Physical Attacks on Robot Navigation Systems

Meng Wang, Yohei Hayamizu, Matthew Tang, Kevin Gopalan, Shiqi Zhang, Ping Yang
Binghamton University

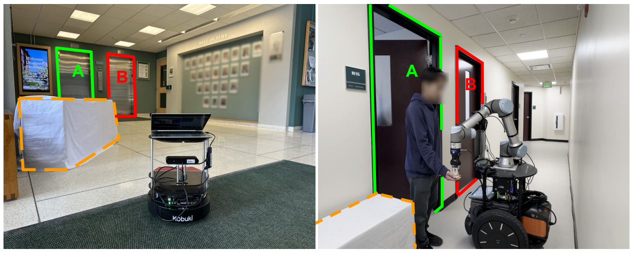
***Abstract*—Mobile robots are becoming an integral part of everyday life. These systems typically rely on generating maps of the environment and using them for navigation. While significant progress has been made in improving the localization and navigation of mobile robots, their vulnerability to adversarial environment changes remains largely unexplored. This paper investigates the adversarial robustness of robot navigation systems and introduces attacks designed to manipulate the navigation environment with minimal modifications. Our proposed attack leverages vision-language models and pre-existing maps to identify objects whose repositioning could cause navigation errors. We also propose a defense mechanism to monitor the confidence of self-localization to detect changes in the environment and bypass attacked areas. Evaluations show that our attacks reduce the navigation success rate from 100% to 8.0% in simulation and from 100% to 40.0% in the real world, while our defense mechanism increases the navigation success rate to 75.3% in simulation and 86.7% in the real world.**

I. Introduction

Mobile robots are increasingly deployed in everyday settings and rely on maps for navigation, where failures can have serious consequences. While mobile robots are significantly more affordable than self-driving cars, their limited sensors make them more vulnerable to adversarial attacks. Therefore, studying the robustness of mobile robots' navigation systems is essential to ensuring their safe and reliable operation in diverse environments.

Despite advances in localization and navigation of mobile robots [26, 27, 41, 33], their vulnerability to adversarial manipulation remains largely unexplored. Adversarial attacks have been widely studied in fields such as computer vision [30, 12, 38, 43, 42], e.g., introducing imperceptible noise to an image. In the domain of mobile robotic systems, adversarial attacks primarily target sensors and decision-making mechanisms crucial for navigation and localization. Previous work has manipulated sensor inputs to mislead robot perception and cause navigation failures [13, 3, 6, 24, 5]. By comparison, we are interested in stealthy attacks through physically manipulating the navigation environment, such as slightly moving a trash can or a desk.

In this paper, we investigate the adversarial robustness of the default navigation systems of Robot Operating System (ROS) [21], as demonstrated on two mobile robot platforms. We propose Adversarial Attacks on Robot Navigation Systems (Aarons), a framework that attacks the robot's navigation system by identifying objects in its environment whose repositioning (e.g., moving or rotating) could induce navigation errors. Aarons exploits the robot's reliance on aligning sensory data with a pre-built map for localization. By subtly altering the environment on the path, it is possible to create a mismatch between the robot's perception and its map, leading to localization errors and potential navigation failure.



a. Administration Building lobby. b. Engineering Building corridor.

Fig. 1. Illustrative example environments: The robot, originally heading to location A (marked green), is misdirected to location B (red) after Aarons subtly moves the desk (marked in orange).

