PREDICTIVE ANALYSIS

Project- Report

TEXT TO IMAGE CONVERTER

Name:

Aastha Prakash :500101844

Aditi Mupalla: 500105986

CODE:

```
import tkinter as tk

import customtkinter as ctk


from PIL import ImageTk

from authtoken import auth_token


import torch

from torch import autocast

from diffusers import StableDiffusionPipeline


# Create the app

app = tk.Tk()

app.geometry("532x632")

app.title("Stable Bud")

ctk.set_appearance_mode("dark")


prompt = ctk.CTkEntry(height=40, width=512, text_font=("Arial", 20), text_color="black",
fg_color="white")

prompt.place(x=10, y=10)


lmain = ctk.CTkLabel(height=512, width=512)

lmain.place(x=10, y=110)


modelid = "CompVis/stable-diffusion-v1-4"

device = "cuda"

pipe = StableDiffusionPipeline.from_pretrained(modelid, revision="fp16",
torch_dtype=torch.float16, use_auth_token=auth_token)

pipe.to(device)
```

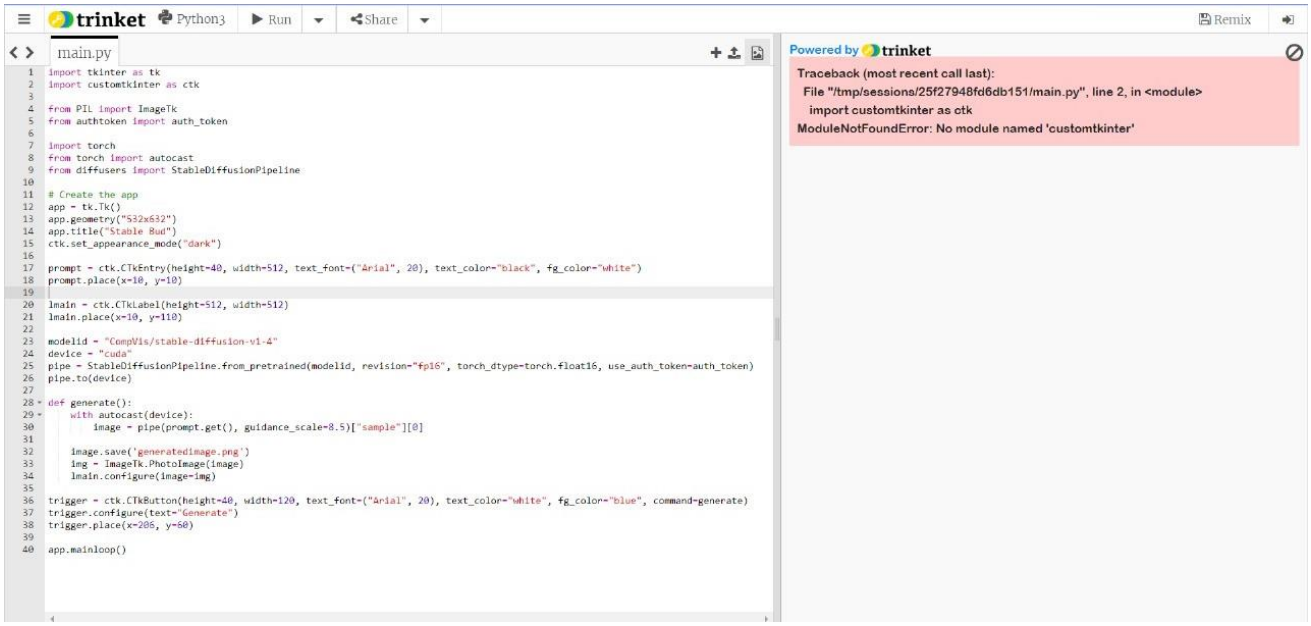
```
def generate():
    with autocast(device):
        image = pipe(prompt.get(), guidance_scale=8.5)["sample"][0]

    image.save('generatedimage.png')
    img = ImageTk.PhotoImage(image)
    lmain.configure(image=img)

trigger = ctk.CTkButton(height=40, width=120, text_font=("Arial", 20), text_color="white", fg_color="blue", command=generate)
trigger.configure(text="Generate")
trigger.place(x=206, y=60)

app.mainloop()
```

OUTPUT:



Project explanation:

We used a pre-trained model called stable diffusion by huggingface since we discovered very late in the project that It would've been very hard to find a sizeable open source dataset and then train on such limited computational power.

We used customtkinter to set up a GUI for writing the prompt.

