of tune. Rather than music that has been synthesized by a machine, it sounds like the piano is being played by an unskilled player.

As the quality of speech synthesis improves and anyone can synthesize a voice at low cost, there are concerns about the technology being misused. This is already becoming an issue in countries like the US and India, where people routinely upload speech to social media. In 2019, the CEO of a company in the UK was tricked into transferring money by a phone call from the parent company’s CEO, which turned out to be an artful scam using a fake voice.

If a phone is set to be activated by a voice command like “Hey Siri,” a fake voice could be used to bypass the authentication to operate someone else’s phone.

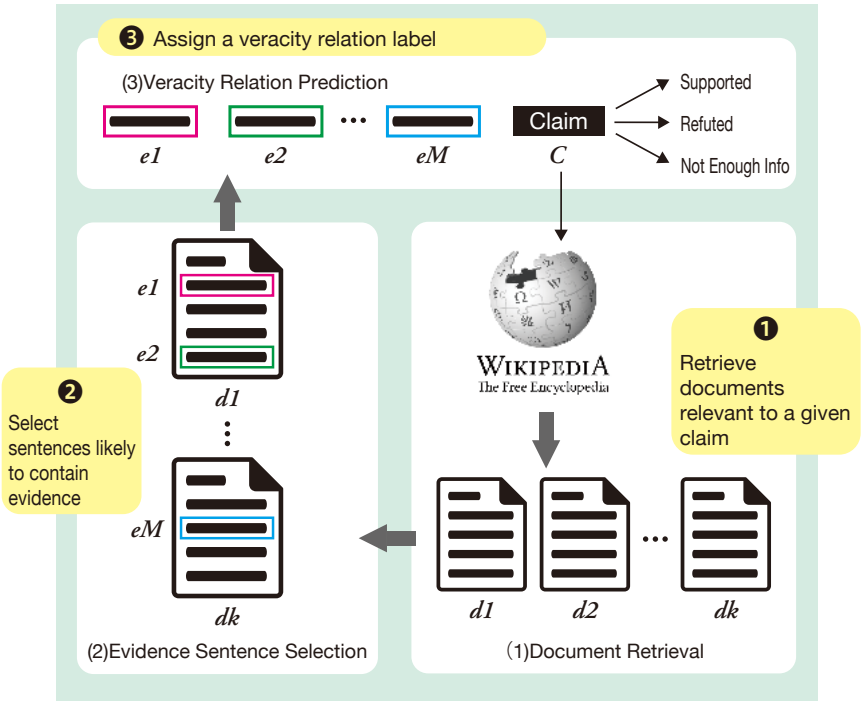
New areas to explore

Dr. YAMAGISHI is developing ways to defend against such attacks. A common benchmark is required when testing and comparing the performance of new detection technologies, and a large-scale database has now been created for this purpose. ASVspoof is a huge public database consisting of pairs of synthetic and natural speech samples that was created in cooperation with companies that include NTT and Google. Users can use the database to objectively evaluate the performance of a system by testing whether it can distinguish between natural speech and synthetic speech.

“It was very hard to create it, but it has been enthusiastically welcomed by people working in this field of research,” says Dr. YAMAGISHI. This work has made a significant contribution to establishing the intellectual infrastructure to enhance this area of research.

The scope of Dr. YAMAGISHI’s research has recently expanded to fact checking. It takes a lot of work to manually fact-check news reports and statements. The idea is to use AI to perform automatic verification based on information from a reliable database. It is often difficult to distinguish between fact and opinion, and there is no guarantee that a database is 100% true. Although there are many issues, if posts on Twitter (now called X), for example, could be automatically monitored 24 hours a day to warn users about fake news and propaganda, it would have a big impact. “As I have been researching deepfakes for many years, I have come to realize that we need to think about how to tell if content is real.” Dr. YAMAGISHI’s research seems to be leading away from speech research. I hope he will continue to explore new areas without hesitation.

