Proceedings of the Annual General Donald R. Keith Memorial Conference
West Point, New York, USA
May 2, 2024
A Regional Conference of the Society for Industrial and Systems Engineering

Autonomous Ground Vehicle Simulation: Usability and Evaluation Study

Jalen Caldwell, Jared Cross, Patrick Oglesbee, and Brandon Thompson

Department of Systems Engineering,
United States Military Academy,
West Point, NY 10996

Corresponding author's Email: poglesbee21@gmail.com

Author Note: The authors would like to extend our thanks to the Engineer Research and Development Center (ERDC) Information Technology Laboratory (ITL) and Institute for Systems Engineering Research (ISER) for providing the opportunity to participate in Autonomous Ground Vehicle (AGV) simulation development and research. The views expressed herein are those of the authors and do not reflect the position of the United States Military Academy, the Department of the Army, or the Department of Defense.

Abstract: Autonomous Ground Vehicles (AGVs) have the potential to be a significant force multiplier on the battlefield, reducing risk to soldiers and increasing logistical capability. These AGVs must be thoroughly tested under various conditions before being incorporated into missions alongside soldiers. This requires not only physical testing but virtual testing through simulation to reduce cost and increase the scope of testing environments and conditions. The Engineering Research and Development Center (ERDC) has developed a Software-in-the-Loop (SIL) simulation to test AGVs and supplied the research team with a SIL. The researchers in this study conducted a usability study of the SIL to provide feedback to ERDC. Then a design of experiments was conducted to test the impacts of simulated weather effects on AGV performance under three Operational Design Domains (ODDs). The results showed the virtual AGV performed better under simulated rain and fog weather conditions as compared to normal operating conditions.

Keywords: Simulation, Autonomous Ground Vehicle, Virtual Testing

1. Introduction

The United States Army is developing Autonomous Ground Vehicles (AGVs) to operate alongside humans to provide new ways of fighting and offer new capabilities on the battlefield (Heckmann, Magnuson, & Park, 2023). Developing, testing, and integrating AGVs within the Army is a time-consuming and costly endeavor. The Engineer Research and Development Center (ERDC) is tasked with developing an AGV simulation that can measure and test vehicle performance in a virtual environment under various Operational Design Domains (ODDs) to reduce testing costs and enable a broader range of testing scenarios. ODDs are distinct scenarios defined by factors including weather, terrain, and vehicle configurations that are used to create a comprehensive set of conditions to test and validate autonomous behavior (Sun, Deng, Chu, Li, & Cao, 2022). By using simulation, users can run numerous tests while being able to control the environment and systematically assess the effects of independent variables on AGV performance. It also offers a potential increase in the efficiency of military operations by increasing speed and accuracy and decreasing materials needed to ensure mission success. Leading this endeavor, ERDC developed a Virtual Autonomous Navigation Environment (VANE) Software-in-the-Loop (SIL) simulation, a high performance computing simulation testbed (Jones et. al., 2008). Comprised of four integral systems, VANE SIL includes an autonomy stack, the Environmental and Sensory Engine, the VIGOR Unreal Engine 4, and the Warfighter Machine Interface. These interconnected computer systems facilitate seamless communication between hardware and software components, thereby generating a comprehensive virtual simulation testing environment. The research team aimed to augment ERDC's virtual AGV testing through this study. The study was conducted in three phases. In Phase I, the team worked with researchers from ERDC to establish a Software-in-the-Loop simulation at the United States Military Academy at West Point. The task for Phase II involved developing a comprehensive user guide for the simulation and conducting a usability study to provide feedback to ERDC on not only the ease of use of the SIL simulation but also to gain useful insights on potential software improvements. Phase III consisted of a design of experiments to measure the effects of weather and terrain on AGV performance within the simulation.

