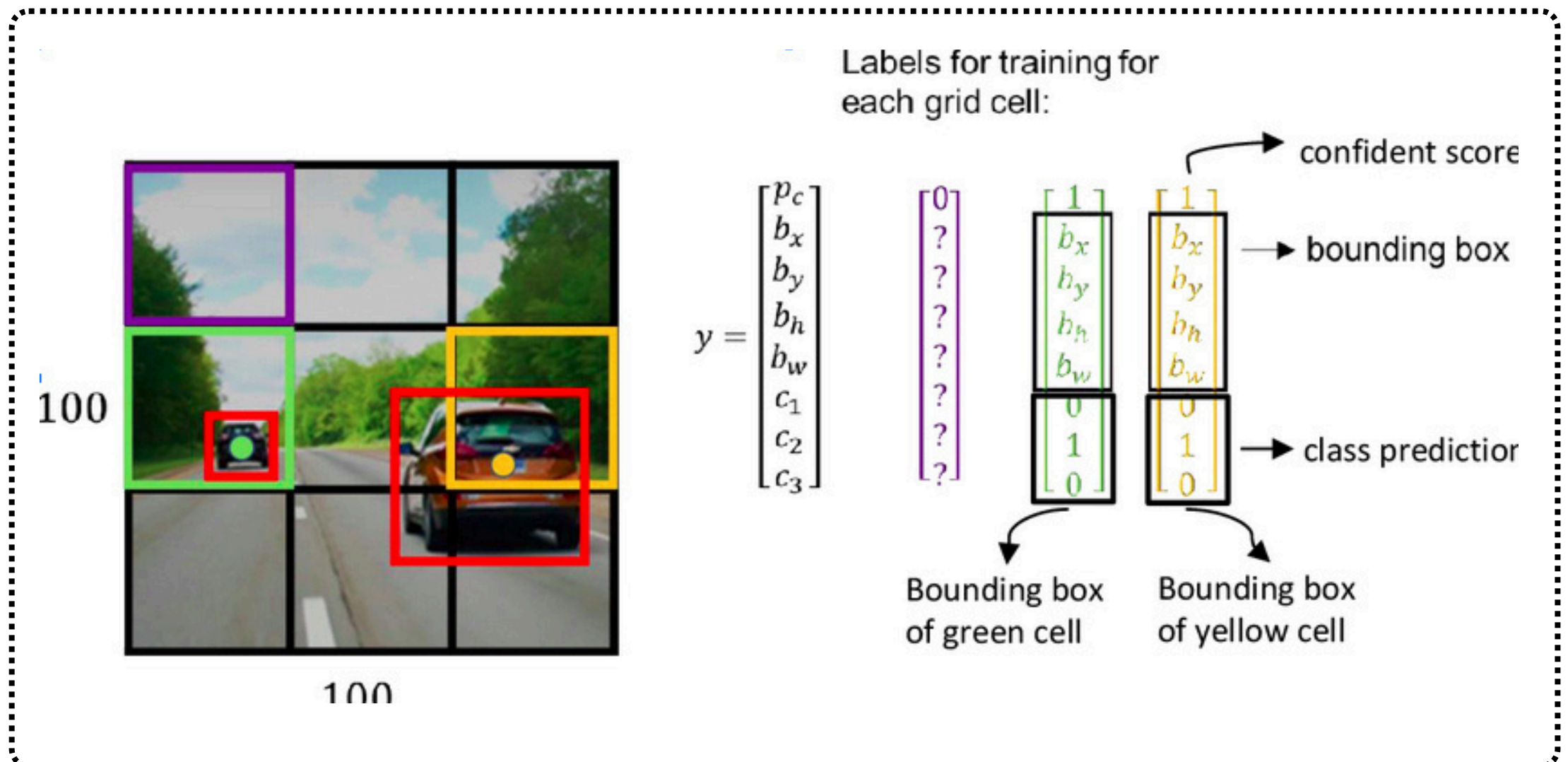


How to Use Google Gemini Models

for Computer Vision Tasks?



