



Holo: Gestured-Controlled Generative-AI Painting Application

Zachary Cowan

Faculty Mentor: Nathan P. Johnson
Bellarmine University Spring 2025

Introduction

Project Holo is a multimodal artificial intelligence (AI) system that integrates multiple AI technologies into a dynamic, hand-gesture-controlled canvas painting and image generation application. With the increasing integration of AI in consumer applications, Holo provides users with an intuitive and interactive platform for AI-driven image creation.

Objectives

Human Computer Interaction (HCI)

Holo seeks to explore a new avenue of Human Computer Interaction allowing users to interact without a traditional keyboard, mouse, and display.

- Mediapipe, a Google Solution, is utilized to enable hand gesture control – a hands-off means of interaction that replaces a physical mouse.
- Vosk speech recognition allows users to type text via speech replacing the traditional method of a physical keyboard.
- A projected display allows for unique interaction that a traditional display cannot provide.

Artificial Intelligence (AI) Integration

Holo integrates multiple AI models allowing users to generate images on the fly that are inspired by their input. Inputs include both images and text prompts.

- Using Image and Text to generate a new image.
 - Input image is manipulated based on text prompt.
 - Can produce new image or variant.
- Models used:
 - “qr2ai / outline”
 - “black-forest-labs/flux-schnell”
 - “adirik/t2i-adapter-sdxl-sketch”

Sketch and Text to Image Generative Model

Holo allows users to sketch an input image from scratch using various painting tools or to sketch over an already existing image.

- Input sketch and text prompt synthesize new image.
- AI model produces a new image based on the prompt guided by the user's sketch
- Can iterate multiple times and sketch over any image, previously generated images included (Compound AI System)

