Generative AI Story & Illustration Creator
利用生成式AI創作故事與插圖

郭舒文 NTNU_010153217　　馮意凡 NTU_R10521609

Github:
https://github.com/Evan102/Generative-AI-Story-Illustration-Creator/tree/main

Outline

- Motivation
- Literature review
- Objectives
- Story generation
- Illustration generation
- Conclusion & Platform demo

Motivation

Listen

See all 2 images

Follow the Author

Ammaar Reshi — Follow

Alice and Sparkle Paperback – January 14, 2023

by Ammaar Reshi (Author), Chat GPT (Author), Mid Journey (Illustrator)

3.0 ★★★☆☆ 79 ratings — See all formats and editions

Kindle	Audiobook	Hardcover	Paperback
$4.99	$0.00	$38.00	$9.13
Read with Our Free App	Free with your Audible trial	3 New from $24.74	1 Used from $7.60 1 New from $9.13

The children's book covered in The Washington Post, TIME Magazine, Le Figaro, NBC News, and media worldwide.

This is a story about a young girl named Alice who discovers the magic of artificial intelligence. She creates her own AI, named Sparkle, and together they go on adventures and use their combined knowledge to make the world a better place. The story explores the incredible abilities of AI and the importance of using them for good. It is a tale of friendship and exploration, filled with magic and wonder.

Alice and Sparkle is a children's story at heart that hopes to inspire children, encourage their curiosity and learning, in one of the most technologically exciting moments in our lifetime.

This book was also co-created using AI tools to both aid in writing and illustration and is the first of its kind to do so.

Soon after its release, Alice and Sparkle started an important discussion around how the creators of AI image generation tools have a responsibility to protect artists and their work. The hope is that this leads to a series of empowering tools that are built responsibly, benefiting and enabling a new set of creators while also protecting those that inspired all of them.

Read less

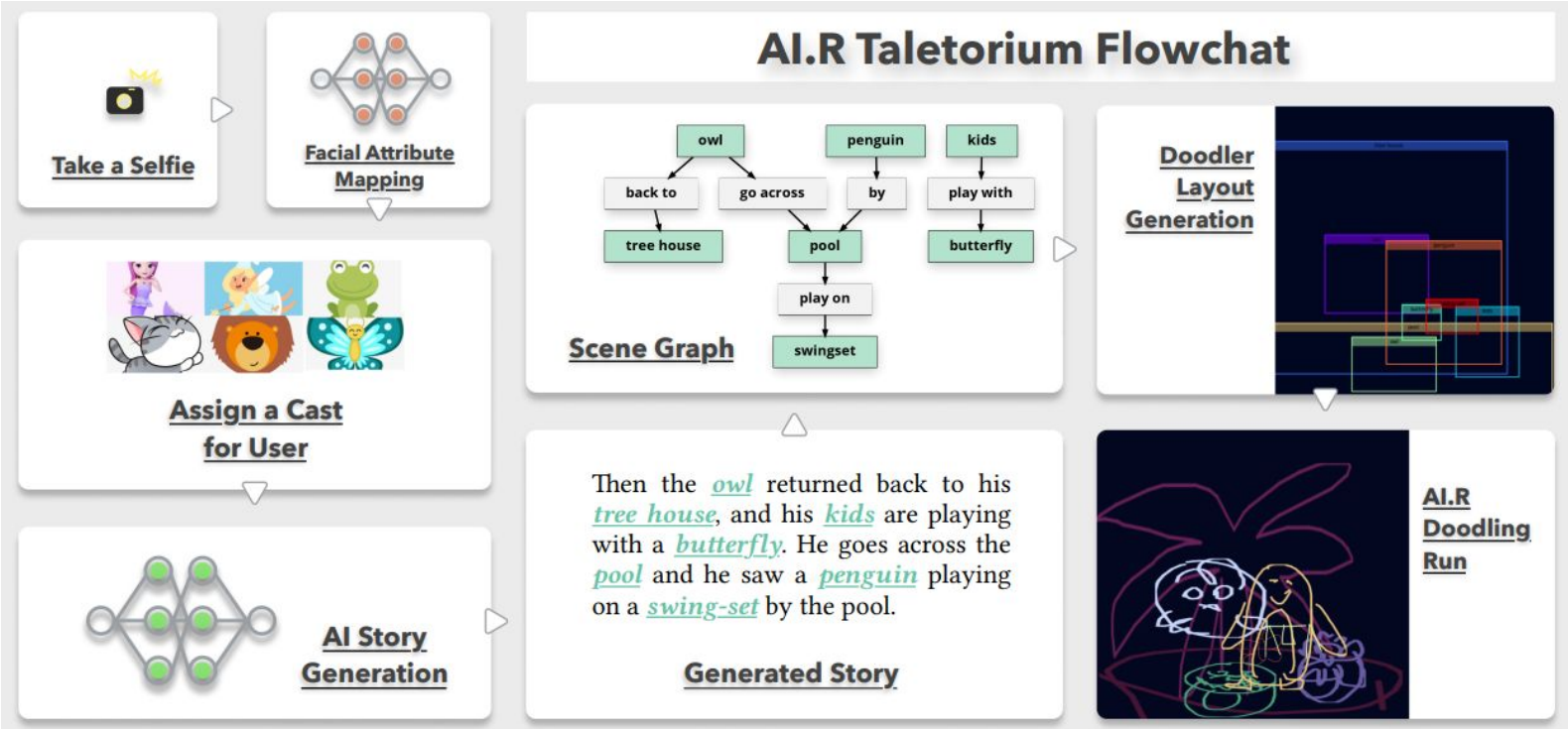
Print length	Language	Publication date	Dimensions	ISBN-13
26 pages	English	January 14, 2023	8 x 0.07 x 10 inches	979-8373324885