Gemini

Why Use Google Gemini for Computer Vision?

Google Gemini makes computer vision easier and more powerful. Here's how it helps:

1. **Works with Both Images and Text:** You can send both **images** and **text** in one request, allowing tasks like detecting objects and describing scenes.
2. **Can Think and Follow Instructions:** Gemini analyzes images and follows your instructions, giving you **accurate results** based on both text and visuals.
3. **One Model for All Tasks:** No need to create different models for each task. Use **one API** for multiple vision tasks, saving time.
4. **Easy to Use:** Developers can quickly integrate Gemini with a simple API, no **deep technical knowledge required**.
5. **Custom Training:** For special tasks, use tools like **YOLO Annotator** to label data and train the model to be more accurate for specific needs.

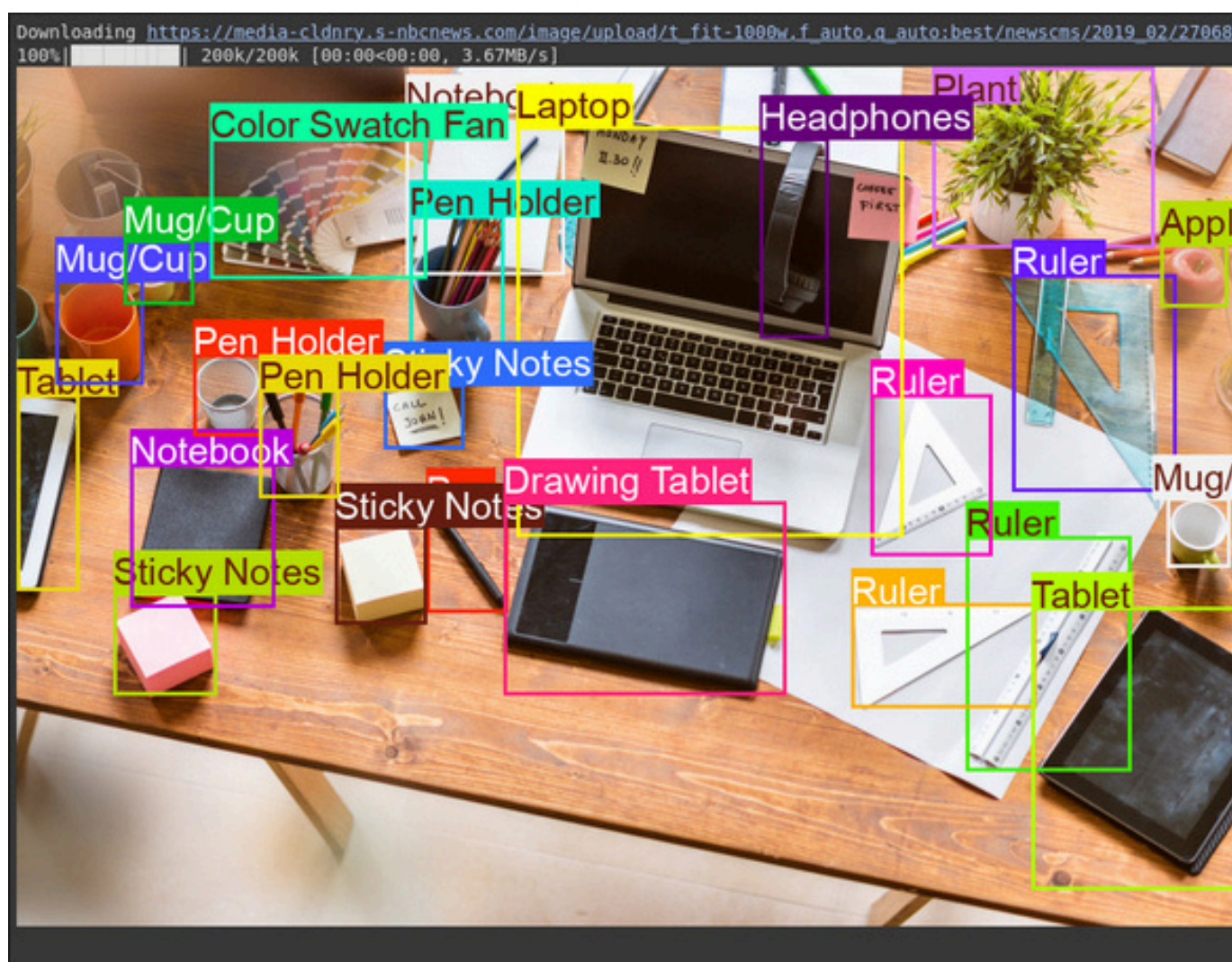
Getting Started

- Install **required libraries** to communicate with the **Gemini API** and handle images.
- Import necessary Python libraries for **image handling, API interaction**, and **analysis**.
- Set up your Google AI API key to authenticate requests to Gemini.
- Set up a function to **send images and prompts** to Gemini and retrieve results.
- Load and format images correctly before sending them to the model.
- **Clean and structure** the output from Gemini for visualization or analysis.

```
def read_image(filename):  
    image_name = safe_download(filename)  
  
    # Read image with opencv  
  