Features

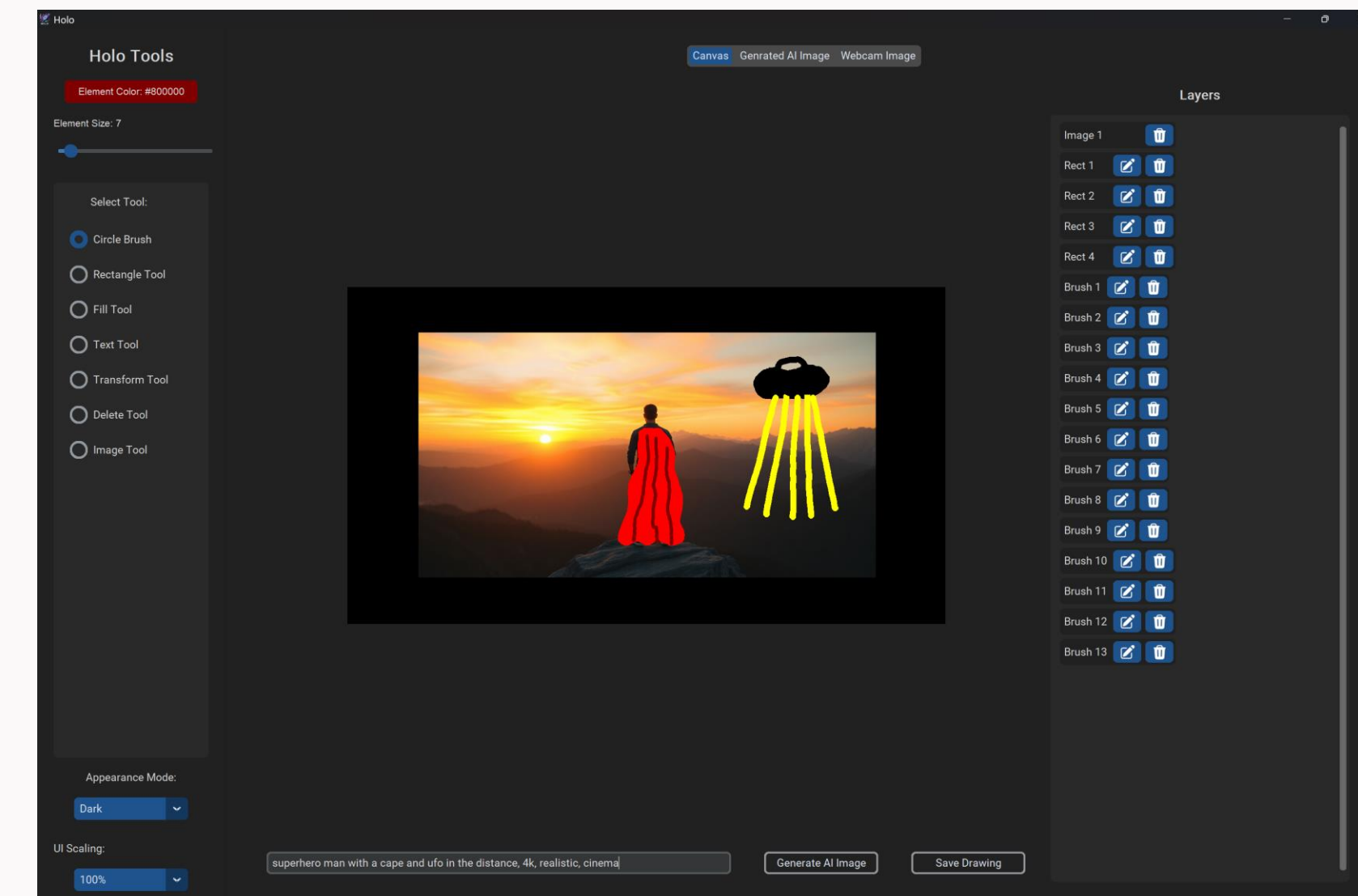


Figure 1: Basic sketch drawn over an image with a prompt that describes what the sketch is.

