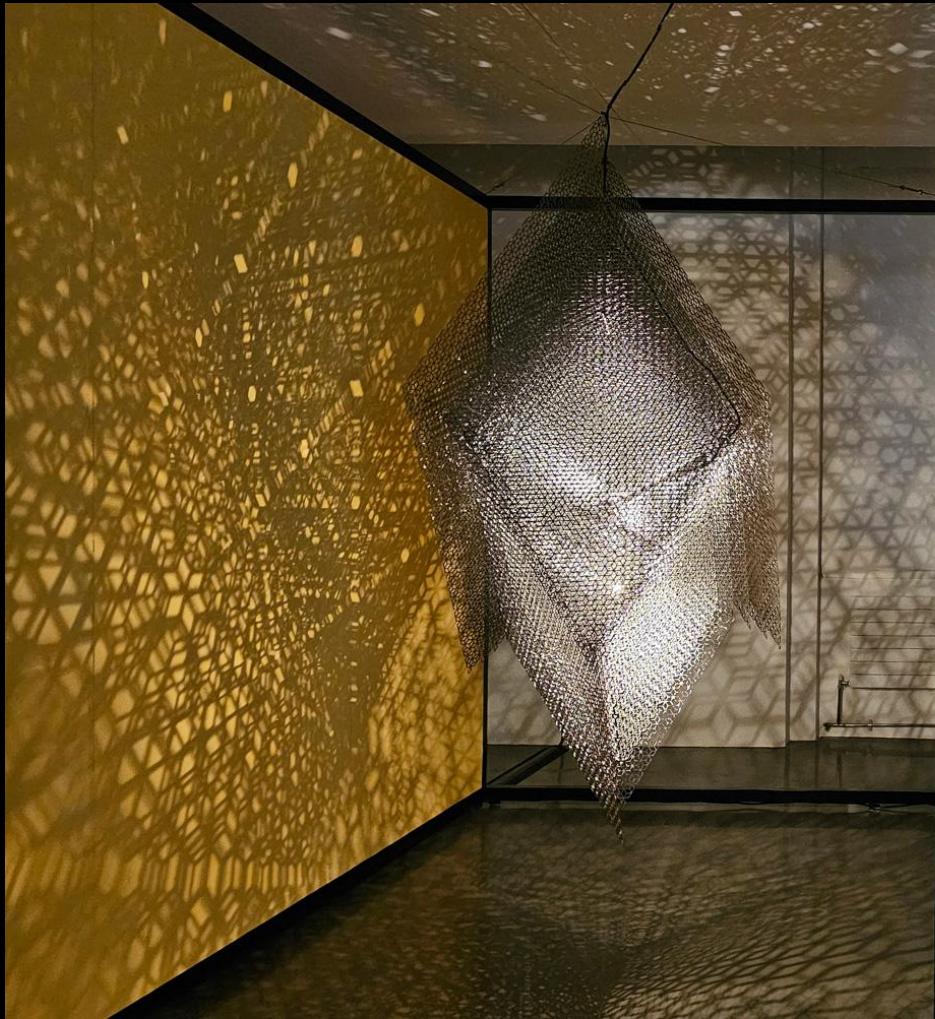
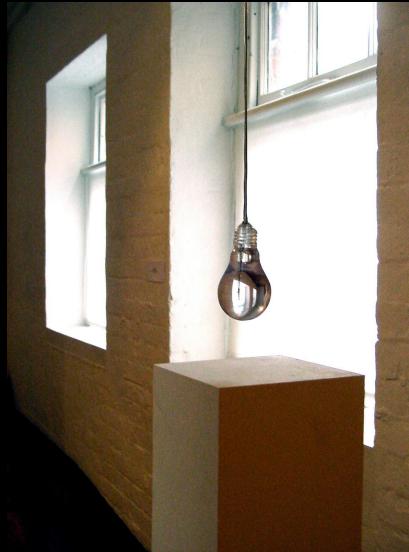