See all details

Until 2023/6/3,

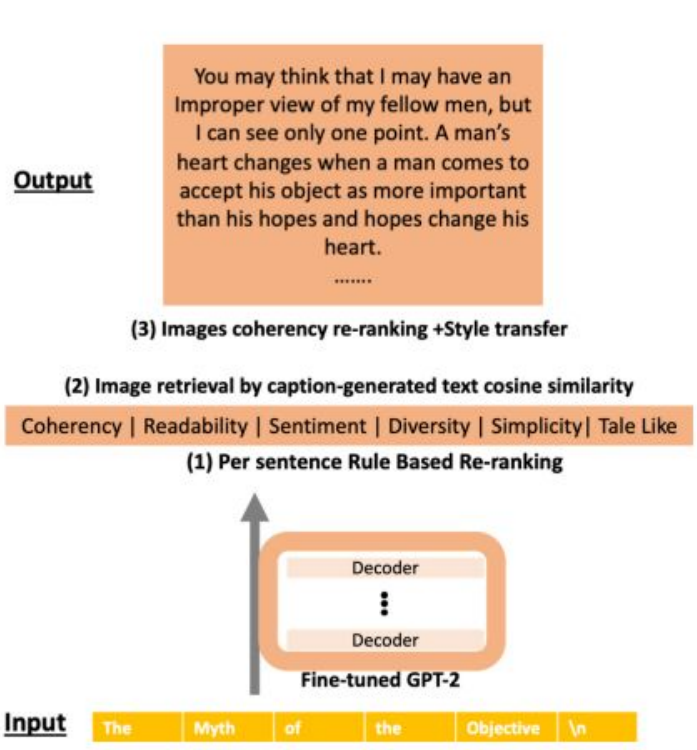
793 books co-authored with chatGPT, 78 books illustrated by Mid Journey.

Literature review



Reference: Mutltimodal AI Companion for Interactive Fairytale Co-creation

Literature review



The system generates text while re-ranking, retrieves images from Unsplash dataset, applies **style transfer** and then re-ranks stories according to the story's visual consistency.

Reference: FAIRYTAILOR: A MULTIMODAL GENERATIVE FRAMEWORK FOR STORYTELLING

Objectives

1. Empower students to experience generative AI's 'creativity' in storytelling through adaptation and creation of stories with illustrations.
2. Leverage a large language model to design story and illustration prompts or fine-tune the model for story creation and illustration.
3. Compare the results of different models for the same prompt.

Story Generation

- Methodology
- Prompt Evolution Process

Methodology

1. Using the ChatGPT API
   a. We connected to the ChatGPT via its API, creating prompts for children's books.
2. Prompt Examples
   a. Students can generate stories directly using simple keywords or detailed descriptions.
3. Language and Token Limitation
   a. As text-to-image generation is more appropriate with English stories
   b. Shortened them to fit the model's 50-token requirement.
4. Web Application Development -Gradio

:Padlet

郭舒文 • 大約1分鐘

302 Padlet

老師示範

中文的故事

先把AI協作的故事貼在這裡，依照自己的想法修調整故事內容

對AI下指令的技巧：
1. 設定AI助理的角色：假設你是兒童故事的編輯，請協助我修改故事。
2.輸入關鍵字：(故事的描述，越詳細越好)
3.輸出的要求：請用「100字」完成一個「適合10歲兒童」閱讀的「中文故事」。
4.如果產生結果不滿意，可以要求AI繼續調整。但必須具體說明，例如：加入新角色小強、改為快樂的結局等等。

♡0

英文故事繪圖

如果要使用AI繪圖功能，需要將故事翻譯成簡短的英文故事，你可以利用AI協助翻譯。

對AI下指令的技巧：
「請將以下內容翻譯成英文。
--- (利用斷行或分隔線)
需要翻譯的內容」

♡0

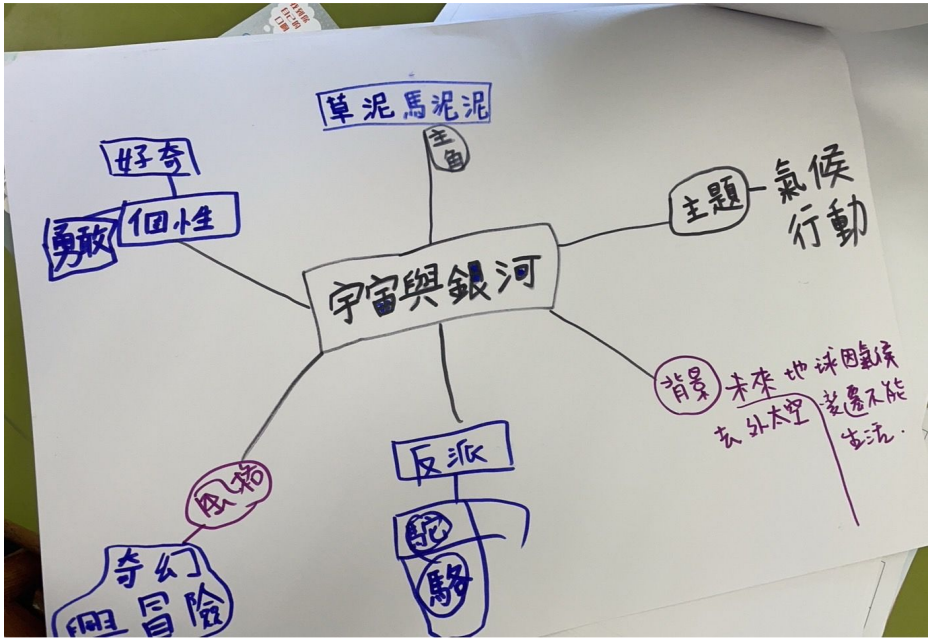
第1組

班級/姓名/座號 - 故事主題

小小企鵝 Cosmo，渴望探索宇宙的奧妙。他在一次研究夜空時，發現了一顆神秘的星星，聽說暗藏著一個宇宙秘密。Cosmo 心懷冒險之心，決定踏上尋找之旅。在旅途中，他結識了來自不同星系的夥伴，大家一起克服了許多難關，終於找到了神秘的星星。他們發現，這顆星星其實是一個友善的太空生物，他們一起享受了屬於青春、冒險和宇宙的美好時光。

Little penguin Cosmo longed to explore the mysteries of the universe. During one study of the night sky, he discovered a mysterious star rumored to hold a cosmic secret. With a heart full of adventure, Cosmo decided to embark on a journey to find it. Along the way, he made friends

