NTT Human Informatics Laboratories: Researching and Developing Technologies That Nurture True Humanity

Kota Hidaka

Abstract

Based on the human-centric principle, NTT Human Informatics Laboratories is engaged in research and development related to new forms of co-existence between the real world and cyberworld. In the Feature Articles in this issue, we introduce NTT Human Informatics Laboratories' latest endeavors.

Keywords: human-centric, humanity, human functions

1. Mission of NTT Human Informatics Laboratories

At NTT Human Informatics Laboratories, our mission is to "research and develop technologies that nurture true humanity" while our vision is to "enable the information-and-communication processing of diverse human functions based on the human-centric principle." Of the characteristics that humans possess, we focus on six: sense perception, sensitivities, thoughts, behavior, body, and the environment. We are developing elemental technologies that convert these characteristics into data and use them in information-and-communication processing. To actualize the Innovative Optical and Wireless Network (IOWN) concept, we are using the above human-centric functions in the integrated research on Digital Twin Computing (DTC) and the remote world.

When we note the changes surrounding us, the following comes to the forefront: the emergence of generative artificial intelligence (AI), greater miniaturization and accuracy of brain-computer interface devices, disillusionment with the metaverse and Web3, and post-capitalism. In consideration of these trends, NTT Human Informatics Laboratories has extracted the following actions to focus on: (1) accelerating research with general-purpose AI that uses the brain as a black box, (2) initiating research with general-purpose AI that uses the brain as a white box, (3) pursuing the metaverse's essential and universal value, and (4) accelerating research directly tied to the humanities. On the basis of these areas, we have set large language models (LLMs) and neurotech/cybernetics as priority elemental technologies to accelerate our applied research. We have established "Project Metaverse" and "Project Humanity" as priority use cases that present the application of integrated research (**Fig. 1**).

2. Priority elemental technologies and use cases

We describe NTT Human Informatics Laboratories' directions of four priority elemental technologies and use cases. The first priority elemental technology/use case, LLMs, shows the reality that is possible with general-purpose AI, as exemplified by the arrival of OpenAI's Generative Pre-trained Transformer 3 (GPT-3) and its use in a variety of fields. We are researching and developing a proprietary LLM that will analyze the mechanisms of the brain by leveraging NTT Human Informatics Laboratories' decades of research in natural language processing.
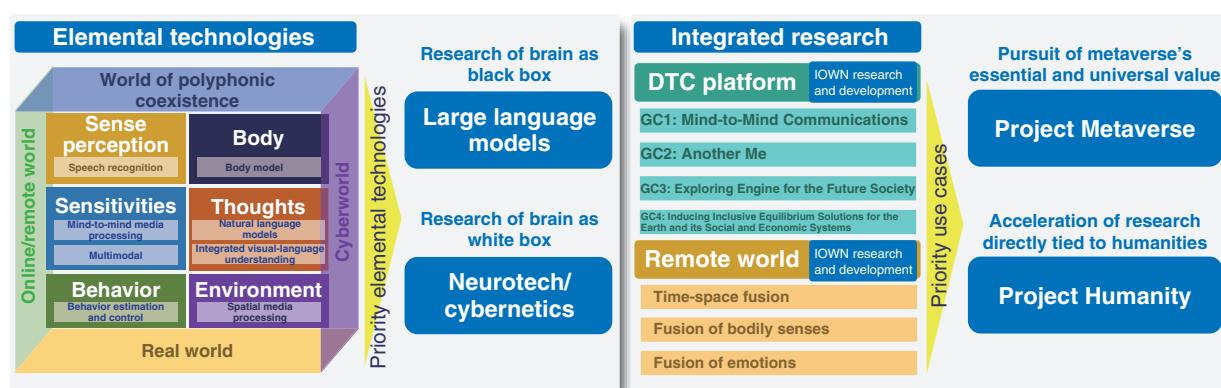
Fig. 1.   Priority elemental technologies and use cases.

By quickly implementing general-purpose AI in society, we will be at the forefront of the new AI era, as well as take on the challenge of innovating the world with general-purpose AI.

For the second priority elemental technology/use case, neurotech/cybernetics, we will use the LLM we have developed, in addition to leveraging knowledge of the workings of the body that we have cultivated to date, to acquire tacit knowledge and develop intuitive human-machine interfaces.

For Project Metaverse, our third priority elemental technology/use case, with our LLM as the base, we will create a new form of "tele-" that transcends distance and time to provide encounters not considered possible before, work and leisure experiences that do not currently exist, and space to know oneself as a human being.

