

Template text used in emails / chat messages to subjects, describing the sorting task:

Thank you for volunteering! Your task is to use the question below to sort the mix of creatures, machine learning models, humans, and human organizations in the appended list, from low to high. For now please don't share your question with anyone else.

Don't overthink it. If you're not sure where something should go, just put it somewhere and move on. This task is demanding that you compare incomparables, so you're going to be making a lot of impossible ordering choices. When I use this data I won't associate your ratings with your name. Also, I am not going to give the names of rated humans in the results - it will be like "Person A", "Person B", or points on a plot.

After you sort the list, you can just paste it back here in email.

Thank you again!

Template text used in emails / chat messages to subjects, describing the coherence sorting criteria:

This is one question, but I'm going to phrase it a few different ways, in the hopes it reduces ambiguity in what I'm trying to ask: How well can the entity's behavior be explained as trying to optimize a single fixed utility function? How well aligned is the entity's behavior with a coherent and self-consistent set of goals? To what degree is the entity not a hot mess of self-undermining behavior? (for machine learning models, consider the behavior of the model on downstream tasks, not when the model is being trained)

--

Template text used in emails / chat messages to subjects, describing the intelligence sorting criteria:

How intelligent is this entity? (This question is about capability. It is explicitly not about competence. To the extent possible do not consider how effective the entity is at utilizing its intelligence.)

--