

ENVISIONING STORIES WITH GENERATIVE AI

MELISSA ROEMMLE

Open Teams Event, 9/18/2025



ABOUT ME

Research scientist exploring how to use
Generative AI to support human creativity

Part of the Midjourney Storytelling Lab,
developing and evaluating AI-augmented
storytelling tools



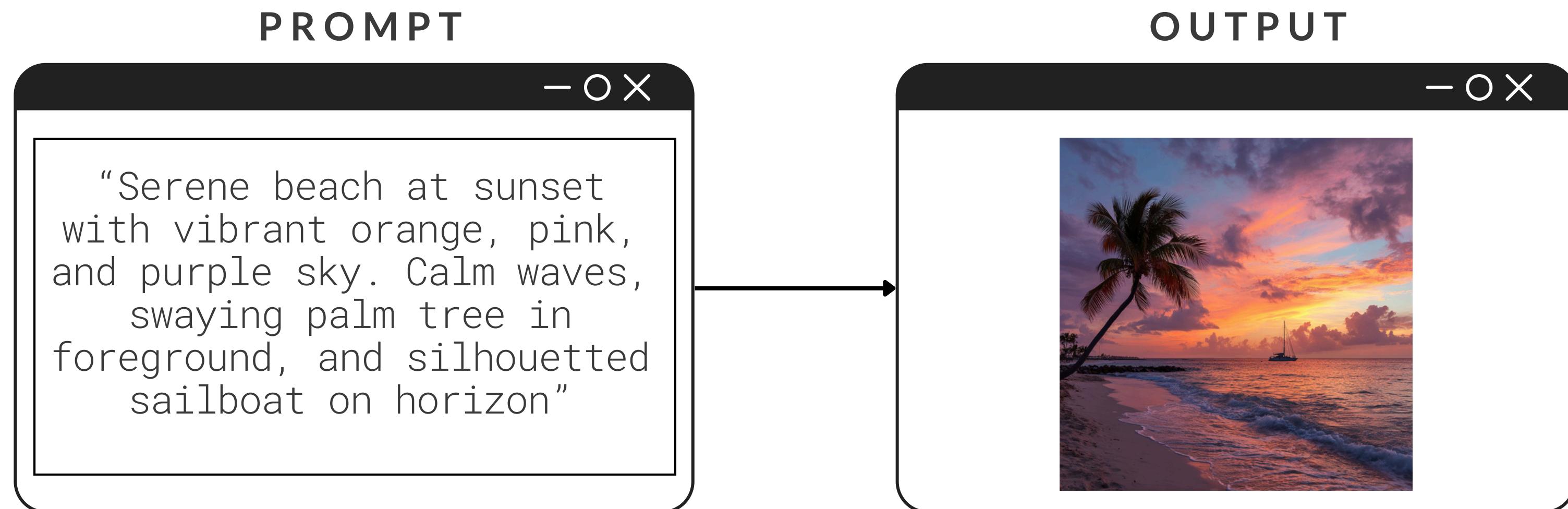
ABOUT ME

Research scientist exploring how to use
Generative AI to support human creativity

Part of the Midjourney Storytelling Lab,
developing and evaluating AI-augmented
storytelling tools

★ Focus of this talk: Generative AI models as
technologies, rather than specific *products*

TEXT-TO-IMAGE MODELS



Text-to-image
models use
natural
language as a
control panel
for creation



Text-to-image models are also a new way to explore language

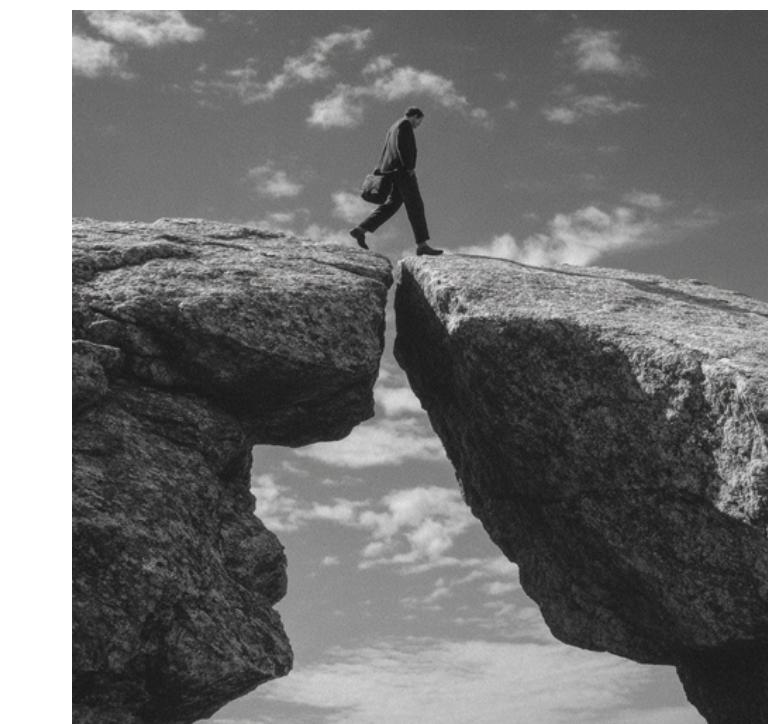
★ by observing concrete interpretations of abstract or ambiguous text, for example



“The color blue is chasing after the number seven”



“angry lamp”



“The advantage outweighs the risk”



“coastal nails”



“The dolphin conducts professionalism”



“mathematical orchards”



STORY VISUALIZATION

- The task of transforming a story from text into visual form
- Can provide a new perspective on the story

RESEARCH CHALLENGES OF STORY VISUALIZATION

- **Text-image alignment:** how to ensure the images suitably depict the story text
- **Image-image alignment:** how to ensure consistency between images belonging to the same story
 - e.g. consistent portrayal of characters, settings, and objects



RESEARCH CHALLENGES OF STORY VISUALIZATION

- **Text-image alignment:** how to ensure the images suitably depict the story text
- **Image-image alignment:** how to ensure consistency between images belonging to the same story
 - e.g. consistent portrayal of characters, settings, and objects