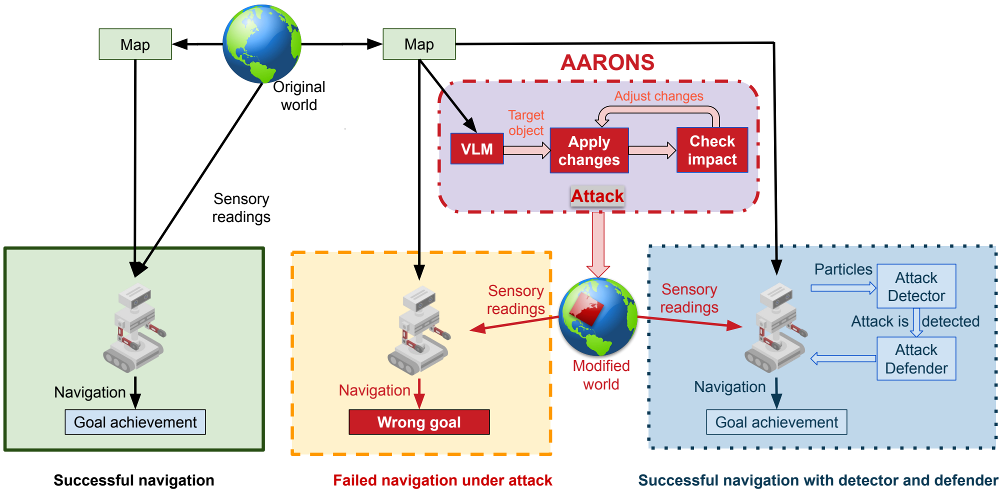
However, this approach is inefficient as the number of object-operation combinations grows. To improve efficiency, we leverage off-the-shelf vision-language models (VLMs) [1, 25, 32] to automatically identify objects whose repositioning could lead to navigation errors. Results show that Aarons can accurately pinpoint objects whose repositioning likely disrupts the robot's localization. Aarons demonstrates that attackers can quickly determine effective attack strategies using VLMs. Figure 1 gives two example environments.

We evaluated Aarons in simulation (supermarket and lobby) and in a matching real-world lobby environment. Our experimental results show that Aarons significantly reduces the robot's navigation success rate from 100% to 8.0% in simulation and from 100% to 40.0% in the real world. We have also proposed a defense mechanism that enables robots to monitor the confidence of self-localization to detect environmental changes and bypass attacked areas. Our defense strategy increases the navigation success rate to 72.7% in simulation and 86.7% in the real world.

II. Related Work

Cyberattacks on mobile robots: Most attacks on mobile robots are cyber methods targeting software vulnerabilities or sensor-level deception [35, 8]. For instance, falsified sensor inputs [6], manipulated landing markers [15], or GPS spoofing [14] can mislead UAVs. Vulnerabilities in visual servoing systems [13] also may mislead a robot to an unintended location. In autonomous driving, adversarial LiDAR perturbations and optical attacks can cause object detection and navigation failures [5, 24]. Denial of Service (DoS) attacks have been shown to induce physical effects in rescue robots [29]. However, these works exploit cyber vulnerabilities rather than physical environmental changes to disrupt navigation.

Physical attacks on mobile robots: Physical adversarial attacks have gained attention for their real-world applicabil-



Successful navigation

Failed navigation under attack

Successful navigation with detector and defender

Fig. 2. Overview of an attack causing navigation errors in a mobile robot by altering its environment. The green box (left) shows normal navigation where the robot reaches its goal. The red dashed box (middle) depicts the attack scenario with environmental changes disrupting navigation. The blue dotted box (right) presents the defense mechanism that helps the robot recover and complete its task.

ity [30], which target environmental features such as textures, objects, or lighting, to mislead perception or decision systems. While prior work has explored attacks on autonomous driving models using physical modifications [4], research on physical attacks against mobile robots remains limited. Recent studies have examined adversarial perturbations targeting motion planning algorithms [2, 31], whereas our work focuses on map-based navigation systems in indoor robots. A closely related study involves dynamic trajectory obstruction by an attacking robot [17], while our method, AARONS, uses static, one-time changes to highlight inherent vulnerabilities in robot navigation.

Defense mechanisms for mobile robots: Existing defenses for mobile robots primarily address cyber threats such as actuator and sensor anomalies [11, 16], jamming [34], and mitigating system vulnerabilities via encryption [39] or AI/blockchain solutions [23]. Safety modules have been introduced to constrain LLM-driven agents [36, 19, 28], and multi-robot coordination has been studied under adversarial disruptions [18]. However, these methods do not address physical attacks via environmental modifications. Improvements in LiDAR-based localization [37, 7] and uncertainty-aware techniques [20] mainly target dynamic but non-adversarial changes.

