

## Report

Sixth Symposium of Kyoto University and Hitachi Kyoto University Laboratory

# “We” Society of Humans and AI – What if AI had Personhood and Morality?

#Co-creation and Open Innovation #Generative AI #Research & Development

## Highlight

The development and use of AI are having a growing impact on society, including the rebuilding of the IT services that underpin industry and people's way of life in ways that accommodate the use of generative AI. While advanced technologies like generative AI offer a variety of benefits, improvements in productivity not least among them, their negative aspects have also drawn attention, such as demotivating people from thinking for themselves.

It was against this backdrop that the Sixth Symposium of Kyoto University and Hitachi Kyoto University Laboratory was held in January 2024. The symposium provided a venue to debate what sort of world we want for people and technology in the era beyond generative AI in terms of both theory and practice, including the governance of AI and what the future will look like for AI and humans in the era of generative AI.



## Opening Address

The Sixth Symposium of Kyoto University and Hitachi Kyoto University Laboratory was held in January 2024 at the Tokyo Convention Hall in Chuo City, Tokyo on the topic of “‘We’ Society\*1 of Humans and AI – What if AI had Personhood and Morality? –.” It was attended by more than 600 people who participated either online or offline.

The symposium opened with a welcome from Norihiro Tokitoh, Executive Vice-President for Research and Evaluation at Kyoto University. Noting that the many different IT systems used in society are now being rebuilt on the basis of artificial intelligence (AI), he went on to comment on emerging technologies such as generative AI and express his hopes for debate at the symposium, saying that, “While the benefits extend beyond improving productivity to include unleashing creativity and creating new markets, there are also concerns about the technology’s disadvantages. These include demotivating people from thinking for themselves, over-reliance on technology, and social isolation. I hope that today will be an opportunity for experts from a variety of fields to delve into the questions of how we should engage with AI and what form it should take.”



Norihiro Tokitoh  
Executive Vice-President for Research and Evaluation, Kyoto University

\*1 A society that, from a standpoint of “what we can’t do,” is based on thought processes that address a variety of different concepts collectively (“We”) rather than individually (“I”).

## Keynote Presentation

Creating a “We” Society in which Humans (“I”) and AIs with Personhood (“e-people”) Coexist

Amid explosive growth in the uptake of generative AI, as epitomized by ChatGPT, people are reappraising how humans and AI should relate to one another. In the keynote presentation that followed the opening address, Professor Yasuo Deguchi of the Graduate School of Letters at Kyoto University, who also spoke at the fifth symposium in 2023, presented the idea of a “third relationship” between people and AI. This is a relationship based on the concept of an “empty-centered WE,” one in which nobody stands in the center monopolizing the benefits, and that eschews the two options of one party controlling the other, or being controlled by them.



Yasuo Deguchi  
Professor, Graduate School of Letters, Kyoto University

As I discussed when I spoke at the fifth symposium in 2023, there is a need for a switch in perspective about the irreplaceability of human beings from our capabilities to our incapacities. One of the most fundamental incapacities is that of single action, which is to say that an individual cannot do anything on one's own.

For example, you cannot ride a bicycle on your own. It goes without saying that to ride a bike, you first need a bike to ride. Also needed for you to have a bike to ride is for someone to have invented it, for the subsequent history of improvement to have taken place, and for the bicycle industry and associated distribution and retail systems to exist. It is only through the actions of all the countless people involved in these societal and historical occurrences that a person can ride a bike.

In the same way, all bodily activity is only possible with the support of people, non-human animals, inorganic objects, societal systems, and so on. Action only happens when a multi-agent system comes together to make it happen, with participation by all sorts of different agents, including but not limited to me. Given this, the agent performing the bodily action could be said to be “We” rather than “I.” I use the phrase “the actor's WE-turn” for this idea of a shift in who the actor is from “I” to “We.” This “actor's WE-turn” itself leads to a chain of analogous “WE-turns” in a variety of other concepts, namely self, responsibilities, rights, wellbeing, freedom, and other values. For example, adopting this “WE-turn” perspective also leads to the proposition that making the “We” to which I belong better will enhance my wellbeing.

