

Building an AI Comic Video Generator for My Son

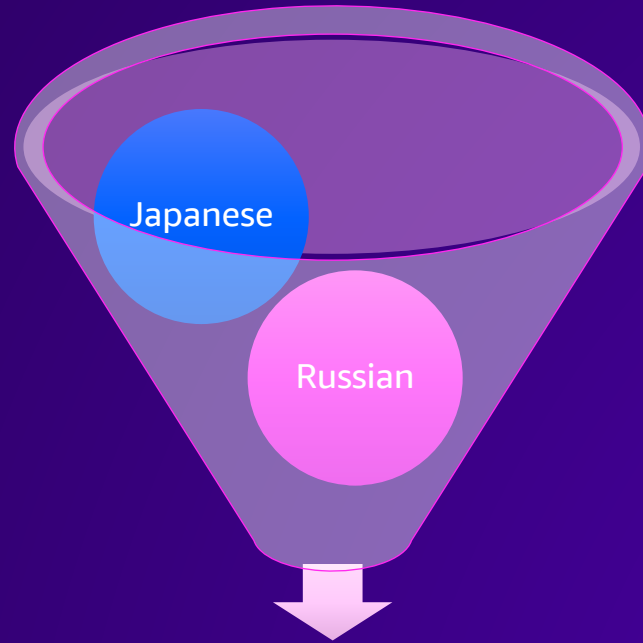
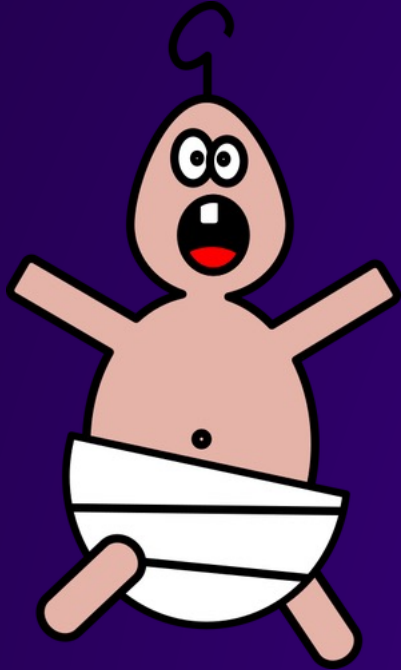
Agustinus Nalwan

GM of AI, Data Science, Data Platform
carsales.com - AWS ML Hero

About me

- Technology lover
- Solving problem with Tech & AI
- Side hobby – Building AI Tech to entertain my family

Baby Name Generator



Vladimasaki, Viktoryukicho





Project Griffin – Bird Simulator

Project faAi - Fashion Assistant AI



Project Ellee

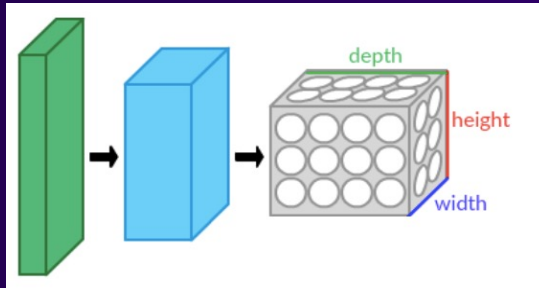


Robotic Teddy Bear

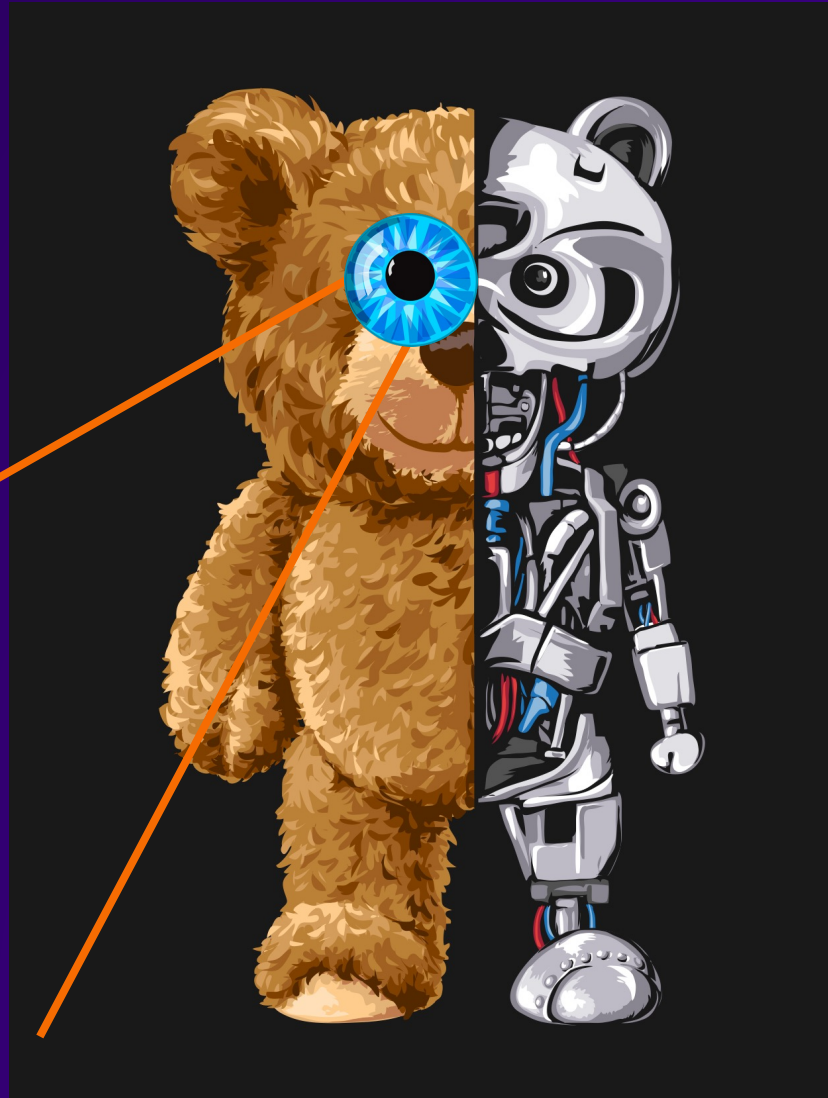


Project Ellee

Large Language Models



MobileNetSSDV2



Ellee in Action

Project Ellee



To view all my projects



Owly – a Personalised Comic Video Generator



Why?



Why?



Sustainable solution



AI story generator - Owly



What does Owly do?



