

New Developments in Communication Science Research in the Generative AI Era— Exploring the Nature of Humans and Information for Co-creating Human- centered Technologies with AI

Futoshi Naya

Abstract

NTT Communication Science Laboratories (CS Labs) is dedicated to the advancement of “heart-to-heart communication” between humans and computer systems. Our research focuses on the development of fundamental theories that explore the nature of information and humans, as well as the creation of innovative technologies that will revolutionize society. This article highlights some of CS Labs’ efforts toward the coexistence of humans and artificial intelligence (AI), taking into account the recent and rapidly advancing trend of generative AI.

Keywords: communication science, artificial intelligence, brain science

1. Introduction

The news that OpenAI has reached 100 million users in just two months since the release of ChatGPT, an interactive generative artificial intelligence (AI), in November 2022, has had a huge impact around the world. As the term “interactive” implies, it has the ability to interact naturally with humans. In addition to writing, summarizing, and translating natural sentences, it can also read and answer questions about charts and diagrams. Furthermore, it can automatically generate desired images, videos, music, and even programs based on user instructions. These capabilities are attracting significant interest. This is due to the significant improvement in the computational power of graphics processing units (GPUs), which has supported the development of deep learning, as well as a major breakthrough in natural lan-

guage processing called “transformer,” a technology that enables learning from a vast and diverse language resource. The Generative Pretrained Transformer (GPT) is a large language model (LLM) that is the result of learning a large amount of data, and the GPT is equipped with an interface that interacts with users, making it suitable not only for researchers and engineers but also for a wide range of general users by providing an environment that can be easily used with smartphones and other devices.

With improvements in accuracy and reliability, such generative AI is sure to permeate our daily lives in the future. The recently announced GPT-4o has attracted further attention for its ability to handle multimodal input-output responses, such as real-time spoken dialogue and questions while capturing and analyzing video from a smartphone camera. However, current generative AI requires users to verbally

provide detailed information about the situation through prompts, depending on the type of response they want. Although a multimodal LLM is of course multimodal, the LLM is built on the basis of linguistic information, and multimodal input information such as video and audio are tokenized and processed to fit the input format of the LLM. Even before humans are able to understand and speak, their brains process all types of sensory information to understand and remember things and concepts, and abstracted symbolic systems such as language are associated with these representations in the brain. In other words, the current multimodal LLM has the limitation of not being able to handle sensory information that cannot be captured by language. This is called the *Symbol Grounding Problem* [1] proposed by cognitive scientist Stevan Harnard in 1990, and it is a fundamental problem that asks whether AI can understand things represented by symbols expressed in language in the same way that humans do by associating real-world concepts and meanings. The current LLM has not yet reached the stage where it can recognize the preferences, feelings, intentions, knowledge, and beliefs of others through facial expressions, attitudes, and conversations, and respond to them in a thoughtful way, or respond based on its own personality and beliefs, as humans do. There remains major philosophical issue, the *Theory of Mind* [2], which asks whether AI can infer and understand the state of mind of others as humans do.

Since its establishment in 1991, NTT Communication Science Laboratories (CS Labs) has been conducting fundamental research to pioneer the field of communication science that connects information and humans, with the goal of “interdisciplinary elucidation of the mechanisms of human understanding” [3]. In the 33 years since its founding, CS Labs has continued to create new discoveries and innovative technologies based on a deep understanding of the nature of information and human beings. As research advances, we recognize the need for a broad approach that spans a variety of interdisciplinary fields, including computer science, engineering, neuroscience, psychology, social sciences, philosophy, medicine, biology, and even mathematics.

In this article, I present representative examples of our recent research activities at CS Labs, which is expanding its interdisciplinary research fields, from the four perspectives of “mastering the essence of information,” “mastering the essence of human nature,” “getting close to people and society,” and

“pursuing fundamental theories.”

