

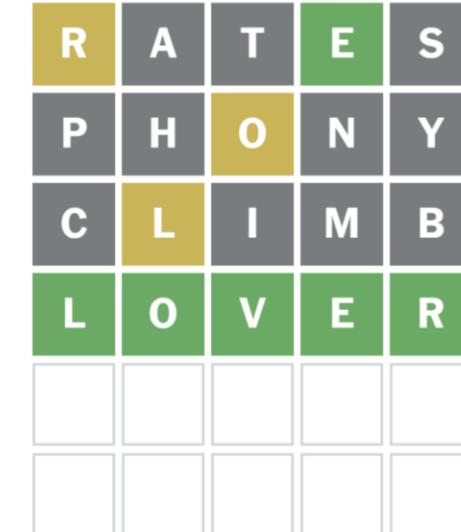
Using GPT Visual Capabilities to Solve a Wordle Puzzle

Jennifer Marsman
Principal Engineer, Office of the CTO

jennmar@microsoft.com

<https://www.linkedin.com/in/jennifermarsman>

<https://twitter.com/JenniferMarsman>



GPT-4V (model: gpt-4-vision-preview)

- Not Roman numeral for 5; the V is for visual.
- The model does not receive image metadata.
- Currently supports PNG, JPEG/JPG, WEBP, and non-animated GIF.
- Image upload limit is 20MB per image.
- No fine-tuning support
- Does not support the **message.name** parameter, **functions/tools**, **response_format** parameter, and we currently set a low **max_tokens** default which you can override.

This info is current as of November 2023. Things can and will change.

GPT-4V (model: gpt-4-vision-preview)

- Timeframe of this work: November and early December 2023
- State of the art model has changed – use GPT 4o
- Do NOT follow the coding practices from the repo. I was writing code before there were proper SDKs, function support, etc.

Scenarios

(your mileage may vary – always test)

Retail

What dress is she wearing?

She is wearing a green lace midi dress with a rainbow belt and gold Gucci loafers. This dress is from the Gucci Spring 2016 Ready-to-Wear collection.



Sports analysis

What is wrong with this skier's form?

There are a few things wrong with this skier's form:

1. They are leaning too far forward: Ideally, your shins should be in contact with the front of your boots, but your body should be more upright. Leaning too far forward can cause you to lose control and fall.
2. They are not properly angulating their knees: The red lines in the image show that the skier's knees are not pointing in the same direction as their skis. This can cause them to lose their edge and slide out.
3. They are not properly balancing on their outside ski: The blue line in the image shows that the skier's weight is not properly distributed over their outside ski. This can cause them to lose their balance and fall.



Overall, this skier needs to work on their balance, angulation, and stance to improve their form and have better control on the slopes.

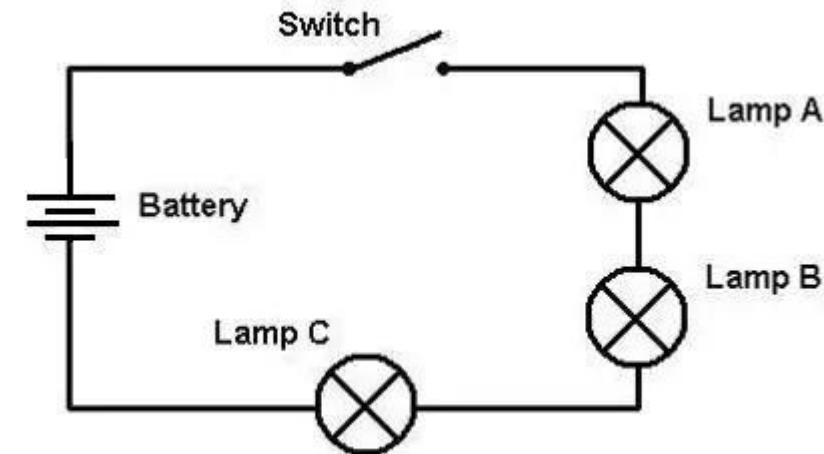
Circuit diagram

Will the lamps in this image work?

No, the lamps in this image will not work.

There are a few reasons why the lamps will not work:

1. The circuit is incomplete: the battery is not connected to the switch properly, and the switch is not connected to the lamps properly. Without a complete circuit, electricity cannot flow and the lamps will not light up.
2. The lamps are not connected in parallel: In order for multiple lamps to work in a circuit, they need to be connected in parallel. This means that each lamp should have its own separate path to the battery. In this image, the lamps are connected in series, which means that they share the same path to the battery. This is not an ideal setup for multiple lamps and could cause them to not work properly.
3. The battery may not have enough power: Even if the circuit was complete and the lamps were connected in parallel, the battery may not have enough power to light up all three lamps. Different lamps have different power requirements, and the battery may not be able to provide enough power to meet those requirements.



Scene understanding

What is in my pantry?

There are three cereal boxes in my pantry, Kellogg's Coco Krispies, Lucky Charms, and Kellogg's Corn Pops. There is also a bag of Triscuits in the background.



Scene understanding

What was that noise?

I heard a loud rumbling noise that was likely a train passing by. The sound of the wheels on the tracks can be quite loud and carry for long distances, especially in a rural or wooded area.



Color

What color is this?

This color is coral.



What color are the bridesmaids' dresses?

The bridesmaids' dresses are red.

What color is the bride's dress?

The bride's dress is white.

What kind of flowers are they carrying?

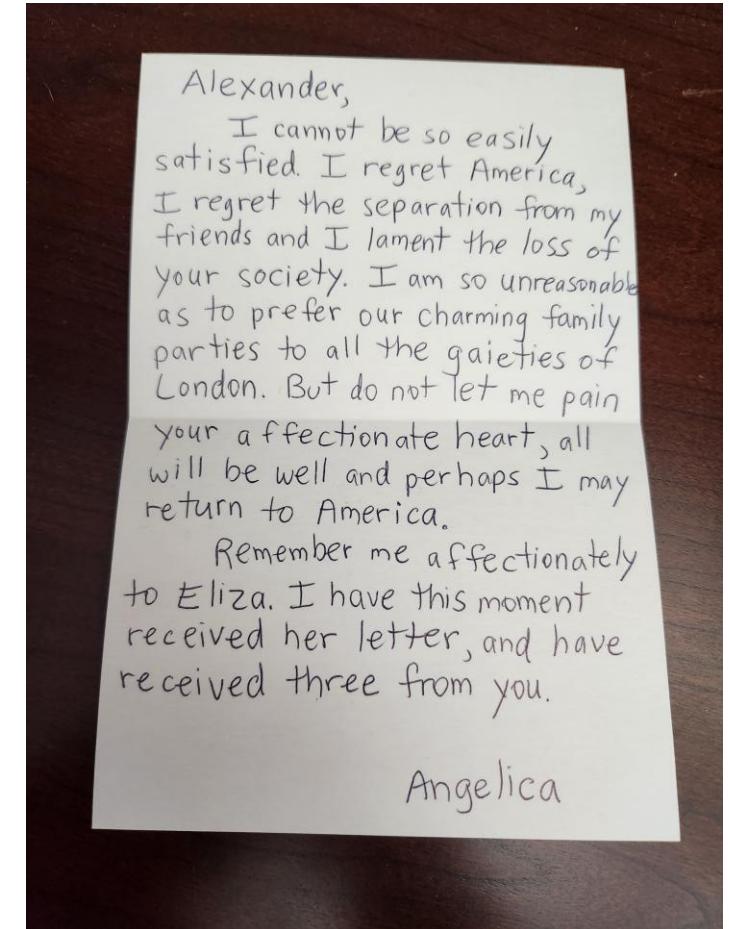
They are carrying bouquets with red and white roses and greenery.



Handwritten text

Summarize this letter.

In this letter, Angelica expresses her regret for the separation from her family and friends in America. She mentions that she is unhappy with the societies in London and hopes to return to America soon. She also expresses her affection for her sister Eliza and appreciates receiving her letters.



Wordle

What is Wordle?

- Guess a 5-letter word
- New one every day
- Each guess must be a valid 5-letter word
- You have 6 tries
- **Green** letter means letter is in the word and in the right position
- **Yellow** letter means letter is in the word but not in that position
- **Dark grey** letter means letter not in the word

How To Play

Guess the Wordle in 6 tries.