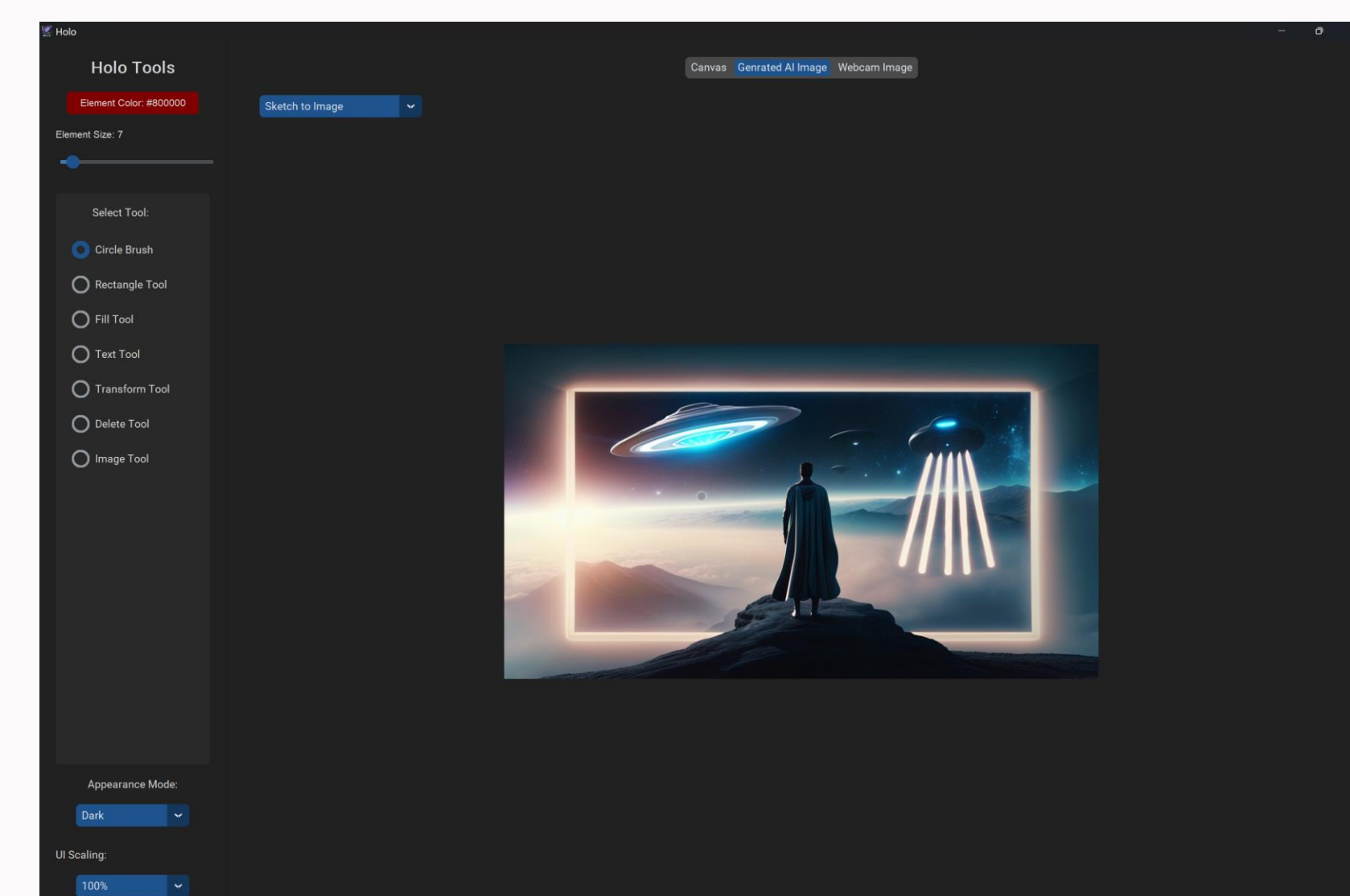


Figure 2: The AI Generated image based on the prompt and sketch from Figure 1.

