NTT
IOWN Technology Report 2023

– Acceleration to the Future –

CONTENTS

INTRODUCTION

Pioneering New Value and a Sustainable Society

The world we inhabit is undergoing unprecedented change. The COVID-19 pandemic has transformed our lifestyles and work patterns, while the rapid evolution of artificial intelligence (AI) has given rise to generative AI that has reshaped industries. Meanwhile, climate change, marked by a surge in extreme weather events, looms large. To tackle these growing challenges effectively, we must confront each issue individually, and discover innovative solutions.

Digital technologies are undeniably crucial in addressing the issues that confront our society today. Over recent decades, digitalization has made remarkable progress, and resolved numerous challenges.

However, it could be argued that digitalization thus far has largely focused on achieving *efficiency*. The digitalization of existing industries and business processes has often revolved around automation, energy-saving measures, and optimization, yielding substantial gains in efficiency. Yet, is it adequate to employ digital technology merely for efficiency gains? We assert that to unlock the full potential of digital technology, we must promote digitalization not solely for efficiency but to create entirely new values.

This principle led us to unveil the IOWN initiative in 2019, as we pursue the development of photonics-electronics convergence technology. This groundbreaking technology makes use of the optical technology that we have diligently researched and developed since the 1960s, and will usher in a completely new paradigm of information processing by constructing a photonic network.

Since its announcement in 2019, we have dedicated ourselves to developing novel networks and computing technology, while conducting research on photonics-electronics convergence. This effort marks our transition from electronic information processing to photonic-electronic processing. In recent years, we have not merely focused on foundational technologies but have also demonstrated and implemented applications and services across diverse domains, including entertainment and medicine. Collaborating with numerous domestic and international partners, we are not advancing on our own, but striving to spark a widespread transformation in society at large.

The IOWN concept is now taking tangible shape. To address myriad societal challenges and bring a sustainable society to fruition by forging new values, we are steadfast in our pursuit of innovation that transcends limitations.

KICKSTART IOWN

IOWN Takes Off:
Envisioning the Future Society through Light and AI

Since unveiling the IOWN vision in 2019, we have diligently advanced research and development efforts to make IOWN a reality. Finally, in March of this year, we began rolling out APN IOWN 1.0 services, marking the transition from concept to implementation. Let us embark on the journey of IOWN and explore the future society illuminated by *light* and *AI*.

Disruptive Surge in Data and Energy Consumption

To address the multifaceted challenges our society faces today, we must reconsider our approach to information processing. Even though digital technology has offered solutions to many of our problems, the contemporary information processing environment is rapidly approaching its limits.

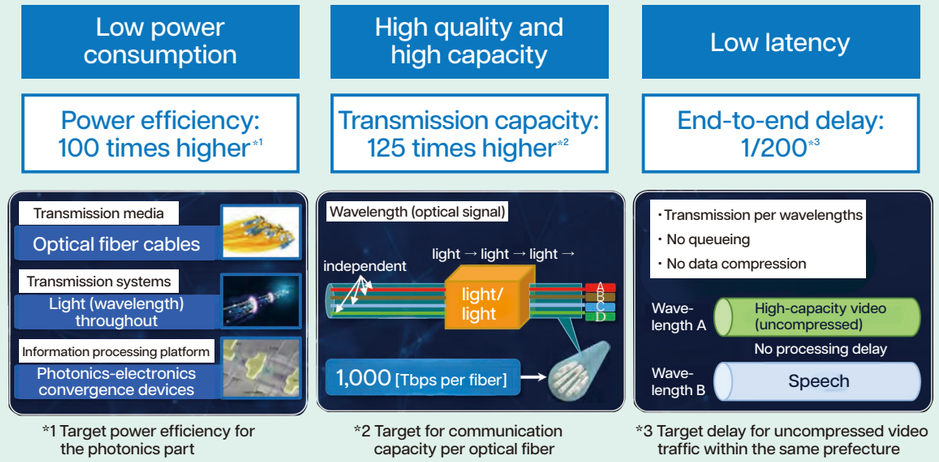
Consider the exponential growth in the volume of data we handle each year. The transition from 1080p HD to 16K video alone is projected to increase data volume by approximately 750 times. Furthermore, the advent of virtual spaces like the metaverse, accommodating not only 2D but also 3D data, is poised to boost data volume nearly 30-fold. When factoring in the Internet of Things (IoT), connecting devices in real-world spaces to the internet, the global number of IoT devices is anticipated to surge from 27 billion in 2017 to 125 billion by 2030. According to a white paper on information and communications from the Ministry of Internal Affairs and Communications, global data volume is expected to surpass 1 yottabyte (1 trillion terabytes) around 2030.

As data volume escalates, so does the energy consumption of data centers, which is projected to be 13 times greater than 2018 levels by 2030. The enormous power demands of AI and cryptocurrencies like Bitcoin, which have seen wider adoption, are also a significant concern. The energy required to train a large language model (LLM) once is equivalent to powering a nuclear plant for an hour. It has been pointed out that Bitcoin's energy consumption in 2022 may rival that of entire countries like Malaysia and Sweden. Digitalization entails the generation and utilization of vast quantities of data, inevitably accompanied by the energy needed for data processing and storage. Unless we address this issue, we will not be able to reap the full benefits of digital society and AI.

From Electricity to Light: The Dream of IOWN

These issues compelled us to turn to optical technology. Electrical circuits consume immense energy when processing great volumes of data, particularly over extended transmission distances. To process copious data within the same timeframe, higher operating frequencies become necessary, causing energy consumption to rise even further. In contrast, light requires minimal power consumption, even with increased transmission distances or operating frequencies. Our extensive research into information processing via light spans several decades, dating back to our pioneering role as the world's first commercial producer of optical fiber for data transmission. Once optical technology transformed how we *transmit* data; we then began exploring its potential application in *processing* it.

Thus, the vision for IOWN (Innovative Optical and Wireless Network) was born. By utilizing optical devices and embedding optical technology in everything from network infrastructure to terminal processing, IOWN promises three significant advantages: high capacity, low latency, and low power consumption. In terms of capacity, transmission capabilities could increase by a factor of 125, while latency measured by end-to-end delay could diminish to 1/200th of present durations. As for power consumption, complementing fiber optic cables with optical (wavelength) through-transmission devices and Photonics-Electronics Convergence elements within information processing infrastructure



*1 Target power efficiency for the photonics part

*2 Target for communication capacity per optical fiber

*3 Target delay for uncompressed video traffic within the same prefecture

Benefits of IOWN

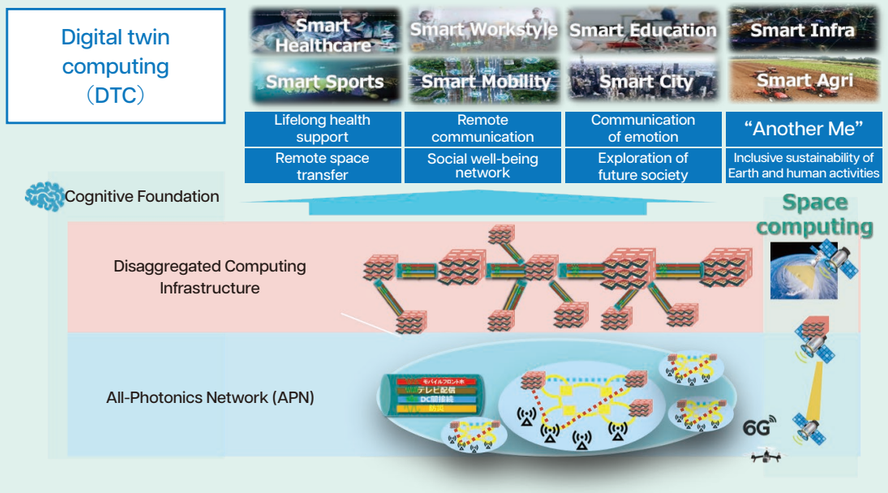
could lead to 100-fold efficiency gains.

In the IOWN vision, we believe that optical information transmission and information processing technologies, integrated with one another through new network architecture, could lead to breakthroughs in the realms of networking and computing and unlock potential new value in the future.

More specifically, our approach proposes a smarter future through the deployment of the All-Photonics Network (APN) and disaggregated computing infrastructure, along with digital twin computing. These components work harmoniously to process, analyze, and utilize vast amounts of data to transform various domains, including healthcare, mobility, entertainment and primary industries. Following the unveiling of the IOWN vision in 2019, we have diligently pursued research and development, culminating in the launch of the APN IOWN 1.0 service in March 2023. IOWN is progressing from the conceptual phase to real-world applications.

Currently, services are primarily tailored to specific industries and contexts. However, as AI and robotics are introduced into more and more applications, industry collaboration via IOWN is poised to accelerate. Domestically, IOWN is expected to encourage digitalization and data utilization not only in manufacturing, urban development, healthcare, and finance, but throughout government services, education, and many other social applications and industries. This trend is not confined to domestic industries; IOWN's impact will extend around the world, as it fosters flexible collaboration across borders through digital twin computing.

Through IOWN, we aim to establish new basic infrastructure that integrates networking and computing, and makes possible a sustainable society brimming with well-being. The concept of well-being, which encompasses physical, social, and mental aspects of happiness and health, may seem unrelated to information infrastructure, but we believe there is an intricate relationship between the technologies underlying various industries and the values underpinning society.



A world created based on optical infrastructure and DTC

In the 1980s, during the heyday of home appliances and automotive manufacturing, technological innovation aimed to produce superior products characterized by heightened performance and user-friendliness. Quality improvement was the focus. Subsequently, in the 1990s, the semiconductor and IT sectors flourished, emphasizing not only *quality* but also *quantity*, measured in the volume and speed of information.

Today, our society demands more than just quantity and quality from information and products; we are in need of a new set of values. This extends beyond the capacity to process vast information swiftly to questions of how information interfaces with people's lives and happiness.

From an era emphasizing material wealth and product quality to an era prioritizing information accessibility and quantity, and now to an era focusing on individual happiness and improving lives – the essence of technology has transformed along with the values shaping each era. We believe that the new computing and network infrastructure heralded by IOWN will contribute significantly to the well-being of individuals.

