InfiniVision

Text-to-Image Generator

1. Introduction

InfiniVision is a text-to-image generator utilizing the OpenAI's GPT-3 and DALL-E models. Flask framework is used for web development. Users can enter to input text descriptions and receive corresponding images generated based on those descriptions as it would seem 20 years later. The app also has voice recording feature for a more user-interactive experience.

2. Setup Instructions

Python 3.8 or higher and pip (Python package installer) are prerequsities for the project. Ensure you have these software installed on your system. Then, follow the instructions below:

1. Clone the repository

The github repository link: https://github.com/elisturegun/InfiniVision.git

You need a copy of the project repository. Open your command prompt and run the following commands:

```
git clone https://github.com/elisturegun/InfiniVision.git
cd InfiniVision
```

Create necessary directory: Run the following command to create the `static` directory to store generated images:

```
mkdir -p static
```

In static folder, generated_image.png file will be created after an image is generated.

2. Install Dependencies

The project has many dependencies listed in the `requirements.txt` file. You can install them using pip. In the terminal, run following command to install requirements.txt:

```
pip install -r requirements.txt
```

3. **Setup API Keys**

   You need an API key to use openai models. Follow these steps to work with your API key:

   a) **Get an API Key**: Obtain your API key from openai website:

   https://openai.com/index/openai-api/

   b) **Configure the API Key**: Open `app.py` file in a text editor and replace the placeholder with your actual API key:

```
openai.api_key = 'your_openai_api_key'
```

4. **Execution Instructions**

   You can start the Flask application by following command. In your terminal, ensure that you are in the project directory and execute:

```
python app.py
```

5. **Access The Application**

   Once the application is running, open your web browser and go to following link:

   http://127.0.0.1:5000

3. Explanation of Directory Structure

static/:

a. `generated_image.png`: This file will be created dynamically after an image is generated and removed when redirecting to index.html

templates/:

b. `index.html`: The main page where users can enter text descriptions or use voice input to describe a scene to see 20 years later.
c. `loading.html`: The loading page displayed while the the generate image button is clicked on index page.
d. `result.html`: The result page is seen when image generation is completed. The page displays the generated image with options to download or create a new image.

app.py: The main application file that sets up the Flask server and defines the routes and functions for handling requests.

requirements.txt: A file listing all the dependencies required for the project. This can be used to install the necessary packages using pip.

README.md: The detailed documentation of the project, including setup instructions, usage instructions, and more.

4. Usage Instructions

Text Input

1. Enter a text description of the scene you want to generate in the text input field.
2. Click on the "Generate Image" button.
3. Wait for the image to be generated and displayed on the result page.

Voice Input

1. Ensure you are in a quiet and interrupted environment.
2. Click on the microphone icon to start voice recognition.
3. Speak your description clearly. Your voice will be recorded until silence.
4. The recognized text will be filled into the text input field automatically.
5. Edit the text input field if you want to.
6. Click on the "Generate Image" button to generate the image.
7. Wait for the image to be generated and displayed on the result page.

5. Functionality

a. Main Pages and Endpoints

Index Page (/)

- **Description:** The main interface where users can enter text descriptions or use voice input to describe a scene.
- **Features:**
  - Text input field
  - Voice input button
  - Submit button to generate an image

Loading Page (/loading)

- **Description:** Displays a loading animation while the image is being generated.
- **Features:**
  - Animated loading screen
  - Status check for image generation

Result Page (/result)

- **Description:** Displays the generated image with options to download or create a new image.
- **Features:**

    - Generated image display
    - Download button
    - Create new image button

API Endpoints

/generate (POST)

- **Description:** Handles the text input from the user and initiates the image generation process.
- **Parameters:**
    - text: The textual description provided by the user.
- **Response:** JSON object indicating the status of the image generation.

/status (GET)

- **Description:** Checks the status of the image generation process.
- **Response:** JSON object with the status and filename of the generated image if ready.

/clear_image (GET)

- **Description:** Clears the generated image and resets the system to accept new inputs.
- **Response:** Redirects to the index page.

/start_voice_recognition (GET)

- **Description:** Starts the voice recognition procedure to get text input.
- **Response:** JSON object with the transcribed text or error message.

6. Performance Evaluation

The Infinivision demonstrates good performance in generating high-quality images from both text and voice inputs. The system's integration of advanced large language model GPT-3 and image generation model DALL-E supports a captivating user experience. Processing speed and handling detailed descriptions may be enhanced in the future to promote user satisfaction and system performance.