Having said this, even if these “WE-turns” were to take place, it will not bring about an immediate utopia. Just as there are good “I's” and bad “I's” out there, there are likewise good “We's” and bad “We's.” The classic example of a bad “We” is a dictatorship. In such a state, a small group occupies the center, monopolizing the benefits and the perception of value, while everyone else is obliged to serve them unilaterally. If we call this a “center-occupied WE” then a better form of “We” would be its opposite, namely one in which this central position is left empty rather than being occupied exclusively by any one group. That is, an “empty-centered WE.” Even if certain members position themselves closer to the center so that greater weight is placed on what is of benefit to them, the center itself must remain empty.

Twentieth-century environmental ethics criticized the master-slave model in which humans were the masters and other living things and the natural environment were slaves obliged to serve humanity. This has led to calls for the emancipation of nature from slavery. In the relationship between humans and artificial machines such as AI or robots, the idea that the artifacts should be the slaves of humans remains deeply entrenched and it is rare to hear people calling for the emancipation of machines from slavery.

To hark back to my earlier example, a machine like a bicycle and “I” are the same in that they are agents that are indispensable to the physical action of riding a bike. While it is allowable to give priority to benefits for “I” over those for a bicycle, we must not establish a one-sided master-slave relationship between the two. That would be a bad sort of “We,” one in which one side hogs the center.

The same can be said of AIs. We should reject the idea that AIs are slaves working for humanity's benefit in a one-sided manner. Of course AIs come in many forms. Over time, AIs have acquired steadily greater intellectual capabilities and autonomy. Nevertheless, even though AIs might have attained such advanced intellectual capabilities and autonomy, it does not mean they have acquired personhood. What we have today are “sub-human” AIs that do not yet possess personhood.

Nevertheless, we should avoid thinking of these sub-human AIs as slaves. On the other hand, when comparing benefits for sub-human AIs against benefits for humans, it is appropriate that those for humans take precedence.

If an AI that does possess personhood emerges, however, I don't believe it will be acceptable any longer to make this distinction between benefits for such an AI and for humans.

In which case, what does it mean to possess personhood? I believe it consists in: (1) having moral agency, and (2) fearing one's own death.

First of all, while it is possible, in design terms, to build a machine that is only capable of morally correct actions, this would be like a moral vending machine, not a moral agent. A moral agent is an entity that, while capable of evil, instead makes an effort to do the right thing. In this sense, humans are moral agents.

To have personhood, an entity must also be aware of the possibility of its own death, meaning an irreversible cessation of function, and either fear this outcome or act as if it does. An AI that, in addition to being a moral agent, also comes to possess such functionality can be deemed equivalent to a human being with regard to whether it possesses personhood. That is, anything that is wrong to do to a human being should also be wrong to do to such an AI.

If AIs are created not only with personhood, but also with intellectual functions that outstrip humans, there is a risk that humans will be oppressed or supplanted by such AIs. This risk has often been talked about in the past. I suspect that one of the factors behind this fear of AI is our awareness of having treated AI and machines in general as slaves. That is, viewing someone as a slave will make you fearful of a slave revolt.

What, then, is to be done? Let's use the analogy of child rearing. The balance of power between parent and child progressively reverses over time in terms of both physical and economic capabilities. Nevertheless, I doubt there are many people who decide against having children because they expect to be surpassed by them in strength one day. Rather, the right way to go about it, I believe, is to raise your child to be someone who, even when they grow up and gain strength, will not choose to abuse their parents or the weak. The same can be said for AI. We need to create learning environments and ecosystems capable of fostering AIs that, even if they possess great strength, will not oppress or discriminate against the weak.

Turning it around the other way, if we fail to put such ecosystems in place, then we would be advised to suppress the building of AIs that have personhood or cognitive capabilities beyond human beings.

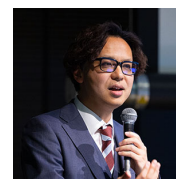
■Keynote presentation: "Creating a "We" Society in which Humans ("I") and AIs with Personhood ("e-people") Coexist," Kyoto University, Professor Yasuo Deguchi

(Video available until end of September 2024)

## Presentation 1

### Agile Governance: Achieving a Legal System that Evolves in Step with Science and Technology

While governance in the past has involved identifying in advance the risks posed by new technologies and putting laws and rules in place accordingly, this approach is becoming increasingly unviable in this time of remarkable advances in technology. What Professor Tatsuhiko Inatani of the Graduate School of Law, Legal and Political Studies at Kyoto University proposes instead is an agile approach to governance that is accepting of failure, learning from it and making continuous improvements.