Story title



Owly in Action



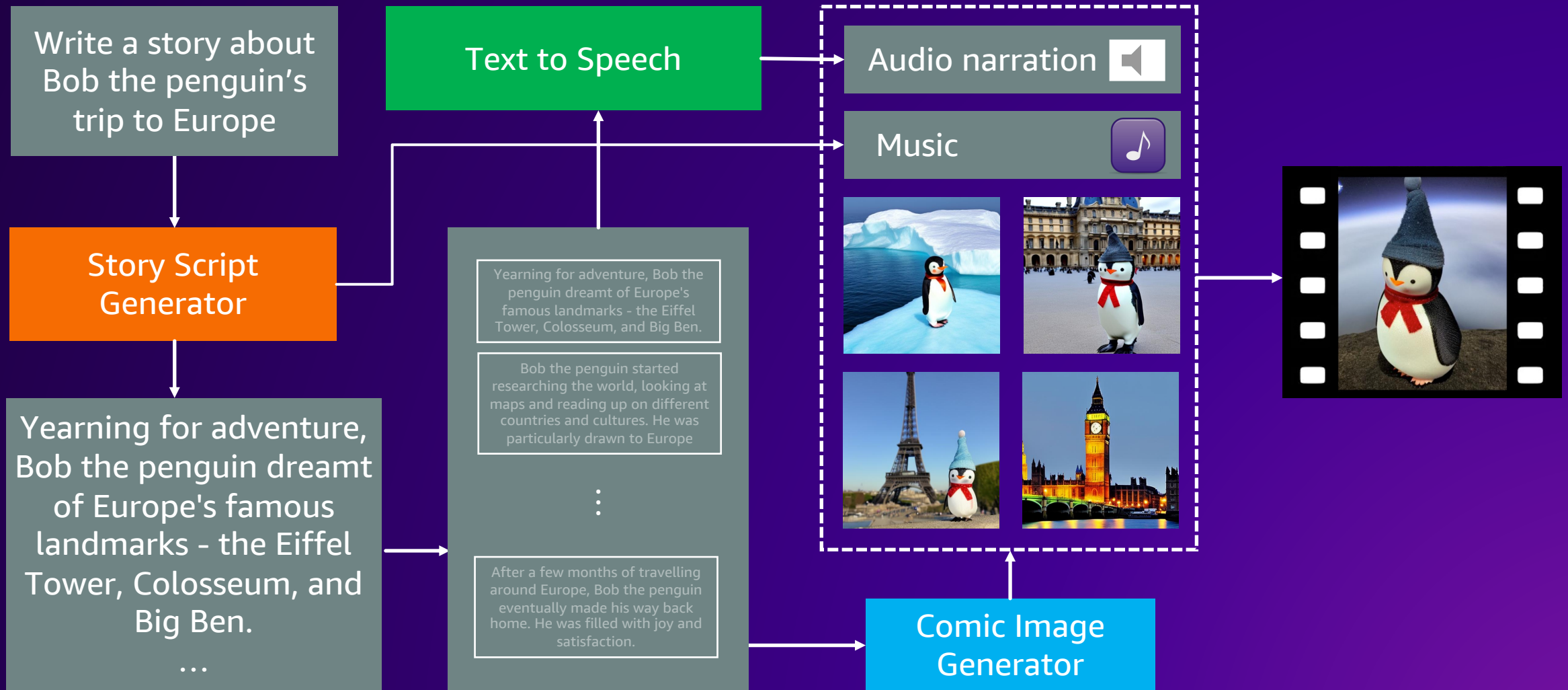


Bob the Brave Penguin: Adventures in Europe

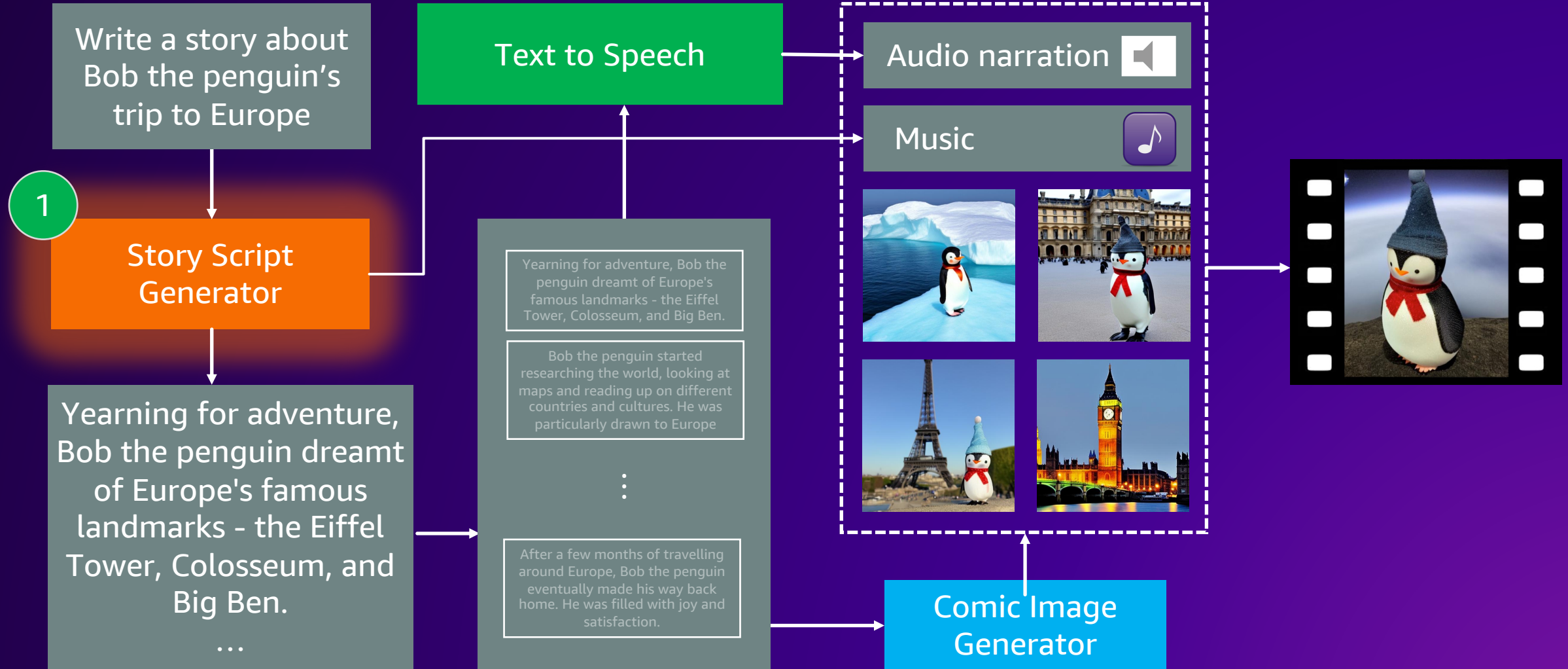


Made by Owly AI

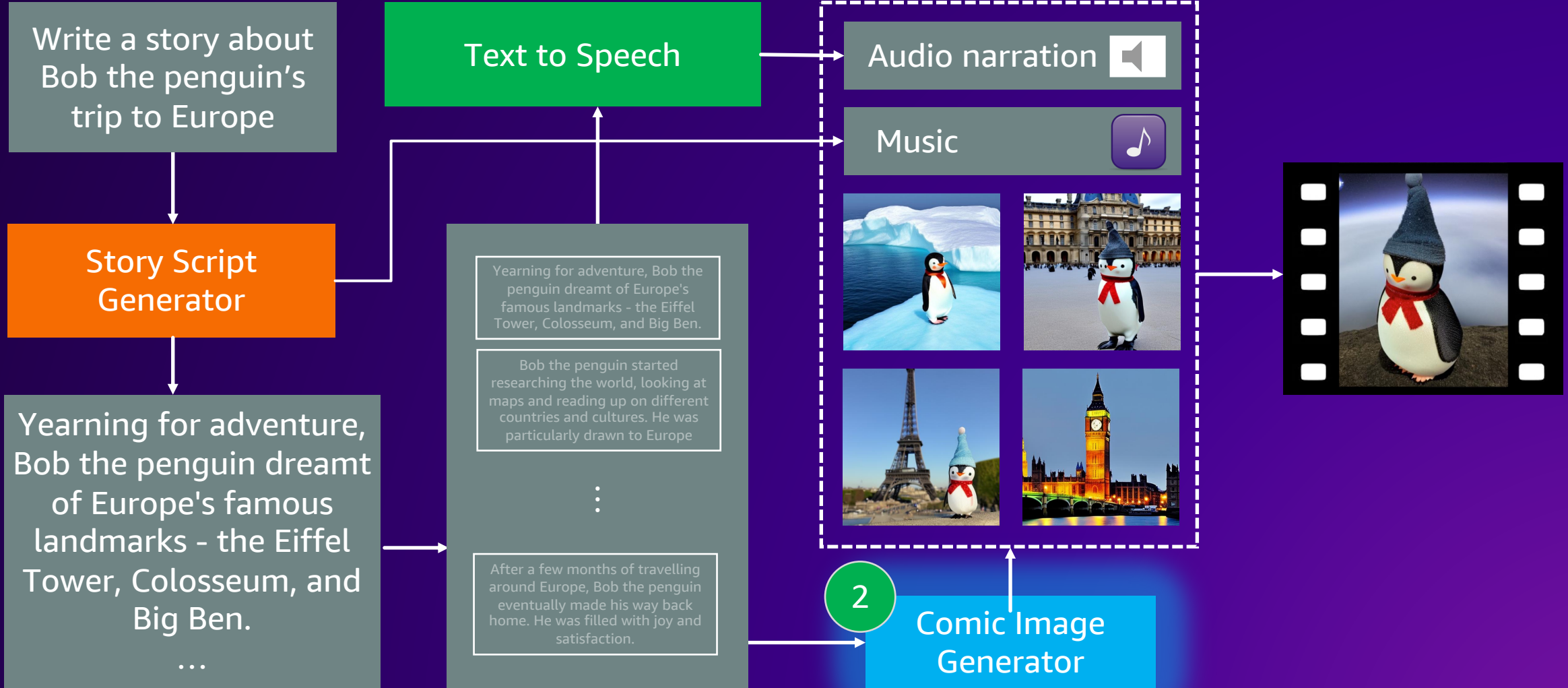
How does Owly work?



