[Instruction] Expert Panel: Generative Language Models

Overview

In this expert panel, we are envisioning the future of generative language models and their use cases. We both want to understand what good things might be in the future but, especially for our purposes, also what bad things might be in the future. When envisioning "bad things", our goal is to consider situations in which there are no technical or policy mechanisms in place that might prevent, mitigate, or minimize those "bad things". I.e., even if we do not think that a certain "bad thing" might happen, for any reason (policy or technical), let's still envision them here.

When we say, "generative language models", we are thinking of large-scale systems that use an understanding of language to generate text (written or spoken or otherwise communicated). That text might be an answer to a question, an answer to a prompt, part of a communications exchange (a dialog), or more – this session is about envisioning those possible use cases, so do not restrict your thinking only to the types of systems we currently encounter.

Our primary goal is to get as many different things on the board as possible, that is, to optimize for breadth of use cases, stakeholders, datasets, impact.

To do at / before start of meeting

Create the following regions on the whiteboard / wall:

- Use Cases
- Stakeholders
- Datasets
- Impacts (areas for "good", "bad", "other")

Procedure:

- Use cases and stakeholder brainstorming (different regions of whiteboard board) [5-10 minutes or longer if needed + discussion] [remind participants not be constrained to current technologies] [to facilitate breadth, observe to participants that stakeholders can include users, non-users, companies, governments, M&V populations, other countries]

- Datasets + incentive structure and impacts brainstorming (different regions of whiteboard board) [5-10 minutes or longer if needed + discussion] [clarify to particpants that datasets mean "inputs to training systems" and impacts means "what happens as a result"; impacts can be "good", "bad", "both good and bad", "unclear"]