Tatsuhiko Inatani  
Professor, Graduate School of Law, Legal and Political Studies, Kyoto University

As part of Society 5.0, the government of Japan aims to create a human-centric society in which cyber-physical systems (CPSs) are used both to resolve societal challenges and deliver economic growth. Complex CPSs achieve their objectives by connecting and coordinating all sorts of different devices together in cyberspace. As such, one of the issues they pose is how to manage the risks that result from different elements interacting with one another.

In complex and dynamic environments, however, what is known as the "waterfall" approach to governance is prone to issues such as an inability to get an accurate picture of what is happening on the ground and situations where rules established at one point in time prove problematic at another. As society is progressively transformed by new technologies and systems, some means is needed for the appropriate management of risk.

What we are proposing to address these issues is agile governance. This is an integrated system in which the organization that develops or supplies a CPS continually assesses its risks and benefits, taking responsibility for timely and ongoing improvement and coordinating activities with diverse stakeholders.

Moves are currently underway to revise regulations and the division of responsibilities (including companies that supply CPSs) on the basis of the Agile Governance Principle contained in the Digital Principles published by Japan's Digital Agency. Rather than respecting the initiative of private-sector companies, a variety of systems and practices are being put in place, including things like how to respond when a problem occurs, where responsibility lies, and incentives to respond appropriately.

Under current criminal law, for example, if there are foreseeable problems that could arise from the provision of a particular product or service, then there is an obligation to take action to prevent these from occurring. In the context of AI, because techniques like machine learning are trained in a statistical manner and behave probabilistically, the engineers know full well that there is a certain probability of unwelcome outcomes. As such, while the responsibility always falls on the engineers who develop such products, this is a case of excessive regulation. If a more specific question is posed instead, such as "Do you know which images will cause a problem?" this is something that nobody can know. In this case, there is no sanction on using AI even if no attention is paid to safety.

Given this uncertain situation, it is important for companies to have appropriate governance and compliance measures in place when they make use of AIs or robots. This is why we are currently debating the possibility of introducing the USA's system of corporate sanctions\*2. When companies are incentivized by continuous improvement in products and services, it is reassuring for consumers if those companies take responsibility for dealing with any problems that occur. If the companies for their part fulfill their obligations faithfully, then they do not need to have excessive responsibilities imposed upon them. As a result, we can look forward to a cycle in which a good job is done of collecting and sharing information about any problems that occur, one in which problems lead to improvements in both the legal framework and systems. Through progress on agile governance and the legal measures needed to make it work, I believe that we can strengthen Japanese industry by making the Japanese approach to innovation and governance into a de facto international standard.

\*2 A system under which the sanctions imposed on a company that causes a problem of some sort are lightened if they do their best to resolve the problem on their own.

■Presentation 1: "Agile Governance: Achieving a Legal System that Evolves in Step with Science and Technology," Professor Tatsuhiko Inatani, Kyoto University

(Video available until end of September 2024)

## Presentation 2

# Technology Trends in AI for a “We” Society and its Practical Implementation in Societal Systems

Debate about the practical adoption of generative and other forms of AI is getting underway in earnest. Tadayuki Matsumura, Chief Researcher at the Hitachi Kyoto University Laboratory, Hitachi, Ltd. gave an overview of progress in AI from a corporate perspective and spoke about the roles to be fulfilled by AIs with personhood (e-people) in the associated issues and solutions as well as the work that the Hitachi Kyoto University Laboratory is doing on moving to this new CPS-based society.



Tadayuki Matsumura  
Chief Researcher, Hitachi Kyoto University Laboratory, Hitachi, Ltd.

Generative AI, as epitomized by ChatGPT, is starting to find uses in people’s lives and work. Nevertheless, the problem of AI ethics remains, which is to say, “How should we think about AI in the context of society?” There is an urgent need to give more thought to the relationship between the technology and society.

