

Reasoning with images

(Summary from openai tutorials by JW 2.1.2025, sample code available)

Image understanding with GPT-4o can perform well, but often needs chain-of-thought prompting, few shot examples or fine tuning to achieve the best results.

Conversely, o1 performs well at understanding images out-of-the-box.

This is due to the test and learn approach it follows in its reasoning, which gives it multiple chances to detect hallucinations before providing an answer.

One use case that is emerging is to incur the latency and cost hit for o1 upfront.

Preprocessing the image and indexing it with rich details so that it can be used for Q&A later.

Organizational structure use case

For this image reasoning task an org structure of a fictional organization.

1. Ask o1 to tell us what this is (image)
2. Use o1 to do image understanding and extract a detailed Json that describes that image and what's going on in it in a lot of detail
3. Then you can do text-only follow-ups to then interpret the information.

So you are not paying the extra cost and latency of processing the image every time.

You effectively preprocess it with o1 in a very high quality, detailed manner, and then use it for follow-up Q&A.

