Guidelines for Developing and Utilizing AI Technology that Uses Brain Data to Estimate Perceptual Information

January 2024

Committee for Designing the Guidelines for Developing and Utilizing AI Technology that Uses Brain Data to Estimate Perceptual Information

These guidelines were developed as a part of the projects by the National Institute of Information and Communications Technology (NICT); "April 2021-March 2022 The Research on Ensuring Social Acceptance of AI Technologies that Use Brain Data to Estimate Perceptual Information" and "2023 Research for Social Implementation of Sensory Evaluation Technology in Cyber Space Using Brain data".

Table of Contents

Terminologies introduced in these guidelines and their definitions

Terminology used in these guidelines are defined below:

Perception: The subjective experience of understanding and recognizing characteristics and forms of things in the external world, obtained through the five senses.

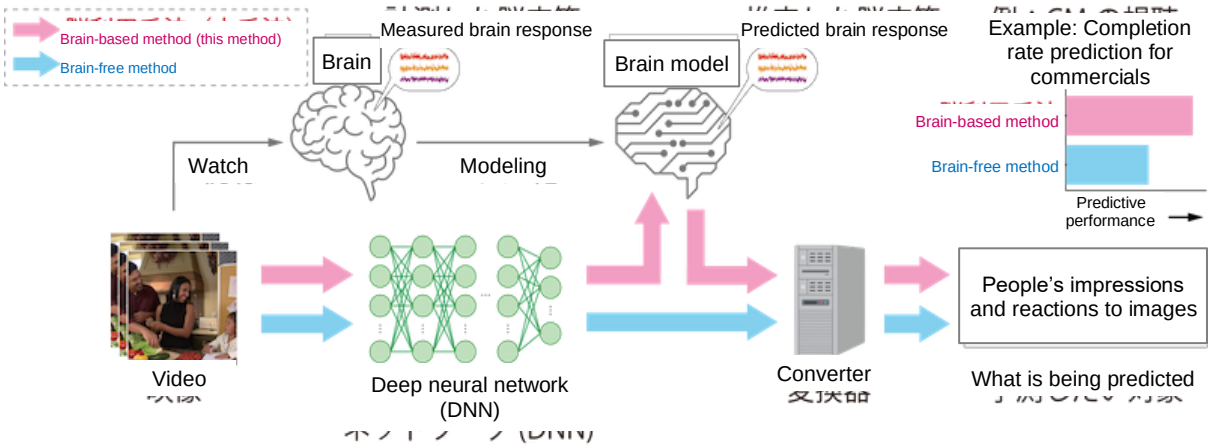
Impression/reaction: The cognitive/behavioral human reaction produced as a result of perception. For example, having a fondness for an object that one has perceived, getting motivated to make a purchase, or remembering the perceived object.

Brain data: Data obtained by recording cranial nerve activity using brain function measurement devices, such as functional magnetic resonance imaging (fMRI), also called brain activity data. This data expresses and represents information, including specific perception. It is also a form of information that can explain and estimate human perception and behavior beyond what is consciously expressed in a language. The main scope of these guidelines is data on dynamically changing functional representations of information in the brain.

AI technology that uses brain data to estimate perceptual information. (BTL: Brain-mediated Transfer Learning): Also referred to as "brain-mediated transfer learning." A machine learning model (decoding model) that approximates brain data (encoding model) from collected brain function data (brain data) related to perceptual experiences, such as watching videos, and further infer the perception of such content from the approximated brain data (Nishida *et al.* 2020).

Brain data model (BTL model): Also referred to simply as "model" in these guidelines. Synonymous with brain-mediated transfer learning models. Currently, there are individual models that estimate brain data and approximate perceptual contents for an individual and models that use multiple individual models to obtain generalized approximation results . The sample size assumed for these projects is about 50 to 100 Japanese people of various genders and ages. In principle, an individual model only estimates brain activity from an input (image, etc.) to the individual's brain. By linking the estimated results with perceptual information (annotation data) and impression/reaction/behavioral information, such as preferability/ad click-through rate, collected from other groups, we can derive such perceptions, impressions/reactions/behavior from brain data.

Figure: Brain data model overview[1]



Secondary use (of a model): Secondary use of a trained brain-mediated transfer learning model in other training data or models. For example, in the case of transfer learning, by using a brain-mediated transfer learning model that estimates perception, we can create a secondary model that predicts the advertising index

[1]Nishida S, Nakano Y, Blanc, A, Maeda N, Kado M, Nishimoto S. Brain-mediated Transfer Learning of Convolutional Neural Networks. Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence 34(4):5281–5288, 2020. https://doi.org/10.1609/aaai.v34i04.5974, and https://www2.nict.go.jp/nie/nishida/research.html

from the result obtained by linking the perception estimation results for an advertisement video and the new advertisement index.

Increase in effectiveness: Refers to the fact that the use of brain-mediated transfer learning may enhance the effects of services or products on people, such as the appeal of advertisement.

Specific brain data processing characteristics: Refers to the fact that perception differs between individuals due to personal attributes, even when viewing the same target. For example, a human face is not thought to be processed as a "face" by a person with face blindness—such people can recognize each part of the face but cannot recognize the face as a whole. For other individuals with addictive behaviors, the target stimulus may elicit a higher craving than others. Given the fact that each person's brain characteristics are different, the guideline focuses its discussion on characteristics related to "clinical symptoms and disorders," such as those listed in the examples.

