
Ryota Takatsuki

Email: R.Takatsuki@sussex.ac.uk

EDUCATION

Sussex Centre for Consciousness Science, University of Sussex (Brighton, UK) - *Ph.D. in Informatics*

Sep 2025 - Mar 2029 (Planned)

- Supervised by Prof. Anil Seth and Prof. Chris Buckley
- Studying the mechanisms of optical illusions in Vision Transformers and predictive coding networks using mechanistic interpretability techniques

The University of Tokyo (Tokyo, Japan) - *M.A. in Applied Computer Science, Interdisciplinary Information Studies*

Apr 2023 - Sep 2025

- Developed a light-weight real-time 3D video transmission system for drone holoportation applications, supervised by Prof. Akihiro Nakao
- Withdrew in Sep 2025 to begin Ph.D. studies after a gap year since Apr 2024

The University of Tokyo (Tokyo, Japan) - *B.S. in Systems Innovation, Engineering*

Apr 2019 - Mar 2023

Thesis: 圧縮RGBD画像のLocal5G超低遅延伝送による高臨場感通信 (High-presence communication through Local 5G ultra-low latency transmission of RGBD images)

- Thesis project (April 2022 - Mar 2023) supervised by Prof. Akihiro Nakao
- Worked on mesh generation, RGBD image compression, multi-camera registration, and application-layer protocol.
- Implemented a light-weight real-time 3D video transmission system running over Local 5G.

EXPERIENCE

Research and Development Center for Large Language Models, National Institute of Informatics (Tokyo, Japan) - *Technical Assistant*

Jun 2024 - Sep 2025

-
- Working on bistable perception in vision language models in collaboration with Yanaka Laboratory (UTokyo)

AI Alignment Network (Tokyo, Japan) - *Research Fellow*

Jun 2024 - PRESENT

- Engaging in community-building activities for AI Alignment, such as hosting a study group on mechanistic interpretability, writing educational blogs

Araya Inc (Tokyo, Japan) - *Moonshot program R&D intern*

Dec 2021 - Sep 2025

- Supervised by Dr. Ippei Fujisawa and Dr. Ryota Kanai
- Working on mechanistically interpreting the mechanism of optical illusions in transformer-based models and developing interpretability tools, in collaboration with Sonia Joseph (McGill Univ.)
- Worked on meta-networks in the context of AI interpretability and Consciousness

Araya Inc (Tokyo, Japan) - *R&D intern*

Dec 2021 - Sep 2023

- Supervised by Dr. Ippei Fujisawa and Dr. Ryota Kanai
- Conducted research on meta-networks in the context of AI interpretability and Consciousness

The University of Tokyo (Tokyo, Japan) - *Tutor for international students*

Apr 2023 - Mar 2024

RESEARCH

Publications

- **Takatsuki, R.**, Doi, T., Watahiki, A., Seth, A., & Yanaka, H., (How) Do MLLMs Report Bistable Images Like Humans? (under review)
- Aoki, K., **Takatsuki, R.**, Minegishi, G., Haruki, Y., & Kawahara, D., In-Context Neurofeedback: Can Large Language Models Control Their Internal Representations Requiring Privileged Access? (under review)
- Aoki, K., **Takatsuki, R.**, Minegishi, G., & Cho, H., Mechanistic Interpretability: A New Trend in Interpretability Research (accepted in Transactions of the Japanese Society for Artificial Intelligence)

-
- Kanai, R., **Takatsuki, R.***, & Fujisawa, I. (2025). Meta-representations as representations of processes. *Neuroscience of Consciousness*, 2025(1), niaf038.
 - **Takatsuki, R.**, Joseph, S., Fujisawa, I., & Kanai, R. (2025). Decoding Vision Transformers: the Diffusion Steering Lens. *The First Workshop on Mechanistic Interpretability for Vision, CVPR 2025*.
 - **Takatsuki, R.**, Kanai, K., & Nakao, A. (2024). A real-time 3D video transmission and display system for drone holoportation applications. *IEICE Technical Report; IEICE Tech. Rep.*, 123(367), 13–18.
 - **Takatsuki, R.**, & Nakao, A. (2023). High-presence communication through Local 5G ultra-low latency transmission of RGBD images. *IEICE Technical Report; IEICE Tech. Rep.*, 122(406), 157–162.
 - **Takatsuki, R.**, & Nakao, A. (2023). Quantitative evaluation of the impact of the inter-access point interference on unlicensed communication in high-capacity, low-latency transmission. *IEICE Technical Report; IEICE Tech. Rep.*, 122(406), 227–232.

