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

NII Today

67
Sep. 2018
National Institute of Informatics News

NII Interview

Tackling Uncertainty in Machine Learning

Fuyuki Ishikawa [Associate Professor, Information Systems Architecture Science Research Division / Organizer, Workshop on Machine Learning Systems Engineering]

One Research Paper Will Change the World Tackling Problems Through Domestic and International Collaboration

Masashi Sugiyama [Director, RIKEN Center for Advanced Intelligence Project]

Identifying the Nature of Machine Learning to Protect Against Attacks and Faults

Shin Nakajima [Professor, Information and Society Research Division]

Dialog

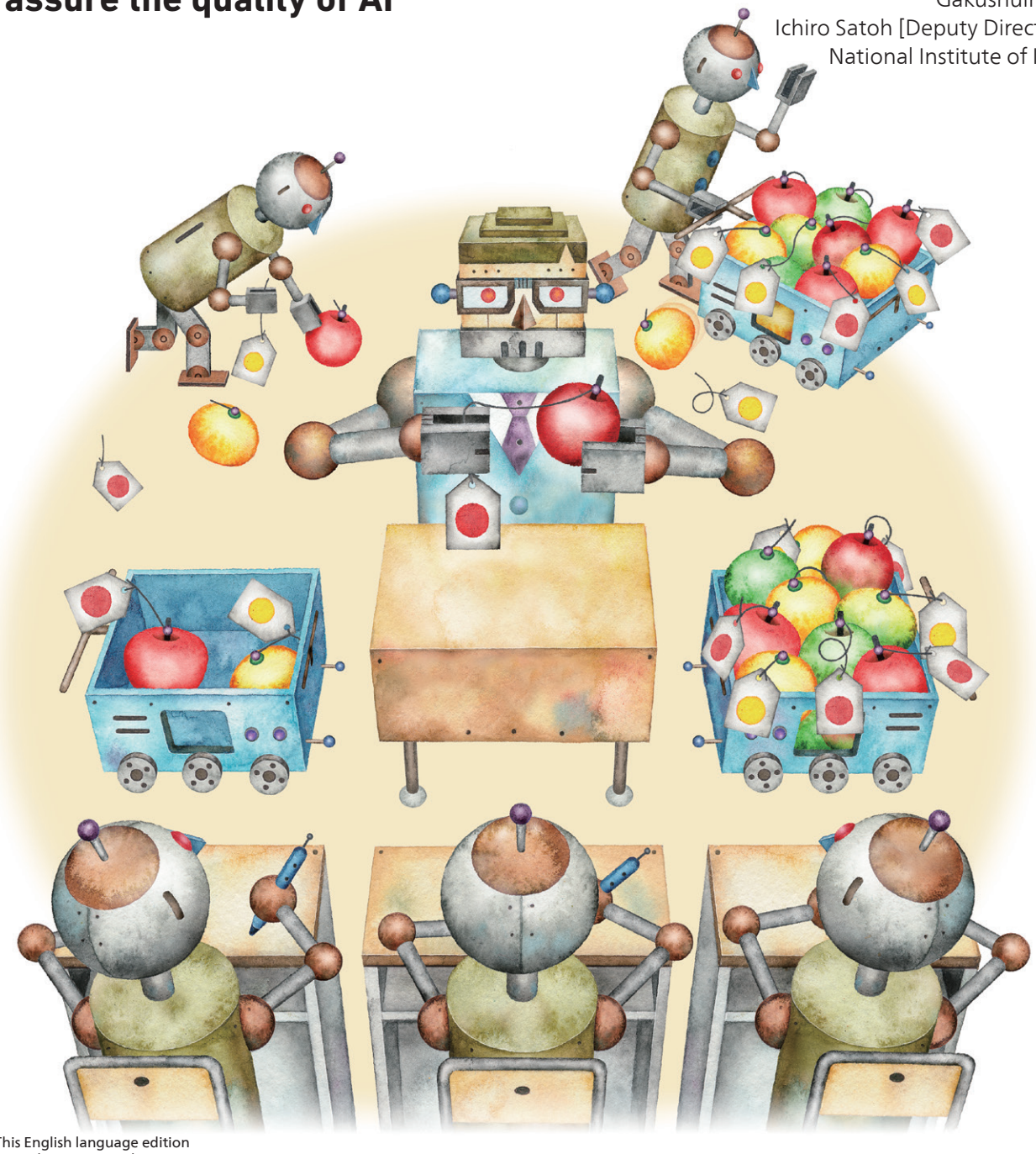
Quality of Machine Learning from a Legal Perspective

Souichirou Kozuka [Professor, Faculty of Law, Gakushuin University]
Ichiro Satoh [Deputy Director General, National Institute of Informatics]

Feature

New Software Engineering for Machine Learning

How to assure the quality of AI



This English language edition
NII Today corresponds to
No. 81 of the Japanese edition

Tackling Uncertainty in Machine Learning

Toward utilization of AI products and services in society

Fuyuki Ishikawa Associate Professor, Information Systems Architecture Science Research Division, National Institute of Informatics / Visiting Associate Professor, Graduate School of Informatics and Engineering, University of Electro-Communications / Visiting Associate Professor, School of Multidisciplinary Sciences, Graduate University for Advanced Studies

Interviewer: Keiichi Murayama Nikkei commentator

Interest in products and services using artificial intelligence (AI) is growing. Machine learning technologies are advancing, and there are great hopes for the new value that data will create. However, because of the "uncertainty" inherent in machine learning, engineers working in corporations at the cutting edge of system development are facing problems never encountered before. The Special Interest Group (SIG) on Machine Learning Systems Engineering (MLSE) in Japan Society for Software Science and Technology was launched to provide an opportunity for discussions aimed at resolving these problems. Associate Professor Fuyuki Ishikawa of NII played a central role in launching this SIG, and I asked him about its background.

Murayama Interest in AI is increasing in society as a whole, and there are growing expectations for the creation of new businesses in the industrial sector and elsewhere.