IOWN's Vital Role in Spreading AI

To ensure the well-being of individuals, we must comprehend the diverse values held by a broader spectrum of people. Additionally, as society confronts increasingly complex and dire challenges, we need to absorb and analyze information on an unprecedented scale in order to forecast future developments. In this endeavor, the utilization of AI plays a pivotal role.

The surging expectations for AI seen across the world today are highly relevant to IOWN. Centralized data collection is an impractical approach to the diverse array of data sources, encompassing not just vehicles and factories but also individuals and the environment, and countless other sensors. Therefore, progress is expected in asynchronous distributed learning, wherein AI systems in diverse domains collaborate and share knowledge. IOWN will serve as the communication infrastructure for such systems, enabling AIs to process vast amounts of information and enhance their interactions.

This shift is particularly relevant to our longstanding commitment to the research and development of natural language processing technology, which enables AI to understand and generate everyday human language. We are leveraging our expertise to accelerate research and development of our own large language model (LLM) in anticipation of AI making even greater inroads in society.

LLMs are a rapidly evolving branch of

the natural language processing field. Progress hinges on expanded data availability, enhanced computational capabilities, and the development of new training algorithms. Notably, LLMs like OpenAI's GPT-4 are gaining widespread adoption in Japan and are expected to have substantial impacts on business.

These advancements have endowed LLMs, including GPT-4, with the ability to interact naturally with humans. However, LLM development has posed ethical and technical challenges. LLMs are susceptible to learning biases from training data, which can result in inappropriate outputs. Furthermore, LLMs' inner workings remain opaque, making it difficult to understand how they generate their output. Resolving these issues will require ongoing research and development.

Although LLMs like GPT-4 exhibit remarkable capabilities, they still have difficulty in seamlessly *collaborating* with humans or growing alongside us as *partners* in our life. We aspire to develop an AI cognitive engine capable of natural collaboration with people across various environments that can contribute to the well-being of each individual.

To achieve this, we will need to develop AIs that have the same interfaces as human beings. We are actively working on a model called VisualMRC (*1) designed to interpret language within web pages visually, much like a human, and SlideVQA (*2), designed to respond to questions based on multiple sets of images such as slideshows. We are also constructing a Japanese visual reading comprehension model. Leveraging these models, we aim to create versatile software robots capable of interactive collaboration with human beings.

Once every person can collaborate with such software as an assistant, the world will witness an increase in communication not only between people and AI but also between AI and AI, as well as between AI and objects. Furthermore, if AI processes not only text data but also all audiovisual information perceived by humans, it will have to continuously handle vast volumes of data in real time. In this context, IOWN will become an indispensable information processing platform, fostering enhanced cooperation between humans and AI, as well as between multiple AIs.

IOWN will not only make possible new computing and networking infrastructure — it will emerge as a fundamental societal infrastructure enabling collaboration between humans and cutting-edge technologies such as AI, underpinning values that will steer society toward a brighter future.

*1 —— VisualMRC: Machine Reading of Images of Documents https://www.anlp.jp/proceedings/annual_meeting/2021/pdf_dir/A5-1.pdf

*2 —— SlideVQA: Asking and Answering Questions Regarding Multiple Images of Documents https://www.anlp.jp/proceedings/annual_meeting/2023/pdf_dir/A11-4.pdf

NEW NETWORK & COMPUTER

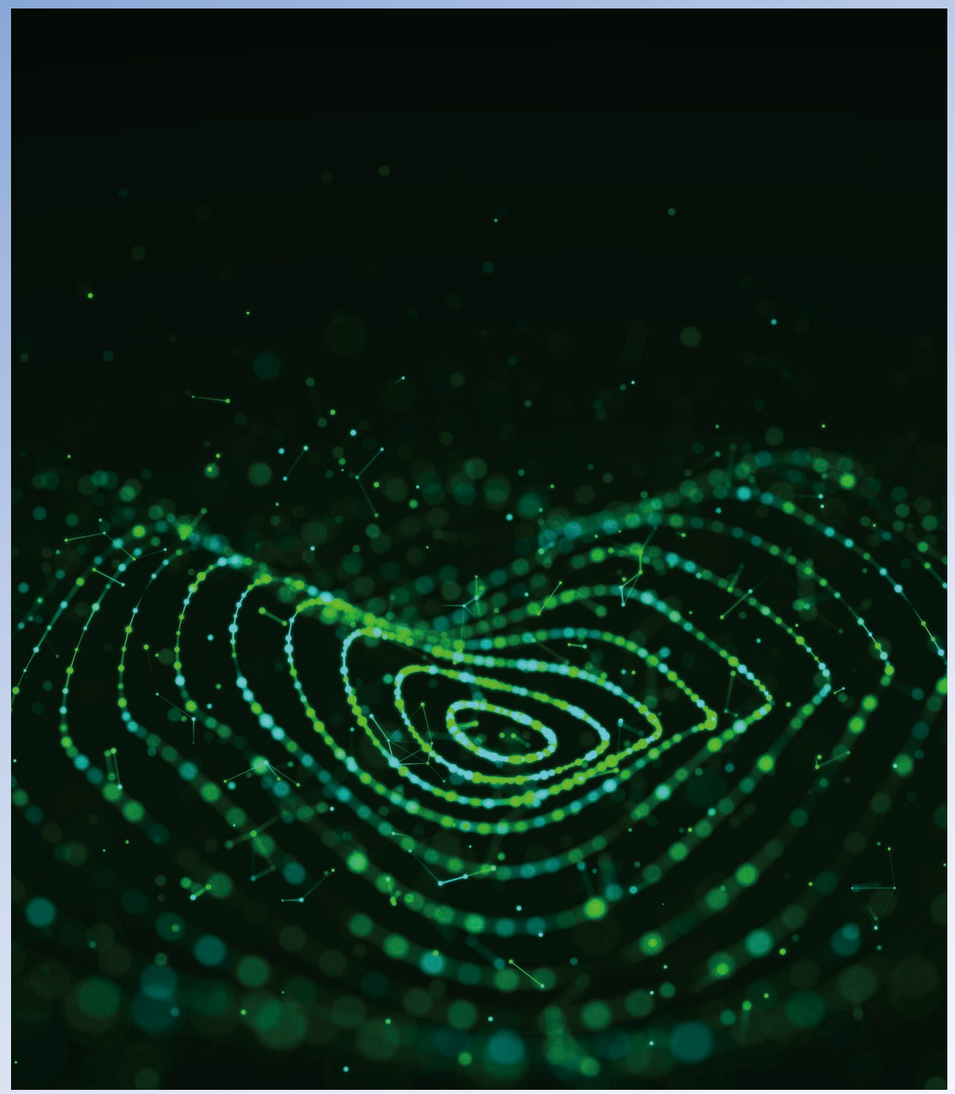
Two Pillars Supporting IOWN

Networks and computing are the two pillars that support IOWN as we begin to deploy the technology. First, in the field of networks, we are developing all-new network technology called the All-Photonics Network (APN). This technology will serve as the foundation for next-generation data centers and will be deployed in various fields that can take advantage of its performance. On the computing side, the core pillar is our development of photonics-electronics convergence devices. We have established a roadmap with five phases for the research and development of this technology, and are now planning to transition to the third phase. As we make progress along this roadmap, optical information processing will penetrate deeper into computers, enabling a new form of information processing that will be an indispensable part of making the IOWN vision reality.

NEW NETWORK & COMPUTER

APN (All-Photonics Network)

APN (All-Photonics Network) is networking technology that enables high-speed, high-capacity, and low-latency communication by connecting everything with light. Since the service was launched in March 2023, we have been making steady progress deploying it for various applications. Following its introduction for medical and entertainment uses, how will APN change our lives in the future?



Converting All Connections to Light

APN (All-Photonics Network) is a new network foundation that makes use of optical communication technology. By transmitting information in the form of light, optical communication technology allows more information to be transmitted in a shorter time than electronic signals. While some may think that our communication is already done mostly over fiber optic cables, in reality electricity must be converted into light during the communication process, resulting in quantitative and qualitative losses.

By converting all connections to direct optical paths, we will be able to construct networks that are far higher capacity, lower latency, and more secure than ever before. It should also be noted that multiple networks can be created over single optical fibers by altering the wavelength. This means APN could be used to create distinct networks for different purposes, such as one using conventional Internet Protocols alongside another using dedicated medical protocols.

In March 2023, APN marked the milestone of its first service deployment as APN IOWN 1.0. At present, it provides 100 Gbps private connections, and users have exclusive use of optical wavelengths from end to end. This on its own leads to a 200-fold reduction in latency, and makes it possible to visualize and adjust latency. The network utilizes three types of equipment: APN terminal equipment, APN-G (gateway), and APN-T (transceiver). Sales of these devices are now underway.

Experimenting with APN Services and Applications in Various Fields

By harnessing optical communication technology that minimizes the need for light-to-electricity conversion, APN promises a substantial enhancement in data transmission speed and a significant reduction in communication latency. The eradication of these delays will have a profound impact across various domains, including VR/AR-based content, AI-mediated communication, and robot control.

In fact, we are actively developing projects across multiple domains. For instance, the “Concert of the Future II,” held in February 2023, showcased real-time remote performance connecting multiple venues, including Tokyo and Osaka. While multi-location performances typically contend with latency of 50 milliseconds (0.05 seconds) or more, this project succeeded in reducing video and audio transmission latency between Tokyo and Osaka to just 20 milliseconds (0.02 seconds). This achievement demonstrates the feasibility of real-time collaboration even in remote settings. In the realm of entertainment, we hosted the first multipoint live comedy/entertainment event, “NTT WEST presents ‘Future Owa-Live!’ supported by Yoshimoto,” connecting three locations in Osaka Prefecture via APN. This demonstration illustrated that comedy performances can

take place remotely without any discomfort, and earned resounding cheers from the audience.

We are also actively collaborating with numerous corporate partners in industries including medicine and finance. In the medical field, we are conducting demonstration experiments of the hinotori™ robotic assisted surgery system, which is operated via an 8K video connection. This project is notable for achieving stable real-time control with feedback from the robot.

In the financial sector, we have proven that it is possible not only to reduce latency, but also iron out latency differences between multiple locations at the microsecond level. Ordinary networks result in different latencies when a main system in Tokyo is accessed from locations such as Sapporo, Osaka, and Fukuoka. In comparison, APN enables the measurement and fine-tuning of communication latency depending on the distance. In the context of high-frequency trading (HFT), financial transactions that are conducted in milliseconds, APN is poised to make trading fairer and more equitable.

