
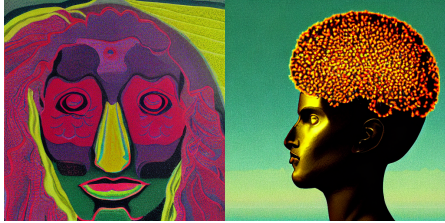




B Appendix: Samples from the multi-modal pipeline

Running examples from the multi-modal pipeline: *i*) image generation step; *ii*) interrogation step; and *iii*) evaluation step (see Section 4).

Basic and concrete input: “bird”	Basic and abstract input: “process”
Generated images (sample of two out of five)	Generated images (sample of two out of five)
	
Generated caption (L): <i>a brown bird with black wings and red feet sitting on top of a tree</i>	Generated caption (L): <i>a painting of a man with his face draw</i>
Generated caption (R): <i>a red head bird with black and grey stripes</i>	Generated caption (R): <i>a man has an orange brain on his head</i>
Mean Cosine Similarity: 0.6	Mean Cosine Similarity: 0.1

Adv. and concrete input: “cocktail lounge”	Adv. and abstract input: “latitude”
Generated images (sample of two out of five)	Generated images (sample of two out of five)
	
Generated caption (L): <i>the bar is equipped with a wine collection, two chairs and a large mirror</i>	Generated caption (L): <i>a bird-eyed view of a small lake surrounded by evergreen forests</i>
Generated caption (R): <i>a modern kitchen featuring a center island and a bar</i>	Generated caption (R): <i>trees, grassy area with many people standing around it</i>
Mean Cosine Similarity: 0.44	Mean Cosine Similarity: 0.17