Ishikawa We can do amazing things with machine learning and new business opportunities will arise, as has been pointed out. Technology libraries, frameworks, and other tools have appeared. Options are increasing, and the functionality of these tools is expanding. Using these technologies, it is becoming easier to build systems employing machine learning. That is the background to the current boom in AI, but from an engineer's standpoint, it is not all dreams and magic. Many of the engineers that I regularly come into contact with at companies have concerns.

Fuyuki Ishikawa

Murayama What kinds of concerns?

Ishikawa Building some sort of system has certainly become easier. However, putting that system on the market as a product and assuming product liability are a different story. Up until now, there were certain conventions in software development, and there were standards that dictated what had to be done so that the software would be accepted by society and so that we could be confident that we had done a good job, but that is all up in the air at the moment.

Conventional software development was deductive: people wrote rules according to the purpose of the system and programmed those rules. However, creating machine learning systems from training data can be described as inductive software construction. This is an indirect method in which, rather than the system working according to rules determined by people, people provide the system with lots of data and let it make the rules. The system is a so-called black box, which cannot be controlled directly. It could betray people's expectations. For engineers, this means that although they have created the system themselves, they cannot say for sure what it can or cannot do, because they do not know until they try something.

Whether it was calculating taxes or recording employee information, up until now, the system developer was able to enter into a contract with their client by discussing the value of the system and working out the cost from there. But in the case of machine learning, it is impossible to determine how accurate the system will be until it is built. Conventional methods of deciding what will be built through discussions with the client are

not applicable.

It was with the hope of doing something about this situation that the MLSE group was launched. As facilitators, we intend to provide a "forum" by hosting study sessions and other activities. We expect that engineers and researchers working in corporations will make up the majority of participants. We also intend to introduce research from overseas and put together feature articles for magazines.

Murayama Meanwhile, various systems are already being developed using machine learning. Are engineers working on development with mixed feelings?

Ishikawa Often, a trial version is made in a proof-of-concept (PoC) phase first, and if that goes well, then the development moves into the actual project phase, but this is also problematic. In some cases, too much effort is devoted to PoC. A characteristic of inductive software engineering is that 100% correct answers cannot be guaranteed. This means that the client can later point out that the software is not adequately accurate or that its capabilities are insufficient, which is a source of worry for engineers and contractors.

The data used in tests in the PoC phase are trial data, and there are limits to what can be done with such data. A system that works well with trial data will not necessarily work well with actual data. Some of the worries of engineers originate after the PoC success. This problem is why it is important to create tools for proper discussions with the client, establish techniques for managing information, and so on.

Murayama We need a social consensus to build systems, with the understanding that machine learning cannot be 100% accurate.

Ishikawa That is correct. Of course, engineers are doing their best in development and are endeavoring to fill in the gaps in the technology. However, additionally, it is important that system users and the people who fund development understand that this is how machine learning is. Nevertheless, there are many extremely useful ways of using machine learning. The reality that we face is that nothing is risk-free, so I think that we need to discuss this calmly and everyone needs to accept the true situation.

For example, it is great that car manufacturers are working to make self-driving cars a reality, but this must include building consensus about machine learning. Without that consensus, accidents, when they occur, could result in overly negative reactions and excessive regulation. It is important to create an environment that allows engineers to work on development without unreasonable worries.

Murayama After considering the risks, society must advance in a direction that maximizes the benefits of machine learning.

Ishikawa The most amazing thing about machine learning is that systems can do vaguely defined things that cannot be put into words. Until now, we had to express our requirements in words and define what we were creating as rules, but the things that people really want to achieve tend to be quite nebulous. These kinds of obscure things can be done by a system using machine learning.

Engineers are enjoying their jobs, despite worrying about new issues, since creating new things, producing something through trial and error, is exciting. I want us to resolve the issues by using the power of technological development and the power of consensus building. Through the MLSE activities, I hope that we will enliven the field, while being aware of the problems.

Murayama Research and development in AI is being spearheaded by the United States with its many influential IT corporations, while China is closing in on the them. If, with the top two countries being the United States and China, Japan is forgotten, will this pose a problem in terms of Japan's competitiveness and security?

Ishikawa What we, as researchers, have to do is create entire systems containing machine learning well. While combining inductive and deductive reasoning, we should ensure security for highly critical parts with the rule base. By taking the best from both approaches, the ultimate goal is to raise the overall level of completion. Machine learning has reached the stage of practical application, and the next step is seeking to combine it with conventional methods.

Great things can be achieved when things, machine learning, conventional systems, and software engage. That is what we are really aiming for. For example, computer operating systems are all made by American companies, but when it comes to the computers themselves, Japan has the advantage in manufacturing. There are many people who are able to compete on a global level. Hopefully, the entry of AI into this field will make it even stronger. Many companies within the manufacturing industry are experimenting with machine learning. I hope that Japan's strength will be demonstrated by our collective capabilities.

(Photography by Yusuke Sato)

A Word from the Interviewer

Machine learning—with machines learning from vast amounts of data—is revolutionary in the sense that it has expanded the possibilities of systemization to fields where that was once difficult. Understandably, society is watching AI intently, but the systems that are created will certainly not be all-powerful and they will have the potential to make mistakes. The question now is whether system creators and system users will have the wisdom to build trusting, understanding relationships.

Keiichi Murayama Graduated from the School of Law, Tohoku University, in 1992 and joined Nikkei Inc. in the same year. In the Industry Department, covered areas including IT/electronics, automotive, and medical news. After studying abroad at Harvard University and working at the US Silicon Valley Bureau, became Senior Staff Writer in 2012. Also served as Editorial Writer from 2015. Assumed his present post in 2017. Responsible for IT and startups. Most recent book is *Startup kigyoka no riaru* (The Real Lives of Startup Entrepreneurs).