Conceptual diagram of fact-checking system



Copyright and Privacy in the Era of Generative AI

Professor of Business Law,
Hitotsubashi University School of Law

IKEGAI, Naoto

Interviewer

IDA, Kanako

Editorial Writer,
Asahi Shimbun

Flexible legislation to allow innovation

Generative AI learns from existing data and creates new content. There are questions about the rights to the data used to train large language models (LLMs) and other deep learning models, and whether personal information is protected. Professor IKEGAI, Naoto of Hitotsubashi University, who specializes in information law and policy, explains the fast-moving debate.

—— As generative AI rapidly gains ground, I get the impression that the legal system is failing to keep up. What is the current situation?

The debate about the legal system surrounding generative AI was sparked by the launch of the AI chatbot "ChatGPT" in November 2022. The European Union (EU)'s proposed AI regulation is in the spotlight as the world's first comprehensive regulation of AI. The first draft did not take the existence of generative AI into account; new clauses specific to generative AI were added in 2023. The global debate has moved on rapidly in less than a year.

—— How about the situation in Japan?

The Personal Information Protection Commission published "Warnings for the use of generative AI services" in June 2023, and issued a statement asking OpenAI, USA, the operator of ChatGPT, to comply with Japan's personal information laws. The Japanese government, which is taking the lead in the Hiroshima AI Process agreed at the G7 Summit, has formed an AI strategy council and started discussing the issues in earnest. The government is at the stage of figuring out what kinds of rules are needed to respond to the various risks while fostering innovation by generative AI.

With regard to copyright, a proactive measure was taken. A revision to Japan's copyright law in 2018 includes a new Article 30-4 which permits the broad use of copyrighted works for machine learning, including generative AI, without the permission of the copyright holder.

The legal position of generative AI is a gray area

—— Why was that revision considered necessary at the time?

Discussions about legislation were stepped up around 2015, when the focus was on promoting AI development in general, based on deep learning. Generative AI was not taken into account at that time, but there seems to have been an awareness of the risk of copyright violation by AI-generated works. But of course, socially, the discussion did not go as far as the possibility of generative AI systems capable of fully replacing human creators, like ChatGPT and image generation AI "Stable Diffusion."

—— Do the new regulations mean that if a work is used for training purposes only, it is not subject to copyright protection?

As long as it is used for purposes of data analysis, rather than to en-

IKEGAI, Naoto

Completed doctoral course at the University of Tokyo Graduate School of Interdisciplinary Information Studies in 2012. Ph.D. in socio-information and communication studies. After working as a specially appointed researcher at the Center for Interdisciplinary Research, Research Institute of Information and Systems, joined the Department of Business Law at Hitotsubashi University in 2021 (current position). Specializes in information law and policy.

joy the thoughts or sentiments expressed in the work, in principle, it can be used without the **permission** of the copyright holder.

—— I believe there are exceptions, so it is rather difficult to understand.

One part that is particularly open to debate is that copyright restrictions do not apply if it would "unreasonably prejudice the interests of the copyright owner." The Agency for Cultural Affairs suggested in 2019 that this applies to using works in a way that affects the market or could potentially affect it in future. The example given was datasets sold for data analysis, but the status of generative AI is still a gray area.

What's more, Article 30-4 only applies to copyright restrictions at the training stage. Whether a work generated by AI infringes copyright is a separate question. When determining whether copyright has been infringed, the issues are similarity to and reliance on the existing copyrighted work. With a human creation, reliance can be judged by whether the creator saw the prior work; but with generative AI, it is debatable whether a creation can be said to "rely on" a copyrighted work included in the training data.

The latest generative AI systems can now perform various tasks to a high level. For example, if trained on a large quantity of artworks by a particular artist, AI can generate art in the style of that artist. Under what conditions would we accept that copyright has been infringed in a case like this? So far, there have been no cases leading to legal action in Japan, but I think the debate about these rules is heating up.

—— How will the interpretation of the regulations be decided?