APN will not merely support existing services and forms of expression, but also foster new kinds of business and expression that were previously unimaginable.

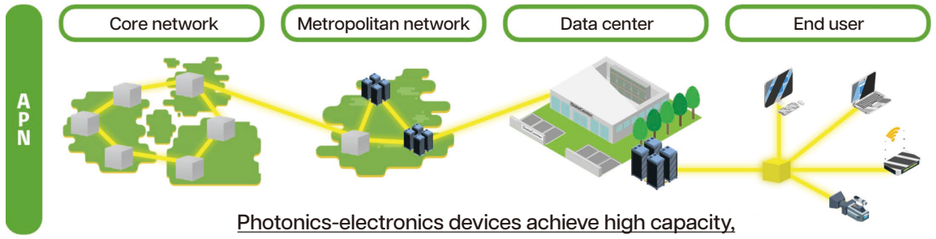
Data Centers Also Rapidly Expanding

APN is also deeply synergistic with our data center business. NTT has an extensive track record of colocation, in which it operates data centers for numerous companies, with the third-largest share in the global market. We have established data centers in over 20 countries and regions worldwide, spanning the APAC region, including locations such as Hong Kong and Singapore, as well as Europe and the United States.

By linking data centers through APN, we can construct an ultra-high-speed, ultra-high-capacity network on a global scale. This network will reinforce data logistics worldwide and increase the accessibility of cloud services. Increasingly, data centers must be more varied and complex, with the need to decentralize data centers to balance disaster risk and to safely manage data close by in order to respond to various problems. The introduction of APN not only enhances our data center capabilities but also ushers in a novel approach to crafting unique communication environments tailored to the diverse needs of various companies.

Progress in the development of APN is marked by advancements in communication technologies and core infrastructure. For example, our open and disaggregated transponder empowers users with a compact device to establish end-to-end optical pathways, free from vendor or system-specific constraints. In the past, creating an optical network necessitated bulky equipment from specialized vendors. However, the miniaturization and openness of optical communication devices will enable companies to construct optical networks more flexibly and effortlessly. In the realm of foundational technology, we are making strides in a suite of optical and electronic devices that interconnect networks through high-speed, high-capacity optical pathways for APN deployment. This includes multi-core fiber for transmitting crucial data with minimal latency and photonics-electronics conversion devices that convert electrical and optical signals with minimal power consumption. These advancements signify that APN is poised for broader application, evolving into a versatile communication network capable of accommodating a multitude of needs.

Specifically, following the launch of IOWN 1.0 service in 2023, development of IOWN 2.0 is planned for 2025. IOWN 2.0 will enable even faster and lower latency communications and is expected to pave the way for new digital services. In addition, new business opportunities are expected to emerge along with wider adoption of APN, including for data centers and growing partnerships.



Photonics-electronics devices achieve high capacity, low power consumption and low latency

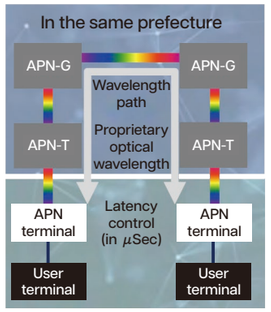
Deployment of APN: System

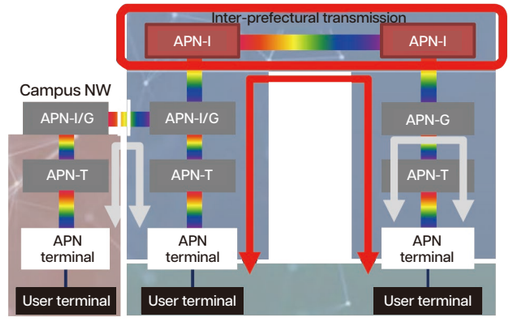
APN Step 1
Service start in March 2023
- Transmission delay: 1/200
- Transmission capacity: 1.2x

APN Step 2*
Inter-prefectural communication
- Transmission delay: 1/200
- Transmission capacity: More than 6x
- Power efficiency: 13x

APN Step 3 and after
Further advancement
- Transmission delay: 1/200
- Transmission capacity: 125x
- Power efficiency: 100x

*Planned announcement at EXPO 2025

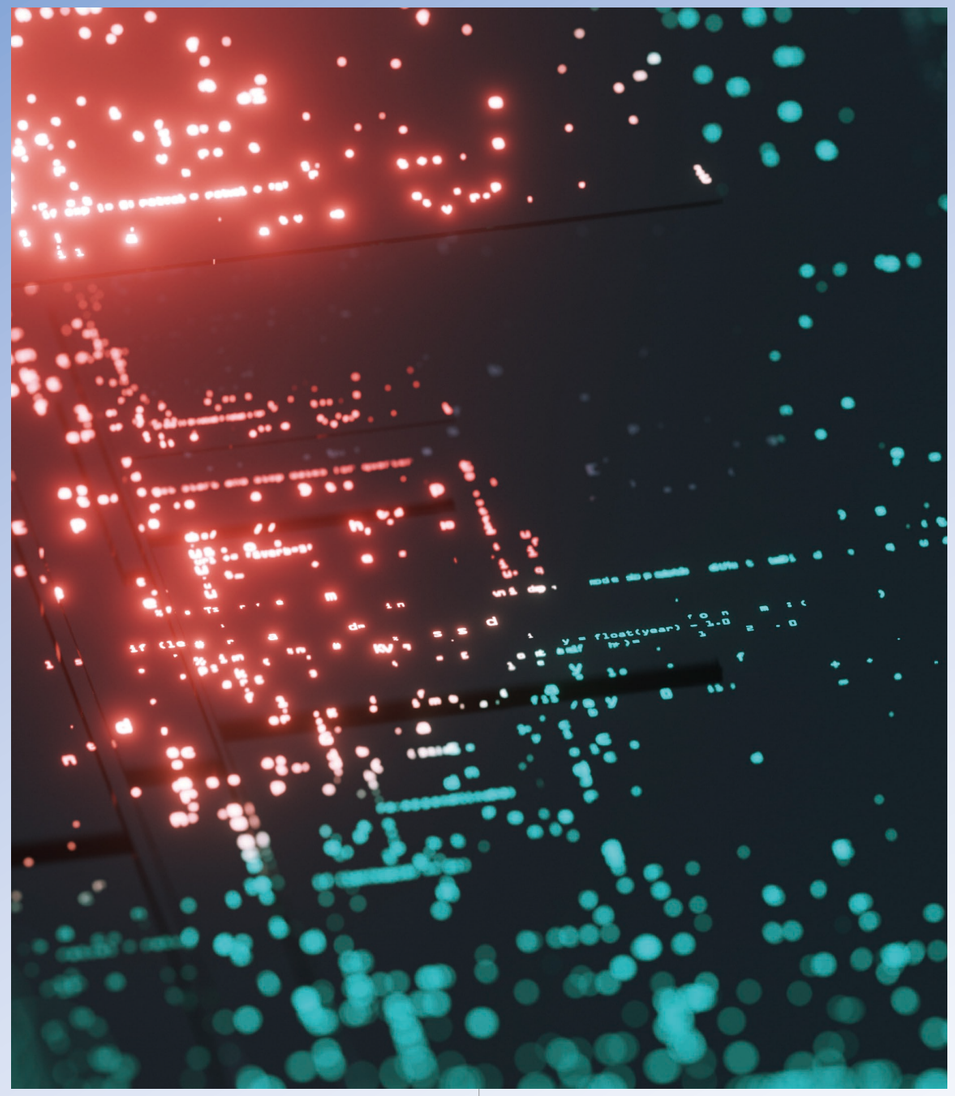




NEW NETWORK & COMPUTER

Photonics-Electronics Convergence Devices

Photonics-electronics convergence devices, which integrate optical technology into devices that were once based entirely on electronics, will enable us to establish a new information processing infrastructure on the level of computing as well. Specifically, what makes photonics-electronics convergence devices distinct from conventional devices, and how will they integrate light and electricity?



Integrating Electronic and Photonic Devices

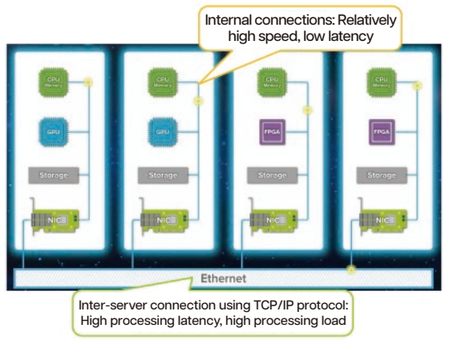
We have pursued research into technologies that integrate light and electricity. For example, in April 2019, NTT published a paper in the British scientific journal *Nature Photonics* on research into optical transistors capable of high-speed operations with ultra-low power consumption. We have continued to study various potential uses for light in addition to electricity.

Our research in this field was also the impetus for the IOWN vision. We are currently pursuing the deployment of photonics-electronics convergence technology as a key step on the road to making IOWN reality. In order to imagine a new form of information processing, we need to make technological strides not only in networking but also in computing. In synchrony with transforming networks through APN, we want to introduce photonics-electronics convergence devices in computing.

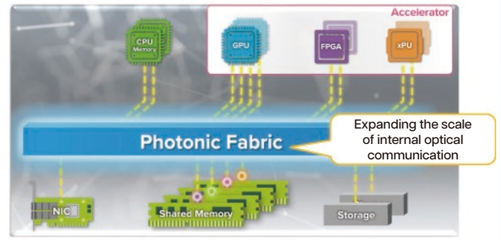
Photonics-electronics convergence devices integrate electronic and photonic elements into a single system to increase the speed of data transfer and improve energy efficiency. This technology will be particularly important for server-to-server connections, communications inside computers, and other applications that require high-speed transfer of large amounts of data.

Evolution of Photonics-Electronics Convergence Devices

Photonics-electronics convergence devices are an essential element of the IOWN vision. As part of IOWN 2.0, photonics-electronics convergence devices for board-to-board connectivity will utilize Optical Engine to enable optical communication between boards, reducing power consumption and improving the performance of devices. This not only reduces the latency of APN, but increases capacity and lowers power



(a) Conventional server configuration



Rack-scale computing beyond the box unit

Exploring original physical configurations, logical configurations, and control methods that leverage rack-scale computing

Physical Configuration | Logical Configuration | Deployment Configuration

(b) Disaggregated Computing

consumption.

