Sosuke Nishikawa

Software/Research Engineer

✉ sosuken115@gmail.com | 🏠 sosuke.info | ⓞ Sosuke115 | in sosuke-nishikawa-2100581a0

Education

- **M.S. in Information Science and Technology, The University of Tokyo** — *Mar. 2022*
  - Supervisors: Isao Echizen, Yoshimasa Tsuruoka
  - Department Head's Award
- **B.E. in Electrical Engineering and Computer Science, The University of Tokyo** — *Mar. 2020*
  - Supervisor: Yoshimasa Tsuruoka
  - GPA: 3.8/4.3

Work Experience

LY Corporation — *Tokyo, Japan*
Research Engineer, Science Group — *Apr. 2022 - PRESENT*

- Conducted research and development to enhance core NLP/ML libraries for Japanese text/queries using C++ and Python. Focus areas included improving Query Understanding, Text Embeddings, and Language Model Processing.
- Developed and operated ML infrastructure for NLP using MLOps practices. This included automating model training with PyTorch and Trino, optimizing inference servers with ONNX Runtime, and deploying with Docker and Kubernetes.

Studio Ousia — *Tokyo, Japan*
Software / Research Engineer (Part Time) — *Jan. 2020 - PRESENT*

- Conducted research and development on Japanese LLMs (e.g., Instruction Tuning, RAG)
- Developed NLP SaaS products utilizing Flask, React, and MySQL
- Conducted research on introducing entity representations into pre-trained language models [2], [3]

Echizen Lab, National Institute of Informatics — *Tokyo, Japan*
Research Assistant — *Apr. 2020 - Mar. 2022*

- Conducted research on Adversarial Examples for addressing AI vulnerabilities [1]

Stockmark Inc. — *Tokyo, Japan*
Software Engineer (Part Time) — *Dec. 2018 - Dec. 2019*

- Developed a web application (Anews) enabling users to quickly discern individuals' areas of expertise

Internship

Integral AI — *San Francisco, CA*
Research Engineer Internship (three-month) — *Jan. 2022 - Mar. 2022*

- Conducted research and development on the integration of robotics and large-scale multimodal models

Sony Corporation — *Tokyo, Japan*
Research Engineer Internship (one-month) — *Sep. 2020 - Oct. 2020*

- Conducted research on dialogue systems and the classification of conversations

Cookpad Inc. — *Tokyo, Japan*
Machine Learning Engineer Internship (two-week) — *Sep. 2020 - Sep. 2020*

- Developed an API for predicting ingredients from recipes [8]

Publications

My full publication list is available on my Google Scholar profile.

Refereed

[1] Closer Look at the Transferability of Adversarial Examples: How They Fool Different Models Differently

- Futa Waseda, **Sosuke Nishikawa**, Trung-Nghia Le, Huy H. Nguyen, Isao Echizen
- *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2023

[2] EASE: Entity-Aware Contrastive Learning of Sentence Embedding
- **Sosuke Nishikawa**, Ryokan Ri, Ikuya Yamada, Yoshimasa Tsuruoka, Isao Echizen
- *Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, Oral, 2022
- OSS: https://github.com/studio-ousia/ease

[3] A Multilingual Bag-of-Entities Model for Zero-Shot Cross-Lingual Text Classification
- **Sosuke Nishikawa**, Ikuya Yamada, Yoshimasa Tsuruoka, Isao Echizen
- *SIGNLL Conference on Computational Natural Language Learning (CoNLL)*, Oral, 2022

[4] Data Augmentation with Unsupervised Machine Translation Improves the Structural Similarity of Cross-lingual Word Embeddings
- **Sosuke Nishikawa**, Ryokan Ri and Yoshimasa Tsuruoka
- *Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: Student Research Workshop (ACL-IJCNLP SRW)*, 2021

Non-Refereed

[5] 検索クエリにおける共起情報を活用した非曖昧ドメイン固有語辞書の構築: ランドマークの事例
- EN title: Building a Dictionary of Unambiguous Domain-Specific Words Leveraging Co-occurrence Information in Search Queries: A Case Study of Landmark
- **Sosuke Nishikawa**, Sota Yamashiro, Hiroki Asano, Shumpei Sano, Manabu Sassano
- *Annual Meeting of the Association for Natural Language Processing (NLP)*, 2023, Domestic Conference

[6] エンティティの言語間リンクに基づく多言語モデルの構築
- EN title: Building Multilingual Models based on Inter-Language Links of Entities
- **Sosuke Nishikawa**, Ikuya Yamada, Yoshimasa Tsuruoka, Isao Echizen
- *Annual Meeting of the Association for Natural Language Processing (NLP)*, 2021, Domestic Conference

[7] マルチラベル分類による材料推薦モデル
- EN title: A Model for Recommending Cooking Ingredients Using Multi-Label Classification
- Yusuke Fukasawa, Jun Harashima, **Sosuke Nishikawa**
- *Annual Meeting of the Association for Natural Language Processing (NLP)*, 2021, Domestic Conference

[8] Studio Ousia at the NTCIR-15 SHINRA2020-ML Task
- NTCIR Conference on Evaluation of Information Access Technologies (NTCIR), 2020
- **Sosuke Nishikawa**, Ikuya Yamada
- 1st Place in German and French

[9] 教師なし機械翻訳の出力結果から作成した多言語分散表現の分析
- EN title: Analysis of Cross-Lingual Word Embeddings using Pseudo Corpus Generated from Unsupervised Machine Translation
- **Sosuke Nishikawa**, Yoshimasa Tsuruoka
- *Annual Meeting of the Association for Natural Language Processing (NLP)*, 2020, Domestic Conference

Books

大規模言語モデル入門 II 〜生成型 LLM の実装と評価（to appear）
- 山田育矢、鈴木正敏、西川荘介、藤井一喜、山田康輔、李凌寒
- 技術評論社, 2024

Awards

- Exempted from the Japan Student Services Organization's loan-based scholarship due to outstanding academic achievements — *2022*
- Received the Department Head's Award, Graduate school of information science and technology, The University of Tokyo — *2022*

Skills

Languages	Japanese (*native*), English (*conversant*, IELTS 6.0)
Programming Languages	Python, C/C++, JavaScript, HTML/CSS, MySQL
Libraries and Services	Pytorch, huggingface/transformers, Docker, Kubernetes, React.js

Visit https://sosuke.info to view more details on my projects and experiences.