One Research Paper Will Change the World Tackling Problems Through Domestic and International Collaboration

Fulfilling our responsibility for quality as researchers involved in methodology

Masashi Sugiyama Director, RIKEN Center for Advanced Intelligence Project (AIP Center)
Professor, Department of Complexity Science and Engineering,
Graduate School of Frontier Sciences, University of Tokyo

Interviewer: Nobuyuki Yajima Senior Researcher, Nikkei BP Intelligence Group, Nikkei Business Publications, Inc.

"The machine learning that some researchers have been working on is in the spotlight, and its application in various fields is anticipated. We researchers involved in the foundations of machine learning are very pleased. We do feel some pressure due to issues such as safety, but in this field, we can do research that will change the world." So says Masashi Sugiyama, University of Tokyo graduate school professor and leading researcher in the field of machine learning. At the RIKEN Center for Advanced Intelligence Project, where Professor Sugiyama serves as Director, some 650 researchers collaborate with organizations in Japan and abroad to carry out research with awareness of applications and problems, including methodologies not related to machine learning. I asked Professor Sugiyama about the current situation, issues, and the role of researchers.

The world has changed completely

In the communities where machine learning researchers gather, we sometimes talk about how the world has changed. I have been engaged in machine learning research for about 20 years, and because it was an unfashionable field, I was able to do whatever research I wanted exactly as I pleased. However, then machine learning using deep models, so-called deep learning, became hugely popular and began to be applied throughout society. This was widely reported by the media, and we researchers started to be approached by many people. Algorithms invented by researchers are being used in society. It is a heavy responsibility. We say to each other that we cannot turn the clock back to the good old days.

Major global corporations now invest huge amounts of research money in the field, but there are still many problems that cannot be solved. Extension of machine learning and deep learning is necessary, and completely new methods are possible, so the field is still challenging and rewarding for researchers.

I am working on algorithms for solving problems through statistical processing of collected data by computer. I have invented a general-purpose algorithm that can be used for multiple problems, and an algorithm that allows machines to learn from incomplete data. Meanwhile, some researchers are working on the mathematical theory behind algorithms. The underlying methodologies are being researched by all of us.

The creation of excellent methodologies depends on the ideas of researchers, so the one research paper that changes the world

Masashi Sugiyama

Received a PhD (engineering) from the Graduate School of Information Science and Engineering, Tokyo Institute of Technology in 2001, and became a research assistant there in the same year. Became an Associate Professor at the same university in 2003. Professor at the University of Tokyo since 2014. Has served concurrently as Director of the RIKEN Center for Advanced Intelligence Project (AIP Center) since 2016. Work includes theoretical research on machine learning and data mining, and development of algorithms, as well as their application to signal processing, image processing, robot control, etc.

could come out of Japan. After all, the surge in popularity of deep learning came about because of a single research paper published in 2006.

Tackling problems that no one else is researching

My role as Director of the Center for Advanced Intelligence Project (AIP Center) involves finding researchers, establishing partnerships with universities overseas, and building relationships with companies for collaborative research.

The AIP Center currently has 150 full-time and 500 part-time researchers. They are not just researching machine learning; but also engaged in research closely related to application, as well as methodology.

At the AIP Center, researchers have 10 years during which they can do careful, in-depth research. This research environment is unequaled anywhere else in the world. To researchers doing work on the fundamentals, we say that we do not mind if it does not make any sense to anybody else; we just find problems that cannot be solved and attempt to solve them and achieve a successful result in 10 years' time. There are some fun areas of fundamental research, such as image processing and automatic translation, that we are purposely avoiding because major global corporations are frantically working on them.

The reason why we do not mind if the research is incomprehensible to other people is that it is important to find problems that make people ask, "Why work on that problem?" We hope that while a researcher is grappling with a problem that they really do not know how to solve, they will arrive at a new methodology.

The researchers who are doing work related to application are engaged in research that supports other scientific fields, such as medicine and materials science, as well as research focused on social problems such as disaster resilience, social infrastructure management, and health care for the elderly. Although their research is related to application, the researchers at the AIP Center are information researchers, so they collaborate with researchers and companies in medicine or materials, and partners working on social problems.

The issue is robustness

When one carries out research as an individual researcher and sees a huge amount of research as the Director of the AIP Center, the issues come into clear focus. The hottest topic among researchers at the moment is *robustness*. How do we provide methodologies that are robust against change? It is a matter of considering safety, or quality.

When we do research, we abstract and model a certain problem and think about an algorithm to solve it. In other words, we provide a neat solution using neat assumptions. However, in the world of application, assumptions soon change. Therefore, it becomes necessary to provide algorithms that can produce answers even with a weak model where everything is not neatly decided, and it is assumed that the assumptions will waver.

There are also issues of fairness and transparency. When allowing a computer to make a decision based on past data, if the decisions made by humans in the past were not particularly fair, then the computer could make an unfair decision in the future. Fairness is not mentioned much in Japan, but it is given a great deal of importance in Europe and the United States. Transparency means making it clear why decisions were made.

There are various issues, but as a researcher involved in methodology, I believe that it would be good to understand the universal core. In fact, with the methods of understanding the core, I have been working on the problem that it is difficult to implement algorithms and good data are not collected. This problem came up in the implementation stage, prior to safety.

With regard to implementation, I prepared a framework of multiple algorithms with a certain degree of universality using an approach called "probability density ratio estimation," and I got some companies to use the framework. Implementing a specific algorithm for a specific problem takes time and money, as engineers have to study the algorithm in each case.

In terms of data collection, I am researching algorithms that can make decisions from incomplete data and "small data." We are said to be in an era of big data, but that is only really true in the field of Internet services. In other fields such as manufacturing, medicine, and social infrastructure, everyone is struggling to collect data.

When thinking about robustness, methodology research alone is not enough. Lawyers and other legal experts come to the AIP Center, and they research the ethical and legal issues that arise when AI is used in society. It is also necessary to build closer collaborative relationships with researchers who are doing work related to application.

The other day, I was invited to an MLSE meeting (see Pages 02, 03), and I was struck by the enthusiasm of the people there. Methodology is important, but it is after all algorithms, so software engineers are essential for its implementation. When implemented, the engineers discover various problems, and we re-examine the algorithms and theories.