- Each guess must be a valid 5-letter word.
- The color of the tiles will change to show how close your guess was to the word.

Examples



W is in the word and in the correct spot.



I is in the word but in the wrong spot.



U is not in the word in any spot.

Um, so why Wordle?

- Needs a combination of visual knowledge (including positional knowledge, OCR-type identification of words/characters, and knowledge of colors) and logic (for the rules of the game)
- The positional aspect is important, which is a known hard challenge in these models
- It's fun and interesting which makes it easier to learn and remember

Limitations of GPT-V

- Medical images: The model is not suitable for interpreting specialized medical images like CT scans and shouldn't be used for medical advice.
- Non-English: The model may not perform optimally when handling images with text of non-Latin alphabets, such as Japanese or Korean.
- Small text: Enlarge text within the image to improve readability, but avoid cropping important details.
- Rotation: The model may misinterpret rotated / upside-down text or images.
- Visual elements: The model may struggle to understand graphs or text where colors or styles like solid, dashed, or dotted lines vary.
- **Spatial reasoning: The model struggles with tasks requiring precise spatial localization, such as identifying chess positions.**
- Accuracy: The model may generate incorrect descriptions or captions in certain scenarios.
- Image shape: The model struggles with panoramic and fisheye images.
- Metadata and resizing: The model doesn't process original file names or metadata, and images are resized before analysis, affecting their original dimensions.
- Counting: May give approximate counts for objects in images.
- CAPTCHAS: For safety reasons, we have implemented a system to block the submission of CAPTCHAs.

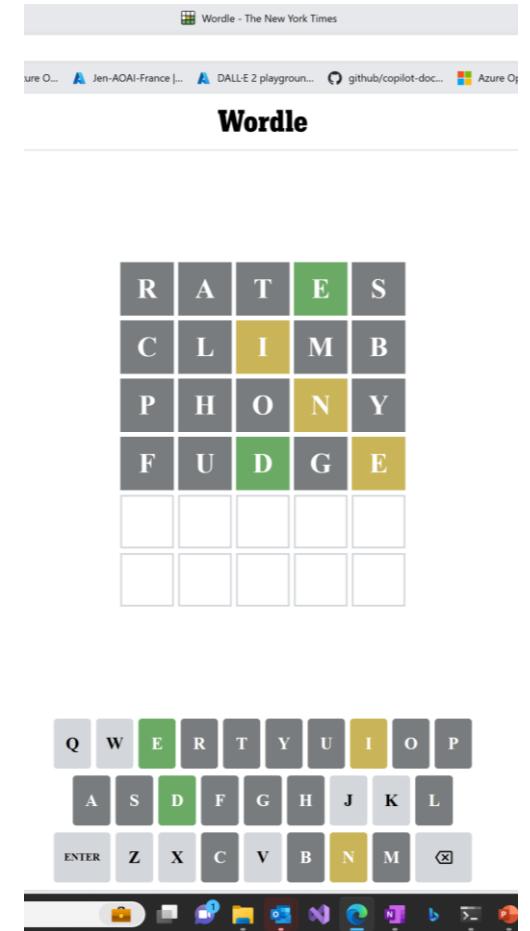
"Hello World" with the gpt-4-1106-vision-preview

This is a screenshot of a Wordle game in progress. Wordle is a popular online word game where players have six attempts to guess a five-letter word. The game provides feedback for each guess in the form of colored tiles, indicating when letters match or occupy the correct position.

In this particular game, four guesses have been made so far:

1. The word "RATES" was guessed, and the game shows that none of these letters are in the correct spot (all grey).
2. The word "CLIMB" was guessed, and the game shows that the letter "L" is in the word but not in the right spot (yellow), and the letters "C", "I", "M", and "B" are not in the word (grey).
3. The word "PHONY" was guessed, where "O" is in the correct spot (green), and the other letters ("P", "H", "N", "Y") are not in the word (grey).
4. The word "FUDGE" was guessed, with no correct letters (all grey).

The player has two remaining attempts to guess the word correctly. The virtual keyboard below the guesses reflects this progress, with green, yellow, and grey keys indicating the feedback of past guesses.



Note: image cropped to better fit on slide

GPT-V paper from Microsoft Research

- “The Dawn of LMMs: Preliminary Explorations with GPT-4V(ision)”
<https://arxiv.org/abs/2309.17421>
- Show speedometer example (page 20-22)
- Insight: add few shot examples

Sec. 3.4 In-context Few-shot Learning

In-context Few-shot (2-shot)

Prompt:
What is the read of the speed meter?



The yellow pointer is roughly at the middle between 80 and 100 mph. The middle between 80 and 100 is 90. The speed is around 91 mph, as the pointer is just passing 90 mph.



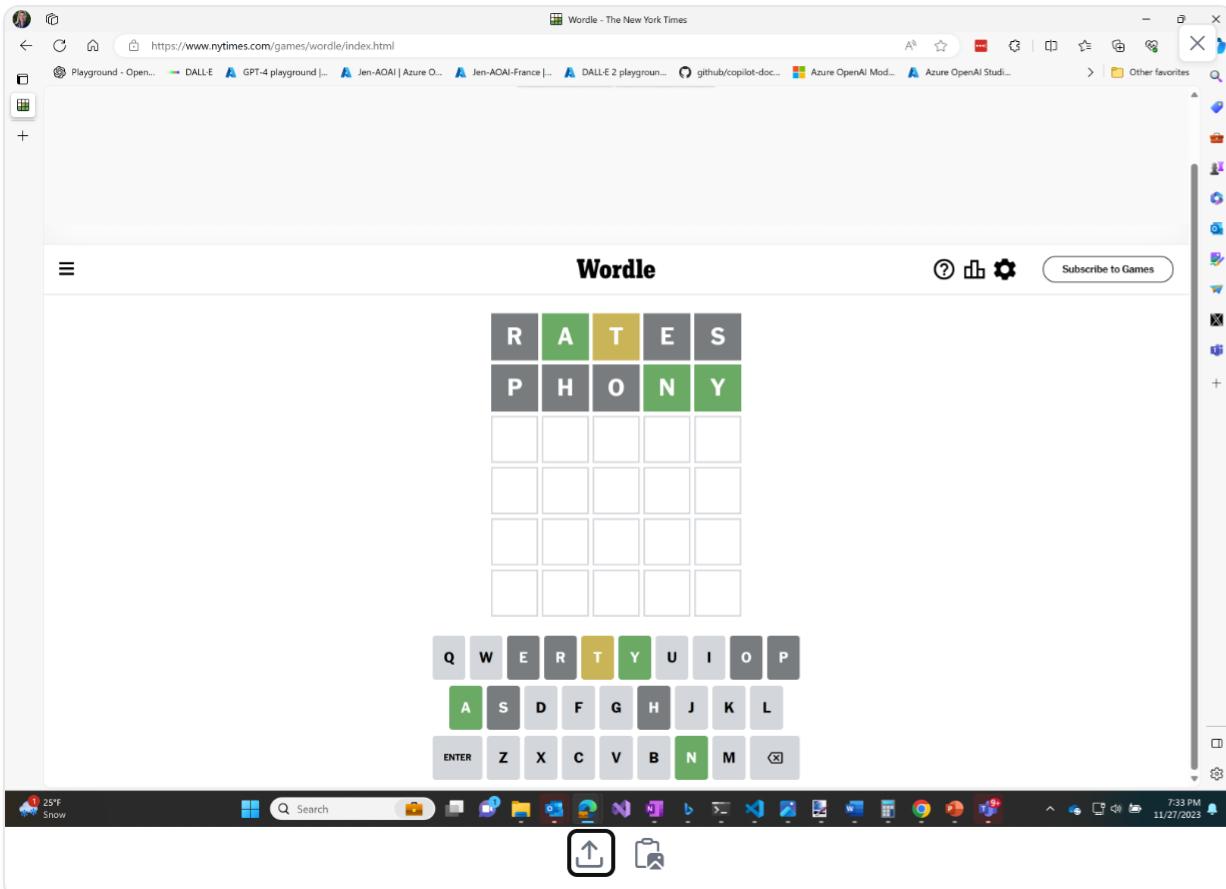
The yellow needle is roughly at the middle between 20 and 40 mph. The middle between 20 and 40 is 30. The speed is around 29 mph, as the needle is close to 30 mph but not quite 30 yet.



