AISHWARYA AGRAWAL aishwarya.agrawal@mila.quebec • https://www.iro.umontreal.ca/~agrawal/

Appointments

• University of Montreal, QC, Canada Assistant Professor	Start: Decemb	ber 2020
Department of Computer Science and Operations Research (DIRO)		
• Mila, QC, Canada Core Member Canada CUEAD, AL Chain	Start: Decemb	per 2020
		2020
• Google DeepMind, Montreal, Canada 20% Research Scientist	Start: Decemb	ber 2020
• DeepMind, London, UK Research Scientist	Aug 2019 - D	Dec 2020
EDUCATION		
• Georgia Institute of Technology, Atlanta, USA Ph.D., Computer Science GPA: 4.0/4.0, Advisor: Dhruv Batra	Aug 2017 - A	ug 2019
• Virginia Tech, Blacksburg, USA Aug 20 Ph.D., Computer Engineering GPA: 4.0/4.0, Advisor: Dhruy Batra	14 - Aug 2017 (tran	sferred)
• Indian Institute of Technology (IIT), Gandhinagar, India Bachelor of Technology, Electrical Engineering; Minor, Computer Science and GPA: 9.42/10, B.Tech Project Advisor: Shanmuganathan Raman Institute Silver Medal	July 2010 - M Engineering	ay 2014
Honors and Awards		
• Young Alumni Excellence Award 2023 for Outstanding Academic Achievement (Awarded by IIT Gandhinagar to recognize the achievements of their alumni outstanding contributions to their respective fields.)	its who have made	2023
• Ruppers Up of 2010 A AAI / ACM SICAI Discortation Award		2020
(One of 2 runners-up from nominations from universities across the globe)		2020
• Georgia Tech 2020 College of Computing Dissertation Award		2020
(One of 4 awardees from nominations from across College of Computing at G	eorgia Tech)	
• Georgia Tech 2020 Sigma Xi Best Ph.D. Thesis Award		2020
(One of 10 awardees from nominations from across all schools at Georgia Tech	1)	
• Canada CIFAR AI Chair Award		2019
(One of 34 awardees from nominations from across Canada)	Vicion (dealined)	2010
• Google renowship in Machine Perception, speech rechnology and computer (One of 54 awardees from PhD applicants from all over the world)	vision (decimed)	2019
• Facebook Fellowship in Computer Vision (declined)		2019
(One of 21 awardees from more than 900 PhD applicants from all over the wo	orld)	
• Foley Scholar Finalist		2018
(One of 8 finalists of the 2018 Foley Scholars Awards – an annual fellowship a students involved in exceptional research at the GVU center at Georgia Tech)	warded to outstand	ling
• Rising Stars in EECS		2018
(An Academic Career Workshop for Women, hosted at MIT) (One of 60+ selected candidates with funded accommodation and travel)		
• NVIDIA Graduate Fellowship		2018
(One of 11 awardees from more than 120 applicants from a host of countries)		
• Invitation to Women in Research Lean In event		2018

(An annual event bringing together outstanding computer science and engineerin post-docs, organized by Facebook)	g Ph.D. students and
• Outstanding Reviewer award, NIPS (Awarded to ~3.6% of reviewers)	2017
• Travel award, Women in Machine Learning Workshop	2017
• Finalist, Microsoft Research PhD Fellowship	2017
• Finalist, Adobe Research PhD Fellowship	2017
• Outstanding Reviewer award, CVPR (Awarded to ~8.5% of reviewers)	2017
• Travel award, Women in Computer Vision Workshop, CVPR	2017
• Travel award, Women in Machine Learning Workshop	2016
• Best Discussion Participant Award, Advanced Computer Vision Course, Virginia	Tech 2016
• Best Poster award, Object Understanding for Interaction Workshop, ICCV (Awarded for VQA: Visual Question Answering)	2015
• Travel award, Women in Computer Vision Workshop, CVPR	2015
• Institute Silver Medal, IIT Gandhinagar (Awarded for second highest cumulative performance index in the batch (~ 45 str	2014 udents))
 Scholarship for academic excellence, IIT Gandhinagar (Awarded to branch topper – 1 out of ~45 students) 	2011, 2012
• Dean's List for academic excellence, IIT Gandhinagar	2011,2012,2013,2014
• Among top 0.76% of students in the IIT-JEE examination (Out of \sim 0.4 million students who appeared for the examination)	2010
TEACHING	