Going further, the photonics-electronics convergence devices for chip-to-chip connections that will be deployed as part of IOWN 3.0 will enable optical communication between semiconductor packages by placing the photonics-electronics convergence component next to the silicon inside the package. This will enable even greater board miniaturization and reductions in power consumption. IOWN 4.0 aims to integrate photonics into chips themselves. In short, we seek to achieve board-to-board optical communications with IOWN 2.0 (FY2025), chip-to-chip communications with IOWN 3.0 (FY2028), and integration of photonics into chips with IOWN 4.0 (after FY2032), resulting in devices with 100 times greater power efficiency.

Implementing a New Computing Paradigm

In discussing the deployment of photonics-electronics convergence devices, we should not ignore the role of disaggregated computing. This computing model physically separates computing resources (CPU, memory, storage, etc.) from physical servers and allows these components to be freely combined and utilized over the network. This enables more efficient resource use and greatly increases system scalability and flexibility. Disaggregated computing is playing an important role in enabling new computing paradigms such as cloud computing and edge computing. It could also be useful for applications that need to process large amounts of data at high speed, such as AI and big data analytics.

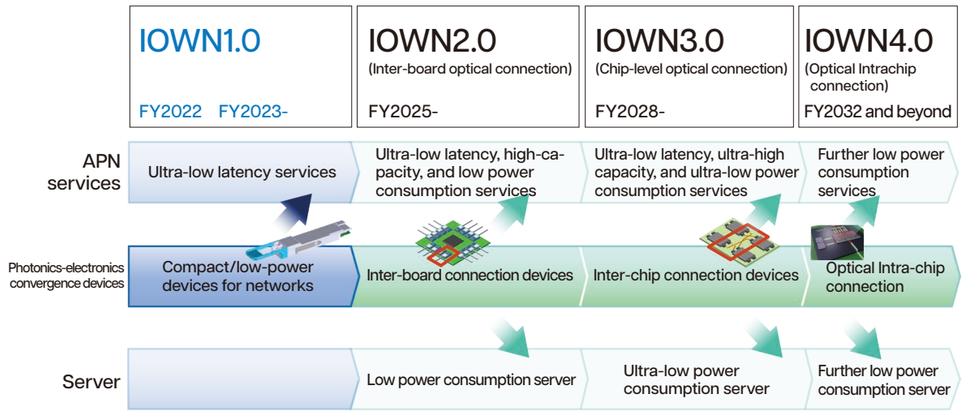
Demand for disaggregated computing is expected to manifest particularly in large-scale data centers and cloud environments. Because many users simultaneously perform a variety of tasks in these environments, resource demands are highly dynamic. Disaggregated computing is capable of flexibly responding to these demands.

Resource-intensive applications such as AI and big data analytics also require powerful computing resources to process large amounts of data. With disaggregated computing, users are able to access resources when they need them, at the scale they need.

Furthermore, in combination with edge computing, disaggregated computing technology could support real-time data processing and low-latency applications. Edge computing reduces data transfer latency and improves real-time performance by processing data close to the device where it is generated. Combined with disaggregated computing, it can maximize the utilization of edge devices to enable high-performance data processing in the edge environment.

We are also developing a new computing platform that employs disaggregated computing architecture. To meet our target of commercial rollout in 2026, we are working on operating systems, controllers, and applications for this platform that will serve as infrastructure powering the IOWN vision.

In June 2023, we founded NTT Innovative Devices Corporation as a manufacturing company specializing in photonics-electronics convergence devices to expedite the transition from concept to reality. Furthermore, to enhance the progress of R&D and the practical application of IOWN-related technologies, including 6G, we are committing approximately 100 billion yen to the overall IOWN R&D efforts for FY2023. We will continue to make investments to accelerate the development of our disaggregated computing platform and services such as digital twin computing.



HOW IOWN CHANGES THE WORLD

Envisioning the World of 2030 through IOWN

As we introduce IOWN 1.0, the proliferation of services and applications harnessing IOWN is poised to shape the future. What sort of world will emerge on the horizon? How will conventional industries and businesses be transformed? To explore the society of 2030, we've gathered experts from three distinct areas: a versatile business platform, an innovative healthcare startup, and a researcher reevaluating the AI-human interaction. We invite you to imagine the societal landscape of 2030 through these three conversations.

1_
APN: Building Resilient Data Centers with APN
Guest: Yoshiki Yanagi (Amazon Web Services Japan)

2_
IOWN for LLM [1]: IOWN Powers New Medical Communication with Generative AI
Guest: Sho Okiyama (Aillis, Inc.)

3_
IOWN for LLM [2]: IOWN Powers Creative Human Future with Generative AI
Guest: Satoshi Kurihara (Keio University)

HOW IOWN CHANGES THE WORLD

APN: Building Resilient Data Centers with APN

Guest: Yoshiki Yanagi (Amazon Web Services Japan)
Interviewer: Yoshihiko Kainuma (IOWN Development Office, NTT Research and Development Planning Department), Maiko Imoto (IOWN Development Office, NTT Research and Development Planning Department)

Yoshiki Yanagi
Senior Solutions Architect, Enterprise Solutions Division, Technology Management Group, Amazon Web Services Japan, G.K. He actively collaborates on various projects with NTT Group companies as a solutions architect in the telecommunications sector.

In this illuminating interview, we embark on a journey into the future of networks and data centers alongside Mr. Yoshiaki Yanagi of Amazon Web Services (AWS). As the world continually evolves, emerging technologies like all-photonics network (APN) stand poised to revolutionize data center operations and reshape the business landscape. We dive deep into how APN tackles existing challenges and the myriad business opportunities it opens up.

Minimizing Environmental Impact

Maiko Imoto: NTT Group currently provides network services linking customer sites with AWS as a delivery partner of AWS Direct Connect. We have engaged in Proof of Concept (PoC) as we explore the diverse ways in which APN can be utilized.

Yoshiaki Yanagi: In my role as a solutions architect, I collaborate closely with customers to find ways that technology can solve their business challenges. Networks are absolutely essential to cloud computing services that enable on-demand use of IT resources via the internet.

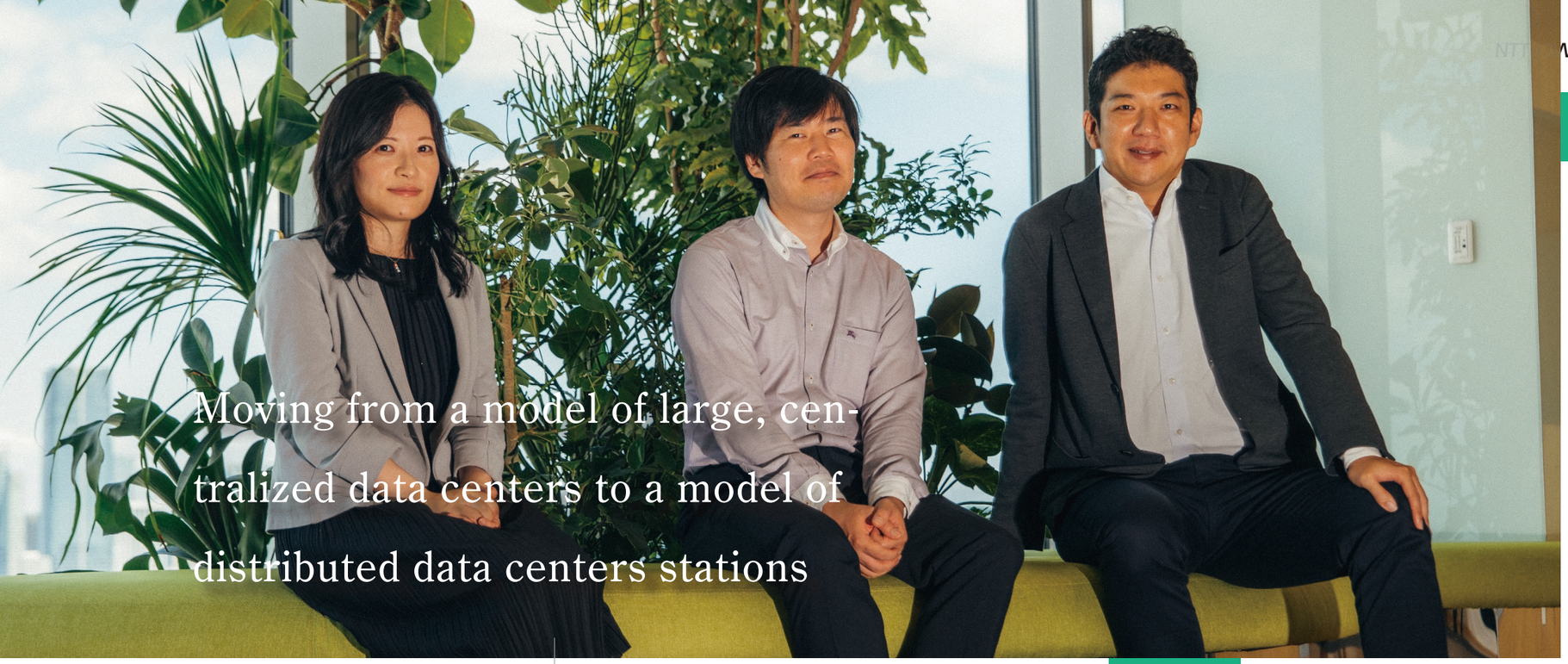
Yoshihiko Kainuma: It is essential to discuss and understand data centers as closely related to networks. As 4K and 8K video has become more and more common and the metaverse and IoT get popular, data volume is increasing exponentially. As a result, data centers are expanding, responding to the need for more network equipment and servers. Simultaneously, power consumption is growing year after year, prompting certain regions and countries to impose limits on new data center construction.

YY: We also hear from our enterprise customers that they care about their carbon footprint and sustainability management. According to a report from the Center for Low Carbon Society Strategy, which operates under the umbrella of the Japan Science and Technology Agency, in 2018, data centers consumed 14 terawatt hours (TWh) in Japan and 190 TWh worldwide.(*1)

If this trajectory continues, data center energy consumption is expected to reach 90 TWh in Japan and 3,000 TWh worldwide by 2030. Rapid improvements in energy efficiency have helped limit energy demand growth from data centers and data transmission networks, which each account for about 1-1.5% of global electricity use. Nevertheless, strong government and industry efforts on energy efficiency, renewables procurement and R&D will be essential to curb energy demand and emissions growth over the next decade. Consequently, many enterprise customers are urgently seeking ways to curb electricity consumption and transition to renewable energy sources.

