While the technology behind autonomous vehicles (AVs) is only getting better over time, it still isn't perfect. Some accidents are being caused by AVs being put into non standard driving scenarios where they have no idea what to do. The AV doesn't know what to do because it either hasn't seen a similar scenario or enough similar examples, so it isn't able to ascertain what the right action is. One recent example of this is a recall of Waymo vehicles due to them having multiple collisions with semi-stationary road barriers such as chains or gates [1]. Around 1,200 vehicles were recalled and given software updates to address the safety issues

This is where large language models (LLMs) differ from autonomous vehicles. LLMs are capable of more human-like reasoning than AVs so in these confusing driving scenarios a LLM can have a greater capacity to reason about what should be done and can end up making safer decisions.

The work I will be doing is based off of the work that William Self has already done in his paper *Towards AV and Human Driving Alignment With LLM Reasoning*. William was able to create a framework that allowed ChatGPT 4-o to receive updates from the environment every timestep and make a decision on what the car should do. ChatGPT was able to drive in a safer manner than the DQN agent that William created, however getting information to and from ChatGPT was not happening fast enough.

For my research I plan to develop a framework similar to William's that connects LLMs to the environment and tests its ability to make decisions in a plethora of scenarios. Before each instance is run I will give the LLM a prompt explaining its goal, which will be to drive as safely as possible while trying to follow normal driving practices. One metric of success will be the speed at which the LLM agent is able to progress through timesteps. In William's work he said that it took about 1.5 seconds of processing for each timestep. When running the environment normally it would likely have between 30 and 60 timesteps occur each second, meaning that the LLM would take between 45 and 90 seconds to process one real life second of data. The speed at which the LLM processes data makes it currently unusable in real scenarios due to how long the processing takes. If I am able to find ways to shrink this number down it would make this method of autonomous driving closer to being usable in the real world. Another metric of success will be how good the decision making of the LLMs are. This is something that will be hard to quantify, because driving performance is somewhat subjective, however it will

1. Anne Marie D. Lee, "Waymo recalls roughly 1,200 self-driving vehicles prone to hitting road barriers," *CBS,* May 14, 2025. [Online]. Available: https://www.cbsnews.com/news/waymo-car-recall-software-crash-self-driving/

Week	Goal / Activity	Deliverable	Points	Notes
1	Draft research proposal	1 page research proposal	5	
2	Optimize the agent I created over the summer as a baseline to compare the LLM's performance against	Graphs showing improvements from original to updated agent	10	
3	Learn how to ask ChatGPT questions through API calls	Program showing basic interaction with ChatGPT through API calls	10	I have never used an API before so I need to spend some time figuring out how everything works
4	Interface ChatGPT with OpenAI highway environment	N/A	10	
5	Interface ChatGPT with environment cont'd	Program that allows ChatGPT to control the car in the environment and image / text logs that show what ChatGPT did	10	
6	Optimize ChatGPT interactions by trying to speed up how quickly the program runs by changing the prompt given to the LLM and playing around with not calling the LLM every frame	N/A	10	
7	Optimize ChatGPT interactions cont'd	N/A	10	
8	Optimize ChatGPT interactions cont'd	Comparison between original and updated program showing decreased runtime	10	
9	Create working github codebase for all the programs I have created	Github repository with all working programs I have so far	5	I want something easier to do during the week of Thanksgiving because I will be traveling a bit
10	Begin drafting final	N/A	10	I want to get started

	deliverable			on this before finals week because I felt rushed when writing my final report in the auditor program, and want to make sure I have enough time for this one.
Final	Finish final deliverable	5 page report	10	