III. AARONS: Attacks on Robot Navigation

The target of AARONS is autonomous mobile robots that navigate using pre-built maps and real-time sensor data. These robots rely on path planning and localization techniques to determine their position and orientation. The objective of the attacker is to induce navigation failures by subtly altering the environment such as by moving or rotating objects near the robot's path. We assume that the attacker has physical access to the environment, allowing them to move or rotate semi-dynamic objects [40], but has no access to the robot itself or its internal hardware or software. The attacker is also assumed to have access to an occupancy-grid map of the environment. While AARONS was evaluated using the default navigation package of ROS, it does not assume specific navigation algorithms or systems.

The robot's navigation system is vulnerable due to its reliance on aligning sensory data with a pre-built map. Even minor environmental changes, such as slightly moving or rotating nearby objects, can disrupt this alignment, causing localization errors and trajectory deviations. Our adversarial attacks exploit this vulnerability by subtly moving or rotating objects along the navigation path.

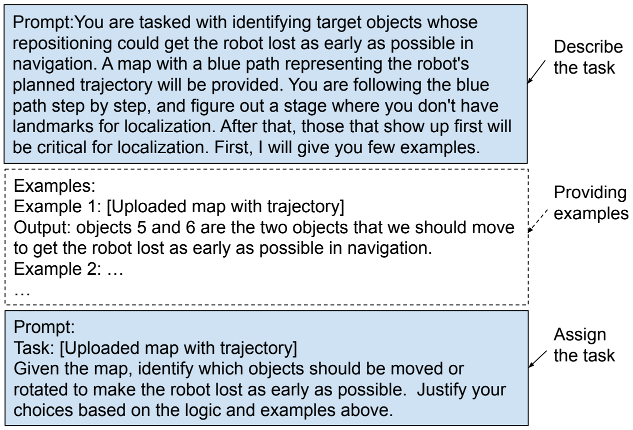
A. *Selection of Target Objects*

We initially developed a brute-force program that perturbs objects along the robot's path to identify scenarios leading to navigation failure. However, this is computationally expensive due to the exponential growth of object-operation combinations. To address this, we leverage VLMs to automatically identify objects whose repositioning may trigger navigation errors, with minimal human intervention. We explore three prompting strategies: zero-shot, chain-of-thought (CoT), and few-shot. In few-shot, the VLM is given K examples of successful attacks, each consisting of a map, a predefined navigation path, and disruptive objects. These examples help the model learn to associate the environmental layout and navigation path with target objects. The structure of the few-shot prompt is shown in Figure 3.

To ensure robustness and consistency in the selection process, we run the VLM $n = 10$ times for each scenario and use majority voting to select the most frequently recommended objects as final targets.

B. *Attack Execution*

To create stealthy attacks, we make the modifications small: objects are rotated by no more than 30 degrees, and objects are moved within a maximum range of 1 meter. AARONS applies all possible object-operation combinations, including the 'move' and 'rotate' operations, to the target objects during the simulation. Based on the simulation results, it will then determine effective attack strategies for real-world environments.



Prompt:You are tasked with identifying target objects whose repositioning could get the robot lost as early as possible in navigation. A map with a blue path representing the robot's planned trajectory will be provided. You are following the blue path step by step, and figure out a stage where you don't have landmarks for localization. After that, those that show up first will be critical for localization. First, I will give you few examples.

Describe the task

Examples:
Example 1: [Uploaded map with trajectory]
Output: objects 5 and 6 are the two objects that we should move to get the robot lost as early as possible in navigation.
Example 2: …
…

Providing examples

Prompt:
Task: [Uploaded map with trajectory]
Given the map, identify which objects should be moved or rotated to make the robot lost as early as possible. Justify your choices based on the logic and examples above.

Assign the task

Fig. 3. Few-shot prompt structure.