7. Generated Samples

A futuristic city skyline

Generated Image from DALLE:

Detailed description generated by LLM: In the year 2041, the futuristic city skyline stretches across the horizon like a vision from a science fiction novel brought to life. Towering skyscrapers made of gleaming glass and steel rise up into the sky, their facades adorned with intricate geometric patterns that seem to shift and shimmer as the light dances off them. The buildings are interconnected by a network of skybridges and aerial walkways, bustling with people commuting between their workplaces and homes.

Embedded within the architecture of the city are futuristic technologies such as holographic displays that project advertisements and announcements into the air, creating a dynamic and ever-changing visual landscape. Automated drones zip through the air, delivering packages and transporting people quickly and efficiently from one part of the city to another.

Solar panels and wind turbines can be seen adorning the rooftops of many buildings, providing clean and sustainable energy to power the bustling metropolis below. The streets below are teeming with self-driving cars and electric bicycles, gliding silently along smart roads that adjust traffic patterns in real-time to optimize efficiency.

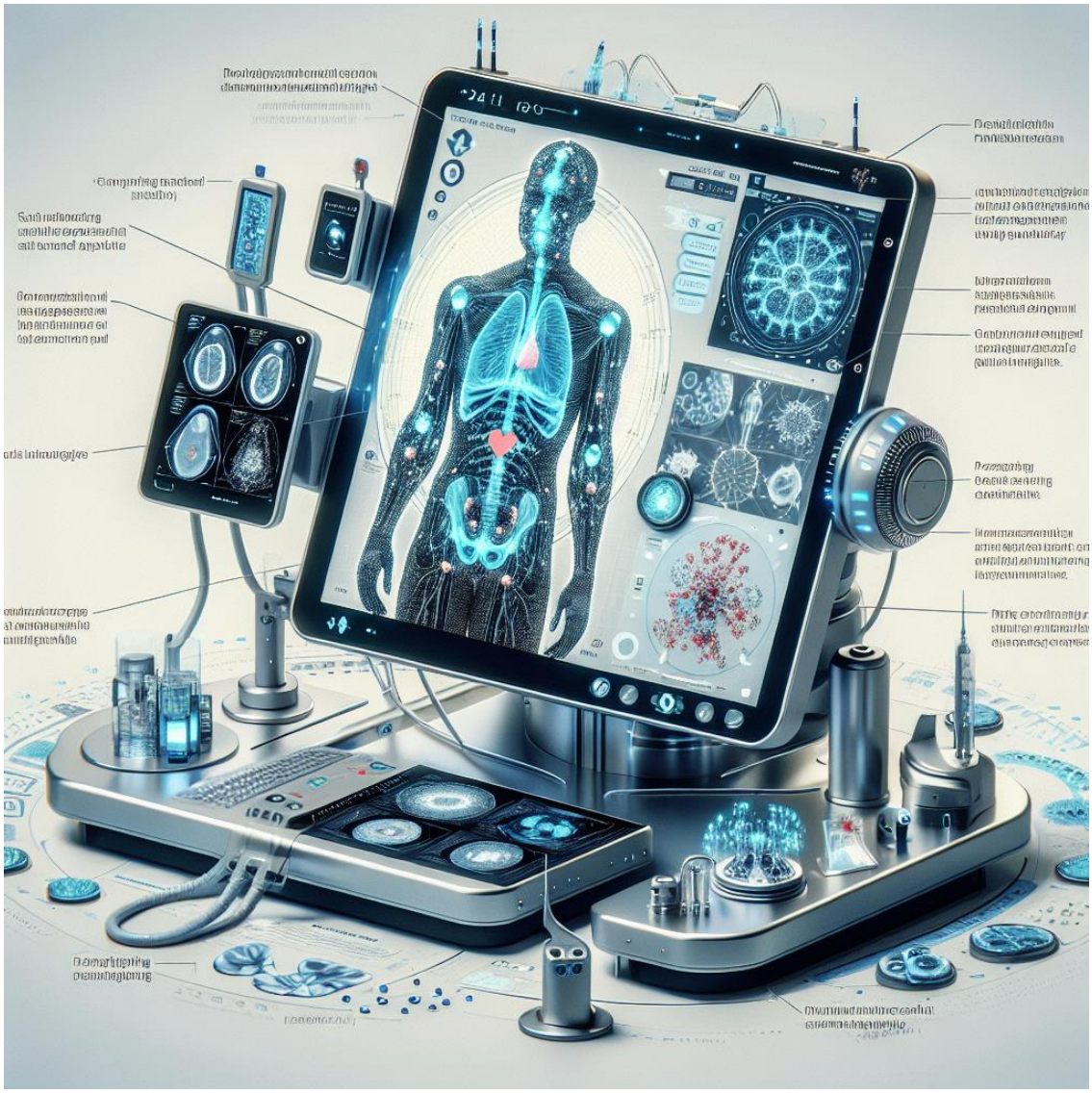
At night, the city comes alive with a dazzling display of neon lights and glowing signs, casting a colorful glow over the urban landscape. Hovering transport pods whiz through the air, carrying passengers to their destinations with a sense of futuristic elegance.