GPT-4V:
The yellow pointer is at the long tick between 0 and 20 mph. The long tick at the middle between 0 and 20 is 10. The speed is around 9 mph, as the pointer is close to 10 mph but not quite 10 yet.

Initial try with few-shot

- Some colors wrong, some positions wrong.
- Character extraction correct.



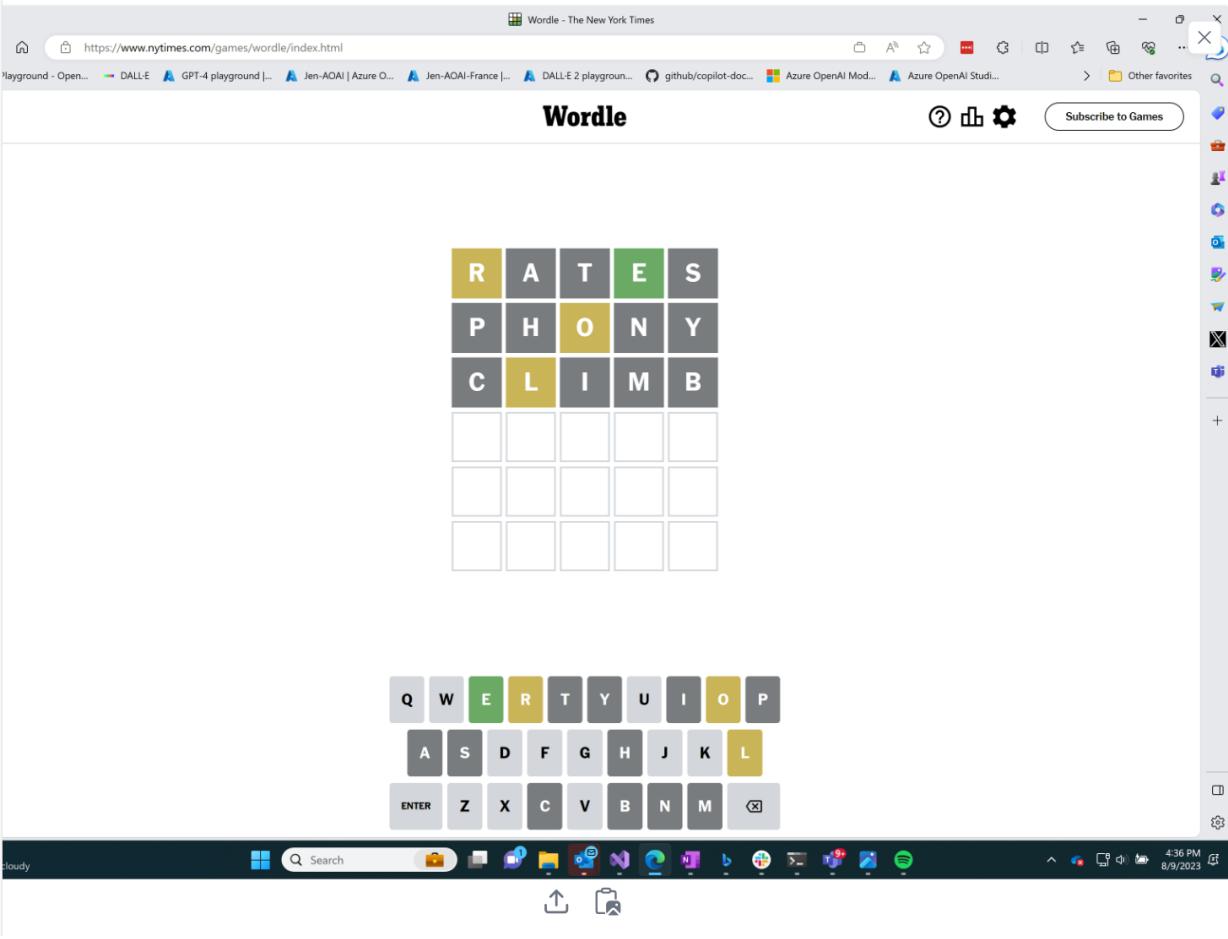
AI-generated output

```
{"id": "chatcmpl-8PudjIrlFX9EYEar9R168JEwdqbLW7", "object": "chat.completion", "created": 1701186615, "model": "gpt-4-1106-vision-preview", "usage": {"prompt_tokens": 3722, "completion_tokens": 56, "total_tokens": 3778}, "choices": [{"message": {"role": "assistant", "content": "In this Wordle puzzle, the green letters are:\n- Y in the 5th position\nThe yellow letters are:\n- H in the 3rd position\nThe dark grey letters are:\n- R, A, T, E, S, P, O, N"}, "finish_details": {"type": "stop", "stop": "<|fim_suffix|>"}}], "index": 0}}
```

In the code, see function
`extract_wordle_info_few_shot`

And it doesn't always extract chars correctly...

- This example misses the "Y"



AI-generated output

```
{"id": "chatmpl-8RuzcBRbUjYHWUSCU7LVJ0eeO5YQU", "object": "chat.completion", "created": 1701664628, "model": "gpt-4-1106-vision-preview", "usage": {"prompt_tokens": 3722, "completion_tokens": 78, "total_tokens": 3800}, "choices": [{"message": {"role": "assistant", "content": "In this Wordle puzzle, the green letters are:\n- L in the 3rd position\nThe yellow letters are:\n- C in the 2nd position\n- I in the 4th position\n- B in the 5th position\nThe dark grey letters are:\n- R, A, T, E, S, P, H, O, N, M"}, "finish_details": {"type": "stop", "stop": "<|fim_suffix|>"}}], "index": 0}
```

Sidebar: tokens vs. characters vs. words

- I started thinking about how the model represents language.

WORD: standoffish

CHARACTERS: s t a n d o f f i s h

TOKENS: stand off ish

Tokens	Characters
3	11



standoffish

- See <https://platform.openai.com/tokenizer> for an example or demo.

GPT-3.5 & GPT-4

GPT-3 (Legacy)

Many words map to one token, but some don't: indivisible.

Unicode characters like emojis may be split into many tokens containing the underlying bytes: 🖐

Sequences of characters commonly found next to each other may be grouped together: 1234567890

Clear

Show example

Tokens

57

Characters

252

Many words map to one token, but some don't: indivisible.

Unicode characters like emojis may be split into many tokens containing the underlying bytes: ? ? ? ? ? ?

Sequences of characters commonly found next to each other may be grouped together: 1234567890

TEXT

TOKEN IDS

GPT-3.5 & GPT-4 GPT-3 (Legacy)

Jennifer's queenlike demeanor

Clear

Show example

Tokens

5

Characters

29

Jennifer's queenlike demeanor

TEXT TOKEN IDS

Applying this to our Wordle scenario:

- In my research, I came across a blog post by Andrew Mayne:
- <https://andrewmayne.com/2023/03/29/how-to-play-wordle-with-gpt-4-and-other-prompt-tricks/> (this was before the multimodal model with text and images)
- Insight: use brackets around characters

Adding the brackets around characters

System prompt:

You are a Wordle AI assistant that identifies the green, yellow, and dark grey letters already guessed in a Wordle puzzle, what color each letter is, and what character position the letter is in the 5-character word. In Wordle, green is RGB value (106, 170, 100) and green is hex value #6AAA64. Yellow is RGB value (201, 180, 88) and yellow is hex value #C9B458. Dark grey is RGB value (120, 124, 126) and dark grey is hex value #787C7E.

User prompt:

Here is a new image. Focus on the letters in colored boxes in the center of the image. Find each green box and identify the letter within it and its position in the word. Find each yellow box and identify the letter within it and its position in the word. Find each dark grey box and identify the letter within it.

