

Kohei Ozaki

Tokyo, Japan | eowner[at]gmail.com | ho.lc | linkedin.com/in/koheiozaki | github.com/smly

Summary

- ML Engineer and Tech Lead with 10+ years in research-driven and production-grade machine learning.
- Delivered ML solutions and built high-quality datasets across industries ranging from finance and advertising to healthcare and manufacturing.
- 4x Kaggle Grandmaster (20+ gold medals). Published at CoNLL and CVPR workshops and contributed datasets and tools used by the global AI community.

Education

Nara Institute of Science and Technology, Ph.D in Computer Science Apr 2011 – Mar 2014

- Computational Linguistics Laboratory (Supervisor: Prof. Yuji Matsumoto)
- Completed coursework without degree.

Nara Institute of Science and Technology, MS in Computer Science Apr 2009 – Mar 2011

- Computational Linguistics Laboratory (Supervisor: Prof. Yuji Matsumoto)
- **Thesis:** "Mutual k-Nearest Neighbor Graphs for Semi-Supervised Learning" (NAIST-IS-MT0951027)

Tokyo University of Science, BS in Computer Science Apr 2005 – Mar 2009

- Artificial Intelligence Laboratory (Supervisor: Prof. Ikuo Tahara)
- **Thesis:** "Breaking CAPTCHAs with Directed Spanning Tree"

Experience

AI Fellow & Senior Manager, Rist – Tokyo, Japan Mar 2023 – Present

- Contributed to the full project lifecycle in various capacities, from initial proposal and development to serving as Project Manager and Tech Lead.
- Drove a joint research partnership for building foundation models for printed circuit board inspection systems, delivering technical proposals.

Self-employed, Freelance Dec 2017 – Oct 2023

- Advisory service and software development service.
- Worked as a committee member of AI Edge Contest, organized by METI (Ministry of Economy, Trade and Industry) from 2017 to 2023.

Machine Learning Engineer, Ubie – Tokyo, Japan Mar 2022 – Feb 2023

- Developed dialogue algorithms and curated datasets for a symptom search engine.
- Built BI dashboards with data pipelines to support customer success operations.

Software Engineer, Preferred Networks – Tokyo, Japan Jul 2019 – Feb 2022

- Contributed to the development of Optuna (a hyperparameter optimization framework) as a committer, primarily focusing on integration modules and the LightGBM Tuner.
- Delivered data science solutions for projects in cancer screening, plant control, and retail services.

Senior Software Engineer, Recruit Technologies – Tokyo, Japan Nov 2015 – Sep 2016,
May 2017 – Jun 2019

- R&D on image retrieval and 3D reconstruction.
- Technical advice on voice recognition features for an educational application.
- Technical advice and support for new business projects.

Principal Software Engineer, Magne-Max Capital Management Oct 2016 – Apr 2017

- Development of mutual funds based on behavioral economics.

- Analyst**, AIG Holdings – Tokyo, Japan Mar 2015 – Oct 2015
- Built a fraud scoring system for property and casualty insurance, collaborating closely with the special investigation team.
- Software Engineer**, FreakOut – Tokyo, Japan Jan 2014 – Feb 2015
- Developed ML-based algorithms for real-time bidding in target advertising market.
- Software Engineer**, mixi – Tokyo, Japan May 2012 – Dec 2013
- Research and development for recommendation on social network service.
 - Developed a real-time bidding system for online advertising.

Projects

- RiichiLab – An Open Arena for Riichi Mahjong AI** (riichi.dev) 2022 - Present
- Developed and open-sourced a game simulator and game viewer for Riichi Mahjong, an imperfect information game (github.com/smly/RiichiEnv).
 - Built and operated an open arena server where Mahjong AI agents can compete against each other in real time.

Kaggle Community

- 4x Kaggle Grandmaster. Over 20 competition gold medals.
- Authored a book about Kaggle. Organized 9 meetups. Gave 10+ talks at conferences, companies and meetups.
- Google Developer Expert for Kaggle since 2025.

Publications

(CVPR'20 WS) Shuhei Yokoo, **Kohei Ozaki**, Edgar Simo-Serra, Satoshi Iizuka, "**Two-stage discriminative re-ranking for large scale landmark retrieval**", Proc of the CVPR'20 workshop.

(CVPR'19 WS) **Kohei Ozaki**, Shuhei Yokoo, "**Large-scale landmark retrieval and recognition under a noisy and diverse dataset**", Technical report for CVPR'19 workshop.

- Developed GLDv2-clean, the largest high-quality public dataset for image retrieval tasks that was later used as a seed dataset for DINOv2. github.com/cvdfoundation/google-landmark.

(CVPR'19 WS) Yusuke Niitani, Toru Ogawa, Shuji Suzuki, Takuya Akiba, Tommi Kerola, **Kohei Ozaki**, Shotaro Sano, "**Team PFDet's methods for Open Images Challenge 2019**", Technical report for CVPR'19 workshop.

(CoNLL'11) **Kohei Ozaki**, Masashi Shimbo, Mamoru Komachi, Yuji Matsumoto, "**Using the mutual k-nearest neighbor graphs for semi-supervised classification of natural language data**", Proc of the CoNLL'11.

Awards (Selected)

- 4th place** - (NeurIPS'25 WS) Google CodeGolf Championship, 2025. (gcfg.pages.dev)
- 4th place** - (CVPR'23 WS) Image Matching Challenge 2023, 2023.
- 5th place** - xView3: Detecting IUU Fishing Vessels (hosted by DIU), 2021.
- 3rd place** - NGA Overhead Geopose Challenge (hosted by NGA), 2021.
- 1st place** - (CVPR'21 WS) Hotel-ID to Combat Human Trafficking, 2021.
- 1st place** - (CVPR'19 WS) Google Landmark Recognition Challenge, 2019.
- 1st place** - (KDD'15 WS) ACM KDD Cup'15: Massive Open Online Courses Dropout Prediction, 2015.
- IPSJ Yamashita SIG Memorial Research Award, 2012.

Lecture

"Data Science Exercises", Kumamoto University, 2025.

- 90min lecture titled "Introduction to LLMs" & 540min exercise on data science.

"Advanced Lectures on Information Science", Tokyo Metropolitan University, 2019.

"Big Data Analysis", Osaka University, 2018.