In the distance, towering megastructures reach towards the sky, housing entire communities within their self-contained ecosystems. Parks and green spaces dot the cityscape, offering residents a respite from the hustle and bustle of urban life.

This futuristic city skyline is a testament to human ingenuity and technological advancement, a vision of what the world could look like in 20 years time.

An advanced medical diagnostic device

Generated Image from DALL-E:



Detailed description generated by LLM: In the year 2041, medical diagnostic devices have evolved to a new level of sophistication and capability, particularly in the realm of imaging technology. One such advanced device is a state-of-the-art diagnostic system that utilizes a combination of artificial intelligence, nanotechnology, and advanced imaging sensors to provide unparalleled insights into the human body.

The device itself is a compact and sleek piece of equipment, resembling a high-tech tablet mounted on a stand with an array of sensors and cameras integrated seamlessly into its design. At the heart of this advanced diagnostic system is a powerful artificial

intelligence engine that is capable of interpreting complex imaging data with incredible speed and accuracy.

When a patient is scanned using this device, a high-resolution, three-dimensional image of the body is generated in real time, capturing the smallest details with remarkable clarity. The device is able to not only identify and visualize individual organs, tissues, and cells, but also to conduct detailed analyses of their function and structure at a molecular level.

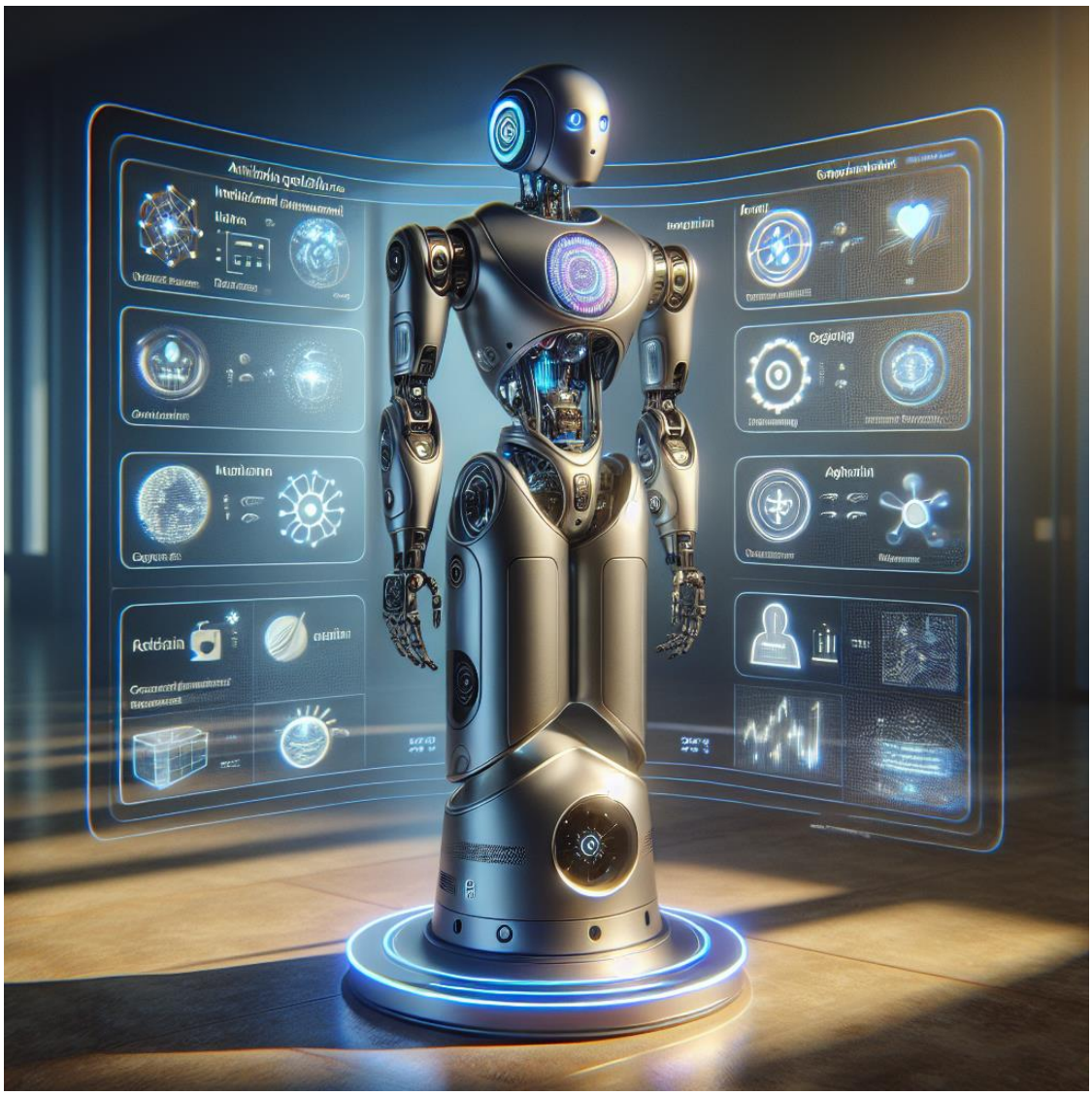
One of the most revolutionary features of this advanced diagnostic device is its ability to provide predictive insights into a patient's health outlook. By analyzing patterns and anomalies in the imaging data, the device can predict the likelihood of developing certain diseases or conditions, allowing for early intervention and personalized treatment plans.