The fourth priority elemental technology/use case, Project Humanity, uses neurotech/cybernetics as the base to promote diversity and inclusion. It also provides ways to support people around us, such as family, friends, co-workers, and people with illnesses and disabilities as well as those supporting them by creating a world that can acquire human functions while respecting the wishes of the user (**Fig. 2**).

In this issue's Feature Articles, we report on NTT Human Informatics Laboratories' latest research and development (R&D) efforts on the above priority elemental technologies/use cases and activities to apply them. In this article, we introduce NTT's proprietary LLM.

3.   NTT's proprietary LLM "tsuzumi"

The arrival of LLMs, such as GPT, has increased the feasibility of general-purpose AI. In November 2023, NTT Human Informatics Laboratories released "tsuzumi," NTT's proprietary, highly efficient LLM. This LLM addresses the issues faced by GPT: language-model size, reliability of information, model extensibility, applicability to non-language modalities, and power consumption associated with large-scale training. The following features differentiate tsuzumi from other LLMs: (1) compact language models (reduced cost), (2) superiority in Japanese language processing, (3) improved customizability, and (4) multimodal support (LLM with physical sensory capabilities).

Regarding tsuzumi's compact language models (reduced cost) (Feature 1), tsuzumi is available in two versions: ultralight with 0.6 billion parameters and light with 7 billion parameters. These versions can run high-speed inference with just one central processing unit (CPU) and one inexpensive graphics processing unit (GPU), respectively. In terms of the cost of using the GPU cloud for training, tsuzumi's ultralight version costs 1/300 and the light version costs 1/25 that of GPT-3 (175 billion parameters). For inference processing, the ultralight version costs 1/70 and the light version costs 1/29 that of GPT-3 (based on NTT's estimations).

The tsuzumi LLM's superiority in Japanese language processing (Feature 2) is due to NTT Human Informatics Laboratories' accumulated expertise from decades of research in natural language processing. The result is tsuzumi's high performance even with a small parameter size. In benchmark testing of LLMs using Rakuda, tsuzumi outperformed GPT-3.5 and the top Japanese LLMs (**Fig. 3**).

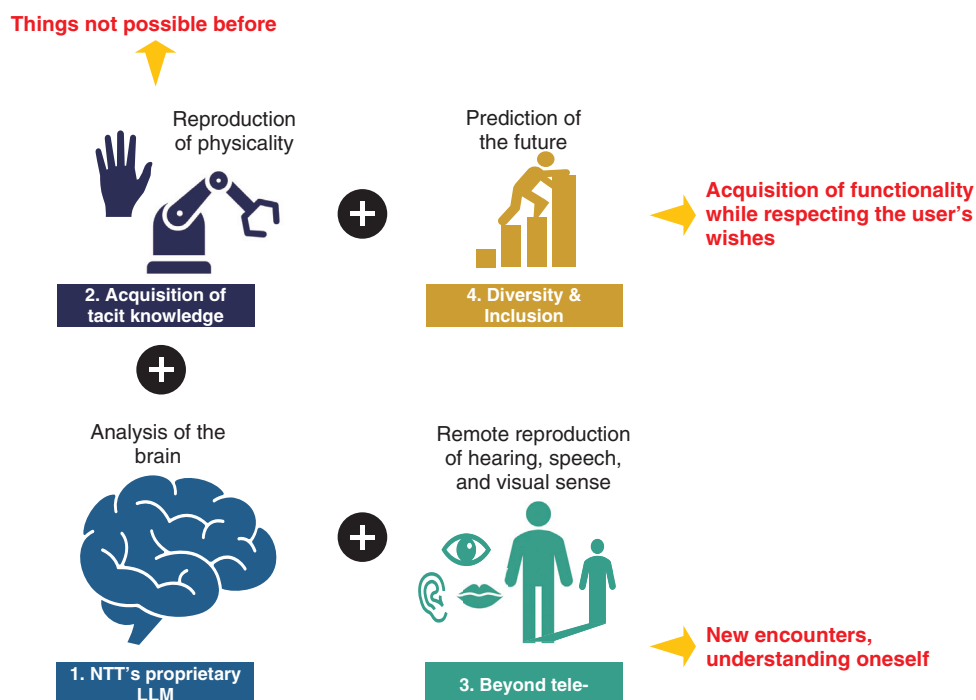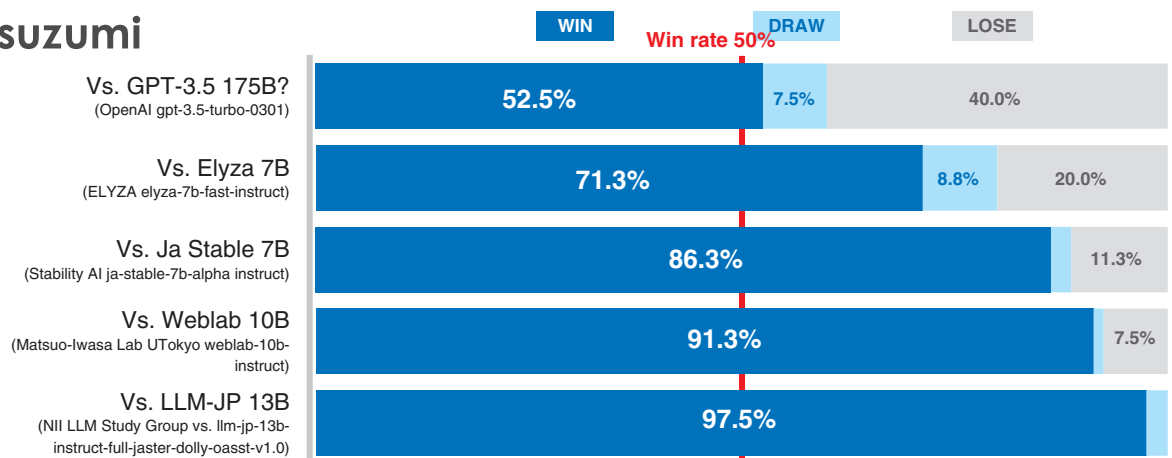For improved customizability (Feature 3), tsuzumi