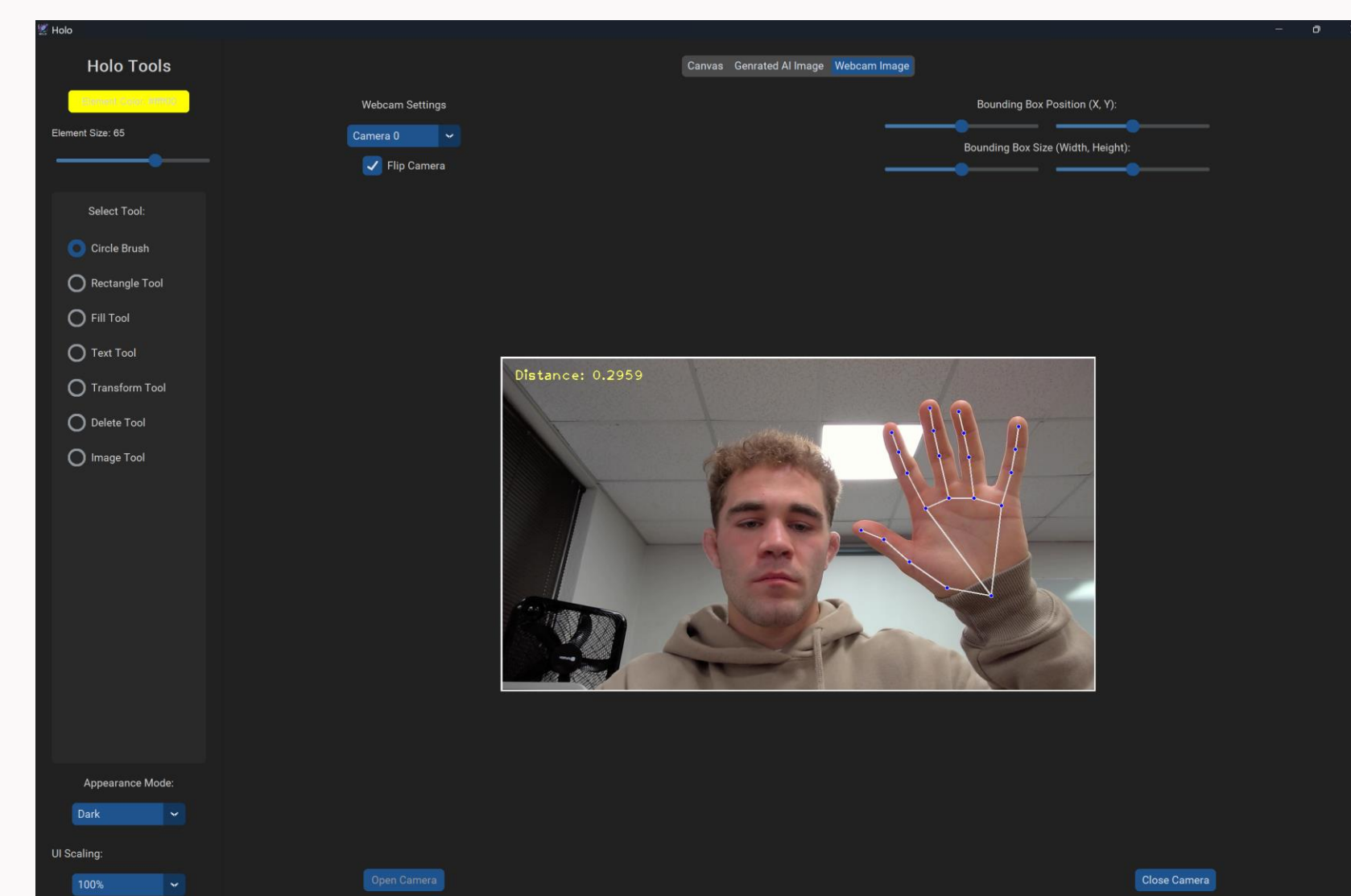


Figure 3: Demonstration of Mediapipe hand tracking. The respective joints and segments of the fingerbones are identified and tracked in real time.