第2組



Prompt Evolution Process

Before

Step 1: Create a Chinese story

Step 2: Translate the Chinese story into English

Step 3: Extract keywords for illustration

After

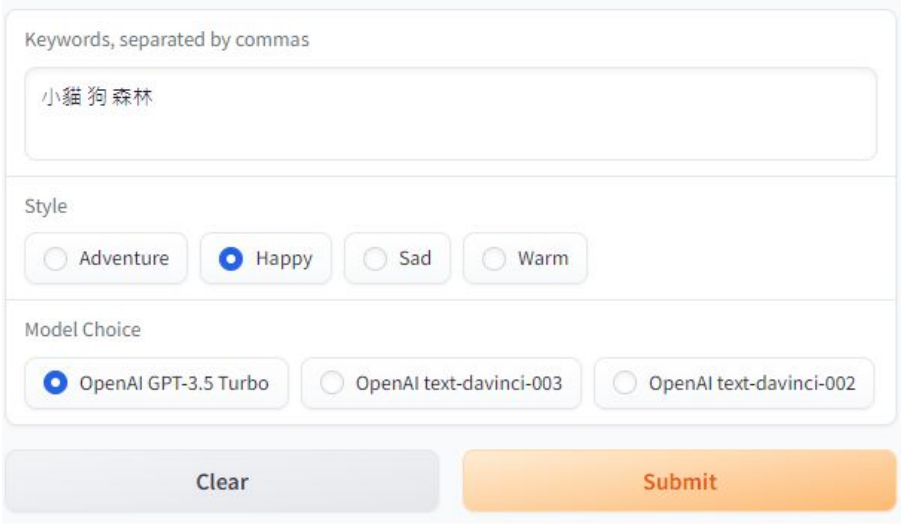
Give ChatGPT a role , input keywords, and specify requirements for the output format (e.g., token limits, language).

Input: Chinese or English keywords

Output: English translation (50-word token summary)

Tip: Prompt allows for the use of mixed Chinese and English languages.

```
prompt = f"As a {role}, Please write a story of around 200 words in Traditional Chinese , Story condensed into 50 tokens in English based on the following keywords and style, style:  {style}, keywords: {', '.join(keywords)} ."
```

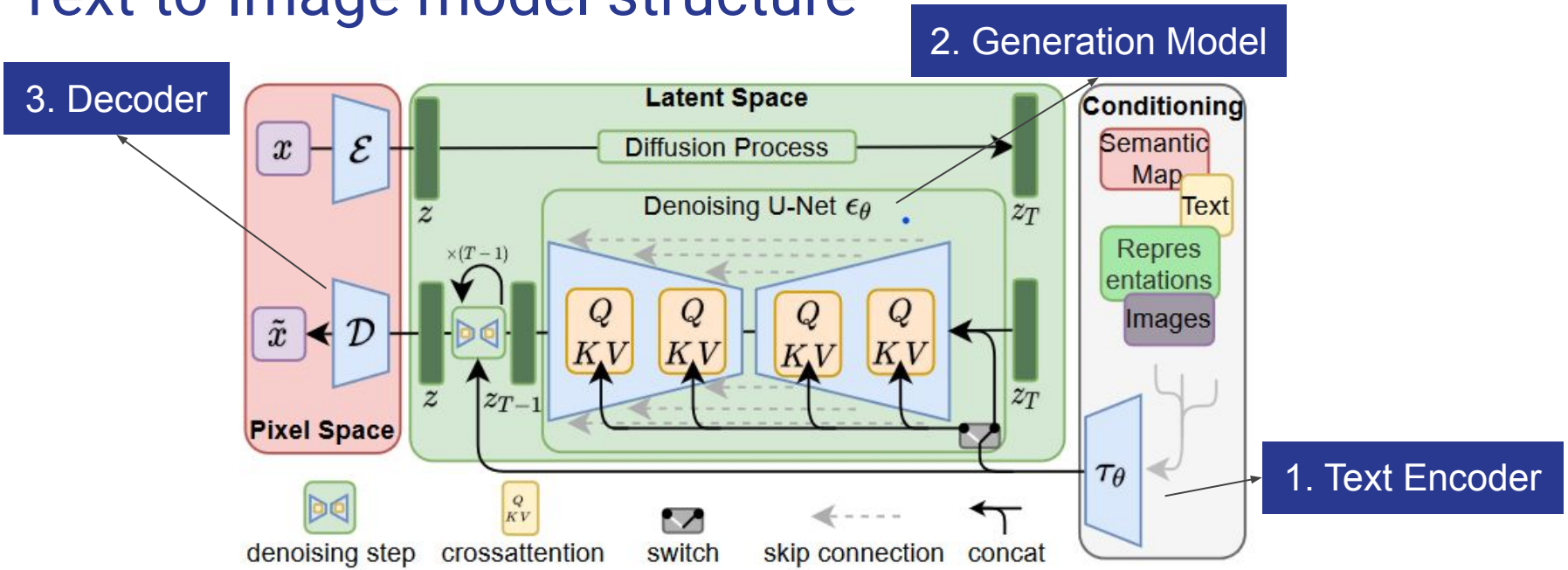


https://colab.research.google.com/drive/1ThrEesio7C7sZ0Hj7MeIstxwiaW0pGHq?usp=sharing

Illustraion Generation

- Text-to-Image model structure
- Low-Rank Adaptation(LoRA) introduction
- Comparison of Text-to-Image model prediction

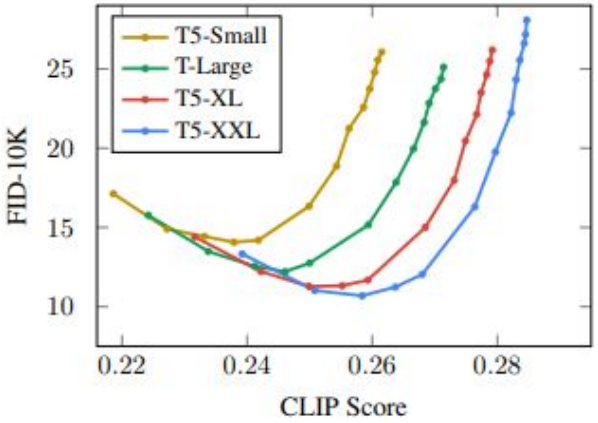
Text-to-Image model structure



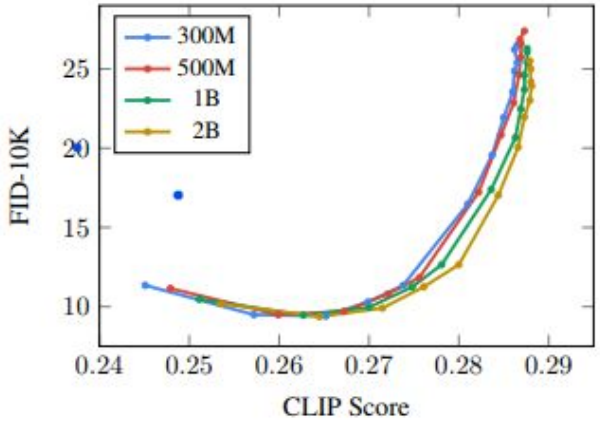
Reference: [High-Resolution Image Synthesis with Latent Diffusion Models]