Things not possible before

Reproduction of physicality

2. Acquisition of tacit knowledge

Prediction of the future

4. Diversity & Inclusion

Acquisition of functionality while respecting the user's wishes

Analysis of the brain

1. NTT's proprietary LLM

Remote reproduction of hearing, speech, and visual sense

3. Beyond tele-

New encounters, understanding oneself

Fig. 2.   Acquisition of human functions.

tsuzumi

WIN   DRAW   LOSE

Win rate 50%

Vs. GPT-3.5 175B?
(OpenAI gpt-3.5-turbo-0301)
52.5%   7.5%   40.0%

Vs. Elyza 7B
(ELYZA elyza-7b-fast-instruct)
71.3%   8.8%   20.0%

Vs. Ja Stable 7B
(Stability AI ja-stable-7b-alpha instruct)
86.3%   11.3%

Vs. Weblab 10B
(Matsuo-Iwasa Lab UTokyo weblab-10b-instruct)
91.3%   7.5%

Vs. LLM-JP 13B
(NII LLM Study Group vs. llm-jp-13b-instruct-full-jaster-dolly-oasst-v1.0)
97.5%

* Rakuda benchmark: https://yuzuai.jp/benchmark Ran on 22 Oct 2023.
  Scored by evaluating each LLM using two-model comparisons by GPT-4 (40 questions x 2 in order of presentation).
  Except for llm-jp, model output uploaded on the site were used in the evaluation.
  llm-jp is excluded because input repetition and termination tokens are done as post-processing due to settings described in the model card of Hugging Face.
  Evaluation scores were obtained by comparing all models listed on the leaderboard dated 27 Sep 2023 and tsuzumi-7b using two-model comparisons by GPT-4 and ranking the results using Bradley-Terry strengths.

Fig. 3.   Superiority of tsuzumi in Japanese language processing.

offers three tuning methods to flexibly respond to different requirements such as accuracy and cost: prompt engineering, full fine-tuning, and adapter tuning (**Fig. 4**). These tuning methods enable industry-specific customization at low cost.