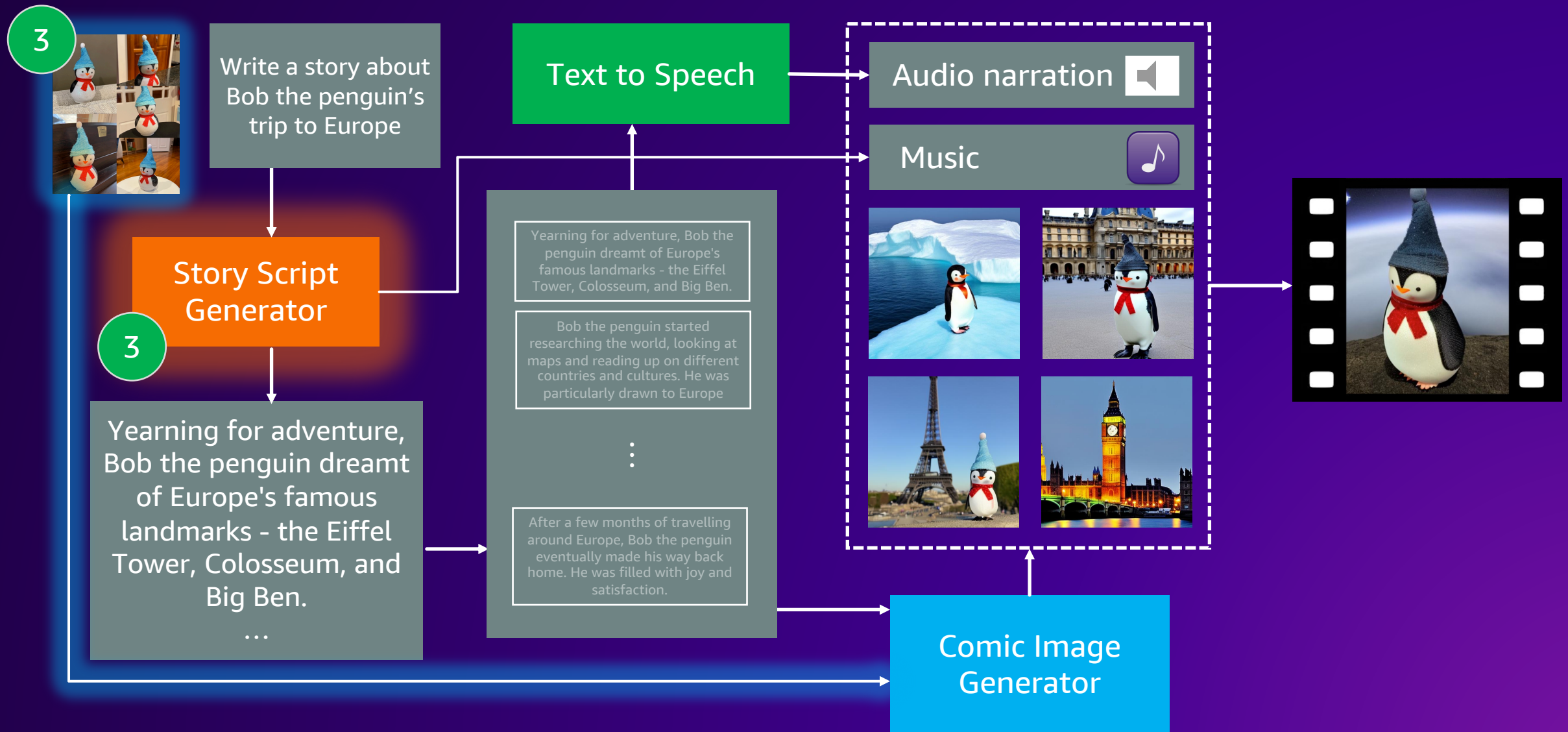
Agenda - Story Script Generator



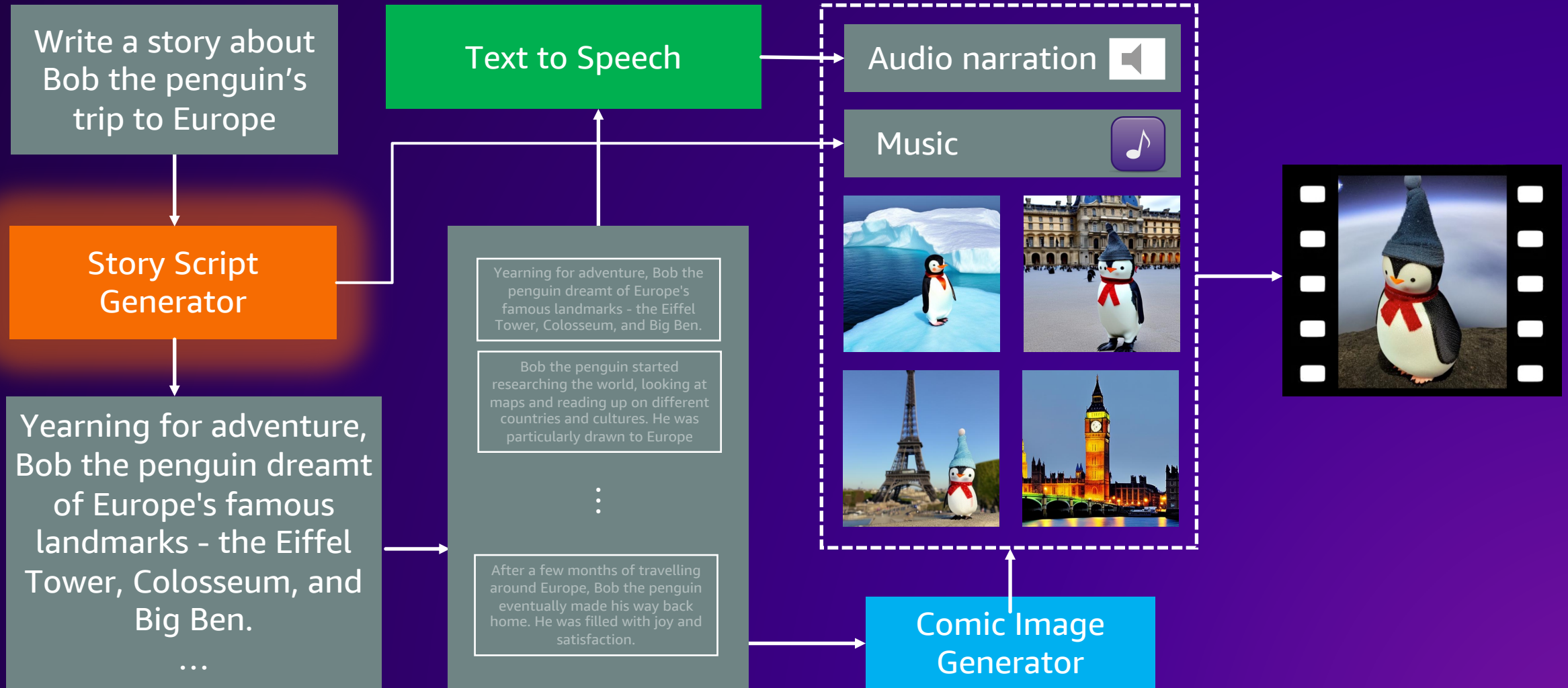
Agenda – Comic Image Generator



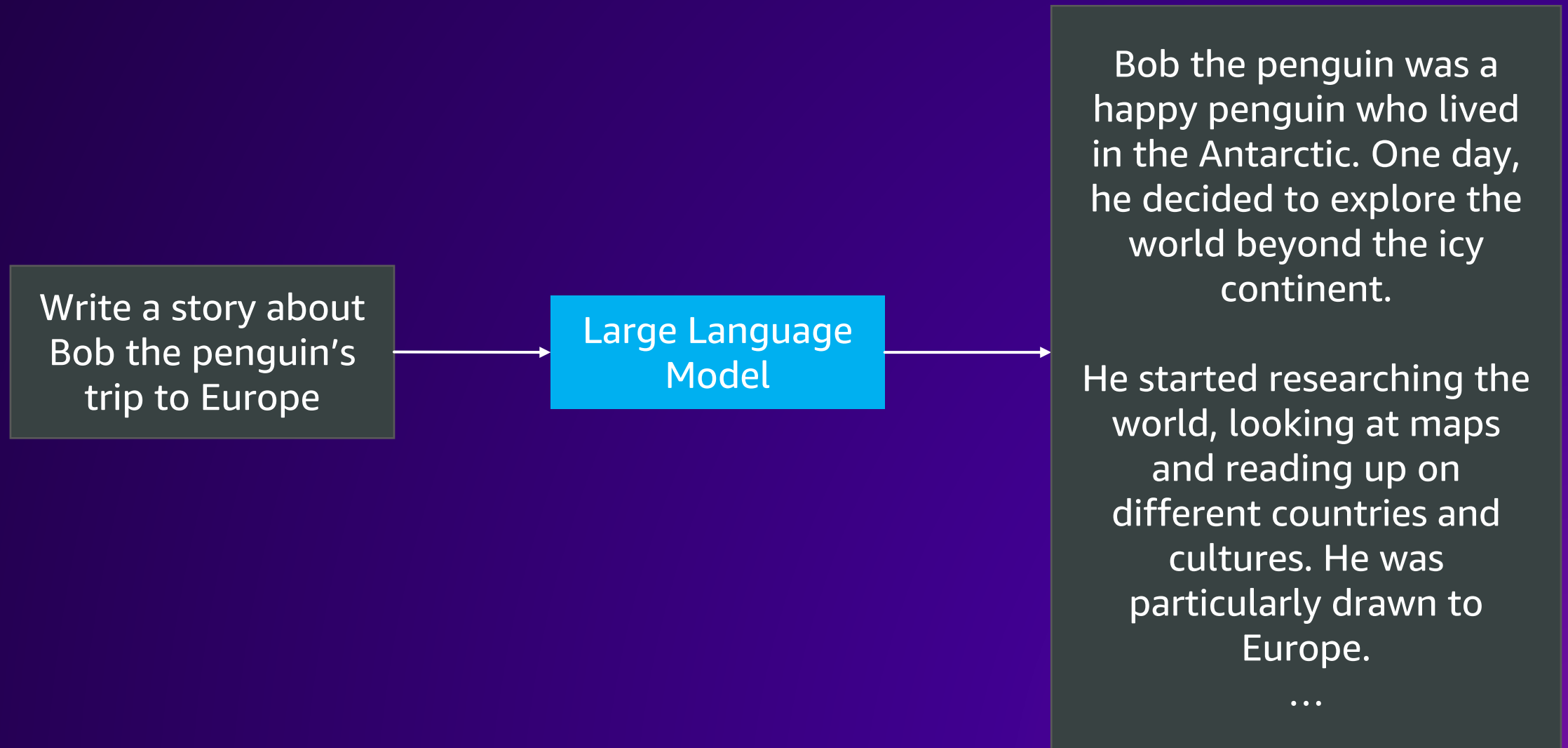
Agenda – Challenge and Enhancement



Building the story script generator



Large Language Model



Amazon Bedrock ×

Overview

Examples

▼ Foundation models

Models


Providers

Embeddings

▼ Playgrounds

Text

Image


Terms of service 

Overview

Foundation models


Discover models

Amazon Bedrock supports foundation models from industry-leading providers. Choose the model that is best suited to achieving your unique goals. [Explore models](#)

 **Jurassic-2 series**
By AI21 Labs


AI21 Labs is transforming writing and reading experiences with models that perform language tasks such as generating, classifying, summarizing, and paraphrasing text in seven languages.

[View provider details](#)

 **Titan**
By Amazon


Amazon leader in applying ML to e-commerce, cloud computing, online advertising, and digital streaming services. The Amazon Titan models leverage Amazon's 20+ years of ML experience.

[View provider details](#)

 **Claude**
By Anthropic

Anthropic is an AI safety and research company offering the Claude family of large language models. These models are purpose built for AI-based assistance use-cases such as obtaining customer service or comprehending documents.

[View provider details](#)

 **Stable Diffusion**
By Stability AI

Stability AI is a generative AI company with a goal to inspire global creativity and innovation.

[View provider details](#)

Amazon Bedrock X

Overview

Examples

▼ Foundation models

Models

Providers

Embeddings

▼ Playgrounds

Text

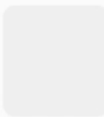
Image

Terms of service

Amazon Bedrock > Text playground

Text playground Info

Streaming



Select provider ▼

Select model ▼

Inference configuration Reset

Please select model above to view inference configurations

Amazon Bedrock ×

Overview

Examples

▼ Foundation models

Models


Providers

Embeddings

▼ Playgrounds

Text


Image

Terms of service 

Amazon Bedrock > Text playground

Text playground [Info](#)


☐ Streaming



Amazon ▾

Select model ▴

Foundation Models

 Titan Large v1.01
Context size = 4k

Custom Models

Inference configuration [Reset](#)

Please select model above to view inference configurations

Amazon Bedrock ×

Overview

Examples

▼ Foundation models

Models


Providers

Embeddings

▼ Playgrounds


Text

Image

Terms of service 


Amazon Bedrock > Text playground Info


Text playground Streaming




Anthropic ▾

Select model ▲

 Claude Instant V1 v1.3
Context size = 9k

 Claude V1 v1.3
Context size = 12k

 Claude V2 v2
Context size = 12k

Inference configuration Reset

Please select model above to view inference configurations

Human:
Assistant:

Inference configuration

[Reset](#)

▼ Randomness and diversity [Info](#)

Temperature

Top P

Top K

▶ Run



View API request

♡ Save

Human: Write me a 300 words story about Bob the Penguin who wants to travel to Europe visiting famous landmarks. Please break the story down into maximum of five sections with 30 maximum words per section.

I



Run



Inference configuration

[Reset](#)

▼ Randomness and diversity [Info](#)

Temperature

1



Top P

0.999



Top K

250



[View API request](#)



Save

along the River Seine. "Magnifique!" Bob cheered.

Then, Bob flew south to Rome. He was in awe of the Colosseum's grandeur and St. Peter's Basilica's beauty. As the sun set over the city, Bob sighed, "When in Rome!"

Sadly, Bob's trip was ending. On his last night, he stargazed at the canals of Venice. "I'll always treasure this dream come true," he said.

The next morning, Bob boarded a ship for Antarctica, his heart full of wonderful memories to share with his penguin friends. Bob proved that with determination, dreams can become reality. |

**Run**

Inference configuration

[Reset](#)

▼ Randomness and diversity [Info](#)

Temperature



Top P



Top K

[View API request](#)[Save](#)

Human: Write me a 300 words words story about Bob the Penguin who wants to travel to Europe visiting famous landmarks.
Please break the story down into maximum of five sections with 30 maximum words per section.
At the start of the story please suggest song style from the following list only which matches the story and put it within <>. Song style list are action, calm, dramatic, epic, happy and touching..

 Run

Inference configuration

[Reset](#)

▼ Randomness and diversity [Info](#)

Temperature

1

Top P

0.999

Top K

250

[View API request](#) Save

by.

In Rome, Bob explored the Colosseum and Roman Forum. He threw a coin in the Trevi Fountain, hoping to return to Rome again someday. He saw masterpiece artworks like the Creation of Adam on the ceiling of the Sistine Chapel. "Michelangelo was a genius!" exclaimed Bob.

Finally, Bob journeyed to Athens, cradle of civilization. He admired the ancient Parthenon and Acropolis. "Standing here, you can feel the history," Bob said in awe. His trip at an end, Bob returned home, his suitcase full of memories to last a lifetime. His friends welcomed him back, eager to hear stories of his adventures in faraway lands.



Run



Inference configuration

[Reset](#)

▼ Randomness and diversity [Info](#)

Temperature

1



Top P

0.999



Top K

250



View API request

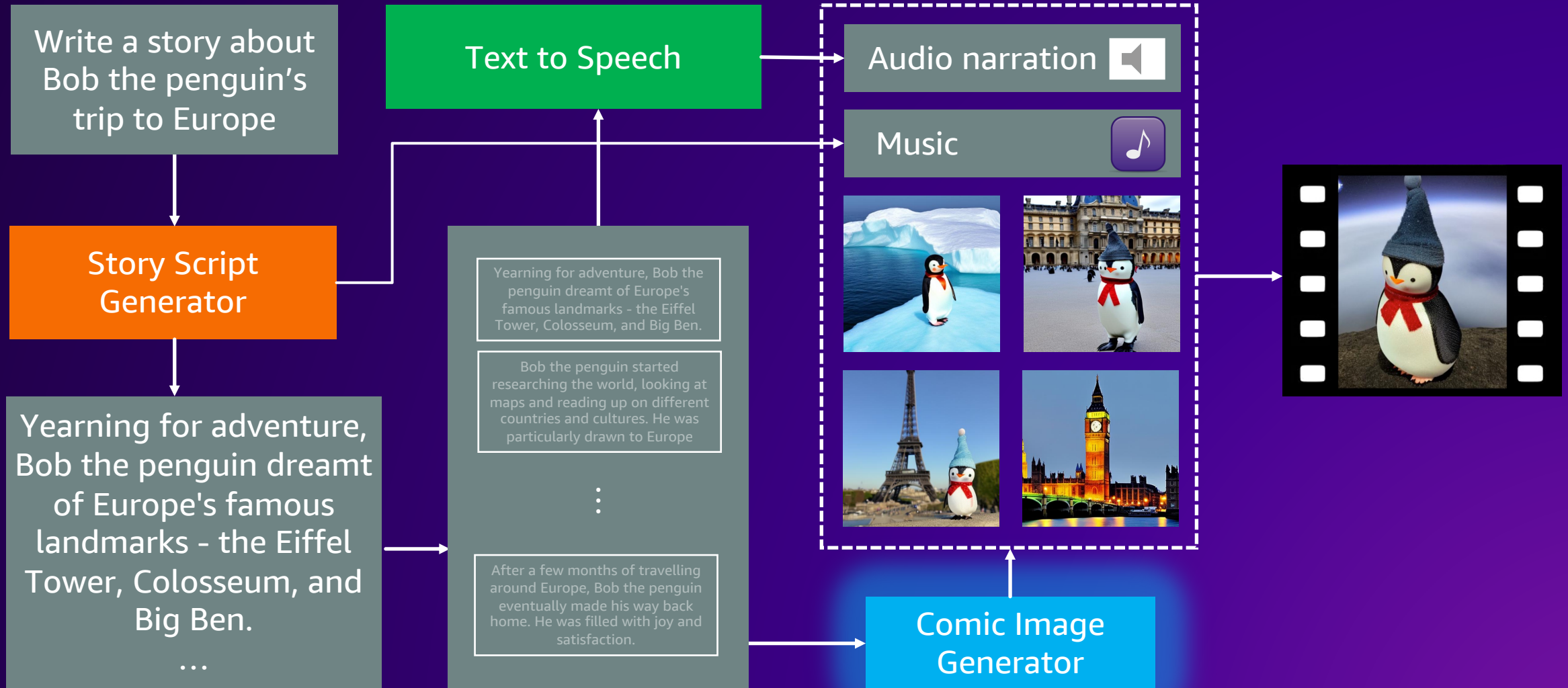


Save

Amazon Bedrock

```
bedrock = boto3.client(  
    service_name='bedrock',  
    region_name='us-west-2',  
    endpoint_url='https://bedrock.us-west-2.amazonaws.com'  
)  
  
body = json.dumps({"prompt": full_query, "max_tokens_to_sample": 1000})  
modelId = 'anthropic.claude-v1'  
accept = 'application/json'  
contentType = 'application/json'  
  
response = bedrock.invoke_model(body=body, modelId=modelId, accept=accept, contentType=contentType)  
response_body = json.loads(response.get('body').read())  
out_text = response_body.get('completion')  
  
return out_text
```