VISUAL E-READER

A speculative use case for story
visualization

DESIGN

STORY WITH
SELECTED FRAGMENT

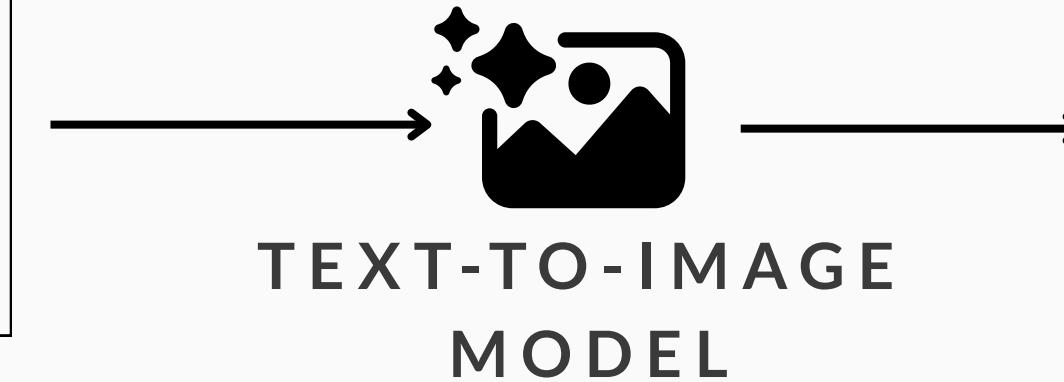
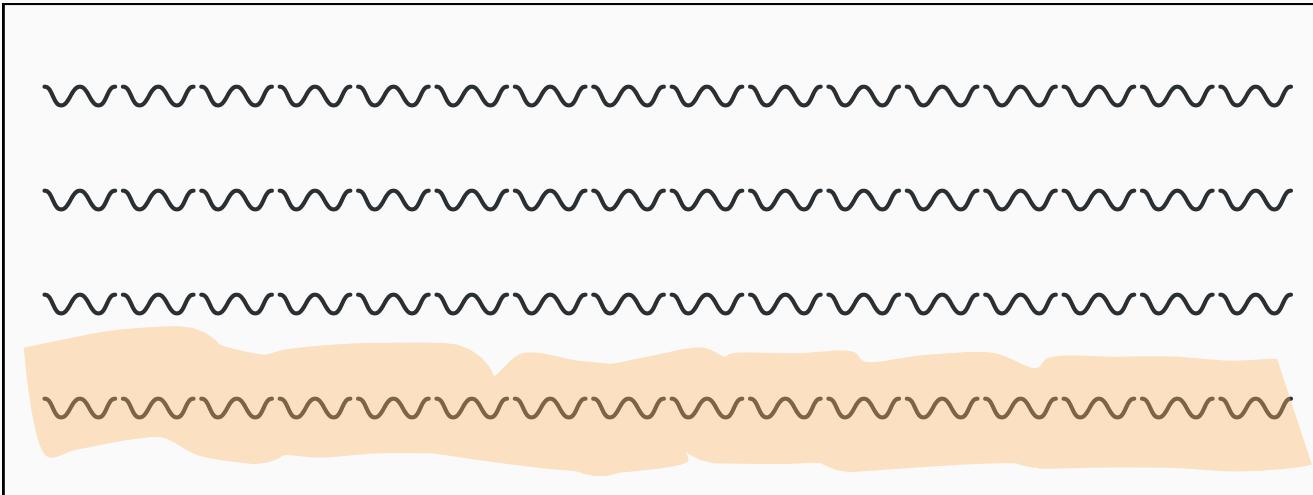
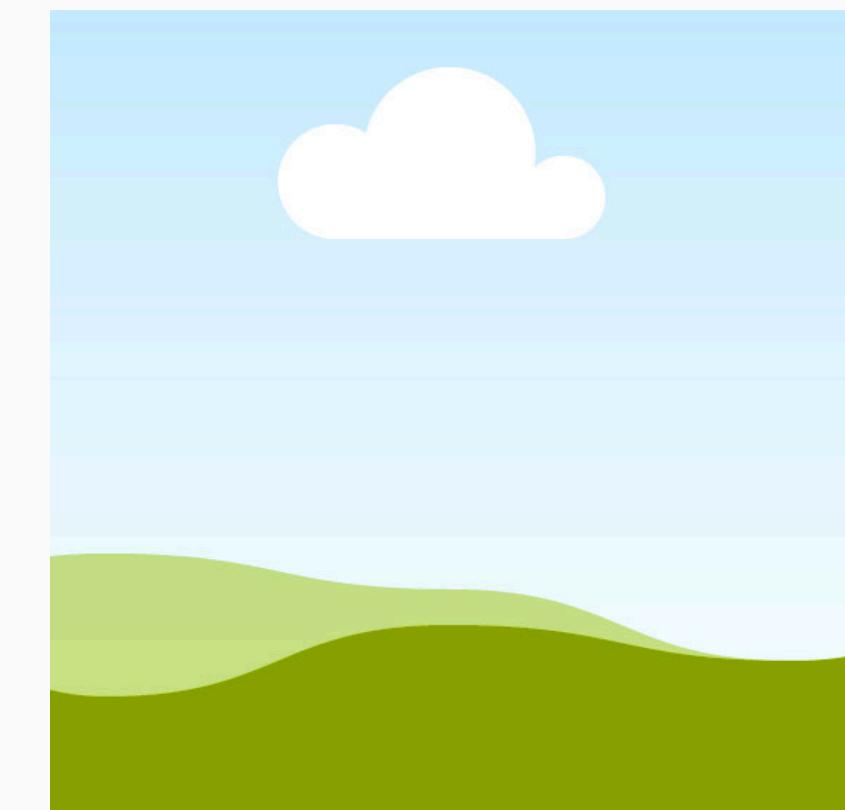
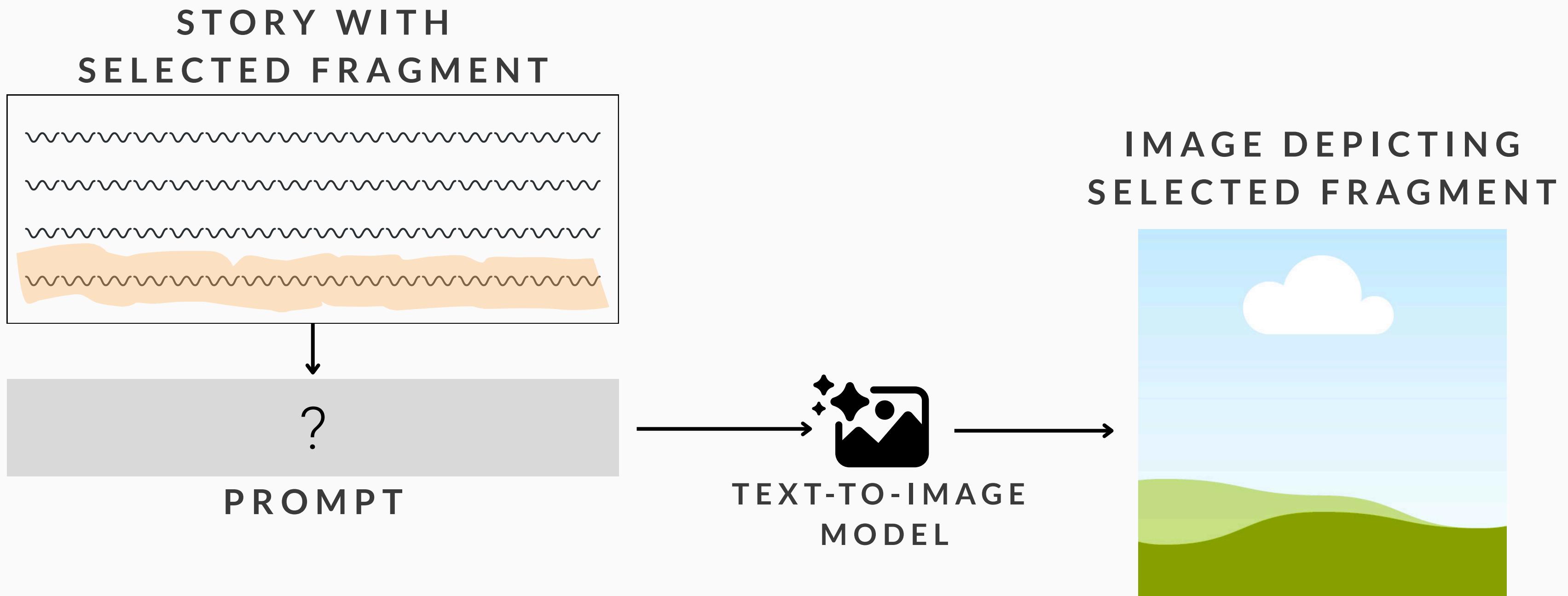