2. Phase I: Establish the West Point VANE SIL

The West Point VANE SIL, shown in Figure 1, is a set of integrated computers and software programs developed by ERDC to facilitate testing of autonomy navigation systems within the Department of Defense (DoD). VANE merges detailed models of vehicles, sensors, and environments (Jones et al., 2008). This allows for realistic computer-based tests of autonomous vehicle systems working in complex virtual settings. Four computers work in unison within the West Point VANE SIL, each containing a specific software program: the autonomy software, virtual environment software, sensor software, and user interface software. The autonomy software uses algorithms to continuously develop a navigation path for the autonomous vehicle. The virtual environment software provides the virtual terrain model and is used to simulate the vehicle dynamics and provides the vehicle interface for the autonomy software to monitor and control performance. The user interface provides the human-computer interaction, allowing a user to input a mission, execute a mission, and remote operate the vehicle. The last essential component of the simulation is the sensor system which creates a virtual Light Detection and Ranging (LiDAR) scanner that interprets the virtual environment. LiDAR is essential because it provides precise three-dimensional information about the surrounding environment, enabling autonomous systems to navigate safely and efficiently (Dominguez, 2011). The Environmental and Sensor Engine (sensor software) computer contains software that simulates the LiDAR of the virtual environment for the SIL system. The research team worked with ERDC personnel to test and troubleshoot the SIL. The team recommended improvements to the initial SIL concept to improve the user interface to include adding two large monitors allowing the user to view the LiDAR sensor image, virtual environment, and user interface simultaneously.

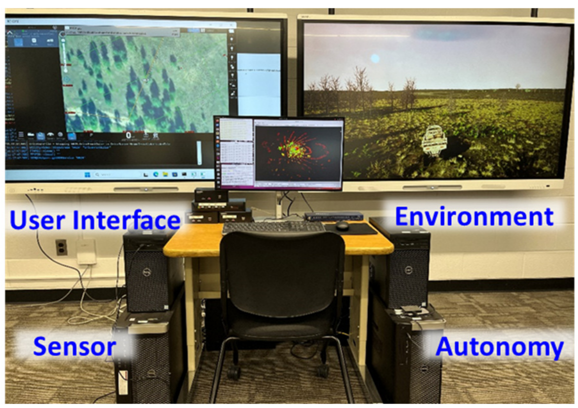

Figure 1. West Point VANE SIL Simulation Setup

3. Phase II: Usability Study

The research team developed a West Point SIL User Guide detailing step-by-step directions including powering on the system, operating and troubleshooting the simulation, collecting data, and shutting down the system. The goal of the user guide was to facilitate an effortless navigation through the entire simulation process, ensuring that participants can successfully execute the simulation without assistance, avoiding any significant challenges, and within a reasonable timeframe.

We conducted a small-scale usability study with five participants to evaluate the ease of use of the SIL and provide feedback to ERDC. In addition, the usability study was used to recognize and resolve any possible challenges or constraints in the execution of the SIL simulation prior to conducting the design of experiments. We provided the participants with a brief orientation of the SIL setup to explain the hardware components. The research team developed and used a script to ensure consistency in directions given to participants. Participants were then instructed to follow the user guide to execute the SIL. The instructions had four main categories: startup, data recording, troubleshooting, and shutdown. On average, it took the participants approximately 40 minutes to complete one simulation run. After completion, participants provided feedback using a questionnaire designed by the research team that enabled easy conversion of participant responses to a Technical Support and Operational Analysis (TSOA) evaluation matrix. The survey primarily utilized Likert scale questions to gauge participants' thoughts and emotions while navigating the SIL and was followed by a discussion with the team to capture any additional feedback (Brooke, 1986). The TSOA evaluation matrix was created to evaluate innovative technologies and capabilities that can be used in challenging environments to improve military operations (Blinde, 2023). Figure 2 show a portion of the TSOA evaluation matrix displaying the overall ease of use assessment. The team developed a qualitative assessment as well as a

ISBN: 97819384962-4-0

quantitative score on a scale of 1(lowest score) to 10 (highest score) from Extremely Low probability, Most Likely probability, and Extremely High probability.