Presentations

Oral

- **Takatsuki, R.**, Joseph, S., Fujisawa, I., & Kanai, R. “Towards Reverse-engineering the Kanizsa Illusion: A Mechanistic Study of Vision Transformer”. *The 39th Annual Conference of the Japanese Society for Artificial Intelligence*. Osaka, Japan. May 29, 2025.
- Aoki, K., **Takatsuki, R.**, Minegishi, G. “Mechanistic Interpretability: A New Trend in Interpretability Research”. *The 39th Annual Conference of the Japanese Society for Artificial Intelligence*. Osaka, Japan. May 29, 2025.
- **Takatsuki, R.**, Fujisawa, I., & Kanai, R. “Towards elucidating the mechanism of illusory contours through mechanistic interpretability”. *Vision Society of Japan*. Tokyo, Japan. Sep 15, 2024.
- **Takatsuki, R.**, Fujisawa, I., & Kanai, R. “Qualia as Representations of Processes”. *The 27th annual meeting of Association for the Scientific Study of Consciousness..* Tokyo, Japan. Jul 4, 2024.

Poster

- **Takatsuki, R.**, Nishimoto, S., Minegishi, G., & Seth, A. “Do Autoregressive LLMs Exhibit Postdiction-Like Integration of Later Tokens?”. *The 29th annual meeting of Association for the Scientific Study of Consciousness*. Santiago, Chile. Jul 1st, 2026. (Planned)

- **Takatsuki, R.**, Fujisawa, I., & Kanai, R. “Meta-Representations and Conceptual IIT”. *Post ASSC satellite symposium: Mathematical and philosophical models of consciousness: from classical to quantum-like*. Kyoto, Japan. Jul 8, 2024.
- **Takatsuki, R.**, Fujisawa, I., & Kanai, R. “Weight structures in ANNs and their modality-specificity”. *Consciousness Research Network 2023*. Taipei, Taiwan. Aug 26, 2023.

Invited

- **Takatsuki, R.** “Introduction to Mechanistic Interpretability” *Student Session, International Conference on Robot Ethics and Standards 2024*. Jul 31, 2024.

Service

- Reviewer, [The First Workshop on Mechanistic Interpretability for Vision](#), CVPR 2025
- Organizing Committee, [Organized Session ‘Mechanistic Interpretability’](#) at the 39th Annual Conference of the Japanese Society for Artificial Intelligence
- Organizing Committee and Reviewer, AAAI-25 Workshop Program [‘1st Workshop on Post-Singularity Symbiosis’](#)

AWARDS

- **Academist Prize 5th “Understanding the World as Perceived by AI: Toward a Future Where Humans and AI Coexist!”**
A crowdfunding research grant of around 2 million yen by 238 supporters ([link](#))
Additional grant of 300k yen ([link](#))
- **Sussex Centre for Consciousness Science PhD Studentship**, University of Sussex
Stipend and fee-waiver for 3.5 years and a research grant
- **Young Researcher Encouragement Award**, IEICE Technical Committee on Network Systems, 2024
This award is given to the presenter of an encouraging lecture given by a young researcher at a Type 1 Research Meeting organized by the Network Systems Research Special Committee, which is recognized as particularly excellent and promising for future activities. The award is given to approximately 10% of the presentations eligible for the Young Researchers’ Encouragement Award.

OTHER ACTIVITIES

- Participated in [Qstr-IIT summer school](#) (Oct 2025) - partially funded
- Participated in [Autumn School of Computational Neuroscience](#) (Nov 2024) - fully funded
- Contributed to [AI safety Notes](#)
- Completed Neuromatch Academy, Computational Neuroscience (Jul 2024)
- Host a weekly Mechanistic Interpretability Study Group at AI Alignment Network (May 2024 - Present)

-
- Co-host a consciousness journal club, 'Consciousness Olympia' (Mar 2024 - Present)
 - Authored [the first introductory article on mechanistic interpretability in Japanese](#)
 - Participated in EAGx Philippines 2023 - fully funded
 - Participated in AGI fundamentals reading group (Apr 2023 - Apr 2024)
 - Served as ITU delegate to [Huawei Asia-Pacific Seeds For The Future Program 2022](#), Bangkok/Hua Hin, Thailand - fully funded
 - Served as Japanese delegate to [the Generation Connect Global Youth Summit 2022](#), Kigali, Rwanda - fully funded

LANGUAGE PROFICIENCY

- IELTS Academic OA 8.0 (Dec 2021)