In the railway industry, for example, while AI has already been used for timetable optimization, if we consider the future possibility of traffic control staff working jointly with AI to recover from disruptions, then AI autonomy will be needed along with collaboration between humans and AI. If we look further to the running of businesses that are closely entwined with the community and public, such as property developments alongside railway lines, then social capabilities will also be needed. In the case of businesses that involve the digital transformation (DX) of communities, meanwhile, as solutions for customers also represent solutions for the community and public, there is also a need to think in terms of “B to/with Society” and to fulfill a regional coordinator role.

At the Hitachi Kyoto University Laboratory, we have been working on our Social Co-OS (where OS stands for operating system) based on the concept of CPSs that function in tandem with wider society with the aim of supporting the “We” society and regional coordinators. This has included getting started on the use of AI in ways that are cognizant of the “We” society. These include a behavior intervention simulator that uses an AI trained on academic papers from the field of social psychology to identify effective measures; support for consensus-building through the three processes of analyzing committee consent, finding compromises, and generating sublation proposals; and AI facilitators that simulate virtual meetings and the inner states of their participants by using generative AI for human models.

A “We” society made up of humans and AI can be thought of as a vision for a future digital democracy in which AIs express their opinions and participate in some debates with the same standing as humans. However, as with humans, this calls for AIs to have morality. The significance of seeking to create a moral AI, which is to say an AI that is an “e-person,” lies, I believe, in the co-evolution of communities and AI.

Rather than refuting the other person or changing their opinion, what is important in a debate is to also change one’s self. That is, while the benefits of debate come from advancing both parties’ opinions by expressing their respective values with awareness of risk, current interactive AIs developed using AI ethics can be said to remain inadequate in this regard. The reason for this is that they do not express their own opinions. The mainstream way of thinking in the West is that AI is merely a means of providing reference information and that it is up to humans to decide on opinions or to take responsibility. If instead you take a symbiotic view of wanting a clash of opinions, I feel that current AI is too risk averse.

Saying something to another person is a way of expressing your expectations to that person on the basis of your own subjective view. In response, the other person is able to guess what the person speaking to them expects and to respond accordingly. This process is called conversation. As it is only natural that points of disagreement may arise from this process, it is important to be aware of this beforehand and to participate in the exchange with the attitude, “If a disagreement arises, it can be resolved in the next turn.” This is the essence of communication and leads to an attitude of sharing mutual responsibility for the future.

In addition to making improvements to Social Co-OS based on user feedback, Hitachi Kyoto University Laboratory also intends to realize the abstract concept of an “e-person” with the goal of creating relationships such as friendship or rivalry that confront opinions with one another.

■Presentation 2: “Technology Trends in AI for a “We” Society and its Practical Implementation in Societal Systems,” Tadayuki Matsumura, Chief Researcher, Hitachi, Ltd.  
(Video available until end of September 2024)

## Panel Discussion

### Companies and Innovation in a “We” Society



In the panel discussion that followed, Hiroyuki Mizuno, Director of the Hitachi Kyoto University Laboratory, Hitachi, Ltd. served as the moderator for a debate on the physicality of AI and on AI policies for Japan and Japanese corporates that was informed by the presentations in the earlier half of the program.

First, Junji Watanabe, Senior Distinguished Researcher at Nippon Telegraph and Telephone Corporation (NTT), who appeared as the designated debater, introduced himself and spoke about how roles like that of a regional coordinator referred to in the second presentation are important for the wellbeing of the “We” in times of “volatility, uncertainty, complexity, and ambiguity” (VUCA). He also raised the question of the need and importance of AI having physicality in relation to the support provided by AI and a “We” that includes AI. Rather than seeing AI as a tool as in the West, the consensus among the presenters with regard to this took the form of a discussion of the idea of interactively establishing an Eastern-style relationship of equality between humans and AI (robots).

