Kotaro Yoshida

M.Sc. in Engineering

yoshida.k.0253@m.isct.ac.jp | Homepage | GitHub | Google Scholor

*"I am a first-year master's student at the Institute of Science Tokyo, specializing in deep learning. My primary interests are improving generalization in deep neural networks and developing model merging methods for flexible model editing. During my undergraduate studies, I received the Outstanding Student Award in recognition of my academic achievements and research contributions. I have also competed in multiple machine learning competitions—earning medals each time—which strengthened both my theoretical grounding and practical skills. My academic experience includes an exchange program at the National University of Singapore and research internships at A*STAR."*

Education

Institute of Science Tokyo — *Tokyo, Japan*

M.Sc. in Engineering — *Supervisor: Prof. Konstantinos Slavakis* — *04/25 - 03/27 (Expected)*

- Topic: Efficient AI and Mechanistic Interpretability
- **Thesis Proposal Excellence Award**

Institute of Science Tokyo — *Tokyo, Japan*

B.Sc. in Engineering — *Supervisor: Prof. Konstantinos Slavakis* — *04/21 - 03/25*

- Topic: Out-of-Distribution Generalization in Deep Learning
- **Outstanding Student Award**

National University of Singapore — *Singapore*

Exchange Program in Computer Science — *08/23 - 05/24*

Ueda High School — *Nagano, Japan*

High School Diploma — *04/18 - 03/21*

Skills

Programming	Python, C, Java, LaTeX
Framework and Tools	PyTorch, TensorFlow, Wandb, Slurm, Git
Languages	Japanese (Native), English (Proficient), Korean (Beginner), Mandarin (Beginner)

Research Experience

Sakana AI — *Tokyo, Japan*

Student Researcher — *Supervisor: So Kuroki* — *02/26 - Present*

- Topic: LLM Agent

Agency for Science, Technology and Research (A*STAR) — *Singapore, Singapore*

Singapore International Pre-Graduate Award (SIPGA) — *Supervisor: Atsushi Nitanda* — *06/25 - 09/25*

- Topic: Learning Theory

ProPlace Inc — *Tokyo, Japan*

Research Intern — *Supervisor: Hiroki Naganuma* — *06/23 - 01/26*

- Topic: Out-of-Distribution Generalization, Calibration, and Model Merging in Deep Learning

Agency for Science, Technology and Research (A*STAR) — *Singapore, Singapore*

Research Intern — *Supervisor: Atsushi Nitanda* — *02/24 - 05/24*

Honors & Awards

Academic Awards

2025	**Singapore International Pre-Graduate Award (SIPGA),** Agency for Science, Technology and Research (A*STAR)	*Singapore*
2024	**Student Encouragement Award,** The 65th National Convention of Information Processing Society of Japan (IPSJ2024)	*Yokohama Japan (Online)*

Others

2023 **Bronze medal [Top 10%],** Kaggle ICR – Identifying Age-Related Conditions *Online*

2024 **Bronze medal [Top 4.5%],** SIGNATE SMBC Group GREEN×DATA Challenge *Online*

2023 **Gold medal [10th place : Top 1.6%],** SIGNATE Student Cup 2023 *Online*

Publication

* denotes equal contribution.

Preprint / Papers under review

"How Does Preconditioning Guide Feature Learning in Deep Neural Networks?", *Kotaro Yoshida, Atsushi Nitanda,*

Under Review *09/25*

"Robust Invariant Representation Learning by Distribution Extrapolation", *Kotaro Yoshida, Konstantinos Slavakis,*

Arxiv Preprint *05/25*

"Augmenting NER Datasets with LLMs: Towards Automated and Refined Annotation", *Yuji Naraki*, Ryosuke Yamaki*, Yoshikazu Ikeda, Takafumi Horie, Kotaro Yoshida, Ryotaro Shimizu, Hiroki Naganuma (* denotes equal contribution),*

Arxiv Preprint *03/24*

Journal

"An Empirical Study of Pre-trained Model Selection for Out-of-Distribution Generalization and Calibration", *Hiroki Naganuma, Ryuichiro Hataya, Kotaro Yoshida, Ioannis Mitliagkas,* **[TMLR2025]**

Transactions on Machine Learning Research *04/25*

"Towards Understanding Variants of Invariant Risk Minimization from the Perspective of Calibration", *Kotaro Yoshida, Hiroki Naganuma,* **[TMLR2024]**

