

# Amanpreet Singh

---

<https://apsdehal.in>  
me@apsdehal.in  
March 2023

## Research Interests

Multimodal, Computer Vision, Natural Language Processing, Multiagent Systems

## Professional Experience

**HuggingFace, *Research Lead - Palo Alto, CA*** 2022 - 2023

- Working on cutting-edge multimodal research with emphasis on open science.

**Facebook, *AI Researcher - Menlo Park, CA*** 2018 - 2022

- Designed and developed a framework (MMF) for multimodal research and production. Used heavily internally (>50 teams/100M qps) and externally.
- Led and executed different research directions with emphasis on large scale vision and language applications.

**Amazon, The Climate Corporation, Google, Wingify** Various  
***Software Engineering Intern***

## Publications & Preprints

- [1] Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, François Yvon, **Amanpreet Singh** Gallé, Matthias, et al. Bloom: A 176b-parameter open-access multilingual language model. *arXiv preprint arXiv:2211.05100*, 2022.
- [2] Mingyang Zhou, Licheng Yu, **Amanpreet Singh**, Mengjiao Wang, Zhou Yu, and Ning Zhang. Unsupervised vision-and-language pre-training via retrieval-based multi-granular alignment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022.
- [3] **Amanpreet Singh**, Ronghang Hu, Vedanuj Goswami, Guillaume Couairon, Wojciech Galuba, Marcus Rohrbach, and Douwe Kiela. Flava: A foundational language and vision alignment model. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022.
- [4] Sasha Sheng, **Amanpreet Singh**, Vedanuj Goswami, Jose Magana, Tristan Thrush, Wojciech Galuba, Devi Parikh, and Douwe Kiela. Human-adversarial visual question answering. *Advances in Neural Information Processing Systems*, 34, 2021.
- [5] Douwe Kiela, Max Bartolo, Yixin Nie, Divyansh Kaushik, Atticus Geiger, Zhengxuan Wu, Bertie Vidgen, Grusha Prasad, **Amanpreet Singh**, Pratik Ringshia, et al. Dynabench: Rethinking benchmarking in nlp. In *NAACL-HLT*, 2021.
- [6] Ronghang Hu and **Amanpreet Singh**. Unit: Multimodal multitask learning with a unified transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1439–1449, 2021.
- [7] **Amanpreet Singh**, Guan Pang, Mandy Toh, Jing Huang, Wojciech Galuba, and Tal Hassner. Textocr: Towards large-scale end-to-end reasoning for arbitrary-shaped scene text. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8802–8812, 2021.

- [8] Medhini Narasimhan, Erik Wijmans, Xinlei Chen, Trevor Darrell, Dhruv Batra, Devi Parikh, and **Amanpreet Singh**. Seeing the un-scene: Learning amodal semantic maps for room navigation. In *Proceedings of IEEE/CVF European Conference on Computer Vision (ECCV)*, 2020.
- [9] Oleksii Sidorov, Ronghang Hu, Marcus Rohrbach, and **Amanpreet Singh**. Textcaps: a dataset for image captioning with reading comprehension. In *Proceedings of IEEE/CVF European Conference on Computer Vision (ECCV)*, 2020.
- [10] **Amanpreet Singh**, Vedanuj Goswami, and Devi Parikh. Are we pretraining it right? digging deeper into visio-linguistic pretraining. *arXiv preprint arXiv:2004.08744*, 2020.
- [11] Douwe Kiela, Hamed Firooz, Aravind Mohan, Vedanuj Goswami, **Amanpreet Singh**, Pratik Ringshia, and Davide Testuggine. The hateful memes challenge: Detecting hate speech in multimodal memes. *Advances in Neural Information Processing Systems*, 33:2611–2624, 2020.
- [12] Ronghang Hu, **Amanpreet Singh**, Trevor Darrell, and Marcus Rohrbach. Iterative answer prediction with pointer-augmented multimodal transformers for textvqa. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9992–10002, 2020.
- [13] Alex Wang\*, Yada Pruksachatkun\*, Nikita Nangia\*, **Amanpreet Singh\***, Julian Michael, Felix Hill, Omer Levy, and Samuel Bowman. Superglue: A stickier benchmark for general-purpose language understanding systems. In *Advances in Neural Information Processing Systems*, pages 3266–3280, 2019.
- [14] **Amanpreet Singh**, Vivek Natarajan, Meet Shah, Yu Jiang, Xinlei Chen, Dhruv Batra, Devi Parikh, and Marcus Rohrbach. Towards vqa models that can read. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8317–8326, 2019.
- [15] **Amanpreet Singh**, Tushar Jain, and Sainbayar Sukhbaatar. Learning when to communicate at scale in multiagent cooperative and competitive tasks. In *Proceedings of International Conference on Learning Representations (ICLR)*, 2019.
- [16] Nan Rosemary Ke, **Amanpreet Singh**, Ahmed Touati, Anirudh Goyal, Yoshua Bengio, Devi Parikh, and Dhruv Batra. Learning dynamics model in reinforcement learning by incorporating the long term future. In *Proceedings of International Conference on Learning Representations (ICLR)*, 2019.
- [17] Alex Wang, **Amanpreet Singh**, Julian Michael, Felix Hill, Omer Levy, and Samuel R Bowman. Glue: A multi-task benchmark and analysis platform for natural language understanding. In *Proceedings of International Conference on Learning Representations (ICLR)*, 2019.
- [18] Alex Warstadt, **Amanpreet Singh**, and Samuel R Bowman. Neural network acceptability judgments. *Transactions of the Association for Computational Linguistics*, 7:625–641, 2019.
- [19] **Amanpreet Singh**, Vivek Natarajan, Yu Jiang, Xinlei Chen, Meet Shah, Marcus Rohrbach, Dhruv Batra, and Devi Parikh. Mmf - a multimodal framework for vision & language research. In *SysML Workshop, NeurIPS*, volume 2018, 2018.
- [20] **Amanpreet Singh** and Sharan Agrawal. Canvasgan: A simple baseline for text to image generation by incrementally patching a canvas. In *Science and Information Conference*, pages 86–98. Springer, 2019.