2. Mastering the essence of information

At CS Labs, we are conducting research on information processing technologies for all media that convey information in communication. The development of sensing technology has made it possible to visualize and convert previously unobservable phenomena into meaningful information. *Sound* is a very familiar part of our daily lives, but it is usually measured using microphones. Sound travels as waves through the air, and while microphones can measure sound at the point where they are placed, it is difficult to know in detail how sound is generated and how it travels through space. In recent research at CS Labs, we have used optical technology with laser beams and a high-speed camera to capture sound waves as moving images, creating a *sound visualization system* that can be used to measure the sound field [4]. The phase of the laser light is modulated in accordance with the coarseness and density of the sound as it passes through the sound field. Our technology images the sound field by interfering with a laser beam that is unmodulated by the original sound and measures the modulation by capturing the interfering light with a high-speed camera. However, the noise in the measurement is very large, so by combining simulation-generated sound field datasets with deep learning techniques, we trained the model by varying the number of sound sources and noise conditions to eliminate this noise and provide a clear visualization of the sound field. This technology is expected to have a wide range of applications, including not only the visualization of various sounds but also the design of acoustic devices that reproduce pleasant sounds and the analysis of noise sources. Instead of using microphones to measure sound vibrations, which have had large individual differences in accuracy, we expect the technology to develop into an ultra-precise sound pressure measurement technology that aims to become the next-generation standard based on physical quantities such as optical frequencies.

In addition to the sound and light mentioned above, we have recently been developing a technology that uses AI to infer information about the behavior of the heart muscle cells that are the source of the electrocardiogram (ECG) [5]. Although the rough correspondence between ECG waveforms and diseases has been elucidated in medicine, it has been difficult to estimate what is happening at the level of myocardial cells, which is the basis of the ECG waveforms.

In this research, we developed a machine learning technique to solve the inverse problem of accurately estimating parameters generated from artificially generated ECG waveforms using the *Fugaku* supercomputer, which simulates a physical model of the heart and generates ECG waveforms using parameters such as Na (sodium) and Ca (calcium) ions in heart muscle cells, conductivity, and heart geometry as inputs. This research aims to create a bio-digital twin that enables the simulation of tailor-made medicine by measuring various biological data, including ECGs, and analyzing the factors and relationships among them to predict health conditions that reflect individual genetic characteristics and lifestyle habits and verify which drugs and treatments are effective. It is extremely difficult to model the heart as an organ, the relationships between blood and other organs from the microscopic behavior of heart muscle cells and biochemical reactions in the body, and also to model the macroscopic behavior of the entire human body as an individual. To solve this problem, it is necessary not only to capture the various events that occur in the body but also to have a technology that can quickly find accurate information from a large number of combinations about which events are interdependent and causally related to each other. The article in this issue introduces large-scale data analysis with high speed and rigor as a springboard for solving such problems [6].

3. Mastering the essence of human nature

CS Labs promotes research on the human senses (vision, hearing, touch, etc.), motor control, and emotions with the goal of scientifically elucidating the mechanisms of human information processing. We conduct a wide range of research, including studies that investigate the *universality* of human sensory, motor, and emotional functions, as well as research that seeks answers to fundamental questions about the *diversity* of innate characteristics of each individual and acquired experience and learning.

For example, it is known that when we walk, we unconsciously estimate our walking speed based on the information we receive from our eyes and adjust our walking motion to walk at the optimal speed. We have conducted research to determine whether the sense of speed we feel when we watch a first-person camera image of another person riding a bicycle on television is the same as the sense of speed we would feel when we were actually riding a bicycle. We experimentally investigated the change in walking

speed of participants wearing a head-mounted display as they walked through a virtual corridor, moving the striped walls back and forth. The results indicated that the coarser the stripes, the faster the participants felt they were walking, and the finer the stripes, the slower they felt they were walking. However, when the participants did not walk but simply watched the stripes flow, the finer the stripes, the faster they felt they were moving, and the coarser the stripes, the slower they felt they were moving [7].

This suggests the possibility that the process of speed estimation in the human brain differs between motion and perception (multiple speedometers exist in the brain). This finding may contribute to the design of interfaces that provide highly immersive experiences in virtual spaces such as the Metaverse or present visual images that do not cause virtual-reality sickness.

The above is an example of research on the universality of human sensory and motor abilities. A familiar example of the diversity of human abilities is the dominant hand and foot. Many people develop differences in the dexterity of their hands and feet as they grow up. Right-handed people have great difficulty writing with their left hand, while people who have been corrected from left-handed to right-handed can move both hands with some dexterity, depending on the type of movement. Recent research at CS Labs has developed a method for easily measuring and quantitatively evaluating limb dexterity by simply rotating a smartphone in a circular motion. Experimental results indicated that the dominant hand has less variability in repetitive movements than the non-dominant hand, how this variability changes with age, and the essential mechanism by which this variability occurs. The article in this issue explains the above in detail [8].

