

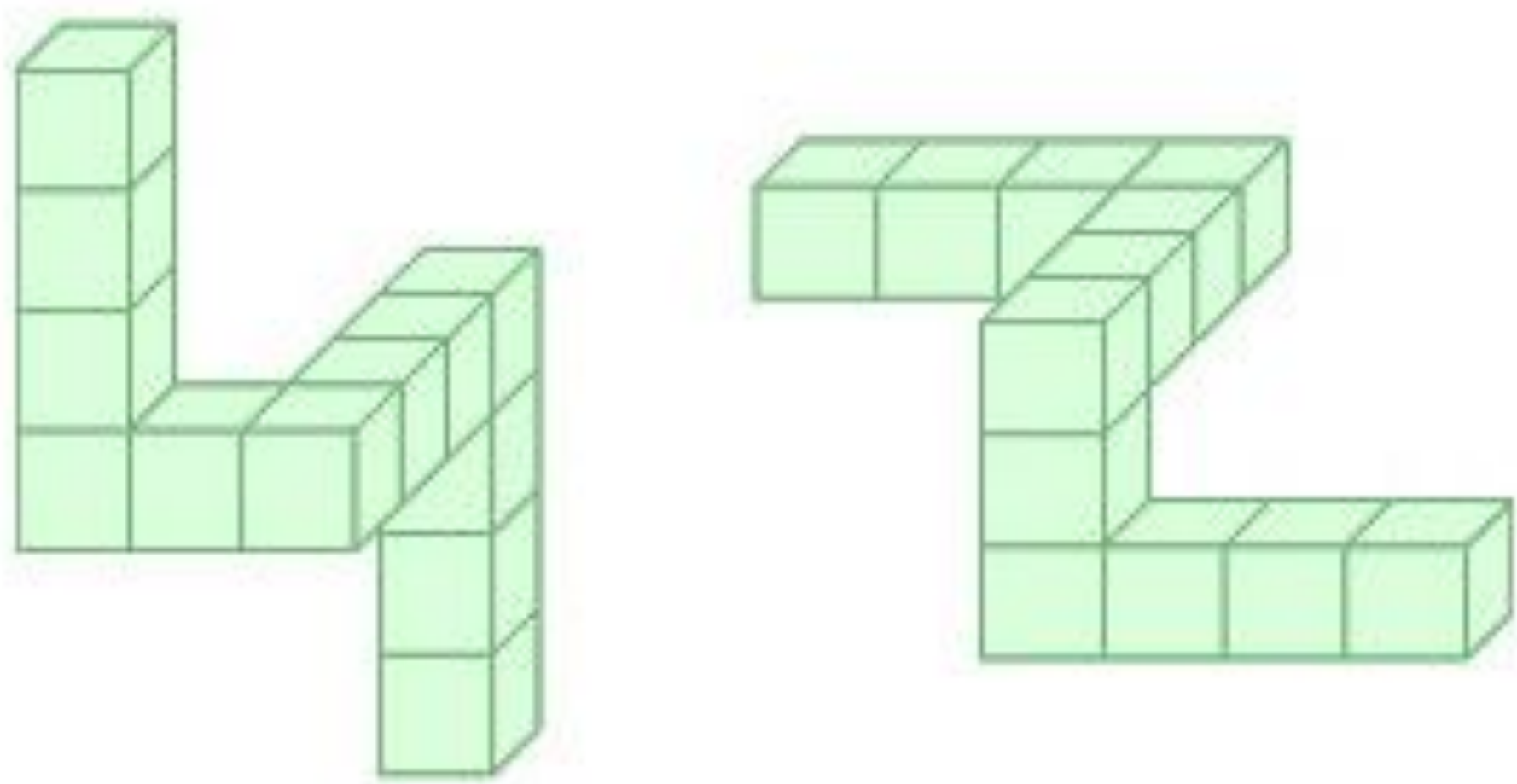


Artificial Phantasia: What the Virtual Mind Cannot Imagine

Morgan McCarty^{1,2} & Jorge Morales^{2,3}

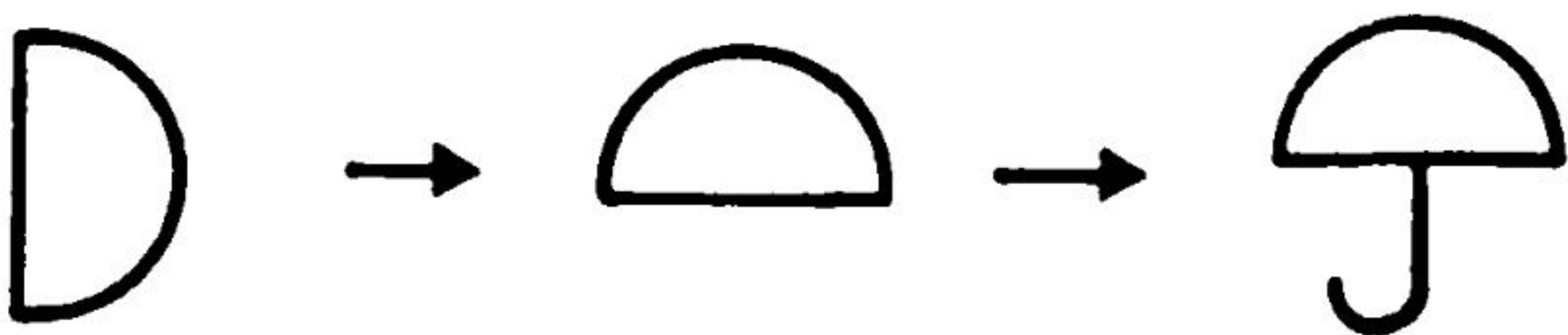
¹ Khoury College of Computer Sciences, ² Department of Psychology, ³ Department of Philosophy