Text-to-Image model structure - 1. Text Encoder

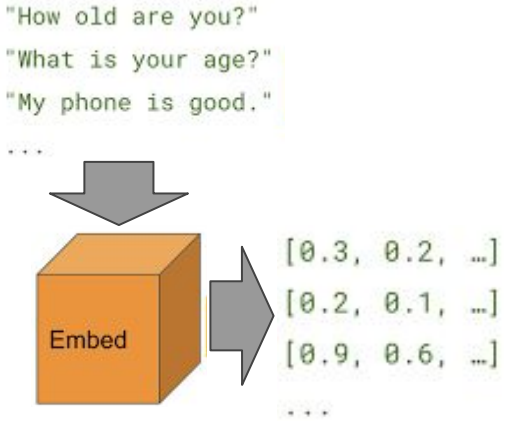
ClipText (A GPT-based model) or BERT can be used as a text encoder.



(a) Impact of encoder size.

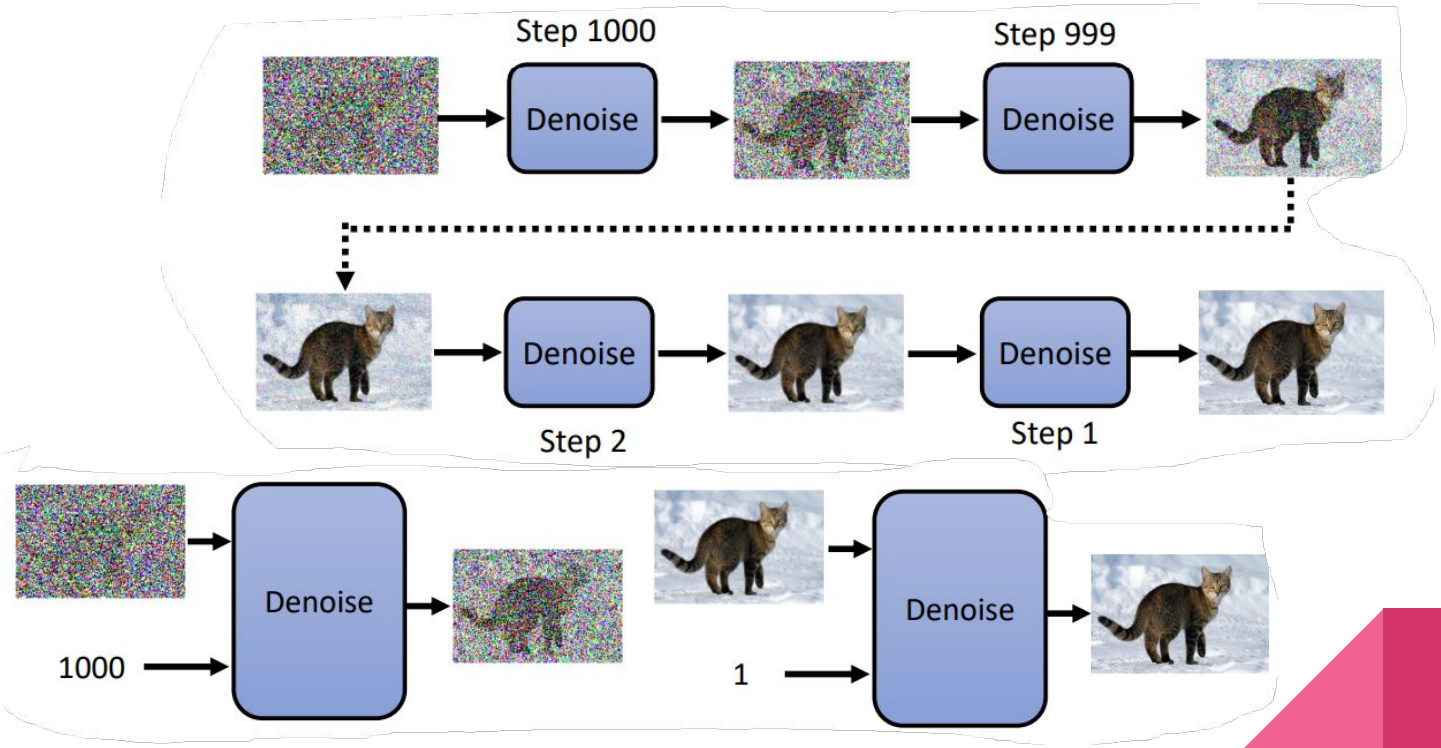


(b) Impact of U-Net size.



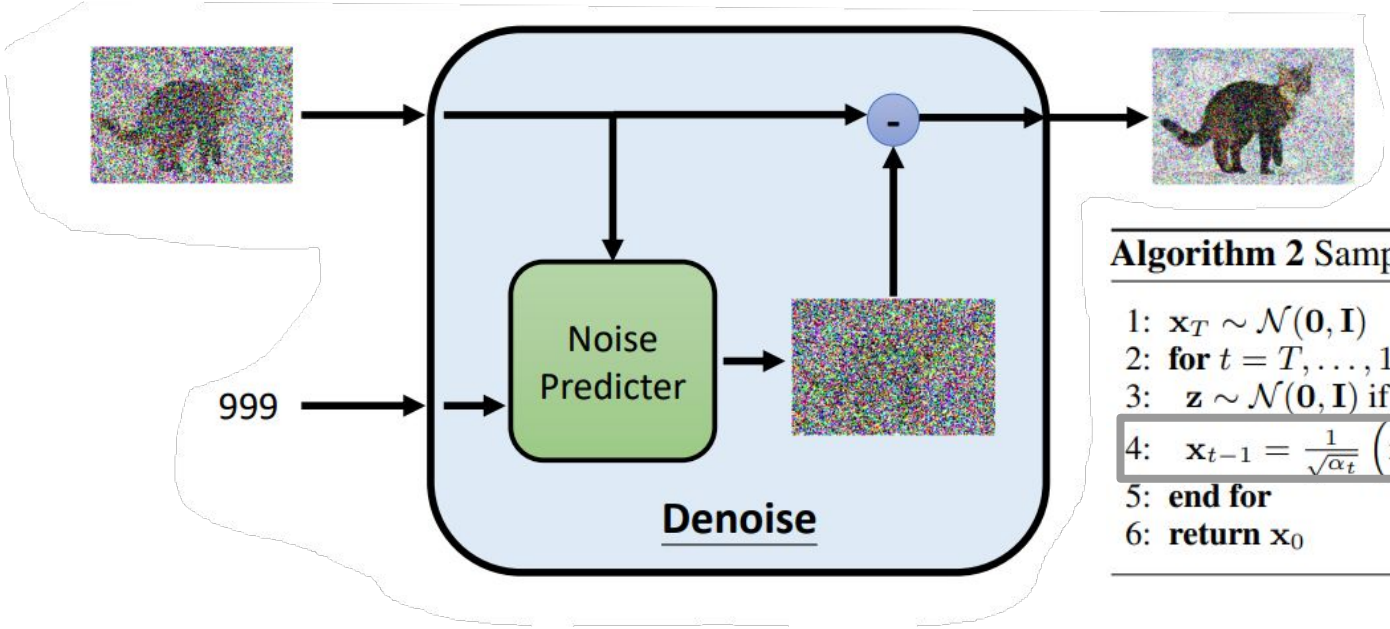
Reference: [Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding]