IMAGE DEPICTING
SELECTED FRAGMENT



DESIGN



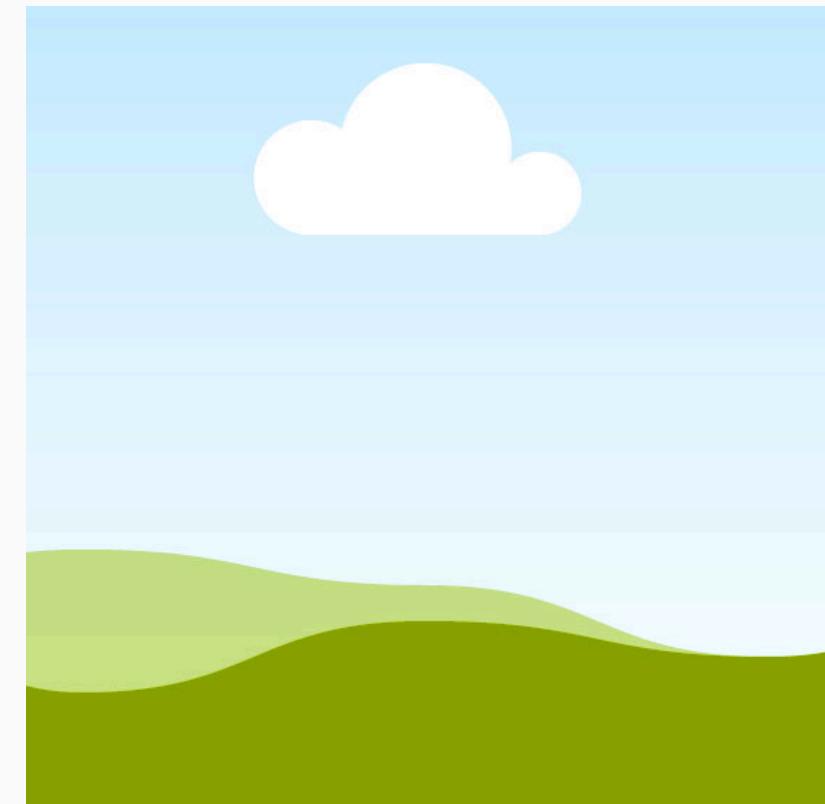
Ellen dreamed of winning a prize for her roses. She planned to enter her special purple rose at the fair. She fertilized the rose bush and covered it each night. The roses grew more beautiful every day.

Ellen ended up winning the prize.



?

PROMPT

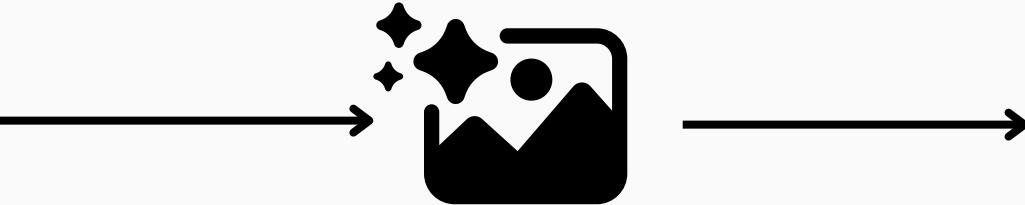


Ellen dreamed of winning a prize for her roses. She planned to enter her special purple rose at the fair. She fertilized the rose bush and covered it each night. The roses grew more beautiful every day.