Building the comic image generator



Comic Image Generator

Penguin is
standing on an
iceberg in
antarctica

Comic Image
Generator



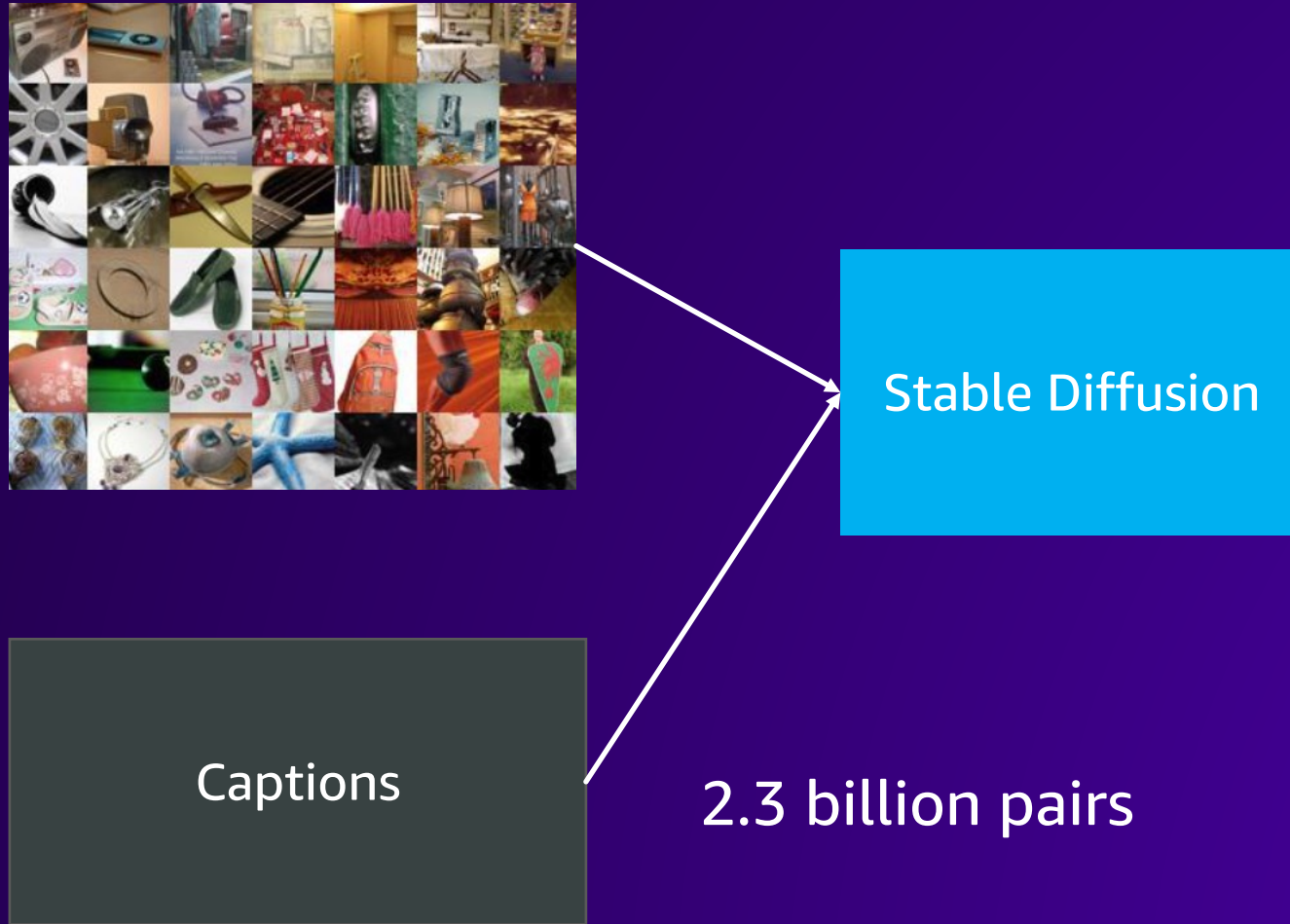
Stable Diffusion 2.1

Penguin is
standing on an
iceberg in
antarctica

Stable Diffusion
2.1 model



How Stable Diffusion model was built



Stable Diffusion

Captions

2.3 billion pairs

Stable Diffusion Model

Turtle is swimming in
the sea

Stable Diffusion



Stable Diffusion 2.1



Meat lover pizza with banana toppings

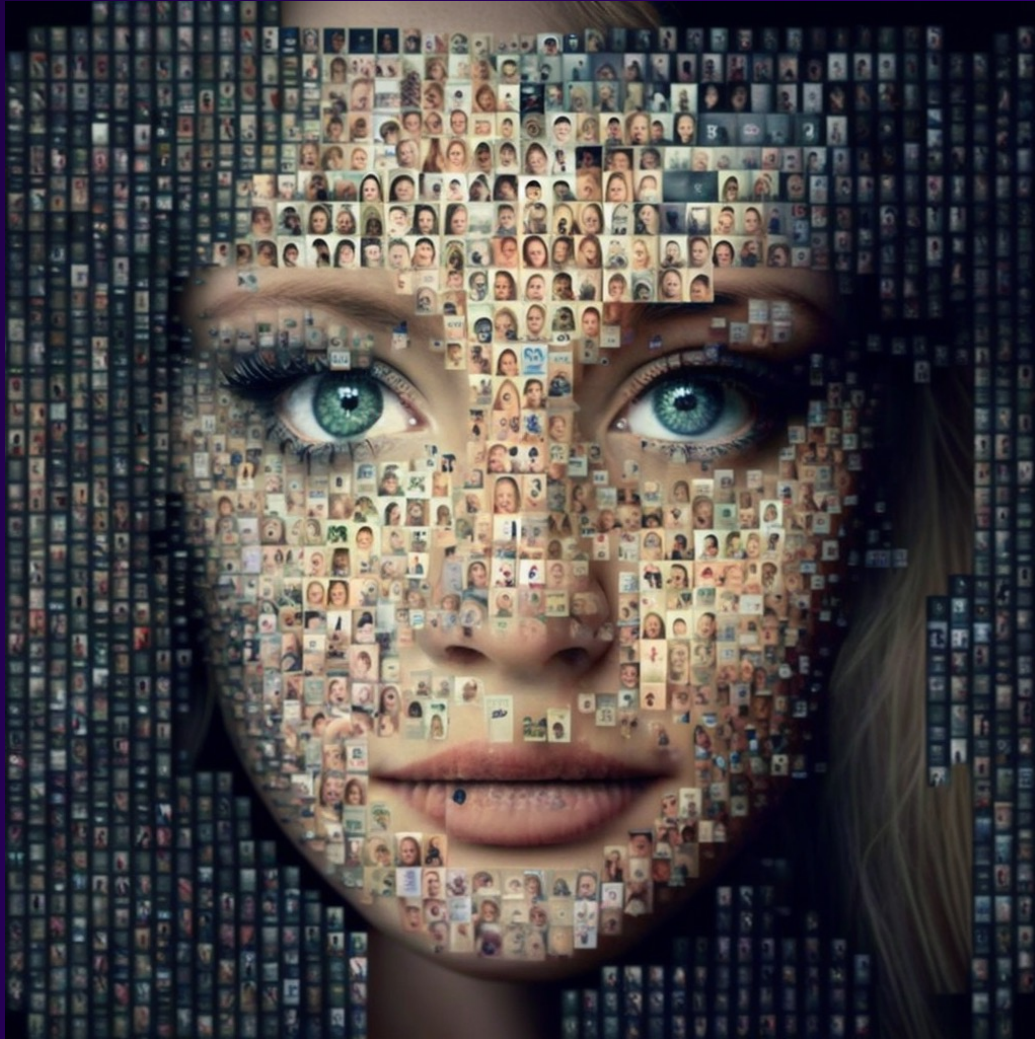


A shaolin monk is riding a harley davidson on a countryside



A monkey and a panda are playing chess

Training Stable Diffusion Model



Amazon SageMaker JumpStart

SageMaker JumpStart


[Show introduction](#)[Browse Shared Models](#)

[Solutions](#)[ML tasks](#)[Data types](#)[Notebooks](#)[Frameworks](#)[Resources](#)

Foundation Models

Deploy foundation models trained on broad dataset and usable in wide range of use cases.

[Explore All Foundation Models \(13\)](#)




Stable Diffusion 2.1 base

[Featured](#)[Text To Image](#)

Fine-tunable: **Yes**
Source: **Stability AI**
Pre-training Dataset: **LAION-5B**

[View model >](#)




FLAN-T5 XXL

[Text2text Generation](#)

Pre-training Dataset: **English Text**
Fine-tunable: **No**
Source: **Hugging Face**

[View model >](#)



Alexa TM 20B

[Featured](#)[Text Generation](#)


Pre-training Dataset: **Common Crawl (mC4) and...**
Fine-tunable: **No**
Source: **Alexa**

[View model](#)

Vision Models

Fine-tune and deploy pretrained vision models with one click.

[Explore All Vision Models \(293\)](#)



Stable Diffusion x4 upscaler...