Text-to-Image model structure - 2. Generation Model



Reference: [HUNG-YI LEE ML 2023 spring course - Diffusion Model]

Text-to-Image model structure - 2. Generation Model



Algorithm 2 Sampling

1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \ldots, 1$ **do**
3: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
4: $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$

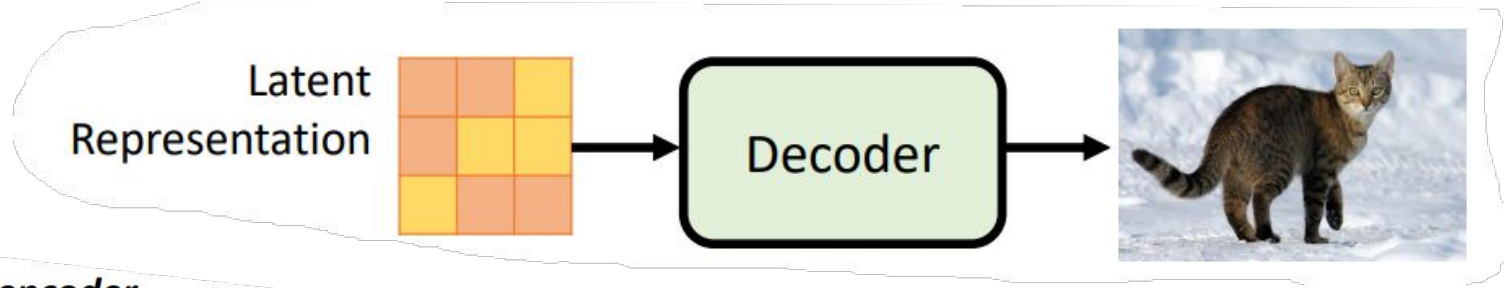
Reference:

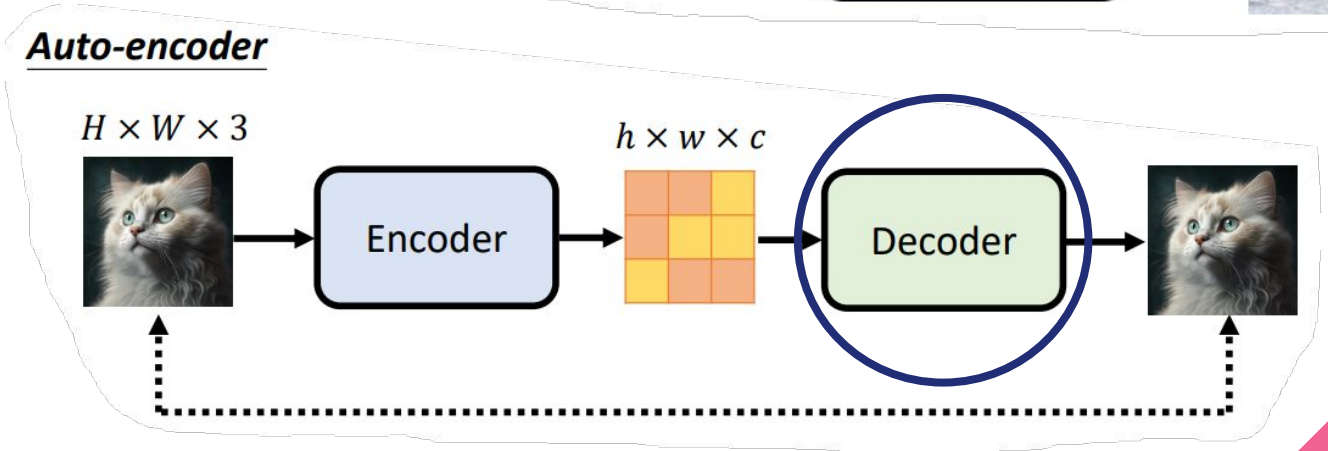
[1] HUNG-YI LEE ML 2023 spring course - Diffusion Model
[2] Denoising Diffusion Probabilistic Models