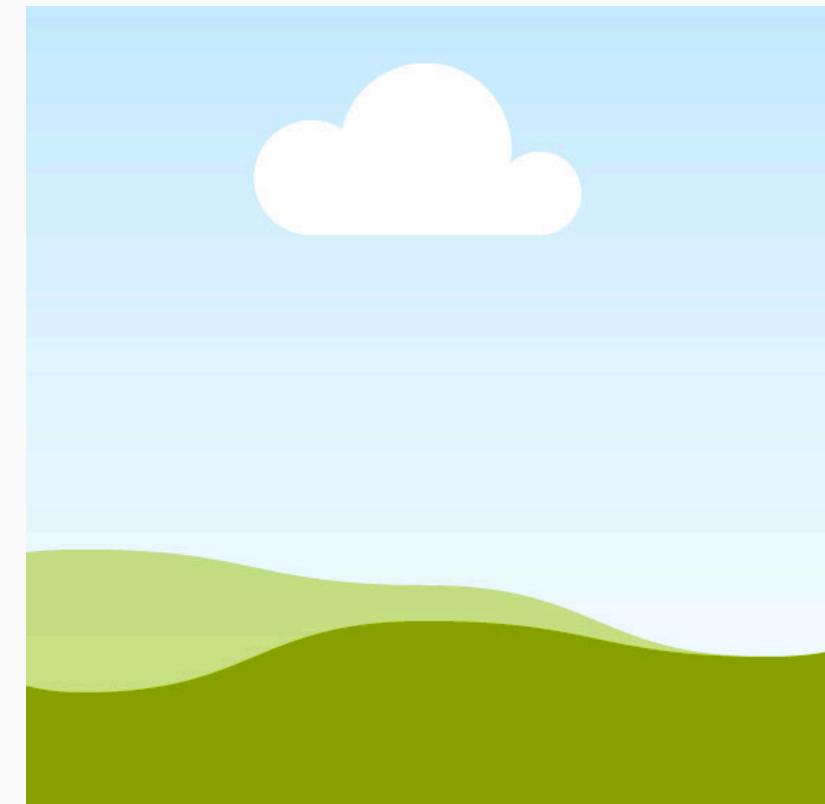
Ellen ended up winning the prize.



“Ellen ended up winning the prize.”



Use fragment itself as prompt
("No-Context" Strategy)



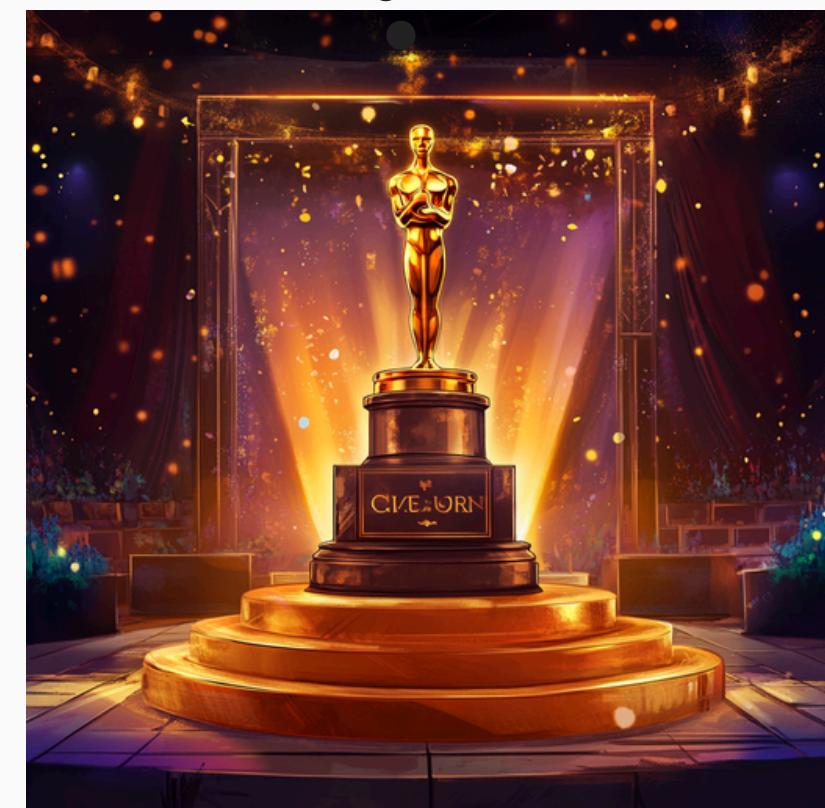
Ellen dreamed of winning a prize for her roses. She planned to enter her special purple rose at the fair. She fertilized the rose bush and covered it each night. The roses grew more beautiful every day.

Ellen ended up winning the prize.

Obviously, image is missing important context

“Ellen ended up winning the prize.”

Use fragment itself as prompt (“No-Context” Strategy)

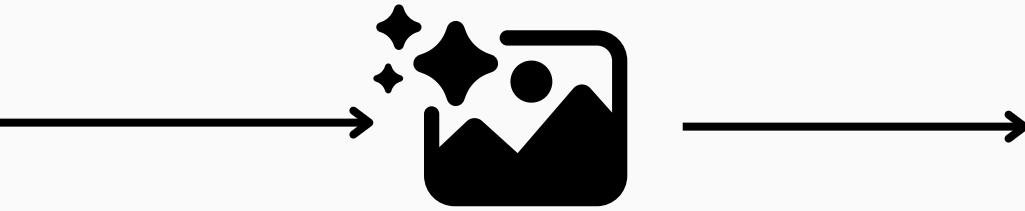


Ellen dreamed of winning a prize for her roses. She planned to enter her special purple rose at the fair. She fertilized the rose bush and covered it each night. The roses grew more beautiful every day.