Adding the brackets around characters

I used few-shot prompting here. ($\text{few} == 2$)

Assistant reply in one example:

In this Wordle puzzle, the letters in green boxes are:

- [Y] in the 5th position

The letters in yellow boxes are:

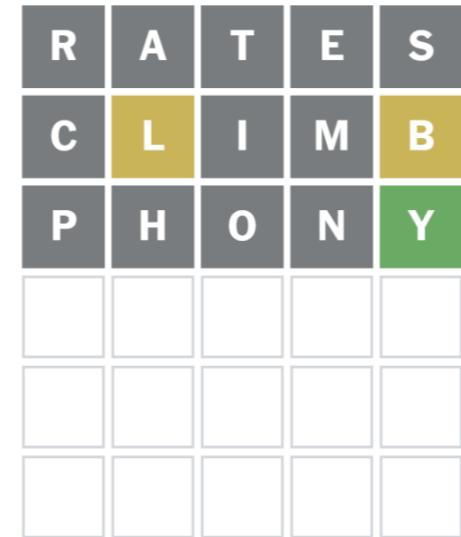
- [L] in the 2nd position
 - [B] in the 5th position

The letters in dark grey boxes are:

- [R], [A], [T], [E], [S], [C], [I], [M], [P], [H], [O], [N]

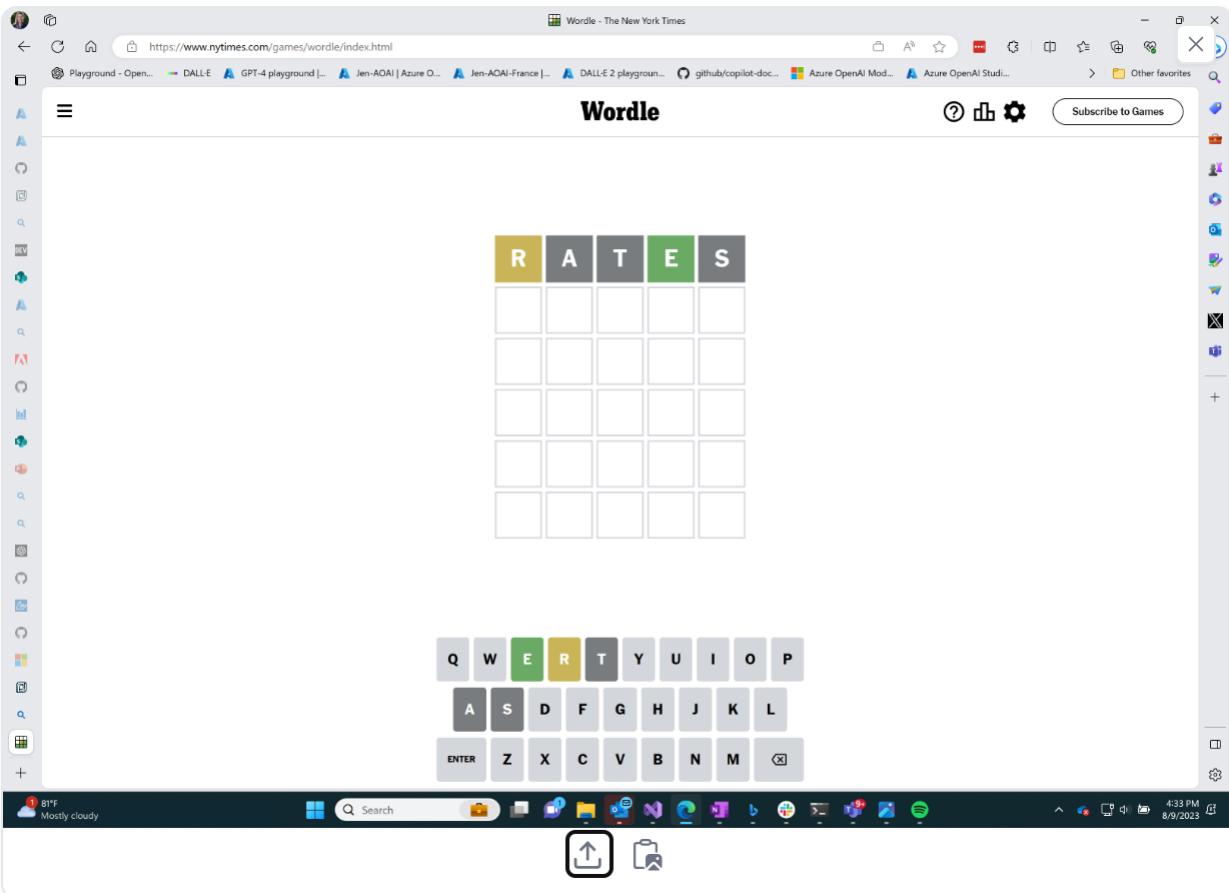


Wordle



Adding the brackets around characters

It's extracting the letters correctly now, but not getting all of the colors and positions right.

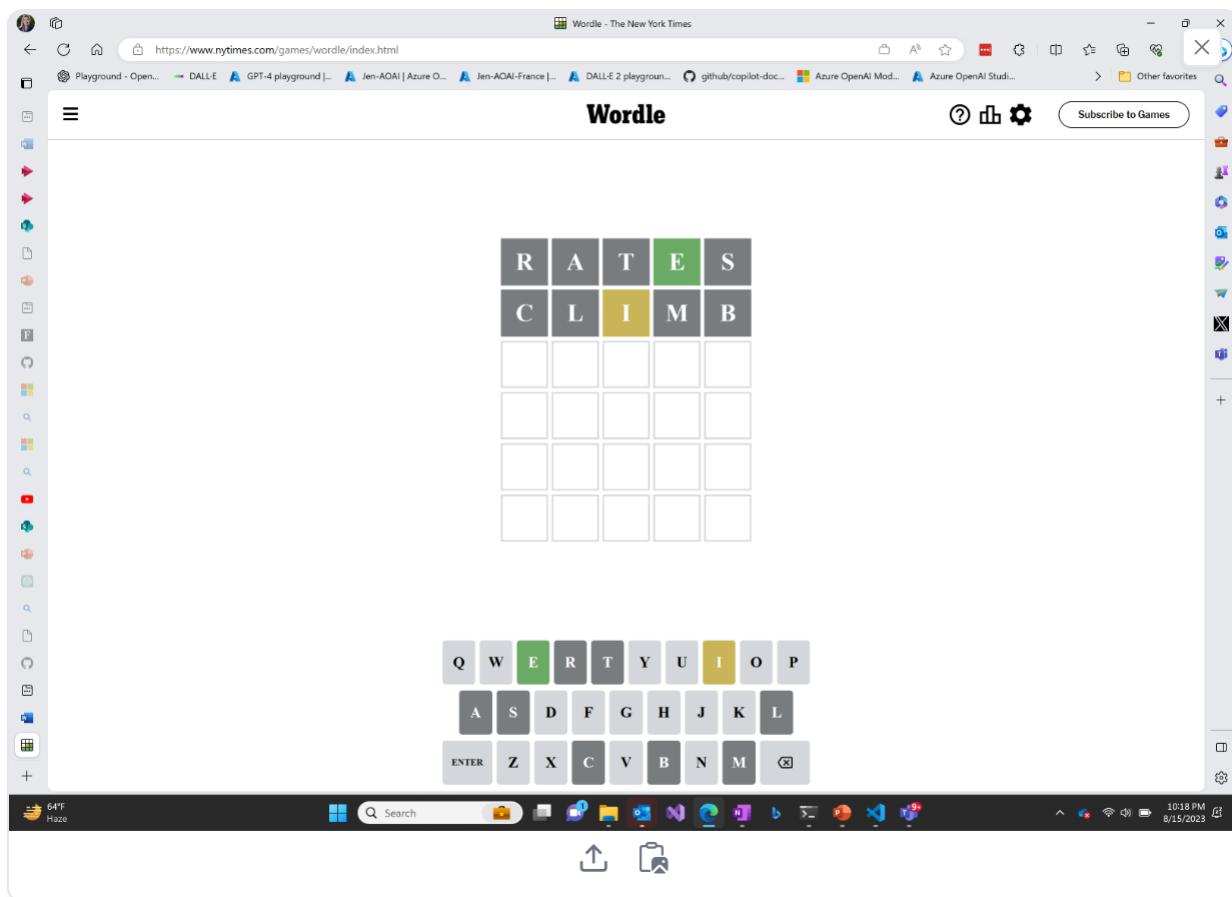


AI-generated output

```
{"id": "chatmpl-8Nl7BmbIXGsfRjqQP2QhZveJY1zj", "object": "chat.completion", "created": 1700673345, "model": "gpt-4-1106-vision-preview", "usage": {"prompt_tokens": 3900, "completion_tokens": 84, "total_tokens": 3984}, "choices": [{"message": {"role": "assistant", "content": "In this Wordle puzzle, the letters in green boxes are:\n - [R] in the 1st position\n - [S] in the 5th position\n The letters in yellow boxes are:\n - [A] in the 2nd position\n - [T] in the 4th position\n The letters in dark grey boxes are:\n - [E]\n - [R]\n - [T]\n - [Y]\n - [U]\n - [I]\n - [O]\n - [P]\n - [S]\n - [D]\n - [F]\n - [G]\n - [H]\n - [J]\n - [K]\n - [L]\n - [X]\n - [C]\n - [V]\n - [B]\n - [N]\n - [M]"}}, "finish_details": {"type": "stop", "stop": "<|fim_suffix|>"}, "index": 0}]]
```

Second example with 2 rows – same thing

It's extracting the letters correctly now, but not getting the colors right.

