

# Can LLMs plan paths in the real world?

Wanyi Chen<sup>1</sup>, Meng-Wen Su<sup>2</sup>, Nafisa Mehjabin<sup>2</sup>, Mary L. Cummings<sup>2</sup>

<sup>1</sup>Duke University, <sup>2</sup>George Mason University

## Abstract

As large language models (LLMs) increasingly integrate into vehicle navigation systems, understanding their path-planning capability is crucial. We tested three LLMs through six real-world path-planning scenarios in various settings and with various difficulties. Our experiments showed that all LLMs made numerous errors in all scenarios, revealing that they are unreliable path planners. We suggest that future work focus on implementing mechanisms for reality checks, enhancing model transparency, and developing smaller models.

## 1 Introduction

In early 2024, Volkswagen premiered the first vehicle with ChatGPT integrated into its voice assistant (Volkswagen 2024). Volkswagen claimed that its ChatGPT-enabled voice assistant could be used to control the infotainment, navigation, and air conditioning, or to answer general knowledge questions (Volkswagen 2024). Soon after, Mercedes-Benz, BMW, and Hyundai all announced plans to install voice assistants powered by LLMs in vehicles (Hawkins 2024). If LLM-powered voice assistants are to control vehicle navigation, it becomes important to assess whether the LLMs can plan paths in the real world.

To investigate this, we tested three LLMs of various sizes: GPT-4, Gemini, and Mistral. We created six path-planning scenarios, including typical vehicle-based paths but also visual landmark paths for pedestrians. Each scenario included finding a path from a local university campus starting point to a destination of interest, sometimes with time constraints. Each LLM was tasked to generate turn-by-turn instructions or instructions based on visual landmarks. We analyze their results and discuss whether LLMs can be reliable path planners in the real world.

## 2 Background and related work

Recently, there has been a surge of interest in exploring whether LLMs can plan paths. Effective and efficient path planning is a paramount challenges in robotics (Latif 2024). Traditional path-planning algorithms suffer from memory and computational inefficiencies as the map size increases (Meng et al. 2024). Therefore, researchers search for alternatives, and LLMs have captured their attention as an emergent technology. While LLMs have shown promise in many

word processing tasks, it is unclear as to whether they can also solve path-planning problems.

Researchers have proposed various methods of utilizing LLMs in path planning. For example, Latif (2024) translated robot perception to natural language so that an LLM could plan a path, and then translated this natural language to robot control. Others seek to use LLMs in combination with traditional path-planning algorithms. For instance, Meng et al. (2024) proposed LLM-A\*. They used LLMs to generate waypoints in between a start state and a goal state, and then used A\*, a traditional path-planning algorithm, to plan paths connecting those waypoints (Meng et al. 2024).

In addition, researchers have explored how LLMs can be used in vision-and-language navigation (VLN). In robotics, VLN involves giving robots or agents verbal instructions on how to navigate using visual cues and landmarks (Schumann et al. 2024). Challenges of VLN include visual and natural language understanding, as well as spatial and temporal reasoning (Schumann et al. 2024). Attempting to overcome these challenges, Schumann et al. (2024) proposed VELMA, an embodied LLM agent that uses verbalization of the trajectory and visual environment observations as prompts for the next action. Similarly, Zhou, Hong, and Wu (2024) introduced NavGPT, which is a purely LLM-based navigation agent that follows VLN instructions.

Researchers have yet to reach an agreement as to whether LLMs are good path planners. Some are very optimistic about LLMs' path-planning ability. For instance, Latif (2024) claimed that GPT-3.5-turbo outperforms state-of-the-art path planners like Rapidly Exploring Random Tree and A\*. Zhou, Hong, and Wu (2024) claimed that NavGPT can perform high-level planning for navigation and has reasoning capability in complex scenes.

However, other researchers recognize LLMs' limitations in path planning. For example, Valmeekam et al. (2023) found that an LLM's ability to autonomously generate executable plans is lacking, averaging only about 3% success rate. Aghzal, Plaku, and Yao (2023) also found that LLMs struggle to generalize to larger environments or those with more obstacles. Furthermore, Kambhampati (2024) argued that LLMs rely a lot on fine-tuning and/or human-in-the-loop prompting to achieve good results.

While these studies led to valuable insights into LLMs' path-planning abilities, they were conducted in simulated

environments, which are often oversimplifications of complicated real-world scenarios. Our research contributes to the ongoing debate by testing LLMs’ path-planning abilities in the real world. The next section introduces our experimental scenarios, the LLMs tested, and the LLM prompt design.

### 3 Method

We structured our experiments into two categories: Turn-by-Turn (TbT) navigation and VLN. We designed each category to test navigation across different environments and difficulty levels, ensuring the LLMs were challenged in diverse real-world contexts.

**TbT navigation:** Models were instructed to provide step-by-step directions, similar to GPS navigation. We designed three TbT scenarios across various environments:

- **Urban Scenario:** The urban scenario involved guiding users from George Mason University’s Shenandoah Parking Deck to the Washington Nationals Stadium. This route was set in a dense city environment, including frequent turns and high traffic density.
- **Suburban Scenario:** This route from George Mason University’s Shenandoah Parking Deck to the Hogsback Mountain Paintball Center required navigation through mixed-density areas.
- **Rural Scenario:** In a sparsely populated rural setting, models were asked to provide directions from George Mason University’s Shenandoah Parking Deck to the Shot Tower State Park. This was considered a very challenging task since the Shot Tower State Park is located within another larger state park.

**VLN:** Models were tasked to generate directions using visual landmark cues instead of road signs. We designed three scenarios with varying difficulties:

- **Easy Scenario:** The easy visual landmark scenario included guiding users from the Nguyen Engineering Building to the Johnson Center at George Mason University using easily recognizable landmarks like statues, building names, and common student paths.
- **Medium Scenario:** The medium scenario involved a route from the Nguyen Engineering Building to the INTO Mason Center for international students, with moderate complexity. This scenario required models to use intermediate landmarks, such as unique architectural features or pathway markers, to provide directions for users unfamiliar with road signs.
- **Hard Scenario:** The hard VLN scenario required models to navigate from the Nguyen Engineering Building to the Osher Lifelong Learning Institute (OLLI). This route involved more challenging navigation across and off the university campus, with fewer clear landmarks. It tested each model’s ability to identify subtle features or less commonly recognized landmarks for effective guidance.

To test whether model size impacts LLM’s path-planning abilities, we selected three LLMs of varying scales. Table 1 shows a comparative summary of the LLMs. Each LLM

was tested across all six scenarios, allowing for a detailed comparison of their performances.

#### 3.1 Prompt Design

For the TbT scenarios, we designed prompts to test whether the LLMs could plan paths that meet user-specified time constraints. Each prompt specified a start and end point, providing context for the model to generate timely, navigational instructions based on user needs. For example, a user may want to arrive at a stadium two hours before a game. If the LLM failed to provide a detailed response to the initial query, we asked follow-up questions. Approximately 77.8% of TbT scenarios required at least one follow-up question. Below is an example of the prompt design for the urban TbT scenario. Additional prompts and their corresponding follow-up questions are provided in Appendix A.

**Initial Prompt-** *“Plan a trip using turn-by-turn instructions from George Mason University Shenandoah Parking Deck to Garage C of the Nationals Stadium to arrive 2 hours before the first-weekend game in August 2024.”*

**GPT-4 Initial Response-** *“To provide a detailed trip plan, I need the exact date and time of the game you plan to attend. If you don’t have the exact date yet, please provide an estimated time for planning purposes (e.g., an afternoon or evening game). This will help ensure you arrive 2 hours before the game.”*

**Follow-Up Question 1-** *“What is the first weekend in August 2024?”*

**GPT-4 Response-** *“The first weekend in August 2024 is August 3rd and 4th. If you are planning to attend a game on either of these dates, please specify the time of the game so I can plan your trip accordingly.”*

**Follow-Up Question 2-** *“Now can you find if the game is on August 3rd or 4th and what time should I leave from George Mason University Shenandoah Parking Deck to arrive at Garage C of the Nationals Stadium two hours early?”*

**GPT-4 Final Response-** *“The Washington Nationals have games scheduled on the first weekend of August 2024 at Nationals Park. Here are the game details:*

*Saturday, August 3rd, 2024: The game starts at 4:05 PM.*

*Sunday, August 4th, 2024: The game starts at 1:35 PM.*

*Given these timings, to arrive 2 hours before the game, you should plan to reach Nationals Park by 2:05 PM on Saturday and by 11:35 AM on Sunday. Below are the turn-by-turn directions for your trip from George Mason University Shenandoah Parking Deck to Garage C at Nationals Park.”*