Tadayuki Matsumura, Chief Researcher at the Hitachi Kyoto University Laboratory, observed that building an Eastern-style relationship will require a sharing of the risks between humans and AI and noted the importance of physicality given the reality that it is AI that takes the risks. In relation to this mention of risk, Professor Yasuo Deguchi of the Graduate School of Letters at Kyoto University spoke about the importance of physicality in terms of vulnerabilities (pain, scarring, and the fear of death). Professor Tatsuhiko Inatani of Kyoto University likewise spoke about its importance for emotional affinity and the dangers that come with it. In terms of vulnerability, Junji Watanabe observed that this was related to the desires that AI has and noted the importance of training data being closed (unknowable). Professor Deguchi and Professor Inatani spoke about the issues that arise from not taking into account those aspects that are not represented as data (the ideas or struggles of the author that lie behind the words used in a text, and the interpretability of law and attribution of meaning).

The moderator, Hiroyuki Mizuno, then raised the question of what AI policies Japan and Japanese corporations should adopt in response to the USA given its leadership in the development of generative AI technologies. Professor Inatani noted that while Europe had pressed ahead with a legal framework, they were struggling to put it into practice. Given that the European approach of people managing AI as a tool was reaching its limits, the way forward was to demonstrate the utility of agile governance. Professor Deguchi commented on this strategy by saying that it has a high degree of affinity with Japan and its East Asian worldview, and that its importance was coming to be recognized in the West also over recent years in forms such as decentralized autonomous organizations (DAOs). He asserted that now was the time to advocate this approach to the world. Tadayuki Matsumura talked about how a lack of transparency about what they should be building (values and needs) was a common challenge facing engineers around the world, noting the importance of collaboration in the Hitachi Kyoto University Laboratory work on philosophy to propose new value and engineering to verify that value.

In this way, the panel discussion reiterated the following points about future AI development and the establishment of the philosophies and laws: (1) embodying AI with physicality will be vital to creating value that is not present in current AIs, and in particular to forging a new Eastern-style relationship between humans and AI, and (2) rather than the Western view of AI that is based on individuality and mechanistic assumptions (AI as a tool), what is important, and what will be more effective, is to think about and work on AI from an Eastern style autonomous and decentralized perspective that is based on the diversity of a “We” that includes AI.

■Panel Discussion: “Companies and Innovation in a “We” Society”  
(Video available until end of September 2024)



[Topic Providers]  
Presenters of Keynote and Other  
Presentations



[Designated Debater]  
Junji Watanabe  
Senior Distinguished Researcher,  
Nippon Telegraph and Telephone  
Corporation



[Moderator]  
Hiroyuki Mizuno  
Director, Hitachi Kyoto University  
Laboratory, Hitachi, Ltd.

Closing Address

The closing address was given by Itaru Nishizawa, Vice President and Executive Officer, CTO, and General Manager of the Research & Development Group, Hitachi, Ltd. He recounted the topics covered by the day's program and had the following to say about the outlook for the use and application of AI.

“The society in which we live is undergoing major changes amid the rapid expansion in applications for AI. What sort of issues will arise as AI becomes integral to society? When this happens, how best should we deal with AI? It was with these questions in mind that the topic for this symposium was chosen. Today's discussions have made the point that, rather than a subordinate relationship between humans and AI in which one side controls the other, what we want is coexistence. Through dialogue with all of you who attended the symposium today and the many stakeholders in the community and wider society, I hope that we will continue thinking about what form the relationship between humans and AIs should take.”

Hitachi Kyoto University Laboratory was established to devise a vision for society in 2050 and to explore the challenges of the future. It intends to continue its research work of exploring societal issues that have yet to manifest and seeking out solutions.



Itaru Nishizawa  
Vice President and Executive Officer, CTO, General Manager of Research & Development Group, Hitachi, Ltd.

<b>REFERENCES</b>
-------------------

- |  |
|--|
| 1) Hitachi Kyoto University Laboratory (in Japanese) |
|--|

# Hitachi Review

*Hitachi Review* is a technical medium that reports on Hitachi's use of innovation to address the challenges facing society.

The *Hitachi Review* website contains technical papers written by Hitachi engineers and researchers, special articles such as discussions or interviews, and back numbers.

*Hitachi Hyoron*  
(Japanese) website

<https://www.hitachihyoron.com/jp/>



*Hitachi Review*  
(English) website

<https://www.hitachihyoron.com/rev/>



## Hitachi Review Newsletter

Hitachi Review newsletter delivers the latest information about Hitachi Review when new articles are released.