There are few software engineering researchers at the AIP Center, so we are grateful to be invited by people in software engineering. I hope that we will be able to contribute to software engineering too.

(Photography by Yusuke Sato)

A Word from the Interviewer

Working as both a researcher and the director of an organization is tough. While continuing to conduct research, Professor Sugiyama compiles his results into books in English and flies all over the world looking for talented people. Ever since the days when he was just a researcher, he has been enthusiastic about collaborating and proposing his methodologies to corporations and other organizations. It is perhaps this attitude that makes him well suited to the role of Director. I wish him success both as a researcher and as a center Director.

Nobuyuki Yajima

Born in 1960. Studied mathematics at university with the aim of becoming a computer engineer, but joined Nikkei McGraw-Hill, Inc. (now, Nikkei Business Publications, Inc.) in 1985 and became a reporter for *the Nikkei Computer* magazine. Worked as editor-in-chief *of Nikkei Computer* from 2009. Assumed his present post in 2016.

Identifying the Nature of Machine Learning to Protect Against Attacks and Faults

An approach considering product quality, service quality, and platform quality

Shin Nakajima Professor, Information and Society Research Division, National Institute of Informatics
Professor, School of Multidisciplinary Sciences, Graduate University for Advanced Studies

Interviewer: Masako Wakae Senior Staff Writer, Yomiuri Shimbun

Recent years have seen a succession of reports on how artificial intelligence (AI) can be deceived. AI appears to be surprisingly easy to fool. Examples include an image of a panda that is falsely recognized as a gibbon when noise is added to the image, and a stop sign that is recognized as a speed limit sign when graffiti is placed on the sign. If AI permeates our lives in the future, then its reliability will be a matter of life and death. However, we do not really understand why AI makes mistakes. Then what can we do? I asked Professor Shin Nakajima, who researches quality issues in one of the techniques used in AI, machine learning.

Wakae There have been many reports recently about research on methods of attacking machine learning.

Nakajima Various studies were reported after a research group in Canada announced a method of attack called “adversarial examples” in 2014. This method causes trained learning models to mistakenly recognize images by adding a particular type of noise to the images when giving data to the model that makes inference. A famous example was published by the same research group in 2015. In this example, noise was added to an image of a panda, and although it still looked like a panda to the human eye, the inference program concluded that it was a gibbon. The road sign case published by researchers in the U.S.A. in 2017 is also well known. Black graffiti on a stop sign caused it to be mistaken for a speed limit sign.

Shin Nakajima

Wakae That is a matter of life and death. Why does this kind of thing happen?

Nakajima In these two studies, there was a problem with the deep neural networks, which is one type of machine learning, and recognition errors occurred when data of a certain kind were given. Actually, though, the cause of mistaken recognition often cannot be fully explained. There are further cases where machine learning software systems make implausible inferences.

Wakae Why is it not understood?

Nakajima To put it very simply, the inference, or judgment, of machine learning models is dependent on the algorithms and data used. So even when the learning algorithms are correct, if the model learns from odd data, then it may make strange decisions; and even when the data are correct, if the algorithms are wrong, then the model may make strange decisions. Conversely, even when the algorithms are wrong, the model may produce results that appear to be correct. In short, it is often impossible to know where the problem lies simply by looking at the results inferred by the model.

Wakae That makes things difficult, I would think. How should it be dealt with?

Nakajima I believe that a three-pronged approach is important when thinking about the quality of machine learning software: product quality, service quality, and platform quality. Product

quality concerns the reliability of the product, such as whether programs are bug-free. Service quality concerns the dependability of the inference results, as to whether they meet expectations. Finally, in the case of machine learning, I believe that platform quality is also important.

Wakae Platform quality is not a term that I am familiar with.

Nakajima For example, let us assume that we are creating a learning program that will recognize handwritten Roman numerals from 1 to 10, and we make it learn various numbers. If the program learns properly, then it should be able to produce a correct answer even if the number is poorly written. However, if we enter a Chinese numeral instead of a Roman numeral, the program may give us an incorrect result. We can think about this in two ways. The first is to consider it wrong that we did not think about Chinese numerals, and to decide to repeat the learning process while including Chinese numerals. The second is to consider it wrong that we entered Chinese numerals into a program developed so as to recognize Roman numerals, and therefore, to decide to place limits on the way in which the program is used. In machine learning, it is necessary to judge the adequacy of the program from the results in this way, and repeat the learning process or adjusting the data used for learning. Whether the platform that supports this process, which usually involves human activities, functions properly is a question of platform quality. It is important to approach improving the dependability of AI by considering which of these three qualities is causing a problem.

Wakae Will the problem of AI dependability be resolved by improving the dependability of each of these qualities?

Nakajima Unfortunately, the situation is not that simple, because, in machine learning, data are entered while the program is being used, and it learns from that data, but it is often impossible to predict what data will be entered.

Wakae How about setting conditions in which the AI can be used in advance?

Nakajima The problem is that stipulating such conditions is quite difficult in machine learning. The mistaken recognition of a road sign that we talked about earlier happened when black graffiti was added, but other patterns could appear on a road sign, such as dirt caused by rainwater. If we were to say that for this reason, as a condition of use, autonomous driving can only be used in places without road signs, then that would render it useless. If the condition of use excluded using it on streets with black graffiti on the road signs and an accident happened as a result of mistaken recognition caused by dirt due to rainwater, then who would be responsible? The person who created the machine learning software system, or the person who used it?

Wakae That is certainly difficult. What can be done?

Nakajima I suggest that first we classify the ways in which machine learning is used. This means classifying its usage in three dimensions by the following indicators: (1) Are the data that come in predictable? (2) Can the adequacy of the inference results be judged only by humans, or is automatic judgment possible? and (3) How serious will it be if a failure occurs? Then, we start using it in application service sectors where the impact of any failure that may occur will be negligible. If a failure would have serious consequences, involving human life, for example, then the use of machine learning should be confined to supporting human decisions. For example, if a recommendation system were used for advertising, then it probably would not be a major problem if it made a mistake, but if it were used for medical treatment, then it would be a matter of life or death. Entrusting autonomous driving entirely to AI is too risky as well. The decision on how to use AI in these areas cannot be made only from the perspective of technology. It should be considered from the perspective of social acceptance, and for that reason, I think that it needs to be considered not just by people in technologies but together with those in the social sciences.