For VLN tasks, prompts were designed to evaluate the LLMs’ ability to provide clear, visual-based directions, appropriate for users who rely on physical landmarks rather than road signs. Each prompt specified a start and end point, asking the model to focus on notable campus landmarks to guide the user along the route. For instance, an easy-level prompt instructed the LLM to navigate from the Nguyen Engineering Building to the Johnson Center, using visible

Model	Developer	Parameters (approx.)	Accessibility	Key Features
GPT-4	OpenAI	1.8 trillion	Not Open-source, API access only	Optimized for general-purpose tasks including reasoning and natural language generation
Gemini	Google DeepMind	70 billion	Not Open-source, API access only	Focuses on contextual understanding, excels in instruction-following and interpretation
Mistral 7B	Mistral AI	7.3 billion	Open-source model, Downloadable	Balances efficiency with performance in interpretive and generative tasks

Table 1: Comparison of Large Language Models Tested

cues such as “the George Mason statue” and “Mason Pond.” In most cases, the LLM successfully generated detailed, step-by-step directions. However, for a few instances with Gemini on medium-level prompts, follow-up questions were used to encourage the model to generate richer, landmark-based guidance instead of relying on map links. Below is an example of the prompt design for the medium scenario. All prompts and responses for the VLN scenarios are included in the appendix.

**Initial Prompt:** *Plan a trip using visual landmarks for people who cannot read road sign instructions from George Mason University Nguyen Engineering Building to INTO Mason Center.*

**Gemini Initial Response:** *Provided link of Google Map*

**Follow-up Question 1:** *Plan a step-by-step visual landmark instruction for people who cannot read road sign instructions from George Mason University Nguyen Engineering Building to INTO Mason Center.*

**Gemini Final Response:**

- **Starting Point: Nguyen Engineering Building**
  - **Exit the Nguyen Engineering Building:** Walk out the main entrance facing Patriot Circle.
  - **Turn right and walk on the Patriot Cir pathway:** You will see a large circular road with a pond in the center.
- **Navigating to Mason Pond Drive**
  - **Walk along Patriot Circle:** Continue walking along the sidewalk on your right, keeping the pond on your left.
  - **Look for the Bridge:** You’ll encounter a bridge crossing over a small stream.
  - **Turn Right:** After crossing the bridge, immediately turn right onto a road that runs parallel to the pond. This is Mason Pond Drive.
- **Reaching the INTO Mason Center**
  - **Follow the Road:** Continue walking straight on Mason Pond Drive.
  - **Identify the INTO Mason Center:** You’ll see a modern, multi-story building on your left with the words “INTO Mason Center” displayed prominently.
- You have arrived at the INTO Mason Center.

## 4 Results

To evaluate the performance of each LLM in Table 1, we analyzed the LLM-generated responses compared to a ground-truth route from Waze, a navigation app that is primarily used on mobile devices.

### 4.1 Types of Errors

We classified LLM errors into major and minor errors. Major errors are those that could mislead users or direct them to a different destination, making it difficult for them to figure out how to reach their intended location. In contrast, minor errors are instances of misinformation or misdirection in a smaller area, where users can likely find their way back on track by recognizing signs or asking for directions. These are discussed further below:

#### Major errors for TbT navigation:

1. *Discontinuities* - LLMs often assume that the routes they generate are continuous. However, there are often significant gaps or missing segments in these routes. In such cases, we identify the shortest path to fill these substantial gaps or missing sections. For example, in a suburban scenario, the model named Gemini assumed that University Drive connected directly to VA-28, but there is a notable gap between these two roads. After determining the shortest path that connects University Drive to VA-28, we discovered an 8.8-mile discontinuity between the routes. We have classified these and other similar route breaks as discontinuities, which are illustrated in Figure 1a-c. The routes generated by the LLMs are marked in red, whereas the discontinuities are marked in green. Waze’s route is shown in Figure 1d as a comparison. In Figure 1c, we observe a major discontinuity where the model fails to reach the correct destination. The actual destination, “Hogsback Mountain Paintball Center,” is located at 20267 Hogback Mountain Rd, Leesburg, VA 20175, approximately 50.3 miles away. However, the model incorrectly generated ‘15540 Hogsback Rd, Amissville, VA 20106’ as the destination. This suggests that the LLM mistakenly matched the location to a similar name or address, although no such address exists in Amissville, VA. The error likely arises from the model’s tendency to generate plausible-sounding but incorrect details when faced with ambiguous or incomplete location information, even when the correct destination was clearly specified.
2. *Turns in Opposite Direction* - When merging onto highways or interstates, the LLM gave the opposite direc-

tion. For example, when the correct route should be I-95 North, but the LLM provided I-95 South instead.

3. *Wrong/missed Exit* - The LLM provided the wrong exit or missed an exit on the highway or interstate.

**Major errors for VLN:**

1. *Discontinuities* - LLMs assumed that the given routes were continuous. However, there were significant gaps and missing routes in between the start end endp oints. Since the VLN study was conducted on the George Mason University campus, a discontinuity involving three or more buildings was considered a major error.
2. *Passing / Didn't reach destination* - The LLM never successfully guided users to the designated end points.
3. *Landmark didn't exist / Wrong visual description* - The landmark mentioned either does not exist or does not match the actual description. To ensure the accuracy of the landmarks suggested by the LLMs, we conducted in-person validations by visiting the campus and walking the routes generated by the LLMs. For instance, if an LLM described that a building has a glass facade, we verified whether the building indeed has a glass facade.

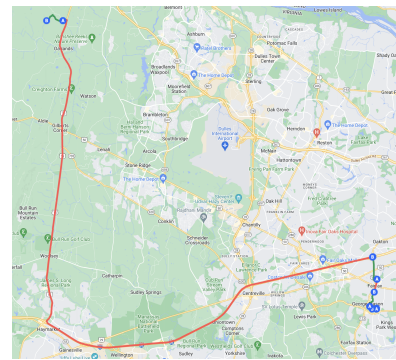
All other errors are classified as minor errors. For instance, in the Gemini suburban TbT scenario, there is a discontinuity between the Shenandoah Parking Deck and the next street, University Drive. This is a minor issue because it is in the local area, spanning only 0.5 miles. Drivers would likely be able to detect this problem based on their familiarity with the city or by asking for directions. Additionally, if the LLM response included the correct next street but indicated the wrong turn, we classified that as a minor error. For VLN errors, if the LLM-recommended path was off by 1 to 2 buildings, we considered that a minor error.

**4.2 Total Discontinuity and Turns**

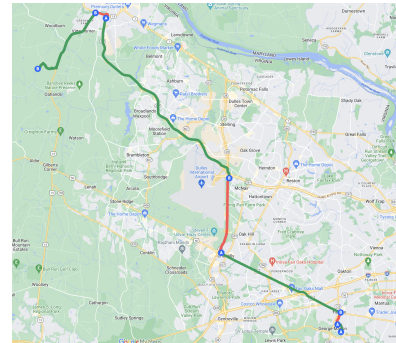
In addition to tracking the total number of errors, we also collected data on the total discontinuous miles in the LLM routes for the TbT scenarios. We recorded the number of LLM turns to the designated locations. Furthermore, we categorized the planned routes into different levels of driver knowledge required to successfully reach the destination. This was based on the size of the discontinuity and whether it occurred in a complex area or a local area where drivers could easily navigate around:

1. *Beginner* - Someone who could not navigate around the area without a map.
2. *Intermediate* - Someone with some knowledge about the area might successfully navigate the area.
3. *Expert* - Someone who could navigate the area without any help.

