Seiji Shimizu

Email: shimizu.seiji.so8@is.naist.jp Website: seiji-shimizu.github.io Google Scholar: link

Research Interests

Clinical Natural Language Processing, Large Language Models, Robustness under Domain Shift, Synthetic Data Generation, Privacy-Preserving Methods, Clinical Information Extraction, Multi-Agent Systems

Education

PhD in Information Science 2024 – Present
Nara Institute of Science and Technology (NAIST)
Advisor: Prof. Eiji Aramaki
Thesis: LLM-based Clinical Text Processing for Reliable Healthcare Applications (working title)

MSc in Information Science 2022–2023
Nara Institute of Science and Technology (NAIST)

BA in Language and Area Studies 2017–2021
Tokyo University of Foreign Studies

- Exchange Student, University of Texas at Austin, USA Aug 2018 – May 2019

Publications

1 A Herd of Language Models Makes a Better Zero-shot Annotator for Clinical Named Entity Recognition.
Seiji Shimizu, Shoko Wakamiya, Eiji Aramaki.
Findings of ACL, 2026.

2 J-ClinicalBench: A Benchmark for Evaluating Large Language Models on Practical Clinical Tasks in Japanese.
Seiji Shimizu et al.
LREC, 2026.

3 Exploring LLM Annotation for Adaptation of Clinical Information Extraction Models under Data-Sharing Restrictions.
Seiji Shimizu, Shohei Hisada, Yutaka Uno, Shuntaro Yada, Shoko Wakamiya, Eiji Aramaki.
Findings of ACL, 2025.

4 RecordTwin: Towards Creating Safe Synthetic Clinical Corpora.
Seiji Shimizu, Ibrahim Baroud, Lisa Raithel, Shuntaro Yada, Shoko Wakamiya, Eiji Aramaki.
Findings of ACL, 2025.

5 Toward Cross-Hospital Deployment of NLP Systems: Model Development and Validation of Fine-Tuned Large Language Models for Disease Name Recognition in Japanese.
Seiji Shimizu, Tomohiro Nishiyama, Hiroyuki Nagai, Shoko Wakamiya, Eiji Aramaki.
JMIR Medical Informatics, 2025.

6 Generating Distributable Surrogate Corpus for Medical Multi-Label Classification.
Seiji Shimizu, Shuntaro Yada, Shoko Wakamiya, Eiji Aramaki.
CL4Health @ LREC-COLING, 2024.

7 QA-based Event Start-Points Ordering for Clinical Temporal Relation Annotation.
Seiji Shimizu, Lis Pereira, Shuntaro Yada, Eiji Aramaki.
LREC-COLING, 2024.

8 Improving Self-Training with Prototypical Learning for Source-Free Domain Adaptation on Clinical Text.
Seiji Shimizu, Shuntaro Yada, Lisa Raithel, Eiji Aramaki.
BioNLP, 2024.

Research Experience

PhD Researcher, Social Computing Lab 2023 – Present
Nara Institute of Science and Technology

- Developed multi-LLM annotation methods for zero-shot clinical information extraction [1]
- Developed a benchmark for evaluating large language models on electronic health records (EHRs) across multiple real-world clinical tasks, as part of the JST CREST (Core Research for Evolutionary Science and Technology) BioDX project [Project Link] [2]
- Developed temporal relation extraction models for drug–disease interactions, integrated into a clinical knowledge base within the SIP (Cross-ministerial Strategic Innovation Promotion Program) integrated healthcare system project [Project Link] [7]
- Developed synthetic clinical corpus generation methods in collaboration with the German Research Center for Artificial Intelligence (DFKI), enabling privacy-preserving data sharing [4]
- Investigated domain shift across hospitals and document types in clinical NLP, and developed robust adaptation methods for real-world deployment [5, 8]
- Developed a clinical event extraction model for automated identification of lung cancer patients eligible for clinical trials, in collaboration with Fujitsu (manuscript under review)
- Developed a multi-agent system for clinical hypothesis generation targeting drug repurposing for chemotherapy-induced adverse events, as part of the JST CREST BioDX project (manuscript under review)

Research Intern, NEC Corporation Sep 2024 – Nov 2024

- Developed and evaluated large language model-based annotation methods for source-free domain adaptation under clinical data-sharing constraints [3]

Grants & Awards

NAIST Granite-AI Fellowship (Research Incentive + Research Grant) 2024 – Present
Program Link
- Total funding: JPY 3,900,000/year (JPY 3,000,000 stipend + JPY 900,000 research grant)

JST AIP Challenge Program (API Challenge Award) 2024
Program Link
- Awarded JPY 1,000,000

Outstanding Research Award, IPSJ SIG-NL 2024
Award Link

Technical Skills

Programming: Python
ML/NLP: PyTorch, Transformers
Languages: English, Japanese