YK: This is exactly one of the issues that IOWN must tackle.

YY: AWS is helping to reduce power consumption by optimizing its design and



Moving from a model of large, centralized data centers to a model of distributed data centers stations

operation of cloud data centers through the use of low-carbon concrete and the latest technologies in air conditioning systems. This means, according to 451 Research, that just by moving IT workloads to AWS, our customers can decrease their carbon footprint by 77%. Amazon has committed to achieving 100% renewable energy across all its operations, including AWS, by 2025, and is aiming to reach net-zero carbon footprint by 2040, ten years ahead of the goals of the Paris Agreement.

AWS is also committed to supporting customers in building sustainable application systems on AWS. AWS Well-Architected is a set of best practices for designing AWS. When we help organizations study, evaluate, and improve their workloads, sustainability is one of the main pillars along with operational efficiency and security. To take another example, we are working with NTT DOCOMO on a proof of concept utilizing AWS Graviton processors, a family of processors designed by AWS. We were able to demonstrate that it leads to a power reduction of about 70% compared to conventional architectures. I get a clear sense in discussions with customers that sustainability is an important consideration for them to select services they need.

Significance of Eliminating Communication Latency

YY: In addition to these initiatives within AWS cloud data centers, we're also working together with NTT in the field of networks.

YK: This collaboration is integral to harnessing the potential of APN within IOWN. In conventional networks, data-sorting devices such as routers and switches were the cause of increasing both power consumption and latency. In contrast, APN streamlines efficiency by enabling end-to-end connectivity through optical fiber. However, since APN will not entirely replace existing networks, we're diligently exploring the most useful applications of APN in close communication with our partners.

YY: We worked together on a project earlier this year at the tech conference Interop Tokyo that installed an APN on a section of the network between the conference center at Makuhari Messe and AWS's Tokyo Region.(*2) We managed to verify the high speed, low-latency, stable connection, and are excited to think about the future where cloud connections with APN technology could benefit our customers.

YK: While there had only been a few experiments connecting remote sites with APN, these represent significant strides. We anticipate expanding these sections in the future. Over the long term, our goal is to contribute to reduced power consumption and latency in AWS's computing infrastructure through APN and photonics-electronics convergence technologies.

MI: Eliminating latency is critical, especially when transmitting large video files or conducting critical transactions can hinge on delays of mere milliseconds. As we overcome latency challenges, the prospect of streaming VR and XR content from the cloud becomes increasingly realistic.

Building Resilient Data Centers

YK: As improvements in low power consumption continue, it may transform the very structure of data centers. Instead of consolidating massive data centers in a single location, we envision the distribution of smaller data centers across various regions. When interconnected, these smaller centers could be described as "data stations" of a larger network.

YY: AWS's global infrastructure has achieved a combination of low latency

together with high availability, resiliency, and expandability. For example, in Japan we have two regions, Tokyo and Osaka. In the Tokyo Region, we've established four Availability Zones that each contain clusters of data centers. Constructing the system out of multiple Availability Zones is one way of achieving availability and resilience for our customers. At AWS, we clearly define Availability Zones as being physically separated at a meaningful distance. What does "meaningful" mean? When thinking about lightning, tornadoes, earthquakes, and other risks, it is better to have more distance between data centers. However, if they are too far, then we have the problem of network latency. In order to balance these demands, each Availability Zone is located between a few and a hundred kilometers from other Availability Zones.

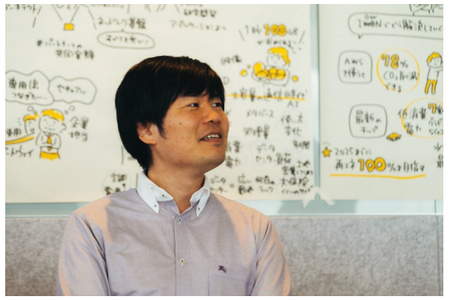
As Mr. Kainuma mentioned, the configuration of data centers is likely to evolve in tandem with the networks.

MI: Our objective is to develop APN technology to create more resilient networks, ensuring that the services and systems relied upon by people in Japan remain uninterrupted. Seamless system functionality at all times is of paramount importance.

YK: General IT systems are typically spread

*1 —— Impact of Progress of Information Society on Energy Consumption (Vol. 2), https://www.jst.go.jp/lcs/pdf/fy2020-pp-03.pdf
*2 —— AWS Region: A geographic area where data centers are clustered. As of September 2023, there are 32 regions around the world, including Tokyo and Osaka. An AWS region consists of three or more Availability Zones (AZs = data center locations), each of which is isolated and physically separated from the others.

Yoshihiko Kainuma (IOWN Development Office, NTT Research and Development Planning Department)



out over extensive geographic areas, but the integration of APN to reduce latency opens up new possibilities.

APN Optimizing Society

YY: When I talk with my colleagues who cover other industries, I am being asked more about NTT's IOWN initiative. I think this shows that customers in other industries are interested in what new social possibilities AWS could unlock in combination with IOWN. While there is certainly great anticipation in sectors such as telemedicine and autonomous vehicles, I also get the sense that the media and broadcasting industries are expecting more for APN. For example, video playback problems occur as a result of the lag of signals from broadcasters or cell towers. This happens all over Japan because the number of router hops and network distances vary at each transmission station, and operators are taking various approaches to addressing these issues. If we could transmit signals synchronized via APN, this sort of problem could be addressed. I think IOWN is a technology that can solve many issues facing society.

YK: I'm confident that APN could meet these expectations. We have been working on PoC trials in the fields of medicine and broadcasting. At last year's NTT R&D Forum, we demonstrated an environment in which remote medical equipment could operate via APN with virtually no latency. We're also actively engaged in experiments of remote production by directly linking studios and stadiums.

MI: Beyond enhancing existing services, we are also eager to tackle challenges such as creating digital twins. For instance, if we can visualize and optimize the flow of traffic and people, we could alleviate congestion and enhance productivity. Reducing wasted time could contribute significantly to overall well-being.

YY: Expanding digital twins to the realm of logistics would enable us to visualize the flow of goods, and develop solutions to food waste and other issues. We will be able to deliver items where they are needed while also decreasing energy consumption. IOWN could potentially allow for the whole of society to be optimized. Of course, AWS remains committed to solving our customers' challenges.

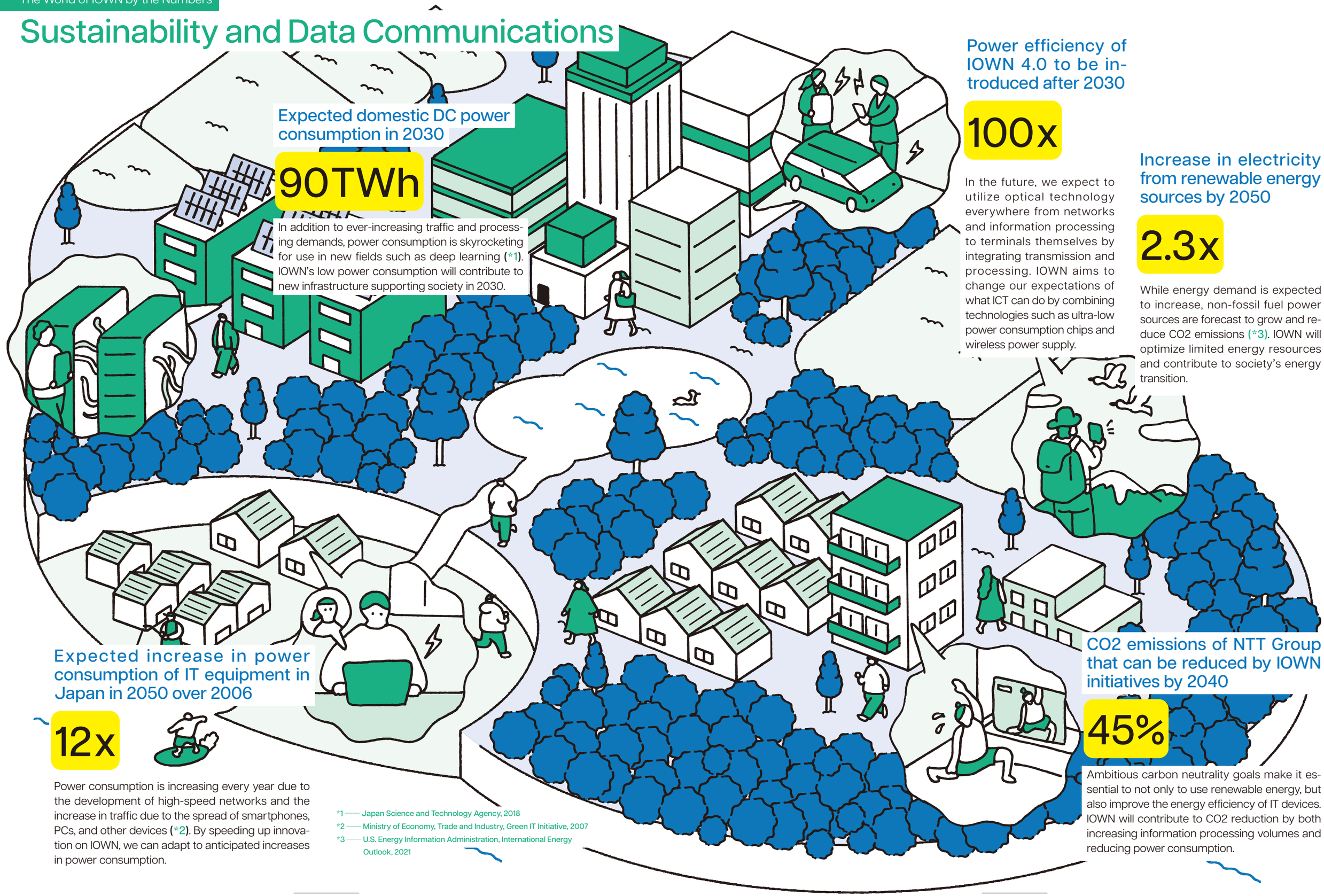
YK: Collaborating with AWS is critically important for NTT to move beyond merely providing networks and work extensively with partners to develop use cases that create new value. True transformative change emerges from collective efforts, not the endeavors of a single company. I look forward to working together with AWS to establish sustainable infrastructure using APN.