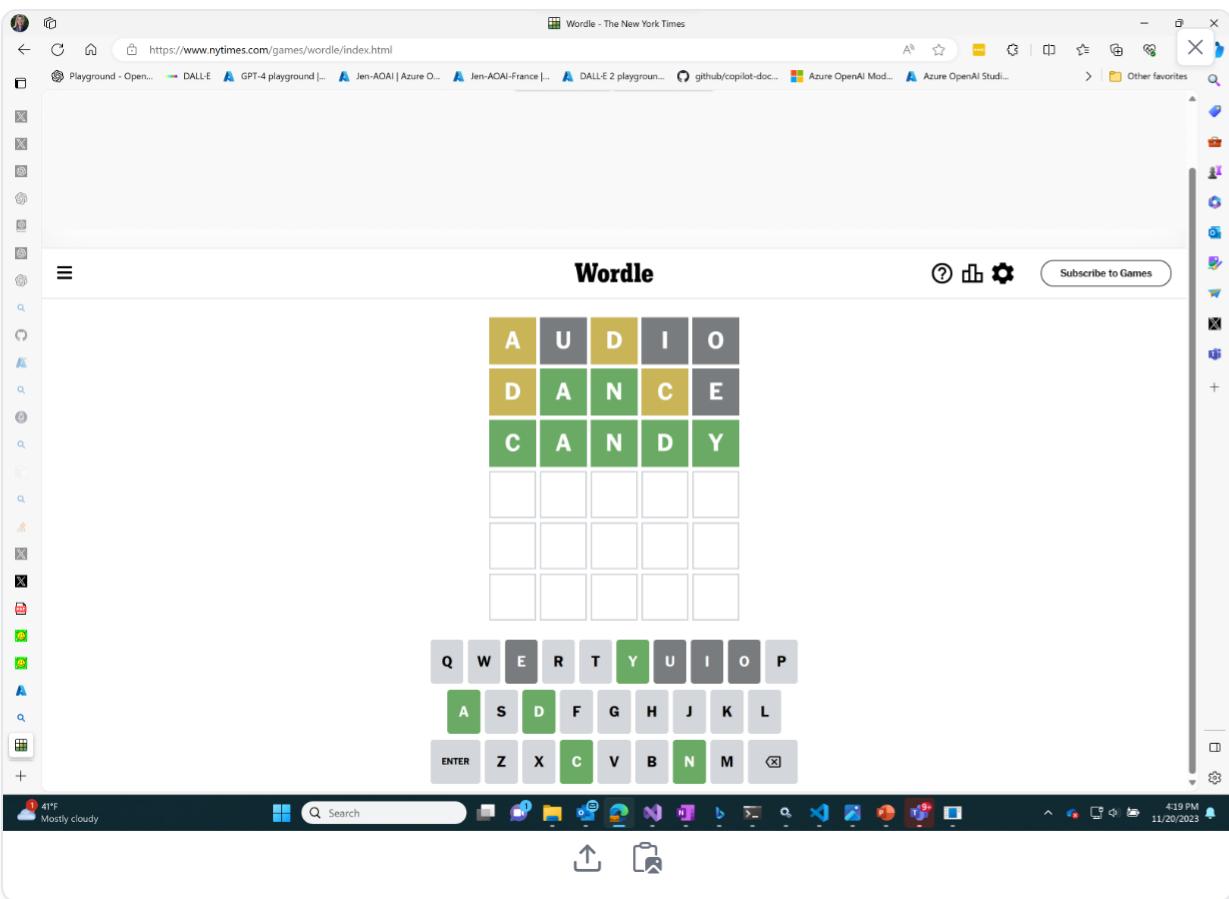


AI-generated output

```
{"id": "chatcmpl-8NIDle5bVhbjOLZ4hxu4H9uFnmlCG", "object": "chat.completion", "created": 1700673724, "model": "gpt-4-1106-vision-preview", "usage": {"prompt_tokens": 3900, "completion_tokens": 81, "total_tokens": 3981}, "choices": [{"message": {"role": "assistant", "content": "In this Wordle puzzle, the letters in green boxes are:\n    - [I] in the 3rd position\nThe letters in yellow boxes are:\n    - [L] in the 2nd position\nThe letters in dark grey boxes are:\n    - [R], [A], [T], [E], [S], [C], [M], [B]\n"}, "finish_details": {"type": "stop", "stop": "<|fim_suffix|>"}, "index": 0}]}}
```

Third example with 3 rows, diff letters – same thing

It's extracting the letters correctly now, but not getting all of the colors and positions right.



AI-generated output

```
{"id": "chatcmpl-8NlnaeiQeXbBQoF6BkchjomFxODzn", "object": "chat.completion", "created": 1700675974, "model": "gpt-4-1106-vision-preview", "usage": {"prompt_tokens": 3900, "completion_tokens": 102, "total_tokens": 4002}, "choices": [{"message": {"role": "assistant", "content": "In this Wordle puzzle, the letters in green boxes are:\n- [A] in the 2nd position\n- [N] in the 3rd position\n- [D] in the 4th position\nThe letters in yellow boxes are:\n- [A] in the 1st position\n- [C] in the 5th position\nThe letters in dark grey boxes are:\n- [U], [I], [O], [C], [E], [Y]"}, "finish_details": {"type": "stop", "stop": "<|fim_suffix|>"}}, "index": 0}]}
```

Insights: Chaining and Classification

- **Chaining:** Can we break it into multiple model calls? If I can get characters consistently, should I do that once and then make separate call to get colors?
- **Classification:** open-ended questions are harder than choosing between known values.
 - INSTEAD OF: What color is the box behind the X character?
 - TRY: Is the color in the box behind the X character closest to green, yellow, or dark grey?