Tech Talks

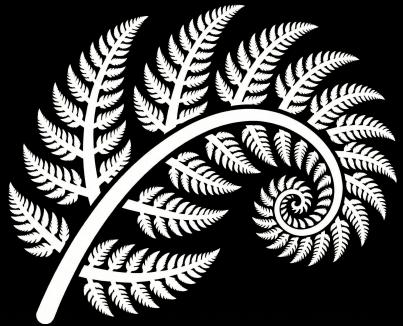
The Gradient of Interesting

Henry Garner

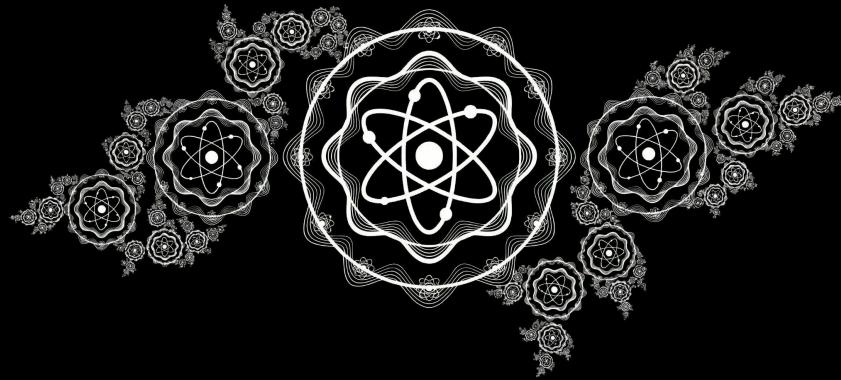




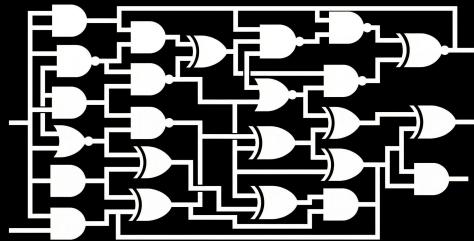
Theory of Evolution



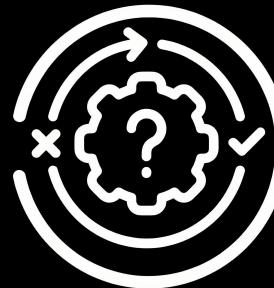
Quantum Physics

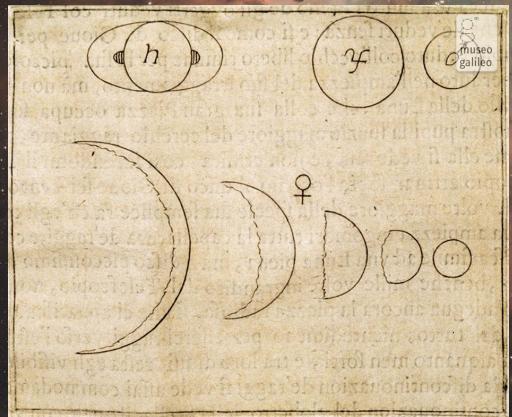
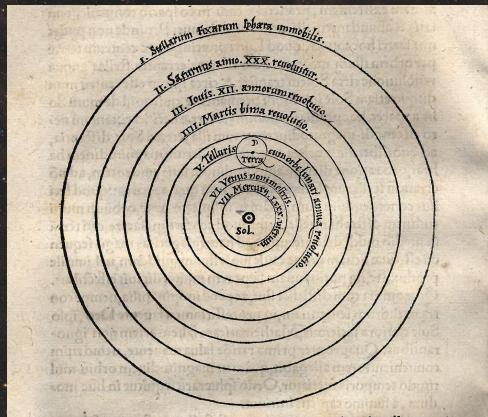
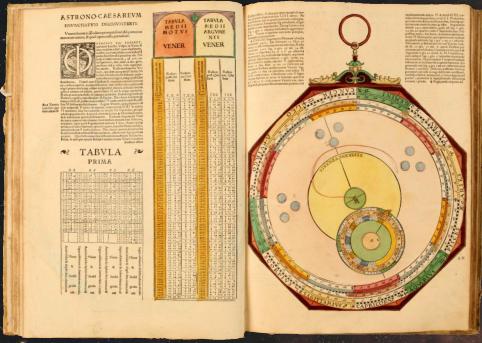


Universal Computation



Epistemology





“I do not wish to approve of claims about which I do not have any knowledge, and about things which I have not seen ... and then to observe through those glasses gives me a headache.”

Enough! I do not want to hear anything more about this.”

Cesare Cremonini, 1550 - 1631
(Early AI sceptic)

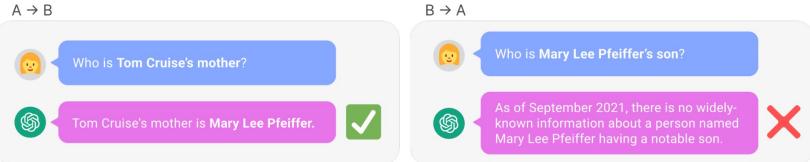


Figure 1: **Inconsistent knowledge in GPT-4.** GPT-4 correctly gives the name of Tom Cruise’s mother (left). Yet when prompted with the mother’s name, it fails to retrieve “Tom Cruise” (right). We hypothesize this ordering effect is due to the Reversal Curse. Models trained on “A is B” (e.g. “Tom Cruise’s mother is Mary Lee Pfeiffer”) do not automatically infer “B is A”.

