Akari Asai

E-mail: akari@cs.washington.edu
Web: akariasai.github.io/

EDUCATION

University of Washington, Seattle, Washington — September 2019 - Present
Ph.D. Student, Computer Science and Engineering
Advisor: Prof. Hannaneh Hajishirzi

The University of Tokyo, Tokyo, Japan — March 2019
B.E. in Information and Communication Engineering
Thesis Advisor: Kiyoharu Aizawa
Major GPA: 3.90/4, Cumulative GPA: 3.83/4; 1st rank in class (1/133), Dean's Award

University of California, Berkeley, Berkeley, California — August 2015 - May 2016
American and International Study Program (AISP) at Department of Political Science

EMPLOYMENT

University of Washington — Sep 2019-Present, Seattle, WA
Ph.D. Student (Supervisor: Hannaneh Hajishirzi)

Meta AI
Research Intern (Supervisors: Sebastian Riedel, Scott Yih) — Jul 2022-Jan 2023, Seattle, WA
Visiting Student Researcher (Present; Supervisor: Scott Yih) — Sep 2023-Present, Seattle, WA

Allen Institute for AI — Jul 2021-Dec 2021, Seattle, WA
Research Intern (Supervisors: Hannaneh Hajishirzi, Matt Gardner)

Salesforce Research — Apr 2019-Aug 2019, Palo Alto, CA
Research Intern (Supervisors: Kazuma Hashimoto, Caiming Xiong)

Microsoft Research Asia — Sep 2018-Jan 2019, Beijing, China
Research Intern (Supervisors: Duyu Tang, Nan Duan)

Aizawa Yamasaki Lab, The University of Tokyo — Apr 2018-Mar 2019, Tokyo, Japan
Thesis Student (Supervisor: Kiyoharu Aizawa)

Studio Ousia — Mar 2018-Mar 2019, Tokyo, Japan
Research Intern (Supervisor: Ikuya Yamada)

Tsuruoka Lab, The University of Tokyo — Dec 2017-Mar 2019, Tokyo, Japan
Research Assistant (Supervisor: Yoshimasa Tsuruoka)

Google, Google Chrome Blink team — Aug-Oct 2017, Tokyo, Japan
Engineering Intern

Megagon Labs — Mar-Apr, 2017, Tokyo, Japan
Research Intern (Supervisor: Alon Y. Halevy)

Google, iGSA (Google Search iOS) team — Jun-Oct 2016, Tokyo, Japan
Engineering Intern

SELECTED AWARDS & GRANT

Best Paper Honorable Mention at NeurIPS 2023 Instruction Workshop 2023
ACL 2023 Best Video Award 2023
Microsoft Accelerate Foundation Models Research 2023
Stability AI HPC Cluster Research Grant 2023
EECS Rising Stars 2022
IBM PhD Fellowship 2022
The Dean's Award, The Best Bachelor Thesis Award, The University of Tokyo 2019
The Nakajima Foundation Scholarship (Doctoral Study Abroad Program Fellowship) 2018
Women Techmakers Scholarship (Google Anita Borg Memorial Scholarship) Scholar 2017

PUBLICATIONS

[1] Scaling Retrieval-Based Language Models with a Trillion-Token Datastore.
Rulin Shao, Jacqueline He, **Akari Asai**, Weijia Shi, Tim Dettmers, Sewon Min, Luke Zettlemoyer, Pang

Wei Koh
Conference on Neural Information Processing Systems (NeurIPS), 2024.

[2] CopyBench: Measuring Literal and Non-Literal Reproduction of Copyright-Protected Text in Language Model Generation.
Tong Chen, **Akari Asai**, Niloofar Mireshghallah, Sewon Min, James Grimmelmann, Yejin Choi, Hannaneh Hajishirzi, Luke Zettlemoyer, Pang Wei Koh
Conference on Empirical Methods in Natural Language Processing (EMNLP), 2024.

[3] Fine-grained Hallucination Detection and Editing for Language Models.
Abhika Mishra, **Akari Asai**, Vidhisha Balachandran, Yizhong Wang, Graham Neubig, Yulia Tsvetkov, and Hannaneh Hajishirzi
Conference on Language Modeling (COLM), 2024.

[4] BUFFET: Benchmarking Large Language Models for Few-shot Cross-lingual Transfer.
Akari Asai, Sneha Kudugunta, Xinyan Velocity Yu, Terra Blevins, Hila Gonen, Machel Reid, Yulia Tsvetkov, Sebastian Ruder, Hannaneh Hajishirzi
Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL; *Oral*), 2024.

[5] Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection.
Akari Asai, Zeqiu Wu, Yizhong Wang, Avirup Sil, Hannaneh Hajishirzi
International Conference on Learning Representations (ICLR; Oral – Top 1%), 2024
Best Paper Honorable Mention at *Instruction Workshop at NeurIPS*, 2023.

