Hayate **Iso**

PhD in Engineering · NLP Researcher

444 Castro Street Suite 720 Mountain View, CA 94041

(+1) 650-439-7418 | hyate.iso@gmail.com | isomap.github.io | isomap | hayate-iso | Google Scholar

Research Interests

Large Language Models	Retrieval Augmented Generation[2], Long Context Understanding[5], Uncertainty in Instruction[1,4]
Text Generation	Opinion Summarization[7,9,10], Data-to-Text Generation[14], Fact-based Text Editing[13]

Work Experience

Megagon Labs *Mountain View, CA*

Senior Research Scientist *Jan. 2021 - Present*

- Led the comparative opinion summarization project, resulting in an ACL Findings paper and an internal R&D Award. Integrated this project into the hotel reservation service chatbot.
- Presented work at the largest internal knowledge-sharing event, recognized as top 0.1% of employees and awarded Best Presenter.
- Mentored four interns and two junior researchers, leading to five published papers with significant contributions as second author.
- Implemented state-of-the-art LLM tools (vLLM, LLaMA-index, FAISS), enhancing productivity lab-wide.
- **Selected Artifacts:** [NAACL 2024] [LREC-COLING 2024] [EACL 2024] [IJCNLP-AACL 2023] [ACL 2022 Findings] [R&D Award (Internal)] [Best Presenter at Cross/Engine Forum (Internal)]

Amazon *Seattle, WA (remote)*

Applied Scientist Intern *Jul. 2020 - Oct. 2020*

- Developed a text editing model for Alexa, correcting Speech-to-Text errors using external knowledge, enhancing user interaction quality.

Megagon Labs *Mountain View, CA*

Research Intern *Feb. 2020 - Jul. 2020*

- Developed an unsupervised opinion summarization method using VAE and a convex aggregation approach, achieving state-of-the-art performance.
- **Artifacts:** [EMNLP 2021 Findings] [Code] [R&D Award (Internal)]

ByteDance *Beijing, China*

Research Intern *May. 2019 - Nov. 2019*

- Worked on an automated system for updating outdated text using a knowledge base, enhancing content accuracy and relevance.
- Proposed a new text editing task, fact-based text editing, in which the goal is to revise a given document to better describe the facts in a knowledge base.
- **Artifacts:** [ACL 2020] [Code] [US Patent]

Advanced Industrial Science and Technology *Tokyo, Japan*

Technical Trainee *Oct. 2018 - Dec. 2018*

- Developed a data-to-text model that tracks entity state while generating long texts from large structured data.
- **Artifacts:** [ACL 2019] [Code] [Data]

Japan Society for the Promotion of Science *Tokyo, Japan*

Research Fellowships for Young Scientists (DC1) *Apr. 2018 - Feb. 2020*

- Awarded the most prestigious governmental research fellowship in Japan.
- Received a monthly stipend of 200,000 JPY and an annual research grant of approximately 800,000 JPY for up to three years. Resigned in February 2020 to participate in a research internship.

RaRe Technologies *Remote*

Student Incubator *Jul. 2016 - Sep. 2016*

- Enhanced the GenSim library by implementing online vocabulary updating functionality for Word2Vec, supporting continual learning.
- Developed comprehensive tutorial notebooks to guide users on the effective application of Doc2Vec for continual learning scenarios.

