

# Yuta Oshima

PH.D. STUDENT · THE UNIVERSITY OF TOKYO

Engineering bldg. 2nd 9F 92C1, GCI Matsuo Lab. 7-3-1, Hongo, Bunkyo-ku, Tokyo, JAPAN, 113-0033

☎ (+81) 90-3523-0449 | ✉ [yuta.oshima@weblab.t.u-tokyo.ac.jp](mailto:yuta.oshima@weblab.t.u-tokyo.ac.jp) | 🏠 <https://shim0114.github.io/> | 🎓 Yuta Oshima

## Summary

I'm a first-year PhD student at The University of Tokyo, advised by Yutaka Matsuo, and I closely collaborated with Hiroki Furuta.

My ultimate goal is to build **real-time, interactive world models** that seamlessly translate human imagination into reality—empowering anyone to craft and mold their envisioned worlds with intuitive and precise control.

Toward this vision, my research focuses on controllable generation with **diffusion models** and **world models**, including work on **alignment** and **instruction-following generation** for fine-grained control.

## Education

### The University of Tokyo

PH.D. IN COMPUTER SCIENCE

Advisor: Prof. Yutaka Matsuo

Tokyo, Japan

Apr. 2025 - Present

### The University of Tokyo

M.S. IN COMPUTER SCIENCE

Tokyo, Japan

Apr. 2023 - Mar. 2025

### The University of Tokyo

B.S. IN ENGINEERING

Tokyo, Japan

Apr. 2019 - Mar. 2023

## Selected Publications

### CONFERENCE

#### MultiBanana: A Challenging Benchmark for Multi-Reference Text-to-Image Generation

[YUTA OSHIMA](#), DAIKI MIYAKE, KOHSEI MATSUTANI, YUSUKE IWASAWA, MASAHIRO SUZUKI, YUTAKA MATSUO, HIROKI FURUTA

CVPR 2026 Main

2026

#### Inference-Time Text-to-Video Alignment with Diffusion Latent Beam Search

[YUTA OSHIMA](#), MASAHIRO SUZUKI, YUTAKA MATSUO, HIROKI FURUTA

NeurIPS 2025

2025

#### ADOPT: Modified Adam Can Converge with Any $\beta_2$ with the Optimal Rate

SHOHEI TANIGUCHI, KENO HARADA, GOUKI MINEGISHI, [YUTA OSHIMA](#), SEONG CHEOL JEONG, GO NAGAHARA, TOMOSHI IYAMA, MASAHIRO SUZUKI, YUSUKE IWASAWA, YUTAKA MATSUO

NeurIPS 2024

2024

### CONFERENCE WORKSHOP

#### AKITalk: Audio-Implicit Keypoints for Identity-Preserving Talking-Head Video Synthesis

RIKU TAKAHASHI, RONGZHI LI, [YUTA OSHIMA](#), SHO KUNO, RYUGO MORITA, ISSEY SUKEDA

3rd Place in The ACM Multimedia Asia 2025 Grand Challenge

MMAsia 2025 Workshop

2025

#### SSM Meets Video Diffusion Models: Efficient Video Generation with Structured State Spaces

[YUTA OSHIMA](#), SHOHEI TANIGUCHI, MASAHIRO SUZUKI, YUTAKA MATSUO

ICLR 2024 Workshop

2024

#### Tactile In-Hand Pose Estimation through Perceptual Inference

TATSUYA KAMIJO, TOMOSHI IYAMA, [YUTA OSHIMA](#), GENTIANE VENTURE, TATSUYA MATSUSHIMA, YUTAKA MATSUO, YUSUKE IWASAWA

IROS 2023 Workshop

2023

Spotlight Presentation

### PREPRINT

## WorldPack: Compressed Memory Improves Spatial Consistency in Video World Modeling

YUTA OSHIMA, YUSUKE IWASAWA, MASAHIRO SUZUKI, YUTAKA MATSUO, HIROKI FURUTA

Preprint

Dec. 2025

## Enhancing Unimodal Latent Representations in Multimodal VAEs through Iterative Amortized Inference

YUTA OSHIMA, MASAHIRO SUZUKI, YUTAKA MATSUO

Preprint

Oct. 2024

## Work Experience

---

### EQUES Inc.

AI RESEARCHER

Research on anime-style video generation within the GENIAC project, supported by Japan's Ministry of Economy, Trade and Industry (METI).

Tokyo, Japan

Jul. 2025 - Present

### World-leading Innovative Graduate Study Program Co-designing Future Society (WINGS-CFS), The University of Tokyo

RESEARCH ASSISTANT

Tokyo, Japan

Dec. 2023 - Present

## Academic Honors

---

### AWARDS

- 2025 **3rd Place**, The ACM Multimedia Asia 2025 Grand Challenge: Multimodal Multiethnic Talking-Head Video Generation *Kuala Lumpur, Malaysia*
- 2024 **Outstanding Presentation Award**, JSAI Annual Conference *Shizuoka, Japan*
- 2023 **Spotlight Presentation**, IROS 2023 Workshop on World Models and Predictive Coding in Cognitive Robotics *Detroit, U.S.A*

### GRANTS

- 2025 **1,000,000 JPY**, Leadership Development Program for Ph.D. (LDPP), the University of Tokyo

## Academic Activities

---

- 2026 **Reviewer**, European Conference on Computer Vision (ECCV)
- 2026 **Reviewer**, International Conference on Machine Learning (ICML)
- 2026 **Reviewer**, The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)
- 2026 **Reviewer**, International Conference on Learning Representations (ICLR)