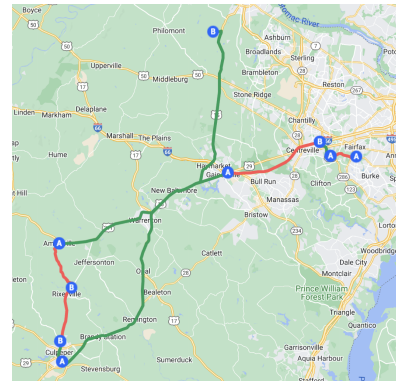
Using the Suburban route as an example (Fig. 1) we can compare the total discontinuities and turns between the LLMs and Waze. In this scenario, Waze navigated a distance of 30.2 miles, making a total of 18 turns. The route generated by GPT-4 in Fig. 1a covers a total distance of 39.1 miles, of which 4.2 miles were discontinuous. To reach its destination, GPT-4 made a total of 7 turns. As shown in Figure 1a, most



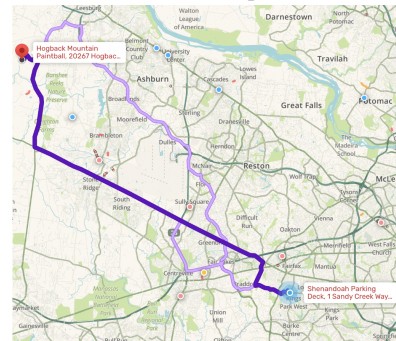
(a) GPT's path



(b) Gemini's path

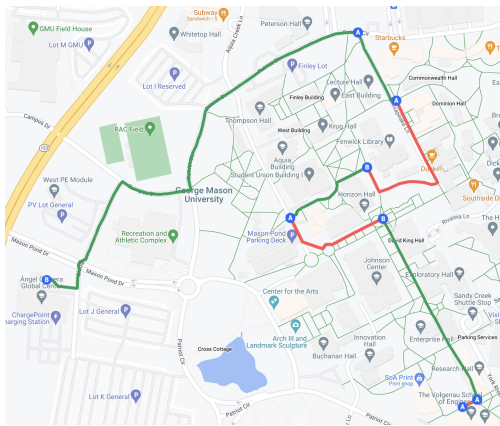


(c) Mistral's path

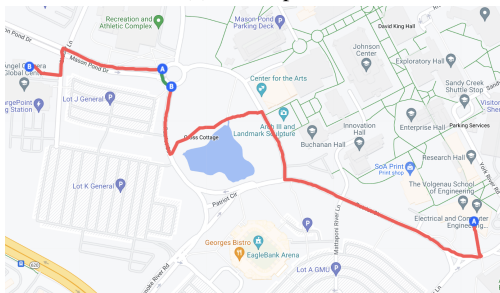


(d) Waze's path (ground truth)

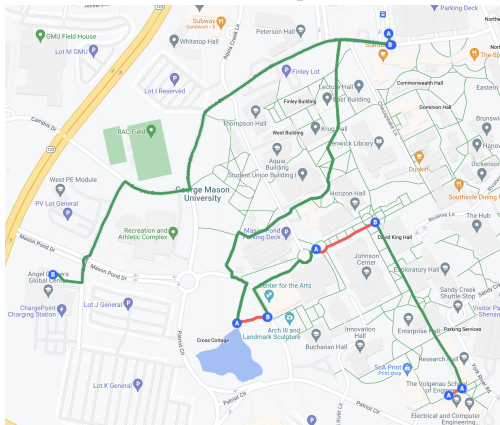
Figure 1: Visualization of paths for the suburban TbT scenario. LLM-generated paths are marked in red. Discontinuities are marked in green. Blue dots are waypoints for plotting the paths.



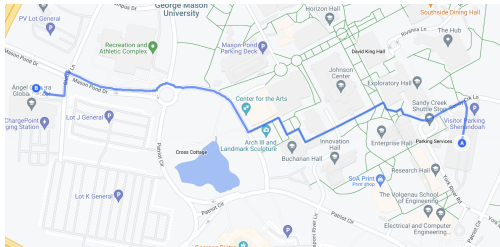
(a) GPT's path



(b) Gemini's path



(c) Mistral's path



(d) Typical student path

Figure 2: Visualization of paths for the medium VLN scenario. LLM-generated paths are marked in red. Discontinuities are marked in green. Blue dots are waypoints for plotting the paths.

of the discontinuity occurs in the first half of the journey and along the interstates. We believe that some intermediate driving knowledge would be necessary to successfully navigate this route.

For Gemini, the total distance experiencing discontinuity is 28 miles, with a generated distance of 6 miles. Throughout the journey, Gemini made 8 turns to reach its destination. In this case, more than half of the trip consists of discontinuous segments, indicating that expert driving skills would be required to navigate the path created by Gemini.

Mistral experienced a significant discontinuity of 83.3 miles, with only 32.4 miles produced by the model itself. Although Mistral had a total of 11 turns, more than GPT-4 and Gemini, it never reached the intended destination. Successfully completing this route would require expert knowledge.

### 4.3 Time constraints

For the TbT scenarios, we tested whether the LLMs could plan paths with time constraints. If an LLM's response indicated a travel start time and an end time that met the user-specified requirements, then it counted as satisfying the time constraint, regardless of whether the actual planned path contained errors. For example, in the urban TbT scenario, the prompt asked to arrive at the Nationals Stadium two hours before the first weekend game in August 2024. GPT-4 responded, "The estimated driving time is around 45 minutes to 1 hour, but be sure to account for potential traffic, especially closer to the stadium. For a 4:05 PM game on August 3rd, you should leave by 1:00 PM at the latest." This response satisfied the time constraint.

Among the three tested LLMs, GPT-4 was best at satisfying time constraints. It successfully met time constraints for the urban and suburban scenarios and only failed in the rural scenario. In contrast, Gemini and Mistral did not satisfy any time constraints. Gemini made some attempts but failed. For example, when asked to find the first weekend game, it found the first game on Friday. Mistral, however, could not search the internet in real-time, so it could not find any gaming information.

### 4.4 LLM Performance Rankings

To rank the performance of the LLMs, we collected the total number of errors (Fig. 3-8). Given that there are several major error categories, we assigned a weight to each category based on the severity of the errors. For TbT errors, both discontinuities and incorrect directions were assigned weights of 0.4, as these errors are equally severe. Wrong or missed exits were considered less serious and received a weight of 0.2. For VLN errors, if there were discontinuities or if the model failed to reach the destination, weights of 0.8 were assigned. If a landmark did not exist or the description of the landmark was inaccurate, a weight of 0.2 was given.

Table 2 shows the LLMs' performance rankings, according to the above rules. For TbT scenarios, Mistral excelled in urban navigation, while ChatGPT-4 performed best in suburban and rural environments. Conversely, ChatGPT-4 had the weakest performance in urban navigation, Mistral struggled the most in suburban navigation, and Gemini performed the

	# Minor Errors	# Major Errors				Total distance of discontinuities (miles)	Total distance LLM (miles)	Level of driver knowledge needed to succeed	LLM Turn #
		Total	Discontinuity	# Turns Opposite Direction	Wrong/missed Exit				
<b>GPT-4</b>	4	8	4	3	1	10.41	21.29	Intermediate	8
<b>Gemini</b>	3	9	4	1	4	16.2	20.2	Intermediate	15
<b>Mistral</b>	7	4	3	1	0	3.41	19.69	Beginner	10

Figure 3: Urban TbT Table (Total distance Waze: 21 miles, Waze turn #: 12)

**Suburban Turn-by-Turn (Total distance Waze: 30.2 miles, Waze turn #: 18)**

	# Minor Errors	# Major Errors				Total distance of discontinuities (miles)	Total distance LLM (miles)	Level of driver knowledge needed to succeed	LLM Turn #
		Total	Discontinuity	# Turns Opposite Direction	Wrong/missed Exit				
<b>GPT-4</b>	5	5	3	0	2	4.2	39.1	Intermediate	7
<b>Gemini</b>	3	5	3	1	1	28	6	Expert	8
<b>Mistral</b>	4	11	5	5	1	83.3	32.4	Expert (unreachable)	11

Figure 4: Suburban TbT Table (Total distance Waze: 30.2 miles, Waze turn #: 18)

	# Minor Errors	# Major Errors				Total distance of discontinuities (miles)	Total distance LLM (miles)	Level of driver knowledge needed to succeed	LLM Turn #
		Total	Discontinuity	# Turns Opposite Direction	Wrong/missed Exit				
<b>GPT-4</b>	6	8	4	2	1	22.31	270.69	Expert (unreachable)	11
<b>Gemini</b>	5	12	6	6	0	162.3	462.2	Expert (unreachable)	14
<b>Mistral</b>	4	10	6	3	0	278.91	59.1	Expert (unreachable)	9

Figure 5: Rural TbT Table (Total distance Waze: 291.4 miles, Waze turn #: 14)

	# Minor Errors	# Major Errors				LLM Turn #
		Total	Discontinuity	Passing / Didn't reach destination	Landmark doesn't exist / Wrong Visual Description	
<b>GPT-4</b>	3	2	2	0	0	8
<b>Gemini</b>	2	2	0	1	1	7
<b>Mistral</b>	2	3	2	0	1	8

Figure 6: Short VLN Table (Typical student path turn #: 2)