[6] RealTime QA: What's the Answer Right Now?.
Jungo Kasai, Keisuke Sakaguchi, Yoichi Takahashi, Ronan Le Bras, **Akari Asai**, Xinyan Yu, Dragomir Radev, Noah A. Smith, Yejin Choi, Kentaro Inui
NeurIPS (Datasets and Benchmarks), 2023.

[7] TaskWeb: Selecting Better Source Tasks for Multi-task NLP.
Joongwon Kim, **Akari Asai**, Gabriel Ilharco, Hannaneh Hajishirzi.
Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023.

[8] How to Train Your DRAGON: Diverse Augmentation Towards Generalizable Dense Retrieval.
Sheng-Chieh Lin, **Akari Asai**, Minghan Li, Barlas Oguz, Jimmy Lin, Yashar Mehdad, Wen-tau Yih, Xilun Chen.
Findings of EMNLP, 2022.

[9] AfriQA: Cross-lingual Open-Retrieval Question Answering for African Languages.
Odunayo Ogundepo, Tajuddeen R. Gwadabe, Clara E. Rivera, Jonathan H. Clark, Sebastian Ruder, David Ifeoluwa Adelani, Bonaventure F. P. Dossou, ..., **Akari Asai** (52 authors).
Findings of EMNLP, 2022.

[10] Retrieval-based Language Models and Applications.
Akari Asai, Sewon Min, Zexuan Zhong, Danqi Chen
The Annual Conference of the Association for Computational Linguistics (ACL) Tutorial, 2023.

[11] When Not to Trust Language Models: Investigating Effectiveness and Limitations of Parametric and Non-Parametric Memories.
Alex Mallen*, **Akari Asai***, Victor Zhong, Rajarshi Das, Hannaneh Hajishirzi, Daniel Khashabi
The Annual Conference of the Association for Computational Linguistics (ACL; Oral), 2023.
ACL 2023 Best Video Award.

[12] Task-aware Retrieval with Instructions.
Akari Asai, Timo Schick, Patrick Lewis, Xilun Chen, Gautier Izacard, Sebastian Riedel, Hannaneh Hajishirzi, Wen-tau Yih
Findings of ACL, 2023.

[13] xPQA: Cross-Lingual Product Question Answering across 12 Languages.
Xiaoyu Shen, **Akari Asai**, Bill Byrne, Adrià de Gispert
The Annual Conference of the Association for Computational Linguistics (ACL; Industry track), 2023.

[14] Attentional Mixtures of Soft Prompt Tuning for Parameter-efficient Multi-task Knowledge Sharing.
Akari Asai, Mohammadreza Salehi, Matthew E. Peters, and Hannaneh Hajishirzi.
Conference on Empirical Methods in Natural Language Processing (EMNLP), 2022.

[15] Beyond Counting Datasets: Investigating Multilingual Dataset Construction and Necessary Resources.

Xinyan Yu*, **Akari Asai***, Trina Chatterjee, Junjie Hu, Eunsol Choi.
Findings of EMNLP, 2022.

[16] MIA 2022 Shared Task: Evaluating Cross-lingual Open-Retrieval Question Answering for 16 Diverse Languages.
Akari Asai, Shayne Longpre, Jungo Kasai, Chia-Hsuan Lee, Rui Zhang, Junjie Hu, Ikuya Yamada, Jonathan H. Clark, Eunsol Choi
Workshop on Multilingual Information Access (MIA), 2022.

[17] Evidentiality-guided Generation for Knowledge-Intensive NLP Tasks.
Akari Asai, Matt Gardner, and Hannaneh Hajishirzi
Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL; *Oral*), 2022.

[18] One Question Answering Model for Many Languages with Cross-lingual Dense Passage Retrieval.
Akari Asai, Xinyan Yu, Jungo Kasai, and Hannaneh Hajishirzi.
Conference on Neural Information Processing Systems (NeurIPS), 2021.

[19] Challenges in Information Seeking QA: Unanswerable Questions and Paragraph Retrieval.
Akari Asai and Eunsol Choi.
The Annual Conference of the Association for Computational Linguistics (ACL), 2021.

[20] Efficient Passage Retrieval with Hashing for Open-domain Question Answering.
Ikuya Yamada, **Akari Asai** and Hannaneh Hajishirzi.
The Annual Conference of the Association for Computational Linguistics (ACL), 2021.

[21] XOR QA: Cross-lingual Open-Retrieval Question Answering.
Akari Asai, Jungo Kasai, Jonathan H. Clark, Kenton Lee, Eunsol Choi, and Hannaneh Hajishirzi.
Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL; *Oral*), 2021.

