

# Akshita Bhagia

Email : [akshita23bhagia@gmail.com](mailto:akshita23bhagia@gmail.com)

<https://akshitab.github.io>

Natural Language Processing, Machine Learning, Deep Learning

## EXPERIENCE

---

- **Allen Institute for Artificial Intelligence (AI2)** Seattle, WA  
*Senior Research Engineer* *Jul 2021 - Present*  
*Research Engineer 2* *Jul 2020 - Jun 2021*  
Academic R&D in natural language processing (NLP). See list below for publications.
  - **Open Language Model (OLMo)**: Core contributions in dataset creation, model training, evaluation, and inference.
  - **Open-source libraries**: [OLMo](#) , [dolma](#) [ai2-olmo-eval](#) , [wimbd](#) , [ai2-catwalk](#) , [allennlp](#) , [ai2-tango](#) , [ai2-tailor](#)
  - **Datasets**: [PALOMA](#) , [DOLMa](#) , [WMT22 African \(release\)](#) , [No Language Left Behind NLLB \(release\)](#)
- **Cerebellum Capital** San Francisco, CA  
*Machine Learning Intern* *May 2019 - Aug 2019*
  - Adapted state-of-the-art deep learning models for financial time-series forecasting using Keras and Tensorflow.
- **InFoCusp** Ahmedabad, India  
*Lead Platform Development Engineer* *Feb 2018 - Jun 2018*  
*Research Programmer* *Jul 2015 - Jan 2018*
  - Engineered the core infrastructure of a data science platform used for R&D as well as production of financial models.

## EDUCATION

---

- **University of Washington** Seattle, WA  
*Non-degree course enrollment (NLP for Cultural Analytics); GPA: 4.00/4.00* *Jan 2023 - March 2023*
- **University of Massachusetts, Amherst** Amherst, MA  
*Master of Science in Computer Science; GPA: 4.00/4.00* *Sep 2018 - May 2020*
- **Dhirubhai Ambani Institute of Information and Communication Technology** Gandhinagar, India  
*Bachelor of Technology in Information and Communication Technology; GPA: 8.95/10* *Jul 2011 - May 2015*

## RELEVANT COURSES

---

NLP for Cultural Analytics, Machine Learning, Neural Networks, Deep Learning for NLP, Systems for Data Science, Advanced Algorithms

## PUBLICATIONS

---

\* denotes equal contributors, † denotes core contributors

### Peer-reviewed Conference Papers

- Niklas Muennighoff, Luca Soldaini, Dirk Groeneveld, Kyle Lo, Jacob Daniel Morrison, Sewon Min, Weijia Shi, and others. [OLMoE: Open Mixture-of-Experts Language Models](#). *Accepted to ICLR* (2025).
- Yanai Elazar, **Akshita Bhagia**, Ian Magnusson, Abhilasha Ravichander, Dustin Schwenk, Alane Suhr, Pete Walsh, and others. [What's In My Big Data?](#) *ICLR* (2024). **Spotlight** (5% papers).
- Dirk Groeneveld, Iz Beltagy, Pete Walsh, **Akshita Bhagia**, Rodney Kinney, Oyvind Tafjord, A. Jha, and others. [OLMo: Accelerating the Science of Language Models](#). *ACL* (2024). **Best Theme Paper Award**.
- Ian Magnusson, **Akshita Bhagia**, Valentin Hofmann, Luca Soldaini, A. Jha, Oyvind Tafjord, Dustin Schwenk, and others. [Paloma: A Benchmark for Evaluating Language Model Fit](#). *NeurIPS* (2024).

- Luca Soldaini<sup>†</sup>, Rodney Kinney<sup>†</sup>, **Akshita Bhagia<sup>†</sup>**, Dustin Schwenk<sup>†</sup>, David Atkinson, Russell Authur, Ben Bogin, and others. [Dolma: an Open Corpus of Three Trillion Tokens for Language Model Pretraining Research](#). *ACL* (2024). **Best Resource Paper Award**.
- Hamish Ivison, **Akshita Bhagia**, Yizhong Wang, Hannaneh Hajishirzi, and Matthew Peters. [HINT: Hypernetwork Instruction Tuning for Efficient Zero- and Few-Shot Generalisation](#). *ACL* (2023).
- Shruti Palaskar, **Akshita Bhagia**, Yonatan Bisk, Florian Metzger, Alan W. Black, and Ana Marasović. [On Advances in Text Generation from Images Beyond Captioning: A Case Study in Self-Rationalization](#). *Findings of EMNLP* (2022).
- Zhaofeng Wu, Robert L. Logan IV, Pete Walsh, **Akshita Bhagia**, Dirk Groeneveld, Sameer Singh, and Iz Beltagy. [Continued Pretraining for Better Zero- and Few-Shot Promptability](#). *EMNLP* (2022).