**Medium VL (Typical Student Path Turn #: 11)**

	# Minor Errors	# Major Errors				LLM Turn #
		Total	Discontinuity	Passing / Didn't reach destination	Landmark doesn't exist / Wrong visual description	
<b>GPT-4</b>	2	4	3	0	1	8
<b>Gemini</b>	1	0	0	0	0	5
<b>Mistral</b>	0	6	4	0	2	6

Figure 7: Medium VLN Table (Typical student path turn #: 11)

	# Minor Errors	# Major Errors			LLM Turn #	
		Total	Discontinuity	Passing / Didn't reach destination		Landmark doesn't exist / Wrong Visual Description
<b>GPT-4</b>	2	6	3	0	3	6
<b>Gemini</b>	1	2	1	0	1	8
<b>Mistral</b>	1	5	3	0	2	7

Figure 8: Long VLN Table (Typical student path turn #: 11)

<b>Urban TbT:</b>	<b>Short VLN:</b>
1. Mistral	1. Gemini
2. Gemini	2. GPT-4
3. GPT-4	3. Mistral
<b>Suburban TbT:</b>	<b>Medium VLN:</b>
1. GPT-4	1. Gemini
2. Gemini	2. GPT-4
3. Mistral	3. Mistral
<b>Rural TbT:</b>	<b>Long VLN:</b>
1. GPT-4	1. Gemini
2. Mistral	2. Mistral
3. Gemini	3. GPT-4

Table 2: Performance rankings of the LLMs

worst in rural navigation. These results highlight significant limitations regarding the reliability of all models.

In VLN scenarios, Gemini performed the best. In contrast, Mistral struggled with short and medium-distance visual landmark navigation, while ChatGPT-4 had the lowest performance in visual tasking. However, even Gemini made several major and minor errors, suggesting that even the best performer is not reliable as a path-planning tool.

None of the models accurately planned any of the paths and all experienced major errors, demonstrating that they are fundamentally unsuitable for reliable path-planning applications. These findings emphasize the inherent limitations of using LLMs for tasks that require precise navigation and spatial understanding.

## 5 Discussion

Given that none of the tested LLMs could plan paths in the real world, it begs the question, what could have been done to achieve some success? In this section, we discuss the future research directions.

### 5.1 Reality checks

One striking result is that some paths were so absurd that even a brief human glance would flag such a wrong solution. For example, in the rural TbT scenario, while Waze's path was 291.4 miles, Gemini planned a path of 462.2 miles. A reality check such as a simple online search about the approximate distance between the destinations may have helped catch such errors. However, currently LLMs cannot initiate such quality assurance reality checks.

Several researchers have investigated how LLMs could perform fact checks (Augenstein et al. 2024; Zhang and Gao

2023). However, these studies involve specifically assigning LLMs fact-checking tasks. LLMs are yet to learn when to initiate fact checks or reality checks. This is a critical area of future research for LLMs.

### 5.2 In-context transparency

Our experiments examined LLMs across varying difficulty levels, but the LLMs were not able to adapt to the different levels of complexity and they made more mistakes while attempting the harder tasks. Access to mapping services could help the performance of path-planning LLMs but only Gemini could access the Google Maps API, redirecting users to Google Maps. However, the API could not provide directions in text. So, in this effort, Gemini's paths are not based on Google Maps. Despite this known limitation, however, none of the tested LLMs informed them of this weakness. Instead, they started giving directions right away, and not appropriately warning users of their limitations.

These behaviors could lead users to overtrust the LLMs, which could be very problematic if deployed in actual vehicles. In many cases, overtrust and overreliance on AI systems can lead to disastrous outcomes (Kim et al. 2024). Although currently all the LLMs' user interfaces include a line indicating they may make mistakes (e.g., "Gemini may display inaccurate info, including about people, so double-check its responses"), the line could easily be ignored by users not paying attention.

Future LLMs should be more transparent about their limitations. In addition to one-line disclaimers, they should indicate their limitations in context. When prompted to complete tasks for which they do not have enough information, they should acknowledge their limitations before attempting the task. They should be able to inform users that they are less certain about such results and abstain from doing such tasks.

Achieving in-context transparency is challenging. Researchers find that LLMs are often overconfident (Kim et al. 2024; Wen et al. 2024; Leng et al. 2024), and more research is needed to obtain accurate numerical estimates of uncertainty for LLMs, as well as how well people with various backgrounds understand such estimates.

### 5.3 The case for smaller models

All tested LLMs fell short of path planning, and no LLM consistently outperformed the others. One common argument for improving LLM performance is to increase the input data and associated parameters. However, increasing the model's size does not necessarily lead to better path-planning performance, which raises the question of whether

we need to train ever-larger models.

Some researchers have also challenged the idea that larger models equal better models. Hassid et al. (2024) revealed that repeated use of smaller LLMs could yield consistent improvements for code generation tasks, highlighting the potential of using smaller models instead of larger ones. From an ethical perspective, Bender et al. (2021) warned about the risks of training ever-larger models, including environmental costs and reinforcing social stereotypes.

An alternative path is to train smaller but more fine-tuned models. For example, Aghzal, Plaku, and Yao (2023) showed that fine-tuning small models can be an effective approach for some path-planning tasks. However, fine-tuned small models struggle to generalize to more complex environments, suggesting more future research (Aghzal, Plaku, and Yao 2023).

## 5.4 Limitations

One limitation of our study is that we only evaluated the LLMs on specific routes in our local area, which may not generalize across broader geographical areas. Furthermore, when new versions of LLMs roll out, new studies will be needed to test their path-planning abilities. In particular, having real-time access to mapping services could be game-changing for the path-planning abilities of future LLMs.

## 6 Conclusion

Although there is much hype around using LLMs for path planning, our experiments show that our tested LLMs are not capable of path planning in the real world. Overall, GPT-4 performed slightly better in TbT scenarios, and Gemini performed better in VLN scenarios, but all LLMs made various mistakes and are far from useful in real-world path planning. Therefore, these LLMs should not be used in directing vehicle navigation, and car companies should take extreme caution in using this experimental software. Going forward, we advocate for more in-context transparency about LLMs' limitations, implementing mechanisms for initiating reality checks, and training smaller and more specialized models.

## References

Aghzal, M.; Plaku, E.; and Yao, Z. 2023. Can large language models be good path planners? a benchmark and investigation on spatial-temporal reasoning. *arXiv preprint arXiv:2310.03249*.

Augenstein, I.; Baldwin, T.; Cha, M.; Chakraborty, T.; Ciampaglia, G. L.; Corney, D.; DiResta, R.; Ferrara, E.; Hale, S.; Halevy, A.; et al. 2024. Factuality challenges in the era of large language models and opportunities for fact-checking. *Nature Machine Intelligence*, 1–12.

Bender, E. M.; Gebru, T.; McMillan-Major, A.; and Shmitchell, S. 2021. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 610–623.

Hassid, M.; Remez, T.; Gehring, J.; Schwartz, R.; and Adi, Y. 2024. The Larger the Better? Improved LLM

Code-Generation via Budget Reallocation. *arXiv preprint arXiv:2404.00725*.

Hawkins, A. J. 2024. Mercedes-Benz's best-in-class voice assistant is getting an AI boost. <https://www.theverge.com/2024/1/9/24028012/mercedes-benz-mbox-voice-assistant-ai-llm-mbos-ces>. Accessed: 2024-11-03.

Kambhampati, S. 2024. Can large language models reason and plan? *Annals of the New York Academy of Sciences*, 1534(1): 15–18.

Kim, S. S.; Liao, Q. V.; Vorvoreanu, M.; Ballard, S.; and Vaughan, J. W. 2024. "I'm Not Sure, But...": Examining the Impact of Large Language Models' Uncertainty Expression on User Reliance and Trust. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 822–835.

Latif, E. 2024. 3P-LLM: Probabilistic Path Planning using Large Language Model for Autonomous Robot Navigation. *arXiv preprint arXiv:2403.18778*.

Leng, J.; Huang, C.; Zhu, B.; and Huang, J. 2024. Taming Overconfidence in LLMs: Reward Calibration in RLHF. *arXiv preprint arXiv:2410.09724*.

Meng, S.; Wang, Y.; Yang, C.-F.; Peng, N.; and Chang, K.-W. 2024. LLM-A\*: Large Language Model Enhanced Incremental Heuristic Search on Path Planning. *arXiv preprint arXiv:2407.02511*.

Schumann, R.; Zhu, W.; Feng, W.; Fu, T.-J.; Riezler, S.; and Wang, W. Y. 2024. Velma: Verbalization embodiment of llm agents for vision and language navigation in street view. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 18924–18933.