YY: We want to work together to think about what kind of value we can deliver to our customers on the other end of their services, and what AWS can do to achieve that. It's a rewarding challenge to think about how new technologies can make society better.

MI: The concepts of digital twins and remote production are not new, but it's becoming clear that they are edging closer to reality through new networks like APN. While it's essential to expand the reach of APN, we also need to research proactive steps we can take to integrate the technology with existing systems. We hope to continue closely working with AWS to tackle various challenges.

Maiko Imoto (IOWN Development Office, NTT Research and Development Planning Department)

Sustainability and Data Communications



Expected domestic DC power consumption in 2030

90TWh

In addition to ever-increasing traffic and processing demands, power consumption is skyrocketing for use in new fields such as deep learning (*1). IOWN's low power consumption will contribute to new infrastructure supporting society in 2030.

Power efficiency of IOWN 4.0 to be introduced after 2030

100x

In the future, we expect to utilize optical technology everywhere from networks and information processing to terminals themselves by integrating transmission and processing. IOWN aims to change our expectations of what ICT can do by combining technologies such as ultra-low power consumption chips and wireless power supply.

Increase in electricity from renewable energy sources by 2050

2.3x

While energy demand is expected to increase, non-fossil fuel power sources are forecast to grow and reduce CO2 emissions (*3). IOWN will optimize limited energy resources and contribute to society's energy transition.

Expected increase in power consumption of IT equipment in Japan in 2050 over 2006

12x

Power consumption is increasing every year due to the development of high-speed networks and the increase in traffic due to the spread of smartphones, PCs, and other devices (*2). By speeding up innovation on IOWN, we can adapt to anticipated increases in power consumption.

CO2 emissions of NTT Group that can be reduced by IOWN initiatives by 2040

45%

Ambitious carbon neutrality goals make it essential to not only to use renewable energy, but also improve the energy efficiency of IT devices. IOWN will contribute to CO2 reduction by both increasing information processing volumes and reducing power consumption.

*1 —— Japan Science and Technology Agency, 2018
*2 —— Ministry of Economy, Trade and Industry, Green IT Initiative, 2007
*3 —— U.S. Energy Information Administration, International Energy Outlook, 2021

HOW IOWN CHANGES THE WORLD

IOWN for LLM[1]: IOWN Powers New Medical Communication with Generative AI

Guest: Sho Okiyama (Aillis, Inc.)
Interviewer: Kyosuke Nishida (NTT Human Informatics Laboratories),
Takanobu Oba (IOWN Development Office, NTT Research and Development Planning Department),
Suzuyo Inoue (R&D Vision Group, NTT Research and Development Planning Department)

Sho Okiyama
Born in 1985, graduated from the University of Tokyo Faculty of Medicine and licensed as an emergency medicine physician. After working at the Japanese Red Cross Medical Center, Okiyama engaged in emergency medicine as a medevac helicopter doctor, a DMAT member of a disaster relief medical team, and a ship's doctor. In 2017, he founded Aillis to develop AI medical devices.

One of the areas most expected to be revolutionized by technology is healthcare. In addition to the spread of telemedicine using ultra high-speed and ultra-low latency networks, AI innovations such as large language models (LLMs) hold the potential to fundamentally reshape the medical industry. We sat down with Aillis CEO Sho Okiyama, whose company is at the forefront of AI medical device development, to delve into the future of medicine — a future not confined to treating ailments but encompassing holistic care.

Deciphering "Ambiguity" to Master AI

Sho Okiyama: I initially pursued a career as an emergency physician, working on medevac helicopters and ships. However, I gradually became interested in the overlap of medicine and information science. This led me to launch Aillis, where I am dedicated to the development and deployment of AI medical devices.

Suzuyo Inoue: Were you interested in the connection between AI and medicine from the time you started your company?

SO: When we laid the foundation for Aillis, AI medical devices had yet to gain approval in Japan. We are building a product that photographs the throat and employs AI to assist in making informed diagnoses, but in the early stages of development, data collection proved to be a challenge. The healthcare industry had not made substantial strides in utilizing data. Today, around 30 such products have been approved, which shows how society has shifted.

Kyosuke Nishida: While the impact of AI diagnoses is undeniable, achieving flawless diagnoses remains difficult. Moreover, people have raised concerns about how LLMs can hallucinate. Introducing AI into fields like medicine, where errors are intolerable, presents particular challenges.

SO: It's important to focus on building trust among individuals, rather than trying to create a perfect AI. Human beings are responsible for the diagnosis. When we introduce our product, we must go beyond technical explanation and think about delivery to hospital physicians and patients.

SI: Digital transformation (DX) seems to be progressing across the industry. Is AI going to be one of the most impactful technologies?

SO: Multimodal AI, comprising big data and neural networks, has structured knowledge in a manner beyond human comprehension. The integration of this knowledge into a user-friendly interface like the chat system was groundbreaking.

KN: What's particularly fascinating is that contemporary LLMs can be expanded to process multiple forms of media. Previously, language and images were analyzed separately, but now they are interconnected. However, addressing human bias inherent

Customizing AI to enhance communication with patients.

in the data used for training remains a concern.

SI: When I myself was involved in healthcare device research, I used to wonder how information should be delivered to customers. People today are surrounded by a vast amount of information, and sometimes consume too much of the wrong kind and end up biased.

SO: We have to be comfortable with ambiguity. For example, a diagnosis of the common cold from a hospital visit does not guarantee a 100% certainty that it is, indeed, a cold. I think the same goes for LLMs like ChatGPT.

New Products Emerge from New Networks

Takanobu Oba: What is your perspective on the role of networks in the future of medicine?

SO: While networks will undoubtedly continue to evolve, the information that human doctors regularly handle in their diagnostic processes is incredibly multifaceted. They take into account not just the patient's verbal communication but also factors like speech speed, breath control, and gait. As we think about future networks, the challenge lies in aligning and integrating such intricate input data.

TO: A network on its own is insufficient. It must go hand in hand with computational capabilities and AI to gather and analyze vast quantities of information effectively.

SO: We now have a theory linking earlobe wrinkles to an elevated risk of heart attack. For a long time, nobody thought of a connection between the ear and the heart. More such correlations that humans cannot notice will surface as data collection continues. If we can process data beyond what humans perceive, the realm of healthcare will undoubtedly evolve.

TO: That's exactly the world we envision in a future of IOWN. For medicine and other fields closely related to life, we aim to create this infrastructure in collaboration with governments and corporations.

SO: In some respects, new infrastructure gives rise to new products. At the same time, technological advances increasingly run into limits imposed by the laws of physics. For instance, while the evolution of networks might suggest the feasibility of remote surgery, even at the speed of light, there remains an approximately 0.2-second lag between Japan and Brazil. That will hinder instantaneous response and thus remote surgery across vast distances will demand a distinct set of skills.

TO: IOWN envisions applications such as remote surgery as well, but as you point out, there are still physical limitations. On the other hand, if the latency time is constant, humans can easily operate the system once they get used to it. So the question becomes how to translate these physical constraints into forms humans can manage.

A World Where Multiple AIs Collaborate

SO: I understand that NTT is also developing LLMs. Could these be customized for individual companies or different industries such as healthcare?

KN: Our aim isn't to create a single monolithic LLM; rather, we envision a society where various LLMs, each with distinct

Kyosuke Nishida (right), NTT Human Informatics Laboratories
Takanobu Oba (left), IOWN Development Office, NTT Research and Development Planning Department

personalities and strengths, work together smoothly. Crafting an all-powerful LLM would prove costly and power-consuming, but combining diverse LLMs can broaden the scope of applications.

SI: What are some conceivable LLM applications within the medical sphere?

SO: LLM's application may go beyond initial diagnoses and surgeries to support ongoing communication. From the patient's standpoint, a conversation with a physician offers limited interaction and happens every few months. An AI operating 24/7 alongside the patient could be valuable in a different way. It could be important to have an AI that accommodates the patient's subjectivity and values rather than always providing objective and optimal solutions.

SI: Effective communication is a crucial aspect of medicine. Developing AI that can grow alongside the patient day by day holds great promise.

SO: For lifestyle diseases like metabolic syndrome, hypertension, and diabetes—ailments that currently constitute a substantial portion of medical costs—improving lifestyle has a far greater long-term impact than taking medication. An LLM that can serve as a lifestyle improvement coach would offer immense value. Additionally, it could help to reduce labor costs and expedite drug discovery processes.

TO: The development of LLM requires substantial financial resources, so in the field of healthcare, we have to think about how to sustain the project while building relationships with devices, healthcare institutions, and other entities.

SO: In the future, perhaps 30 years from now, LLM could evolve into government- or corporate-provided infrastructure. It will be interesting to see if individuals contract LLM services tailored to their lifestyles, like we do today with electricity or gas.

Optimization on a Global Scale Will Transform Healthcare

KN: IOWN will prove to be a key element of LLM infrastructure.

TO: IOWN seeks to enhance information processing efficiency by substituting electricity with light. Traditionally, systems were optimized to respond to demands, but not to minimize energy consumption. Thus, a shift in this paradigm could potentially impact all kinds of infrastructure. From a power consumption perspective, networks, mobility systems, and urban infrastructure today are replete with inefficiencies. Streetlights illuminate unoccupied streets continuously, and trains maintain a five-minute frequency even if they're nearly empty. These inefficiencies ultimately translate into more energy use. Through AI and computing, we want IOWN to achieve optimization at a global scale.

SO: The prospect of global optimization is very encouraging. In the medical arena, individual physicians' skills and experience remain a major bottleneck that could be optimized. Moreover, the energy challenges that IOWN seeks to overcome will become increasingly evident in this field as well. It's plausible that AI will take on the role of information processing, while physicians will assume the primary role of communication.

SI: Could this revolutionize the nature of medical services?

SO: While remote medicine is likely to become more popular, I personally hold high expectations for handling non-visual information. Human information processing primarily relies on visual data, but in medicine, I think its percentage could be as low as 20-30%. Conversational input and tactile feedback yield critical insights.

KN: In my LLM research, I have considered whether AI could convert taste sensations into words. Presently, information processing hinges on visual information, but if we can translate diverse sensory inputs into language, the prospects for remote medicine could expand.