Wakae Could it be decided that, even though machine learning is useful, we should not use it?

Nakajima In his book, *Normal Accidents: Living With High-Risk Technologies,* an American sociologist Charles Perrow analyzed the Three Mile Island accident in 1979 as a catastrophic accident that resulted from an accumulation of ordinary, trivial events, and argued that we must not create systems dependent on advanced technologies that cannot be adequately controlled, even if they are technically feasible. I think that the same could be said about AI, but the pros and cons should be considered by society as a whole.

Wakae Finally, please tell me about the third-party evaluation system that you are researching.

Nakajima I believe that an independent third-party evaluation and accreditation organization is needed for the peace of mind of both those providing AI systems and those using them. I envisage an organization that checks resistance to adversarial example attacks, clarifies conditions of use, and identifies responsibility in case of failure. In a similar fashion to the sharing of information about aviation accidents around the world, I think that a framework for reporting/sharing information about accidents involving AI is necessary for expansion of the technology.

(Photography by Yusuke Sato)

A Word from the Interviewer

In the midst of an AI boom, we tend to overlook the fact that AI is prone to attacks and that assuring its quality is difficult. It is frightening that quality problems that are a matter of life or death are still unresolved. I was struck by Professor Nakajima's view that this is why we must not create certain things, even if they are technically feasible, and dialogue between people in social sciences and technologies is essential in drawing the line.

Masako Wakae

Graduated from Aoyama Gakuin University in 1988 and joined the Yomiuri Shimbun in the same year. Worked in the Society Division before becoming Senior Staff Writer in 2014.

Quality of Machine Learning from a Legal Perspective

Speaking with an expert in product liability (PL) about AI issues

Souichirou Kozuka
Professor, Faculty of Law, Gakushuin University

×

Ichiro Satoh
Deputy Director General, National Institute of Informatics
Professor, Information and Society Research Division
Professor, School of Multidisciplinary Sciences,
Graduate University for Advanced Studies

Today's artificial intelligence (AI) systems use massive amounts of training data in development, and as a result, decision-making criteria in these systems also depend to a great extent on that data. It is technically difficult, however, to guarantee the quality of training data. There are also cases in which the user adds training data during system operation making it difficult to determine exactly where responsibility lies with the quality of that data. Today, we talk about quality assurance in the field of AI with Professor Souichirou Kozuka, an expert on the relationship between the law and advanced technologies such as space commerce, autonomous driving, and AI.

From Software-defined to data-defined

Satoh Professor Kozuka, how did you develop an interest in legal issues associated with advanced technologies?

Kozuka My initial area of research was franchise agreements, which, in Japan, are not regulated by civil law or elsewhere. They are a type of agreement that is difficult to understand. Since then, I have had an ongoing interest in legal issues associated with various forms of transactions that are not easily resolved by traditional ways of thinking. Furthermore, on coauthoring the book "Intelligent Transport Systems (ITS) and the Law" (edited by Tomonobu Yamashita, Yuhikaku Publishing Co., Ltd. 2005), I had the opportunity of writing on the theme of "information," and since then, I've taken up related themes such as autonomous driving, AI, and other advanced technologies.

Satoh Today, I would like to ask you about quality assurance from a legal perspective in machine learning and deep learning, technologies that provide the technical platform supporting AI. I believe this is an important topic since requirements for these technologies are quite different from those of conventional software development.

If I could describe machine learning in just a few words, I would say that it's a mechanism for obtaining output by providing input to a "pre-trained model" trained from a massive amount of data. That is, in contrast to explicit setting of classification criteria (conditions) by a program as in conventional software, machine learning automatically learns classification criteria (pre-trained model) from massive amounts of training data. This definition of machine learning reflects a shift from software-defined to data-defined systems in which the results obtained are statistically based. Understanding quality assurance from a legal perspective under such conditions can be quite challenging.

Kozuka Product liability (PL) law stipulates that the party manufacturing a defective product bears the responsibility of providing compensation for damages resulting from defects in the manufactured product. This law targets "manufactured or processed movables (objects)," which means that software itself is not a direct target although the software embedded in an object can be an issue. The same line of thinking applies to products in which software is replaced by data. In the end, if a

Souichirou Kozuka

Professor Kozuka graduated from the Faculty of Law at the University of Tokyo and served as a research assistant at the same university and as a professor at Sophia Law School before assuming his present position. He is a committee member of the Conference Toward AI Network Society held by the Ministry of Internal Affairs and Communications (MIC). He coauthored "Introduction to Space Law for Entrepreneurs" and "Payment Law" among various works and recently contributed to "Autonomous Driving and the Law" (edited by Tomotaka Fujita).

product used by someone has a defect and lacks "safety that the product ordinarily should provide," responsibility lies with the manufacturer. For this reason, I believe that discussion is first needed to determine what is meant by "safety that the product ordinarily should provide" in the case of tools that incorporate AI.

Satoh To the extent that machine learning is dependent on statistical models, a 100% safety guarantee cannot, in principle, be provided. Furthermore, given that training data is necessarily past data, machine learning cannot be applied to a future situation that has not occurred in the past.

Kozuka In general, it has never been possible to provide a 100% safety guarantee for any manufactured product. This is why a rocket, for example, incorporates redundancy that allows for a launch to be aborted via multiple lines of control in the event of an unforeseen problem. In other words, it is sufficient to have safeguards that prevent people from being harmed. Similarly, scissors can be a deadly weapon depending on how they are used, but that would not be grounds for banishing them from society. Rather, it would be desirable for society if users were to become more skillful in the use of products while being warned of any dangers.

Satoh That's encouraging—you explained it well.