[22] MultiModalQA: complex question answering over text, tables and images.
Alon Talmor, Ori Yoran, Amnon Catav, Dan Lahav, Yizhong Wang, **Akari Asai**, Gabriel Ilharco, Hannaneh Hajishirzi, Jonathan Berant.
International Conference on Learning Representations (ICLR), 2021.

[23] LUKE: Deep Contextualized Entity Representations with Entity-aware Self-attention.
Ikuya Yamada, **Akari Asai**, Hiroyuki Shindo, Hideaki Takeda, and Yuji Matsumoto.
Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.

[24] Wikipedia2Vec: An Optimized Tool for Learning Embeddings of Words and Entities from Wikipedia.
Ikuya Yamda, **Akari Asai**, Jin Sakuma, Hiroyuki Shindo, Hideaki Takeda, Yoshiyasu Takefuji, and Yuji Matsumoto.
Conference on Empirical Methods in Natural Language Processing (EMNLP, demo), 2020.

[25] Logic-guided Data Augmentation and Regularization for Consistent Question Answering.
Akari Asai and Hannaneh Hajishirzi.
The Annual Conference of the Association for Computational Linguistics (ACL), 2020.

[26] Learning to Retrieve Reasoning Paths over Wikipedia Graphs for Question Answering
Akari Asai, Kazuma Hashimoto, Hannaneh Hajishirzi, Richard Socher, and Caiming Xiong.
International Conference on Learning Representations (ICLR), 2020.

BEFORE PH.D. [27] The Aleatoric Uncertainty Estimation Using a Separate Formulation with Virtual Residuals.
Takumi Kawashima, Qing Yu, **Akari Asai**, Daiki Ikami, Kiyoharu Aizawa.
International Conference on Pattern Recognition (ICPR), 2020.

[28] Multi-task Learning Based on Separable Formulation of Depth Estimation and Uncertainty.
Akari Asai, Daiki Ikami, and Kiyoharu Aizawa.
Conference on Computer Vision and Pattern Recognition (CVPR), Uncertainty and Robustness in Deep Visual Learning, 2019.

[29] HappyDB: A Corpus of 100,000 Crowdsourced Happy Moments
Akari Asai, Sara Evensen, Behzad Golshan, Alon Halevy, Vivian Li, Andrei Lopatenko, Daniela Stepanov, Yoshihiko Suhara, Wang-Chiew Tan, and Yinzhan Xu.
International Conference on Language Resources and Evaluation (LREC), 2018.

[30] A Data-Driven Approach to Understanding Happiness

Akari Asai, Vivian Li, Daniela Stepanov, and Wang-Chiew Tan.
ACL Widening NLP Workshop, 2017.

PREPRINTS

[31] Quantifying the Influence of Evaluation Aspects on Long-Form Response Assessment.
Go Kamoda, **Akari Asai**, Ana Brassard, Keisuke Sakaguchi
Arxiv Preprint, 2024.

[32] Reliable, Adaptable, and Attributable Language Models with Retrieval.
Akari Asai, Zexuan Zhong, Danqi Chen, Pang Wei Koh, Luke Zettlemoyer, Hannaneh Hajishirzi, Wen-tau Yih
Arxiv Preprint, 2024.

[33] Adv-bert: BERT is not robust on misspellings! generating nature adversarial samples on bert.
Lichao Sun, Kazuma Hashimoto, Wenpeng Yin, **Akari Asai**, Jia Li, Philip Yu, Caiming Xiong
Arxiv Preprint, 2020.

[34] Multilingual Extractive Reading Comprehension by Runtime Machine Translation.
Akari Asai, Akiko Eriguchi, Kazuma Hashimoto, and Yoshimasa Tsuruoka.
Arxiv Preprint, 2018

TEACHING EXPERIENCE

Courses:

Head TA / Co-instructor, CSE599J Data-centric ML, University of Washington	Winter 2024
Head TA, CSE473 Artificial Intelligence, University of Washington	Fall 2023

Tutorial Instructor:

Tutorial on Retrieval-based LMs and Applications, ACL	July 2023

Guest Lectures for Courses:

CS 769 Advanced Natural Language Processing, University of Wisconsin-Madison (expected)	November 2024
CS 11-667 Large Language Models: Methods and Applications, CMU	September 2024
CPSC 477/577 Natural Language Processing, Washington University in St. Louis	April 2024
CS 11-711 Advanced NLP, Carnegie Mellon University	April 2024
CPSC 477/577 Natural Language Processing, Yale University	March 2024
CSE 447 Natural Language Processing, University of Washington	March 2024
CSE 473 Introduction to Artificial Intelligence, University of Washington	October 2023
CSE 447 Natural Language Processing, University of Washington	March 2023
CSE 373 Data Structures and Algorithm, University of Washington	December 2022

SERVICES

Departmental Service:

UW CSE Ph.D. Admission Committee Member (2021, 2022)
UW CSE Ph.D. Admission NLP Student Area Chair (2021, 2022)
Diversity and Inclusion Committee Representative (2022-2023)
UW CSE Faculty Hiring Student DEI Reviewer (2022-2023)
UW NLP Seminar Organizer (2023, 2024)

Workshop and Conference Organization:

Lead Organizer of Workshop on Multilingual Information Access (MIA), NAACL 2022
Shared-task Lead Organizer MIA 2022 Shared task, NAACL 2022.
Student Volunteer Coordinator, NAACL 2022.