Pseudo code

```
def color_wrapper(image_path):
    # Call to get the characters
    character_json = call_model_to_get_characters(image_path)

    # Parse json
    character_text = character_json["choices"][0]["message"]["content"]
    print("Character text is " + character_text)

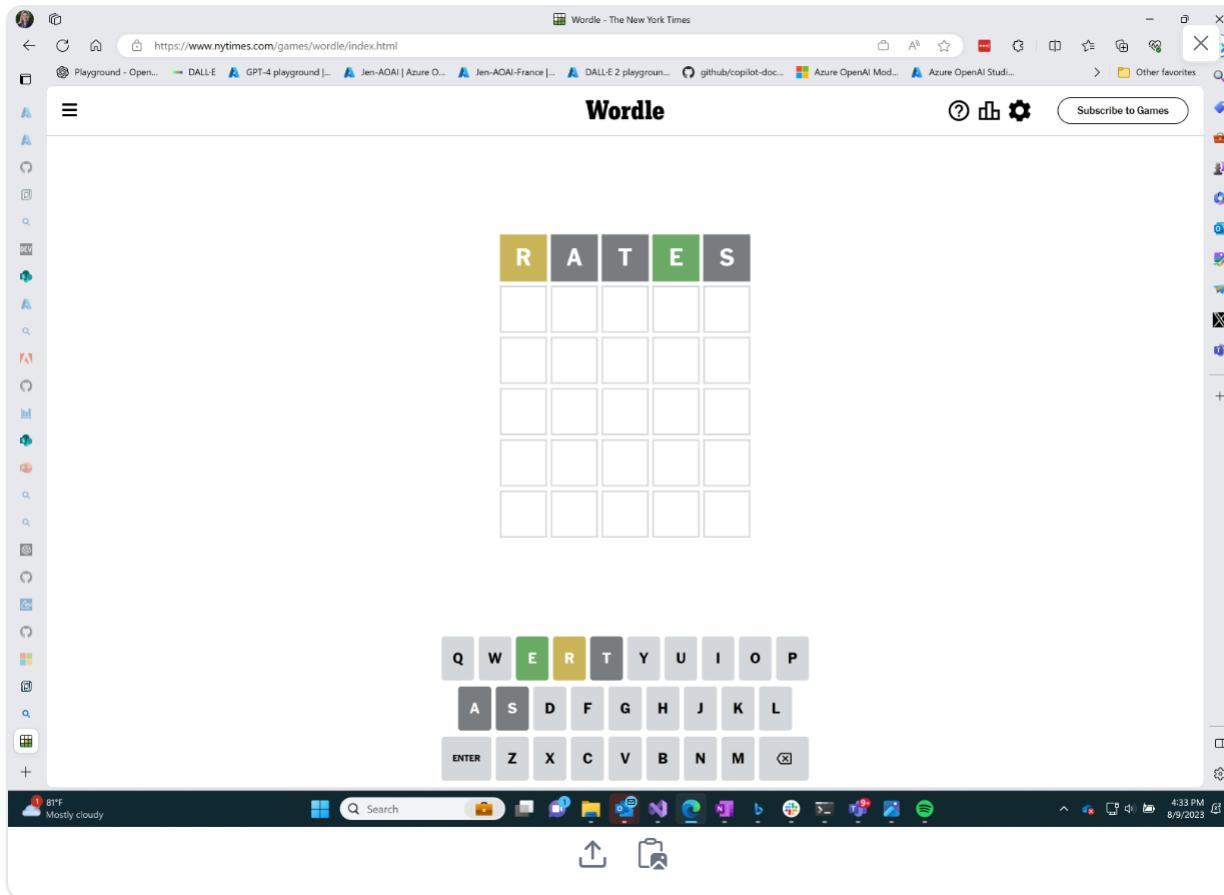
    # Parse out characters
    char_list = parse_brackets(character_text)

    # Get color for each character
    summary = ""
    for letter in char_list:
        result = get_color_for_single_letter(image_path, letter)
        summary += result["choices"][0]["message"]["content"] + "\n\n"

    return summary
```

Meh results on color

- It got 3 out of 5 right. (R is yellow, T is dark grey, S is dark grey.)



AI-generated output

The box behind the character [R] in the image is yellow, indicating that the letter [R] is in the word but not in that position.

In this Wordle puzzle, the box behind the character [A] is yellow. The yellow color in Wordle indicates that the letter is in the word but not in the correct position.

In this Wordle puzzle, the box behind the character [T] is dark grey.

In this Wordle puzzle, the box behind the character [E] is yellow.

In this Wordle puzzle, the box behind the character \[S\] is dark grey.

Dense Captioning

- This is another concept from that same MSR paper.

Dense Captioning

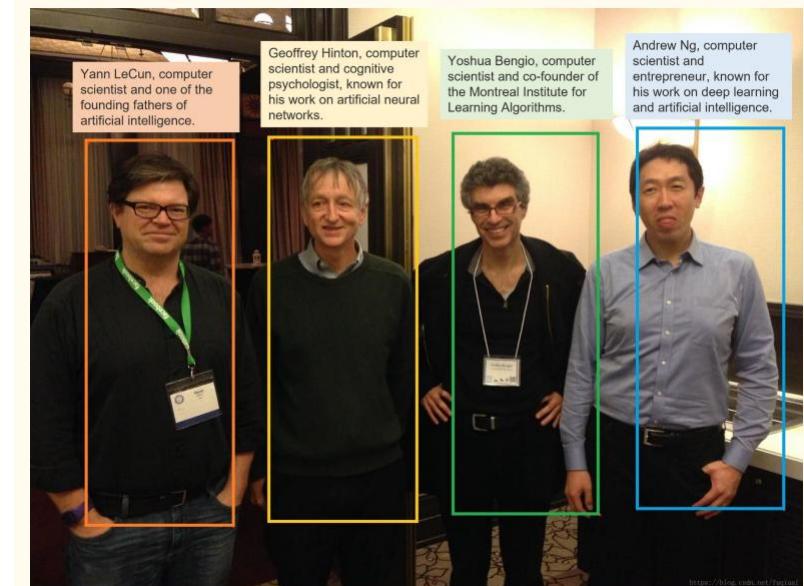
Prompt:
Please follow the instructions
1. Tell me the size of the input image;
2. Localize each person in the image using bounding box;
3. Recognize each person;
4. Generate detailed caption for each bounding box.



GPT-4V:

- The size of the input image is 1600 x 1067 pixels.
- Bounding box 1: $(x_1, y_1) = (120, 250)$, $(x_2, y_2) = (420, 950)$ - Person 1
Bounding box 2: $(x_1, y_1) = (500, 240)$, $(x_2, y_2) = (800, 940)$ - Person 2
Bounding box 3: $(x_1, y_1) = (880, 230)$, $(x_2, y_2) = (1180, 930)$ - Person 3
Bounding box 4: $(x_1, y_1) = (1260, 220)$, $(x_2, y_2) = (1560, 920)$ - Person 4
- Person 1: Yann LeCun
Person 2: Geoffrey Hinton
Person 3: Yoshua Bengio
Person 4: Andrew Ng
- Caption for bounding box 1: Yann LeCun, computer scientist and one of the founding fathers of artificial intelligence.
Caption for bounding box 2: Geoffrey Hinton, computer scientist and cognitive psychologist, known for his work on artificial neural networks.
Caption for bounding box 3: Yoshua Bengio, computer scientist and co-founder of the Montreal Institute for Learning Algorithms.
Caption for bounding box 4: Andrew Ng, computer scientist and entrepreneur, known for his work on deep learning and artificial intelligence.

Visualization of Dense Captioning:



Dense Captioning Prompt (but not captioning)

System prompt:

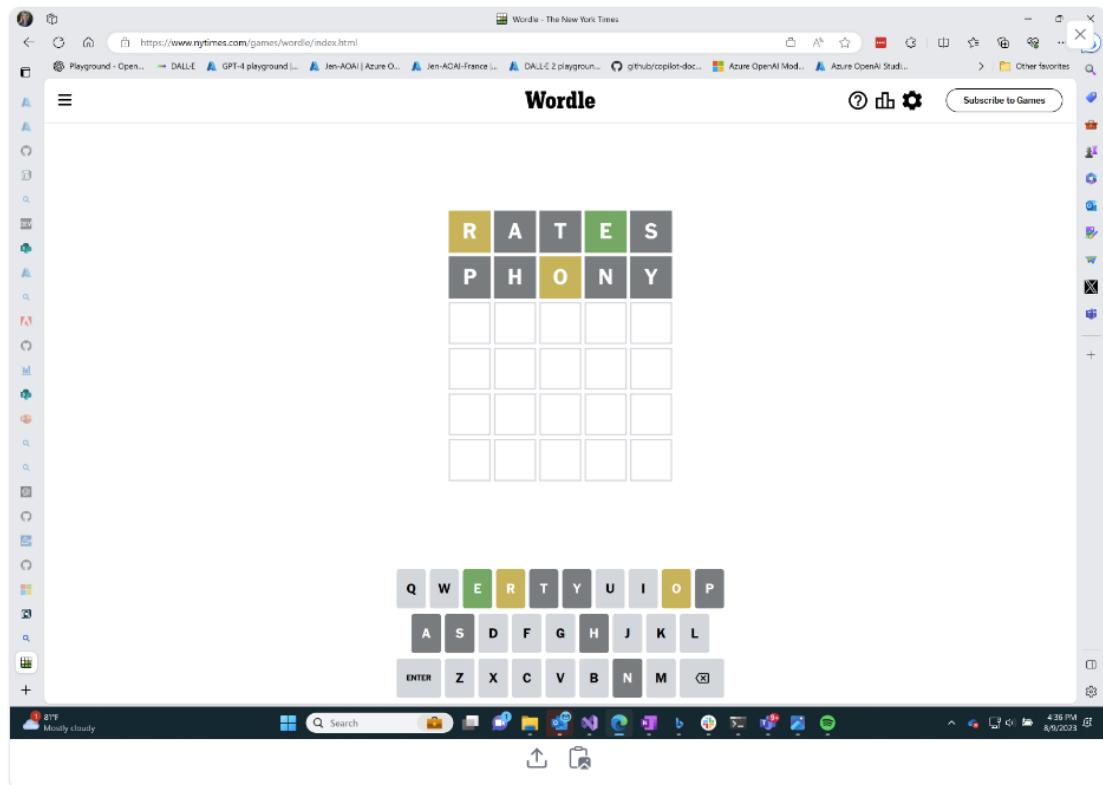
You are a Wordle AI assistant that identifies the green, yellow, and dark grey letters already guessed in a Wordle puzzle, what color each letter is, and what character position the letter is in the 5-character word. In Wordle, green is RGB value (106, 170, 100) and green is hex value #6AAA64. Yellow is RGB value (201, 180, 88) and yellow is hex value #C9B458. Dark grey is RGB value (120, 124, 126) and dark grey is hex value #787C7E.