Improving user literacy is essential

Satoh Also of concern is the fact that AI cannot deal with unusual situations for the most part. If input deviates from training data, reliability suddenly drops. Simply speaking, the user has no idea what type of training data has been provided.

Kozuka The first consideration here is whether the user can easily react when unexpected results are output to prevent any mishaps. If that is difficult, some type of measure must be taken. There may be a need, for example, for professional users that can assist users in the proper use of an AI system. In this regard, I'm concerned that users are unable to keep up with technical progress, as is the case at present.

Satoh Today, AI systems are centered about the business-to-business (B2B) model, but as ordinary consumers come to use AI, a variety of problems will surely arise.

Kozuka I wonder about the degree of AI literacy even in the corporate world. Taking for example the lawsuit filed by Suruga Bank against IBM in relation to development of a banking system, I believe that the two parties had a different understanding of what to be accomplished. It is essential that initiatives be launched for improving user literacy in AI.

How should data, the ultimate source of value, be handled?

Satoh Another matter of concern in AI is that new training data may come to be added by the user or client after the system has been put into operation. Of course, that added data is of no concern to the original system developers. In some cases, the addition of attack-related data can cause erroneous operation. In addition, the pre-trained model of machine learning implies a mass of statistical values that would be meaningless to people reading them from the outside. On the other hand, the benefits of AI could not be demonstrated if users were prevented from adding new data, so I think that some types of conventions or rules are needed.

Kozuka But in that case, it would be necessary to stipulate the conditions as to what range of data would be acceptable for addition. This, however, could be technically difficult.

Satoh Difficult indeed! For example, in the case of AI for autonomous driving, I think that only broad limits could be established such as no nighttime data in a daytime system. Similarly, a self-driving car trained on country roads may run very well in the country but may cause an accident in the city.

Kozuka In such cases, due diligence on the part of developers and users becomes all the more important.

Satoh On the other hand, it could be said that the value of a system increases the more it accumulates good training data. In other words, a used system could fetch a higher price. There is also the problem of intellectual property when a pre-trained model comes to be distributed.

Kozuka As a legal expert, that's a thought-provoking issue. Would it be better to entrust risk control to intermediate second-hand dealers rather than to the manufacturer that developed the original pre-trained model? A distributer of second-hand goods, however, is generally smaller in scale with fewer resources than the manufacturer. Perhaps a mechanism similar to a certificate of authenticity in the sale of antiques would be necessary. At the same time, distribution of a pre-trained model itself would no doubt be difficult.

Satoh The difficulty in sorting out legal liability and rights in the case of pre-trained models is a problem.

Kozuka Bringing concerned parties together to decide on standard processes is one approach to solving this problem. However, that could stifle competition, so whether it would be a good thing for any such standards to become industry practice is a troubling issue. In any case, producing a product while putting risk aside lacks due diligence and is therefore a legal matter. Discussion is essential.

Meanwhile, Germany and the United States have a history of disseminating new technologies throughout society on the premise that liability for any damages must be accepted. In Japan, no one is willing to take on that risk, so in the event of an accident, the parties to that accident tend to be blamed more than necessary by society. The concept of strict liability should perhaps be studied in relation to the development and diffusion of new technologies.

Satoh Professor Kozuka, thank you very much for sharing your valuable insights with us today.

(Interview/Report by Madoka Tainaka, Photography by Yusuke Sato)

Ichiro Satoh

NEWS 1

Hosting NII Week, packed with academic events

Having designated the week from June 18 (Mon) to 23 (Sat) "NII Week," NII held a variety of academic events at its National Center of Sciences (Chiyoda-ku, Tokyo)

A round-up of NII's research and initiatives

Japan Open Science Summit 2018 (JOSS2018)

Panel discussion at JOSS2018

The first event was the Japan Open Science Summit 2018 (JOSS2018) held jointly by NII, the Japan Science and Technology Agency (JST), the National Institute for Materials Science (NIMS), the National Institute of Science and Technology Policy (NISTEP), the National Institute of Information and Communications Technology (NICT), and Re*poN: Academic Repository Network, on June 18 and 19. One of Japan's largest conferences, JOSS2018, gathered together all stakeholders promoting open science, including researchers, citizen science groups, and organizations supporting research such as libraries and research administration offices, in order to disseminate the latest open science trends in Japan and the world.

On the first day, the keynote address was given by Dr. Ross Wilkinson, Executive Director of the Australian National Data Service (ANDS), with a speech titled "The Opportunities of FAIR and Open Data: an Australian and International Perspective." On the second day, Dr. Makoto Suematsu, President of the Japan Agency for Medical Research and Development (AMED), gave a lecture on "AMED's Mission: Why is Data Sharing Difficult?" Also, representatives of the six sponsoring organizations took to the stage for a panel discussion on "The Current State and Future Outlook of Activities Related to Research Data Use in Japan."

Academic Information Infrastructure Open Forum 2018

The Academic Information Infrastructure Open Forum 2018 was held on June 20 and 21. The aim of this forum was to quickly share with relevant parties a clear picture of the education and research environments at universities and research institutions made possible by NII's Science Information Network (SINET5), and to promote the infrastructure together with its users.

The two-day program was made up of nine tracks of sessions, including a contents track, cloud computing track, SINET track, open science track, learning analytics track, and a security track. The sessions consisted of lectures and lively discussions. In the security track, participants discussed the future vision and direction of the NII Security Operation Collaboration Services (NII-SOCS), a project fully launched in July 2017 to build the foundation for an information security system based on inter-university cooperation.

Open House 2018

The final event in NII Week was NII Open House 2018, aimed at informing the wider public about NII's research results and business activities. It was held over two days on June 22 and 23.

On the first day, the keynote speech was given by University of Tokyo President Makoto Gonokami, a key figure in the government's Council on Investments for the Future, which discusses the national strategy for growth. His speech was titled "Expectations for NII Toward Realizing Society 5.0."