Although the code is ready and we are sure it should work, we were unable to test it because our laptop was throwing SSL certificate errors due to which pip installations of required libraries did not work. If it was tested on another machine, it'll work

A text-to-image converter is an advanced tool powered by artificial intelligence (AI) that transforms written descriptions into visual images. This technology leverages deep learning models, enabling the system to interpret the text and generate corresponding visual representations, which closely align with the provided description.

Key Concepts in Text-to-Image Conversion:

1. Natural Language Understanding (NLU):

Natural language understanding (NLU) is a branch of AI focused on comprehending human language. When a text description is input into a text-to-image converter, the system first needs to process and understand the various components of the description. The model analyzes:

Objects: What specific items or entities are described? (e.g., a tree, a car, a person)

Actions: Is there any movement or activity involved? (e.g., running, jumping, flying)

Attributes: What are the details associated with the objects? (e.g., color, size, texture, shape)

Spatial Relationships: How are these objects positioned relative to each other? (e.g., the car is parked next to the tree)

Emotions or Atmosphere: Does the text imply any mood or tone? (e.g., calm, eerie, festive)

Through NLU, the model identifies the key elements described in the text and prepares to visualize them accurately.

2. Image Synthesis:

Once the system understands the text, it moves to the image synthesis phase. This involves using AI-based techniques to generate visual content from the extracted information. The most common models employed for this task include:

Generative Adversarial Networks (GANs): GANs are a type of deep learning model where two networks (a generator and a discriminator) work together. The generator creates images based on the text input, and the discriminator

evaluates whether the generated images look real. Over time, the generator gets better at producing realistic images.

Diffusion Models: These models generate images by gradually converting noise into detailed visuals based on the text description, refining the image through each iteration.

The system "learns" these processes by being trained on large datasets that pair text descriptions with corresponding images. Through this training, the model develops the ability to generate visually coherent and contextually accurate images based on the text prompts it receives.

Applications of Text-to-Image Converters:

1. Art and Design:

Text-to-image tools are revolutionizing creative industries by offering artists and designers a powerful way to generate visual concepts from mere descriptions. For example, a designer could input a simple prompt like "a futuristic city at night with neon lights," and the system would generate an image that embodies this concept. This allows artists to rapidly prototype ideas, explore creative variations, and draw inspiration from AI-generated visuals. It's especially useful for concept art, digital painting, and character design, where initial ideas can be generated quickly and iteratively.

2. Advertising:

In marketing and advertising, visuals are key to capturing audience attention. With text-to-image conversion, marketers can generate custom images for their campaigns based on a brief description of the brand or product. For example, an advertising team could input a text prompt such as "a refreshing glass of orange juice on a sunny beach," and the converter would produce an image that matches this scene. This not only speeds up the creative process but also offers flexibility in generating a variety of visuals for different campaigns.

3. Storytelling and Media:

For writers, filmmakers, and game developers, text-to-image tools enable visualization of story elements and scenes. A writer could describe a scene like "a dragon flying over a medieval castle during sunset," and the system would generate an image that reflects this description. This is particularly helpful in the planning stages of storytelling, such as creating storyboards, concept art for animated films, or illustrations for books. It enhances the creative process by bringing written ideas to life in a visual format, fostering new perspectives on narrative construction.

4. Accessibility:

Text-to-image conversion has profound implications for improving accessibility, especially for visually impaired individuals. By converting text into images, the tool allows people with visual impairments to "see" concepts through generated visuals, offering an enhanced understanding of the world around them. For instance, a user can input a description of an environment or scene, and the system will produce a corresponding image, which can then be described using assistive technologies (like screen readers or tactile feedback devices).

Benefits of Text-to-Image Conversion:

- Creativity on Demand: Text-to-image tools open up endless possibilities for creative exploration. Artists, writers, and marketers can use these tools to quickly visualize their ideas, bypassing the need for complex graphic design or illustration skills.

- Time Efficiency: The ability to generate custom visuals from text descriptions allows businesses and individuals to save time in creating content, accelerating the design and production process.

- Customization: These tools provide a high degree of customization, enabling users to fine-tune their descriptions and generate images tailored to their specific needs or visions.

- Democratization of Visual Art: Text-to-image technology makes the creation of high-quality visuals more accessible to non-designers. Anyone with a creative idea can bring it to life without needing advanced artistic skills.

Future Implications:

As text-to-image conversion technology continues to evolve, we can expect even more sophisticated and accurate image generation. The technology might:

Support higher levels of detail, allowing for complex, photorealistic images based on intricate descriptions.

Enable interactive design processes, where users can modify images in real-time by adjusting their text inputs or selecting specific aspects to change.

Integrate more seamlessly into virtual reality (VR) or augmented reality (AR) environments, allowing users to create 3D models or immersive scenes based on text descriptions.