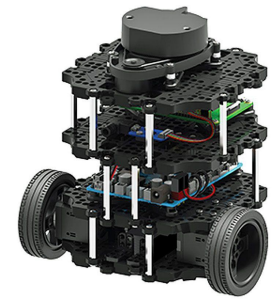




Foundation Model Driven Robotic Embodied Navigation



-
- SARHANA ADHIKARI
 - PROFESSOR YI FANG

Robotic Navigation in Outdoor Environment

Autonomous Vehicles and Drones

- Mapless Navigation
- Being able to decipher the environment outdoor
- Localization of the robot

Overview of the Foundation Model (LLM and LVM)

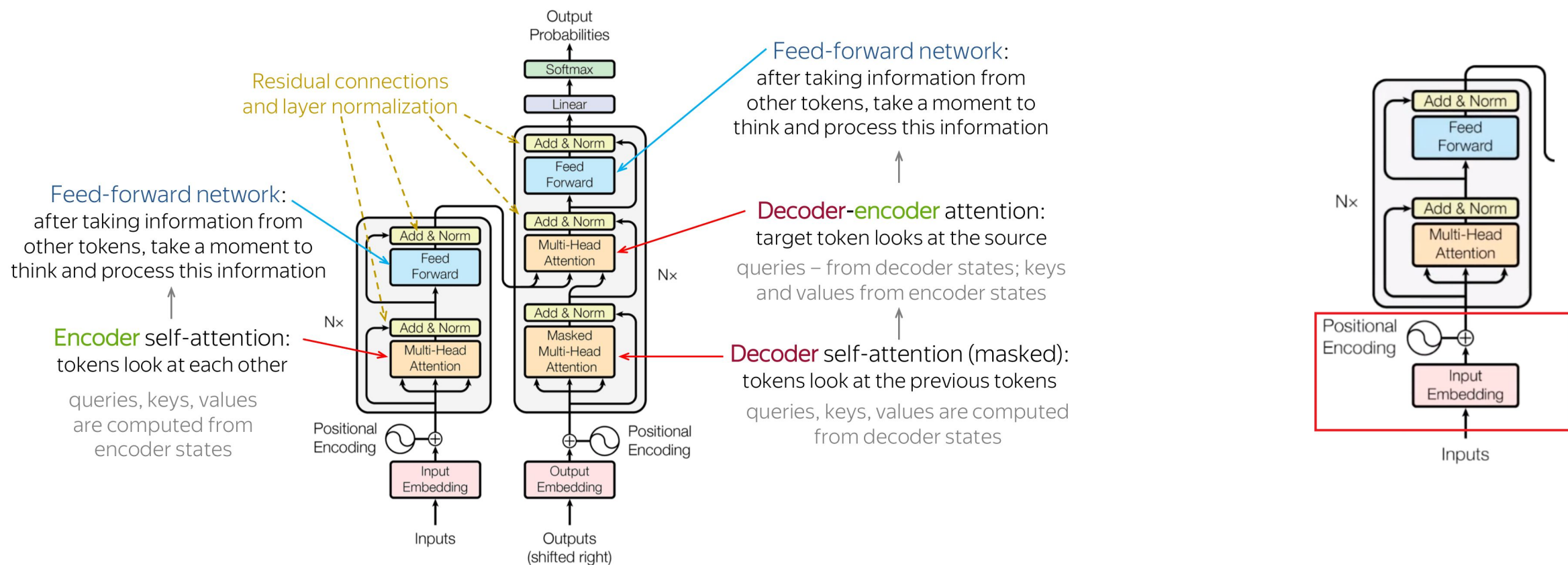
Key Features of Large Language Model (LLM):

- Natural language understanding
- Contextual text generation
- Answering questions
- spatial functional reasoning

Key Features of Language Vision Model(LVM):