Usability Factor 3 (UF3): Ease of Use					
UF3-A: Intuitiveness	**UF3-B: End User Assumptions**	**Notes & Comments**	**EL**	**ML**	**EH**
End user can rapidly discover system functionality without explanation or user guide	System interface assumes common knowledge widely present in the end user population	**UF3-A Notes & Comments:** The system is fairly complex and requires specific instructions on how to operate properly. Most average users are not familiar with using a terminal command prompt in a Linux computer.	10	10	10
			9	9	9
			8	8	8
			7	7	7
End user can quickly understand system functionality, operations, and interface with minimal explanation when equipped with user guide	System interface design has some number of user assumptions that are unreasonable to expect among the end user population	**UF3-B Notes & Comments:** Understanding how the order of starting the simulation and its software components require detailed instructions. Hardware configurations will differ between SIL locations which will require further instruction/orientation. Sequencing order and timing bewteen different tasks also requires detailed instruction.	6	6	6
			5	5	5
			4	4	4
			3	3	3
End user finds system or its interface difficult to understand regardless of explanation with user guide	Design requires insights and understanding well-beyond the end user population		2	2	2
			1	1	1

Figure 2. Usability Factor: Ease of Use from the TSOA

Participants reported a high level of ease in coding within the simulation environment. Those with prior experience in coding terminals demonstrated faster execution of the simulation. Notably, 75% of participants found it effortless to modify the simulation environment. Additionally, participants clearly understood the interface and controls, feeling confident in their ability to operate the simulation with assistance from the user guide. Throughout the study, no participants encountered significant difficulties. However, minor challenges were noted, particularly in navigating the software components and understanding their interrelationships. Minor difficulties were observed during the startup and shutdown processes, primarily due to the strict sequential nature of these procedures. Of the two, participants found the startup process slightly more challenging than shutdown. The research team also found that when troubleshooting, participants needed clear instructions on how to use the remote controller and guidance on returning to specific steps to complete the simulation. Nonetheless, grasping the software's intricacies significantly enhanced participants' overall experience and success. The research team updated and improved the user guide based on feedback and findings of the small-scale usability study.

4. Phase III: AGV Performance Study

4.1 Methodology

The research team was tasked with testing the effects of simulated weather conditions and terrain type on AGV performance. ERDC provided the virtual environment and a mission set consisting of a course with programed waypoints for the AGV to navigate. The course measured almost a full kilometer in total distance through various types of vegetation ranging from semi-dense forest to open terrain with low grass and dirt roads with varying degrees of elevation. Over the last several years, the Army has conducted both physical and virtual tests of AGVs on the same testing course. However, the long distances within each test run limited the number of successful completions of the course without requiring human intervention. In addition, long testing durations resulted in increased variability that made it difficult to compare performance or identify the cause of variation (Adams, Bakre, Kramer, Riekena, & Thompson, 2023). Similar issues and results were found during the initial testing of the West Point SIL and during the usability study. As a result, the testing runs were divided into shorter distances with distinct terrain and vegetation characteristics, known as ODDs, to better isolate their effects on the autonomy, sensor, and vehicle performance.

The research team identified three distinct ODDs based on the type of vegetation to test AGV performance. Figure 3 displays the three ODDs overlayed onto an aerial photograph of the terrain. Researchers from ERDC previously mapped and constructed the same vegetation patterns as seen in the aerial photograph into the virtual environment. ODD Heavy consisted of a semi-densely forested area with a gradual incline at the end of the route. ODD Medium contained less trees, had a gradual decline, and featured a dirt road handrailed by a small berm. Lastly, ODD Light was made up of fairly flat terrain with sparse, small trees at the start and open terrain with low grass and a dirt path for the last half. These three ODDs served as one of the independent measures in the AGV performance study. Weather condition was the second independent measure tested in the AGV performance study. The researchers at ERDC simulated the effects of weather on the AGV by changing the LiDAR sensor attenuation within the sensor software to reduce the perceived friction in the sensor scan. The team was provided with three levels of weather condition, Normal, Rain, and Fog. Figure 4 displays the virtual environment at the start of ODD Heavy and the LiDAR sensor scans at the start location for all three weather conditions.