Humans are emotional creatures, and their subjective evaluations of their impressions of things and others are subject to ambiguity, depending on their own physical and psychological state at the time. Questionnaires that ask about subjective impressions involve the individual's habitual emotional expressions and perceptions, fluctuating responses to the same question, and other ambiguities. Such uncertainty is an inherent part of human nature, but is it possible to know a person's true feelings from the pattern of their responses? In a recent study, we proposed a method for estimating the reliability of responses by statistically extracting response patterns and eliminating ambiguities [9].

4. Getting close to people and society

Thus far, I have introduced research on the nature of information and human nature. In this section, I introduce our research on the nature of communication, which is to connect information and people or to connect people to each other and society through information.

Today, information and telecommunication networks are deeply embedded in our daily lives and support all societal activities. Failure of network infrastructures, not only due to natural disasters but also due to human error, can have devastating effects. Designing more robust and reliable networks requires reliability analysis to analyze the durability of components against failures and disasters and vulnerability analysis to identify which components need to be strengthened. However, even with only 50 network components, there are up to 1000 trillion combinations, and conventional techniques can only be applied with approximate analysis methods. To evaluate the reliability of network design on a realistic scale, rigorous solution methods without approximations are required. In an article in this issue, we introduce in detail the rigorous solution method for network analysis using a data structure called decision graph [10].

In human sciences, it has become possible to quantitatively evaluate indices such as unnaturalness and comfort that people perceive in images by mathematically modeling and simulating human visual information processing mechanisms using deep learning technology. The article in this issue presents a technology for automatically generating images with which people feel natural and comfortable on the basis of a model of human visual information processing [11].

As reported in a recent press release, we were the first to discover that the electroencephalogram (EEG) of e-sports players just before a fighting game match reveal patterns that are strongly related to subsequent victory or defeat [12]. Unlike machines, humans are not only physically skilled, but their performance in a game is also affected by *mental* factors such as extreme pressure and tension. This study revealed EEG patterns associated with strategic decision-making about how to respond to the opponent's moves, and EEG patterns associated with emotional control to remain unperturbed in the face of adversity. It was also shown that a machine learning model can predict whether the player will win or lose with about 80% probability on the basis of the EEG state imme-

diately before the game. This suggests that there is an ideal brain state for competitive gaming. This fact may lead to a new mental training that approaches the ideal brain state not only for sports but also for people who need to cope with pressure and require high levels of skill and ability, such as doctors performing surgery and pilots who need to make accurate judgments and control their aircraft. The IOWN (Innovative Optical and Wireless Network) concept promoted by NTT aims to enable Digital Twin Computing for humans. The results of this research will lead to the realization of skill enhancement by digitizing the brain state of a skilled person and simulating the process of approaching that of a skilled person from the brain state of a trainee.

In childcare and education, it is important to study the development of social skills with others in addition to the development of language in infants. The closeness that young children feel toward others is the basis for friendships and is an important topic in developmental psychology. However, the traditional method of behavior observation by adults is very time consuming, and the purpose is easily revealed to young children. In a recent study at CS Labs, infants aged 3–6 years were asked to draw pictures of themselves and others, and it was found that the smaller the horizontal distance between the closest points between the drawn figures, the greater the intimacy the infants felt towards the others, and that these were significantly correlated [13]. We plan to develop this finding into an assistive technology that can quickly detect interpersonal problems such as isolation and bullying in today's society, where there are many situations where adult supervision is not sufficient and make people aware of the need to prevent such problems.

5. Pursuing fundamental theories

In October 2021, CS Labs established the Institute for Fundamental Mathematics (IFM) [14], an organization dedicated to researching the fundamental theories of modern mathematics, to accelerate long-term research and development and further strengthen the source of the “fountain of knowledge.” As of July 2024, there are eight mathematicians working at the IFM. Mathematicians from different specialties, such as number theory, algebra, geometry, representation theory, analysis, and dynamical systems, cross and connect with each other to explore unknown mathematical truths and solve important unsolved problems. Its mission is to propose approaches using

modern mathematics to important problems in other interdisciplinary fields, such as physics, biology, and medicine, and to discover new mathematical objects. The September 2024 issue of NTT Technical Review [15] presents recent important research results and perspectives from members of the IFM.

6. Conclusion

In this article, I have presented representative examples of recent research from CS Labs. As technologies such as generative AI become more sophisticated, the importance of human-centered technologies, such as understanding the nature of each person's diversity and how to process and convey useful information according to individual characteristics, will become increasingly important. We will continue to develop other areas of research toward a future in which humans and computers can truly understand each other and in which humans and AI can work together to create a better society.