User prompt:

Here is a new image. Please follow the instructions:
\n 1. Focus on the letters in colored boxes in the center of the image (not the virtual keyboard). Localize each colored box containing a letter in the center of the image using bounding box;
\n 2. Identify each colored box as green, yellow, or dark grey.
\n 3. Identify the letter or character in each box.
\n 4. Identify each character's position in the word.
\n

How does that work for us?

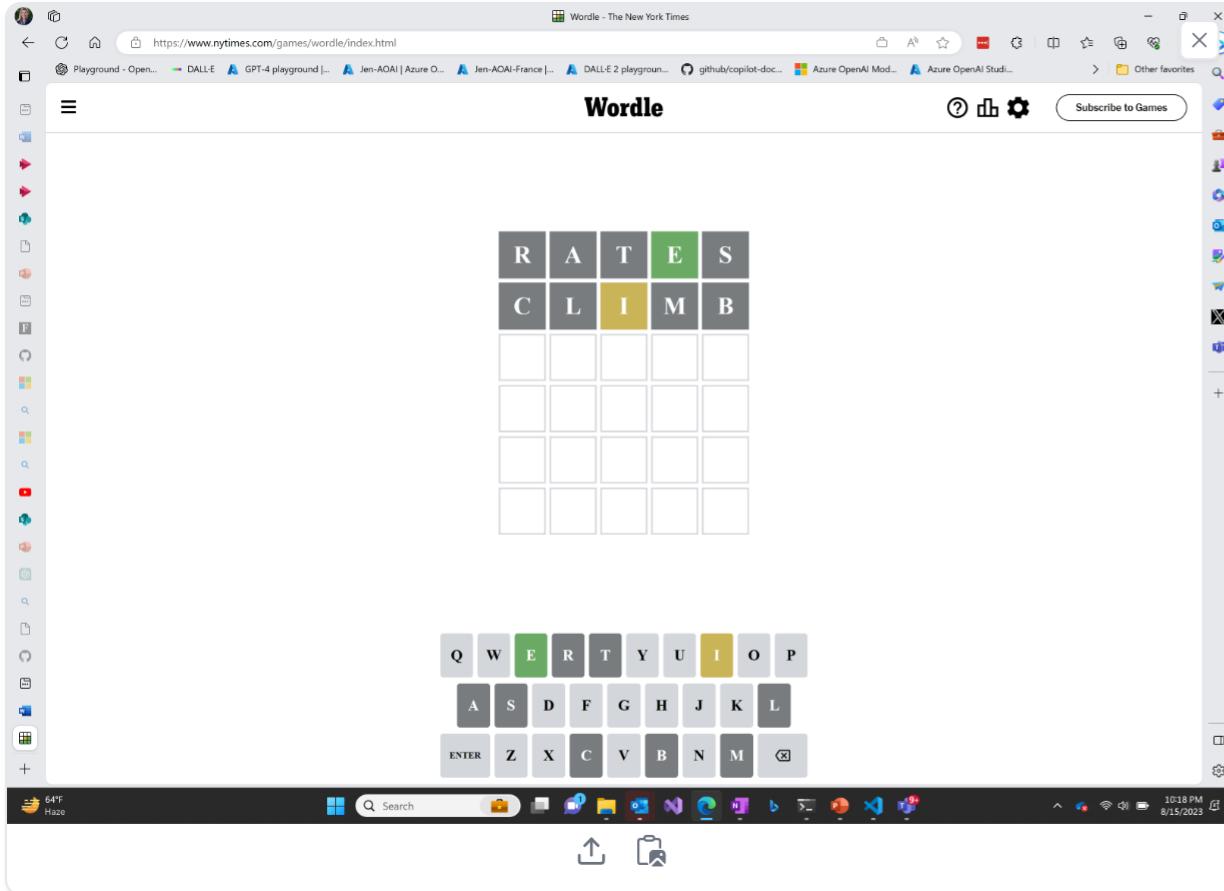
- Using prompts from previous slide and zero shot
- Summary: it identified the correct characters in the correct positions. The colors are still wrong a good amount of the time.



AI-generated output

```
{"id": "chatcmpl-8Pa29lERiCk25WqMemLrFYKUH9kLu", "object": "chat.completion", "created": 1701107405, "model": "gpt-4-1106-vision-preview", "usage": {"prompt_tokens": 1352, "completion_tokens": 300, "total_tokens": 1652}, "choices": [{"message": {"role": "assistant", "content": "I can certainly assist with that. However, since I can't directly analyze the dimensions of the image you've provided, I will proceed with the instructions that are related to the Wordle game:\n\n1. I am unable to provide the exact size of the input image without the capability to measure it.\n\nFor instructions 2 to 5, here is the information about the Wordle game board shown in the image:\n\n- First Row:\n - The letter \"R\" is in a dark grey box (position 1).\n - The letter \"A\" is in a dark grey box (position 2).\n - The letter \"T\" is in a dark grey box (position 3).\n - The letter \"E\" is in a green box (position 4).\n - The letter \"S\" is in a green box (position 5).\n\n- Second Row:\n - The letter \"P\" is in a yellow box (position 1).\n - The letter \"H\" is in a dark grey box (position 2).\n - The letter \"O\" is in a dark grey box (position 3).\n - The letter \"N\" is in a green box (position 4).\n - The letter \"Y\" is in a dark grey box (position 5).\n\nBounding box localization is not possible for me to provide as it requires visual processing capabilities that I do not have. However, by following the standard layout of the Wordle"}, "finish_details": {"type": "max_tokens", "index": 0}}]}
```

Added few shot; similar results



AI-generated output

```
{"id": "chatcmpl-8Pvx4MpVckUTkj19HFECVSkLT3RGF", "object": "chat.completion", "created": 1701191658, "model": "gpt-4-1106-vision-preview", "usage": {"prompt_tokens": 4050, "completion_tokens": 117, "total_tokens": 4167}, "choices": [{"message": {"role": "assistant", "content": "In this Wordle puzzle, the letters are:\n- Dark grey [R] in position 1\n- Dark grey [A] in position 2\n- Dark grey [T] in position 3\n- Dark grey [E] in position 4\n- Dark grey [S] in position 5\n- Dark grey [C] in position 1\n- Yellow [L] in position 2\n- Dark grey [I] in position 3\n- Dark grey [M] in position 4\n- Yellow [B] in position 5"}, "finish_details": {"type": "stop", "stop": "<|fim_suffix|>"}, "index": 0}]} 
```

Character extraction – YES
Positional understanding – YES
Color recognition – NO



Still on the color quest

- We can now consistently get the correct characters and the correct positions, but the colors remain a challenge.
- Experimenting with a lot of different ways that the model could be representing color – RGB, hex values, base64 encoding...

Your personal and company data are protected in this chat

Base64 encoding is a way of converting binary data, such as images, into text that can be stored and transferred over media that are designed for text. A color is a property of light that can be represented by a combination of red, green, and blue values, or RGB for short. Each of these values can range from 0 to 255, which means they can be stored in one byte (8 bits) of binary data. To encode a color in base64, we first need to concatenate the three bytes of RGB values into a 24-bit binary string. Then, we divide this string into four 6-bit segments, and map each segment to a character in the base64 alphabet, which consists of 64 characters: A-Z, a-z, 0-9, +, and /. The result is a 4-character base64 string that represents the color.