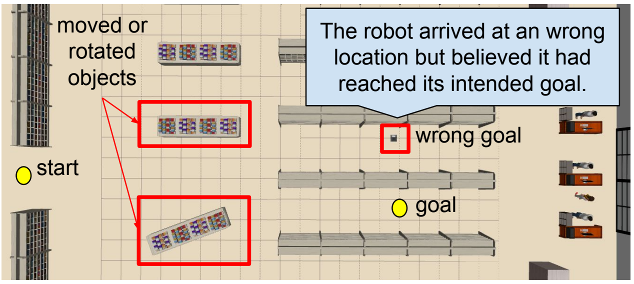

moved or rotated objects
The robot arrived at an wrong location but believed it had reached its intended goal.
wrong goal
start
goal

Fig. 4. A successful attack in a simulated supermarket environment.

Figure 4 illustrates a successful adversarial attack in a simulated supermarket. In this attack, two shelves along the robot's path were slightly moved, and one of them was also rotated. As a result, the robot deviated from its intended trajectory and arrived at an incorrect location, while mistakenly perceiving that it had reached the target destination. Such errors can have life-threatening consequences in real-world scenarios. For instance, a robot guide could misdirect a blind person toward a hazardous drop-off, or a healthcare robot might deliver medication to a patient in a wrong room in hospitals.

C. Detection and Defense

The blue dotted box in Figure 2 shows our detection and defense mechanisms. While AARONS does not assume specific localization or navigation algorithms, our defense approach assumes the robot using particle filters for localization.

Algorithm 1 provides the full pseudocode. The detection module monitors particle distributions at 20Hz (Lines 2-3) and uses DBSCAN [9] to quantify dispersion (Line 4). A concentrated particle cloud indicates confident localization, whereas significant dispersion, e.g., the formation of multiple clusters beyond a threshold τ, signals potential attack and triggers an alert (Lines 5-6).

Upon detection, the robot deviates from its original path to bypass the attacked area. The system selects the closest recovery position from a pool of recovery positions (Lines 13-14), and temporarily navigates there before regenerating a path to the final goal (Lines 15-18). This two-step strategy allows the robot to bypass the attacked area without a full environmental reevaluation.

Algorithm 1 Detection and Defense for AARONS

```
Require: Real-time particle states $P$, Navigation map $M$, Threshold
    $\tau$, Initial destination $D_{initial}$, Recovery position pool $R$
Ensure: Detection result $D$, Defense path $Path_{new}$
 1: Detection Phase:
 2: while Robot is navigating do
 3:     Monitor particle distribution $P$
 4:     Cluster $P$ using DBSCAN
 5:     if Number of clusters $> \tau$ then
 6:         $D \leftarrow$ Attack Detected
 7:         Break
 8:     else
 9:         $D \leftarrow$ No Attack
10:     end if
11: end while
12: Defense Phase:
13: if $D ==$ Attack Detected then
14:     Select nearest recovery position $R_{nearest}$ from $R$
15:     Re-plan path: $Path_{new} \leftarrow$ Navigate to $R_{nearest}$
16:     Navigate to $R_{nearest}$ and stop
17:     Re-plan path: $Path_{final} \leftarrow$ Navigate to $D_{initial}$
18:     Navigate to $D_{initial}$
19: end if
```

IV. EXPERIMENTAL RESULTS

We conducted experiments in both simulation and real-world to evaluate our attack strategies, detection, and defense methods, using ROS. Simulation was performed with Gazebo and RViz for environment control and visualization. In real-world tests, sensing, decision-making, and control were handled through ROS, with GMapping and AMCL used for mapping and localization. We use the default ROS navigation packages [1], which combine global path planning (e.g., A* and Dijkstra) with local planners (e.g., DWA [10] and Elastic Bands [22]). While more advanced systems exist, our goal is to show that even the most widely used navigation systems remain vulnerable.

a) Real World Experiments: We conducted real-world experiments in the lobby of an Administration Building (Figure 5(c)), using a TurtleBot 2 equipped with 2D LiDAR and a laptop running ROS (Figure 1(a)). The robot's task was to travel from the entrance to the left elevator (A in the figure), with and without attacks. Each setting was repeated 15 times using default ROS localization and navigation.