If we just wait for a case to go to court, the unpredictable legal status will continue. The government and stakeholders could produce a set of guidelines. When Article 35 of the Copyright Act was revised in 2018 to allow copyrighted works to be used for digital education, education-related organizations and rights holder organizations held extensive discussions resulting in the publication of guidelines. But unlike that case, where functional guidelines were established by agreement between the two parties involved (the rights holders and the schools), in the case of AI, it is unclear which groups should discuss the issue.

There are discussions going on around the world between the side that develops and provides generative AI and the side that produces the content it is trained on, including the question of how the proceeds should be distributed. There are calls for legislation to resolve the debate, but for the time being, I think it is important for the parties to clarify the issues by "soft law" to see how problems should be resolved.

—— Rights holders may have doubts about copyright restrictions during the development and training stages of generative AI.

News media websites have always operated on the assumption that users click on links from search results to see information, which leads to advertising revenue and subscriptions. But with an architecture like ChatGPT, information is provided without the user needing to click. This must be upsetting the delicate balance of the existing media business model. We must make sure this does not result in nobody providing high-quality information. I think this needs to be considered as an issue affecting the foundations of democracy.

Who is responsible for AI-generated content?

—— If there is some issue with a work generated by generative AI, who would be held responsible?

There are various layers: from the provider of the AI foundation model, to the players who provide specific services incorporating this model, to general users. As well as those who transmit the information itself, the digital platform with which information is spread has a major role in guarding against illegal and harmful information online, not just copyright infringements. Similarly, for generative AI, each layer has a role to play, such as filtering or adjusting the training process so that the system will not generate problematic content. But in some cases, it is difficult for service providers to take measures for technical reasons, so I think the role of the foundation model provider should be emphasized.

The EU's AI regulation clearly assigns responsibility to the foundation model provider, and the EU

aims to introduce regulations in line with its technical characteristics.

—— There are also concerns about whether personal information is properly protected by generative AI, which handles huge amounts of data.

When companies or universities use data in the course of business, they must comply with the Personal Information Protection Act. If personal data entered as part of a prompt can be used for machine learning on the generative AI side, for example, this would be classed as third-party provision of personal data, which might require the consent of the person in question. Just like using regular cloud services in business, it is necessary to check the terms of service when using generative AI. We also need to be aware of the possibility of personal information being used for unintended purposes.

General users should decide whether to use AI with an awareness of the risk of the personal data they enter being used for machine learning, and avoid entering highly private information if they do decide to use AI. There is a possibility of sensitive information about yourself or other people being reflected in generative AI output. Users should be aware of this as a matter of AI literacy.

Researchers must consider ethics

—— Generative AI could produce information that promotes discrimination or prejudice, incorrect information, or false information.

How much the law should get involved with these kinds of issues is a difficult question, as addressing them needs to be balanced against freedom of expression.

As a provider of a service with considerable social influence, I think the focus should be on the provider's responsibility not to cause social disorder by spreading incorrect or false information. But it is not practical or desirable for the law to define specific requirements like "use this technology" or "do not output this keyword." Perhaps a desirable system would be to create incentives for providers to voluntarily take measures like ongoing action to deal with inappropriate AI-generated content, responding properly to behavior that violates the terms of service, and training the model in a way that avoids prejudice as much as possible. At the same time, society should monitor and verify whether these initiatives are sufficient.

The EU's proposed AI regulation will require the foundation model provider to "identify and mitigate reasonably foreseeable risks to health, safety, fundamental rights, the environment, democracy, and the rule of law." Flexible legislation that allows innovation while ensuring businesses fulfill their social responsibilities is required.

—— How will the fairness and impartiality of technology be ensured?

Globally, an increasingly small number of players provide the majority of fundamental technologies and services. This could become a new source of power. We need regulations to stop services that affect the whole of society from being opaquely and arbitrarily shaped by certain players.

A limited number of companies have access to huge datasets, so intervention may be required in terms of competition policy.

—At the UN Security Council's meeting on AI, concerns were expressed about states using generative AI for censorship or repression.