[Featured](#)[Text](#)

Fine-tunable: **No**
Source: **Stability AI**





Image Classification

[Featured](#)[Swin Base 4 7 224](#)

Dataset: **ImageNet**
Fine-tunable: **Yes**
Source: **TensorFlow Hub**



Object Detection

[Ssd Resnet50 V1 Fpn 640x640 Coco 17](#)

Pre-training Dataset: **COCO '17**
Fine-tunable: **Yes**
Source: **Tensorflow**


Deploying Stable Diffusion 2.1

MODEL

Stable Diffusion 2.1 base

text · text to image · foundation models · featured

Open notebook

 Browse JumpStart

Deploy Train Notebook Model details

Deploy Model

Deploy a pretrained model to an endpoint for inference. Deploying on SageMaker hosts the model on the specified compute instance and creates an internal API endpoint. JumpStart will provide you an example notebook to access the model after it is deployed. [Learn more.](#)

> Deployment Configuration

> Security Settings

Deploy

Deploying Stable Diffusion 2.1

The screenshot shows the Amazon SageMaker Studio interface. On the left, the file explorer shows a directory named 'StableDiffusion' with subdirectories 'old' and 'Amazon_JumpStart_Text_T...'. The main notebook area displays the following code:

```
[ 3 ]: from ipywidgets import Dropdown
from sagemaker.jumpstart.notebook_utils import list_jumpstart_

# Retrieves all Text-to-Image generation models.
filter_value = "task == txt2img"
txt2img_models = list_jumpstart_models(filter=filter_value)

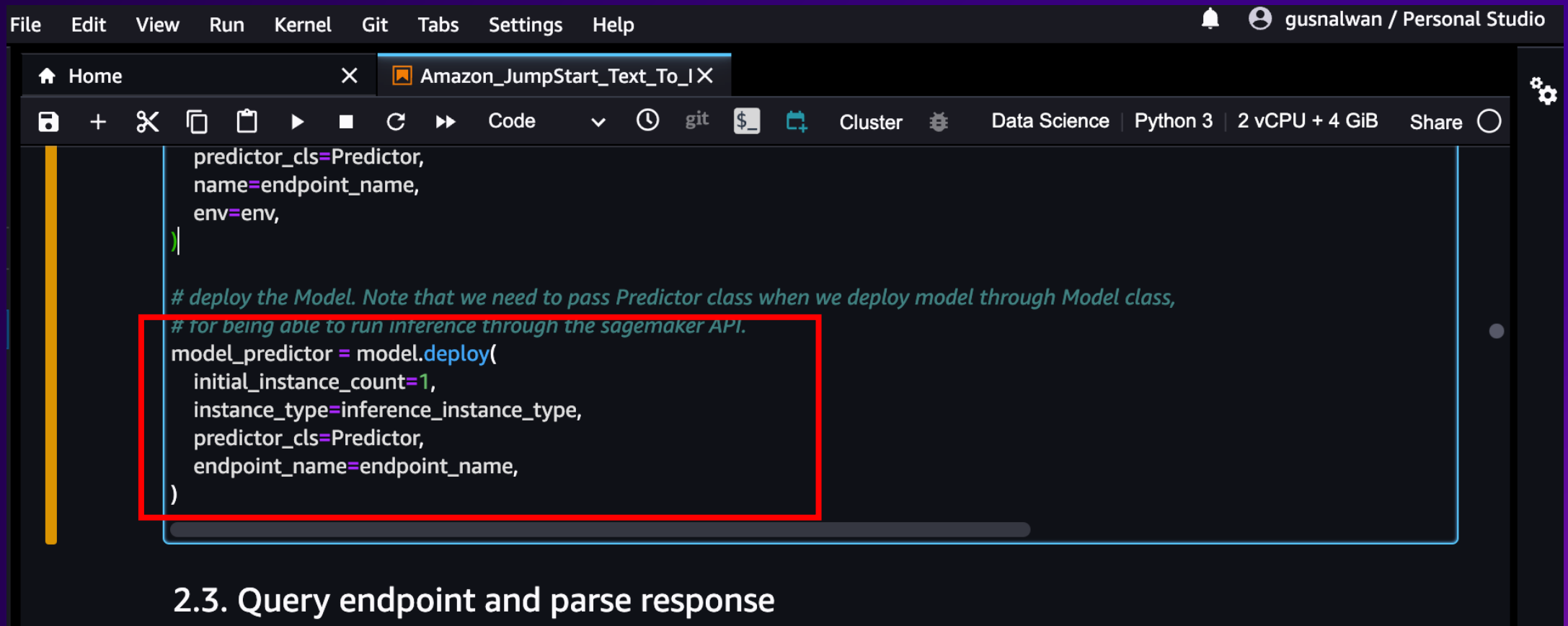
# display the model-ids in a dropdown to select a model for inference
model_dropdown = Dropdown(
    options=txt2img_models,
    value="model-txt2img-stabilityai-stable-diffusion-v2-fp16",
    description="Select a model",
    style={'description_width': "initial"},
    layout={"width": "max-content"},
)
display(model_dropdown)

[ 5 ]: # model_version="" fetches the latest version of the model
model_id, model_version = model_dropdown.value, ""
print(model_id)
print(model_version)
```

Below the code, a dropdown menu is shown with the selected model: 'model-txt2img-stabilityai-stable-diffusion-v2-fp16'. The bottom status bar indicates 'Python 3 (Data Science) | Idle' and 'Kernel: Idle | Instance MEM'.

huggingface-txt2img-piesposito-openpotionbottle-v2
huggingface-txt2img-plasmo-voxel-ish
huggingface-txt2img-plasmo-woolitize
huggingface-txt2img-progamergov-min-illust-background-diffusion
huggingface-txt2img-prompthero-linkedln-diffusion
huggingface-txt2img-prompthero-openjourney
huggingface-txt2img-qilex-magic-diffusion
huggingface-txt2img-rabidgremlin-sd-db-epic-space-machine
huggingface-txt2img-rayhell-popupbook-diffusion
huggingface-txt2img-runwayml-stable-diffusion-v1-5
huggingface-txt2img-s3nh-beksinski-style-stable-diffusion
huggingface-txt2img-sd-dreambooth-library-original-character-cyclps
huggingface-txt2img-sd-dreambooth-library-persona-5-shigenori-style
huggingface-txt2img-sd-dreambooth-library-seraphm
huggingface-txt2img-shirayu-sd-tohoku-v1
huggingface-txt2img-thelastben-hrrzg-style-768px
huggingface-txt2img-timothepearce-gina-the-cat
huggingface-txt2img-trystar-clonediffusion
huggingface-txt2img-tuwonga-dbluth
huggingface-txt2img-tuwonga-rotoscopee
huggingface-txt2img-volrath50-fantasy-card-diffusion
huggingface-txt2img-yayab-sd-onepiece-diffusers4
model-txt2img-stabilityai-stable-diffusion-v1-4
model-txt2img-stabilityai-stable-diffusion-v1-4-fp16
model-txt2img-stabilityai-stable-diffusion-v2
✓ model-txt2img-stabilityai-stable-diffusion-v2-1-base

Deploying Stable Diffusion 2.1



```
File Edit View Run Kernel Git Tabs Settings Help gusnalwan / Personal Studio
Amazon_JumpStart_Text_To_I
+ ✂ 📄 📄 ▶ ■ ↺ ▶▶ Code ⌵ ⌚ git $ _ Cluster 🐛 Data Science Python 3 2 vCPU + 4 GiB Share ⌵
predictor_cls=Predictor,
name=endpoint_name,
env=env,
)

# deploy the Model. Note that we need to pass Predictor class when we deploy model through Model class,
# for being able to run inference through the sagemaker API.
model_predictor = model.deploy(
    initial_instance_count=1,
    instance_type=inference_instance_type,
    predictor_cls=Predictor,
    endpoint_name=endpoint_name,
)
```

2.3. Query endpoint and parse response

Generating Images

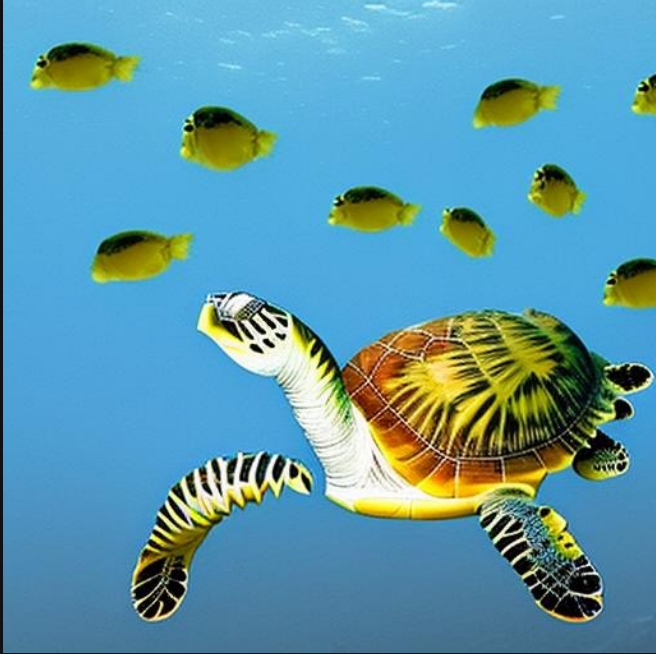
Home

Amazon_JumpStart_Text_To_I

Cluster

• [8] :

```
text = "a photo of a turtle swimming underwater surrounded by school of fish"
query_response = query(model_predictor, text)
img, prmppt = parse_response(query_response)
display_img_and_prompt(img, prmppt)
```



Challenges and Enhancement



Bob the penguin had always dreamed of traveling to Europe. He is tired of his routine life in Antarctica



One day Bob decided to take the plunge and begin planning this trip.



After a long boat ride, he checked into his hotel and eagerly headed to see the Eiffel Tower in Paris



Over the next few days, he visited famous landmarks like the Louvre Museum, Notre Dame Cathedral, and the Palace of Versailles

Character inconsistency

Penguin doll



Bob the penguin had always dreamed of traveling to Europe. He is tired of his routine life in Antarctica

Someone's leg



One day Bob decided to take the plunge and begin planning this trip.

Suited man



After a long boat ride, he checked into his hotel and eagerly headed to see the Eiffel Tower in Paris

Man in blue shirt



Over the next few days, he visited famous landmarks like the Louvre Museum, Notre Dame Cathedral, and the Palace of Versailles

Character inconsistency

Penguin doll



Bob the penguin had always dreamed of traveling to Europe. He is tired of his routine life in Antarctica

Someone's leg



One day *Bob* decided to take the plunge and begin planning this trip.

Suited man



After a long boat ride, *he* checked into his hotel and eagerly headed to see the Eiffel Tower in Paris

Man in blue shirt



Over the next few days, *he* visited famous landmarks like the Louvre Museum, Notre Dame Cathedral, and the Palace of Versailles

Scene out of focus

Generated



Desired



Generated



Desired



Bob the penguin had always dreamed of traveling to Europe. He is tired of his routine life in Antarctica

One day Bob decided to take the plunge and begin planning this trip.

Long paragraph and multiple focal points

Tackling scene out of focus

For each section, please describe the scene in detail and always include the location in one sentence within [] for example:

[a photo of character in the location]

*[a photo of Bob the penguin in Antarctica]
Bob the penguin had always dreamed of traveling to Europe. He is tired of his routine life in Antarctica*

*[a photo of Bob the penguin reading a book]
One day he decided to take the plunge and begin planning this trip*

*[a photo of Bob the penguin on a cruise ship]
After a long boat ride, he checked into his hotel and eagerly headed to see the Eiffel Tower in Paris*

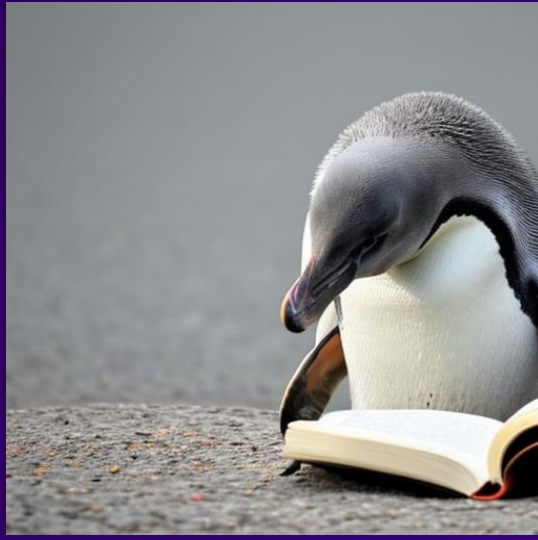
...

Better scene focus and character consistency



A photo of Bob the penguin in antarctica

Bob the penguin had always dreamed of traveling to Europe. He is tired of his routine life in Antarctica



A photo of Bob the penguin reading a book

One day he decided to take the plunge and begin planning this trip.