    image = cv2.cvtColor(cv2.imread(f"/content/{image_name}"),  
cv2.COLOR_BGR2RGB)  
    # Extract width and height  
  
    h, w = image.shape[:2]  
  
    # # Read the image using OpenCV and convert it into the PIL format
```


Task 1: Object Detection

Gemini can find objects in an image and report their locations (bounding boxes) based on your text instructions.



The prompt guides the model on detection and output format. It converts normalized coordinates to pixel values, and the Annotator tool draws boxes and labels on the image.

Task 2: Image Captioning

Gemini can create text descriptions for an image.

```
# Define the text prompt

prompt = """
What's inside the image, generate a detailed captioning in the form of short
story, Make 4-5 lines and start each sentence on a new line.
"""

image, _, _ = read_image("imageurl") # Read image and extract width,
height
plt.imshow(image)

plt.axis('off') # Hide axes

plt.show()

print(inference(image, prompt)) # Display the results
```

Input

The prompt requests a narrative description (4 lines, new lines). The image is shown, and the function returns the generated text, useful for alt text or summaries.



The copper lady stood tall, bathed in the warm glow of the afternoon sun. Her torch reached towards the heavens, a beacon against the pale blue sky. Across the rippling water, the mighty skyline of Manhattan rose like a modern forest. Skyscrapers jostled for space, with the Empire State Building prominent among them. She watched silently, a timeless guardian overlooking the bustling city.

