



# Sotaro Takeshita

PHD CANDIDATE

✉ [oh.sore.sore.soutarou@gmail.com](mailto:oh.sore.sore.soutarou@gmail.com) | 🌐 [sotaro.io/about](https://sotaro.io/about) | 📷 [sobamchan](#) | 📺 [sotaro-takeshita-920a99198](#) | 🐦 [@sobamchan](#) | 🎓 Sotaro Takeshta

## Education

### University of Mannheim

PH.D. STUDENT IN NATURAL LANGUAGE PROCESSING

- Text summarization for scholarly documents.
- Supervisor: Prof. Dr. Simone Paolo Ponzetto

Mannheim, Germany

Sep. 2022 - Present

### The National University of Electro-Communication

M.S. IN COMPUTER SCIENCE

- Thesis: Multimodal models for commonsense reasoning.
- Supervisor: Prof. Dr. Yasuhiro Minami

Tokyo, Japan

Apr. 2018 - Mar. 2020

### The National University of Electro-Communication

B.S. IN COMPUTER SCIENCE

- Thesis: Unsupervised text summarization with reinforcement learning.
- Supervisor: Prof. Dr. Yasuhiro Minami

Tokyo, Japan

Apr. 2013 - Mar. 2018

## Publication

### JOURNAL

#### Cross-lingual extreme summarization of scholarly documents

SOTARO TAKESHITA, TOMMASO GREEN, NIKLAS FRIEDRICH, KAI ECKERT AND SIMONE PAOLO PONZETTO

International Journal on Digital Libraries (IJDL)

2023

### CONFERENCE

#### GenGO: ACL Paper Explorer with Semantic Features

SOTARO TAKESHITA, SIMONE PAOLO PONZETTO, KAI ECKERT

The 62nd Annual Meeting of the Association for Computational Linguistics, System Demonstrations

Bangkok, Thailand

2024

#### ROUGE-K: Do Your Summaries Have Keywords?

SOTARO TAKESHITA, SIMONE PAOLO PONZETTO, KAI ECKERT

The 13th Joint Conference on Lexical and Computational Semantics colocated at NAACL 2024, (Best Long Paper)

Mexico City, Mexico

2024

#### ACLSum: A New Dataset for Aspect-based Summarization of Scientific Publications

SOTARO TAKESHITA, TOMMASO GREEN, INES REINIG, KAI ECKERT AND SIMONE PAOLO PONZETTO

In Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)

Mexico City, Mexico

2024

#### ZusammenQA: Data Augmentation with Specialized Models for Cross-lingual Open-retrieval Question Answering System

CHIA-CHIEN HUNG, TOMMASO GREEN, ROBERT LITSCHKO, TORNIKE TSERETELI, SOTARO TAKESHITA, MARCO BOMBIERI, GORAN GLAVAŠ, SIMONE PAOLO PONZETTO

Workshop on Multilingual Information Access (MIA) at NAACL

Seattle, Washington + Online

2022

#### X-SCITLDR: Cross-Lingual Extreme Summarization of Scholarly Documents

SOTARO TAKESHITA, TOMMASO GREEN, NIKLAS FRIEDRICH, KAI ECKERT AND SIMONE PAOLO PONZETTO

ACM/IEEE Joint Conference on Digital Libraries (JCDL)

Cologne, Germany and Online

2022

### UNDER REVIEW

#### IRSum: One Model to Rule Summarization and Retrieval

SOTARO TAKESHITA, SIMONE PAOLO PONZETTO, KAI ECKERT

2024

#### Zero-cost Text Embedding Dimension Reduction for Text Retrieval

SOTARO TAKESHITA, YURINA TAKESHITA, SIMONE PAOLO PONZETTO, KAI ECKERT

2024

## Experience

---

### VADIS (VARIABLE Detection, Interlinking and Summarization)

RESEARCHER

Germany

2022 - Present

- A project funded by German Research Foundation (DFG)
- Development of a summarization system for social science research articles.

### Build It Inc.

DATA ANALYST & ML ENGINEER

Tokyo, Japan

2015 - 2022

- Analysis of user behaviours using statistical and machine learning tools
- Development of recommendation SaaS

### The University of Campinas, Faculty of Electrical and Computer Engineering

RESEARCH INTERNSHIP

Campinas, Brazil

Nov. 2018

- Developed Neural Network-based Computer Vision system for self-driving wheelchairs.
- Advisor: Prof. Dr. Eric Rohmer

### NTT Media Intelligence Lab

RESEARCH INTERNSHIP

Kanagawa, Japan

Sep. 2018

- Development of an NLP system to automatically generate funny stories in Japanese.
- Advisor: Dr. Ryuichiro Higashinaka

### Recruit Holdings

DATA SCIENCE INTERNSHIP

Tokyo, Japan

Aug. 2017

- Development of an automatic scoring system for essays.
- Prized with a best intern award

### Sección de Estudios de Posgrado e Investigación de ESIME Culhuacan, Instituto Politecnico Nacional

RESEARCH INTERNSHIP

Mexico City, Mexico

Feb. 2016 - Apr. 2016

- Analysis on models of automatic face recognition models.
- Advisor: Prof. Dr. Mariko Nakano Miyatake

## Projects

---

**GenGO: ACL Paper Explorer with Semantic Features**, <https://gengo.sotaro.io>

**GenGO-Chat: LLM-powered NLP paper exploration system.**, <https://gengo-chat.sotaro.io/>

## Talks

---

- |      |                                                                                              |               |
|------|----------------------------------------------------------------------------------------------|---------------|
| 2024 | <b>Scholarly Document Processing: Language Technologies to Help Scientist</b> , Invited Talk | DESY, Germany |
| 2023 | <b>Overview of Tasks and Challenges in Automatic Text Summarization</b> , Invited Talk       | NEC, Germany  |