IFT 6765 (UdeM) - Links between Computer Vision and Language
IFT 6135 (UdeM) - Representation Learning
Winter 2022, 2023, 2024
Fall 2021, 2022, 2023, 2024

Research Group

PhD students	
• Oscar Manas	Fall 2021 - Present
• Le Zhang	Fall 2023 - Present
• Sarvjeet Singh Ghotra (co-supervised with Aaron Courville)	Fall 2023 - Summer 2024 (change of supervisor)
• Qian Yang	Fall 2023 - Present
• Kanishk Jain	Fall 2023 - Present
• Rabiul Awal	Fall 2024 - Present
 MSc students Saba Ahmadi Le Zhang Sarvjeet Singh Ghotra Shravan Nayak Ankur Sikarwar 	Winter 2022 - Dec 2023 (graduated; started internship with me) Fall 2022 - Summer 2023 (fast-tracked to PhD program) Fall 2022 - Summer 2023 (fast-tracked to PhD program) Fall 2023 - Present Fall 2024 - Present
Research internsOscar ManasRabiul AwalSaba Ahmadi	Summer 2021 (started PhD with me) Winter 2023 - Summer 2024 (started PhD with me) Winter 2024 - Present
Publications	

- O. Manas, P. Astolfi, M. Hall, C. Ross, J. Urbanek, A. Williams, A. Agrawal, A. Romero-Soriano, M. Drozdzal. Improving Text-to-Image Consistency via Automatic Prompt Optimization. Accepted to the *Transactions on Machine Learning Research (TMLR)*, 2024. Featured Certification.
- R. Awal, L. Zhang, A. Agrawal. Investigating Prompting Techniques for Zero- and Few-Shot Visual Question Answering. Accepted to the Workshop on Multimodal Algorithmic Reasoning, NeurIPS, 2024.
- R. Awal, S. Ahmadi, L. Zhang, A. Agrawal. VisMin: Visual Minimal-Change Understanding. Accepted to the *Conference on Neural Information Processing Systems (NeurIPS)*, 2024.
- S. Nayak, K. Jain, R. Awal, S. Reddy, S. van Steenkiste, L. Anne Hendricks, K. Stanczak, A. Agrawal. Benchmarking Vision Language Models for Cultural Understanding. Accepted to the *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2024. Oral Presentation.
- Q. Yang, W. Yan, A. Agrawal. Decompose and Compare Consistency: Measuring VLMs' Answer Reliability via Task-Decomposition Consistency Comparison. Accepted to the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2024.
- F. Bordes et al. An Introduction to Vision-Language Modeling. arXiv preprint, arXiv:2405.17247, 2024.
- L. Zhang, R. Awal, A. Agrawal. Contrasting intra-modal and ranking cross-modal hard negatives to enhance visio-linguistic compositional understanding. In the *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.
- S. Ahmadi, A. Agrawal. An Examination of the Robustness of Reference-Free Image Captioning Evaluation Metrics. In the *Findings of the Association for Computational Linguistics: EACL*, 2024.
- O. Manas, B. Krojer, **A. Agrawal**. Improving Automatic VQA Evaluation Using Large Language Models. In the 38th Annual AAAI Conference on Artificial Intelligence, 2024.
- L. Zhang, Y. Wu, F. Mo, J. Nie, A. Agrawal. MoqaGPT: Zero-Shot Multi-modal Open-domain Question Answering with Large Language Model. In the *Findings of the Association for Computational Linguistics (EMNLP)*, 2023.
- E. Bugliarello, L. Sartran, A. Agrawal, L. Anne Hendricks, A. Nematzadeh. Measuring Progress in Fine-grained Vision-and-Language Understanding. In the Association for Computational Linguistics (ACL), 2023.
- O. Manas, P. Rodrguez^{*}, S. Ahmadi^{*}, A. Nematzadeh, Y. Goyal, **A. Agrawal**. MAPL: Parameter-Efficient Adaptation of Unimodal Pre-Trained Models for Vision-Language Few-Shot Prompting. In the *European Chapter of the Association for Computational Linguistics (EACL)*, 2023. Oral Presentation.
- A. Agrawal*, I. Kajic*, E. Bugliarello*, E. Davoodi, A. Gergely, P. Blunsom, A. Nematzadeh*. Rethinking Evaluation Practices in Visual Question Answering: A Case Study on Out-of-Distribution Generalization. In the *Findings of the European Chapter of the Association for Computational Linguistics (EACL)*, 2023.
- A. Agrawal. Visual Question Answering and Beyond. PhD Dissertation, 2019.
- A. Agrawal, M. Malinowski, F. Hill, A. Eslami, O. Vinyals and T. Kulkarni. Generating Diverse Programs with Instruction Conditioned Reinforced Adversarial Learning. *arXiv preprint*, arXiv:1812.00898, 2018.
- S. Ramakrishnan, A. Agrawal and S. Lee. Overcoming Language Priors in Visual Question Answering with Adversarial Regularization. In *Neural Information Processing Systems (NIPS)*, 2018.
- A. Agrawal, D. Batra, D. Parikh and A. Kembhavi. Don't Just Assume; Look and Answer: Overcoming Priors for Visual Question Answering. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

