GPT Image Compare

How well does GPT 4 Vision compare objects in images?

Stephen Witty
switty@level500.com
GitHub: level500.com
870-688-5220

3ken.cc

AutoGro

Project Goal

Determine how well ChatGPT Vision compares objects in images

Measure results empirically

Create a framework that allows different images / objects to be easily tested

Methodology

Create a program that submits images to ChatGPT Vision

Ask ChatGPT to indicate if two objects in the image are the same or different

The program will select images at random to compare from a source pool

Multiple test runs will take place automatically

Results will be scored and reported

Implementation

Program written in Python

The source objects are 10 stuffed animal toys

Each toy was photographed 5 times from different front-facing orientations

50 source photos total

For each test run, two photos were randomly picked and combined

Photos were scaled to 800 x 600 and submitted to ChatGPT 4 Vision

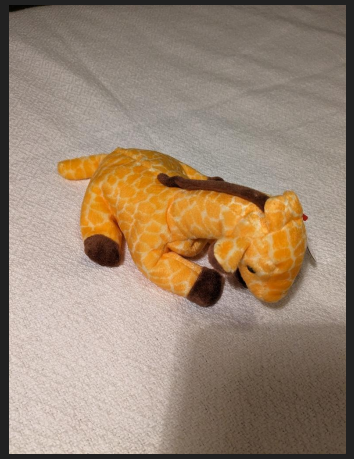
GPT was asked if the objects were the same

Source Material

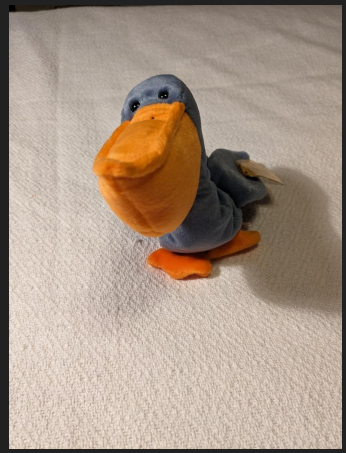






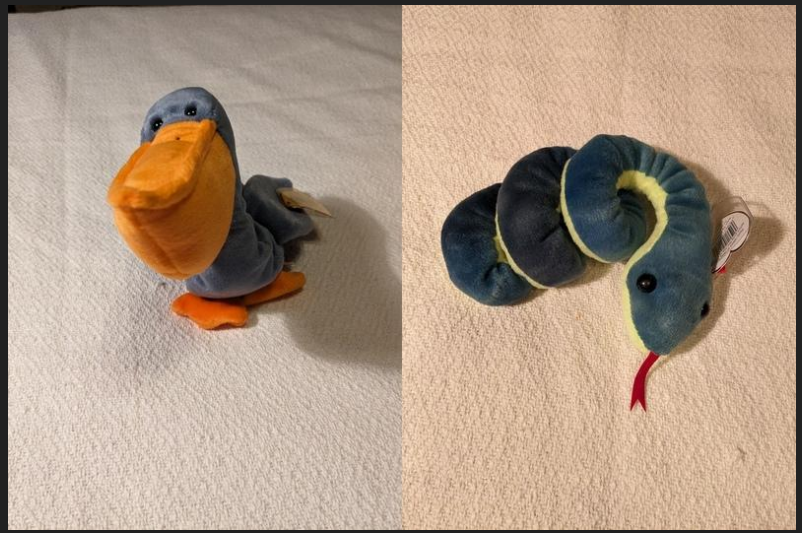



Each stuffed animal was photographed from 5 front positions for 50 photos total across 10 toys



Example successes - GPT identified as the same or different correctly

Identified as the same by GPT



Identified as different by GPT

Example failures - GPT misidentified

GPT identified both sets of images as the same object

Results

Test run of 500

Each run was two combined images

GPT asked if the objects in the images were the same or different

Correct answers: 489

Wrong answers: 11

ChatGPT Vision was correct 97.8% of the time