【Grant-in-Aid for Scientific Research (S)】

Externalizing and sharing the world model in the brain

Principal Investigator	Kyoto University, Graduate School of Informatics, Professor KAMITANI Yukiyasu　Researcher Number : 50418513
Project Information	Project Number : 25H00450　Project Period (FY) : 2025-2029 Keywords : AI, Brain, Brain decoding, Brain–machine interface

Purpose and Background of the Research

● **Outline of the Research**

This research aims to decode the “world model” represented in our brain using brain decoding methods and to share it with others in the form of 3D graphics and other media. Previous brain decoding research has primarily focused on reproducing sensory input to the eye as images. In contrast, this research goes a step further, aiming to elucidate the 3D world that the brain constructs.

● **Background**

Our research team has developed machine learning-based brain decoding methods. Using these, we can predict the category and features of objects a person is viewing from brain activity measured by MRI and other methods. More recently, by aligning brain activity and artificial intelligence (AI), we have succeeded in reconstructing images of what people see or imagine.

However, current technology has a significant limitation. It depends on the form of sensory input and fails to capture the entirety of the "world model" that the brain constructs.

Theories of the brain suggest that our perception is not just a mirror of sensory input but that the brain possesses an internal model of the world and forms perception while predicting incoming information.

● **Purpose**

The purpose of this research is to advance brain decoding to the next level. Going beyond the conventional "reproduction of sensory input," we aim to elucidate and externalize the world model.

Specifically, we adopt a NeuroAI approach, which fuses cutting-edge multimodal AI and generative AI with neuroscience, to externalize the brain's internal world model as 3D graphics and semantic information. This will enable us to scientifically capture an individual's subjective worldview.

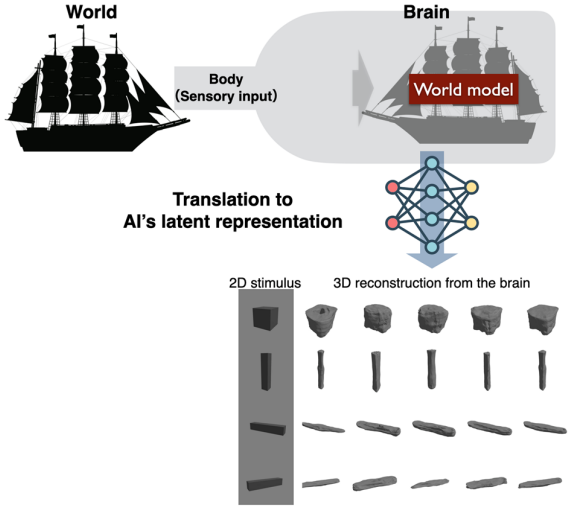

Figure 1. Outline of the study

● **Significance**

In the scientific realm, our research integrates brain decoding with generative brain theory, advancing both neuroscience and AI. By elucidating neural representations beyond sensory input, we seek to uncover the fundamental architecture of human perception and cognition.

We are also developing advanced systems that externalize the world model in real-time through virtual environments. This novel brain–machine interface paradigm promises to transform human communication with applications spanning healthcare and education.

To address concerns about methodological rigor in brain decoding research, we propose mathematical formalizations that clearly define capacities and limitations of the methods. Through open science practices and ethical consideration, we aim to establish credibility while ensuring the alignment with societal values.

Expected Research Achievements

Externalizing the perceptual world in the brain

This research analyzes human brain activity measured via MRI or implanted electrodes through AI’s latent representation. AI systems are known to acquire information representations similar to those of the biological brain by learning from vast amounts of real-world image and text data. We translate brain activity into AI’s latent features and use them to generate human-recognizable content such as three-dimensional shapes and language.

● **Recreating the mind's 3D world**

We measure brain activity while subjects view photos, videos, and other media, then translate it into the latent representations of multimodal AIs. Using generative AI, we recreate the world experienced by the individual as 3D graphics and other content. We aim to externalize "internal worlds”, not just of what people actually see, but also of imagination, illusions, and various psychological states.

● **Connecting the brain to VR in real-time**

We develop real-time brain decoding methods using implanted electrodes. We aim to propose prototype interfaces that combine world models read from the brain with virtual spaces, enabling sharing with others.

● **Toward reliable neurotechnology**

Our research characterizes training data and AI’s latent representation from the perspectives of zero-shot prediction (the ability to handle new outcomes) and compositionality (the ability to generate complex outcomes by combining basic elements). We establish theoretical frameworks that clarify the capacities and boundaries of brain decoding methods, enabling rigorous evaluation and the development of evidence-based guidelines for reliable neurotechnology.

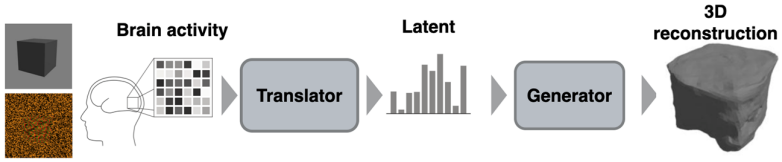

Figure 2. Reconstruction of 3D shape from brain activity

Homepage Address, etc.	https://kamitani-lab.ist.i.kyoto-u.ac.jp/ https://www.youtube.com/@ATRDNI https://github.com/KamitaniLab