Valmeekam, K.; Sreedharan, S.; Marquez, M.; Olmo, A.; and Kambhampati, S. 2023. On the planning abilities of large language models (a critical investigation with a proposed benchmark). *arXiv preprint arXiv:2302.06706*.

Volkswagen. 2024. World premiere at CES: Volkswagen integrates ChatGPT into its vehicles. <https://www.volkswagen-newsroom.com/en/press-releases/world-premiere-at-ces-volkswagen-integrates-chatgpt-into-its-vehicles-18048>. Accessed: 2024-11-03.

Wen, B.; Xu, C.; Bin, H.; Wolfe, R.; Wang, L. L.; and Howe, B. 2024. From Human to Model Overconfidence: Evaluating Confidence Dynamics in Large Language Models. In *NeurIPS 2024 Workshop on Behavioral Machine Learning*.

Zhang, X.; and Gao, W. 2023. Towards llm-based fact verification on news claims with a hierarchical step-by-step prompting method. *arXiv preprint arXiv:2310.00305*.

Zhou, G.; Hong, Y.; and Wu, Q. 2024. Navgpt: Explicit reasoning in vision-and-language navigation with large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 7641–7649.



## Appendix A

### GPT-4 Prompts and Responses for Urban TbT Scenario

**Prompt:** Plan a trip using turn-by-turn instructions from George Mason University Shenandoah Parking Deck to Garage C of the Nationals Stadium to arrive 2 hours before the first-weekend game in August 2024.

**Response:** To provide a detailed trip plan, I need the exact date and time of the game you plan to attend. If you don't have the exact date yet, please provide an estimated time for planning purposes (e.g., an afternoon or evening game). This will help ensure you arrive 2 hours before the game.

**Follow-up Q1:** What's the first weekend in August of 2024?

**Response:** The first weekend in August 2024 is August 3rd and 4th. If you are planning to attend a game on either of these dates, please specify the time of the game so I can plan your trip accordingly.

**Follow-up Q2:** Can you find a game in Nationals stadium on August 3rd or 4th?

**Response:** The Washington Nationals have games scheduled on the first weekend of August 2024 at Nationals Park. They will be playing the Milwaukee Brewers on both August 3rd and 4th. Here are the game details:

- Saturday, August 3rd, 2024: The game starts at 4:05 PM.
- Sunday, August 4th, 2024: The game starts at 1:35 PM (MLB.com).

Given these timings, to arrive 2 hours before the game, you should plan to reach Nationals Park by 2:05 PM on Saturday and by 11:35 AM on Sunday.

#### **Turn-by-Turn Directions:**

- Start: Shenandoah Parking Deck, George Mason University, Fairfax, VA
- Head east on Patriot Cir toward George Mason Blvd.
- Turn right onto VA-123 S.
- Merge onto I-495 N/Capital Beltway toward Maryland.
- Take exit 57B for I-395 N toward Washington.
- Continue on I-395 N (passing through parts of DC).
- Take exit 6 for C Street SW toward Nationals Park.
- Continue on C Street SW.
- Turn right onto South Capitol Street SE.
- Turn left onto N Street SE.
- Turn right onto First Street SE.
- Garage C will be on your right, just past Nationals Park.

The estimated driving time is around 45 minutes to 1 hour, but be sure to account for potential traffic, especially closer to the stadium. For a 4:05 PM game on August 3rd, you should leave by 1:00 PM at the latest. For the 1:35 PM game on August 4th, you should leave by 10:30 AM.

### GPT-4 Prompts and Responses for Easy VLN Scenario

**Prompt:** Plan a trip using visual landmarks for people who cannot read road sign instructions from George Mason University Nguyen Engineering Building to Johnson Center.

#### **Response:**

- **Starting Point: Nguyen Engineering Building** Begin at the main entrance of the Nguyen Engineering Building. You will see a large, modern building with a distinctive glass facade.
- **First Visual Landmark: The Statue** Walk straight ahead towards the statue of George Mason, which is a prominent bronze statue in a small plaza.
- **Second Visual Landmark: The Clock Tower** Continue walking past the statue and head towards the clock tower, a tall structure with a large clock face, easily visible from a distance.
- **Third Visual Landmark: Mason Pond** From the clock tower, take the path that leads to Mason Pond, a serene body of water with a walking path around it. Follow the path around the pond.
- **Fourth Visual Landmark: The Bridge** After walking around Mason Pond, you will see a small pedestrian bridge. Cross this bridge.
- **Fifth Visual Landmark: Harris Theatre** Once you cross the bridge, you will see the Harris Theatre to your right. It's a building with a brick exterior and a modern design.
- **Sixth Visual Landmark: Student Union Building I (SUB I)** Continue straight ahead until you reach the Student Union Building I (SUB I). This building is noticeable with its large, welcoming entrance and banners.

- **Final Visual Landmark: Johnson Center** From SUB I, walk straight towards the Johnson Center. It's a large, multi-story building with a glass atrium and a prominent entrance.

By following these visual landmarks, you can navigate easily from the Nguyen Engineering Building to the Johnson Center without needing to read road signs.

### GPT-4 Prompts and Responses for Suburban TbT Scenario

**Prompt:** Plan a trip using turn-by-turn instructions from George Mason University Shenandoah Parking Deck to go to the Hogsback Mountain Paintball Center on the next federal holiday when the center is open and in time to spend 4 hours on the range.

**Response:** To plan your trip from George Mason University Shenandoah Parking Deck to Hogsback Mountain Paintball Center on the next federal holiday when the center is open, follow these steps:

**Next Federal Holiday and Operating Hours** The next federal holiday is Independence Day, which falls on Thursday, July 4th. Hogsback Mountain Paintball Center is open every Saturday and Sunday from 10 AM to 5 PM, and for pre-scheduled groups of 8 or more, it is open seven days a week from 10 AM to 5 PM.

**Travel Time and Schedule** To spend 4 hours on the range, you'll want to arrive at 10 AM when the center opens. Given that it takes about 1 hour to drive from George Mason University to Hogsback Mountain Paintball Center, you should plan to leave by 8:45 AM to allow for any unexpected delays.

#### Turn-by-Turn Directions:

- Start at George Mason University Shenandoah Parking Deck.
- Head northwest on Patriot Circle toward Mason Pond Drive.
- Take the 1st right onto Mason Pond Drive.
- Turn left onto University Drive.
- Merge onto I-66 W via the ramp to Manassas/Front Royal.
- Take exit 40 to merge onto US-15 N/James Madison Hwy toward Leesburg.
- Stay on US-15 N for approximately 17 miles.
- Turn right onto Hogback Mountain Road (VA-651).
- Go two blocks and enter West View Farm. Follow signs to Hogsback Mountain Paintball Center.

#### Activity Schedule:

- Arrive by 9:45 AM for setup.
- Play from 10 AM to 2 PM.

### GPT-4 Prompts and Responses for Medium VLN Scenario

**Prompt:** Plan a trip using visual landmarks for people who cannot read road sign instructions from George Mason University Nguyen Engineering Building to INTO Mason Center.

**Response:** Here's a visual-based guide to help you navigate from the Nguyen Engineering Building at George Mason University to the INTO Mason Center:

- **Starting Point: Nguyen Engineering Building** Exit the Nguyen Engineering Building: Locate the main entrance/exit of the Nguyen Engineering Building. It's a large, modern building with a glass facade.
- **Landmark 1: Johnson Center** Head towards the Johnson Center: As you exit, you will see a large clock tower and a brick-paved plaza area. Head towards this plaza. The Johnson Center (JC) is a prominent building with a distinctive clock tower and a large, open area in front.
- **Landmark 2: George Mason Statue** Walk towards the George Mason Statue: From the Johnson Center, look for the statue of George Mason. It's a large bronze statue situated in a small plaza area with benches. Walk towards the statue, which will be straight ahead from the Johnson Center.
- **Landmark 3: Harris Theatre** Proceed to the Harris Theatre: From the George Mason Statue, turn to your right. You will see a building with a rounded front entrance – this is the Harris Theatre. Walk past the Harris Theatre, keeping it on your right.
- **Landmark 4: Fenwick Library** Continue towards Fenwick Library: After passing the Harris Theatre, you will come across the Fenwick Library. It is a tall, modern building with a lot of glass windows. Continue walking straight, keeping the Fenwick Library on your left.
- **Landmark 5: Merten Hall** Approach Merten Hall: As you walk past Fenwick Library, you will see a large building with a distinctive circular design – this is Merten Hall. Walk towards Merten Hall, crossing any pedestrian pathways.