## Workshops and Preprints

- **Akshita Bhagia<sup>\*</sup>**, Jiacheng Liu<sup>\*</sup>, Alexander Wettig, David Heineman, Oyvind Tafjord, Ananya Harsh Jha, Luca Soldaini, and others. [Establishing Task Scaling Laws via Compute-Efficient Model Ladders](#). *arXiv preprint* (2024).
- Team OLMo, Pete Walsh<sup>†</sup>, Luca Soldaini<sup>†</sup>, Dirk Groeneveld<sup>†</sup>, Kyle Lo<sup>†</sup>, Shane Arora<sup>†</sup>, Akshita Bhagia<sup>†</sup>, and others. [2 OLMo 2 Furious](#). *arXiv preprint* (2024).
- Dirk Groeneveld, Anas Awadalla, Iz Beltagy, **Akshita Bhagia**, Ian Magnusson, Hao Peng, Oyvind Tafjord, Pete Walsh, Kyle Richardson, and Jesse Dodge. [Catwalk: A Unified Language Model Evaluation Framework for Many Datasets](#). *arXiv preprint* (2023).
- Kyle Richardson, Ian Magnusson, Oyvind Tafjord, **Akshita Bhagia**, Iz Beltagy, Arman Cohan, Pradeep Dasigi, and others. [Robust Tooling and New Resources for Large Language Model Evaluation via Catwalk \(extended abstract\)](#). *GEM Workshop, EMNLP* (2023).

## ACADEMIC SERVICE AND OUTREACH

---

### ◦ Reviewer

- DHTech workshop at DH2024.
- WiNLP workshop at EMNLP (2024).
- ACL Rolling Review (Jun 2024, Apr 2024, Feb 2024); conferences: ACL, EMNLP.
- LREC-COLING (2024); emergency reviewer.

- **Co-organizer** for the [2022 WMT Shared Task on Large Scale Machine Translation Evaluation for African Languages](#).

- **Co-organizer** for [AllenNLP Hacks 2021](#) hackathon for minoritized communities.

- **Master's chair** for the graduate chapter of [UMass Amherst CSWomen](#) (Feb 2019 - May 2020); handled monthly travel grants. The organization was awarded a Women for UMass grant towards travel expenses for graduate students.

- **Student Representative** of the Gender Cell at DA-IICT (2014-2015).

## TALKS AND PANELS

---

- [Tutorial on data preparation, model training, and adaptation](#) at NeurIPS (December 2024).
- Panel on RSEs in domain-specific ecosystems at the US-RSE conference (October 2024).
- Open Source AI panel at the [AI/ALL Summit: Future of Equitable and Inclusive AI](#) (August 2024).
- Presenter on the [Databricks webinar](#) for the OLMo project (March 2024).

- Interviewed by Sam Charrington on the [TWiML podcast](#) about the OLMo project (March 2024).
- Workshop on the AllenNLP library at AllenNLP Hacks 2021.
- Workshop on Running experiments on Google Colaboratory at AllenNLP Hacks 2021.
- Panelist on the GHC Information Panel at UMass CICS in Spring 2020.

## AWARDS

---

- Geekwire Innovation of the Year 2024 Award for OLMo.
- Recipient of the AnitaB.org Grace Hopper Conference Scholarship 2019.

## MENTORING

---

- **Pranjali Basmatkar**, Research Engineering Intern @ AI2 (Summer 2023).
- **Sunitha Selvan**, Research Engineering Intern @ AI2 (Summer 2022).
- **Arjun Subramonian**, Research Engineering Intern @ AI2 - **AI2 Intern of the Year** (Summer 2021).
- **Subhodeep Maji**, Research Programmer Intern @ InFoCusp (Spring 2017).