Among the G7 nations, there is a consensus that generative AI should be used in line with democratic values. But the issue of how new technologies will be used in developing countries or those with authoritarian regimes needs to be discussed on an international level.

The Council of Europe, an international organization made up of 46 member states, has been working on drawing up an AI treaty for a few years now. This is the first move towards an internationally binding treaty.

I think there will be questions about how to involve countries other than

[Copyright Act]

(Exploitation without the Purpose of Enjoying the Thoughts or Sentiments Expressed in a Work)

Article 30-4: It is permissible to exploit a work, in any way and to the extent considered necessary, in any of the following cases, or in any other case in which it is not a person's purpose to personally enjoy or cause another person to enjoy the thoughts or sentiments expressed in that work – provided, however, that this does not apply if the action would unreasonably prejudice the interests of the copyright owner in light of the nature or purpose of the work or the circumstances of its exploitation:

(i) if it is done for use in testing to develop or put into practical use technology that is connected with the recording of sounds or visuals of a work or other such exploitation;

(ii) if it is done for use in data analysis (meaning the extraction, comparison, classification, or other statistical analysis of the constituent language, sounds, images, or other elemental data from a large number of works or a large volume of other such data; the same applies in Article 47-5, paragraph (1), item (ii));

(iii) if it is exploited in the course of computer data processing or otherwise exploited in a way that does not involve what is expressed in the work being perceived by the human senses (for works of computer programming, such exploitation excludes the execution of the work on a computer), beyond as set forth in the preceding two items.

We need to decide what form international rules should take

Western advanced nations in the international framework, including the Hiroshima AI Process.

On the other hand, as there is still no global treaty on personal information protection, this is an issue where there are more than a few fundamental differences in values, even between the USA and Europe. It will naturally take time to reach a broad consensus. As technology is progressing at a rapid pace, nobody knows which approach is correct. Rather than rushing to cement detailed rules in a treaty, perhaps we need to take time to see what works in each country before deciding what form international rules should take.

—— It has been suggested that there should be an international monitoring organization for AI like the International Atomic Energy Agency (IAEA).

It would not need to be as solid a framework as for atomic energy, but it would be worthwhile to form a framework to stop AI from being used for military purposes or disinformation warfare, and to monitor the situation in each country and enable the international community to take appropriate action.

—— What should researchers bear in mind if they are currently involved in developing generative AI?

One suggestion is to follow guidelines such as the NII's "Handbook of Data Management for Open Science"*1, as the rules for personal information protection in academic research have changed, which includes the use of personal information at the training stage of generative AI. Another point is that those who are involved in developing generative AI should think about the ethical considerations and how to provide technology of generative AI, as part of the process of putting rules into place. For generative AI to be widely accepted and utilized by society, I believe it is vital for researchers themselves to discuss ethics going beyond the legal system.

*1 Handbook of Data Management for Open Science
https://www.nii.ac.jp/service/handbook/
(Japanese)

A Word from the Interviewer

What merits and demerits will generative AI bring to society? While this is unknown, the path to legislation and international cooperation looks steep, but there are many clues to the way ahead, such as the EU's proposed regulation. AI seems terrifying because it does not innately have the awareness of rules and ethics that we can expect from a human. If used in the wrong way, it could lead to losing the values of fairness and respect for human rights that humanity has built. There is a growing need for transparency from those developing and providing the technology, and an open discussion including both providers and users.

IDA, Kanako
Editorial Writer, Asahi Shimbun

Joined Asahi Shimbun after graduating from the University of Tokyo's Department of Social Psychology. While working, gained a Master's degree (socio-information and communication studies) from the University of Tokyo Graduate School of Interdisciplinary Information Studies in 2012. At Asahi Shimbun, has worked in the Society division and as head of the Brussels Bureau, and is currently responsible for editorials on legal issues. Published articles include "Discussion and reporting of the lay judge system" (Journal of Mass Communication Studies, Volume 82).