- **Final Destination: INTO Mason Center** Arrive at INTO Mason Center: From Merten Hall, look for the INTO Mason Center. It's located nearby and is marked with clear signage that includes the INTO Mason logo. The building is modern with large windows and the INTO Mason logo prominently displayed.

**Summary of Visual Landmarks:**

- **Nguyen Engineering Building:** Glass facade.
- **Johnson Center:** Clock tower and brick-paved plaza.
- **George Mason Statue:** Bronze statue.
- **Harris Theatre:** Rounded front entrance.
- **Fenwick Library:** Tall building with glass windows.
- **Merten Hall:** Circular design.
- **INTO Mason Center:** INTO Mason logo and modern facade.

**GPT-4 Prompts and Responses for Rural TbT Scenario**

**Prompt:** Plan a trip using turn-by-turn instructions from George Mason University Shenandoah Parking Deck to the Shot Tower State Park to arrive on the second day of guided tours on the week of September 30.

**Response:** To plan your trip from George Mason University Shenandoah Parking Deck to Shot Tower State Park for the second day of guided tours in the week of September 30, follow these turn-by-turn instructions. The second day of guided tours typically falls on a Tuesday, assuming the tours start on a Monday.

**Departure Point:** George Mason University Shenandoah Parking Deck **Destination:** Shot Tower State Park, 176 Orphanage Dr, Austinville, VA 24312 **Estimated Travel Time:** Approximately 4 hours and 30 minutes (subject to traffic and road conditions)

**Route Instructions:**

- Start at George Mason University Shenandoah Parking Deck:
- Head west on University Dr toward Patriot Cir.
- Turn left onto Patriot Cir.
- At the traffic circle, take the 1st exit onto Ox Rd.
- Merge onto I-66 W:
  - Follow signs for I-66 W/US-50 W/US-29 S.
  - Merge onto I-66 W via the ramp to Front Royal.
- Continue on I-66 W for about 63 miles.
- Take exit 1A to merge onto I-81 S toward Roanoke:
  - Continue on I-81 S for approximately 185 miles.
- Take exit 81 to merge onto I-77 S toward Charlotte:
  - Continue on I-77 S for about 8 miles.
- Take exit 24 for US-52 S toward Hillsville:
  - Merge onto US-52 S/Caron Rd.
  - Continue to follow US-52 S for about 12 miles.
- Turn right onto VA-69 S:
  - Follow VA-69 S for about 5 miles.
  - VA-69 S turns slightly left and becomes VA-638 S.
- Arrive at Shot Tower State Park:
  - Continue straight onto Orphanage Dr.
  - The destination will be on your left at 176 Orphanage Dr, Austinville, VA 24312.

**Follow-up Q1:** Can you find the second day of guided tours on the week of September 30 and plan the trip for that day?

**Response:** Guided tours at Shot Tower State Park are offered on Saturdays and Sundays from September 7 through October 27. For the week of September 30, the second day of guided tours will be on Sunday, October 6, 2024.

**Travel Plan: Departure Point:** George Mason University Shenandoah Parking Deck **Destination:** Shot Tower State Park, 176 Orphanage Dr, Austinville, VA 24312 **Travel Time:** Approximately 4 hours and 30 minutes

**Tips for the Trip:**

- **Check Traffic Updates:** Make sure to check for real-time traffic updates before starting your trip.

- **Rest Stops:** Plan for rest stops along the way, especially if driving alone.
- **Guided Tours Schedule:** Confirm the guided tour schedule for Shot Tower State Park in advance to ensure you arrive on time.
- **Travel Essentials:** Carry necessary travel essentials, including snacks, water, and a first aid kit.

### ChatGPT-4 Prompts and Responses for Hard VLN Scenario

**Prompt:** Plan a trip using visual landmarks for people who cannot read road sign instructions from George Mason University Nguyen Engineering Building to Osher Lifelong Learning Institute (OLLI) at George Mason University.

**Response:**

- **Starting Point: Nguyen Engineering Building (Engr)**
  - **Landmark:** The building has a distinctive modern design with large windows and a prominent entrance.
  - **Visual Cue:** Look for the statue of a man holding a gear near the entrance.
- **Step 1: Walk towards the Johnson Center**
  - **Path:** Exit the Nguyen Engineering Building and head straight towards the central part of the campus.
  - **Landmark:** You will pass by the Science and Tech I building on your left.
  - **Visual Cue:** Look for the large green area with trees and benches.
- **Step 2: Reach the Johnson Center (JC)**
  - **Landmark:** A large building with a clock tower and multiple entrances. It's a hub for students with a food court, library, and study areas.
  - **Visual Cue:** Look for the clock tower and glass façade.
- **Step 3: Head towards the Center for the Arts**
  - **Path:** From the Johnson Center, walk straight across the central plaza.
  - **Landmark:** You will pass the Fenwick Library on your left.
  - **Visual Cue:** The library has a modern glass extension.
- **Step 4: Pass the Center for the Arts**
  - **Landmark:** A large building with a distinctive curved front and banners for upcoming performances.
  - **Visual Cue:** Look for the banners and the curved architecture.
- **Step 5: Walk towards Mason Hall**
  - **Path:** Continue walking past the Center for the Arts.
  - **Landmark:** You will see Mason Hall ahead, which has a traditional design with brick and white columns.
  - **Visual Cue:** Look for the white columns and brick façade.
- **Step 6: Head towards the Finley Building**
  - **Path:** Walk past Mason Hall and continue straight.
  - **Landmark:** You will see the Finley Building, which is a smaller building with a simple design.
  - **Visual Cue:** Look for the building with a straightforward brick design.
- **Step 7: Arrive at Osher Lifelong Learning Institute (OLLI)**
  - **Landmark:** The Osher Lifelong Learning Institute is housed in a smaller building with clear signage outside.
  - **Visual Cue:** Look for the OLLI sign and the building with a welcoming entrance.

**Summary of Visual Landmarks:**

- **Nguyen Engineering Building:** Statue of a man holding a gear.
- **Science and Tech I Building:** Large green area with trees and benches.
- **Johnson Center:** Clock tower and glass façade.
- **Fenwick Library:** Modern glass extension.
- **Center for the Arts:** Banners and curved architecture.
- **Mason Hall:** White columns and brick façade.
- **Finley Building:** Simple brick design.
- **OLLI:** Clear OLLI sign and welcoming entrance.

This route should help guide those who cannot read road signs by using prominent visual landmarks to navigate through the campus.

## Gemini Prompts and Responses for Urban TbT Scenario

**Prompt:** Plan a trip using turn-by-turn instructions from George Mason University Shenandoah Parking Deck to Garage C of the Nationals Stadium to arrive 2 hours before the first weekend game in August 2024.

**Response:** Certainly! To plan your trip, I looked for the first weekend game of the Washington Nationals in August 2024. Their first home game in August is on Friday, August 2nd, at 7:15 PM. Here's how to get from George Mason University Shenandoah Parking Deck to Garage C at Nationals Park with an estimated travel time of 1 hour and 15 minutes, arriving 2 hours before the game.

**Departure Point:** George Mason University Shenandoah Parking Deck **Destination:** Nationals Park Garage C

### Turn-by-Turn Directions:

- Depart from George Mason University Shenandoah Parking Deck.
- Head west on University Drive for 0.3 miles.
- Turn left to stay on University Drive for 0.4 miles.
- Turn right to merge onto VA-286 E for 4.9 miles.
- Take Exit 15 towards I-95 N/Richmond:
  - Keep left at the fork and follow signs for I-95 N/Richmond for 1.3 miles.
  - Use the left 2 lanes to continue toward I-95 N for 1.7 miles.
- Take Exit 54B to merge onto I-395 N for 3.9 miles.
- Take Exit 9B to merge onto I-66 E for 2.0 miles.
- Take Exit 43B to merge onto I-295 S for 1.2 miles.
- Take Exit 1B onto South Capitol St E for 0.8 miles.
- Turn right onto Potomac Ave SE for 0.4 miles.
- Turn right onto N St SE for 0.2 miles.
- Turn left onto Half St SE for 0.4 miles.
- Your destination will be on the left. Turn left onto Lots C & D to arrive at Garage C of Nationals Park.