References

- [1] S. Tsuchiya, H. Nakashima, H. Nakagawa, K. Hashida, H. Matsubara, Y. Osawa, and Y. Takama (eds.), "AI Encyclopedia (2nd Edition)," Kyoritsu Shuppan, 2003 (in Japanese).
- [2] M. Koyasu, "Theory of Mind," Iwanami Shoten, 2000 (in Japanese).
- [3] "NTT Communication Science Laboratories: Elevate Information and Communication to the Level of Human Thought," NTT Technical Journal, Vol. 4, No. 9, pp. 58–59, 1992 (in Japanese).
- [4] K. Ishikawa, "Imaging and Precise Measurement of Sound by Light," J. IEICE, Vol. 106, No. 9, pp. 849–854, 2023 (in Japanese).
- [5] R. Nishikimi, M. Nakano, K. Kashino, and S. Tsukada, "Variational Autoencoder-based Neural Electrocardiogram Synthesis Trained by FEM-based Heart Simulator," Cardiovascular Digital Health Journal, Vol. 5, No. 1, pp. 19–28, 2024. <https://doi.org/10.1016/j.cvdhj.2023.12.002>
- [6] Y. Fujiwara, "Fast Knowledge Discovery from Big Data—Large-scale Data Analysis with Accuracy Guarantee via Efficient Pruning Methods," NTT Technical Review, Vol. 22, No. 11, pp. 36–40, 2024. <https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202411fa4.html>
- [7] S. Takamuku and H. Gomi, "Vision-based Speedometer Regulates Human Walking," iScience, Vol. 24, No. 12, 2021. <https://doi.org/10.1016/j.isci.2021.103390>
- [8] A. Takagi, "The Crux of Human Movement Variability," NTT Technical Review, Vol. 22, No. 11, pp. 41–46, 2024. <https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202411fa5.html>
- [9] S. Kumano and K. Nomura, "Multitask Item Response Models for Responses Bias Removal from Affective Ratings," Proc. of the 8th International Conference on Affective Computing and Intelligent Interaction (ACII 2019), Cambridge, UK, Sept. 2019.
- [10] K. Nakamura, "Towards Reliable Infrastructures with Compressed Computation," NTT Technical Review, Vol. 22, No. 11, pp. 23–28, 2024. <https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202411fa2.html>
- [11] T. Fukiage, "Human-centric Image Rendering for Natural and Comfortable Viewing—Image Optimization Based on Human Visual Information Processing Models," NTT Technical Review, Vol. 22, No. 11, pp. 29–35, 2024. <https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202411fa3.html>
- [12] Press release issued by NTT, "World's First: Discovery of Neural Patterns Strongly Linked to Esports Match Outcomes—Predicting 'similar-level match' and 'upsets' with approximately 80% accuracy," July 18, 2024. <https://group.ntt/en/newsrelease/2024/07/18/240718a.html>
- [13] A. Shinohara, M. Narazaki, and T. Kobayashi, "Children's Affiliation toward Peers Reflected in Their Picture Drawings," Behavior Research Methods, Vol. 55, pp. 2733–2742, 2023. <https://doi.org/10.3758/s13428-022-01924-2>
- [14] Press release issued by NTT, "NTT has established the Institute for Fundamental Mathematics—Advancing the pace of exploration into the unexplored principles of quantum computing," Oct. 1, 2021. <https://group.ntt/en/newsrelease/2021/10/01/211001a.html>
- [15] "Feature Articles: Challenging the Unknown: Mathematical Research and Its Dreams," NTT Technical Review, Vol. 22, No. 9, pp. 16–72, 2024. <https://ntt-review.jp/archive/2024/202409.html>



Futoshi Naya

Vice President, Head of NTT Communication Science Laboratories.

He received a B.E. in electrical engineering, M.S. in computer science, and Ph.D. in engineering from Keio University, Kanagawa, in 1992, 1994, and 2010. He joined NTT Communication Science Laboratories in 1994. From 2003 to 2009, he was with Intelligent Robotics and Communication Laboratories, Advanced Telecommunications Research Institute International (ATR). His research interests include communication robots, sensor networks, pattern recognition, data mining in cyber-physical systems, and AI-based tailor-made education support. He is a member of the Institute of Electrical and Electronics Engineers (IEEE), the Society of Instrument and Control Engineers, and the Institute of Electronics, Information and Communication Engineers (IEICE).