ISBN: 97819384962-4-0

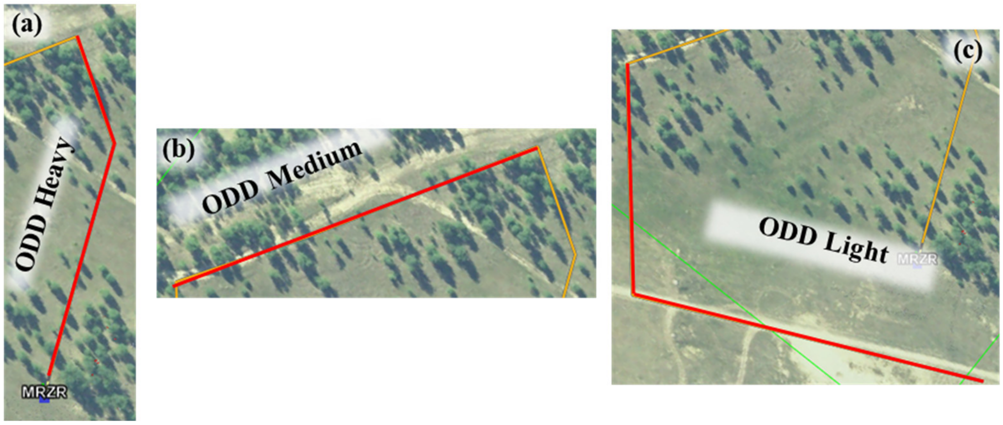

Figure 3. Three Operational Design Domains Tested: Heavy vegetation (a), Medium vegetation (b), Light vegetation (c)

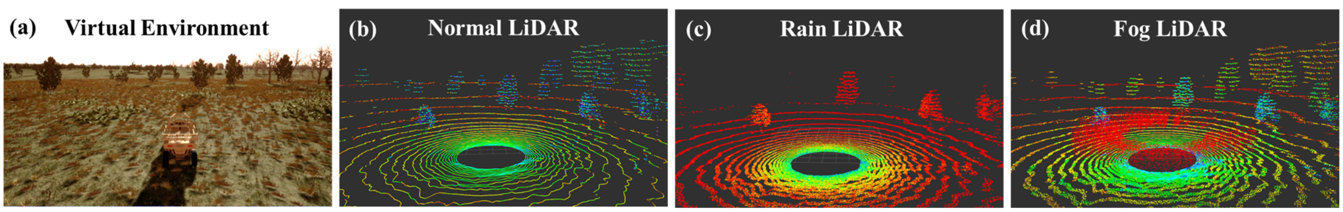

Figure 4. Virtual Environment (a) and LiDAR sensor by Weather Condition, Normal (b), Rain (c), and Fog (d)

A factorial design of experiments was conducted consisting of each combination of dependent measure, terrain with three levels, ODD Heavy, ODD Medium, and ODD Light, and weather condition with three levels, Normal, Rain, and Fog. The researchers simulated 10 successful runs for each combination of factors for a total of 90 successful runs. Simulation runs that failed to complete the course were documented along with the reason for failure and were then replaced by the next successful run. In total, the team conducted 118 simulations to achieve the desired number of successful runs. We recorded data in the form of bag files from executing each virtual run until the vehicle reached the final waypoint of the ODD. We generated new bag files every 25 seconds to limit the file size. Each run contained 5-10 bag files. A bag file acts as a digital container or recording that keeps track of data from various sources, all in one place. It records and keeps track of everything that happens during an experiment or simulation (Turner, 2005). The bag files were combined using Python code and results were exported to a csv file and summarized resulting in an HTML document and Excel® spreadsheet. The HTML document included a visual display of the AGV's route during the simulation and the spreadsheet contained the data for six dependent measures: total distance traveled (measured in meters), maximum speed (kilometers per hour), average speed (kilometers per hour), average moving speed (kilometers per hour), percentage of the time moving (as a percent), and total duration (seconds). The research team hypothesized that the normal weather condition would produce the best results and the AGV would perform optimally on the ODD Light with minimal obstacles.