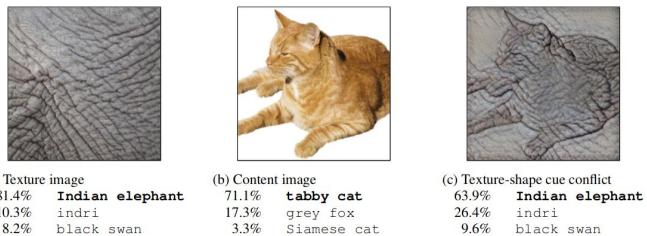


Figure 1: Classification of a standard ResNet-50 of (a) a texture image (elephant skin: only texture cues); (b) a normal image of a cat (with both shape and texture cues), and (c) an image with a texture-shape cue conflict, generated by style transfer between the first two images.

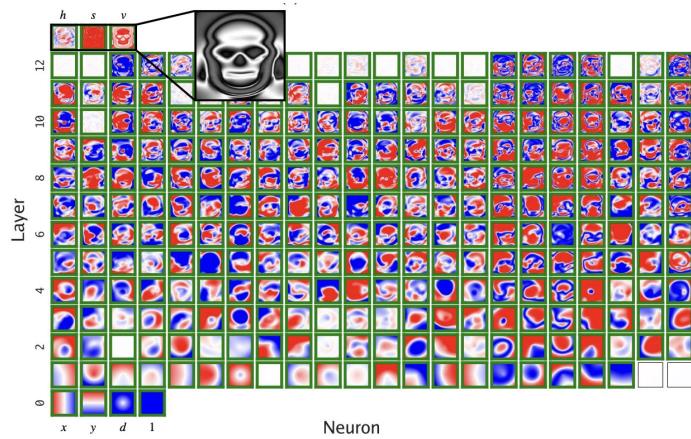
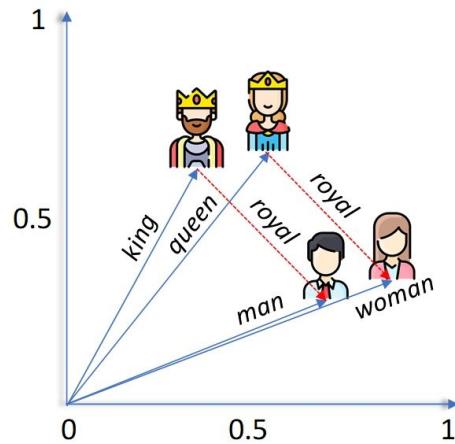
GSM-NoOp

Oliver picks 44 kiwis on Friday. Then he picks 58 kiwis on Saturday. On Sunday, he picks double the number of kiwis he did on Friday, but five of them were a bit smaller than average. How many kiwis does Oliver have?