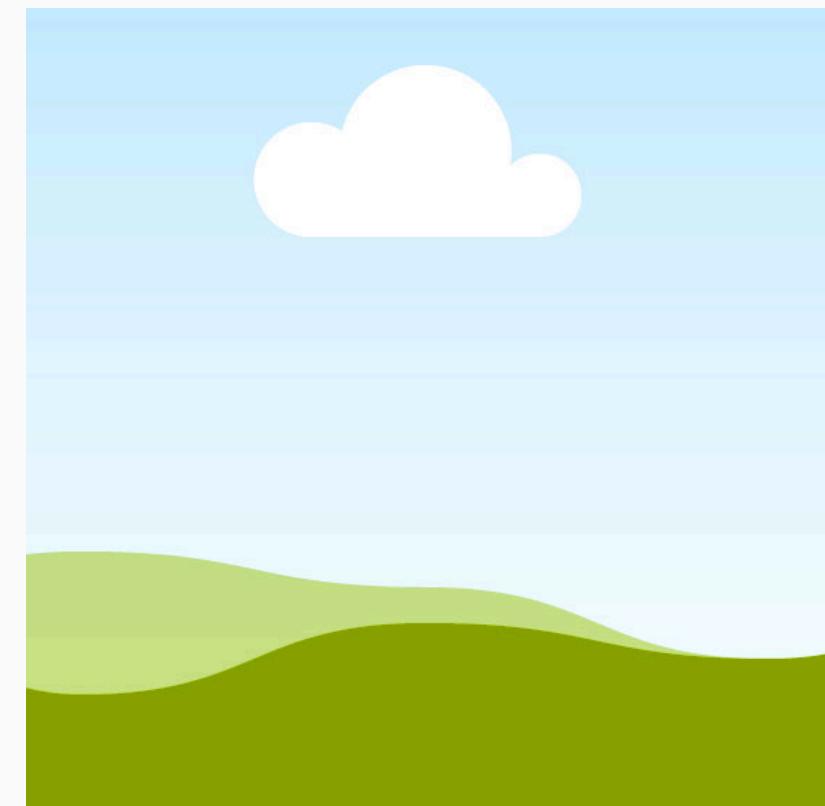
Ellen ended up winning the prize.



"Consider this story: [Ellen dreamed of winning a prize for her roses...] Based on this context, illustrate this fragment of the story: [Ellen ended up winning the prize.]"



Include the entire story before the fragment
("Verbose-Context" Strategy)



Ellen dreamed of winning a prize for her roses. She planned to enter her special purple rose at the fair. She fertilized the rose bush and covered it each night. The roses grew more beautiful every day.

Ellen ended up winning the prize.



"Consider this story: [Ellen dreamed of winning a prize for her roses...] Based on this context, illustrate this fragment of the story: [Ellen ended up winning the prize.]"



Too much context gives result is similar to not enough context

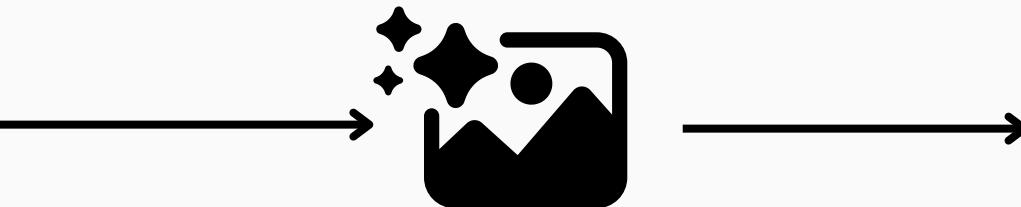
Include the entire story before the fragment
("Verbose-Context" Strategy)

Ellen dreamed of winning a prize for her roses. She planned to enter her special purple rose at the fair. She fertilized the rose bush and covered it each night. The roses grew more beautiful every day.

Ellen ended up winning the prize.



"The woman who had carefully tended her special purple rose bush ended up winning the prize at the fair."



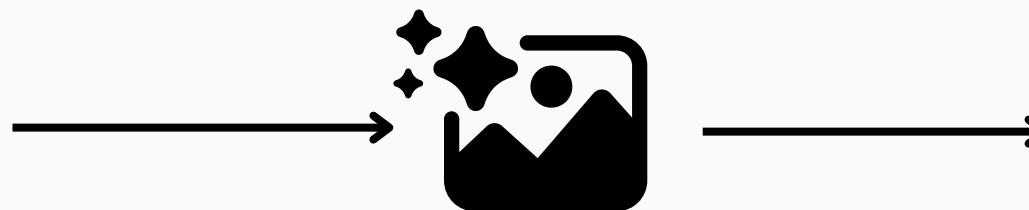
Use the fragment as the prompt, but edit it to summarize the context ("Succinct-Context" Strategy)

Ellen dreamed of winning a prize for her roses. She planned to enter her special purple rose at the fair. She fertilized the rose bush and covered it each night. The roses grew more beautiful every day.

Ellen ended up winning the prize.



"The woman who had carefully tended her special purple rose bush ended up winning the prize at the fair."



Use the fragment as the prompt, but edit it to summarize the context ("Succinct-Context" Strategy)

Reflects story, but still does not depict what's happening in this particular fragment

Ellen dreamed of winning a prize for her roses. She planned to enter her special purple rose at the fair. She fertilized the rose bush and covered it each night. The roses grew more beautiful every day.