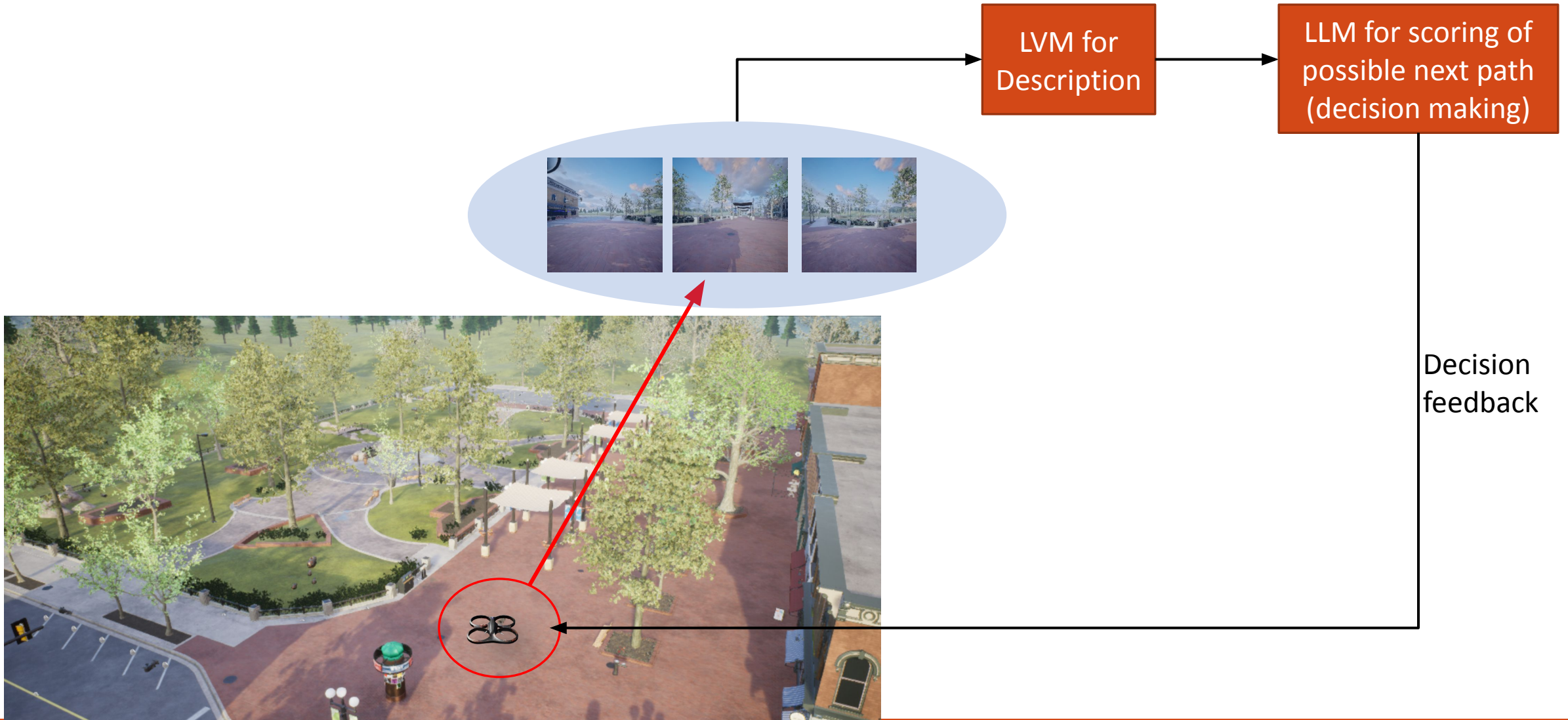
- Object recognition
- Scene understanding
- Text generation [description of scene/ image]