- G. Christie^{*}, A. Laddha^{*}, **A. Agrawal**, S. Antol, Y. Goyal, K. Kochersberger and D. Batra. Resolving Language and Vision Ambiguities Together: Joint Segmentation & Prepositional Attachment Resolution in Captioned Scenes. In the journal of *Computer Vision and Image Understanding (CVIU)*, 2017.
- A. Agrawal, A. Kembhavi, D. Batra and D. Parikh. C-VQA: A Compositional Split of the Visual Question Answering (VQA) v1.0 Dataset. arXiv preprint, arXiv:1704.08243, 2017.
- A. Agrawal*, J. Lu*, S. Antol*, M. Mitchell, C. L. Zitnick, D. Parikh and D. Batra. VQA: Visual Question Answering. In Special Issue on Combined Image and Language Understanding, International Journal of Computer Vision (IJCV), 2017.
- A. Agrawal, D. Batra and D. Parikh. Analyzing the Behavior of Visual Question Answering Models. In Conference on Empirical Methods in Natural Language Processing (EMNLP), 2016.
- C. L. Zitnick, A. Agrawal, S. Antol, M. Mitchell, D. Batra and D. Parikh. Measuring Machine Intelligence Through Visual Question Answering. In AI Magazine, 2016.
- G. Christie^{*}, A. Laddha^{*}, A. Agrawal, S. Antol, Y. Goyal, K. Kochersberger and D. Batra. Resolving Language and Vision Ambiguities Together: Joint Segmentation & Prepositional Attachment Resolution in Captioned Scenes. In Conference on Empirical Methods in Natural Language Processing (EMNLP), 2016.
- T. Huang, F. Ferraro, N. Mostafazadeh, I. Misra, A. Agrawal, J. Devlin, R. Girshick, X. He, P. Kohli, D. Batra, C.L. Zitnick, D. Parikh, L. Vanderwende, M. Galley and M. Mitchell. Visual Storytelling. In North American Chapter of the Association for Computational Linguistics (NAACL), 2016.
- S. Antol^{*}, A. Agrawal^{*}, J. Lu, M. Mitchell, D. Batra, C. L. Zitnick and D. Parikh. VQA: Visual Question Answering. In International Conference on Computer Vision (ICCV), 2015.
- A. Agrawal and S. Raman. A Novel LBP Based Operator for Tone Mapping HDR Images. In International Conference on Signal Processing and Communications (SPCOM-2014) and IEEE Xplore, 2014.
- R. Das, A. Agrawal, M. Upton and E.J. Seibel. Optically clearing tissue as an initial step for 3D imaging of core biopsies to diagnose pancreatic cancer. In SPIE BiOS, pp. 89410N-89410N. International Society for Optics and Photonics, 2014.