4.2 Results

Data was analyzed from the first 10 successful runs for each level of weather (Normal, Rain, and Fog) on each ODD (Heavy, Medium, and Light) for a total of 90 virtual runs. The mean and standard deviation were calculated for all dependent measures (Table 1). Mean data for runs under the Normal condition traveled the farthest distance, had the lowest average speed and average moving speed, and as a result, had the highest total duration for all three ODDs. Mean data for runs under the Rain condition traveled the shortest distance, had the highest average speed, and as a result, had the lowest total duration for all three ODDs.

ISBN: 97819384962-4-0

Table 1. Results data by ODD and Weather Condition, Mean (Standard Deviation)

ODD	Weather	Distance Traveled (meters)	Max Speed (km/hr)	Average Speed (km/hr)	Average Moving Speed (km/hr)	Percent Moving (%)	Total Duration (seconds)
Heavy Vegetation	Normal	250.5 (8.5)	4.86 (0.14)	1.85 (0.15)	2.13 (0.20)	87.31 (7.25)	136.0 (13.4)
	Rain	228.6 (5.5)	4.99 (0.17)	2.16 (0.22)	2.49 (0.12)	86.71 (6.07)	106.8 (12.5
	Fog	242.2 (9.9)	4.96 (0.08)	2.10 (0.17)	2.40 (0.10)	87.83 (5.24)	115.6 (11.3)
Medium Vegetation	Normal	280.2 (25.3)	5.10 (0.20)	1.83 (0.15)	2.03 (0.15)	90.16 (4.28)	154.8 (23.2)
	Rain	250.6 (6.6)	5.18 (0.13)	2.31 (0.27)	2.52 (0.15)	90.99 (2.23)	110.4 (16.6)
	Fog	255.3 (13.6)	5.15 (0.15)	2.24 (0.24)	2.53 (0.12)	88.08 (6.19)	115.2 (12.1)
Light Vegetation	Normal	409.4 (23.8)	5.01 (0.19)	2.13 (0.15)	2.33 (0.13)	91.35 (3.77)	193.1 (21.5)
	Rain	386.5 (7.32)	5.15 (0.13)	2.51 (0.19)	2.70 (0.19)	92.81 (2.57)	155.0 (13.4)
	Fog	396.2 (17.3)	5.08 (0.12)	2.35 (0.12)	2.55 (0.11)	92.13 (2.14)	168.6 (7.1)

One-way analysis of variance (ANOVA) tests were performed on all dependent measures to determine if there was a difference in means between Weather Condition on each ODD. Post hoc Tukey's HSD tests were performed for all significant ANOVAs to determine if there is a significant difference between the means of the Weather Conditions. All statistical analyses used a significance level of $\alpha=0.05$.

On the ODD Heavy, there was a significant difference for distance traveled, $F(2,27)=18.17$, $p<0.00$, with post hoc comparisons showing Rain significantly lower than both Fog and Normal. There was a significant difference for average speed, $F(2,27)=8.35$, $p<0.00$, with post hoc comparisons showing Rain and Fog significantly higher than Normal. A significant difference was found for average moving speed, $F(2,27)=8.35$, $p<0.00$, with post hoc comparisons showing Rain and Fog significantly higher than Normal. Finally, a significant difference was found for total duration, $F(2,27)=16.44$, $p<0.00$, with post hoc comparisons showing Rain and Fog significantly quicker than Normal. No significance difference existed for Max Speed ($p=0.09$) or Percent Moving ($p=0.92$).