Photography by KOYAMA, Kosuke

NII NEWS TOPICS

Period: May 1 (Mon.) to July 31 (Mon.), 2023

More details about news items are available online.

www.nii.ac.jp/news/2023

NEWS RELEASE 2023

Jun.12 NII, with other organizations, provides an open database of teaching materials on quantum technology: Collaboration with Kyushu University, Keio University, Nagoya University, and the University of Tokyo to promote training in quantum technology

May 24 Recruiting a licensing partner for NII's automatic fake facial image detection program SYNTHETIQ VISION: Looking for a partner company to deliver NII's latest AI research results to society

May 23 Series of three keynote speeches on generative AI at NII Open House June 2 (Fri): Introducing the latest fake media detection technology and other latest research

May 15 June 3 (Sat) Learn programming thinking at NII: Computer Science Park in Chiyoda-ku, Tokyo (in-person event)

May 2 NII Weeks 2023 will introduce NII's work to a wide audience. NII Open Forum on Informatics, NII Open House, and Japan Open Science Summit to be held consecutively

AWARDS 2023

July 31 A paper by Prof. ECHIZEN, Isao (Information and Society Research Division), Prof. YAMAGISHI, Junichi (Digital Content and Media Sciences Research Division), Assistant Prof. by Special Appointment NGUYEN Hong Huy (Information and Society Research Division), and others awarded the 2022 IEICE Information and Systems Society Paper Award

July 23 A paper by Prof. KANDO, Noriko (Information and Society Research Division) and others was awarded the Best Paper Award at ICTIR 2023.

July 21 A paper by KORI, Mayuko (NII Research Assistant & SOKENDAI Informatics Course, Hasuo Lab), Prof. HASUO, Ichiro (Information Systems Architecture Science Research Division), and others awarded the CAV Distinguished Paper Award at CAV 2023

July 13 A paper by Prof. INOUE, Katsumi (Principles of Informatics Research Division) and others awarded the 10-Year Test-of-Time Award at ICLP 2023

July 1 KATO, Kanji (Researcher by Special Appointment, Center for Open Data in the Humanities, ROIS-DS) awarded the 2023 Nakasone Seizen Memorial Research Encouragement Award

June 8 A paper by Assistant Prof. by Special Appointment NGUYEN Hong Huy (Information and Society Research Division), Prof. YAMAGISHI, Junichi (Digital Content and Media Sciences Research Division), Prof. ECHIZEN, Isao (Information and Society Research Division), and others awarded the 2022 IEICE Paper Award

May 14 Associate Prof. by Special Appointment, Paolo ARCAINI (Information Systems Architecture Science Research Division) and Former Associate Prof. by Special Appointment Ahmet CETINKAYA awarded first place in the Cyber-Physical Systems (CPS) Tool Competition at SBFT 2023

May 8 An article by Prof. ECHIZEN, Isao (Information and Society Research Division) and others awarded the 2022 Best Article Award by the Journal of the Institute of Image Information and Television Engineers

Facebook
https://www.facebook.com/jouhouken/

X (Twitter)
https://twitter.com/jouhouken

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

NII mascot Bit on X (Twitter)
https://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

Events held by NII this year. Videos of some of the lectures can be viewed on the NII website.

Event report 1

NII Open Forum on Informatics 2023

Dates: May 29 (Mon) to 31 (Wed), 2023
Hybrid event
https://www.nii.ac.jp/openforum/2023/

This event is held annually to promote our academic research platforms, including the SINET6 network and NII RDC (research data cloud), as well as being a forum to exchange opinions with universities and other institutions with a focus on further development. This year it was held as a hybrid event for the first time, with lectures presented at the event venue (Hitotsubashi Lecture Hall, Hitotsubashi University) or online. Over 4,800 people took part.

Event report 2

NII Open House 2023

Dates: June 2 (Fri) and 3 (Sat), 2023
In-person event (Some content streamed online)
https://www.nii.ac.jp/openhouse

NII Open House is held annually to present our research results to the public. This year's event was held in Hitotsubashi, Chiyoda-ku for the first time in four years. Despite bad weather due to an approaching typhoon, many people attended in person. Over 1,200 people took part, including online viewers. Featuring demos and poster sessions by researchers about their research results, a keynote speech on the theme of ChatGPT, an industry-government-academia collaboration seminar, and our Computer Science Park allowing visitors to have fun playing with math and programming, the event was a great success.