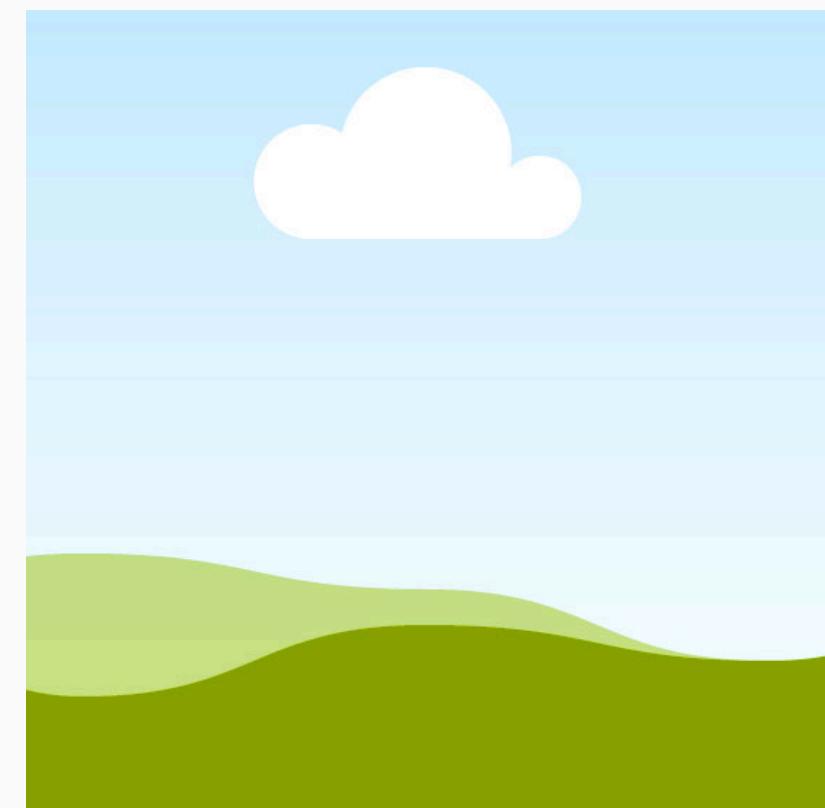
Ellen ended up winning the prize.



"A beaming woman with gardening gloves stands proudly next to a vibrant purple rose bush in full bloom. She's holding a large blue ribbon that reads "First Prize" and a golden trophy. The scene is set at a county fair, with other flower displays and people visible in the background."



Describe the visual content of the *scene* evoked by the fragment
("Scene Description" Strategy)



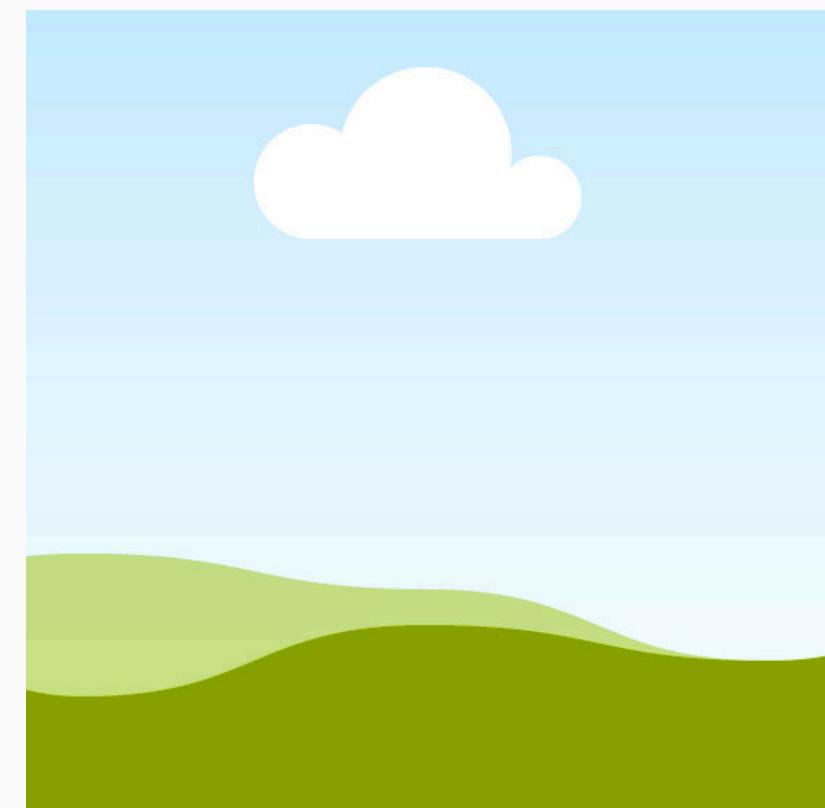
Ellen dreamed of winning a prize for her roses. She planned to enter her special purple rose at the fair. She fertilized the rose bush and covered it each night. The roses grew more beautiful every day.

Ellen ended up winning the prize.



"A beaming woman with gardening gloves stands proudly next to a vibrant purple rose bush in full bloom. She's holding a large blue ribbon that reads "First Prize" and a golden trophy. The scene is set at a county fair, with other flower displays and people visible in the background."

Articulate visual details that are implied



Describe the visual content of the *scene* evoked by the fragment
("Scene Description" Strategy)

Ellen dreamed of winning a prize for her roses. She planned to enter her special purple rose at the fair. She fertilized the rose bush and covered it each night. The roses grew more beautiful every day.

Ellen ended up winning the prize.

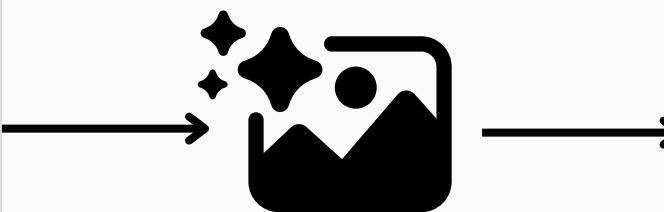


"A beaming woman with gardening gloves stands proudly next to a vibrant purple rose bush in full bloom. She's holding a large blue ribbon that reads "First Prize" and a golden trophy. The scene is set at a county fair, with other flower displays and people visible in the background."