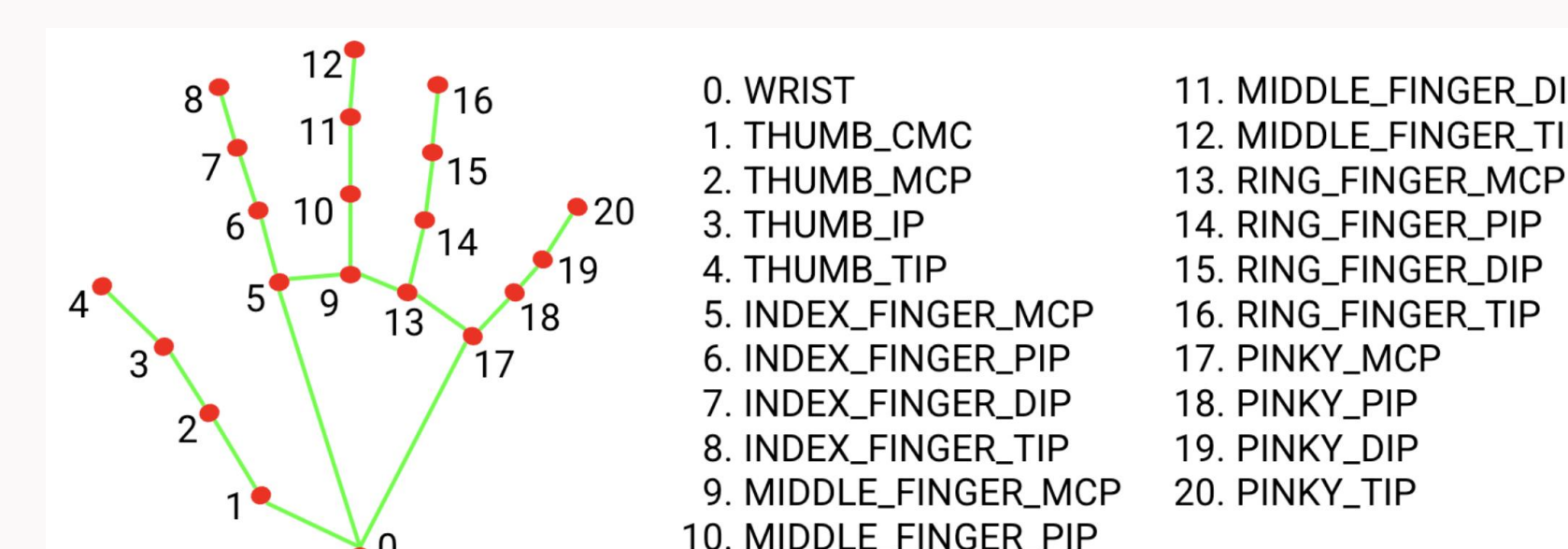


Figure 4: Media pipe hand joints via Googles documentation found at https://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker

Software Design and Dependencies

Holo is written using Python 3.12 with a variety external libraries and tools:

Graphical User Interface

- Tkinter – a binding to the “tk gui toolkit” for use in Python used to build the Holo GUI
- Custom Tkinter – an external custom theme built on top of tkinter to make modern styling quicker.

AI Image Generation

- Replicate – Used to generate AI images based on a given prompt and input image.

Speech Recognition

- Vosk – Used for offline speech recognition to convert speech to text.
- PyAudio – Used to capture audio from the microphone for speech recognition.

Live Webcam Feed and Gesture Control

- OpenCV – Used to capture video frames from the webcam, process images, and draw rectangles and text on the frames.
- PyAutoGUI – Used to control the mouse cursor based on hand tracking data.
- MediaPipe – Used for hand tracking to detect and process hand landmarks in the video feed.

Math and General Tools

- Threading – Used to run the camera loop and speech recognition in separate threads to avoid blocking the main GUI thread.
- NumPy – Used for efficient array operations and interpolation of coordinates.
- Pillow – Python Imaging Library

Refinements and Future Directions

Image Generation

The AI Model used to generate images plays the largest role in the outcome. As image models improve, they become more useful for generating specific and more refinable results. There are various implications in the following:

- Education
- Creative Applications
- Design and Engineering

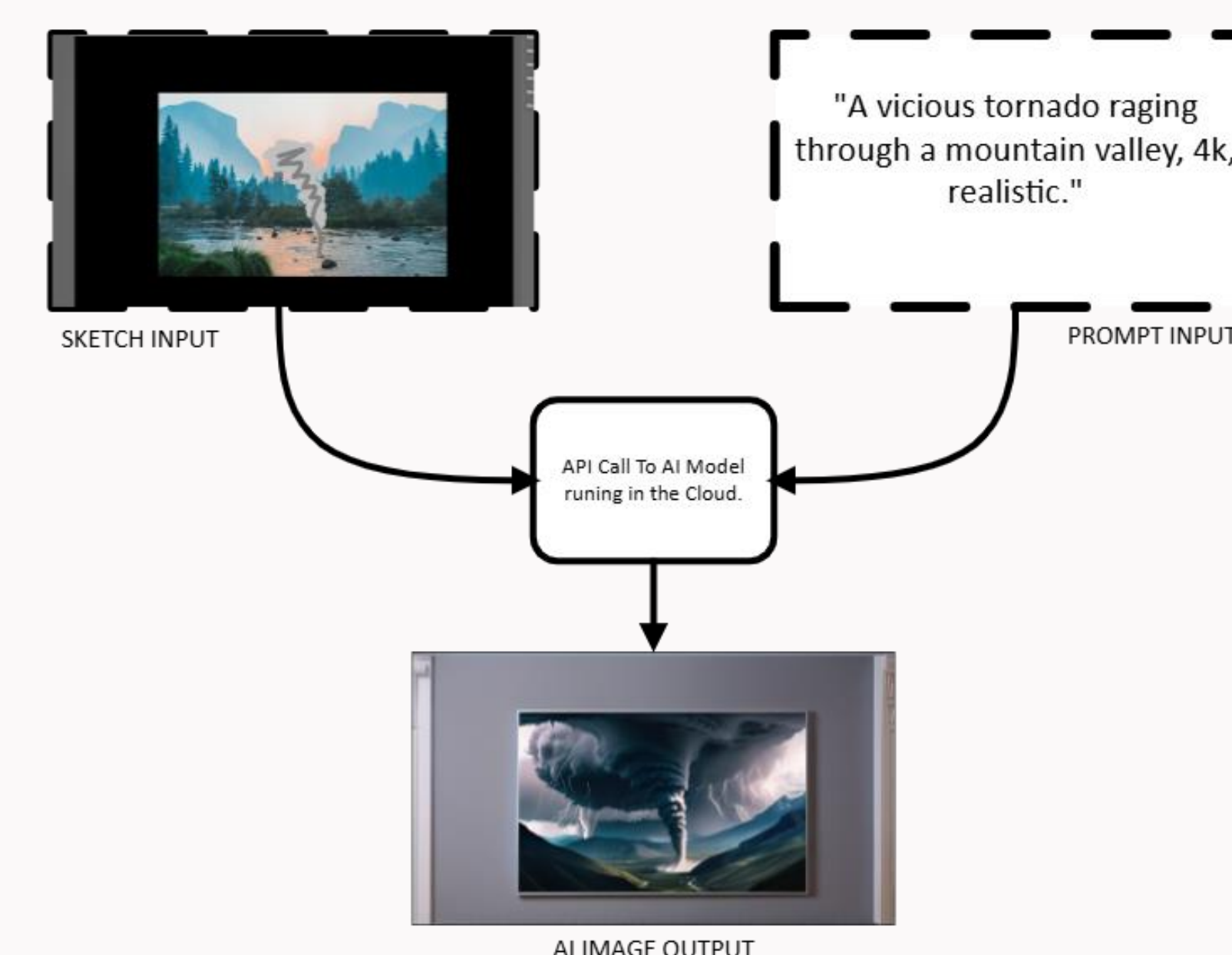


Figure 5: A diagram showing how Holo generates an image from user input.

Conclusions

Human Computer Interaction (HCI)

- With any new means of interacting with a system, learning can take some time. During a Bellarmine Maker Fair in March 2025, many high school students were able to use Holo. Although the interface is intuitive, becoming proficient in drawing with hand gesture control can take time.

Artificial Intelligence (AI) Integration

- Holo allows users to generate images using their sketches and prompts, choosing the model they desire for a given scenario. Holo also allows users to iterate on previously generated images: users may sketch over any given image (AI generated included) to better guide the models to get their desired output.
- With AI systems like Holo, the most effective users learn to predict the style of a given AI model and in turn can edit their prompts for best results.

Sketch and Text to Image Generative Model

- As with many current AI systems, the model training data and architecture strongly correlates to the desirability of the output. The ability to prompt and guide Holo's supported AI models with images and text does play some part in the final output although choosing the right model for image generation makes a significant difference.

References and Documentation

Mediapipe Documentation

https://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker

Custom Tkinter Documentation

<https://customtkinter.tomschimansky.com>

Replicate Documentation

<https://replicate.com/docs>

Vosk Documentation

<https://alphacephei.com/vosk>

Holo Source Code on Github

<https://github.com/zaccowan/Holo>