<b>Honor &amp; Awards</b>	<b>Oral Presentation, TextCaps (2.1% papers)</b> (ECCV, 2020)	2019
	<b>Oral Presentation, M4C (2.3% papers)</b> (CVPR, 2019)	2019
	<b>Spotlight Presentation, SuperGLUE (4.2% papers)</b> (NeurIPS, 2019)	2019
	<b>Fellowship for best research project, NYU</b> (Masters Scholarship)	2018
	<b>Fellowship for best innovation project, NYU</b> (Masters Scholarship)	2017
	<b>Fellowship for top student in state, IITR</b> (Undergraduate Scholarship)	2012
	<b>3rd Place, TopCoder Hackathon</b> (San Francisco)	2015
<b>Invited Talks</b>	<i>Towards unified foundation model for vision-and-language alignment</i> Snorkel	Jan 2022
	<i>Towards unified foundation model for vision-and-language alignment</i> Hugging Face	July 2022
	<i>Towards unified foundation model for vision-and-language alignment</i> Stanford University	July 2022
	<i>Towards unified foundation model for vision-and-language alignment</i> Amazon Research	May 2022
	<i>Towards unified foundation model for vision-and-language alignment</i> Microsoft Research	Mar 2022
	<i>Towards unified foundation model for vision-and-language alignment</i> MLRW, IITG	Feb 2022
	<i>Towards models that can read and reason about scene text</i> DocVQA Workshop, KDD 2021	Sep 2021
	<i>TextCaps dataset and challenge 2021</i> Visual QA and Dialog Workshop, CVPR 2021	Jun 2021
	<i>Building knowledge based VQA systems with MMF</i> KDD 2020	Oct 2020
	<i>Supercharging Vision and Language Research with MMF</i> Visual QA and Dialog Workshop, CVPR 2020	Jun 2020
	<i>TextVQA dataset and challenge 2020</i> Visual QA and Dialog Workshop, CVPR 2020	Jun 2020
	<i>TextVQA dataset and challenge 2019</i> Visual QA and Dialog Workshop, CVPR 2019	Jun 2020
<b>Professional Activities</b>	<i>Reviewer:</i> CVPR 2023, ICLR 2023, ICML 2023, NeurIPS 2022, CVPR 2022, ICLR 2022, ICCV 2021, NeurIPS 2021, ECCV 2020, CVPR 2021, ICLR 2021, CVPR 2020, EMNLP 2020, ACL 2020, ICLR 2020, NeurIPS 2020, AISTATS 2020, MVAP Journal,	

TPAMI, and others

*Workshop Organization:* Visual Question Answering and Dialog Workshop, CVPR 2020; Visual Question Answering and Dialog Workshop, CVPR 2019.

*Panelist:* Visual Question Answering and Dialog Workshop, CVPR 2020.

*Tutorials:* Building knowledge based VQA systems with MMF, KDD 2020.

*Organization:* jQuery Foundation, PyTorch