Publication

1. ***AmbigNLG: Addressing Task Ambiguity in Instruction for NLG***

   Ayana Niwa, **Hayate Iso**

   pre-print arXiv: 2402.17717, 2024 **[Outstanding Young Researcher's Paper (18/427 $\approx$ 4.2%) in ANLP]**

2. ***A Blueprint Architecture of Compound AI Systems for Enterprise***

   Eser Kandogan, Sajjadur Rahman, Nikita Bhutani, Dan Zhang, Rafael Li Chen, Kushan Mitra, Sairam Gurajada, Pouya Pezeshkpour, **Hayate Iso**, Yanlin Feng, Hannah Kim, Chen Shen, Jin Wang, Estevam Hruschka

In *Compound AI Systems workshop*, 2024

3. ***Retrieval Helps or Hurts? A Deeper Dive into the Efficacy of Retrieval Augmentation to Language Models***
Seiji Maekawa, **Hayate Iso**, Sairam Gurajada, Nikita Bhutani
In *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2024 **[Oral, Featured in Engine Forum'24]**

4. ***Distilling Large Language Models using Skill-Occupation Graph Context for HR-Related Tasks***
Pouya Pezeshkpour, **Hayate Iso**, Thom Lake, Nikita Bhutani, Estevam Hruschka
pre-print arXiv:2311.06383, 2023

5. ***XATU: A Fine-grained Instruction-based Benchmark for Explainable Text Updates***
Haopeng Zhang, **Hayate Iso**, Sairam Gurajada, Nikita Bhutani
In *Joint International Conference On Computational Linguistics, Language Resources And Evaluation (LREC-COLING)*, 2024

6. ***Less is More for Long Document Summary Evaluation by LLMs***
Yunshu Wu*, **Hayate Iso***, Pouya Pezeshkpour, Nikita Bhutani, Estevam Hruschka (*equal contribution)
In *European Chapter of the Association for Computational Linguistics (EACL)*, 2024 **[Oral & Poster, Featured in Engine Forum'24]**

7. ***Zero-shot Triplet Extraction by Template Infilling***
Bosung Kim, **Hayate Iso**, Nikita Bhutani, Estevam Hruschka, Ndapa Nakashole, Tom M. Mitchell
In *Asia-Pacific Chapter of the Association for Computational Linguistics and International Joint Conference on Natural Language Processing (AACL-IJCNLP)*, 2023 **[Oral, Featured in Engine Forum'23]**

8. ***Noisy Pairing and Partial Supervision for Opinion Summarization***
Hayate Iso, Xiaolan Wang, Yoshihiko Suhara
In *International Natural Language Generation Conference (INLG)*, 2024

9. ***AutoTemplate: A Simple Recipe for Lexically Constrained Text Generation***
Hayate Iso
In *International Natural Language Generation Conference (INLG)*, 2024

10. ***Comparative Opinion Summarization via Collaborative Decoding***
Hayate Iso, Xiaolan Wang, Stefanos Angelidis, Yoshihiko Suhara
In *Findings of the Association for Computational Linguistics (ACL)*, 2022 **[R&D Award'22, Featured in Cross/Engine Forum'23]**

11. ***Convex Aggregation for Opinion Summarization***
Hayate Iso, Xiaolan Wang, Yoshihiko Suhara, Stefanos Angelidis, Wang-Chiew Tan
In *Findings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2021 **[R&D Award'21]**

12. ***End-to-end Biomedical Entity Linking with Span-based Dictionary Matching***
Shogo Ujiie, **Hayate Iso**, Shuntaro Yada, Shoko Wakamiya, Eiji Aramaki
In *Biomedical Natural Language Processing Workshop (BioNLP)*, 2021

13. ***Biomedical Entity Linking with Contrastive Context Matching***
Shogo Ujiie, **Hayate Iso**, Eiji Aramaki
preprint arXiv:2106.07583, 2021

14. ***Fact-based Text Editinng***
Hayate Iso, Chao Qiao, Hang Li
In *Proceedings of Association for Computational Linguistics (ACL)*, 2020 **[US Patent 11,669,679]**

15. ***Learning to Select, Track, and Generate for Data-to-Text***
Hayate Iso, Yui Uehara, Tatsuya Ishigaki, Hiroshi Noji, Eiji Aramaki, Ichiro Kobayashi, Yusuke Miyao, Naoaki Okazaki, Hiroya Takamura
In *Proceedings of Association for Computational Linguistics (ACL)*, 2019.