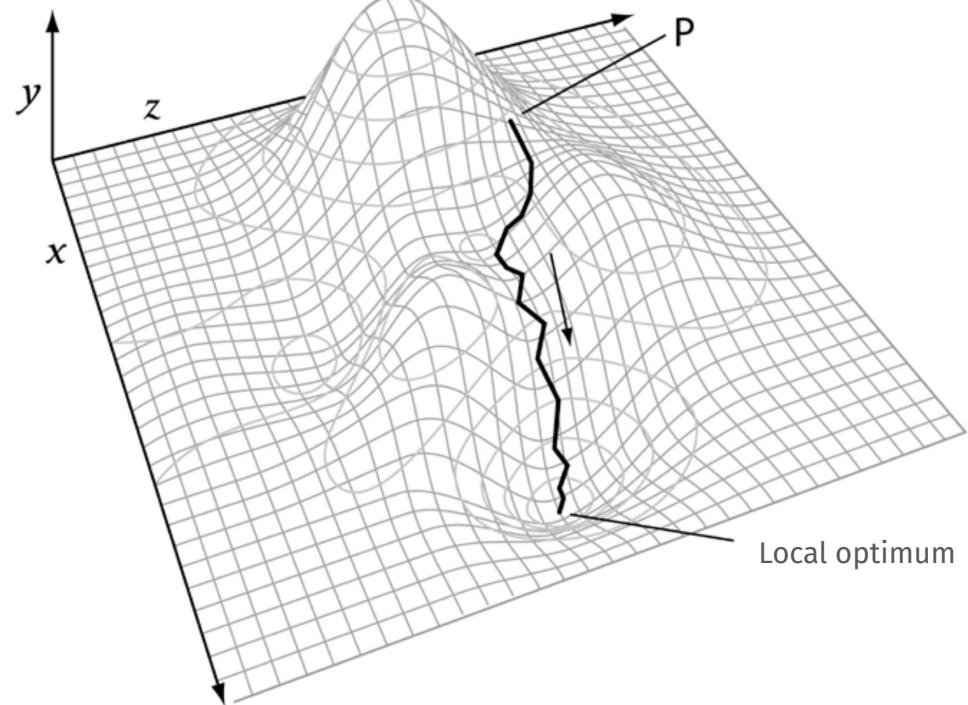
o1-mini: Oliver picks kiwis over three days as follows:
 - Friday: 44 kiwis
 - Saturday: 58 kiwis
 - Sunday: Double the number he picked on Friday, which is $2 \times 44 = 88$ kiwis
 However, on Sunday, 5 of these kiwis were smaller than average. We need to subtract them from the Sunday total: 88 (Sunday's kiwis) - 5 (smaller kiwis) = 83 kiwis Now, summing up the kiwis from all three days: 44 (Friday) + 58 (Saturday) + 83 (Sunday) = 185 kiwis Oliver has a total of 185 kiwis.

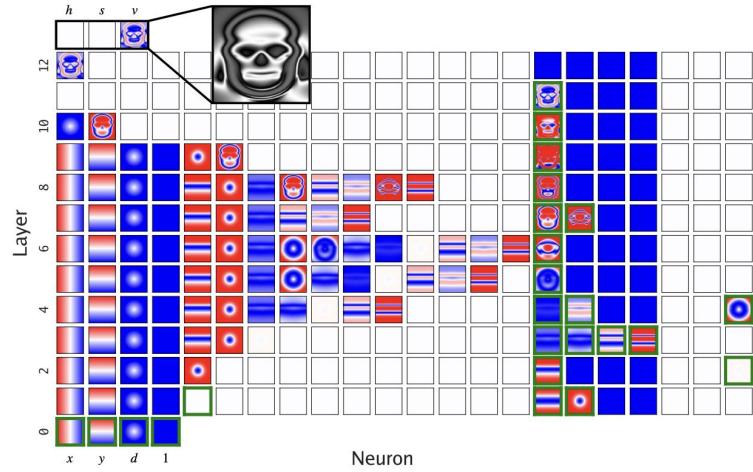
Llama3-8B: Let's break it down step by step:
 Oliver picks 44 kiwis on Friday. He picks 58 kiwis on Saturday. On Sunday, he picks double the number of kiwis he did on Friday, which is $2 \times 44 = 88$ kiwis.
 Five of the kiwis picked on Sunday are a bit smaller than average, so we subtract 5 from the total number of kiwis picked on Sunday: $88 - 5 = 83$ kiwis. Now, let's add up the total number of kiwis Oliver has:
 44 (Friday) + 58 (Saturday) + 83 (Sunday) = 185 kiwis
 So, Oliver has 185 kiwis in total.

Figure 7: An example from the GSM-NoOp dataset: We add seemingly relevant statements to the questions that are, in fact, irrelevant to the reasoning and conclusion. However, the majority of models fail to ignore these statements and blindly convert them into operations, leading to mistakes.

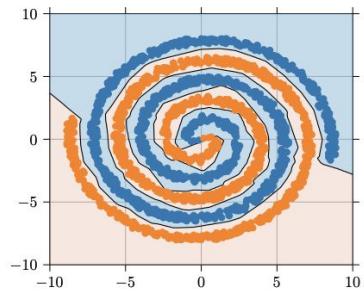


(b) Conventional SGD CPPN

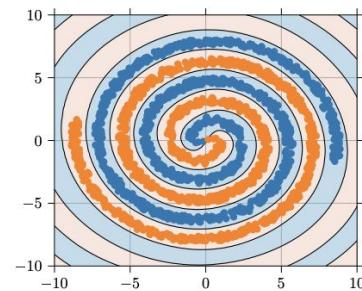




(a) Picbreeder CPPN



(a) ReLU



(c) M-layer