For the ODD Medium, there was a significant difference for distance traveled, $F(2,27)=8.74$, $p<0.00$, with post hoc comparisons showing Rain and Fog significantly lower than Normal. There was a significant difference for average speed, $F(2,27)=13.07$, $p<0.00$, with post hoc comparisons showing Rain and Fog significantly higher than Normal. A significant difference was found for average moving speed, $F(2,27)=43.13$, $p<0.00$, with post hoc comparisons showing Rain and Fog significantly higher than Normal. Finally, a significant difference was found for total duration, $F(2,27)=18.57$, $p<0.00$, with post hoc comparisons showing Rain and Fog significantly quicker than Normal. No significance difference existed for Max Speed ($p=0.50$) or Percent Moving ($p=0.54$).

Lastly, on the ODD Light, there was a significant difference for distance traveled, $F(2,27)=4.34$, $p=0.02$, with post hoc comparisons showing Rain significantly lower than Normal but no difference between Rain and Fog or Fog and Normal. There was a significant difference for average speed, $F(2,27)=14.59$, $p<0.00$, with post hoc comparisons showing Rain and Fog significantly higher than Normal. A significant difference was found for average moving speed, $F(2,27)=15.55$, $p<0.00$, with post hoc comparisons showing Rain and Fog significantly higher than Normal. Again, a significant difference was found for total duration, $F(2,27)=16.14$, $p<0.00$, with post hoc comparisons showing Rain and Fog significantly quicker than Normal. No significance difference existed for Max Speed ($p=0.08$) or Percent Moving ($p=0.92$).

4.3 Discussion

As predicted, the AGV performed best on the ODD Light achieving the highest average speed, average moving speed, and present moving scores. These results were expected as the ODD Light presents fewer obstacles which should result in easier path planning and less stops along the route. Unlike what was hypothesized, the AGV performed best under the Rain weather condition followed by Fog for almost every dependent measure on every terrain type. Performance under Normal weather conditions was worst on each terrain type. Figure 5a displays the path traveled on ODD Medium under each weather condition. As shown, the AGV under Rain and Fog conditions follows a fairly direct route from start to finish whereas the AGV under Normal condition zigzags, weaving around obstacles throughout the route. These unexpected results and autonomous behavior were attributed to a combination of the autonomy software path planning algorithm and the simulated weather effects placed on the virtual sensor. The researchers believed the narrowing of LiDAR perception to simulate weather effects increased performance by limiting obstacle detection, which has a significant influence in path planning decisions. Attenuating the sensor to simulate weather effects reduces the field of view leading to less obstacles detected during each scan,

ISBN: 97819384962-4-0

as seen on the cost map plots in Figures 5b Normal Cost Map and 5c Rain Cost Map, where red, white, and lighter shades of green represent impassible obstacles. The cost maps provide a representation of the obstacles and terrain features within the virtual environment by assigning a value to each pixel representing a point in the horizontal plane around the sensor. The autonomy stack then uses the cost map to plan a path that results in the lowest friction, or lowest pixel values, to the next waypoint. The researchers inspected the cost maps for each run and confirmed that the sensor attenuation under Rain and Fog conditions resulted in fewer fatal obstacles detected compared to Normal weather condition. The cost map in Figure 5c displays the reduced field of view and decreased number of obstacles detected under Rain condition compared to the cost map for Normal, Figure 5b. The reduced sensor visibility and fewer detected obstacles results in the autonomy stack planning a more direct path. The sensor scanning, cost map production, and path planning are continuously updated throughout the duration of the route allowing the AGV to successfully complete a more direct route even with a reduced field of view, or attenuated sensor. The research team shared the results of this study with ERDC in order to improve the path planning algorithms within the autonomy software and verify the effects of weather on the simulated virtual sensor. The researchers recommend expanding the virtual AGV testing to longer routes and different terrains to identify if the results are similar. In addition, the researchers plan to test an updated autonomy stack to determine if a new path planning algorithm alters the AGV performance.