16. ***Wisdom in Adversity: A Twitter Study of the Japanese Tsunami***
Paolo Casani, **Hayate Iso**, Shoko Wakamiya, Eiji Aramaki
Advances in Social Networks Analysis and Mining (ASONAM), 2018

17. ***NTCIR13 MedWeb Task: Multi-label Classification of Tweets using an Ensemble of Neural Networks***
Hayate Iso, Camille Ruiz, Taichi Murayama, Katsuya Taguchi, Ryo Takeuchi, Hideya Yamamoto, Shoko Wakamiya, Eiji Aramaki
In *Proceedings of NTCIR Conference on Evaluation of Information Access Technologies*, 2017 **[1st place in all tracks, Best System Award]**

18. ***Density Estimation for Geolocation via Convolutional Mixture Density Network.***
Hayate Iso, Shoko Wakamiya, Eiji Aramaki
arXiv preprint arXiv:1705.02750, 2017.

19. ***Conditional Density Estimation of Tweet Location: A Feature-Dependent Approach***
   <u>Hayate Iso</u>, Shoko Wakamiya, Eiji Aramaki
   Medinfo, 2017

20. ***Forecasting Word Model: Twitter-based Influenza Surveillance and Prediction***
   <u>Hayate Iso</u>, Shoko Wakamiya, Eiji Aramaki
   In *Proceedings of International Conference on Computational Linguistics (COLING)*, 2016 **[Oral, Outstanding Research in IPSJ-SIGNL]**

Invited Talks

Engine Forum   *Tokyo, Japan*
Leveraging GenAI to Nudge the Last Mile Decision-Making with Comparative Opinion Summarization   *June. 2023*

Cross Forum   *Tokyo, Japan*
Leveraging GenAI to Nudge the Last Mile Decision-Making with Comparative Opinion Summarization   *May. 2023*

Young Researcher Association for NLP Studies   *Remote*
Opinion Summarization from Customer Reviews   *Aug. 2022*

Honors & Awards

2023	**Best Presenter**, Cross/Engine Forum
2022	**R&D Award**, Recruit Holdings
2021	**R&D Award**, Recruit Holdings
2018	**Exemption the full amount of scholarship repayment**, JASSO
2017	**Best System Award (1st place)**, NTCIR13–MedWeb
2016	**Outstanding Research Award**, IPSJ-SIGNL
2016	**Extracurricular Research Funds ($10,000)**, Osaka University

Education

Nara Institute of Science and Technology   *Nara, Japan*
Ph.D. in Information Science   *Oct. 2017 - Mar. 2021*

Nara Institute of Science and Technology   *Nara, Japan*
M.E. in Information Science   *Apr. 2016 - Sep. 2017*

Osaka University   *Osaka, Japan*
B.S. in Mathematical Science   *Apr. 2014 - Mar. 2016*

National Institute of Technology   *Ibaraki, Japan*
A.E. in Electrical and Electronic Engineering   *Apr. 2009 - Mar. 2014*

Program Committees

2024	**Program Committee**, ACL Rolling Review
2023	**Program Committee**, ACL Rolling Review
2023	**Program Committee**, EMNLP Industry Track
2023	**Program Committee**, EMNLP
2023	**Program Committee**, CoNLL
2023	**Program Committee**, EACL
2023	**Program Committee**, AACL-IJCNLP
2022	**Program Committee**, ACL Rolling Review
2022	**Program Committee**, EMNLP
2022	**Program Committee**, AACL-IJCNLP
2021	**Program Committee**, ACL Rolling Review
2021	**Program Committee**, ACL-IJCNLP.
2020	**Program Committee**, AACL-IJCNLP.
2020	**Program Committee**, Young Researcher Association for NLP Studies (YANS).
2019	**Program Committee**, Young Researcher Association for NLP Studies (YANS).
2018	**Program Committee**, COLING.