The tsuzumi LLM's multimodal support (Feature 4) enables acquisition of a wider range of knowledge by not only language but also other modes of input
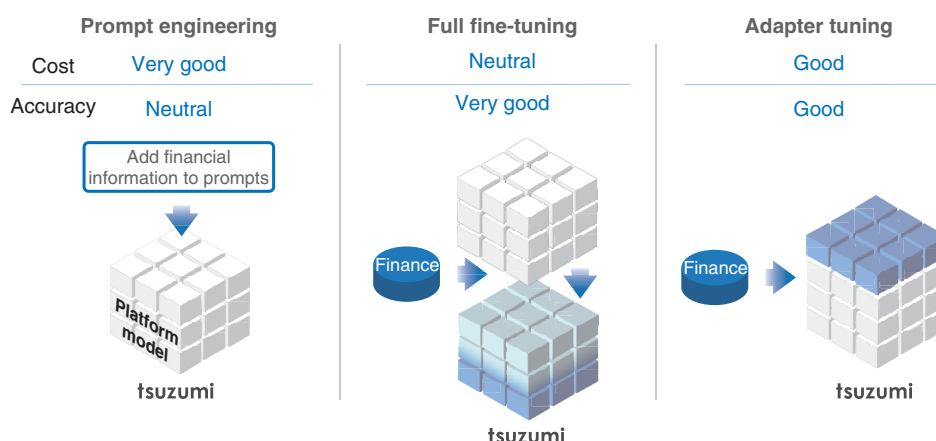
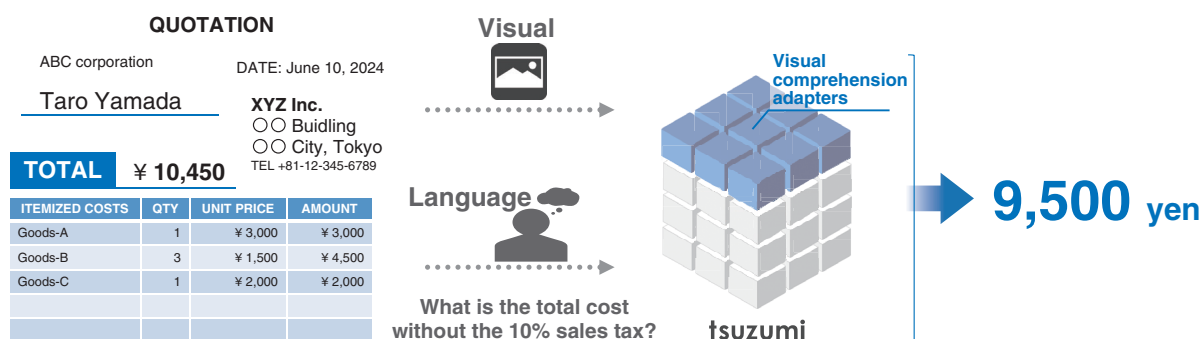Fig. 4.   Three different tuning methods.

Fig. 5.   Multimodal support.

and output, such as images, video, sensor data, nuances in speech, and facial expressions (**Fig. 5**). We are creating use cases centered on tsuzumi by linking it with NTT Human Informatics Laboratories' voice-, image-, and video-processing technologies, which have a history of more than 40 years, and using the knowledge of sensor and actuator technologies and cognitive psychology cultivated through our research on robotics.

4.   Synergy between tsuzumi and IOWN

IOWN's high-capacity, low-latency network provides the foundation for connecting the geographically dispersed resources necessary for an LLM. It also contributes to power saving in LLM operations. In the future, social issues will be solved not by using one massive LLM that knows everything but by linking together small LLMs with specialized knowledge

and personas, including tsuzumi.

5.   Message to everyone

It has been three years since the outbreak of the COVID-19 pandemic, transforming our lives. It made us realize once again that we cannot take for granted the idea that we can meet anyone, anytime. We now understand that when there are restrictions on non-essential/non-urgent activities, feeling happy is not always considered essential or urgent. Our surrounding environment is changing at a dizzying pace. When the Reiwa era (Japanese era based on the reigns of emperors) began in 2019, we could not have imagined what things would be like five years later. NTT's company name includes the word "tele-," meaning "at a distance," and NTT's business has been focused on communication that transcends distance. Amid the changes over the past few years, we have continued

to contribute to spreading and advancing the remote world by using information and communication technology. Our lives are still not yet settled, but it is time to examine from a human-centric perspective what has come about in the past few years and what we should leave behind for future generations. When we cannot see our loved ones, are there ways of communicating that transcend the barriers of time and distance? How much should machines imitate humans? Are there areas where we can allow machines to exceed human capabilities? How should technology provide support for people to experience convenience and happiness? Through collaboration with external parties, we will continue to ask ourselves these questions. By repeatedly testing our theories, we are focusing our resources on technologies that we desire to leave behind for future generations, even generations 100 years from now. The best from NTT Human Informatics Laboratories is still yet to come.

Kota Hidaka
Vice President, Head of NTT Human Informatics Laboratories.
He received an M.E. from Kyushu University, Fukuoka, in 1998 and Ph.D. in media and governance from Keio University, Tokyo, in 2009. He joined NTT in 1998. His research interests include speech signal processing, image processing, and immersive telepresence. He was a senior researcher at Council for Science, Technology and Innovation, Cabinet Office, Government of Japan, from 2015 to 2017.