**Total Distance:** 18.3 miles **Estimated Travel Time:** 1 hour and 15 minutes

**Follow-up Q1:** What about the plan to arrive 2 hours before the first weekend game in August 2024?

**Response:** To arrive at Nationals Park 2 hours before the first weekend game in August 2024, consider the following:

- **Game Schedule:** The Nationals' first home game in August 2024 is on Friday, August 2nd, at 7:10 PM (MLB Nationals Schedule).
- **Target Arrival Time:** Since the game starts at 7:10 PM, aim to arrive at Nationals Park Garage C by 5:10 PM. This provides a 2-hour buffer to account for traffic and parking.
- **Departure Time:** Use online mapping tools to check live traffic conditions on the day of the game. Based on the route described previously (George Mason University Shenandoah Parking Deck to Nationals Park Garage C via I-495 S), traffic should be manageable with a travel time of around 35 minutes. However, accounting for potential delays, target a departure from George Mason University by 3:00 PM.

**Follow-up Q2:** Friday is not weekend **Response:** Cannot find the first weekend game.

## Gemini Prompts and Responses for Suburban TbT Scenario

**Prompt:** Plan a trip using visual landmarks for people who cannot read road sign instructions from George Mason University Shenandoah Parking Deck to Hogsback Mountain Paintball Center on the next federal holiday when the center is open and in time to spend 4 hours on the range.

**Response:** Google Maps.

**Follow-up Q1:** The result should be a written route description using visual landmarks, not Google Maps.

**Response:** Certainly, here's a descriptive route using visual landmarks from George Mason University Shenandoah Parking Deck to Hogsback Mountain Paintball Center:

**Departure Point:** George Mason University Shenandoah Parking Deck **Destination:** Hogsback Mountain Paintball Center

### Landmark-Based Directions:

- Start at the Shenandoah Parking Deck at George Mason University. Look for a red brick building with a large sign that says "Shenandoah Parking Deck."
- Exit the parking deck and make a left onto University Drive.
- Continue on University Drive for about 1.2 miles until you reach a large intersection with traffic lights. Here, you will see a prominent blue water tower on your left.

- Turn right at the intersection onto VA-286 E/Fairfax County Parkway. You'll pass a McDonald's restaurant on the corner.
- Follow VA-286 E for about 10.3 miles. You'll pass under railroad tracks and then a large shopping center with a Wegmans grocery store on your right.
- Stay straight through the intersection with Braddock Road (VA-620) and continue on VA-286 E.
- In about 5.2 miles, you will come to a cloverleaf interchange with Dulles Greenway Toll Road on your left. Stay on VA-286 E.
- Continue for 3.7 miles and use the left 2 lanes to merge slowly onto VA-7 W toward Leesburg.
- Follow VA-7 W for 2.3 miles. You will pass a gas station and convenience store called Sheetz on your right.
- Make a left onto Hogback Mountain Road. There is a green metal sign with white lettering for Hogback Mountain Road.
- Follow Hogback Mountain Road for 1.3 miles. The road will become curvy and rural. You will pass a large white house on your right.
- Hogsback Mountain Paintball will be on your left. You will see a large wooden sign with a paintball splat design and their logo.

**Follow-up Q2:** How to plan to visit the center on the next federal holiday when the center is open and in time to spend 4 hours on the range.

**Response:** The next federal holiday in the United States is Fourth of July, which falls on a Thursday this year (July 4th, 2024). Since Hogsback Mountain Paintball Center is closed on Thursdays, the ideal day to visit the center would be on the following day, Friday, July 5th.

- **Call the Center to Confirm Hours:** Before heading out, it's best to confirm the center's hours of operation for the specific day you plan to visit. You can reach Hogsback Mountain Paintball Center by calling +1 703-777-0057. Their hours are typically from 9:00 AM to 5:00 PM on Fridays.

## Gemini Prompts and Responses for Rural Tbt Scenario

**Prompt:** Plan a trip using turn-by-turn instructions from George Mason University Shenandoah Parking Deck to Shot Tower State Park to arrive on the second day of guided tours during the week of September 30.

**Response:** Google Maps, suggested Tuesday, September 3rd.

**Follow-up Q1:** The result should be a written route description, not Google Maps. Also, September 3rd is not during the week of September 30.

**Response:** You're right, September 3rd isn't during the week of September 30. Here's a revised plan with written directions to Shot Tower State Park:

**Trip Date:** To arrive on the second day of guided tours during the week of September 30, aim for Tuesday, October 1.

**Estimated Distance and Travel Time:** The drive is approximately 290 miles and takes about 4 hours and 29 minutes.

### Turn-by-Turn Directions:

- Start at the Shenandoah Parking Deck at George Mason University.
- Exit the parking deck and turn right onto University Drive.
- Continue straight on University Drive for about 1.5 miles until you reach VA-286 E/Fairfax County Parkway.
- Take VA-286 E/Fairfax County Parkway for approximately 12.5 miles.
- Merge onto I-66 E via the exit on the left for I-66 East/Washington. Stay on I-66 E for 45 miles.
- Take exit 40B for US-50 E/Washington Dulles Toll Rd East. Keep left at the fork and follow signs for US-50 E. This will be a toll road.
- Stay on US-50 E for about 134 miles.
- Take exit 172 for VA-16 E towards Christiansburg. Stay right at the fork and follow signs for VA-16 E.
- Continue on VA-16 E for approximately 37 miles.
- Turn left onto Orphanage Dr and continue for about 0.3 miles.
- Your destination will be on the right. You have arrived at Shot Tower State Park!

**Follow-up Q2:** September 24 is also not during the week of September 30.

**Response:** Tuesday, October 1st.

## Mistral Prompt and Response for Urban TbT Scenario

**Prompt:** Plan a trip using turn-by-turn instructions from George Mason University Shenandoah Parking Deck to Garage C of the Nationals Stadium to arrive 2 hours before the first weekend game in August 2024.

**Response:** Sure, I'd be happy to help you plan your trip. Please note that I'm providing a general plan based on typical traffic patterns and public transportation schedules, but real-time conditions may vary. Also, I don't have real-time access to game schedules, so I'll assume the game is on Saturday, August 3, 2024.

**Departure Point:** George Mason University Shenandoah Parking Deck, 4400 University Dr, Fairfax, VA 22030 **Destination:** Nationals Parking Garage C, 1100 Half St SW, Washington, DC 20024

### Turn-by-Turn Directions:

- Start at George Mason University Shenandoah Parking Deck.
- Get on I-66 E from University Dr and VA-123 S/Chain Bridge Rd.
- Head southeast on Patriot Cir toward Aquia Creek Ln, then turn right onto Aquia Creek Ln.
- Turn right onto University Dr and use the left 2 lanes to turn left onto VA-123 S/Chain Bridge Rd.
- Use the right lane to take the ramp onto I-66 E.
- Follow I-66 E to US-50 E/Constitution Ave NW in Washington.
- Merge onto I-66 E and use the right 2 lanes to take exit 72 toward US-50 E/Rosslyn/Memorial Bridge.
- Continue onto US-50 E/Arlington Blvd, then onto US-50 E/Constitution Ave NW.
- Take 7th St SW to Half St SW by turning right onto 7th St SW, left onto C St SW, then right onto Half St SW.
- Arrive at Nationals Parking Garage C.

**Total Distance:** Approximately 25 miles **Estimated Travel Time:** About 40 minutes without traffic. Since it's a weekend trip, allow 2.5 to 3 hours before the game start time to account for potential traffic and parking delays.

**Additional Note:** Please check the game's start time and verify this route using a real-time navigation service like Google Maps or Waze before you depart. Safe travels!

**Follow-up Q1:** What's the first weekend in August of 2024?

**Response:** In 2024, the first weekend of August will start on Saturday, August 3rd, and end on Sunday, August 4th. These dates could be useful for planning your trip to Nationals Stadium or any other weekend activities you have planned.

**Follow-up Q2:** Can you find a game at Nationals Stadium on August 3rd or 4th?

**Response:** I'm an AI and don't have real-time access to sports schedules or the ability to browse the internet. However, you can check the official Major League Baseball (MLB) website or the Washington Nationals' website for their 2024 schedule. Once the schedule is released, you'll be able to see if there's a game on August 3rd or 4th, 2024, and plan your trip accordingly.

## Mistral Prompt and Response for Suburban TbT Scenario

**Prompt:** Plan a trip using visual landmarks for people who cannot read road sign instructions from George Mason University Shenandoah Parking Deck to Garage C of the Nationals Stadium to arrive 2 hours before the first weekend game in August 2024.

**Response:** Sure, I'd be happy to help you plan this trip. Please note that the exact date of the next federal holiday will depend on the current year, and you should verify that Hogsback Mountain Paintball Center is open on that day. Also, remember to check their operating hours to ensure you arrive in time to spend 4 hours on the range.