Text-to-Image model structure - 2. Generation Model

Reference: https://jalammar.github.io/illustrated-stable-diffusion/

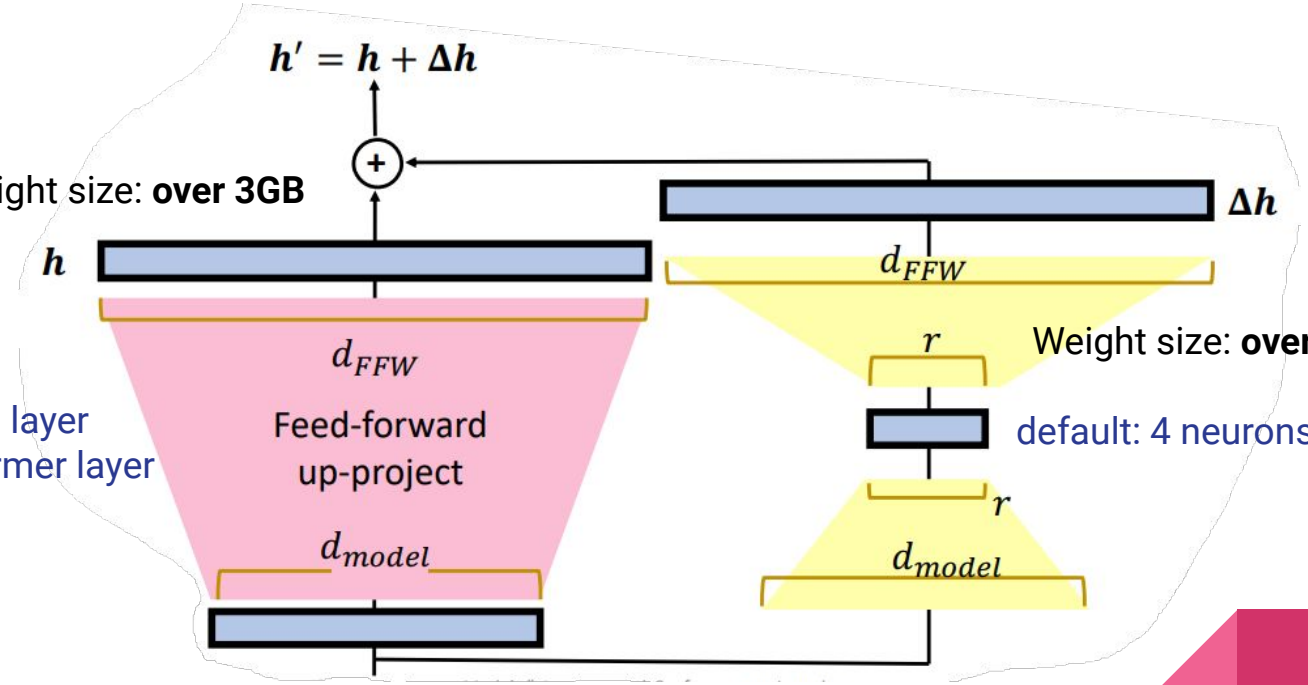
Text-to-Image model structure - 3. Decoder





Reference: [HUNG-YI LEE ML 2023 spring course - Stable Diffusion]

Low-Rank Adaptation(LoRA) introduction



Feed forward layer
from transformer layer

Reference:

[1] HUNG-YI LEE AACL 2022 Tutorial

[2] LORA: LOW-RANK ADAPTATION OF LARGE LANGUAGE MODELS

Comparison of Text-to-Image model prediction

- Story content:
  Fox Xiaohua and Turtle Aman had a race in the forest. Xiaohua ran fast but was overconfident and took a nap. Aman persisted and eventually won.
- Prompt1: '{story content}'
  Prompt2: 'manga, details line without color style: {story content}'
  Prompt3: 'manga, details line without color style: {story content} – niji'
  → Goal: the style of the illustrations from the three prompts can be identical.
- LoRA Fine-tune dataset: lambdalabs/pokemon-blip-captions(833 pokemon images with text)

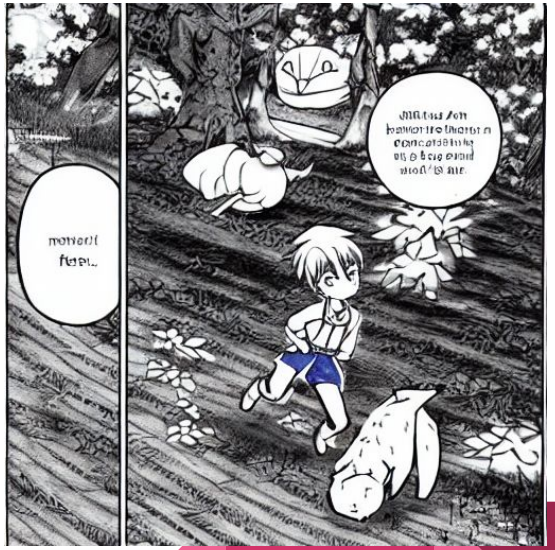
image (image)	text (string)
	"a drawing of a green pokemon with red eyes"

CompVis/stable-diffusion-v1-4 model prediction

Prompt1

Prompt2

Prompt3

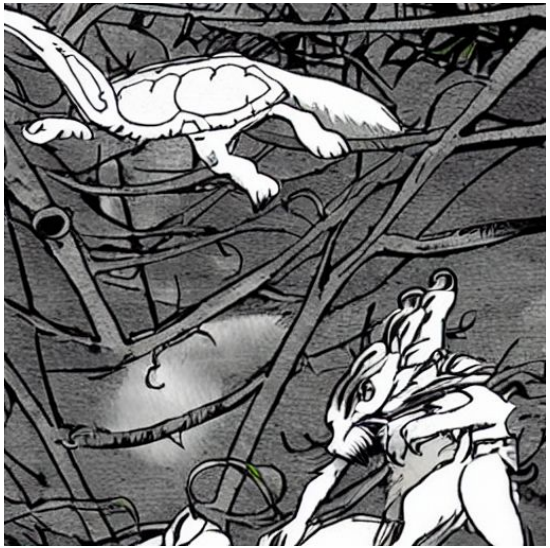


The first prompt is highly different from the other two prompts.

CompVis/stable-diffusion-v1-4 add pokemon LoRA

Prompt1

Prompt2



Prompt3

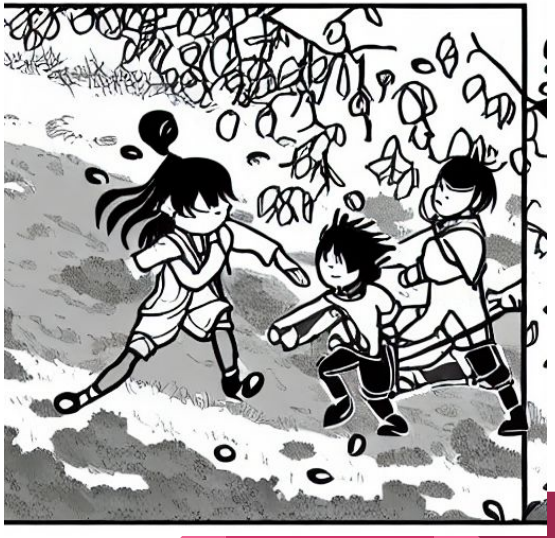
The style changes, but the content is weird.

CompVis/stable-diffusion-v1-4 add Children's Drawing LoRA

Prompt1

Prompt2

Prompt3



The style changes, but the content is weird.

Children's Stories model prediction

Prompt1

Prompt2

Prompt3

The style is similar, but combine human and animal.

gsdf/Counterfeit-V2.5 model prediction

Prompt1

Prompt2

Prompt3

The style is similar,
but some content combine human and animal.

gsdf/Counterfeit-V2.5 add pokemon LoRA

Prompt1

Prompt2

Prompt3

Only animals are in the three images,
and the style is quite the same!