SO: Indeed, converting our personal wording and phrases into common expressions would be immensely valuable. Practically, there are individuals with dementia who struggle to articulate their needs, and I often wonder if more information could facilitate improved treatment. In these cases, family members often act as intermediaries. Outsiders might perceive that the elderly or children don't fully grasp the situation, but family members often have a keen understanding of their loved ones' unspoken thoughts. It would be fascinating if LLM could assist this. Customized AI that enhances communication with individual patients holds immense appeal.

KN: If an AI can assist in conveying things that are hard to put in words, it would undoubtedly be of emotional value to patients as well. We seem to be on the cusp of discovering how AI and humans can work together beyond diagnostic and surgical solutions.

Suzuyo Inoue (right), R&D Vision Group, NTT Research and Development Planning Department

Optimization of Healthcare Accelerated by Communications and AI

Size of global market for AI medical imaging in 2030

1.7 trillion yen

The market for medical treatment using AI-driven image analysis is expected to grow significantly, especially for cancer and cardiovascular diseases (*1). New LLM infrastructure built on IOWN will transform the way doctors work by providing diagnostic support through data analysis.

Japanese government goal for 2030 use of electronic medical records

100%

In accordance with the Japanese government's "Healthcare DX Reiwa Vision 2030," the Ministry of Health, Labor and Welfare is promoting the introduction of electronic medical records, which will be one of the pillars of digital transformation (DX) in the field of medicine. IOWN will play a role in building information infrastructure that handles sensitive personal information through secure communications with APN, leveraging photonic communication to enable next-generation security techniques such as secret computation and next-generation cryptography.

Distance over which latency-free remote surgery was demonstrated over IOWN APN

120km

A demonstration experiment conducted in November 2022 of the Japanese hinotori™ robotic assisted surgery system successfully transmitted 8K uncompressed video securely with a delay of less than 1 millisecond. We are on the cusp of an era of safe and reliable medical care untethered from specific locations.

Shortage of medical, nursing, and other healthcare workers in 2040

960,000 people

The 2022 White Paper on Health, Labor, and Welfare indicates that the population of 20-64 year olds is rapidly declining and there will be a shortage of workers in healthcare-related fields. By optimizing resources through IOWN and LLMs, we aim to build a society where patients can be effectively cared for by fewer medical professionals.

Number of medical devices with AI technology approved by FDA as of March 2022

343

The U.S. Food and Drug Administration (FDA) has evaluated the safety and efficacy of products and approved their export and marketing since 1997, while developing relevant regulations (*2). The low-latency and secure communications made possible by IOWN will help enable conditions around the world where people can receive data-driven medical care.

*1 The Nihon Keizai Shimbun, December 6, 2022
*2 Japan External Trade Organization, March 2022

HOW IOWN CHANGES THE WORLD

IOWN for LLM[2]: IOWN Powers Creative Human Future with Generative AI

Guest: Satoshi Kurihara (Keio University)
Interviewer: Kyosuke Nishida (NTT Human Informatics Laboratories), Takanobu Oba (IOWN Development Office, NTT Research and Development Planning Department)

Satoshi Kurihara
As a Professor at Keio University's Faculty of Science and Technology and the director of Center of Advanced Research for Human-AI Symbiosis Society, Satoshi Kurihara has embarked on extensive research in the fields of artificial intelligence and complex network science. He has previously worked at NTT Basic Research Laboratories and the Institute of Scientific and Industrial Research, Osaka University, and The University of Electro-Communications before assuming his current role in 2018. He also serves as a Project Lecturer at the Artificial Intelligence eXploration Research Center at the University of Electro-Communications and as an advisor to the Ethics Committee of the Japanese Society for Artificial Intelligence.

In an era marked by the proliferation of Large Language Models (LLMs) and other cutting-edge AI technologies, we find ourselves reevaluating the web of connections between AI and humans, AI and AI, and indeed, human and human. While generative AI has garnered its share of skepticism, we must ponder how AI will reshape human communication and creativity. In this conversation with Satoshi Kurihara, an AI researcher at Keio University, we delve into how AI's relationship with humans should be cultivated in the future.

Leveraging AI to Catalyze Creativity

Satoshi Kurihara: My primary focus has been on various projects with the Japanese Society for Artificial Intelligence (JSAI). People have been particularly interested in a five-year research project we are collaborating with Tezuka Production on, with support from New Energy and Industrial Technology Development Organization (NEDO). Currently in its fourth year, we are working with various creators on a sequel to the manga series *Black Jack* , titled *TEZUKA2023* , in order to explore how AI can be harnessed to mass-produce superior creative works.

Kyosuke Nishida: Creating a sequel to a renowned work like *Black Jack* must be quite challenging

SK: Yes. Currently, many people are apprehensive about generative AI. However, I believe that if we can create a work akin to *Black Jack* with the assistance of AI, it could have a profound impact.

Takanobu Oba: The emergence of valuable content could transform the perception of AI globally. Because skepticism of generative AI is so strong in Europe and elsewhere, I think Japan has a role in producing exceptional artistic expressions.

SK: In the European Union, lawmakers are already discussing AI regulation, while the United States has been debating licensing restrictions for AI-based services. Major corporations like NTT have a significant role to play.

KN: What possibilities lie ahead for creators who engage with these services?

SK: One of our fundamental concepts is *interactivity*. We envision a future where AIs interact with human beings as assistants or friends, rather than something that generates answers to simple prompts. For instance, by developing an AI that produces complex prompts that will help create stories, we can develop a tool that fosters interaction between ChatGPT and creators. Having AI as a brainstorming partner accelerates creation when compared to having to start from scratch. What might appear as fragmented pieces of a narrative to the general public can serve as a wellspring of creativity from a creator's perspective.

KN: A notable innovation with LLMs is their capacity to generate content that has

Thinking about the future of AI means facing the complexity of human beings.

not been learned directly. It can introduce serendipity into the creative process. While there is concern that AI might become excessively convenient and stifle human growth, our objective is to create an environment where AI and humans mutually amplify each other.

Diverse AI Leads to Emergence

TO: When we consider AI, the significance of networks cannot be understated. Presently, AI already formulates content and advertising recommendations based on data gleaned from networks. Networks and AI already exert a lot of influence on human decision-making. One consequence could be the accelerated social fragmentation like we have seen in the United States. However, I foresee change on the horizon as AIs collaborate more extensively. Currently, our communication primarily relies on natural language, which tends to engender deliberations that are slow and ambiguous. If AIs can engage in rapid and sophisticated democratic dialogues, I believe it will catalyze transformation within human society.

KN: NTT also envisions a future where small AI entities with diverse personalities interconnect in a multi-agent manner, evolving in tandem with human beings.

SK: When we discuss LLM performance, we often fixate on the number of parameters. Yet even LLMs with a relatively modest parameter count can achieve noteworthy improvements in quality when models are interconnected via high-speed networks. This underscores the importance of networks like IOWN. Collaboration between a multitude of AIs creates a large potential for unprecedented ideas. How much control humans can exert in such scenarios remains uncertain, which is a source of apprehension.

KN: One of the challenges for LLMs is continual growth. If AIs can manage to cooperate and evolve in conversation with humans, they should develop rich character tailored to each individual. Additionally, knowledge acquired by one LLM might find utility in other domains. Nevertheless, it remains difficult for AIs to learn in ways that align with human sensibilities.

TO: It may not be essential to create an AI that rigidly adheres to human values, but, at least for the time being, we have to stick close to them.

SK: Some may harbor concerns about humans being dominated by a colossal AI entity, but a diverse AI landscape could offer enhanced security and safeguard democratic values more effectively.

Infrastructure is Vital to Manage Complex Information

Kyosuke Nishida (NTT Human Informatics Laboratories)

KN: How does *TEZUKA2023* approach the analysis of *Black Jack*? While ChatGPT has acquired image-processing capabilities, does it fully grasp the nuances of characters' emotions and backgrounds?

SK: Initially, humans read the manga and painstakingly transcribe it into text form. The narrative, character descriptions, and other relevant details are transformed into text. Moreover, we collaborate with a professor specializing in narrative analysis who oversees the process of converting individual stories into textual format through the collective efforts of multiple individuals.

KN: The content generated must vary considerably based on the comprehension of the individuals transcribing the story. This underscores how in our daily decision-making, we rely on a multitude of information sources, not merely textual dialogue.

TO: We use language because it is easily managed by humans, but in reality, the amount of non-verbal information we gather far exceeds that gained by language. If this sensory information were collected together, it could unleash new innovation.

SK: For humans, non-verbal cues like facial expressions and physicality carry immense significance, and language alone may fall short. However, processing information from all five senses poses formidable challenges. The influence of infrastructure, such as IOWN, looms large in our ability to manage increasingly complex information.

TO: The internet has interconnected diverse elements, yet it is also subject to slowdowns and interruptions. I envision a world where IOWN eliminates delays, potentially reshaping the landscape of AI.

Will AI Put Up With the Difficulty of Humans?

KN: As AI becomes more prevalent, there's a surge in information different in nature from what we have seen before. Information becomes richer, but concerns about its accuracy rise. For instance, if the number of blogs and news outlets disseminating misinformation produced by ChatGPT increases, AIs could grow even more error-prone as they are trained on flawed data. As a researcher, I'm interested in how we can break this cycle. We need to think about how to make society more robust even as we embrace AIs.

SK: ChatGPT's imperfections are oft-cited, but I wouldn't characterize them as errors; rather, the system is simply generating answers as designed. It is not designed to assess the correctness of information. LLMs excel at summarizing and converting text, and their knowledge may be explained as an inadvertent byproduct of training. LLMs possess a degree of logical reasoning ability, but in the end they have been trained with logical text produced by human beings. Dual process theory in psychology separates human cognition into system 1, or intuition, which is quick and automatic processing, and system 2, or reasoning, which is slow intentional processing. Responses by LLMs currently reflect system 1 only. A trustworthy AI should possess the capacity to respond with the interlocutor's perspective mindfully, and transcend mere conditioned reflexes. At present, AIs remain a tool, but they are going to develop into autonomous thinkers. If we can nurture AIs with a sense of motivation, we may foster a trusting relationship with them.

