

# Masahiro Suzuki

Tokyo | [msuzuki.me](https://msuzuki.me) | [linkedin.com/in/msuzuki7](https://linkedin.com/in/msuzuki7) | [github.com/retarfi](https://github.com/retarfi)

## Summary

---

Masahiro Suzuki is a quantitative analyst at Nikko Asset Management Co., Ltd. and a Ph.D. candidate at the University of Tokyo. His research focuses on natural language processing with applications mainly in finance. He actively contributes to the field through academic research and industry collaboration. He has contributed to open-source communities by developing and releasing Japanese pre-trained language models tailored for the financial sector. Through his work, he aims to advance NLP technologies and bridge the gap between cutting-edge research and real-world applications.

## Education

---

**The University of Tokyo**, Doctorate in Engineering – Hongo, Tokyo, Japan Oct 2022 – present  
Supervisor: Kiyoshi Izumi

- Natural Language Processing, Finance

**The University of Tokyo**, Master in Engineering – Hongo, Tokyo, Japan Apr 2020 – Mar 2022  
Supervisor: Kiyoshi Izumi

- Natural Language Processing, Finance
- GPA: 3.5

**The University of Tokyo**, Bachelor in Engineering – Hongo, Tokyo, Japan Apr 2016 – Mar 2020  
Supervisor: Kiyoshi Izumi

- Natural Language Processing, Finance
- GPA: 3.69

## Experience

---

**Quant Analyst**, Nikko Asset Management Co., Ltd. – Roppongi, Tokyo, Japan Apr 2022 – present

- Engaged in developing quantitative models for Japanese equities using deep learning
- Applied natural language processing to extract alpha and grasp economic trends, which resulted in three publications
- Designed and implemented a PoC to utilize generative AI within the company

**Intern**, Nowcast – Ichigaya, Tokyo, Japan Sept 2020 – Mar 2022  
Responsible for data analysis in the system for economic indices from point of sales (POS) data

**Intern**, Race – Hongo, Tokyo, Japan Mar 2020 – Mar 2022  
Develop a system to extract comments that need to be addressed from employee comments

## Open Source Projects

---

**Economy Watchers Survey Data** June 2024 – present

Developed an easy-to-use framework for the Economy Watchers Survey

- Automatically update the data monthly via GitHub Actions: [GitHub](#)
- The data is also available on Hugging Face Hub: [Dataset](#) / [Task Data](#)

**Japanese LLM Project** Apr 2023 – Dec 2023

Develop Japanese instruction dataset and tune large language model (LLM) pre-trained in English to Japanese with Japanese dataset

- Implementations for the dataset and training are available on GitHub: [Dataset](#) / [Training](#)
- Datasets and models are available on Hugging Face Hub: [Datasets](#) / [Models](#)

**Pre-training Language Models for Japanese** Dec 2020 – July 2023

Develop Japanese pre-trained models adapted to the financial domain

- Developed a framework which allows pre-training with different batch sizes for nodes and published on GitHub along with the pre-training implementations: [GitHub](#)
- Developed a Python package for easy Japanese word segmentation as [jptranstokenizer](#) and published on [PyPI](#)
- Both general and financial models are available on Hugging Face Hub: [General models](#) / [Financial models](#)

## Skills

---

**Programming:** Proficient with Python and Git; understanding of Web, app development

**Languages:** English (proficient), Japanese (native)

**Infrastructures:** Proficient with Ubuntu, Linux. Experienced with SFP modules, QNAP NAS, and YAMAHA RTX router. Managing on-premise servers and networks in the laboratory. Also experience with cloud computing such as Azure, Amazon EC2, GCP, etc

## Awards

---

**Best Paper Award** Feb 2024

Natural Language Understanding and Models of Communication Research Group

- [NLC Research Award 2023 Recipients \(in Japanese\)](#)

**Most Innovative Application of Artificial Intelligence in Japan** Feb 2024

Asia Asset Management 2024 Best of the Best Awards

- [Press Release from Nikko Asset Management Co., Ltd.](#)
- [Asia Asset Management Best of the Best Awards](#)

**Honorable Mention Award** Aug 2023

18th Symposium of Young Researcher Association for NLP Studies (YANS)

- [YANS Award \(in Japanese\)](#)

**ELYZA Award** Aug 2023

18th Symposium of Young Researcher Association for NLP Studies (YANS)

- [YANS Award \(in Japanese\)](#)

**TOYOTA/Dwango AI Scholarship** Apr 2020 – Mar 2021

1 year: 1,200,000 yen / Approx. 11,000 USD

**JEES / SoftBank AI Human Resources Development Scholarship** Apr 2020 – Mar 2021

1 year: 1,000,000 yen / Approx. 9,000 USD

**Best Paper Award** Nov 2019

The 2nd International Workshop on Cross-disciplinary Data Exchange and Collaboration (CDEC)

- [Release from the University of Tokyo](#)

## Publications (Refereed)

---

**Economy Watchers Survey provides Datasets and Tasks for Japanese Financial Domain** Apr 2025

*Masahiro Suzuki*, Hiroki Sakaji

The ACM Web Conference (to appear)

**Sentiment works in small-cap stocks: Japanese stock's sentiment with language models** Jan 2025

*Masahiro Suzuki*, Yasushi Ishikawa, Masayuki Teraguchi, Hiroki Sakaji

[10.1016/j.jjime.2024.100318](#) (International Journal of Information Management Data Insights)

**Impact of Information Disparity between Individual Investors on Profits of Meme Stocks Using an Artificial Market Simulation Approach** Jan 2025

Miyuki Matsumoto, Ryuji Hashimoto, *Masahiro Suzuki*, Yuri Murayama, Kiyoshi Izumi