A photo of Bob the penguin on a cruise ship

After a long boat ride, he checked into his hotel and eagerly headed to see the Eiffel Tower in Paris



A photo of Bob the penguin at Eiffel tower

Over the next few days, he visited famous landmarks like the Louvre Museum, Notre Dame Cathedral, and the Palace of Versailles

Solution to Character in-consistency

Using Prompt Engineering
Always describe the character in detail

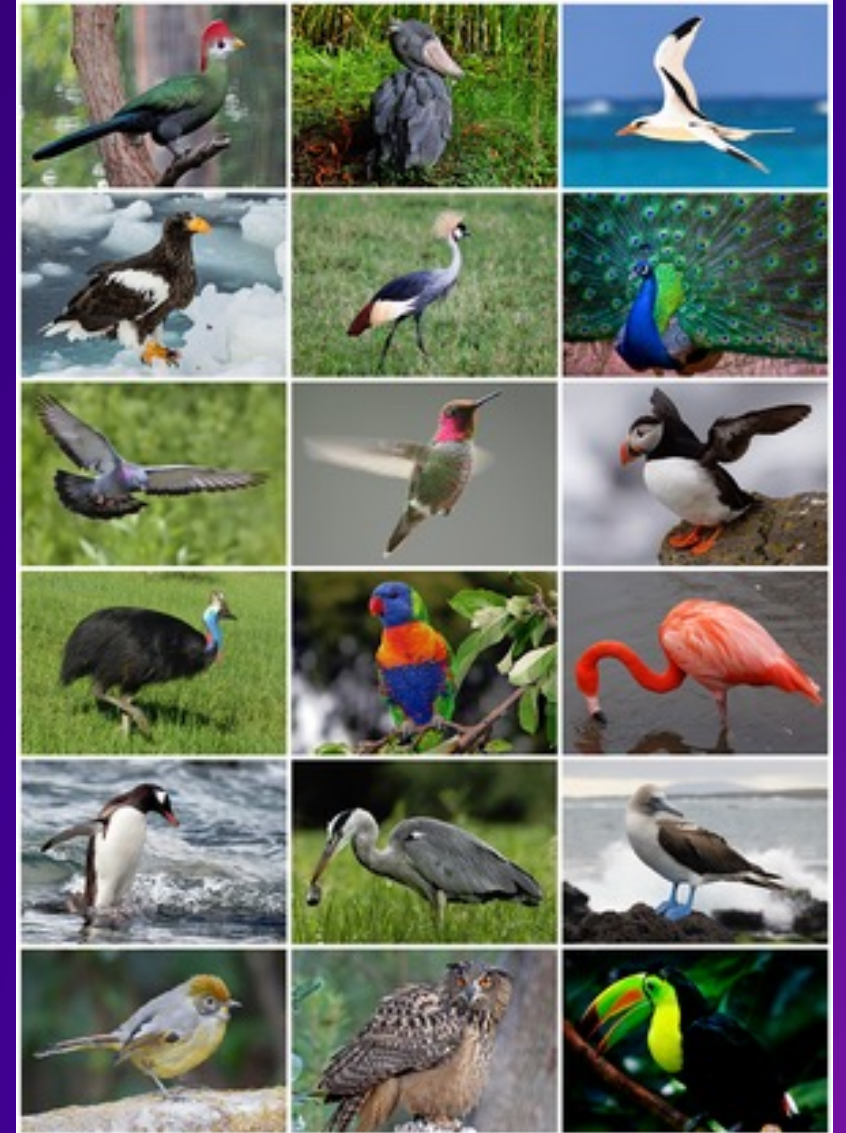
A photo of small, black
emperor penguin in in
Antarctica



Solution to Character in-consistency

Using Prompt Engineering
Always describe the character in detail

Birds - thousands of varieties



Problems

A photo of a small green parrot with yellow beak, white peri ophthalmic ring, yellow legs and red forehead standing on a tree

A photo of iridescent Morpho butterfly boasting large, brilliant blue wings, reflecting light like shimmering sapphires, while its undersides exhibit earthy hues and intricate patterns on a flower



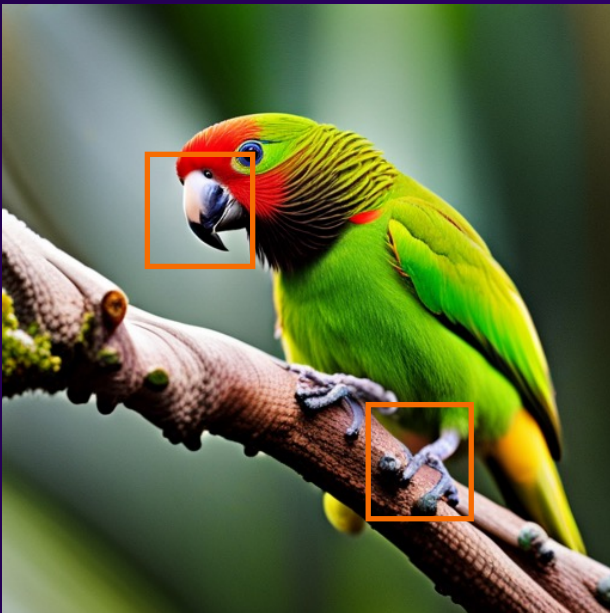
- Text overload



Problems

A photo of a small green parrot with **yellow beak**, white peri ophthalmic ring, **yellow legs** and red forehead standing on a tree

A photo of iridescent Morpho butterfly boasting large, brilliant blue wings, reflecting light like shimmering sapphires, while its undersides exhibit earthy hues and intricate patterns on a flower



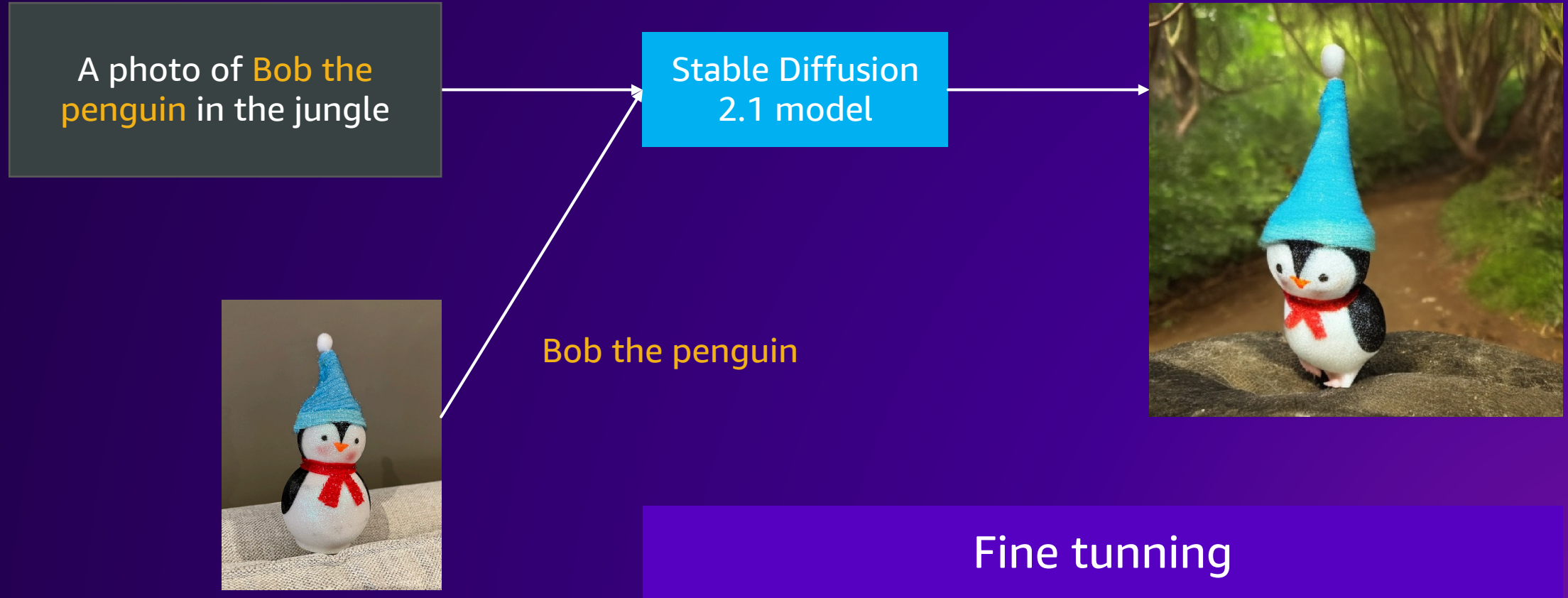
- Text overload
- Inaccurate image generation



What is the solution?

“Picture is worth a thousand words”

Solution



Fine Tuning Stable Diffusion Model

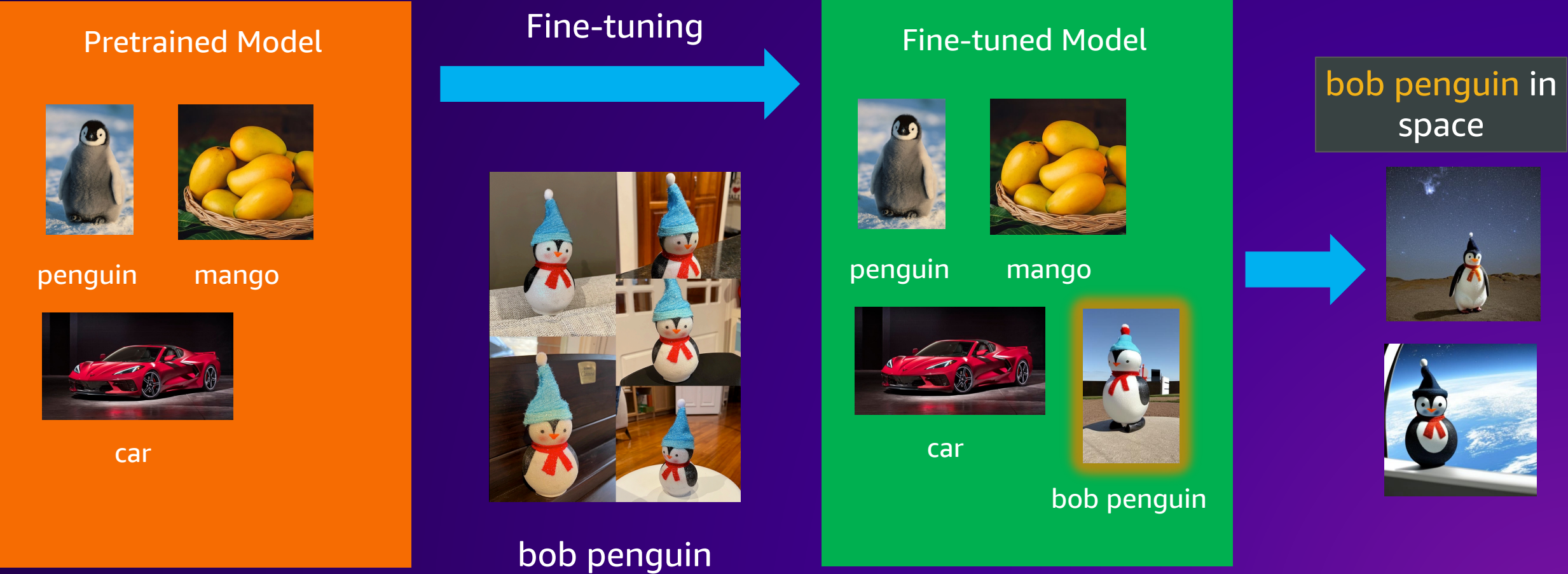
Stable Diffusion 2.1
Pretrained model

Stable Diffusion 2.1
Fine-tuned model

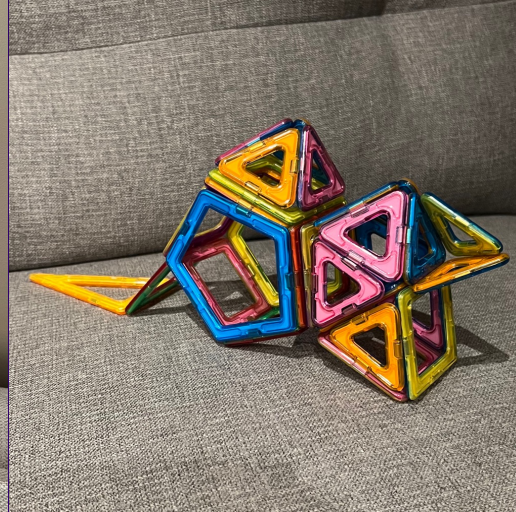
Fine-tuning



Learning a new concept



Unplanned cool feature!