WORKSHOP PAPER PRESENTATIONS

- Enhancing Multi-Agent Multi-Modal Collaboration with Fine-Grained Reward Modeling.
 Q. Yang, W. Yan, A. Agrawal
 To be presented at the workshop on Adaptive Foundation Models, NeurIPS 2024.
- Controlling Multimodal LLMs via Reward-guided Decoding.
 O. Manas, P. D'Oro, K. Sinha, A. Romero-Soriano, M. Drozdzal, A. Agrawal To be presented at the workshop on *Adaptive Foundation Models*, NeurIPS 2024.
- Visual Language Alignment Tuning.
 L. Zhang, Q. Yang, A. Agrawal To be presented at the workshop on Adaptive Foundation Models, NeurIPS 2024.
- CTRL-O: Language-Controllable Object-Centric Visual Representation Learning. A. Rajiv Didolkar, A. Zadaianchuk, R. Awal, M. Seitzer, E. Gavves, **A. Agrawal** To be presented at the workshop on *Adaptive Foundation Models*, NeurIPS 2024.
- Investigating Reliable Question Decomposition for Vision-Language Tasks. Q. Yang, W. Yan, A. Agrawal Presented at the 3rd workshop on *Computer Vision in the Wild*, CVPR 2024.
- CulturalVQA: Benchmarking Vision Language Models for Cultural Understanding. S. Nayak, K. Jain, K. Stanczak, R. Awal, A. Agrawal

Presented at the 3rd workshop on *Computer Vision in the Wild*, CVPR 2024, and at the Montreal AI Symposium, 2024 (**Oral Presentation**).

- An Examination of the Robustness of Reference-Free Image Captioning Evaluation Metrics.
 S. Ahmadi, A. Agrawal Presented at the workshop on *Open-Domain Reasoning Under Multi-Modal Settings (O-DRUM)*, CVPR 2023.
- Investigating Prompting Techniques for Zero- and Few-Shot Visual Question Answering.
 R. Awal, L. Zhang, A. Agrawal
 Presented at the workshop on Open-Domain Reasoning Under Multi-Modal Settings (O-DRUM), CVPR 2023, and accepted at the workshop on Prompting in Vision, CVPR 2023 (Oral Presentation).
- Contrasting Intra-Modal and Ranking Cross-Modal Hard Negatives to Enhance Visio-Linguistic Fine-Grained Understanding.
 L. Zhang, R. Awal, A. Agrawal Presented at the workshop on Open-Domain Reasoning Under Multi-Modal Settings (O-DRUM), CVPR 2023 (Spotlight Presentation).
- MAPL: Parameter-Efficient Adaptation of Unimodal Pre-Trained Models for Vision-Language Few-Shot Prompting.

O. Manas, P. Rodrguez^{*}, S. Ahmadi^{*}, A. Nematzadeh, Y. Goyal, **A. Agrawal** Presented at the workshop on *Open-Domain Reasoning Under Multi-Modal Settings (O-DRUM)*, CVPR 2023.

- Generating Diverse Programs with Instruction Conditioned Reinforced Adversarial Learning
 A. Agrawal, M. Malinowski, F. Hill, A. Eslami, O. Vinyals, T. Kulkarni
 Presented at Visually-Grounded Interaction and Language (Spotlight Presentation), and Learning by Instruction workshops, NIPS 2018.
- Don't Just Assume; Look and Answer: Overcoming Priors for Visual Question Answering