[10.1007/s42001-024-00355-7](#) (Journal of Computational Social Science)

<b>Refined and Segmented Price Sentiment Indices from Survey Comments</b> <i>Masahiro Suzuki</i> , Hiroki Sakaji <a href="https://doi.org/10.1109/BigData62323.2024.10825585">10.1109/BigData62323.2024.10825585</a> (IEEE International Conference on Big Data)	Dec 2024
<b>Is ChatGPT the Future of Causal Text Mining? A Comprehensive Evaluation and Analysis</b> Takehiro Takayanagi, <i>Masahiro Suzuki</i> , Ryotaro Kobayashi, Hiroki Sakaji, Kiyoshi Izumi <a href="https://doi.org/10.1109/BigData62323.2024.10825555">10.1109/BigData62323.2024.10825555</a> (IEEE International Conference on Big Data)	Dec 2024
<b>Sentiment-driven Stock Selection in Japan using Language Models</b> <i>Masahiro Suzuki</i> , Hiroki Sakaji IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr)	Oct 2024
<b>JaFin: Japanese Financial Instruction Dataset</b> Kota Tanabe, <i>Masahiro Suzuki</i> , Hiroki Sakaji, Itsuki Noda <a href="https://doi.org/10.1109/CIFEr62890.2024.10772973">10.1109/CIFEr62890.2024.10772973</a> (IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr))	Oct 2024
<b>FinDeBERTaV2: Word-Segmentation-Free Pre-trained Language Model for Finance</b> <i>Masahiro Suzuki</i> , Hiroki Sakaji, Masanori Hirano, Kiyoshi Izumi <a href="https://doi.org/10.1527/tjsai.39-4_FIN23-G">10.1527/tjsai.39-4_FIN23-G</a> (Transactions of the Japanese Society for Artificial Intelligence (in Japanese))	July 2024
<b>Development and analysis of medical instruction-tuning for Japanese large language models</b> Issey Sukeda, <i>Masahiro Suzuki</i> , Hiroki Sakaji, Satoshi Kodera <a href="https://doi.org/10.36922/aih.2695">10.36922/aih.2695</a> (Artificial Intelligence in Health)	Apr 2024
<b>JMedLoRA: Medical Domain Adaptation on Japanese Large Language Models using Instruction-tuning</b> Issey Sukeda, <i>Masahiro Suzuki</i> , Hiroki Sakaji, Satoshi Kodera <a href="https://openreview.net/forum?id=BfHX0hKRSe">openreview.net/forum?id=BfHX0hKRSe</a> (Deep Generative Models for Health Workshop NeurIPS)	Dec 2023
<b>From Base to Conversational: Japanese Instruction Dataset and Tuning Large Language Models</b> <i>Masahiro Suzuki</i> , Masanori Hirano, Hiroki Sakaji <a href="https://doi.org/10.1109/BigData59044.2023.10386605">10.1109/BigData59044.2023.10386605</a> (IEEE International Conference on Big Data)	Dec 2023
<b>llm-japanese-dataset v0: Construction of Japanese Chat Dataset for Large Language Models and its Methodology</b> Masanori Hirano, <i>Masahiro Suzuki</i> , Hiroki Sakaji <a href="https://doi.org/10.1007/978-3-031-40978-3_47">10.1007/978-3-031-40978-3_47</a> (The 12th International Workshop on Web Services and Social Media in The 26th International Conference on Network-Based Information Systems)	Aug 2023
<b>Constructing and analyzing domain-specific language model for financial text mining</b> <i>Masahiro Suzuki</i> , Hiroki Sakaji, Masanori Hirano, Kiyoshi Izumi <a href="https://doi.org/10.1016/j.ipm.2022.103194">10.1016/j.ipm.2022.103194</a> (Information Processing & Management)	Mar 2023
<b>Gradual Further Pre-training Architecture for Economics/Finance Domain Adaptation of Language Model</b> Hiroki Sakaji, <i>Masahiro Suzuki</i> , Kiyoshi Izumi, Hiroyuki Mitsugi <a href="https://doi.org/10.1109/BigData55660.2022.10020445">10.1109/BigData55660.2022.10020445</a> (IEEE International Conference on Big Data)	Dec 2022
<b>Constructing and analyzing domain-specific language model for financial text mining</b> <i>Masahiro Suzuki</i> , Hiroki Sakaji, Masanori Hirano, Kiyoshi Izumi Information Processing and Management Conference	Oct 2022

- Forecasting Stock Price Trends by Analyzing Economic Reports With Analyst Profiles** June 2022  
*Masahiro Suzuki*, Hiroki Sakaji, Kiyoshi Izumi, Yasushi Ishikawa  
[10.3389/frai.2022.866723](#) (Frontiers in Artificial Intelligence)
- Market Trend Analysis Using Polarity Index Generated from Analyst Reports** Dec 2021  
Rei Taguchi, Hikaru Watanabe, Masanori Hirano, *Masahiro Suzuki*, Hiroki Sakaji, Kiyoshi Izumi  
[10.1109/BigData52589.2021.9671702](#) (IEEE International Conference on Big Data)
- Forecasting Net Income Estimate and Stock Price Using Text Mining from Economic Reports** May 2020  
*Masahiro Suzuki*, Hiroki Sakaji, Kiyoshi Izumi, Yasushi Ishikawa  
[10.3390/info11060292](#) (Information)
- Stock Price Analysis Using Combination of Analyst Reports and Several Document** Nov 2019  
*Masahiro Suzuki*, Toshiya Katagi, Hiroki Sakaji, Kiyoshi Izumi, Yasushi Ishikawa  
[10.1109/ICDMW.2019.00013](#) (The 2nd International Workshop on Cross-disciplinary Data Exchange and Collaboration in International Conference on Data Mining (ICDM))