Transformer Network Structure

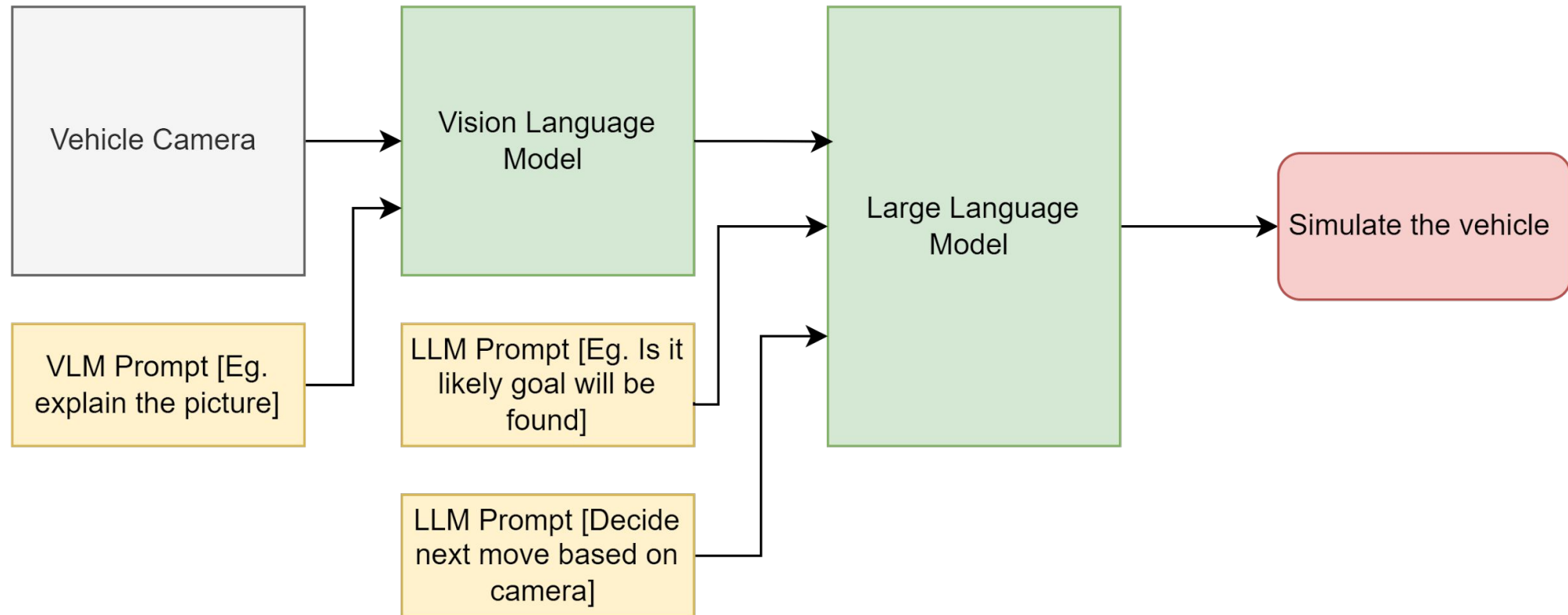


How the LVM contributes the Environmental scene understanding to help make decision

Integration of LVM and LLM



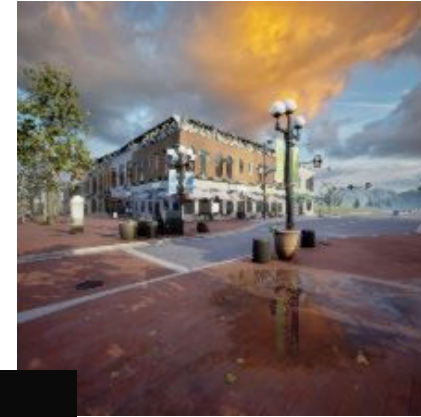
Integration of LVM and LLM



LVM and LLM in every step



Left Camera



Right Camera

```
Score: 1
ComputationalT: 1.043898582458496 seco
Goal Found? No
Score: 1
ComputationalT: 0.663470983505249 seco
Goal Found? No
Score: 3
ComputationalT: 0.6795487403869629 sec
Goal Found? No
```


Demonstration- Find the fountain



UNREAL
ENGINE

TEXTURE STREAMING POOL FULL
Collision Count:0
requestApiControl was succe
API call was not received, ente

2. ****Central Structure****: In the middle of the image, there is a circular structure with a thatched roof supported by a series of tall wooden pillars. This structure resembles a gazebo or pavilion, and it casts a slight shadow on the ground, indicating that the sun is shining.

3. ****Benches****: Surrounding the central structure, there are cylindrical concrete benches. They are spaced evenly around the base of the structure, providing seating for visitors.