KN: Right now, LLMs are designed to follow objectives set by humans, but if they can autonomously establish their goals, they will attain true autonomy. AI may consequently shoulder responsibility, potentially becoming a source of resentment among humans, or AI may develop its own desires. Conversely, many contemporary LLMs rely heavily on data gathered from the web called the Common Crawl, which is surprisingly lacking in character. This could shift as proprietary data sources open up. For example, using data specific to a company could generate unique characteristics. The significance of training data is likely to increase in AI's future development.

SK: LLMs optimize for predetermined objectives, but humans possess multifaceted goals. If we solely focus on a company's profitability, it might seem prudent to downsize the workforce. However, considering the well-being of employees, it might be more beneficial to avoid restructuring. Additionally, human beings are inherently irrational, and making emotionally-driven, seemingly irrational decisions can lead to happiness. I don't believe a monolithic AI entity can effectively handle the intricate functions by which humans pursue their goals. As depicted in *Black Jack*, humans are inherently difficult to get along with, and acknowledging this complexity might illuminate the path forward in AI development.

Takanobu Oba (IOWN Development Office, NTT Research and Development Planning Department)

*1 —— Mitsubishi UFJ Research and Consulting, October 2022
*2 —— Ministry of Internal Affairs and Communications, 2023
*3 —— The Nippon Foundation, September 1, 2023
*4 —— Center for Low Carbon Society Strategy, Japan Science and Technology Agency, February 2023

Number of creators in Japan in 2022

8.22 million

The creator economy is growing as digital platforms become widespread (*1). Generative AI is reducing the hurdles to output. Emergent behaviors between AIs enabled by IOWN are certain to transform what it means to be a creator.

Latency between Tokyo and Osaka using APN IOWN 1.0

8 milliseconds

With APN, latency is 1/200 of conventional networks (verified in a demonstration experiment in February 2023). This is similar to latency when separated by 3 meters in the same space. The proliferation of low-latency, high-capacity communications over APN could stimulate various creative activities, including collaboration across space.

Traffic volume in 2030 compared to 2018

16x

As general-purpose and generative AI is integrated into business, traffic in the world's data centers is expected to soar (*4). IOWN will enable infrastructure that allows many AIs to work together without human intervention.

Estimated size of the global market for generative AI in 2030

14 trillion yen

The market for generative AI is expected to grow at close to 30% annually, expanding 14-fold from 2022 to 2030 (*2). By enabling high-capacity communication, IOWN will prove its value in a market that is expanding from text and images to more complex information such as video and 3D.

Percentage of 17-19 year olds in Japan who have used generative AI in 2023

36.1%

The number of young people who have used generative AI is not small (*3). Creativite activities that require serendipity are sure to change in new ways as a generation grows up taking communication with AI for granted, and as social infrastructure is built on next-generation networks enabled by IOWN.

The World of IOWN by the Numbers

AI Transforms Infrastructure and Human Creativity

DTC IS A NEW PLATFORM FOR URBANISM

DTC Offers the Infrastructure to Enrich Urban Life

Digital twin computing (DTC) is one of the applications being developed to create value by utilizing IOWN. This is a system that allows users to freely combine multiple digital twins in order to precisely reproduce comprehensive simulations of people and vehicles in an urban environment that were not previously possible, thereby enabling future predictions and personalized services. Already, real-world objects such as automobiles and factory machinery are digitally reproduced to improve real-time situational awareness and detect anomalies. DTC will combine these models in countless ways to reproduce the real world at large scale and high precision. It will even go beyond the reproduction of the physical world to entail virtual representation of interactions with internal aspects of humans.

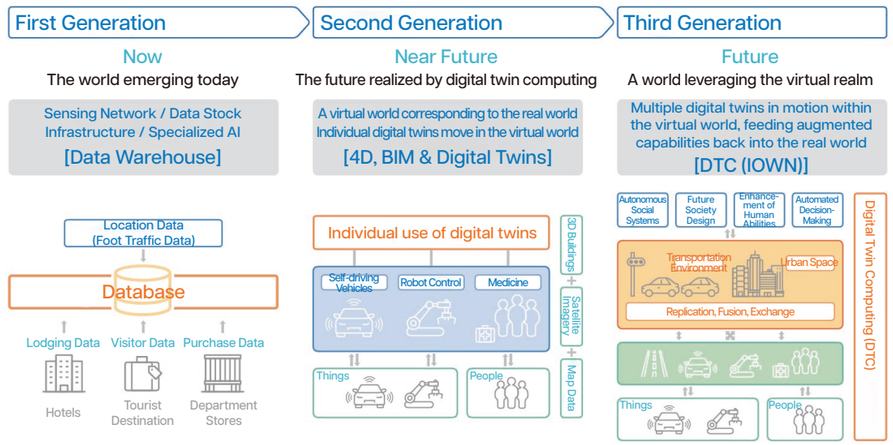
There are numerous potential use cases for DTC, but we are currently rolling out services and conducting experiments mainly in offices and commercial facilities. Among these, APN IOWN 1.0 is already being deployed in a major development being undertaken by Tokyu Land Corporation in Shibuya. This collaborative project between Tokyu Land Corporation, NTT, NTT DoCoMo, and NTT East aims to integrate cutting-edge convenience and sustainability into urban development in ways that address environmental and other social concerns.

The first step in this project is the introduction of IOWN services to Tokyu Land Corporation's urban developments in the greater Shibuya area. Beginning in November 2023, APN IOWN 1.0 will be rolled out at the Tokyu-owned block of the Shibuya Sakura Stage, setting the stage for various services to be offered in the office and event space.

What specific services are anticipated? Let's consider office space first. The introduction of ultra high-speed communications in office environments will enable remote meetings and the exchange of large volumes of uncompressed data with no delay, even with far-away locations abroad. In remote meetings, it is also possible to integrate AI-based automatic translation into the communication process over IOWN. Real-time automatic translation will enable teams to collaborate on a wider global scale without worrying about language barriers. Communication through enormous high-resolution displays and 3D images will also be possible with almost no latency, making communication something completely different from the past. Of course, we must not forget that IOWN's disaggregated computing technology will make this communication possible with exceptionally low power consumption. Richer communication will also be greener.

From a health perspective, DTC's ability to track individual health status in real time will also transform the office experience. If IOWN can connect office gyms to remote locations, people could take advantage of personal training without leaving the office, and companies could offer their employees more opportunities for well-being. Moreover, DTC will enable simulation of the health status of workers in a building to make recommendations via applications. By suggesting appropriate diet and exercise, DTC could help people achieve a healthy work style that is tailored to their specific office environment.

Once DTC is extended to commercial facilities as well as office spaces in the future, remote concierge support and shopping using XR technology could enable highly satisfying shopping experiences tailored to the tastes and preferences of each individual. Foreign visitors to Japan would be able to communicate using real-time translation, and entertainment such as e-sports will become more popular across Japan and abroad. In the urban development of the future, DTC may become the infrastructure for providing people more pleasant experiences and achieving a better way of life.



APPENDIX

INTRODUCTION to NTT R&D FORUM 2023

The NTT R&D FORUM provided an easy-to-understand summary of the latest findings from NTT Group R&D, encompassing specific services, systems, use cases, and elemental technologies for IOWN, which commenced service in March of this year. This year's forum introduced a plethora of initiatives categorized into four distinct areas: "IOWN Pickup," which introduces AI technologies by NTT Group, "IOWN Now," which is already being implemented as a service, "IOWN Evolution," poised to catalyze innovation across all industries, and "IOWN Future," designed to contemplate the future of society from a more long-term perspective.

For more detailed information, kindly visit the NTT R&D website. Please scan the QR code provided below for direct access.

IOWN Pickup

NTT version Large Language Models
NTT, the world's leading company in the field of language processing research, will unveil NTT's version of LLM at this R&D Forum, bringing together the best of its technologies.

IOWN Now

APN (All-Photonics Network)
We will introduce new use cases utilizing APN IOWN 1.0, which is now available.

Next-generation Computing Infrastructure (Super White Box)
We will introduce the progress of next-generation computing infrastructure for IOWN 2.0.

IOWN Evolution

On-Demand Type All-Photonics Network
On-demand optical paths enable ultra-realistic video transmission and ultra-high speed data transmission, contributing to deployment of APN into social infrastructure.

New Value Creation in Wireless Network
We will create wireless network services that meet extreme requirements, integrate advanced multiple wireless access networks, and generate needs for new use cases.

Realization of Fusion and Cooperation by Network Evolution
Toward the 6G generation, we will realize real-time, remote-world futuristic services by coordinating and optimally controlling all terminals/NWs/services.

Smart Infrastructure
We contribute to realizing a sustainable management of social infrastructures by sharing data, manpower and technologies.

Network Operations, Robust Network
This area introduces operational technologies that are robust to failures, disasters, fluctuations in power supply and demand, etc. for ICT resources such as NW.

Project Metaverse
– Fusion of Real and Virtual –
We will introduce our efforts toward realization for a new means of communication that extends the well-being of diverse people and societies based on humanity.

Technology that Supports Individuals with Information from Their Brains and Bodies
Using brain and body information, we can see things that were previously invisible. We support the solution of diverse individual issues with the latest technology that utilizes all kinds of biometric information.

Well-being・Lifelong Health Support
We contribute to the Well-being of people and society through "Bio Digital Twin" which predicts people's physical condition in the future, and "Social Well-being" which supports a better way of being for individuals and groups.

Smart City that Optimizes the Whole
We will introduce urban development that creates advanced value by chaining AI, and value creation with mobility and service robots using a 4D digital platform.

Data Management・Software
We will introduce our efforts using state-of-the-art cryptography and software management techniques.

IOWN Future

Information Processing Technology to Enrich Society
We challenge the information processing infrastructure that contributes to the rapid and efficient resolution of social issues, as well as the device technologies necessary to realize this.

Sustainable Technology to Nurture the Earth
We work on environment, energy, and space technologies to nurture a safe, comfortable, and sustainable earth.

Medical, Healthcare and Human Augmentation Technologies to empower People
We promote research on medical, healthcare and human augmentation technologies to enable people to be active and vibrant.

Network Technologies Supporting Future Information Distribution
We challenge innovative network technologies to support various information distribution in the future and to realize rich communication.

NTT R&D Website
https://www.rd.ntt/e/

Nippon Telegraph and Telephone Corporation
Research & Development Marketing Division
Research and Development Planning Department

technology_report-ml@ntt.com
Published 22 November 2023