

HIYA ROY

(+1) 914-309-6131 ◇ hiyaroy12@gmail.com ◇ <https://hiyaroy12.github.io/>

PROFESSIONAL SUMMARY

Applied AI researcher specializing in **LLMs, agentic AI, and multimodal systems**, with experience developing **ontology-grounded AI systems**, and production-scale generative AI. Expertise spans **structured pruning and knowledge distillation, vision-language and diffusion models**.

EDUCATION

The University of Tokyo, Japan

Sept 2015 – Mar 2021

Ph.D. in Electrical Engineering & Information Systems, GPA: **3.9/4**

2017 – 2021

M.S. in Electrical Engineering & Information Systems, GPA: **3.94/4**

2015 – 2017

WORK EXPERIENCE

Oracle, USA

April 2026 - Present

Position: Senior Applied Scientist, Oracle Analytics

Role: Developed **LLM-based automated Enterprise Ontology and knowledge graph generation** for the **Oracle AI Data Platform**, extracting concepts, relationships, instances, and business rules from enterprise documents and tables with **source provenance**. Evaluated downstream **LLM and agent reasoning**, and validated extracted business rules for **grounded and verifiable AI**.

- Achieved **higher concept and triple recall** than a **Codex skill-based harness** on Text2KGBench benchmark using GPT-5.6-Sol-medium for ontology generation, while delivering significantly **lower latency** and stronger end-to-end performance on a 5,000-document enterprise stress test.

Rakuten USA, Inc.

June 2022 – Jan 2026

Rakuten Group Inc.

April 2021 - June 2022

Position: Research Scientist, Rakuten research lab

Role: Led the development of Rakuten internal efficient LLMs through **structured pruning** and **Nemotron-style distillation**. Delivered **production-scale multimodal systems** for advertising, including **artifact detection and inpainting, SDXL-based ad-image generation, and Qwen-Image-Edit-based Japanese text rendering**. Developed a **research-to-product pipeline** for artifact removal that was adopted by Rakuten's Business Unit in Japan, supporting an estimated **¥918M monthly GMS**.

- **LLM Pruning and Knowledge Distillation:** Conducted research on **structured pruning** and **knowledge distillation** to compress **RakutenAI LLMs** into efficient student models with reduced size and FLOPs while preserving accuracy based on **NVIDIA Nemotron Framework**. **Compressed 3-4x with 7% overall improvement** on standard LLM benchmarks compared to baseline.

- **Vision-Language Model Training for Generative Ad Images:** Trained and fine-tuned **Stable Diffusion** models to generate and edit multi-resolution advertisement banners. Developed for the background generation of **TDA (Top Dynamic Ads)** in the ad ecosystem of Rakuten. Our model showed higher item integrity compared to commercial Seedream-v4.

NASA JPL (CalTech), USA

Oct 2019 – Jan 2020

Position: Visiting Researcher

Multitask learning for Space Applications: Worked with the Machine learning-based Analytics for Autonomous Rover Systems (MAARS) research group on the Robotic Surface Mobility Group (347F). Developed a joint lightweight neural network framework based on PyTorch for multitask learning (image segmentation and captioning) onboard planetary rovers.

NEC Data Science Research Laboratories, Japan

Jan 2017-Mar 2017

Position: Research Intern

- Worked on Automatic Target Recognition on SAR images using deep learning techniques.

Tata Power Company Limited, India

July 2012 – July 2015

Position: Lead engineer

- Worked as a core Electrical Testing engineer to conditionally monitor and test all electrical equipment, to carry out commissioning tests of new switchgear equipment and relay panels.

TECHNICAL SKILL

Frameworks PyTorch, Tensorflow, scikit-learn, CUDA, cuDNN, MXNET, Pandas, Docker.
Language and Tools Python and Matlab; Conda, Jupyter, Git, OpenCV, Latex, Microsoft Office.

SELECTED PUBLICATIONS

Journal

- Chaudhury S., **Roy H.**, Mishra S., Yamasaki T., **Adversarial Training Time Attack against Discriminative and Generative Convolutional Models**, IEEE Access (2021).
- **Roy H.**, Chaudhury S., Yamasaki T., Hashimoto T., **Image inpainting using frequency domain priors**, SPIE Journal of Electronic Imaging, 30(2), 023016 (2021).

Conference

- **Roy H.**, Stenger B., **Text Removal In E-Commerce Images: A Comparison Of Inpainting Methods**, British Machine Vision Conference (BMVC) 2024.
- **Roy H.**, Chaudhury S., Yamasaki T., DeLatte D.M., Ohtake M., Hashimoto T., **Lunar surface image restoration using U-Net based deep neural networks**, 50th Lunar and Planetary Science Conference 2019.
- **Roy H.**, Yamasaki T., Hashimoto T., **Do hashtags help? – Image aesthetics prediction using only hashtags**, WICV@CVPR 2018.
- Chaudhury S., and **Roy H.**, **Can fully convolutional networks perform well for general image restoration problems?**, Intl. Conf. on Machine Vision Applications, 2017.

PATENTS

A comprehensive list of my published patents can be found here: <https://patents.justia.com/search?q=Hiya+Roy>.

AWARDS AND ACHIEVEMENTS

- Awarded **MEXT Scholarship** by the Government of Japan (**Ministry of Education, Culture, Sports, Science and Technology of Japan**) for graduate studies. Oct 2015 - Aug 2020
- Awarded certificate for **Global Leader Program for Social Design and Management(GSDM)** by the University of Tokyo. Oct 2016 - Mar 2021
- Secured **rank 278 (99.722 percentile)** in **WB Engineering Entrance Examination** 2008.
- Participated in CSIR Program on Youth for leadership in Science for outstanding performance **WB Secondary Board Examination** 2006. [**rank:20 (99.997 percentile)**]

MEDIA COVERAGE

Our work on image inpainting using frequency-domain priors has been introduced in multiple tech blogs: **SPIE news**, **Marktechpost.com**, and **TechXplore**.