4. ****Planters****: Interspersed between the benches are cylindrical planters made from the same material as the benches. They contain small trees or shrubs.

5. ****Trees****: Scattered throughout the image, there are several trees with green foliage. Some of these trees are closer to the viewer, while others are in the background.

6. ****Sky****: The sky fills the upper third of the image, featuring scattered clouds with a predominantly blue background, suggesting a pleasant, sunny day.

7. ****Buildings in the Background****: In the far background, beyond the plaza, we can see low

Running Baseline
Score: 5
ComputationalT: 1.1115796566009521 seconds
Goal Found? Yes

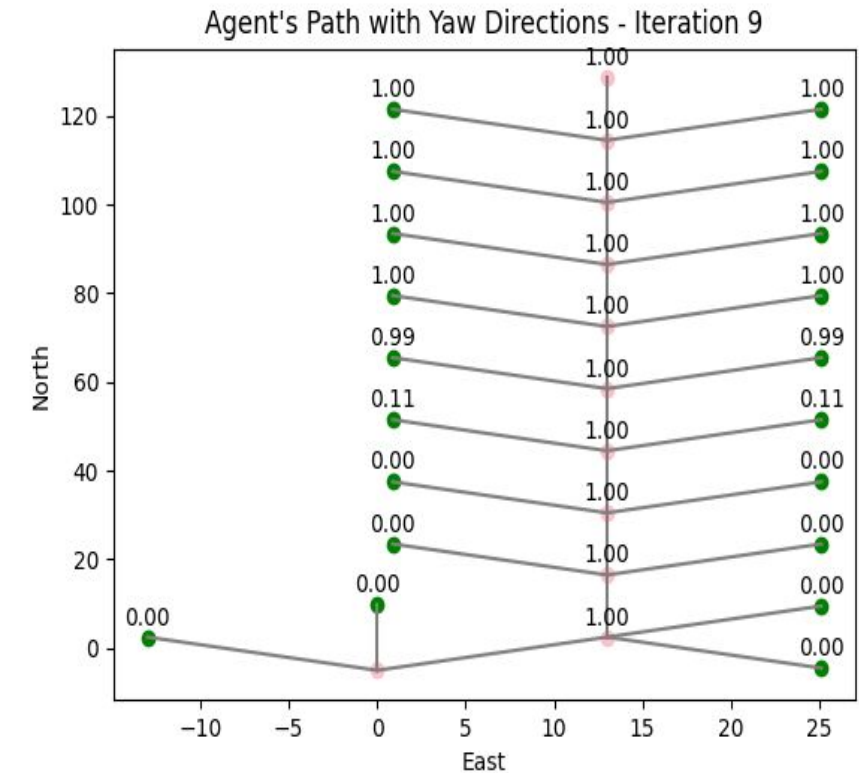
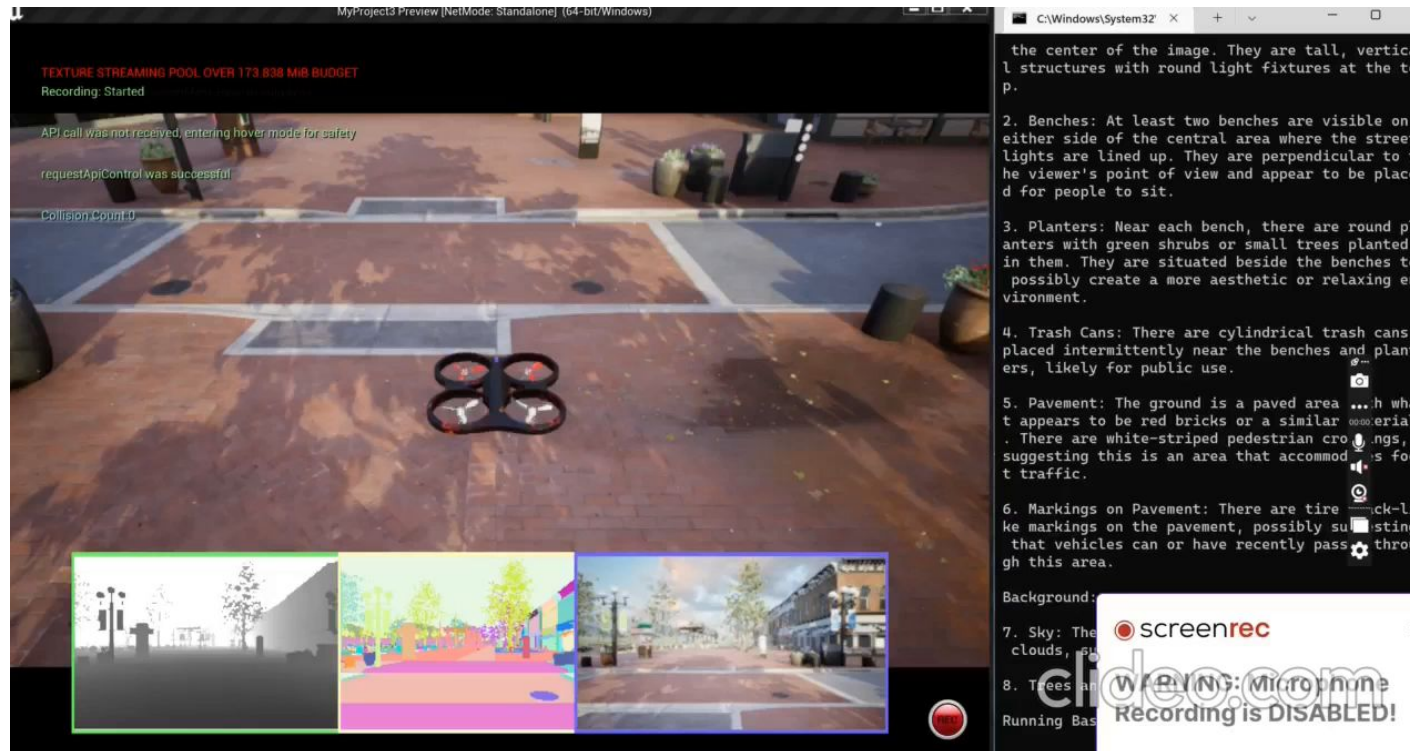
e benches—though they are part of the structures, they are intended for sitting.