 A. Agrawal, D. Batra, D. Parikh, A. Kembhavi
 Presented at Visually-Grounded Interaction and Language workshop, NIPS 2017; at Women in Machine Learning workshop 2017; at Scene Understanding (Spotlight Presentation), Language and Vision (Spotlight Presentation), Women in Computer Vision, and VQA Challenge workshops, CVPR 2017.
- Analyzing the Behavior of Visual Question Answering Models
 A. Agrawal, D.Batra, D.Parikh Presented at *Mid-Atlantic Computer Vision* workshop 2016, *Deep Learning Summer School* 2016 (Oral Presentation) and Women in Machine Learning workshop, 2016.
- VQA: Visual Question Answering
 S. Antol*, A. Agrawal*, J. Lu, M. Mitchell, D. Batra, C. L. Zitnick and D. Parikh
 Presented at the 1st Workshop on Object Understanding for Interaction, ICCV 2015 (Best Poster Award).
- Holistic Scene Understanding via Multiple Structured Hypotheses from Perception Modules G. Christie^{*}, A. Laddha^{*}, **A. Agrawal**, S. Antol, Y. Goyal, K. Kochersberger and D. Batra Presented at *Scene Understanding, Language and Vision* and *Women in Computer Vision* workshops, CVPR 2015.

INVITED TALKS AND PANELS

Advancing multimodal vision-language learning

Area Chair Workshop @ CVPR	2024
• Understanding LLM Understanding Summer School [Video]	2024
• CIFAR AICan Retreat	2024
• World Summit AI Americas [Video]	2023
	-0

 Text in Everything workshop @ ECCV Tea Talks @ Mila [Video] 	$2022 \\ 2022$
 Visual-Language Learning Tutorial on "Visual Recognition Beyond the Comfort Zone: Adapting to Unseen Concepts or at ICCV [Video] 	n the Fly" 2023
Career PanelCIFAR Deep Learning + Reinforcement Learning Summer School	2023
Roundtable lead: RobustnessMontreal AI Symposium (MAIS)	2022
 Are current Vision-language models learning to solve the task or merely learning to dataset Tea Talks @ Mila [Video] 	o solve the 2022
 Vision and Language: Progress and Challenges Research Week with Google (India) Microsoft Research Montreal Montreal AI Symposium Mila Techaide AI Conference 	2022 2021 2021 2021
 Panel: Paths to Leadership in Data Science Women in Data Science Conference (WiDS) [Video] 	2021
 My journey with Artificial Intelligence so far Women in Computer Vision Workshop, CVPR Alumni MasterClass @ Indian Institute of Technology (IIT) Gandhinagar [Video] 	2021 2020
 Visual Grounding in Visual Question Answering GeCKo Symposium [Video] 	2020
Visual Question Answering: Progress and ChallengesUniversity of Oxford (Women in CS Seminar)	2020
 Visual Question Answering and Beyond Vector Institute for Artificial Intelligence Montreal Institute of Learning Algorithms Facebook AI Research Lab, Montreal Google Brain, Montreal Emory University (CS Seminar) Stanford University (NLP Seminar) University of Massachusetts, Amherst (Vision and Graphics Group) Massachusetts Institute of Technology (Computer Vision Group) Boston University (Computer Vision and Learning Group) University of California, Berkeley (Trevor Darrell's Group) Cornell Tech (Pixel Cafe) New York University (Machine Learning for Language Group) 	2018 2018 2018 2018 2018 2018 2018 2018
Don't Just Assume: Look and Answer: Overcoming Priors for Visual Question Ans	wering

• Women in Computer Vision workshop, CVPR [Video (32:33 - 56:08)] 2018

Winner announcements, analysis of results	
• VQA Challenge Workshop, CVPR [Video (47:22 - 1:07:09)]	2017
Analyzing the Behavior of Deep Visual Question Answering Mod	dels
• Deep Learning Summer School, Montreal [Video]	2016
Overview of challenge, winner announcements, analysis of results	5
• VQA Challenge Workshop, CVPR [Video]	2016
VQA: Visual Question AnsweringGPU Technology Conference (GTC) [Video]	2016
Research Internships	
DeepMind, London, UK Collaborators: Mateusz Malinowski, Felix Hill, Ali Eslami, Oriol Vinyals, 7	May 2018 - October 2018 Tejas Kulkarni
 Worked on generating diverse programs with instruction conditioned rei Demonstrated the efficacy of our approach on two domains - 1) drawing software conditioned on instructions and (2) constructing a scene in a 32 instruction. 	nforced adversarial learning. g MNIST digits with a paint D editor that satisfies a certain
Microsoft Research, Redmond, USA Collaborators: Xiaodong He, Jianfeng Gao	May 2017 - August 2017
Worked on developing a novel model for Visual Dialog with the goal of i and textual history.Improved the then state-of-art performance on the Visual Dialog task.	mproved modeling of both visual
Allen Institute for Artificial Intelligence, Seattle, USA Collaborators: Dhruv Batra, Devi Parikh, Aniruddha Kembhavi	January 2017 - May 2017
 Proposed a new setting for VQA and created a new split of the VQA v1 Answering under Changing Priors v1.0 (VQA-CP v1.0) Evaluated the performance of existing VQA models on VQA-CP v1.0 	1.0 dataset – Visual Question
 Developed Grounded VQA (GVQA) – a novel model that outperforms e VQA-CP v1.0. 	existing VQA models on
Microsoft Research, Redmond, USA Collaborators: Xiaodong He, Margaret Mitchell, Dhruv Batra, Devi Parikl	May 2015 - August 2015 h, Larry Zitnick
• Played an active role in releasing the VQA dataset to the public. Development and evaluation code (https://github.com/VT-vision-lab/VQA).	oped and released the VQA API
 Implemented Deep Structured Semantic Model (DSSM) based initial ap Trained and tested DSSM models for image sequences and stories, for a Storytelling) at MSR. 	proaches to solve VQA. n ongoing project (Visual
University of Washington, Seattle, USA Collaborators: Ronnie Das, Eric Seibel	May 2013 - July 2013
• Conducted an investigation comparing degree of optical clearance in par FocusClear against formalin control.	ncreatic tissue using glycerol and
• Implemented filtered backprojection algorithm for 3D reconstruction of and VolView.	OPTM images using MATLAB
Indian Institute of Technology (IIT), Gandhinagar, India Collaborator: Ragavan K.	May 2012 - July 2012
• Programmed FPGA in Verilog to generate firing pulses for SCRs of the control of DC motor.	rectifier for open loop speed

• Developed Verilog code for automatic regulation of SCR's firing angle for closed loop speed control of DC motor.

SERVICE ROLES AND ACADEMIC ACTIVITIES

• Area Chair @ NeurIPS 2024, ECCV 2024, CVPR 2024, NeurIPS 2023, ICCV 2023, NAACL 2021	
• Co-organizer, Workshop on Green Foundation Models @ ECCV	2024
• Co-organizer, tutorial on Visual Recognition Beyond the Comfort Zone: Adapting to Unseen Condon the Fly @ ICCV	cepts 2023
• Co-organizer, CIFAR Deep Learning + Reinforcement Learning Summer School	2023
• Co-organizer, tutorial on Vision-Language Pretraining: Current Trends and the Future @ ACL	2022
• Co-organizer, annual VQA Challenge and Workshop @ CVPR 2016	- 2021
 Reviewer, CVPR 2016, ECCV 2016, IJCV 2017, ICLR 2017, CVPR 2017, ICCV 2017, NIPS 2017, 2018, CVPR 2018, NeurIPS 2019, ACL 2020, CVPR 2020, ECCV 2020, ACL 2023 	ICLR
Graduate Teaching Assistant (Virginia Tech) Fall 2014, Sprin Course: Introduction to Data Structures and Algorithms	g 2015
• Teaching Assistant (IIT Gandhinagar) Fa Course: Computing (Python and C)	ll 2013
Press Coverage	

Visual Storytelling

- New Artificial Intelligence Can Tell Stories Based on Photos. Live Science.
- Will Artificial Intelligence Win the Caption Contest? MIT Technology Review.
- Teaching computers to describe images as people would. Microsoft Blog.
- Microsoft researchers are teaching AI to write stories about groups of photos. Venture Beat.

Visual Question Answering (VQA)

- A Giant Leap for Machine Kind; When Robots Can See. Interview with WVTF Radio IQ.
- Whats in This Picture? AI Becomes as Smart as a Toddler. Bloomberg.