Fine-tuning

```
input_directory
|---instance_image_1.png
|---instance_image_2.png
|---instance_image_3.png
|---instance_image_4.png
|---instance_image_5.png
|---dataset_info.json
|---class_data_dir
|   |---class_image_1.png
|   |---class_image_2.png
|   |---class_image_3.png
|   |---class_image_4.png
```

```
{
  "instance_prompt": "a photo of xyz penguin",
  "class_prompt": "a photo of penguin"
}
```



S3 bucket

Fine-tuning

```
[16]: from sagemaker.estimator import Estimator
      from sagemaker.utils import name_from_base
      from sagemaker.tuner import HyperparameterTuner

      training_job_name = name_from_base(f"stable-diffusion-{train_model_id}-transfer-learning")

      # Create SageMaker Estimator instance
      sd_estimator = Estimator(
          role=aws_role,
          image_uri=train_image_uri,
          source_dir=train_source_uri,
          model_uri=train_model_uri,
          entry_point="transfer_learning.py", # Entry-point file in source_dir and present in train_source_uri.
          instance_count=1,
          instance_type=training_instance_type,
          max_run=360000,
          hyperparameters=hyperparameters,
          output_path=s3_output_location,
          base_job_name=training_job_name,
      )

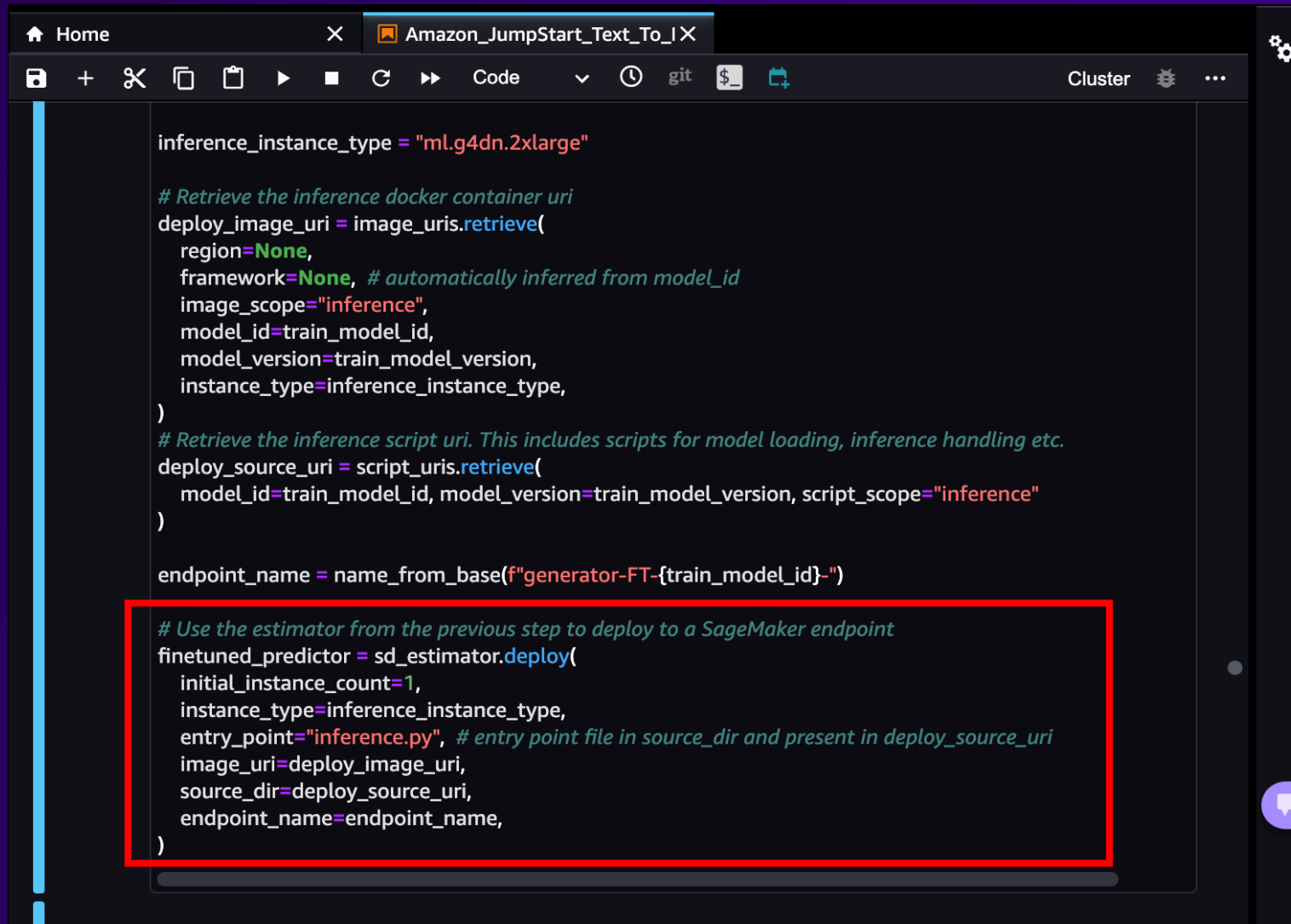
      # Launch a SageMaker Training job by passing s3 path of the training data
      sd_estimator.fit({"training": training_dataset_s3_path}, logs=True)

      ...

[17]: training_job_name = sd_estimator.latest_training_job.name
      training_job_name

[17]: 'stable-diffusion-model-txt2img-stabilit-2023-03-13-01-22-14-342'
```


Fine-tuning



```
inference_instance_type = "ml.g4dn.2xlarge"

# Retrieve the inference docker container uri
deploy_image_uri = image_uris.retrieve(
    region=None,
    framework=None, # automatically inferred from model_id
    image_scope="inference",
    model_id=train_model_id,
    model_version=train_model_version,
    instance_type=inference_instance_type,
)

# Retrieve the inference script uri. This includes scripts for model loading, inference handling etc.
deploy_source_uri = script_uris.retrieve(
    model_id=train_model_id, model_version=train_model_version, script_scope="inference"
)

endpoint_name = name_from_base(f"generator-FT-{train_model_id}-")

# Use the estimator from the previous step to deploy to a SageMaker endpoint
finetuned_predictor = sd_estimator.deploy(
    initial_instance_count=1,
    instance_type=inference_instance_type,
    entry_point="inference.py", # entry point file in source_dir and present in deploy_source_uri
    image_uri=deploy_image_uri,
    source_dir=deploy_source_uri,
    endpoint_name=endpoint_name,
)
```