On the second day, there was a programming workshop called "Aim to be a Future Informatics Olympiad Medalist!" where participants attempted to solve problems geared toward the International Olympiad in Informatics (IOI). The participants were helped by the President of the Japanese Committee for International Olympiad in Informatics, Katsuhiko Kakehi, and four IOI medalists, as well as NII's Associate Professor Yuichi Yoshida and Assistant Professor Yoichi Iwata, both of whom have experience of winning prizes in international programming contests. A male first-year senior high school student who participated said, "I was really happy when the output corresponded to the program I wrote," and a female first-year senior high school student commented, "It was difficult, but fun once I understood the meaning of the programming language."

Now in its fourth year, Open House's popular program "NII Research 100" proved to be exciting. Ten NII researchers presented a total of 100 research findings one after the other, with each researcher having seven minutes to present 10 findings. The audience was amazed by the energetic presentations.

[Top] President of the University of Tokyo, Makoto Gonokami, giving the keynote speech [Bottom] Programming workshop where junior and senior high school students attempted to solve problems geared toward the International Olympiad in Informatics

Shonan Meetings 100th Commemorative Symposium

The Shonan Meetings 100th Commemorative Symposium was held before the opening ceremony of Open House. The NII Shonan Meetings are residential seminars where the world's top researchers gather to intensively discuss issues in the field of informatics, and this symposium was held to commemorate the 100th Shonan Meeting. At the symposium, keynote speeches were given by Professor Raimund Seidel of Saarland University, who is an expert in computational geometry research, and Professor Pierre Casteran of the University of Bordeaux, who is famous for a mechanistic proof of the four-color theorem.

Information Session on the Department of Informatics, The Graduate University for Advanced Studies

An information session on the Department of Informatics in the School of Multidisciplinary Sciences, The Graduate University for Advanced Studies, was held in conjunction with Open House. NII is involved with The Graduate University for Advanced Studies, and established the Department of Informatics in the School of Multidisciplinary Sciences. The department provides graduate school education through 5-year full-term and 3-year second-term doctoral programs.

The information session gave an overview of the Department of Informatics and an explanation of how to apply, and students currently enrolled in the department talked on topics such as student life, their own research, and features of the department.

NEWS 2

All members of Japan's team win medals — one gold, one silver, and two bronze

International Olympiad in Informatics hosted by Japan for the first time

The 30th International Olympiad in Informatics (IOI 2018 JAPAN) was held in Tsukuba, Ibaraki Prefecture, from September 1 to 8. There, 335 senior high school students from 87 countries worldwide went head to head using their programming skills.

Of the Japanese representatives, Wataru Inoue (3rd-year student at National Institute of Technology, Kitakyushu College) tied for 6th place and received a gold medal; while, Hiroaki Hosokawa (3rd-year student at Nada High School) won a silver medal. Both Ikumi Shimizu (3rd-year student at N High School) and Koichi Namekata (2nd-year student at Senior High School at Komaba, University of Tsukuba) won bronze medals. Although not eligible for awards, the special host-country participants also performed extremely well, with Masataka Yoneda (1st-year student at Senior High School at Komaba, University of Tsukuba) achieving the equivalent of a gold medal, Hirotaka Yoneda (1st-year student at Kaisei Senior High School) and Yasutaka Hiraki (1st-year student at Nada High School) achieving the equivalent of silver medals, and Riku Kishida (3rd-year student at Kyoto Municipal Horikawa Senior High School) achieving the equivalent of a bronze medal.

International Olympiad in Informatics (Photo provided by Japanese Committee for International Olympiad in Informatics [JCIOI])

The International Olympiad in Informatics is one of the International Science Olympiads aimed at senior high school students. The competition consisted of individual matches that were held for a total of 10 hours over two days (September 3 and 5). The problems included "Combo," which involved discovering secret commands, "Seats," which involved computing the beauty of international programming contest seating charts, and "Mechanical Doll," which involved creating a circuit to operate a doll. There were restrictions on program execution time and memory used, and not only the ability to write correct programs but the ability to devise algorithms and data structures was required. After scoring, the top one-twelfth (in this contest, 29 people) were awarded gold medals; the next top two-twelfths (in this contest, 55 people) were awarded silver medals; and the next top three-twelfths (in this contest, 83 people) were awarded bronze medals.

Japan's representatives: (From left) Wataru Inoue, Ikumi Shimizu, Koichi Namekata, and Hiroaki Hosokawa (Photo provided by JCIOI)

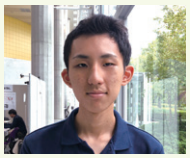
In addition to the competition, there were two days of excursions aimed at strengthening ties between the contestants, who will shape the next generation of mathematical and information sciences. The contestants experienced Japan's cutting-edge science and technology through visits to places such as the Tsukuba Space Center of the Japan Aerospace Exploration Agency and the Science Square Tsukuba of the National Institute of Advanced Industrial Science and Technology, and enjoyed Japan's nature and culture with visits to Hitachi Seaside Park, Aqua World Ibaraki Prefectural Oarai Aquarium, and Oarai Isosaki Shrine.

The next competition will be held in the summer of 2019 in the Republic of Azerbaijan. Applications to enter the national contest for selecting Japan's representatives (18th Japanese Olympiad in Informatics Preliminary Round) will be accepted from October 9. For details, go to https:// www.ioi-jp.org/joi/2018/index.html

NII provides SINET5 network and sponsors a booth

NII supported the contest as a special sponsor. A dedicated virtual private network (L2VPN) was constructed within NII's SINET5 network, and the competition venue was connected to the Tokyo server. Also, NII had a booth in the Tsukuba International Congress Center (near the competition venue) for the duration of the competition, and NII's research and projects were promoted to contestants from all over the world (see photo).

Comments from a Medalist



Wataru Inoue, Gold medal

It is thanks to the All-Japan High School Programming Contest (PC Koshien) that I began programming seriously. The "Seats" problem on the first day was difficult, and I regret that I was not able to get a high score. But I concentrated hard on the second day and achieved a good result. Sixth place is higher than I was aiming for, so I am satisfied. I would like to compete in other programming contests too. In the future, I want to be a programmer or an engineer.

