David Wan

meetdavidwan.github.io | davidwan@cs.unc.edu | Google Scholar

Education

University of North Carolina at Chapel Hill — Chapel Hill, NC
Ph.D. in Computer Science. Advisor: Mohit Bansal — Aug 2021 - Present
Google PhD Fellowship in NLP for 2024 and 2025

Columbia University — New York, NY
M.S. in Computer Science. Advisor: Kathleen McKeown — Sep 2020 - Jul 2021
Thesis: *Methods for Cross-Language Search and Summarization for Low-Resource Languages*

Columbia University — New York, NY
B.A. in Computer Science, Concentration in Linguistics — Sep 2016 - May 2020

Research & Industry Experience

Google Research — New York, NY
Research Intern. Hosts: Sebastien Baur and Gaurav Singh Tomar — May 2025 - Present

Salesforce AI Research — Palo Alto, CA
Research Intern. Hosts: Shafiq Joty and Jesse Vig — May 2024 - Oct 2024

FAIR Labs at Meta — Seattle, WA
Research Intern. Hosts: Ramakanth Pasunuru and Asli Celikyilmaz — May 2023 - Dec 2023

Alexa AI at Amazon — Seattle, WA
Research Intern. Hosts: Mengwen Liu and Markus Dreyer — May 2022 - Oct 2022

Columbia University — New York, NY
Research Assistant. Advisor: Michael Collins — Nov 2018 - May 2019

Research Interest

Natural Language Generation (Summarization, Machine Translation), Factuality, Multimodal

Publications

17. **GenerationPrograms: Fine-grained Attribution with Executable Programs**
David Wan, Eran Hirsch, Elias Stengel-Eskin, Ido Dagan, and Mohit Bansal
In COLM 2025

16. **CLaMR: Multimodal Late-Interaction Retrieval**
David Wan, Han Wang, Elias Stengel-Eskin, Jaemin Cho, and Mohit Bansal
ArXiv Preprint

15. **QAPyramid: Fine-grained Evaluation of Content Selection for Text Summarization**
Shiyue Zhang*, **David Wan***, Arie Cattan, Ayal Klein, Ido Dagan, Mohit Bansal
In COLM 2025

14. **MAMM-Refine: A Recipe for Improving Faithfulness in Generation with Multi-Agent Collaboration**
David Wan, Justin Chih-Yao Chen, Elias Stengel-Eskin, Mohit Bansal
In NAACL 2025

13. **On Positional Bias of Faithfulness for Long-form Summarization**
David Wan, Jesse Vig, Mohit Bansal, Shafiq Joty
In NAACL 2025

12. **Localizing Factual Inconsistencies in Attributable Text Generation**
Arie Cattan, Paul Roit, Shiyue Zhang, **David Wan**, Roee Aharoni, Idan Szpektor, Mohit Bansal, Ido Dagan
In TACL 2025

11. **ACUEVAL: Fine-grained Hallucination Evaluation and Correction for Abstractive Summarization**
David Wan, Koustuv Sinha, Srini Iyer, Asli Celikyilmaz, Mohit Bansal, Ramakanth Pasunuru
In ACL 2024

10. **Contrastive Region Guidance: Improving Grounding in Vision-Language Models without Training**
David Wan, Jaemin Cho, Elias Stengel-Eskin, Mohit Bansal
In ECCV 2024

9. **HISTALIGN: Improving Context Dependency in Language Generation by Aligning with History**
David Wan, Shiyue Zhang, Mohit Bansal
In EMNLP 2023

8. **Extractive is not Faithful: An Investigation of Broad Unfaithfulness Problems in Extractive Summarization**
Shiyue Zhang*, **David Wan***, Mohit Bansal
In ACL 2023

7. **Faithfulness-Aware Decoding Strategies for Abstractive Summarization**
David Wan, Mengwen Liu, Kathleen McKeown, Markus Dreyer, Mohit Bansal
In EACL 2023

6. **Evaluating and Improving Factuality in Multimodal Abstractive Summarization**
David Wan, Mohit Bansal
In EMNLP 2022

5. **Constrained Regeneration for Cross-Lingual Query-Focused Extractive Summarization**
Elsbeth Turcan, **David Wan**, Faisal Ladhak, Petra Galuscakova, Sukanta Sen, Svetlana Tchistiakova, Weijia Xu, Marine Carpuat, Kenneth Heafield, Douglas Oard, Kathleen McKeown
In COLING 2022

4. **FACTPEGASUS: Factuality-Aware Pre-training and Fine-tuning for Abstractive Summarization**
David Wan, Mohit Bansal
In NAACL 2022

3. **Segmenting Subtitles for Correcting ASR Segmentation Errors**
David Wan, Chris Kedzie, Faisal Ladhak, Elsbeth Turcan, Petra Galuscakova, Elena Zotkina, Zhengping Jiang, Peter Bell, Kathleen McKeown.
In EACL 2021

2. **Incorporating Terminology Constraints in Automatic Post-Editing**
David Wan, Chris Kedzie, Faisal Ladhak, Marine Carpuat, Kathleen McKeown
In WMT 2020

1. **Subtitles to Segmentation: Improving Low-Resource Speech-to-Text Translation Pipelines.**
David Wan, Zhengping Jiang, Chris Kedzie, Elsbeth Turcan, Peter Bell, Kathleen McKeown
In CLSSTS 2020

ACHIEVEMENTS AND AWARDS

Google PhD Fellowship in Natural Language Processing, *Google* 2024
One of twelve students globally to receive full funding for two years

Theodore R. Bashkow Award, *Dept. of Computer Science, Columbia University* Apr 2020
One of three undergrads awarded for excelling in independent projects

PROFESSIONAL SERVICES

Reviewer:

- EMNLP NewSumm Workshop 2025
- NeurIPS 2025

- ACL Rolling Review, May 2025
- ACL Rolling Review, December 2024
- ACL Rolling Review, October 2024
- ACL Rolling Review, June 2024
- ACL Rolling Review, April 2024
- ACL Rolling Review, February 2024
- ACL Rolling Review, December 2023
- EMNLP 2023
- ACL 2023
- EMNLP 2022

Invited Talks

Research Trend AI Summer 2025
CLaMR*: Contextualized Late-Interaction for Multimodal Content Retrieval*

Teaching Experience

Columbia University New York, NY
Teaching Assistant for Natural Language Processing Jan 2019 - May 2019, Jan 2020 - May 2020

Columbia University New York, NY
Tutor for Natural Language Processing Sep 2019 - Dec 2019