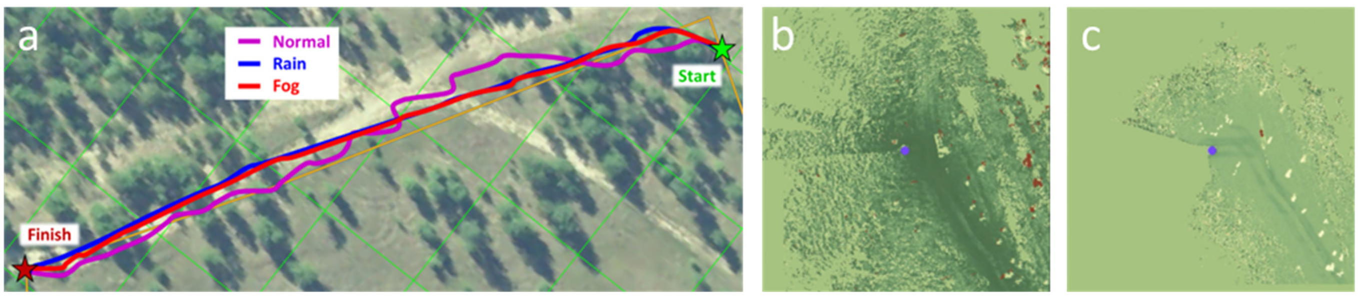

Figure 5. a. ODD Medium with plotted route for each Weather Condition, b. Normal Cost Map, c. Rain Cost Map

5. References

Adams, P., Bakre, A., Kramer, T., Riekena, J., & Thompson, B. (2023). "Evaluating Virtual Simulation for Autonomous Ground Vehicles." Proceedings of the Annual General Donald R. Keith Memorial Conference, West Point, NY, May 4. ISBN: 97819384964-4-8, 282-287. https://www.ieworldconference.org/content/WP2023/Papers/GDRKMCC23_50.pdf

Blinde, L. (2023). Army posts TSOA DiDEX 23 C-sUAS assessment RFI. *Intelligence Community News*. Retreived March 21, 2024, at https://intelligencecommunitynews.com/army-posts-tsoa-didex-23-c-suas-assessment-rfi/

Brooke, J. (1986). SUS -A quick and dirty usability scale Usability and context. Retrieved from https://digital.ahrq.gov/sites/default/files/docs/survey/systemusabilityscale%2528sus%2529_comp%255B1%255D.pdf

Dominguez, R., Onieva, E., Alonso, J., Villagra, J., & Gonzalez, C. (2011). LIDAR based Perception Soluation for Autonomous Vehicles. *11th International Conference on Intelligent Systems Design and Applications* (p. 6). Cordoba: IEEE.

Heckmann, L., Magnuson, S., & Park, A. (2023, December 12). Army Lays Out Plans for Robotic Combat Vehicles. *National Defense*. Retrieved March 20, 2024, from https://www.nationaldefensemagazine.org/articles/2023/12/12/army-lays-out-plans-for-robotic-combat-vehicles.

Jones, R., Priddy, J., Horner, D., Peters, J., Howington, S., Ballard Jr., J., Gates, B., & Cummins, C. (2008). Virtual Autonomous Navigation Environment (VANE). *Earth & Space2008: Engineering, Science, Construction, and Operations in Challenging Environments*. https://doi.org/10.1061/40988(323)114

Sun, C., Deng, Z., Chu, W., Li, S., & Cao, D. (2022) Acclimatizing the Operational Design Domain for Autonomous Driving Systems, *IEEE Intelligent Transportation Systems Magazine*, vol. 14, no. 2, pp. 10-24, March-April, doi: 10.1109/MITS.2021.3070651.

Turner, P. (2005). Unification of digital evidence from disparate sources (Digital Evidence Bags). *ScienceDirect*, vol. 2, issue 3, pp. 223-228, September, doi:10.1016/j.diin.2005.07.001.