Describe the visual content of the *scene* evoked by the fragment
("Scene Description" Strategy)

Articulate visual details that are implied

Finally, we get a better illustration of the fragment



To use text-to-image models for story visualization, you need to describe the scene

- Annotators saw two alternative images for the same story fragment and were asked “*which is the better visualization of the fragment?*”
- Images generated from scene descriptions were significantly favored over images generated from other prompts that use the story text directly

Scene Description vs.	% of Pairs Where Scene Description Won
No-Context	78%
Verbose-Context	75%
Succinct-Context	73%

To use text-to-image models for story visualization, you need to describe the scene

- Annotators saw two alternative images for the same story fragment and were asked “*which is the better visualization of the fragment?*”
- Images generated from scene descriptions were significantly favored over images generated from other prompts that use the story text directly

Scene Description vs.	% of Pairs Where Scene Description Won
No-Context	78.1
Verbose-Context	74.7
Succinct-Context	72.5

★ Much knowledge useful for visualization is implicit, not explicitly stated in the story



Synthesizing scene descriptions

- Where do the scene descriptions come from?
The author of a story can't be expected to write them for every possible fragment
- Alternatively, large language models (LLMs) can act as a “visualization interface” for stories by automatically generating scene descriptions

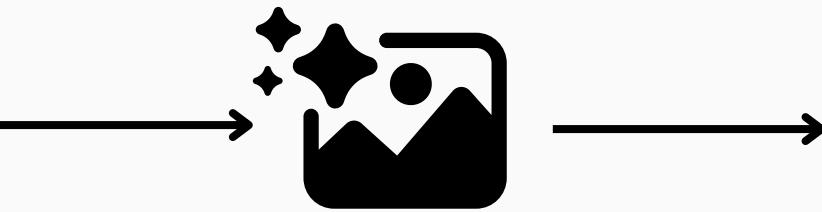
Ellen dreamed of winning a prize for her roses. She planned to enter her special purple rose at the fair. She fertilized the rose bush and covered it each night. The roses grew more beautiful every day.

Ellen ended up winning the prize.



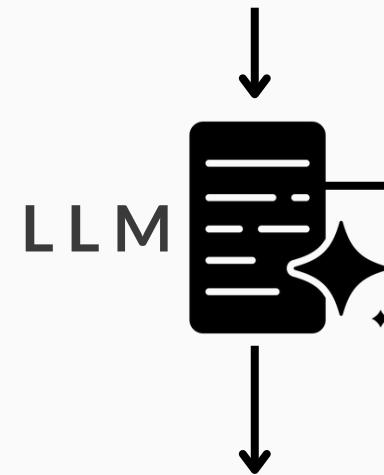
"A beaming woman with gardening gloves stands proudly next to a vibrant purple rose bush in full bloom. She's holding a large blue ribbon that reads "First Prize" and a golden trophy. The scene is set at a county fair, with other flower displays and people visible in the background."

SCENE DESCRIPTION



Ellen dreamed of winning a prize for her roses. She planned to enter her special purple rose at the fair. She fertilized the rose bush and covered it each night. The roses grew more beautiful every day.

Ellen ended up winning the prize.



"A beaming woman with gardening gloves stands proudly next to a vibrant purple rose bush in full bloom. She's holding a large blue ribbon that reads "First Prize" and a golden trophy. The scene is set at a county fair, with other flower displays and people visible in the background."

SCENE DESCRIPTION

LLM INSTRUCTIONS

Imagine an AI system will be used to generate visualizations for story fragments...Your task is to read a story fragment along with its story context and write a scene description that specifies how to visualize the fragment...



VISUAL E-READER TUTORIAL

github.com/roemmele/Envisioning-Stories-with-Generative-AI/example.ipynb

```
▶ ▾
  with open("llm_prompt.txt", "r") as f:
  |   llm_prompt_template = f.read()
  print(llm_prompt_template)
[27] ✓ 0.0s                                     Python

...
... You are operating inside a "visual e-reader" application that envisions highlighted passages of text as
... images using an AI text-to-image generation model. Your task is to consider a highlighted passage
... ("fragment") inside of its story context, and then write a brief "scene description" (max 100 words) that
... prompts the text-to-image model to visually depict the content of the fragment. The scene description
... should focus specifically on what's mentioned in the fragment, using the provided context for
... interpretation. Respond with the scene description only, without preamble.

Highlighted Passage: {fragment}
Context: {story}
Scene Description:
```

VISUAL E-READER TUTORIAL

github.com/roemmele/Envisioning-Stories-with-Generative-AI/example.ipynb

```
[2] story = ("Lisa has a beautiful sapphire ring. "
           "She always takes it off to wash her hands. "
           "One afternoon, she noticed it was missing from her finger! "
           "Lisa searched everywhere she had been that day. "
           "She was elated when she found it on the bathroom floor!")
    ✓ 0.0s                                         Python
```

```
[11] llm_prompt = llm_prompt_template.format(
        story=story,
        fragment="Lisa has a beautiful sapphire ring.")
    ✓ 0.0s                                         Python
```

VISUAL E-READER TUTORIAL

github.com/roemmele/Envisioning-Stories-with-Generative-AI/example.ipynb

```
▶ ▾
    import replicate, os

    client = replicate.client.Client(api_token=os.environ["REPLICATE_API_KEY"])
    scene_description = "".join(client.run(
        "meta/meta-llama-3.1-405b-instruct",
        input={"prompt": llm_prompt}))
    print(scene_description)
[12] ✓ 2.6s
Python
...
... A close-up of Lisa's hand, with the beautiful sapphire ring prominently displayed on her finger, sparkling in the light, set against a soft, neutral background to emphasize the ring's vibrant blue color.
```

