

# Building Agents that can See, Talk, and Act

Monday, February 15, 2021, 16:30 MST

Via Zoom: <https://asu.zoom.us/j/89817732811>

## Abstract

Building intelligent agents that possess the ability to perceive the rich visual environment around us, communicate this understanding in natural language to humans and other agents, and execute actions in a physical environment, is a long-term goal of Artificial Intelligence. In this talk, I will present some of my recent work at various points on this spectrum in connecting vision and language to actions; from Visual Dialog (CVPR17, ICCV17, ECCV20) -- where we develop models capable of holding free-form visually-grounded natural language conversation towards a downstream goal and ways to evaluate them -- to Embodied Question Answering (CVPR18, CoRL18, ICML20) -- where we augment these models to actively navigate in simulated environments and gather visual information necessary for answering questions.

## Bio

Abhishek Das is a Research Scientist at Facebook AI Research (FAIR). He was previously a Computer Science PhD student at Georgia Institute of Technology, advised by Dhruv Batra, and working closely with Devi Parikh.

During and prior to his PhD, he has held visiting research positions at Queensland Brain Institute, Virginia Tech, Facebook AI Research, DeepMind, and Tesla Autopilot. He graduated from Indian Institute of Technology Roorkee in 2015 with a Bachelor's degree in Electrical Engineering.

His research focuses on deep learning and its applications in climate change, and in building agents that can see (computer vision), think (reasoning/interpretability), talk (language modeling), and act (reinforcement learning). He has published at top-tier conferences -- CVPR, ICCV, ICML, IJCAI, CoRL, ECCV, EMNLP, ICASSP -- and journals -- IJCV, PAMI, CVIU. He is a recipient of graduate fellowships from Facebook, Adobe, Snap and top reviewer awards at CVPR and NeurIPS.



**Abhishek Das**  
Facebook AI Research

Host: Tejas Gokhale, Yezhou Yang, Chitta Baral

*The Active Perception Group explores robotic visual learning, tying together the fields of active vision, natural language processing and AI reasoning.*

**ASU** Ira A. Fulton Schools of  
**Engineering**  
Arizona State University

School of Computing, Informatics  
and Decision Systems Engineering