Transactions on Machine Learning Research *06/24*

International conference

"DisTaC: Conditioning Task Vectors via Distillation for Robust Model Merging", *Kotaro Yoshida, Yuji Naraki, Takafumi Horie, Ryotaro Shimizu, Hiroki Naganuma,* **[ICLR2026]** *Brazil*

The 14th International Conference on Learning Representations *01/26*

"On Fairness of Task Arithmetic: The Role of Task Vectors", *Hiroki Naganuma*, Kotaro Yoshida*, Laura Gomezjurado Gonzalez*, Takafumi Horie, Yuji Naraki, Ryotaro Shimizu,* **[ICLR2026]** *Brazil*

The 14th International Conference on Learning Representations *01/26*

"Mastering Task Arithmetic: τJp as a Key Indicator for Weight Disentanglement ", *Kotaro Yoshida, Yuji Naraki, Yoshikazu Ikeda, Takafumi Horie, Ryosuke Yamaki, Ryotaro Shimizu, Yuki Saito, Julian McAuley, Hiroki Naganuma,* **[ICLR2025]** *Singapore*

The 13th International Conference on Learning Representations *01/25*

Domestic conference, workshop

"Evaluating the Effectiveness of Model Linearization in Task Analogies", *Kotaro Yoshida, Yuji Naraki, Takafumi Horie, Ryosuke Yamaki, Ryotaro Shimizu, Yuki Saito, Julian McAuley, Hiroki Naganuma,* *Osaka Japan*

The 39th Annual Conferance Of The Japanese Society For Artificial Intelligence (JSAI2025) *05/25*

"Mastering Task Arithmetic: τJp as a Key Indicator for Weight Disentanglement ", *Kotaro Yoshida, Yuji Naraki, Yoshikazu Ikeda, Takafumi Horie, Ryosuke Yamaki, Ryotaro Shimizu, Yuki Saito, Julian McAuley, Hiroki Naganuma,* *Saitama Japan*

The 27th Information-Based Induction Sciences and Machine Learning (IBIS2024) *10/24*

"Towards Understanding Variants of Invariant Risk Minimization from the Perspective of Calibration", *Kotaro Yoshida, Hiroki Naganuma,* **[Travel Grant]** *Shizuoka Japan*

The 1st Symposium Of Young Researcher Association For Machine Learning (YAML2024) *09/24*

"A Closer Look at Task Analogies: Insights from Function and Parameter Space", *Kotaro Yoshida, Yuji Naraki, Takafumi Horie, Ryosuke Yamaki, Ryotaro Shimizu, Yuki Saito, Hiroki Naganuma,* **[Travel Grant]** *Osaka Japan*

The 19th Symposium Of Young Researcher Association Fof NLP Studies (YANS2024) *09/24*

"Towards Understanding Variants of Invariant Risk Minimization from the Perspective of Calibration", *Kotaro Yoshida*, Hiroki Naganuma* (* denotes equal contribution),* **[Student Encouragement Award]** *Yokohama Japan (Online)*
The 86th National Convention Of Information Processing Society Of Japan (IPSJ2024) *03/24*

"Uncertainty Calibration in Deep Neural Networks through Invariant Risk Minimization", *Osaka Japan (Online)*
Kotaro Yoshida, Hiroki Naganuma* (* denotes equal contribution),*
Forum For Information And Technology 2023 (FIT2023) *09/23*

Fellowship, Scholarship, and Grants-in-Aid

TSUBAME Encouragement Program for Young/Female/Younger Users, Tokyo Institute of Technology

Computational Resource Support [4000 GPU hours] *08/25 - 03/26*

- Topic: Multi-faceted Impact Assessment of Model Merging Techniques in Deep Learning and Guideline Development for Practical Application

TSUBAME Encouragement Program for Young/Female/Younger Users, Tokyo Institute of Technology

Computational Resource Support [4000 GPU hours] *08/24 - 03/25*

- Topic: Robust and Reliable Vision-Language Model Learning by Information Bottleneck

TSUBAME Encouragement Program for Young/Female/Younger Users, Tokyo Institute of Technology

Computational Resource Support [4000 GPU hours] *06/23 - 03/24*

- Topic: Uncertainty Calibration in Deep Neural Networks through Invariant Risk Minimization

HIOKI Scholarship, HIOKI E.E. CORPORATION

Living Expenses *04/21 - 03/25*

Support Scholarship for Study Abroad, Japan Student Services Organization

Living Expenses *08/23 - 05/24*