6. People: There are several people in the background. They are too distant to make out any specific details, but they appear to be enjoying the park space.

7. Pathways: The park has walkways or paths that intertwine between the

Running Baseline
Score: 1
ComputationalT: 1.043898582458496 seconds
Goal Found? No
Score: 1
ComputationalT: 0.663470983505249 seconds
Goal Found? No
Score: 3
ComputationalT: 0.6795487403869629 seconds
Goal Found? No
Current GPS (10.038605690002441, 9.9656820050809233e-08, -0.9026179909706116)
d values [15.019339970856807, 15.019340104522, 11.980743718481884, 11.961394309997559, 11.980743891369995]
Sigma values [1.7083379096939715e-10, 1.13380569428067e-10, 3.738961436687835e-11, 3.7389617598988454e-11]
SCORES [(<exploration_simulator.Node object at 0x000021A42AD45E0>, 1.999999998291662), (<exploration_simulator.Node object at 0x000021A42AE13A0>, 0.999999998291662), (<exploration_simulator.Node object at 0x000021A451247C0>, 0.999999999626104), (<exploration_simulator.Node object at 0x000021A452F70>, 0.999999999626104), (<exploration_simulator.Node object at 0x000021A42ADD970>, 2.999999999626104)]
Drawing nodes!

Demonstration of a failed Simulation and plot of the graph



Results

- Drone Simulation capable of performing simple goal:
“find an object”
- Similar function maybe implemented on Cars, or more accessible robots
- Successful integration of LLM and LVM
- Learned Unreal Engine, AirSim, Prompt Engineering

References

- Reasoning about the Unseen for Efficient Outdoor Object Navigation, 18 Sep 2023 · Quanting Xie, Tianyi Zhang, Kedi Xu, Matthew Johnson-Roberson, Yonatan Bisk
- OpenAI. (2023). *ChatGPT-4V* (Sept 25 version) [Large language model]. <https://chat.openai.com>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is All you Need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf