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

PUBLICATIONS

[1] 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 – top1%), 2024
Best Paper Honorable Mention at *Instruction Workshop at NeurIPS*, 2023.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] When Not to Trust Language Models: Investigating Effectiveness and Limitations of Paramet-
ric 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.

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

[8] 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.

[9] 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.

[10] Beyond Counting Datasets: Investigating Multilingual Dataset Construction and Necessary Resources.
Xinyan Yu*, **Akari Asai***, Trina Chatterjee, Junjie Hu, Eunsol Choi.
Findings of EMNLP, 2022.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

PUBLICATIONS
BEFORE PH.D.

[22] 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.

[23] 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.

[24] 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.

[25] A Data-Driven Approach to Understanding Happiness
Akari Asai, Vivian Li, Daniela Stepanov, and Wang-Chiew Tan.
ACL Widening NLP Workshop, 2017.

PREPRINTS

[26] Fine-grained Hallucination Detection and Editing for Language Models.
Abhika Mishra, **Akari Asai**, Vidhisha Balachandran, Yizhong Wang, Graham Neubig, Yulia Tsvetkov, Hannaneh Hajishirzi
Arxiv Preprint, 2024.

[27] 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
Arxiv Preprint, 2023.

[28] Inferential Text Generation with Multiple Knowledge Sources and Meta-Learning.
Daya Guo*, **Akari Asai***, Duyu Tang, Nan Duan, Ming Gong, Linjun Shou, Daxin Jiang, Jian Yin, and Ming Zhou.
Arxiv Preprint, 2020.

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

RESEARCH
EXPERIENCE

University of Washington   Sep 2019-Present, Seattle, WA
Ph.D. Student (Supervisor: Hannaneh Hajishirzi)
Worked on question answering [21][15] [20], information-seeking multilingual question answering and retrieval [16][13], multi-modal multi-hop question answering [17] and model efficiency [15][9]. Recently focusing on retrieval-based LMs [6][7] and multilingual NLP [27][10].

Meta AI   Jul 2022-Jan 2023 & Sep 2023-Present, Seattle, WA
Research Intern (2022) & Visiting Researcher (Present) (Supervisor: Scott Yih)
Worked on multi-task prompted training for a general-purpose retrieval system that can be adapted to a new task without any additional training [7][4].

Allen Institute for AI   Jul 2021-Dec 2021, Seattle, WA
Research Intern (Supervisors: Hannaneh Hajishirzi, Matt Gardner)
Worked on improving open-domain question answering systems' robustness and performance by learning with evidentiality [12].

Salesforce Research   Apr 2019-Aug 2019, Palo Alto, CA
Research Intern (Supervisors: Kazuma Hashimoto, Caiming Xiong)
Worked on a graph-based recurrent reader-retriever for open-domain question answering [21].

Microsoft Research Asia   Sep 2018-Jan 2019, Beijing, China
Research Intern (Supervisors: Duyu Tang, Nan Duan)
Worked on incorporating knowledge from multiple knowledge sources and meta-learning for few-shot commonsense reasoning [28].

Aizawa Yamasaki Lab, The University of Tokyo   Apr 2018-Mar 2019, Tokyo, Japan
Thesis Student (Supervisor: Kiyoharu Aizawa)
Worked on a separable formulation to estimate uncertainty in deep learning models [22][23].

Studio Ousia   Mar 2018-Mar 2019, Tokyo, Japan
Research Intern (Supervisor: Ikuya Yamada)
Worked on general contextualized entity representations for entity-centric NLP tasks [18]. Conducted a comprehensive study and experiments on the performance of Wikipedia2Vec, a tool to learn static word and entity representations from Wikipedia [19].

Tsuruoka Lab, The University of Tokyo   Dec 2017-Mar 2019, Tokyo, Japan
Research Assistant (Supervisor: Yoshimasa Tsuruoka)
Worked on zero-shot multilingual machine reading comprehension (Japanese and French) [29].

Megagon Labs   Mar-Apr, 2017, Tokyo, Japan
Research Intern (Supervisor: Alon Y. Halevy)
Crowd-sourced HappyDB, a corpus of 100,000 happy moments [24]. HappyDB provided the foundation to the **AAAI 2019 CL-AFF Shared Task**.

INDUSTRY   **Google**, Google Chrome Blink team   Aug-Oct 2017, Tokyo, Japan
EXPERIENCE   *Engineering Intern*
Extended Selection Modifier, the editing module in the Chrome Browser for selection modification corresponding to users' inputs, to support ShadowDOM v1 [slide] [patches].