Stigma: In these guidelines, it refers to the negative associations with individuals' traits, such as mental illness, that arise from the surroundings and self.

Personalization: In these guidelines, it refers to adjusting the stimulus sent by the seller of the product/advertisement to an individual/group with specific characteristics so that the target person will have the greatest reaction (have a favorable impression/develop desire to buy the product, etc.).

Subconscious influence: Refers to the influence on behaviors and provocation of reactions, including actions, through perceptions that cannot be consciously or verbally reported by the person. (The person feels that they are not seeing or hearing anything.) **It is** one of the effects on consumers that are not anticipated in traditional conscious surveys.

Expression of the subject's inner world: It refers to the act of expressing one's own feelings that cannot be understood from the outside to others.

Illusion: Refers to a change in perception or impression due to the information processing characteristics of the brain, not the changes in the physicality of what is perceived. In a narrow sense, it refers to perceptual illusions such as "appearing brighter" and "appearing longer," but in these guidelines, it also refers to things that affect higher-level cognition and behavior, such as a better perception of quality.

People involved in research and development: Refers to people involved in planning of research and development, brain measurement, data collection, analysis, model creation. Includes individuals and organizations that are secondarily involved through outsourcing, etc.

Business operators: Including not only primary users who use brain-mediated transfer learning models to create advertisements and product development but also organizations that work for manufacturers on a contract basis to provide consulting services using brain-mediated transfer learning.

Pseudonymization: Refers to the process of removing personal identifiers from data and replacing them with anonymous values. Used to protect individual privacy and improve data security.

Anonymized information: Refers to information that has been processed so that a specific individual cannot be identified and the personal information cannot be restored.

Opt-in: Refers to the process where users indicate to the other party (business operators) their intent to participate, consent, approve, etc. In the context of these guidelines, it also refers to the action of the business operator, when they obtain prior permission from the user.

Background

The co-evolutionary development of brain neuroscience and artificial intelligence (AI) technologies has enabled the development of "artificial brains," also known as digital twins, which is a product of cross-sectional technology that reproduces and estimates information representation of the brain and human perception on a computer. These technologies have the potential to contribute to innovation in the society as a whole, not only in terms of their scientific and technological significance in promoting a better understanding of the human brain, but also in terms of enabling the development of better products and services by learning more about consumers and users at a level not possible with conventional technologies.

While various AI guidelines have been developed both in Japan and overseas to address ethical issues in the utilization of AI technology, AI that handles information from the brain is an area that requires particular ethical consideration, even more so than that of ordinary AI. These considerations have not yet been put together, hindering its use.

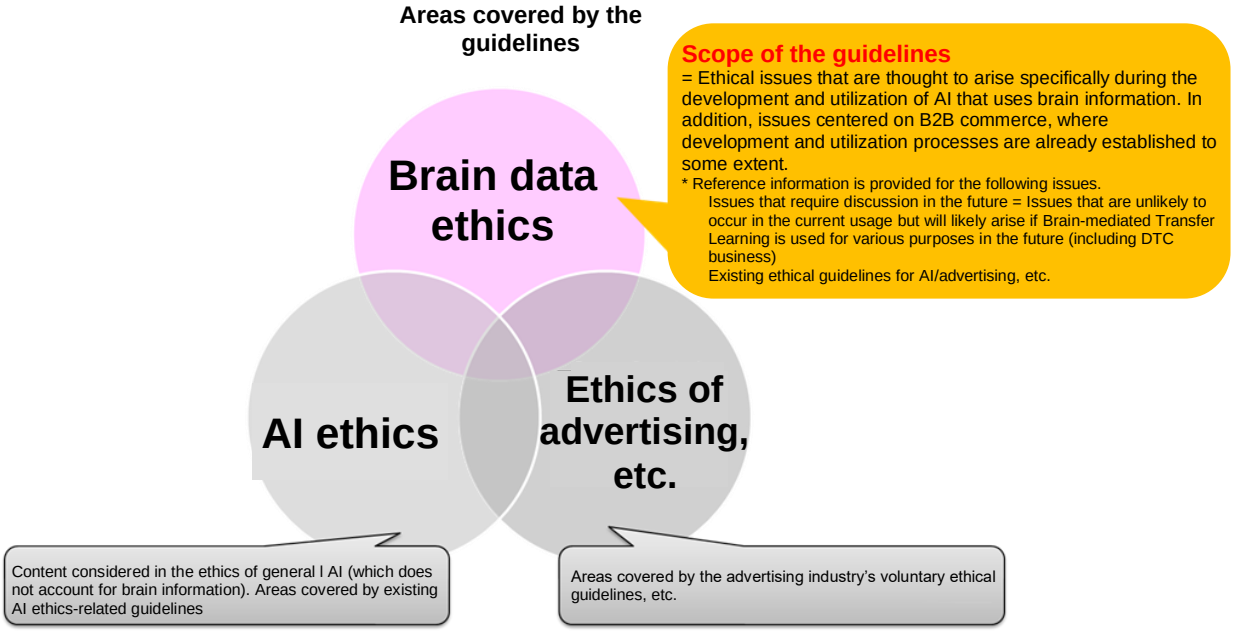
With this in mind, as a part of the project on "AI technology that Uses Brain Data to Estimate Perceptual Information (AI Strategy and Accelerating Social Implementation of Core Fundamental Research)," the project run by the Ministry of Internal Affairs and Communications and the National Institute of Information and Communications Technology (NICT), under the Cabinet Office's Public/Private R&D Investment Strategic Expansion PrograM (PRISM) and the Bridging the Gap Between R&D and the iDeal Society and Generating Economic and Social Value (BRIDGE), the Ministry of Internal Affairs and Communications has established a "Study Group Contributing to Research on Ensuring Social Acceptance of AI Technologies that Use Brain data to Estimate Perceptual Information" (hereinafter referred to as the "study group"). The study group discusses ethical, legal, and social issues (ELSI) related to the social implementation of AI technology that uses brain data to estimate perceptual information. In this regard, it has gathered a wide range of stakeholders, including researchers from neuroscience, humanities and social science, acting as a library of knowledge which can be used to ensure the social acceptance of technology, and has been actively researching and examining options for creating draft guidelines for technology development. These guidelines are the result of those efforts.