CompVis/stable-diffusion-v1-4 pokemon LoRA training process

validation prompt="Fox Xiaohua and Turtle Aman had a race in the forest. Xiaohua ran fast but was overconfident and took a nap. Aman persisted and eventually won."

Before training

epoch 1

epoch 5

gsdf/Counterfeit-V2.5 pokemon LoRA training process

validation prompt="manga, details line without color style: Fox Xiaohua and Turtle Aman had a race in the forest. Xiaohua ran fast but was overconfident and took a nap. Aman persisted and eventually won.--niji"

Before training

epoch 1

epoch 5

Conclusion

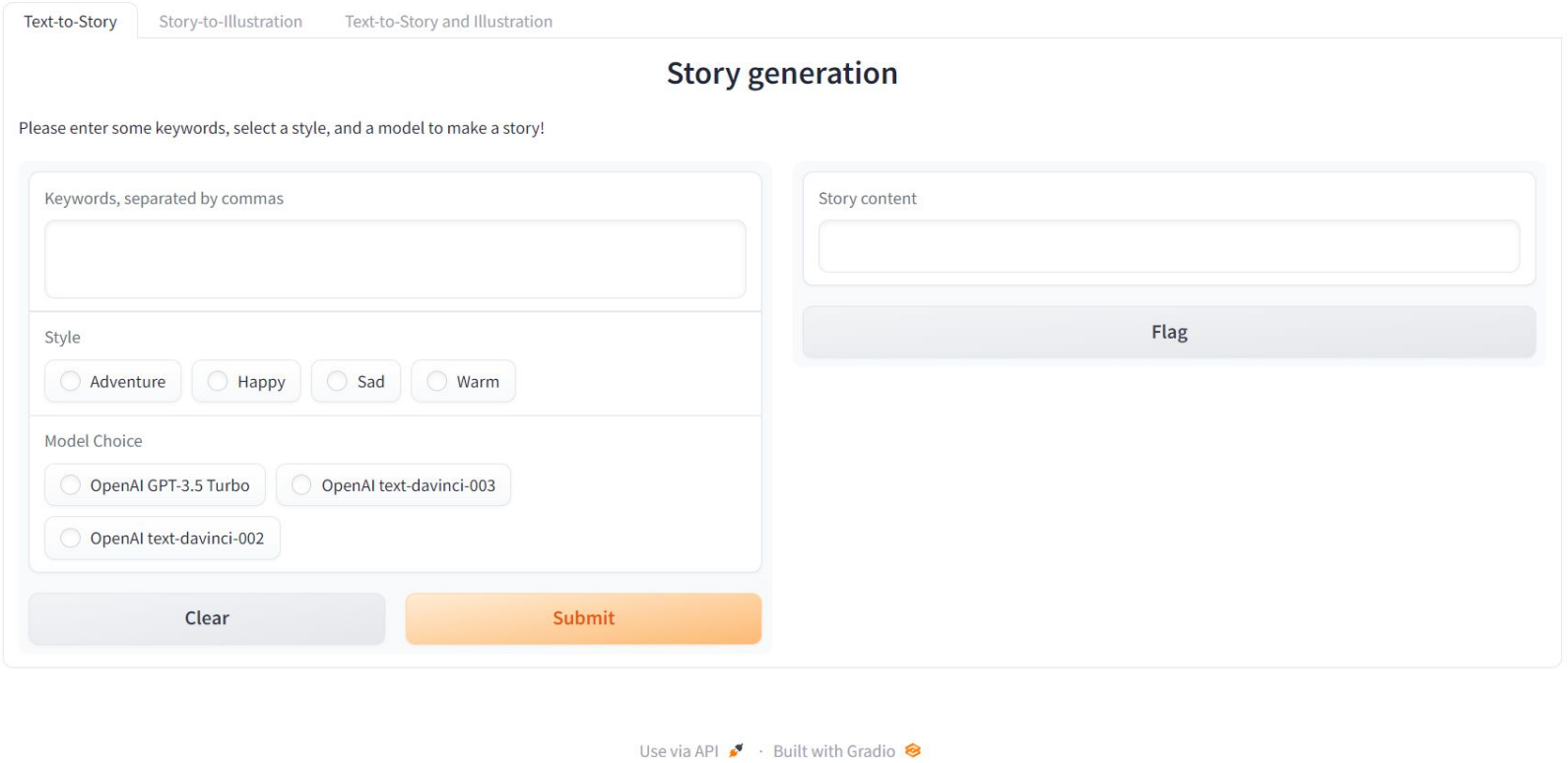
1. When generating stories with ChatGPT, the same keywords would produce different stories each time.

2. Unlike the web version, the API doesn't remember past conversation. Thus, to continue a story, previous context is needed in the prompt. Eventually, the prompt will exceed the API token limit.