Google, iGSA (Google Search iOS) team   Jun-Oct 2016, Tokyo, Japan
Engineering Intern
Designed and developed a feature for the Google Search iOS app that provides users with instant access to transit results on the app's home page using Google Search Featured Snippets.

AWARDS &   **Best Paper Honorable Mention at NeurIPS 2023 Instruction Workshop**   2023
GRANT   **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 Young Researcher Award of the Japanese Association for NLP   2019
            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
            Finalist, Mercari award and Innovator Prize, JP Hacks 2017   2017
            Finalist, Hitachi award, NEC Award and Innovator Prize, JP Hacks 2016   2016
            3rd Place, Social Entrepreneur Contest, UC Berkeley Challenge Lab (IEOR 185)   2016
            1st Place, GEIL, All-Japan policy making contest   2015

TEACHING AND   **Invited Lecturer, CSE 447 Natural Language Processing**, University of Washington   Mar 2024
LEADERSHIP   **Head TA, CSE599J Data-centric ML**, University of Washington   Winter 2024
EXPERIENCE   **Head TA & Guest Lecturer, CSE473 Artificial Intelligence**, University of Washington Fall 2023
               Instructor, Tutorial on Retrieval-based LMs and Applications, ACL   July 2023

Invited Lecturer, CSE 447 Natural Language Processing, University of Washington   Mar 2023
Invited Lecturer, CSE 373 Data Structures and Algorithm, University of Washington Dec 2022
Co-organizer & TA, Todai Girls Hackathon 2017, UTokyo   June - Oct 2017
Invited Speaker, Mind The Gap, Google   Aug 2016 / Sept 2017, Tokyo

SERVICES   **Workshop Lead Organizer:** 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 (Question Answering)
Reviewer/Program Committee: EMNLP 2019-2023, ACL 2020-2023, AKBC 2020, AAAI 2020, IJCAI 2021, NeurIPS 2021-2022, ICLR 2021-2023, ARR 2021-Present.
Department: Ph.D. Admission Committee Member (2021, 2022), Ph.D. Admission NLP Area Chair (2021, 2022), Diversity and Inclusion Committee Representative (2022-Present), UW NLP Seminar Organizer (2023-2024).

MENTORING   Velocity Yu, UW CSE BS student → USC Ph.D. [10][13]   Apr 2021-June 2022
Alex Mallen, a UW CSE BS student → EleutherAI [6]   Apr 2022-July 2023
Abhika Mishra, a UW CSE BS student [26]   January 2023-Present
Daniel Kim, a UW CSE Ph.D. student [3]   Sept 2022-July 2023
Go Kamoda, a Tohoku University CS Master student   March 2023-Present
Weekly Virtual Office hour, offering weekly 1 hour office hours open to public.   Mar 2022-Present
UW CSE Ph.D. Pre-Application Mentorship Service (PAMS)   2022

SELECTED   **University of Edinburgh, Microsoft Medical, UMass Amherst:** Self-reflective Language Models
INVITED TALKS   Virtual, Oct/Nov 2023
Preferred Networks, University of Tokyo, MLDS Unit seminar @Okinawa Institute of Science and Technology: Retrieval-augmented LMs and Applications   Japan, Aug 2023
REML@SIGIR, The University of Queensland: Investigating and Building Efficient and Reliable LMs with Retrieval   Virtual, July 2023
MILA NLP Reading Group: Adaptive and trustworthy NLP with Retrieval Montreal, April 2023
The AI TALK: Adaptive and trustworthy NLP with Retrieval   Virtual, March 2023
AI Quiz King: Towards Better Multilingual Information Access   Virtual, December, 2022
Amazon (Alexa): Towards Better Multilingual Information Access   Virtual, August, 2022
SEA: Search Engines Amsterdam: One Question Answering Model for Many Languages with Cross-lingual Passage Retrieval   Virtual, October 2021
Apple (Web Answers): One Question Answering Model for Many Languages with Cross-lingual Passage Retrieval   Virtual, July 2021
Google Language & Apple (Web Answers): XOR QA: Cross-lingual Open-Retrieval Question Answering   Virtual, January, 2021
Google Research: Learning to Retrieve Reasoning Paths over Wikipedia Graphs   June, 2020

PRESS & MEDIA **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

Competencies **Languages** Japanese (*native*), English (*fluent*, TOEFL 104), Mandarin (*basic*)
Programming Python, C/C++, Java, Matlab, Octave, R, Objective-C, Swift, JavaScript, Lisp, Perl
Libraries and Services Tensorflow, PyTorch, Chainer, Theano, Stanford CoreNLP, NLTK, Amazon Mechanical Turk, Microsoft Azure, Amazon Web Service, Google Cloud Platform