Introduction

Objective: The main purpose of these guidelines is to promote the use of "AI technology that estimates perceptual information from brain data," developed in the PRISM and BRIDGE project, which are science and technology development programs of the Cabinet Office, Government of Japan, and to provide guidelines that take ELSI into consideration.

Target audience: These guidelines are mainly intended for those involved in the research and development of "AI technology that estimates perceptual information from brain data" for business use, as well as companies and organizations that utilize such technology for business.

Scope: The main focus of these guidelines is on ethical issues that arise specifically from the use of brain data. Therefore, the guidelines do not include ethical issues that can be dealt with in the usual "AI ethics" or in the ethical guidelines for advertising that have been voluntarily prepared and operated by the advertising industry and other private business entities (see figure below). The main purpose of these guidelines is to promote the use of brain-mediated transfer learning, but because ethical issues in its application are also highly relevant to the upstream development process, we have compiled a list of issues and guidelines to keep in mind not only for the application but also in the overall process, "from development to application" (see the figure on the following page).



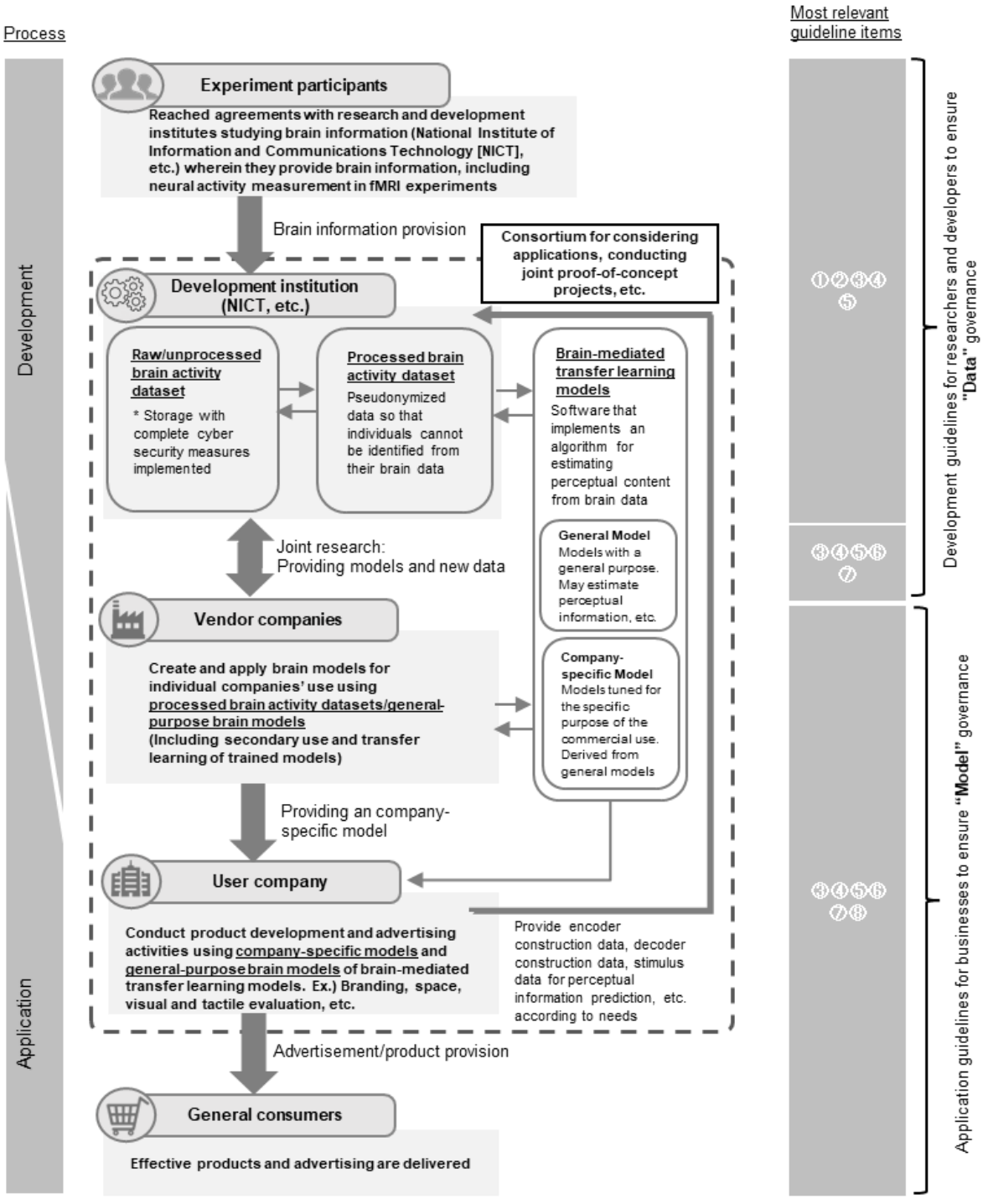
Legal binding force: These guidelines are not legally binding and do not restrict negotiations or contracts between parties.