Output

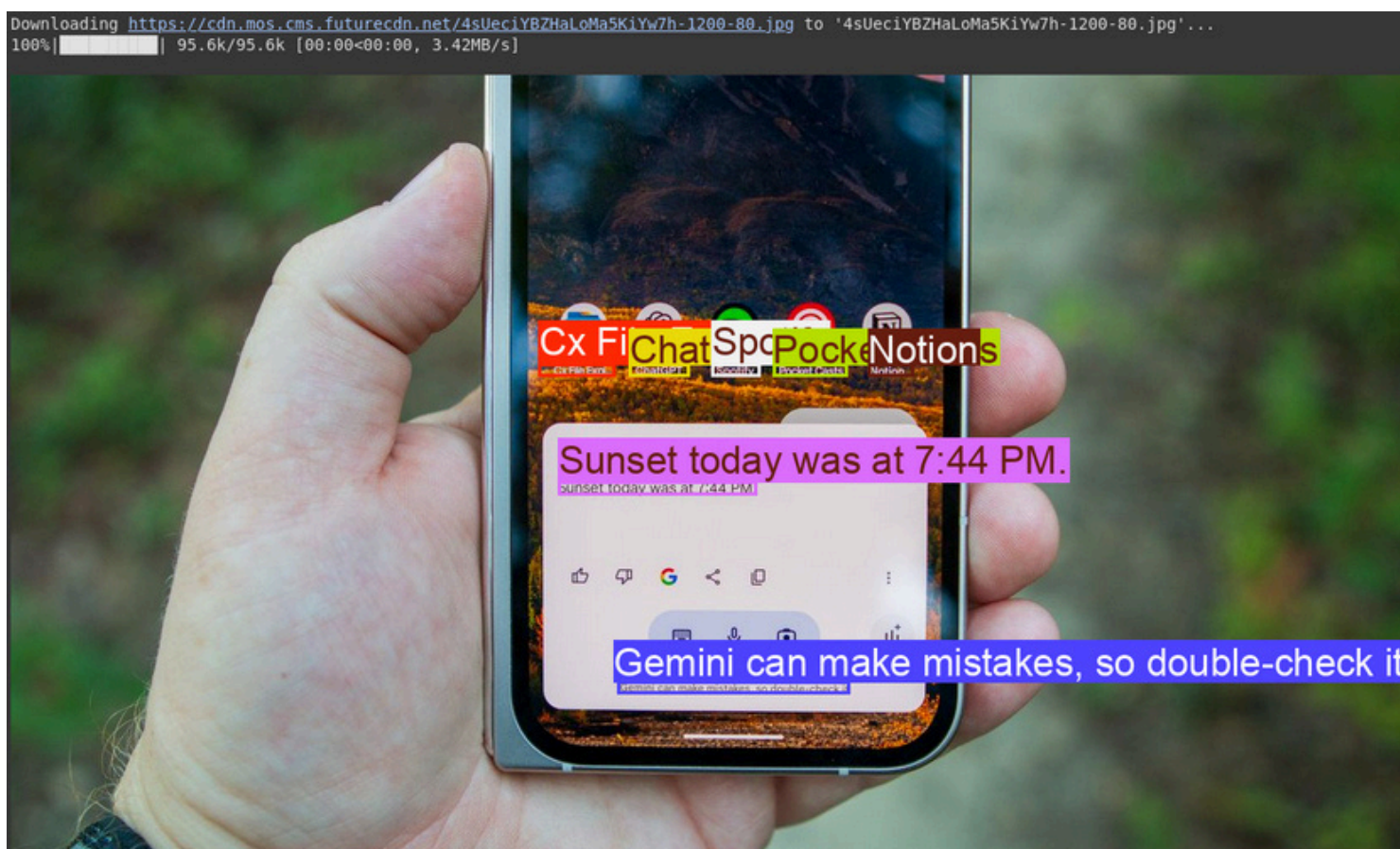
Task 3: OCR

Gemini can read text within an image and tell you where it found the text.

```
prompt = ""  
  
Extract the text from the image  
  
""  
  
# Fixed, plotting function depends on this.  
  
output_prompt = ""  
  
Return just box_2d which will be location of detected text areas + label""  
  
image, w, h = read_image("imageurl") # Read image and extract width,  
height  
results = inference(image, prompt + output_prompt)  
  
# Clean the results and load results in list format  
  
cln_results = json.loads(clean_results(results))  
  
print()  
  
annotator = Annotator(image) # initialize Ultralytics annotator
```