Top: Researchers present their research results at poster sessions.
Middle: Robot island (left) and Math island (right) at the Computer Science Park.
Bottom: Demo of a robot dog at the poster session and visitors at the venue.

[Essay]

Bad Fakes and Good Fakes in Financial Markets

MIZUNO, Takayuki

Associate Professor, Information and Society Research Division, NII

Aside from fundamental fluctuations, financial markets can basically be considered as a zero-sum game. In this competitive environment, investors are constantly looking for strategies to gain an advantage over others, wanting to be the first to get hold of information. Poring over satellite images to gauge oil reserves, reading local newspapers in foreign languages, monitoring shipments from factories… every second counts in the quest for fresh information. But what if you find some juicy information on social media? Do you first check if it is true, or check how far the information has spread? For the mass media, you would probably check whether it is true, but investors are not all that concerned with authenticity. What's more important is how many other investors have seen that information, and being the first to figure out whether to sell or buy stocks.

On the morning of May 22, 2023, an image was posted on Twitter appearing to show an explosion near the Pentagon. The image turned out to be an AI-generated fake. The Russian state propaganda agency RT shared this information with its 3 million followers, and several business news digest accounts then shared it with millions of followers. These accounts included fake news organizations with Twitter Blue status (originally a check mark indicating verified accounts, but now a paid subscription service). As well as being shared on social media, the "news" was even reported on an Indian TV channel. This apparently spooked some investors, causing a sudden drop in US stock prices within minutes.

Investors often act before checking the authenticity of information. If the information later turns out to be fake, the market quickly adjusts. But if insider information, which can be difficult to prove true or false, is presented alongside fake information, there is a risk of confusion in the market. To respond to this, as well as verifying whether information is true, we need to understand how it is spread, how it is interpreted, and what actions it will lead to.

Next, I would like to give an example of "good fakes." Financial models are often criticized for producing inaccurate predictions, but artificially generated or "fake" data could provide a solution to this problem. Stock prices fluctuate so wildly that "two years is a long time in finance." This means that if long-term stock price data are used in an attempt to make a model more robust, the model will end up being adapted to past information. Conversely, if only recent data are used to create a model adapted to the latest information, the model will be less robust, due to insufficient data.

That's why generating pseudo stock price time series data using the GAN algorithm has gained ground among the ICAIF (ACM International Conference on AI in Finance) community. This technique involves using fakes to boost the amount of data. By replacing expensive financial data with fake data, this also removes some of the barriers to model development.

Stock price data have complex properties, like long-term memory and long tail distribution, that are difficult to reproduce with regular time series models. This has so far made it difficult to produce satisfactory fakes. Advances in generative AI technology might help somehow, although the mechanism is unknown. But this would require a huge amount of training data that do not currently exist in the world, so there is still a long way to go.

For better or worse, fakes will provide an opportunity to change financial markets. The rules of financial transactions have hardly changed since the 19th century, and various cracks are starting to appear in the system, not just the issue of bad fakes. What we need is for an innovative financial system to emerge: one that guards against manipulation by bad fake information, without losing the element of speculation.

Weaving Information into Knowledge
NII

National Institute of Informatics News [NII Today] No. 100 Sep. 2023
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
E-mail : kouhou@nii.ac.jp

Publisher: KUROHASHI, Sadao
Editorial Committee Chair: ECHIZEN Isao
Editorial Committee member: IKEHATA, Satoshi; KANEKO, Megumi; KOMIYAMA, Yusuke; TAKEFUSA, Atsuko; MIZUNO, Takayuki
Production: Tech Venture Research Institute; KAJIWARA, Maiko
Design: FROG KING STUDIO
Cover illustration: ICHIMURA, Joe