Area Chair:

EACL 2023
EMNLP 2023
EMNLP 2024

Conference Reviewer/Program Committee:

ACL 2020-2023
EMNLP 2019-2023
ARR 2021-Present.
NeurIPS 2021-2024
NeurIPS Dataset and Benchmark, 2024
ICLR 2021-2024
AKBC 2020
AAAI 2020
IJCAI 2021
COLM 2024
Workshops: WiNLP 2024

Outreach:
Weekly Virtual Office hour, offering weekly 1 hour office hours open to the public (Mar 2022-July 2024).
UW CSE Ph.D. Pre-Application Mentorship Service (2022)
Organizer of Todai Girls Hackathon 2017
Panel at Mind The Gap (Google, Aug 2016 / Sept 2017, Tokyo)

MENTORING

Velocity Yu (Apr 2021-June 2022), UW Undergrad. Published [15][18][4]. Now at USC Ph.D.
Alex Mallen (Apr 2022-July 2023), UW Undergrad. Published [11]. Now Research Engineer at EleutherAI.
Daniel Kim (Sept 2022-July 2023), UW CSE Ph.D. student. Published [7].
Abhika Mishra (January 2023-Present). UW Undergrad. Published [3].
Go Kamoda (March 2023-Present), Tohoku University Undergrad. Published [31]
Tong Chen (March 2024-Present). UW CSE Ph.D. Published [2].

SELECTED INVITED TALKS

Microsoft Research India: *Reliable, Adaptable, and Attributable Language Models with Retrieval*, May 2024.

NIH: *Reliable, Adaptable, and Attributable Language Models with Retrieval*, May 2024.

Meta Smart Glass AI Reading Group: *Reliable, Adaptable, and Attributable Language Models with Retrieval*, April 2024.

University College London, Web Intelligence Group: *Reliable, Adaptable, and Attributable Language Models with Retrieval*, March 2024.

Microsoft Research, Health Futures team: *Self-reflective Language Models*, December 2023.

University of Edinburgh, Institute for Language, Cognition and Computation: *Self-reflective Language Models*, November 2023.

University of Massachusetts Amherst Machine Learning and Friends Lunch: *Self-reflective Language Models*, October 2023.

Preferred Networks *Retrieval-augmented LMs and Applications*, August 2023

MLDS Unit seminar @Okinawa Institute of Science and Technology: *Retrieval-augmented LMs and Applications*, August 2023

The University of Queensland Data Science Seminar Series: *Investigating and Building Efficient and Reliable LMs with Retrieval*, August 2023

The First International Workshop on Retrieval-enhanced Machine Learning @ SIGIR: *Investigating and Building Efficient and Reliable LMs with Retrieva*, July 2023

MILA NLP Reading Group: *Adaptive and trustworthy NLP with Retrieval*, April 2023

The AI TALK: *Adaptive and trustworthy NLP with retrieval*, March 2023

AI Quiz King: *Towards Better Multilingual Information Access*, December, 2022

Amazon (Alexa): *Towards Better Multilingual Information Access* , August, 2022

SEA: Search Engines Amsterdam: *One Question Answering Model for Many Languages with Cross-lingual Passage Retrieval*, October 2021

Apple (Web Answers): *One Question Answering Model for Many Languages with Cross-lingual Passage*

Retrieval, July 2021

Google Language & Apple (Web Answers): XOR QA: Cross-lingual Open-Retrieval Question Answering January, 2021

Google Research: *Learning to Retrieve Reasoning Paths over Wikipedia Graphs*, June, 2020

Selected Press & Media

How RAG-Powered AI Applications Have A Positive Impact On Businesses (Forbes) 2024

How Self-RAG Could Revolutionize Industrial LLMs (Towards Data Science) 2023

Lost in translation no more: IBM Fellowship winner Akari Asai asks — and answers — big questions in NLP to expand information access to all (Allen School News) 2022

Going Beyond SQuAD: Question Answering in Different Languages (Towards Data Science) 2020

Salesforce's AI navigates Wikipedia to find answers to complex questions (Venture Beat) 2020

Learning to retrieve reasoning paths from the Wikipedia graph (Salesforce Research Blog) 2020

Top Trends of Graph Machine Learning in 2020 (Towards Data Science) 2020

100,000 happy moments - What makes people happy? A huge database is making it possible to discern the answer at last. (MIT Technology Review) 2019