Zi Song Yeoh, Malaysian representative, Gold medal

Comments from an International Contestant I started programming because I was encouraged to by a friend with whom I took part in the International Mathematical Olympiad. This is my third IOI, and I am really happy that I was able to win a gold medal for the first time. I usually practice by solving the past problems of different contests. At the Japan Olympiad, the excursions were fun, as well as the competition. The Science Square was very memorable.

SNS

"Hey, this is great!" Hottest articles on Facebook and Twitter (June -August 2018)

National Institute of Informatics, NII (official) Facebook
www.facebook.com/jouhouken/

NII will hold a graduate school information session on June 23 (Sat). If you are interested, please come along! Here is a message from Iino (1st year of second-term doctoral program): "I am doing research on using structured knowledge to support teaching and learning of musical instruments. I look forward to meeting you!" (June 16, 2018)

National Institute of Informatics, NII (official) Twitter
@jouhouken

NII Research 100

Ten NII researchers present a total of 100 research findings, with each researcher presenting 10 findings. Come and watch 75 minutes of intense presentations that will give an overview of the wide field of informatics and consider its future!

Associate Professor Ohmukai (@i2k) and Ayaka Ikezawa (@ikeay) will act as MCs. (June 5, 2018)

Bit on Twitter! Twitter
@NII_Bit

Thank you to the MCs at NII Research 100, Ayaka Ikezawa (@ikeay) and Associate Professor Ohmukai (@i2k)!!

#NIINow (June 22, 2018)

* Some text edited/omitted.

Essay

The True State of the Engineering Research ommunity

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

Software engineering is a field in which researchers examine how to create better software. Needless to say, research on creating software is meaningless unless it is actually used by people who create software. However, in recent years, the software engineering community has seen a decrease in participants from the business sector who are developing software. Declining interaction between industry and academia raises the possibility that university researchers will envisage problems that are unlikely in reality and will invest their efforts in solutions that will go unused.

Meanwhile, more than 70% of people participating in the machine learning engineering community introduced in this edition of *NII Today* are from business. Partly because the community is so new, its current activities are centered on sharing and organizing issues that software developers are actually experiencing. This means that it offers excellent opportunities for researchers to hear firsthand the voices of people developing software, and it is overflowing with problems needing to be solved.

Conventional software engineering grew centered on the theory of organizing knowledge and creating software logically based on that knowledge (deductive software development). This theory is completely different from that of machine learning, in which rules are inductively derived from data (facts) (inductive software construction), and it is difficult to create appropriate, high-quality software using machine learning by applying the methods used in the past. Problems are therefore faced in various areas of software development using machine learning.

The essence of the issue arises from combining machine learning centered on statistical processing with the conventional, logical software descriptions of algorithms, rules, and inferences. Addressing this issue is also challenging and meaningful academically.

If we can create software that combines logical thinking and inductive inference from facts, in the same way that the human brain skillfully combines reasoning with intuitions based on experience to make appropriate decisions, then it could become possible for software to solve various problems by making appropriate decisions.

We are currently listening to opinions from the frontline while managing this machine learning engineering community, but our efforts in the near future will be critical in determining whether we researchers can contribute. Therefore, let us draw on the collective wisdom of researchers to set about addressing the issue.

Future Schedule

October 14 | Inter-University Research Institute Corporation Symposium 2018 (Exhibitor)

October 16—19 | CEATEC JAPAN 2018 (Exhibitor)

October 24 | "The Forefront of Informatics" 4th Public Lecture "Radio Access Networks in the Future: Not Enough Frequencies!" (Speaker: Associate Professor Megumi Kaneko, Information Systems Architecture Science Research Division). For details and registration, go to **https://www.nii.ac.jp/event/shimin/.**

October 25 | 2nd SPARC Japan Seminar 2018. Details at **http://www.nii.ac.jp/sparc/.**

October 30—November 1 | 20th Library Fair & Forum (Exhibitor)

November 9 | 3rd SPARC Japan Seminar 2018

November 19—21 | 2018 Annual Meeting of the Academic Exchange for Information Environment and Strategy (AXIES) (Exhibitor)

November 20 | "The Forefront of Informatics" 5th Public Lecture "Sharing of Real Data: Your Information Used for Academic Research!? No Problem" (Speaker: Professor Keizo Oyama, Digital Content and Media Sciences Research Division)

November 30 | Special SINET Session ① "Development of Production Technology By Cattle Grazing Using AI and ICT" (Speaker: Professor Takafumi Gotoh, Research Field in Agriculture, Agriculture, Fisheries and Veterinary Medicine Area, Department of Agricultural Sciences and Natural Resources, Faculty of Agriculture, Kagoshima University). For details and registration, go to h**ttps://www.nii.ac.jp/event/shimin/.**

December 11 | "The Forefront of Informatics" 6th Public Lecture "Calculation Theory and Reality: Making Difficult Calculations a Breeze" (Speaker: Assistant Professor Yoichi Iwata, Principles of Informatics Research Division)

Notes on cover illustration In connection with machine learning, I drew a scene of a lesson on how to sort apples and oranges. The teacher robot is showing the student robots an apple with a red label attached.Look carefully and you will see that some pieces of fruit have lost their labels.

Weaving Information into Knowledge
NII

National Institute of Informatics News [NII Today]
No. 67 Sep 2018
[This English language edition NII Today corresponds to No. 81 of the Japanese edition.]
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
Publisher | Masaru Kitsuregawa **Editorial Supervisor** | Ichiro Satoh
Cover illustration | Toshiya Shirotani **Copy Editor** | Madoka Tainaka
Production | MATZDA OFFICE CO., LTD., Sci-Tech Communications Inc.
Contact | Publicity Team, Planning Division, General Affairs Department
TEL | +81-3-4212-2028 **FAX** | +81-3-4212-2150 **E-mail** | kouhou@nii.ac.jp

Bit (NII Character)

https://www.nii.ac.jp/en/about/publications/today/