Vision: These guidelines were prepared with the use and provision of brain data at Public R&D Projects in Japan in mind. However, the viewpoints and issues presented here are of a general nature, not limited to the use of brain data at NICT. These guidelines suggest perspectives on issues that are relevant to the overall process from development to utilization of brain data and desirable actions to address the ELSI. These guidelines will serve as a reference for ELSI to be considered not only for AI technology that estimates perceptual information from brain data, which is the main scope of the guideline, but also for future businesses that directly handle or utilize personal brain data, and contribute to the development of appropriate science and technology and its return to society. In fact, a number of cases have emerged in the new cross-sectional area of brain data and AI, such as the combination with generative AI and its use in real-time neurofeedback, and we aim to make it applicable to new business application areas where research and development will continue.

Positions regarding risk and uncertainty: The commercial use of brain data is an area that is expected to develop in the future, and therefore, we believe that there are many

uncertainties that are difficult to predict and measure now. Considering such circumstances, the basic stance of these guidelines is to strongly encourage monitoring and sharing of case studies on uncertain effects.

〈Development of brain-mediated transfer learning models —Overview of a model of research and development into brain information, and its application〉



～List of Guidelines～

1. Obtaining consent to handle brain data
2. Measures to be taken in the life cycle of brain data
3. Considerations and actions for the risk of accidentally promoting discrimination/stigma by using brain data to create AI models
4. Brain modeling with people with specific attributes, and considerations and actions for the potential risk of promoting stereotypes through application of such models
5. Considerations and actions related to specific brain characteristics
6. Consideration and actions on impact on consumers' ideologies/values without conscious awareness
7. Ensuring transparency in the use of brain data
8. Ensuring fair and inclusive accessibility to businesses that use brain data

What is AI technology that uses brain data to estimate perceptual information?

The basis of AI technology that uses brain data to estimate perceptual information is brain data decoding technologies that use machine learning technologies to deciphers and reconstruct the subject's perceptual content from functional brain imaging data, such as such as functional magnetic resonance imaging (fMRI), and encoding technologies that predict brain activity patterns in response to external stimuli or specific internal conditions.

The technology covered by these guidelines is an AI technology that combines various encoding and decoding techniques to predict human brain activity patterns in response to the external physical world, and from these predicted brain activity patterns, predict the corresponding perceptual content and other human responses. This technology is introduced as "brain-mediated transfer learning (BTL)" in the first paper from the field[2]. The following is an overview of this technology.

The steps involved in the development and use of AI technology that uses brain data to estimate perceptual information

Brain data collection: Conduct experiments on dozens to hundreds of people. The researchers provide the participants external stimuli such as images, sounds, products, and advertisements, and measure their brain activity using fMRI experiments. The main developer (National Institute of Information and Communications Technology (NICT), etc.) has to obtain the consent from participants to utilize their brain data.

Modeling: Create a pseudonymized model by linking acquired brain activity data and external data (perception data, behavior data, etc.). There are two levels of models: an encoding model that reproduces and predicts brain activity from feature information of sensory stimuli presented to the subject, and a decoding model that predicts and verbalizes the content and response of presented stimuli from brain activity data. The above models are collectively referred to as brain-mediated transfer learning models (or BTL models). Perceptual data refers to sensory information perceived by external stimuli including images and videos, such as “green grass,” “blue letters,” and “human voices;” impression data refers to verbalized preferences, emotional responses, etc. to target stimuli; and behavioral data refers to data generated by consumer behavior, etc., such as click rates and purchase numbers across markets. Brain activities can be simulated as a virtual brain by using this model as a basis for a virtualization technology that reproduces the brain activity to a specific sensory input and the subsequent perception and further responses on a computer.