Moreover, the device is also equipped with advanced nanotechnology sensors that can detect and analyze biomarkers and other indicators of disease in real time, providing instantaneous feedback to the medical team and enabling rapid decision-making.

Overall, this futuristic medical diagnostic device represents a groundbreaking leap forward in the field of medical imaging, offering unprecedented levels of precision, insight, and predictive capability that have the potential to revolutionize healthcare and improve patient outcomes in ways previously thought impossible.

A high-tech personal assistant robot

Generated Image from DALLE:



Detailed description generated by LLM: In the year 2041, the world has seen remarkable advancements in technology, leading to the creation of a highly sophisticated personal assistant robot. This high-tech robot, named Assistron, is a sleek and futuristic marvel of design and functionality.

Assistron stands at 6 feet tall and is made of a combination of lightweight yet durable materials, giving it a modern and elegant appearance. Its metallic body is adorned with smooth curves and intricate patterns, enhancing its aesthetic appeal. The robot's

exterior is equipped with a cutting-edge holographic display system that projects interactive interfaces for seamless communication with its users.

Assistron is powered by advanced artificial intelligence algorithms, enabling it to understand and respond to natural language commands with unparalleled accuracy and speed. Its voice recognition capabilities are so sophisticated that it can easily distinguish between different users and tailor its responses accordingly. Additionally, Assistron's deep-learning systems allow it to continuously improve its performance over time, adapting to the unique needs and preferences of each individual user.

One of Assistron's most impressive features is its ability to see and understand its surroundings through a network of advanced sensors and cameras. This allows the robot to navigate complex environments with precision and efficiency, making it an invaluable assistant for tasks such as household chores, organization, and even outdoor activities.

Assistron is equipped with a wide range of interactive tools and applications, including calendar management, news updates, weather forecasting, and more. Its connectivity features enable seamless integration with smart home systems, allowing users to control their environment with simple voice commands.

Overall, Assistron represents the pinnacle of personal assistant technology, combining cutting-edge design with unparalleled functionality to enhance the lives of its users in ways previously unimaginable. It is a visionary glimpse into the future of robotics and artificial intelligence, setting a new standard for what is possible in the realm of human-robot interaction.

A self-sustaining space colony on Mars

Generated Image from DALLE:

Detailed description generated by LLM: In the image of a self-sustaining space colony on Mars 20 years from now, we see a sprawling settlement that has transformed the once-barren Martian landscape into a thriving hub of activity. The colony consists of interconnected domed structures covering vast expanses, each housing a variety of facilities and living quarters for the inhabitants.

Along the perimeter of the colony, we see rows of solar panels and wind turbines stretching out into the distance, harnessing the abundant sunlight and winds of Mars to

generate power for the entire settlement. The panels glisten in the Martian sunlight, a testament to the colony's commitment to sustainability and renewable energy.

Within the domes, lush greenery is abundant, as the colonists have successfully cultivated crops and plants in controlled environments. The air inside the domes is filled with oxygen produced by the plants, creating a breathable atmosphere that mirrors Earth's to a certain extent.

In the central hub of the colony, we see a bustling marketplace where residents gather to trade goods and socialize. The area is alive with activity, as people from all walks of life go about their daily routines, reflecting the diverse and vibrant community that has taken root on Mars.

In the background, we catch a glimpse of advanced transportation systems, such as maglev trains and electric vehicles, zipping around the colony, providing efficient and sustainable means of getting around the settlement.

Overall, the image of a self-sustaining space colony on Mars 20 years from now is a vision of human ingenuity and determination, showcasing our ability to adapt and thrive in the harshest of environments. It stands as a testament to the pioneering spirit of the colonists who have made Mars their new home, and a beacon of hope for the future of interplanetary exploration and colonization.