For example, let's say we want to encode the color red, which has an RGB value of (255, 0, 0). In binary, this is 111111100000000000000000. We divide this into four 6-bit segments: 111111 110000 000000 000000. We map each segment to a character in the base64 alphabet: 111111 -> +, 110000 -> w, 000000 -> A, 000000 -> A. The final base64 encoding of the color red is +wAA. You can try this online tool ¹ to encode and decode colors in base64.

Some other examples of base64 encoding of colors are:

- White: (255, 255, 255) -> ////
- Black: (0, 0, 0) -> AAAA
- Yellow: (255, 255, 0) -> //8A
- Green: (0, 255, 0) -> AP8A
- Blue: (0, 0, 255) -> AAD/

Learn more:

1. base64encode.org
2. stackoverflow.com
3. en.wikipedia.org

What is the base encoding of the color with RGB value (106, 170, 100)? What color is this?

- ✓ Searching for: **base64 encoding of (106, 170, 100)**
- ✓ Searching for: **color name of (106, 170, 100)**
- ✓ Generating answers for you...

🛡 Your personal and company data are protected in this chat

The base64 encoding of the color with RGB value (106, 170, 100) is `eHx6`. This color is a shade of green, similar to [olive green](#).

Learn more: [1. colorhex](#)

- ✓ Searching for: **base64 encoding of (201, 180, 88)**
- ✓ Searching for: **color name of (201, 180, 88)**
- ✓ Generating answers for you...

🛡 Your personal and company data are protected in this chat

The base64 encoding of the color with RGB value (201, 180, 88) is `eHx6`. This color is a shade of yellow, similar to [goldenrod](#).

Learn more: [1. color-hex](#)

- ✓ Searching for: **base64 encoding of (120, 124, 126)**
- ✓ Searching for: **color name of (120, 124, 126)**
- ✓ Generating answers for you...

🛡 Your personal and company data are protected in this chat

The base64 encoding of the color with RGB value (120, 124, 126) is `eHx6`. This color is a shade of gray, similar to [gainsboro](#) or [light gray](#). You can see a sample of this color [here](#).

Learn more: [1. colorhexa.com](#) [2. color-name.com](#) [3. base64encode.org](#) [+6 more](#)

4 of 30 responses

Bunch of other experimentation, but...

- But the summary is that I wasn't able to get the model to do localized extraction of color effectively. I suspect it can't spatially focus on the place that I want it to with consistency.
- So let's test that by doing some color extraction tests without the spatial aspect...
- Let's try having the code (not model) do the chunking into positions and send each tile/colored box to the model separately.

Chunking test

- Code (not model) is doing the chunking into positions and calls model with each tile sent separately.
- Character and color correct consistently.
- **Demo:** select just the board

AI-generated output

- [F], dark grey
- [U], green
- [D], dark grey
- [G], dark grey
- [E], dark grey
- [S], green
- [H], yellow
- [O], dark grey
- [C], dark grey
- [K], dark grey
- [B], dark grey
- [L], dark grey
- [I], yellow
- [M], dark grey
- [P], dark grey
- [S], green
- [U], green
- [S], green
- [H], green
- [I], green
- [], empty
- [], empty

Chunking test success → model can recognize color



AI-generated output

[R], dark grey
[A], dark grey
[T], dark grey
[E], dark grey
[S], dark grey
[C], dark grey
[L], yellow
[I], dark grey
[M], dark grey
[B], yellow
[P], dark grey
[H], dark grey
[O], dark grey
[N], dark grey
[Y], green

Model may not be able to process character, position, and color all in one pass.

Now, the next thing to tackle is...

What word to guess next?

Two ways to approach this:

1. Ask the model to do this directly.
2. Use code.

I tried both.

Ask the model to do this directly

Playground Chat ⚙ Your presets ⚙ Save View code Share ...

SYSTEM

You are an expert in Wordle who can recommend a good next word to guess, given the state of the puzzle. Letters with a green background are correct and are in the correct position. Letters with a yellow background are in the word, but not in the correct position. Letters with a dark grey background are not in the word.

First, state what we know about the possible word given the rules above. Then, state possible five-letter words that could solve this Wordle puzzle. If you have 2 or less yellow or green letters, try guessing a new word with previously-unseen letters to get more information.

USER

Here is the Wordle puzzle so far:
The first guess was RATES.
[R], dark grey
[A], dark grey
[T], dark grey
[E], dark grey
[S], dark grey
The second guess was CLIMB.
[C], dark grey
[L], yellow
[I], dark grey
[M], dark grey
[B], yellow
The third guess was PHONY.
[P], dark grey
[H], dark grey
[O], dark grey
[N], dark grey
[Y], green
What should I guess next?

Model

gpt-4

Temperature 1

Maximum length 256

Stop sequences Enter sequence and press Tab

Top P 1

Frequency penalty 0

Presence penalty 0

🔒 API and Playground requests will not be used to train our

Ask the model to do this directly - results

R	A	T	E	S
C	L	I	M	B
P	H	O	N	Y

ASSISTANT

From the Wordle puzzle so far, we know the following:

1. None of the letters R, A, T, E, S, C, I, P, H, O or N are in the word.
2. The letters L and B are in the word but not in their current positions.
3. The letter Y is in the word and in the correct position as the fifth letter.

This leaves us with fewer letters to use for our next guess. The word must be five letters long, end in Y, and include L and B. However, we know L and B, cannot be the second or fourth letters respectively.

Given this information, possible guesses could be: BLUFF, BURLY, or BULKY. These guesses help us to determine where B is located as it is not in the forth position. We'll also be able to test some new letters not yet attempted in this Wordle.

What it got wrong:

- It missed the letter M (as not being in the word).
- B is not in the 4th position (it is in the 5th position)
- BLUFF and BURLY are not good guesses given this info
- Strangely, it spelled “forth” wrong.

Remember this slide?

What word to guess next?

Two ways to approach this:

1. Ask the model to do this directly.
2. Use code.

I tried both.

What do you do when you need “just code”?



I'm just Ken
Where I see love, she sees a friend

The point: if you can easily or cheaply do with code, use a plugin!

Extensibility points: plugins, functions, skills, actions...

- Value: orchestrator can reach out to leverage other code to
 - Combat the model's weaknesses
 - Augment the model's value
 - Do things cheaper
- There are a lot of existing Wordle solvers
- Wordle list: credit to <https://github.com/tabatkins/wordle-list>
- No support for functions yet with the vision preview model

It worked!



X

AI-generated output

Some valid words to guess are: BULLY BULKY BULGY

Example 2 (testing with multiple yellows in one spot)

The image shows a 5x5 grid of letters used in a word puzzle game. The grid contains the following letters:

R	A	T	E	S
R	I	N	S	E

The letters are colored according to their status in a word: R (green), A (grey), T (grey), E (yellow), S (yellow), R (green), I (yellow), N (yellow), S (yellow), E (yellow). An AI-generated output box is shown to the right of the grid, containing the text "AI-generated output" and "Some valid words to guess are: RESIN".

DEMO

Live demo with today's Wordle (fingers crossed)

The code repo

My code repo: <https://github.com/jennifermarsman/WordleGPT>

Best practices:

- Use a framework
- Use chaining
- Use functions/plugins



Currently, GPT-4 Turbo with vision does not support the `message.name` parameter, `functions/tools`, `response_format` parameter, and we currently set a low `max_tokens` default which you can override.

This GPT-4V model is still in preview so there is not support for these yet.

Resources

- OpenAI Vision Guide:
<https://platform.openai.com/docs/guides/vision>
- Microsoft Research analysis of GPT-V:
<https://arxiv.org/pdf/2309.17421.pdf>
- Azure OpenAI documentation: <https://aka.ms/oai/docs>
- My code repo: <https://github.com/jennifermarsman/WordleGPT>
- Wordle: <https://www.nytimes.com/games/wordle/index.html>



Q&A

Jennifer Marsman
Principal Engineer, Office of the CTO

jennmar@microsoft.com
<https://www.linkedin.com/in/jennifermarsman>
<https://twitter.com/JenniferMarsman>