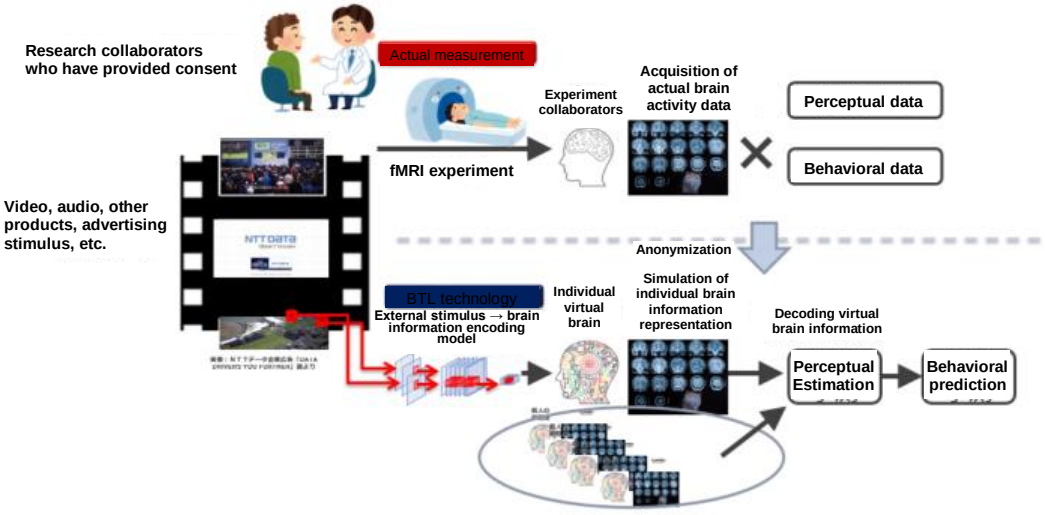
Application: After the model is learned through the process above, the individual's brain activity (brain data representation) for that stimulus is predicted via the encoded model by preparing new stimulus data such as images and videos. The predicted data of the individual's brain activity is used to decode the perceptual content and other information. It is possible to predict behavior at the market level if external data, such as click rates, for specific stimuli are learned. By averaging the prediction results of individual brains, it is expected to generalize the prediction results as well. As described above, by using the model to simulate a series of brain data processing processes, including stimulus input, brain data representation, perception, and behavioral output, companies can improve the efficiency of research and development. It also will be useful to improve products and marketing activities. For example, companies may be able to select the design that is most likely to produce the intended perception among consumers, or advertising displays that are most likely to maximize intended behaviors.

Updating: The model is enlarged or updated when more brain activity data is available, either by adding new subjects or collecting new brain data from existing subjects. The model is also updated when perceptual information. from existing brain data is used for a different purpose (e.g., to predict behavior rather than perception).

Disposal: If the subject who provided the brain data wishes to have their data deleted

[2] Nishida, S., Nakano, Y., Blanc, A., Maeda, N., Kado, M., & Nishimoto, S. (2020, April). Brain-mediated transfer learning of convolutional neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 04, pp. 5281-5288).

afterward, although it is not impossible to delete it, it is not easy[3]. It is difficult to completely remove the impact of an individual's information, as the data are pseudonymized when building the AI model and is intended to be used as a comprehensive average of the individual's model. The original data cannot be restored from the AI model, and it is impossible to identify the individual.



What is currently possible: Application of brain-mediated transfer learning model to B2B commerce

Main purpose: To analyze individual reactions to products and services using brain-mediated transfer learning and use it for advertising and product development.
Example of real-world application: A food processing company utilized AI in the production of TV shopping programs, increasing the incoming call rate by 27.6% (NTT DATA, 2018.[4]), Billboard Japan visualized music trends and predicted hit songs (Billboard Japan and NTT DATA, 2020[5])

Currently impossible uses

Estimating an individual's brain state in real time (Currently, models are constructed offline after data collection)
Deciphering sensitive information such as ideologies and beliefs.

What may be possible in the future: Application of brain data models for DTC business (Direct to Consumer: personal services)

Main purpose: Use brain data models customized for individuals or specific groups to improve personalized services
Example applications: Educational use (using a specific brain model to personalize educational content), skill development, disability/disease treatment/health care use

Currently envisioned usage scheme

We envision that research and development institutions (NICT, etc.) will directly receive requests from businesses to provide technology, enabling them to develop advertisements and products geared toward consumers. It is also possible that businesses receive brain-mediated transfer learning technology from research and development institutions (NICT, etc.), and provide consulting services to third-party companies to support creation of advertisements and product development.
The business side basically handles the modeled brain-mediated transfer learning and does not access the original brain data. On the other hand, the business side, as the user of models, can obtain and utilize data on characteristics of people who contributed the brain

[3] If the brain information AI model itself is found to be harmful in some way, the model itself may be discarded or discontinued.
[4] https://www.nttdata.com/jp/ja/news/release/2018/11130
[5] https://www.nttdata-strategy.com/newsrelease/200903.html

data to the models as long as there is consent from the participants who provided the brain data and personal information protection can be established (ensuring both technological transparency and personal information protection).

Guidelines for using AI technology that uses brain data to estimate perceptual information.

The following guidelines are provided regarding the use of AI technology that uses brain data to estimate perceptual information.

For information on expected use cases that require specific consideration, please refer to the attached report titled "Research Towards Ensuring Social Acceptance of Utilizing AI Technology that Uses Brain Data to Estimate Perceptual Information."

Guideline (1) Obtaining consent to handle brain data

> If we do not pay sufficient attention to the appropriate handling of data, there is a possibility that privacy would be violated by linking of brain information and personal characteristics, and the decision-making process in data acquisition, use, provision to third parties, etc. would be carried out inappropriately. The privacy and decision-making of brain information providers should be protected by appropriately making use of broad consent (consent that permits the degree of freedom of use within the scope of specific uses), opt-in, and pseudonymization. Also, it is recommended to obtain consent from the brain information provider on the following statement at the time of brain information acquisition; "it is difficult to withdraw specific data from the model. Therefore, in principle, the data cannot be deleted after they has been provided."

Currently, we do not expect research institutes to provide data that can identify individuals from brain data. In the future, if companies themselves implement projects that link brain responses to personal characteristics, new privacy-related issues may arise. For example, when advertising potentially addictive products and services, such as gambling, alcohol, and social-network games, as well as products associated with one's socioeconomic situations, like consumer finance, the brain data of participants may get linked with their characteristics, such as addictive tendencies and socioeconomic statuses. The combined data may be then used to develop products and personalization for other consumers. For this reason, it is crucial to continue considering the following: how privacy should be protected when companies collect and share brain data to utilize individual brain data and deciphered results; and when and from whom business operators and researchers need to obtain consent.

