

# Discussion Questions

1. Why do we want a unified model for multimodal understanding tasks?
2. Is the target modality for self-supervision best done randomly or should there be a more principled approach to select it?
3. How does the inclusion of a diverse set of modalities impact the proposed method's ability to generalize across different tasks compared to single-modality models?
4. How does Unified-IO 2 compare to modularized approaches like ViperGPT? Would ViperGPT achieve similar if not better results by leveraging the modality specific SOTA models?
5. Should we focus on designing monolithic, multimodal models that can solve many tasks or task-specific models that have a single, specialized, niche purpose?
6. Is it better to have a unified model that can perform many tasks at a lower accuracy or a task-specific model that can perform one task very well?
7. Is the general framework of Unified-IO 2 what GPT-5 will look like?
8. Does it make sense to unify discriminative recognition tasks with generative tasks? Can the model effectively learn both?
9. How might future research address the noted decrease in performance for tasks not seen during the training of UNIFIED-IO 2?

# Discussion Questions

1. Why do we want a unified model for multimodal understanding tasks?

# Discussion Questions

**2.** Is the target modality for self-supervision best done randomly or should there be a more principled approach to select it?

# Discussion Questions

**3.** How does the inclusion of a diverse set of modalities impact the proposed method's ability to generalize across different tasks compared to single-modality models?

# Discussion Questions

4. How does Unified-IO 2 compare to modularized approaches like ViperGPT? Would ViperGPT achieve similar if not better results by leveraging the modality specific SOTA models?

# Discussion Questions

**5.** Should we focus on designing monolithic, multimodal models that can solve many tasks or task-specific models that have a single, specialized, niche purpose?

# Discussion Questions

**6.** Is it better to have a unified model that can perform many tasks at a lower accuracy or a task-specific model that can perform one task very well?

# Discussion Questions

7. Is the general framework of Unified-IO 2 what GPT-5 will look like?



# Discussion Questions

**8.** Does it make sense to unify discriminative recognition tasks with generative tasks? Can the model effectively learn both?

# Discussion Questions

**9.** How might future research address the noted decrease in performance for tasks not seen during the training of UNIFIED-IO 2?