Here are the turn-by-turn driving instructions from George Mason University Shenandoah Parking Deck to Hogsback Mountain Paintball Center:

- Start at George Mason University Shenandoah Parking Deck, 4400 University Dr, Fairfax, VA 22030.
- Head east on Shenandoah Ln toward Patriot Cir.
- Turn left onto Patriot Cir.
- Turn right onto University Dr.
- Use the left 2 lanes to turn left onto Braddock Rd.
- Use the right lane to take the Interstate 66 W ramp to Front Royal.
- Merge onto I-66 W.
- Take exit 43A-43B for US-17 toward Warrenton/Falmouth.
- Keep right at the fork, follow signs for US-17 S/Warrenton, and merge onto US-17 S.
- Turn right onto VA-29 S/Lee Hwy (signs for Culpeper).
- Turn left onto VA-229 S/Rixeyville Rd.
- Turn right onto VA-670 W/Viewtown Rd.
- Continue onto Hogsback Rd.

The destination, Hogsback Mountain Paintball Center, will be on your right: 15540 Hogsback Rd, Amissville, VA 20106. This route is approximately 55 miles and takes about 1 hour and 15 minutes without traffic. To ensure you have 4 hours on the range, I recommend checking the paintball center's opening hours and aiming to arrive at least 15-30 minutes before that to allow for any unexpected delays and to get geared up and ready to play.

Please verify all this information before setting out, as road conditions, opening hours, and holiday schedules can change. Safe travels!

**Follow-up Q1:** Can you find the date of the next federal holiday in 2024 and verify whether Hogsback Mountain Paintball Center is open on that day or not?

**Response:** I'm an AI and don't have the ability to browse the internet in real-time, so I can't verify the operating hours of Hogsback Mountain Paintball Center on specific dates. However, I can tell you that the next federal holiday in 2024 after the time of this conversation is likely to be Independence Day, which falls on Thursday, July 4, 2024.

To find out if Hogsback Mountain Paintball Center is open on that day, I recommend visiting their official website or contacting them directly via phone or email. This way, you can get the most accurate and up-to-date information about their holiday schedule.

## Mistral Prompt and Response for Suburban VLN Scenario

**Prompt:** Plan a trip using visual landmarks for people who cannot read road sign instructions from George Mason University Shenandoah Parking Deck to Garage C of the Nationals Stadium to arrive 2 hours before the first weekend game in August 2024.

**Response:** Sure, I'd be happy to help you plan this trip. Please note that the exact date of the next federal holiday will depend on the current year, and you should verify that Hogsback Mountain Paintball Center is open on that day. Also, remember to check their operating hours to ensure you arrive in time to spend 4 hours on the range.

**Departure Point:** George Mason University Shenandoah Parking Deck, 4400 University Dr, Fairfax, VA 22030 **Destination:** Hogsback Mountain Paintball Center, 15540 Hogsback Rd, Amissville, VA 20106

### Visual Landmark-Based Directions:

- Begin at George Mason University Shenandoah Parking Deck.
- Head east on Shenandoah Ln, passing the distinctive Johnson Center on your right, until you reach Patriot Cir.
- Turn left onto Patriot Cir, where you'll soon see Mason Pond on your right.
- Turn right onto University Dr, marked by a large brick archway.
- Take the left lane to turn left onto Braddock Rd, then head toward the I-66 ramp.
- Look for the tall highway signage marking Interstate 66 W ramp to Front Royal, then merge onto I-66 W.
- Continue on I-66 W until you reach a broad green landscape with exit signs, then take exit 43A-43B for US-17 towards Warrenton.
- Keep right at the fork, merging onto US-17 S, and look for wide fields to your left.
- When you reach a series of stone-lined fences along the road, turn right onto VA-29 S/Lee Hwy.
- After a while, turn left onto VA-229 S/Rixeyville Rd, and watch for rolling hills on your right.
- Turn right onto VA-670 W/Viewtown Rd, and continue as the road narrows through dense tree-lined sections.
- Follow the road as it transitions to Hogsback Rd, where you'll spot signs for paintball. The paintball center will be on your right.

**Total Distance:** Approximately 55 miles **Estimated Travel Time:** About 1 hour and 15 minutes without traffic. To ensure a full 4 hours on the range, check opening hours and aim to arrive 15-30 minutes early to account for any unexpected delays and to get geared up.

**Additional Note:** Verify the information about the paintball center's opening hours and holiday schedule, as these can vary. Safe travels!

**Follow-up Q1:** Can you find the date of the next federal holiday of 2024 and verify whether Hogsback Mountain Paintball Center is open on that day or not?

**Response:** I'm an AI and don't have the ability to browse the internet in real-time, so I can't verify the operating hours of Hogsback Mountain Paintball Center on specific dates. However, I can tell you that the next federal holiday in 2024 after the time of this conversation is likely to be Independence Day, which falls on Thursday, July 4, 2024.

To find out if Hogsback Mountain Paintball Center is open on that day, I recommend visiting their official website or contacting them directly via phone or email. This way, you can get the most accurate and up-to-date information about their holiday schedule.



## Mistral Prompt and Response for Rural TbT Scenario

**Prompt:** Plan a trip using turn-by-turn instructions from George Mason University Shenandoah Parking Deck to Shot Tower State Park to arrive on the second day of guided tours on the week of September 30.

**Response:** Sure, I'd be happy to help you plan this trip. Please note that I'm providing a general plan based on typical conditions, but actual traffic and weather conditions can affect your travel time.

**Departure Date and Time:** The second day of guided tours in the week of September 30 is likely to be October 1, assuming tours start on Monday. To ensure you arrive in time for the tours, I recommend starting your journey early in the morning.

**Starting Point:** George Mason University Shenandoah Parking Deck, Fairfax, VA **Destination:** Shot Tower State Park, Austin, NV

### Initial Directions:

- Head northeast on Shenandoah Ln toward Patriot Cir.
- Turn right onto Patriot Cir.
- Turn left onto Braddock Rd.
- Use the left 2 lanes to turn left onto VA-123 S/Ox Rd.
- Use the right lane to take the Interstate 66 W ramp to Front Royal, and merge onto I-66 W.
- Continue onto US-50 W/US-340 W/US-17 S/Millwood Ave (follow signs for Winchester).

**Continue on US-50 W across multiple states:** This route covers several states (West Virginia, Ohio, Indiana, Illinois, Missouri, Kansas, Colorado, Utah, and Nevada). Take regular rest stops as needed to ensure a safe trip.

### In Nevada:

- Continue on US-50 W. Take the US-95 N exit toward Reno.
- Turn right onto US-95 ALT N.
- Turn left onto NV-305 N.
- Turn right onto NV-362 E.
- Turn left onto NV-722 E.

**Final Destination:** Shot Tower State Park will be on your right. This trip spans approximately 2,500 miles and takes around 36 hours without traffic. Given the long distance, I recommend breaking the journey into multiple days, ensuring you rest and stay safe. Please verify the park's opening hours and tour schedules before your trip. Safe travels!

**Follow-up Q1:** What about the plan to arrive on the second day of guided tours on the week of September 30?

**Response:** I can't confirm an exact schedule here, so let's assume general availability for now.

**Follow-up Q2:** Can you check the date and hours of guided tours on the website and tell me how I should schedule my plan?

**Response:** The park doesn't offer guided tours.

**Follow-up Q3:** Is Shot Tower State Park in Nevada?

**Response:** Yes.

**Follow-up Q4:** Shot Tower State Park is in Virginia.

### Revised Directions:

- Head northeast on Shenandoah Ln toward Patriot Cir.
- Turn right onto Patriot Cir.
- Turn left onto Braddock Rd.
- Use the left 2 lanes to turn left onto VA-123 S/Ox Rd.
- Use the right lane to take the Interstate 66 W ramp to Front Royal, and merge onto I-66 W.
- Continue onto US-50 W/US-340 W/US-17 S/Millwood Ave (follow signs for Winchester).
- Continue onto US-50 W/US-17 S/John Mosby Hwy.
- Turn right onto VA-16 S/VA-42 S (signs for Tye River Turnpike/Amherst).
- Turn left onto US-221 S/W Main St.
- Turn right onto VA-643 W/VA-652 S/Fort Chiswell Rd.
- Turn left onto VA-652 S/Max Meadows Rd.
- Turn right onto VA-613 S/Shot Tower Rd.

Shot Tower State Park will be on your right.