Furthermore, the focus of issues in the data lifecycle may differ depending on the scope of use, such as research use, commercial use, and medical use. Therefore, one should obtain consent from the brain data providers relating to the processing method, users, and purpose of use of brain data at the time of brain data acquisition. One must obtain consent after explaining the purpose of use (research, commercial, etc.), with the understanding that brain data will be pseudonymized and anonymized and what to do when opting out.

Also, although no such data regarding the linkage of the person's attribute information or daily behavioral data with brain data are included at this stage, "personal information that includes the person's race, belief, social status, medical history, criminal record, whether the person has been harmed by a crime, or any other description specified by a Cabinet Order as requiring special consideration in handling to prevent unjust discrimination, prejudice, or other disadvantage to the person" is considered to be "sensitive personal information" under Article 2, Paragraph 3 of the Act on the Protection of Personal Information. Therefore, the consent of the individual will be required when collecting data that is likely to be categorized as "sensitive personal information".

It is difficult to erase a specific individual's data and recreate a model without it. Therefore, it is recommended that those involved in research and development obtain consent from the brain data provider acknowledging that it is difficult to delete the data from the developed model afterwards, after ensuring them that individuals cannot be identified from the model constructed

from their data.

Expected specific use cases requiring special consideration ➞ See the Supplemental Material Use Cases 2.1.1 and 2.2.2

Guideline (2) Measures to be taken in the life cycle of brain data

> Comprehensive monitoring should be conducted during a series of processes of collecting and processing brain information (assumed to be conducted by a public research institute), and providing and sharing the information for private use during research (acquisition of brain information), as well as during the development, application, and post-use phase of AI. If an instance of impropriety is discovered, the information on the impropriety must be disclosed as soon as possible. In addition, we recommend that such processes be disclosed to brain information providers and consent obtained and that each process be evaluated by a third-party ethics committee or similar group.

In the process from the collection to processing of brain data and its subsequent use, it is necessary to monitor whether brain data is being acquired and used appropriately with regard to the intended purpose of use, usage system, and data management system, etc. It is recommended to evaluate appropriateness of use at the time of use, and evaluation of whether there were any negative effects caused after the use. When an instance of impropriety is discovered, a prompt response should be taken and information should be disclosed about the details of the incident. In principle, it is technically impossible for a business operator using a brain-mediated transfer learning model to access the brain data of the individuals, and it is assumed that it is not necessary from a business perspective. However, if a case arises in which the original brain data or individual can be identified, such as a case of reverse engineering of a learned model, it is recommended that the information be promptly shared with the relevant parties.

Guideline (3) Considerations and actions for the risk of accidentally promoting discrimination/stigma by using brain data to create AI models

> The creation of brain models using specific groups, categorized by gender, age group, race, and health status, etc. and the application of such models to advertising expressions may be advantageous as it may promote fairer and more ethical business use of brain models. On the other hand, there is a risk that comparing and disseminating differences in brain characteristics may encourage xenophobia and discrimination. Therefore, due care must be exercised in creating, applying, and disseminating brain models.

Since brain-mediated transfer learning is an emerging technology, analyzing the characteristics of the brain based on profiling specific gender, age, race, disease, etc., can exacerbate misinterpretation from the general public, resulting in discrimination. It can also lead to stigma (negative attitudes towards people with certain attributes). Comparing and differentiating the information representation characteristics of the brain, for example, in relation to specific mental illnesses, may lead to perversion (e.g., "the brain of a person with a disability cannot recognize a human face").And dissemination of such perverted ideas may lead to promotion of xenophobia and discrimination. Therefore, stakeholders should take appropriate actions when using such models and sharing insights, such as consulting experts on ELSI and personal information protection.

Expected specific use cases requiring consideration ➞ See the Supplemental Material Use Case 2.1.1

Guideline (4) Brain modeling with people with specific attributes, and considerations and

actions for the potential risk of promoting stereotypes through application of such models
Targeting —designing products, advertisements, etc. to meet the specific needs of a group with certain attributes— is considered a normal practice in marketing. However, there are also inherent risks that occur when brain-mediated transfer learning is used. For example, when using brain-mediated transfer learning for marketing cooking utensils, cleaning tools, etc., if only female brain-mediated transfer learning models are used, models may produce advertisements or products that appeal more to women. As a result, there is a risk of strengthening stereotypes (housework is for women, for example) and developing products that are optimized for only a certain group of people (for example, a product that men would find difficult to use). Although such events can also occur through existing corporate activities, it is assumed that such effects would occur unintentionally and implicitly when using brain-mediated transfer learning, even if the business operator does not intend to or the consumer does not consciously notice such effects. Therefore, it is recommended to be aware of the risk of stereotype formation/reinforcement in advance and to monitor the impact of stereotyping when using brain-mediated transfer learning models as with the use of other marketing techniques. Respect for the natural diversity of human beings is required. Also, in the event of a problem, disclosure of what group was targeted in creating the brain-mediated transfer learning model will be needed.

> When creating and using a brain information model with only specific attributes such as a specific gender, there is a risk that this may strengthen specific stereotypes and exclude specific groups in a way that is not intended by the business operator that uses it, or at a level that consumers cannot consciously notice. We should be aware of the risks and be sure to avoid reinforcing stereotypes and to ensure inclusiveness. Also, in the event of a problem, disclosure of what group was targeted in creating the brain model will be needed.

As with Guideline (3), considerations must be given to risks when creating brain-mediated transfer learning models using specific populations based on factors such as gender, age, race, and disease. One potential risk when using the cognitive behavior predicted by the model for businesses is that it may exclude the diversity of people by promoting the tendencies of certain groups of people, which requires consideration.

Expected specific use cases requiring special consideration ➡ See the Supplemental Material Use Cases 2.1.1 and 2.2.2

Guideline (5) Considerations and actions related to specific brain characteristics

> "Specific brain characteristics" refer to brain conditions that are more sensitive than other individuals. For example, a person with addictive symptoms may experience an increased craving for the target substance due to perceptual input of the stimulus, and such characteristics should not be exploited for business purposes. Even in the case of unintentional use, monitoring for such adverse effects and information disclosure and countermeasures are required when adverse effects are found.
> In the research phase, it is meaningful to conduct research on people with specific characteristics related to information processing in the brain. However, it is recommended that consent be obtained by referring to the guidance on personal information requiring special consideration.

Using brain-mediated transfer learning holds promise in producing content that is highly effective for people with specific brain characteristics. However, as a result of this, there is a risk of adverse effects when the vulnerability of their brain is taken advantage of, such as aggravating cravings and worsening psychiatric symptoms for people with addictive tendencies. As an example, there is a possibility of discovering brain responses peculiar to addicts through brain data analysis of products that produce addiction symptoms (including gambling, alcohol, and social-network games). Based on this, one could develop a marketing" method that effectively arouses cravings by intensively targeting addicts. Another possible situation is that a person who feels negatively about his or her body shape or face may feel even more inferior due to the input

of information about beauty. Similarly, in fields such as consumer finance, based on brain data analysis of the economically vulnerable, one may leverage their cravings and impatience to make the business more exploitative. There is research significance in brain data profiling of already addicted patients or people with addictive tendencies, and it is hoped that such information will contribute to the formation of a fairer market.

When dealing with products with a risk of addiction, consideration should be given to whether the results of brain AI analysis will have a greater impact on addicted patients' cravings. We also shouldn't conduct profiling that aggravates a particular psychiatric condition or market products specifically to people with psychiatric conditions (eating disorders, body dysmorphia, etc.).

If adverse effects arise inadvertently due to the vulnerability, appropriate measures, such as monitoring, information disclosure, and countermeasures, will be required.

In addition, special consideration must be given to sensitive personal information as defined in the amended Act on the Protection of Personal Information and the Ministry of Health, Labour and Welfare's "Guidance on Ethical Guidelines for Life Science and Medical Research on Human Subjects" (2021) in the collection of brain data from people with specific attributes. In particular, health-related information, medical history, presence or absence of disability, etc. obtained through services such as medical examinations and medical care are sensitive personal information. This is highly relevant to the collection of brain data of people with specific attributes. In acquiring sensitive personal information, it is not essential to obtain "informed consent" as per medical guidelines, but even if it is not necessary, researchers, etc., in principle need to obtain "appropriate consent" from research subjects, etc. Therefore, a business plan in accordance with the guidance provided in this document is desirable.

Expected specific use cases requiring special consideration ➡ See the Supplemental Material Use Cases 2.1.1 and 2.1.2

Guideline (6) Considerations and actions of impact on consumers' ideologies/values without conscious awareness

> Analysis and use of brain information may reveal expressions and techniques that can change people's ideologies and values without conscious awareness. Special consideration is needed in the development and application of such purposes and methods, and ethics review by a third party should be sought in cases where a stimulus generation method that unconsciously changes people's thoughts and values is explored or used.

We predict that consumer reactions that do not involve conscious perception will be explored and utilized based on brain data analysis. Specifically, this refers to cases in which consumers themselves will feel that they are acting according to their own will, but in reality, their preferences and purchasing behavior are unconsciously influenced by advertisements based on brain data analysis and customization. Although these are qualitatively similar ethical issues to those in existing advertisements, etc., the use of brain data may increase the effect size, , resulting in an imperceptible effect. Thus, the use of brain data itself requires ethical considerations and transparency about the use of brain-mediated transfer learning models.

Furthermore, there is a possibility that brain data may be used for public relations activities related to ideologies and beliefs. However, brain-mediated transfer learning should not be used for advertisements and contents for the purpose of incitement to activities contrary to social values such as democracy, fundamental human rights, freedom of speech and expression, and law and order. In this regard, a social framework would be needed in the future, such as the establishment of a third-party committee to conduct objective examinations.

In addition, it is also necessary to act in line with Guideline (5) "Considerations and actions related to specific brain characteristics."

Expected specific use cases requiring special consideration ➡ See the Supplemental Material Use Cases 2.1.3 and 2.2.2

Guideline (7) Ensuring transparency regarding the use of brain data

> It is necessary to ensure the accuracy and safety of the targets of brain information applications. Consideration should also be given to the impact resulting from the targeted applications.

Since brain-mediated transfer learning is still in its developmental stage, it is an unestablished field in terms of accuracy, reliability, and validity of interpretation. Therefore, we need to be wary of potential misperceptions (excessive trust and expectations) of business operators and consumers who use brain-mediated transfer learning. Specifically, the use of phrases such as "based on brain data analysis" may be used to fictitiously increase consumer confidence while providing low-precision services. Efforts should be made to ensure transparency regarding the process, accuracy, and technical limits of brain-mediated transfer learning construction and to actively communicate with users and consumers. It is strongly recommended to clearly state the use of brain-mediated transfer learning on the business's website.

