

# 11-877 Advanced Topics in Multimodal Machine Learning

## Week 11: Dataset and model biases

**Due date: 11PM EST, Wednesday, March 30 2022**

Submission: <https://forms.gle/RPMdA3LMnpqfWQtn9>

We designed the reading assignments to help you prepare for the live discussions. Discussion probes were drafted related to this week's topic. These were written to help conceptualize the problem and guide your thought process. Take the time to read them first. The goal is not to answer each of these questions and probes individually, but they are meant to be taken as a whole. We also selected research papers relevant to this topic. Required papers should be read completely. Suggested papers should at least be skimmed. The purpose of the reading assignment is to start your critical thinking process, so your responses should demonstrate constructive thoughts, with a good understanding of the current research in this area, and express your own insights.

Your response to this reading assignment should be submitted in the online Google Form (see link above). Your response should consist of four main components:

- (1) **Scouting:** As you start thinking about the discussion probes, it is always good to also scout papers, blog posts, and other resources related to the topic. We ask that you search for related resources and share with us 2 extra links to these new resources. For each extra link, include 1-2 sentences explaining the value and relevance of this extra resource.
- (2) **Reading notes:** As you read the required papers, suggested papers, and the extra resources you scouted, please write down at least 4-6 notes related to the discussion probes. Each note should be 1-3 sentences long. These can be empirical results you observed, ideas or theories expressed by other researchers, or any interesting fact that is worth noting when summarizing your reading.
- (3) **Your thoughts:** Separate from your reading notes, we ask that you reflect more holistically about the discussion probes. Please write 3 discussion points you would like to share on this topic. Each discussion point should be one paragraph (3-5 sentences). These discussion points should go beyond the reading papers, and try to address as many aspects of the discussion probes as you can. We do not expect that you answer all discussion probes. For example, it would be ok to focus on only 1 or 2 probes if these bring the most ideas and thoughts for you.
- (4) **Clarification requests [OPTIONAL]:** Please take a moment to suggest parts of the papers where clarifications would be useful. Try to be as specific as possible in your clarification requests. These requests will be shared with the Reading Leads in charge of creating a short presentation for the beginning of Friday's course and answering other requests directly on Piazza.

### **Week 11 discussion probes:**

- *What could be a taxonomy of biases in multimodal datasets and models?*
- What are some risks related to biases (e.g., social biases) when creating new datasets? How are these risks potentially amplified or reduced when the dataset is multimodal, with heterogeneous modalities? Are there any biases that are specific to multimodal data?
- What are the imperfections that may arise during human annotations? How do these imperfections in data and labels affect multimodal learning of multimodal representations, cross-modal interactions, co-learning, and pre-training?
- Can biases also emerge not only from the multimodal training data, but also from the modeling design decisions themselves? What aspects of multimodal modeling are most prone to learning and possibly emphasizing biases?
- What are potential solutions for tackling these risks and biases in multimodal datasets and models? How can we properly identify, visualize and eventually reduce these biases in multimodal datasets and models?
- How can we better interpret multimodal datasets and models to check for potential biases? What specific dimensions should we strive to understand?
- What are the tradeoffs between large-scale, noisily-collected and annotated multimodal datasets versus small-scale, carefully-curated and annotated datasets? How do these affect multimodal modeling? How does it relate to the popular pre-training paradigm?

### **Required papers (you should read these papers in detail)**

- <https://www.nature.com/articles/s42256-020-00257-z>
- <https://aclanthology.org/2021.naacl-main.78.pdf>

### **Suggested papers (you should skim through these papers, at the minimum)**

- <https://arxiv.org/abs/2110.01963>
- <https://www.aaai.org/AAAI21Papers/AAAI-9821.YeK.pdf>
- <https://arxiv.org/abs/1908.07898>
- <https://arxiv.org/abs/2103.06254>
- <https://arxiv.org/abs/2202.04053>

### **Other relevant papers:**

- <https://arxiv.org/abs/1606.08390>
- <https://arxiv.org/abs/1606.07356>
- <https://arxiv.org/abs/2002.04108>
- <https://arxiv.org/abs/2009.10795>
- <https://arxiv.org/abs/1803.02324>
- <https://arxiv.org/abs/2111.15366>
- <https://arxiv.org/abs/2110.14375>
- <https://arxiv.org/abs/2108.02922>
- <https://www.sciencedirect.com/science/article/pii/S0893608005000407>