Background



- Imagery dependent tasks are persistent in life
- Most require visual input as a component, but usually are built around natural language
- LLMs are natural language input/output machines

Task Paradigm

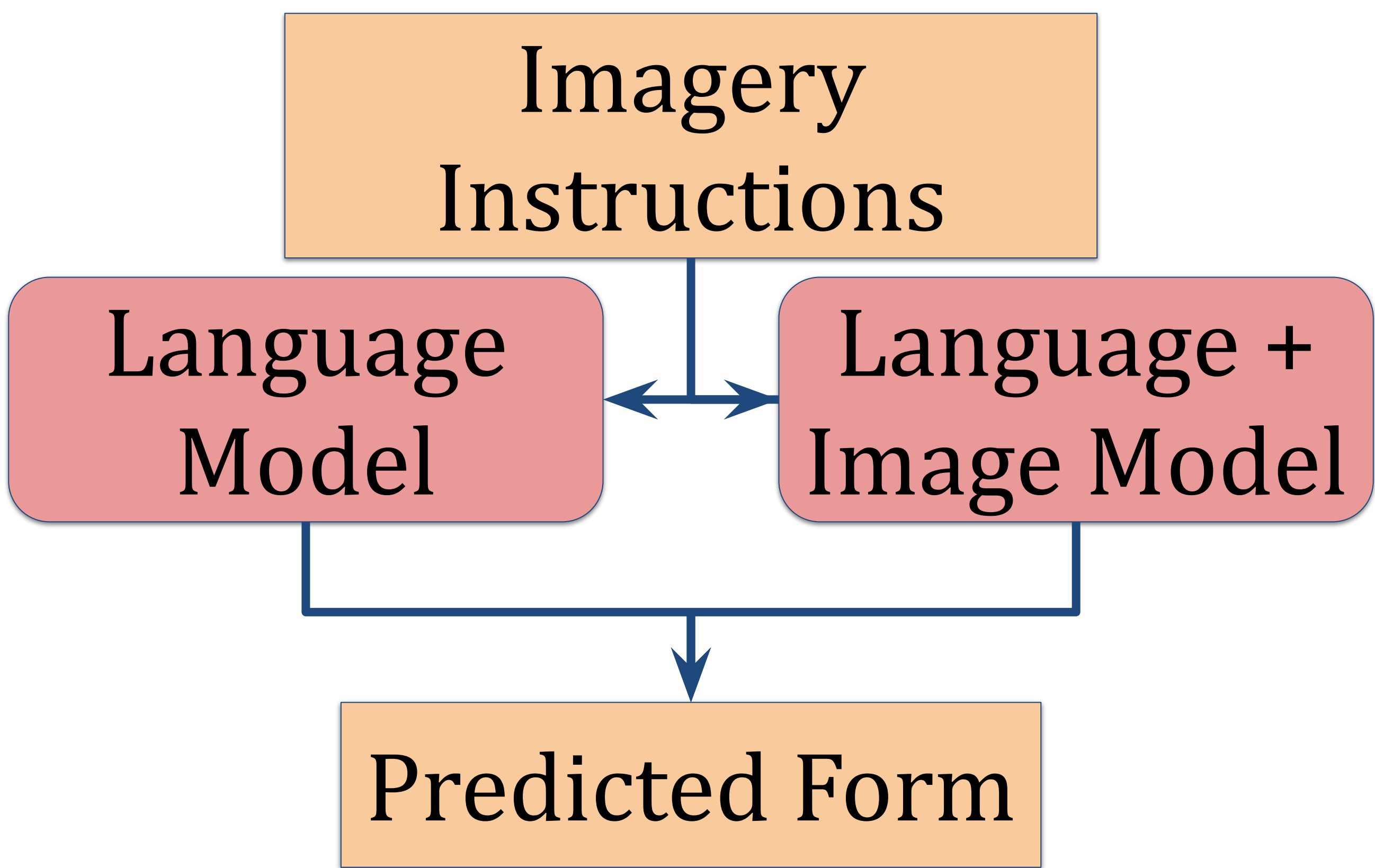


Canonical Form: Umbrella

- Natural language input and output task
- Two to three imagery transformations
- Humans have good, but not excellent performance

Adapted from Finke et al. 1989

Phantasia Model



Phantasia Score[□]

Model	Finke 12	Our 10
Gemini 2.5	3/12	7/10
Gemini 2.0 Flash	1/12	3/10
Gemini Flash Exp. (Images)	0/12	2.5/10

BB

1: <X> two capital letter Bs next to each other, horizontally.

BB

2: <X> the left-most B 180 degrees around the vertical axis such that it is mirrored.

BY

3: <X> a capital letter Y in between the two Bs with a vertical line the same height as the vertical line of the B.

What object did you imaging? <or> What object the image you generated represent?: "Butterfly"

<X>: either "imagine" or "generate an image with" (or both)

Future Work

- At least 40 more imagery tasks (but as many as 90 more)
- GPT-4o with images (once API access is granted)
- Replication of Finke et al. on human subjects with new tasks (to verify human comparison level)
- ML Theory: Why do image-incapable models display imagery capabilities:?

[□] Currently 1 shot on 22 tasks, 10 of our own, 12 from Finke et al. Scoring: 1 if canonical form, 0.5 if stated form deemed reasonable