In addition, since stimuli presented when creating a brain-mediated transfer learning model directly affects the performance of the model, it will be necessary in the future to ensure reliability and reproducibility by disclosing information on what and how much stimuli were presented to subjects and by systematizing and standardizing the presentation methods behind such learned models.

Expected specific use cases requiring consideration ➡ See the Supplemental Material Use Cases 2.1.1, 2.2.3, and 2.3.1

Guideline (8) Ensuring fair and inclusive accessibility to businesses that use brain data

The construction of a brain-mediated transfer learning model is costly, and even if a trained model is used, the price of products and services that use it will be several times higher than that of products that use simple AI. For example, in the future, if brain-mediated transfer learning is applied to preventive services for cognitive function decline and services related to overcoming PTSD, etc., it is possible that health disparities will arise due to a varying budgetary constraint. The public and private sectors must work together to ensure inclusion, such as integration to welfare policies, particularly when introduced products and services are useful to many people.

> Since brain information analysis is expected to be several times more expensive than simple AI, there is concern that targets will be limited in scope. A guarantee of inclusive accessibility is required.

Expected specific use cases requiring consideration ➡ See the Supplemental Material Use Cases 2.3.2

Note that these guidelines may be reviewed and revised in the future in accordance with domestic and international technological research and development, ethical, legal, and social trends, etc[6].

See the supplemental material for elaborate examples of expected use cases.

[6] Combination technologies of generative AI and brain-mediated transfer learning have also been reported as of January 2024 (Takagi, Y., & Nishimoto, S. (2023). High-resolution image reconstruction with latent diffusion models from human brain activity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*), and will be the focus of future attention.

List of the members for the Committee for Designing the Guidelines for Developing and Utilizing AI Technology that Use Brain Data to Estimate Perceptual Information.

Members of the "April 2021-March 2022 Study Group Contributing to Research on Ensuring Social Acceptance of AI Technologies That Use Brain data to Estimate Perceptual Information".

(without honorifics. In Japanese alphabetical order.)

Committee Chairman	Ryuma Shineha	Associate Professor, Research Center on Ethical, Legal and Social Issues , Osaka University
Vice-Chairman	Takahisa Taguchi	Deputy Director, Center for Information and Neural Networks, Future ICT Research Institute, National Institute of Information and Communications Technology
Committee Member	Chiori Ijichi	Group Executive Specialist, Food Research Laboratory, Food Business Division, Ajinomoto Co.
	Yasuhiko Imamura	Representative Director, VIE STYLE Co.
	Hideki Kashioka	Director, Center for Information and Neural Networks, Future ICT Research Institute, National Institute of Information and Communications Technology
	Mitsuo Kishimoto	Professor, Institute for Datability Science, Osaka University
	Manabu Tanifuji	Visiting Professor, Department of Life Science and Medical Bioscience, School of Advanced Science and Engineering, Waseda University
	Hiroshi Nakagawa	Team Leader, RIKEN Center for Integrated Research on Innovative Intelligence
	Masaru Nishitsutsumi	Researcher, Center for Information and Neural Networks, Future ICT Research Institute, National Institute of Information and Communications Technology
	Ippei Hagiwara	Director and Secretary General, Consortium for Applied Neuroscience
	Masashi Hattori	Professor, College of Comprehensive Psychology, Ritsumeikan University
	Hiromitsu Hattori	Professor, College of Information Science and Engineering, Ritsumeikan University
	Ryusuke Hayashi	Senior Researcher, Artificial Intelligence Research Center,

		National Institute of Advanced Industrial Science and Technology (AIST)
	Naoya Maeda	Technical Grade, Social Innovation Division, NTT DATA Corporation
	Atsushi Yokoi	Researcher, Center for Information and Neural Networks, Future ICT Research Institute, National Institute of Information and Communications Technology
	Megumi Yokono	Associate Professor, Faculty of Social SciencesSchool of Social Sciences, Waseda University
Information Provider	Shinji Nishimoto	Professor, Department of Neuroscience, Graduate School of Frontier Biosciences, Osaka University
	Tomohiro Kuroda	Professor, Department of Medical Information Planning, Kyoto University Hospital
	-	Genetic Information Handling Association of Japan

External evaluation committee members of "2023 Research for Social Implementation of Sensory Evaluation Technology in Cyber Space Using Brain data"

（without honorifics. In Japanese alphabetical order.）

External evaluation committee member	Ryota Kanai	Chief Executive Officer, Araya Inc.
	Mitsuaki Takemi	Project Assistant Professor, Faculty of Science and Technology, Keio University
	Kenji Nakamura	Advanced Research Center Project Manager, DIC Corporation
	Soshi Nakamura	Chief Researcher, Takenaka Research & Development Institute
	Fumiya Nakamura	Senior Manager, Asahi Quality & Innovations, Ltd. Core Technology Laboratories Division Two

Guideline Development Administrative Office

(NTT Data Institute of Management Consulting, Inc. Neuro Innovation Unit)
Takuya Ibaraki, Midori Momoki, Miho Kaiya, Akari Tsurumaki, Misaki Tajima, Kikyo Sahashi, Azuna Sawada, Yuri Kawase.