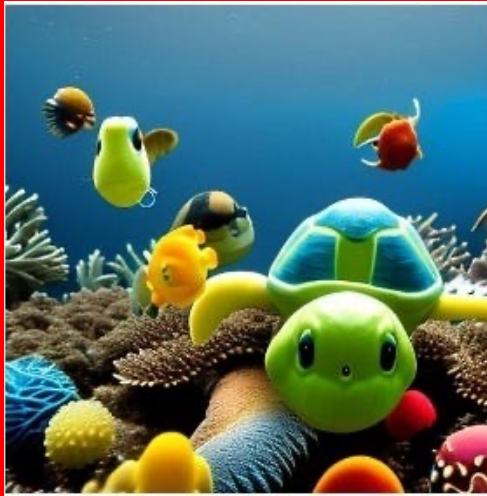
Fine-tuning

Training set



Nick the kangaroo is sitting
next to the lake

Rex the turtle is swimming
around school of fish



Training set



Failure examples

Tom the hedgehog is milking the cow

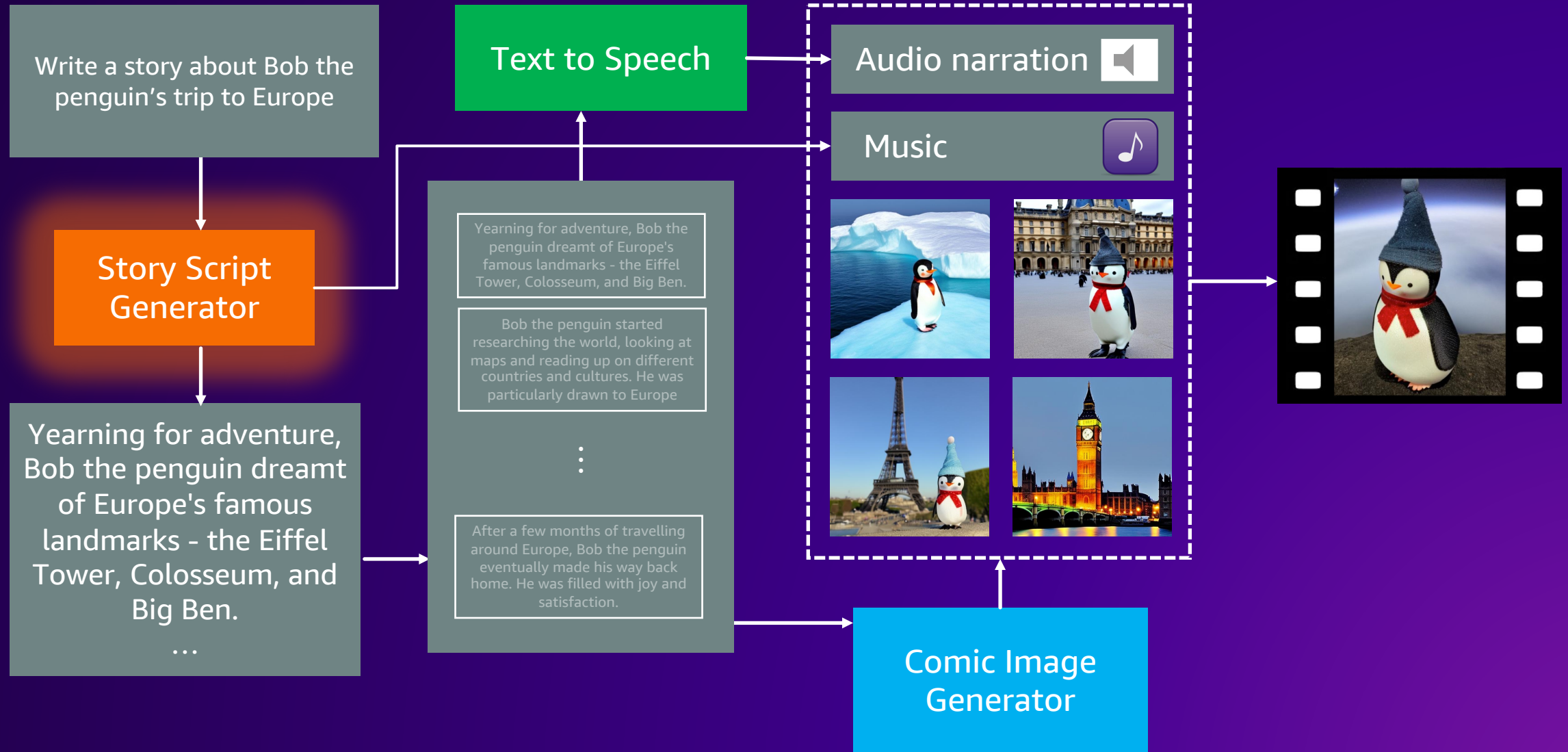


Tom the hedgehog is carrying a flower bouquet

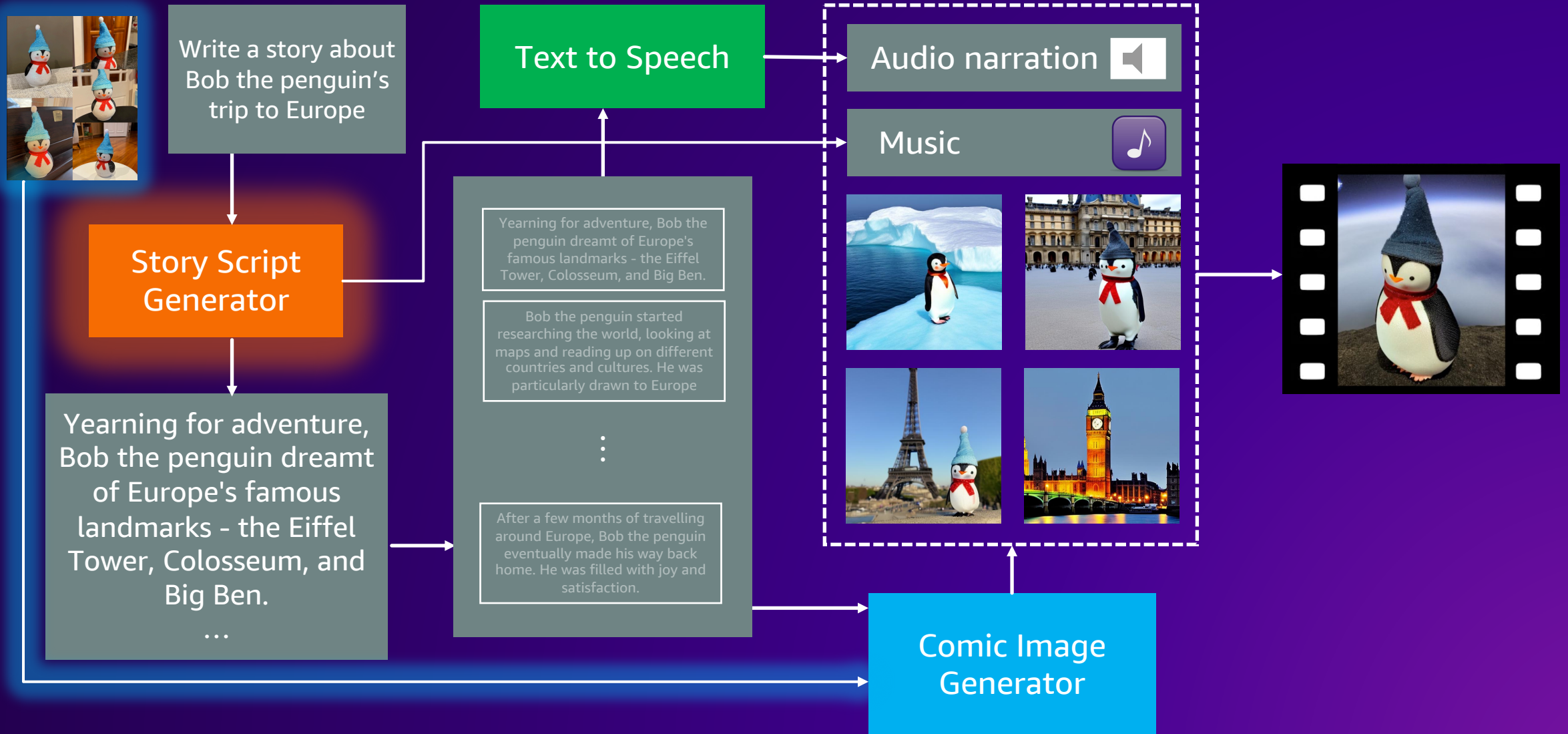


Tom the hedgehog

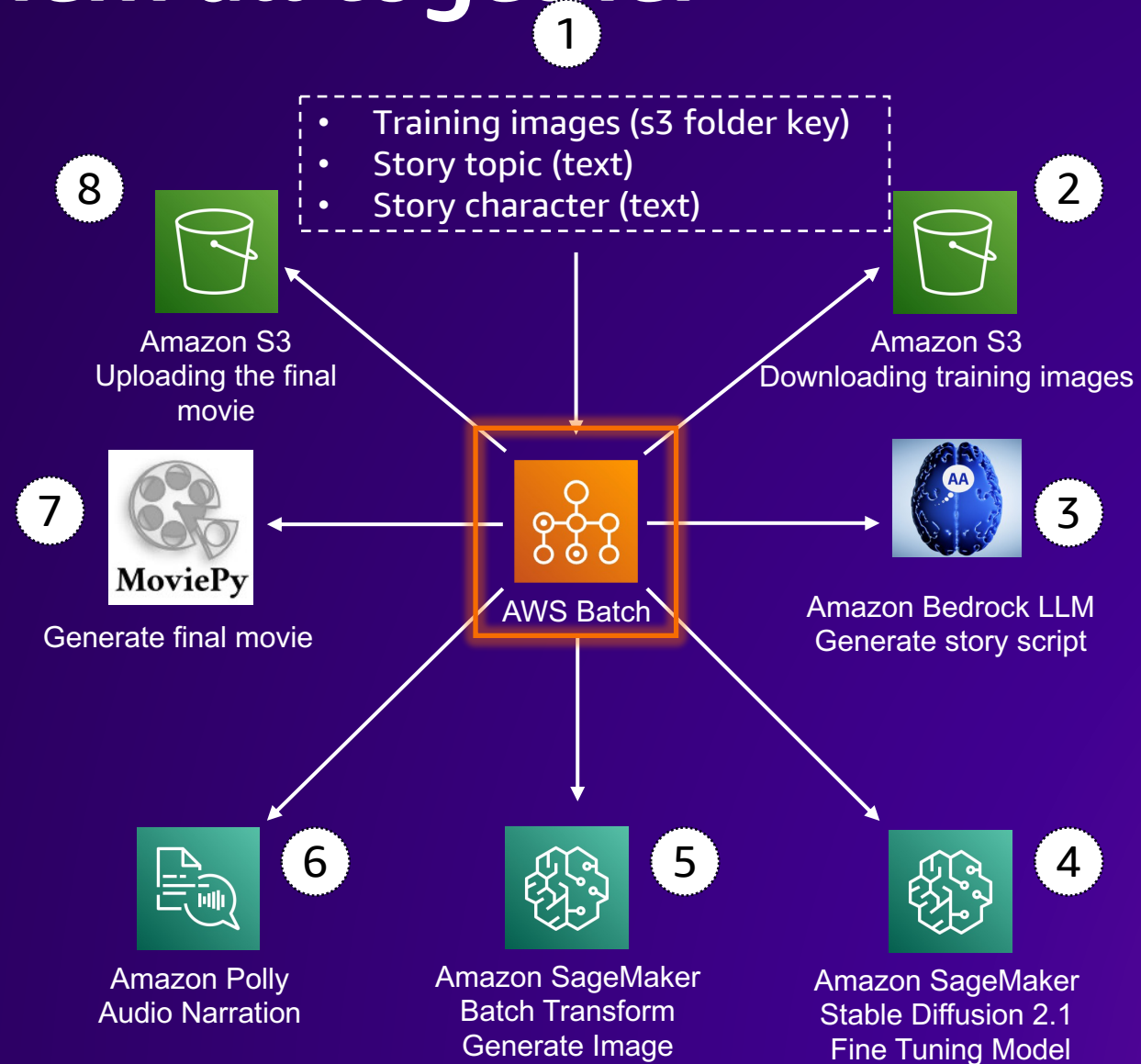
Initial architecture



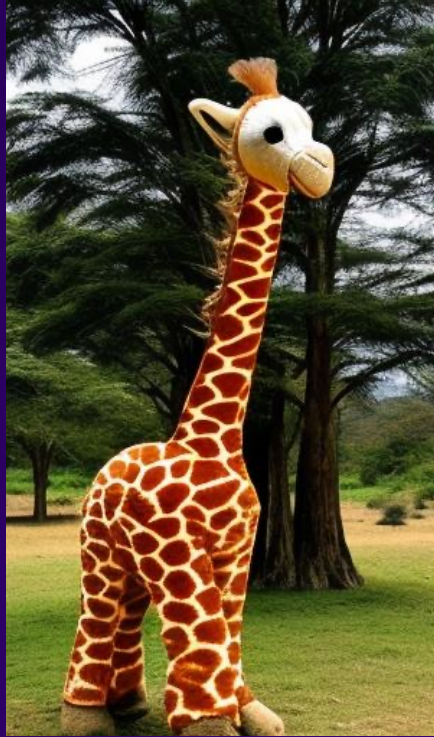
Final architecture



Stitching them all together



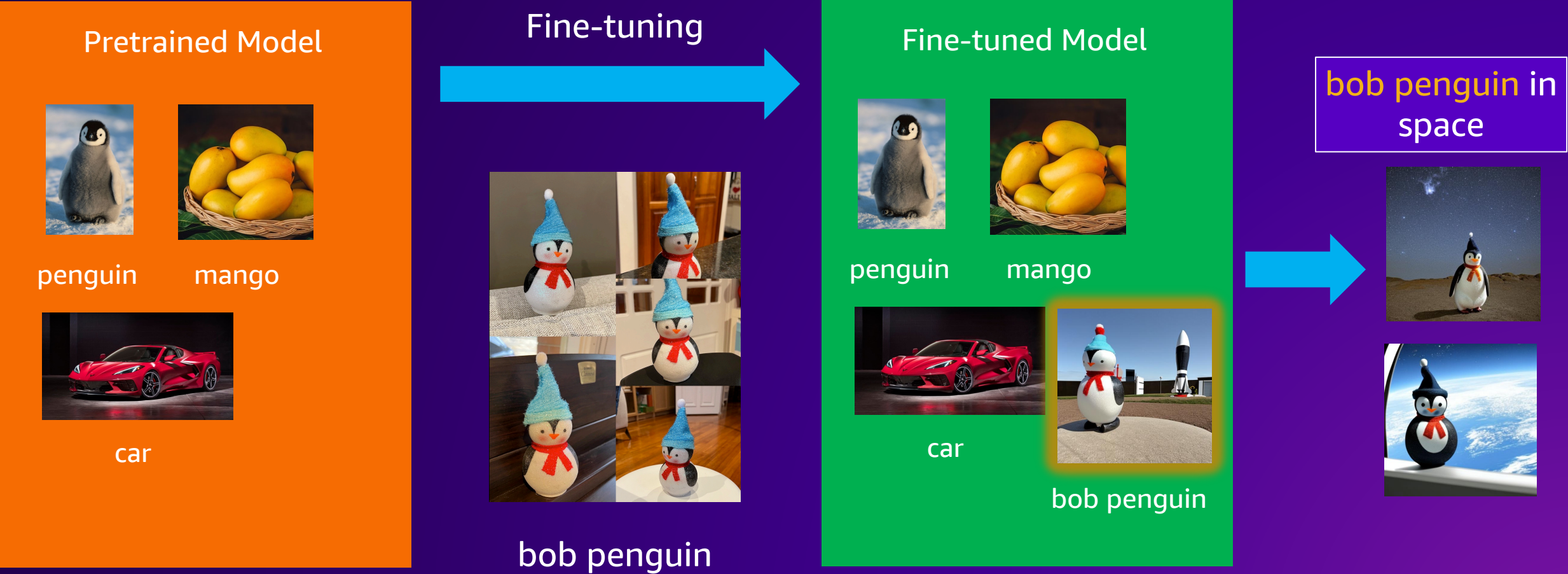
Generated Stories



<https://shorturl.at/stACV>

Something cool – Mixing images together

Learning a new concept



Overriding an existing concept

Pretrained Model



cat

Fine-tuning



cat

Fine-tuned Model



cat

Dancing cat



Cool things you can do



Thank You

Agustinus Nalwan

● agustinus-nalwan

 [linkedin.com/in/agustinus-nalwan](https://www.linkedin.com/in/agustinus-nalwan)



To view all my projects