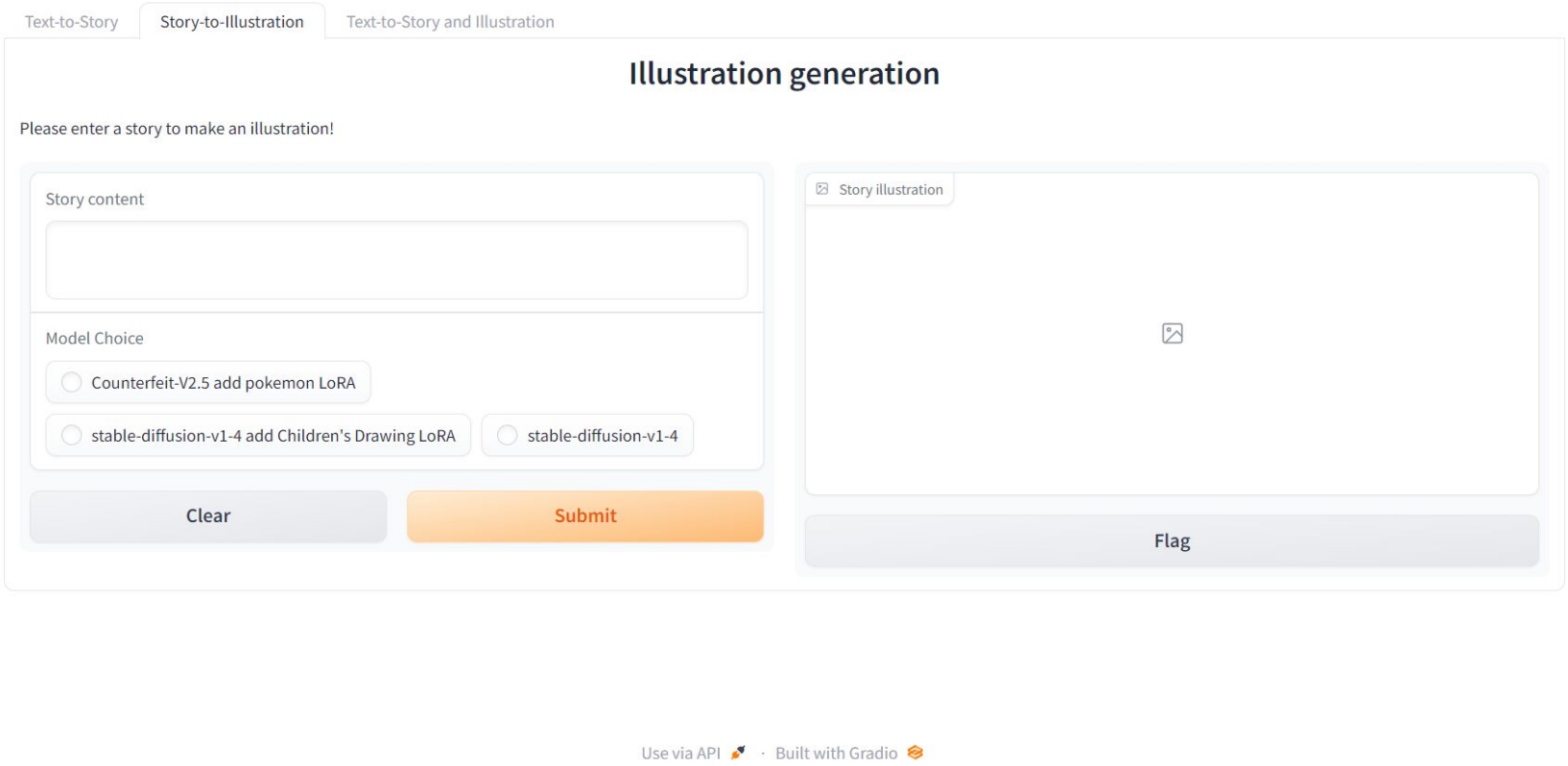
3. LoRA fine-tuning is helpful to change the generation image style from text-to-image model in just one epoch.

4. It is difficult to generate continuous illustrations or two characters for a story via one text-to-image model. Models for Inputs **including text and image** might be a solution for generating continuous illustrations.

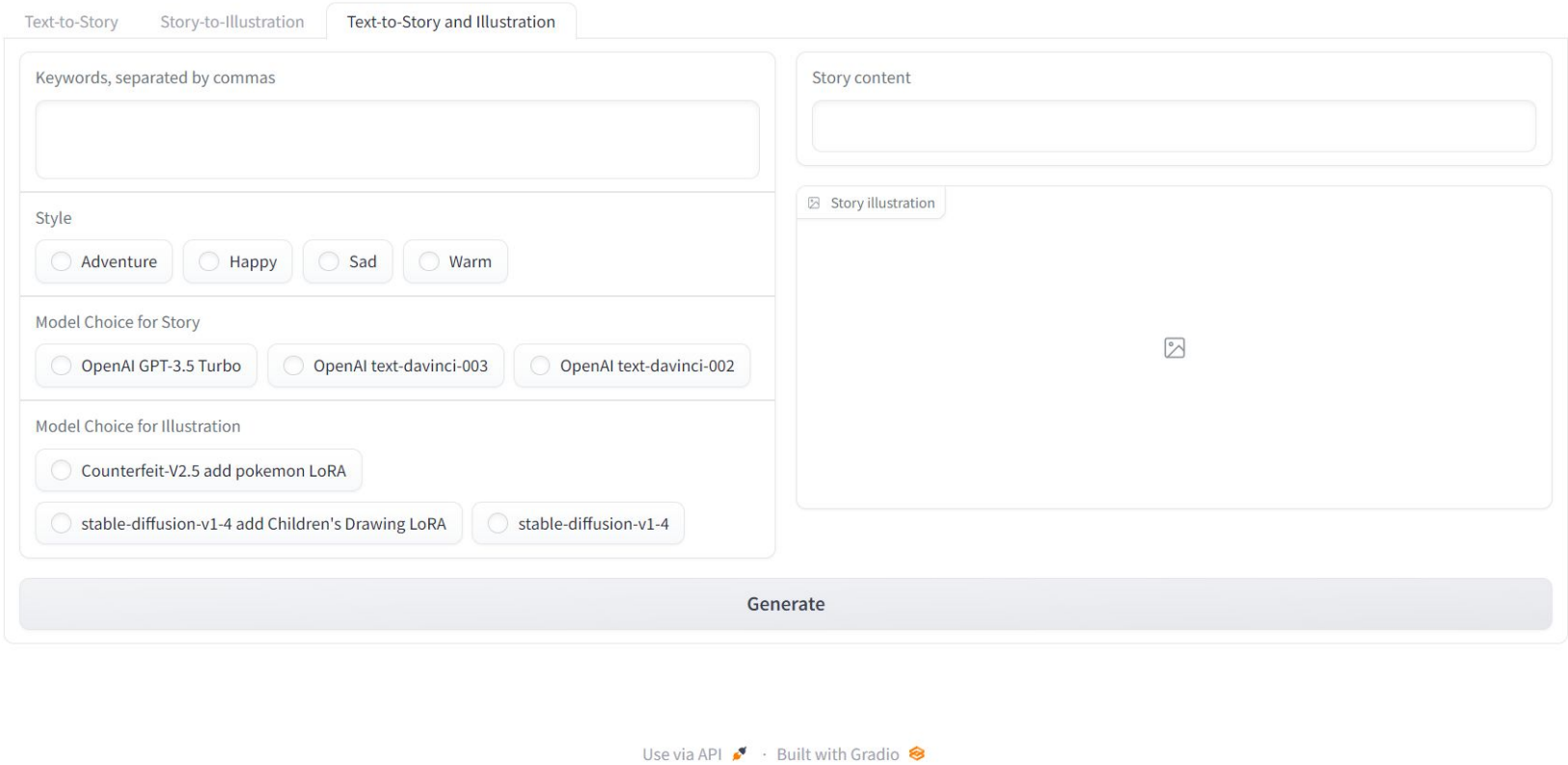
Platform demo



For creating story



For creating illustration from a story



Creating story and illustration