Input

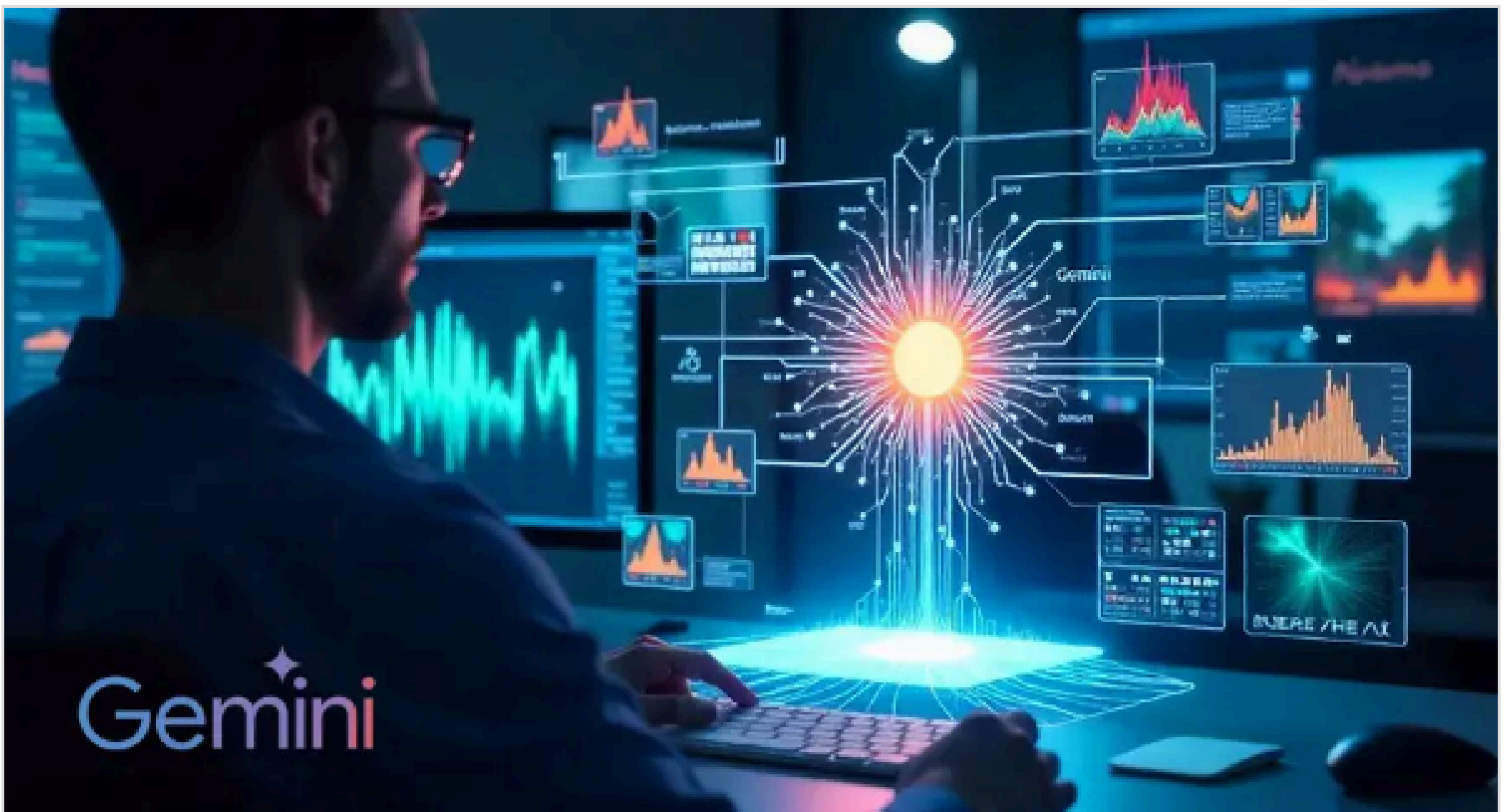
It locates and displays the text, drawing boxes on the image, useful for digitizing documents or reading text from images.



Output

Google Gemini for Computer Vision makes it easy to tackle tasks like object detection, image captioning, and OCR through simple API calls. By sending images along with clear text instructions, you can guide the model's understanding and get usable, real-time results.

To Know More..



How to Use Google Gemini Models for Computer Vision Tasks?

Explore Google Gemini for computer vision and unlock its potential in object detection, caption generation, and more.