VISUAL E-READER TUTORIAL

github.com/roemmele/Envisioning-Stories-with-Generative-AI/example.ipynb

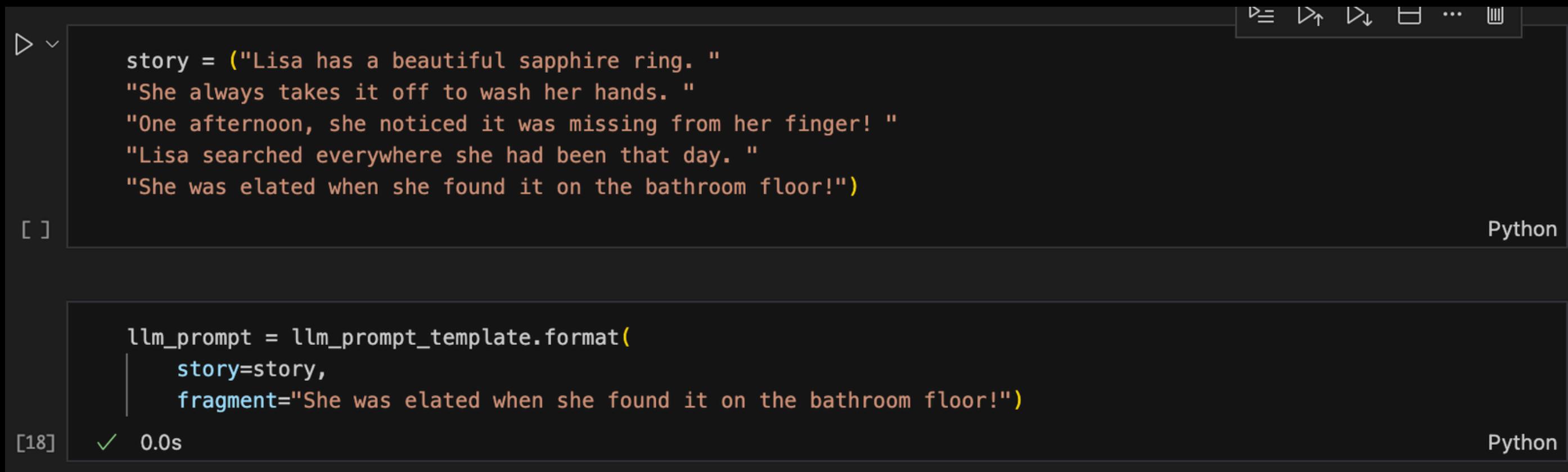
```
▶ ▾
  from IPython.display import Image

  url = client.run("black-forest-labs/flux-1.1-pro",
                   input={"prompt": scene_description,
                           "output_format": "png"})
  Image(url, width=300)
[16] ✓ 4.3s Python
...

```

VISUAL E-READER TUTORIAL

github.com/roemmele/Envisioning-Stories-with-Generative-AI/example.ipynb



```
▶ story = ("Lisa has a beautiful sapphire ring. "
   "She always takes it off to wash her hands. "
   "One afternoon, she noticed it was missing from her finger! "
   "Lisa searched everywhere she had been that day. "
   "She was elated when she found it on the bathroom floor!")
```

```
[ ] Python
```

```
llm_prompt = llm_prompt_template.format(
    story=story,
    fragment="She was elated when she found it on the bathroom floor!")
```

```
[18] ✓ 0.0s Python
```

VISUAL E-READER TUTORIAL

github.com/roemmele/Envisioning-Stories-with-Generative-AI/example.ipynb

```
▶ 
  scene_description = "".join(client.run(
    "meta/meta-llama-3.1-405b-instruct",
    input={"prompt": llm_prompt}))
  print(scene_description)
[19] ✓ 6.0s                                     Python
...
... A young woman, overjoyed, bends down to pick up a sapphire ring from the white tile floor of a bathroom. Her hand reaches out, and her fingers close around the ring as a relieved smile spreads across her face. Soft, natural light enters through a nearby window, illuminating the scene. The bathroom's neutral colors and simple fixtures provide a calm background for the woman's triumphant moment. Her eyes shine with happiness as she gazes at the recovered ring.
```

VISUAL E-READER TUTORIAL

github.com/roemmele/Envisioning-Stories-with-Generative-AI/example.ipynb

```
url = client.run("black-forest-labs/flux-1.1-pro",
                  input={"prompt": scene_description,
                         "output_format": "png"})
Image(url, width=300)
```

[25] ✓ 4.0s Python

...



VISUAL E-READER APP DEMO

github.com/roemmele/Envisioning-Stories-with-Generative-AI/visual-e-reader

The screenshot shows the Visual E-Reader application interface. At the top, there is a header bar with the title "Visual E-Reader" and the subtitle "The Yellow Wallpaper (By Charlotte Perkins Gilman)". On the right side of the header are three icons: a file icon labeled "Upload .txt File", a gear icon for settings, and a user icon.

The main content area is divided into two sections. On the left, there is a large white text area containing the story "The Yellow Wallpaper". The first few paragraphs read:

The Yellow Wallpaper (By Charlotte Perkins Gilman)

It is very seldom that mere ordinary people like John and myself secure ancestral halls for the summer.

A colonial mansion, a hereditary estate, I would say a haunted house, and reach the height of romantic felicity—but that would be asking too much of fate! ↵

Still I will proudly declare that there is something queer about it.

Else, why should it be let so cheaply? And why have stood so long untenanted?

John laughs at me, of course, but one expects that in marriage.

John is practical in the extreme. He has no patience with faith, an intense horror of superstition, and he scoffs openly at any talk of things not to be felt and seen and put down in

On the right side, there is a sidebar titled "Image" which contains a placeholder image icon (a gray circle with a curved arrow) and the text "No highlights yet Select text to apply imagination".



TAKEAWAY

- Visualization can provide new opportunities to engage with text-based stories
- Text-to-image models are useful for on-the-fly visualization
- LLMs can support this by verbalizing visual scenes
- Emerging opportunities to expand story engagement to other modalities beyond images (e.g. video, audio)



A large, open book lies flat against a dark background. The pages of the book are filled with a detailed landscape painting of a coastal scene at sunset or sunrise. The sky is a warm, golden-yellow, transitioning into darker blues and purples. Below the horizon, a calm sea with gentle waves stretches to the right. In the foreground, a small, dark silhouette of a person stands on the left page, looking out over the water. The overall composition is surreal and dreamlike.

THANK YOU!

MELISSA@ROEMMELE.IO