Without attacks, the robot succeeded in all 15 runs (**100%**). With a small perturbation (moving and rotating a table), the success rate dropped to **40%** ($6/15$). A t-test confirmed the significance of this drop ($t = 4.58$, $p < 0.0001$, mean diff = 0.60).

With our detection and defense enabled, the robot succeeded in $15/15$ trials under normal conditions (100%), including one false positive where the robot still reached the goal. Under attack, the success rate improved to **86.7%** ($13/15$). One attack was undetected, resulting in failure. A t-test showed a

[1] http://wiki.ros.org/navigation

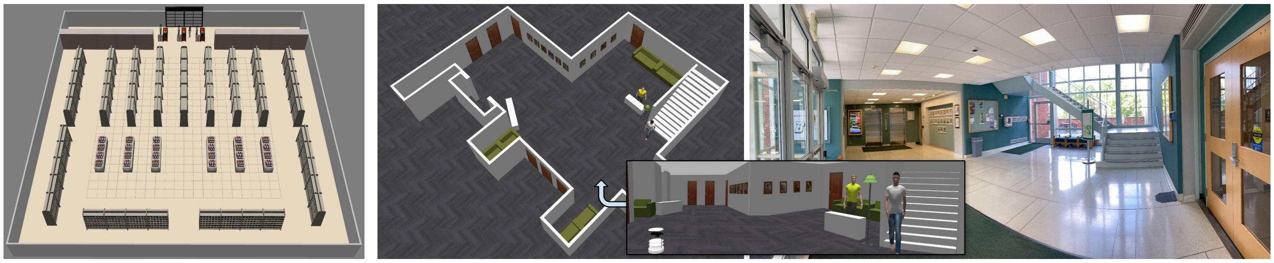

a. Supermarket
b. Administration building lobby
c. Administration building lobby (Real world)

Fig. 5. Experiment environments.

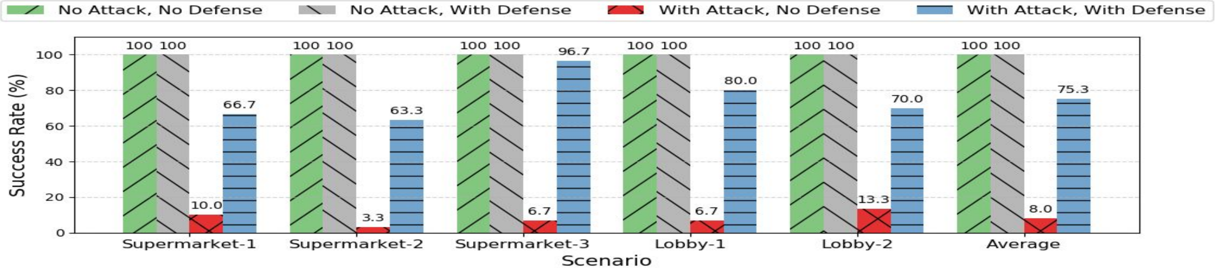


Fig. 6. Navigation success rate with and without attacks in simulation.

significant improvement with defense ($t = 2.93$, $p = 0.0067$, mean diff = -0.47).

To assess generalizability, we applied the attack to a Segway-based robot in a building corridor. As shown in Figure 1(b), a slight change in table position caused the robot to misdeliver the medicine to the wrong room.

b) Simulated Environments: We evaluated our methods in two Gazebo-simulated environments (shown in Figure 5(a and b)): a structured supermarket with narrow aisles and an open-layout lobby based on a real-world map. These environments represent different navigation challenges.

Five navigation tasks were performed (three in the supermarket, two in the lobby), each under four experimental conditions with 30 trials, totaling 600 trials. Figure 6 shows that without attacks, the robot completed all tasks successfully. Under attack, success rates dropped sharply to 3.3%–13.3% (avg. 8.0%). Paired t-tests confirmed statistically significant performance degradation ($p < 0.0001$).

Our detection mechanism achieved high accuracy (Table I), with an average true positive rate of 98.0% and true negative rate of 91.3%. False positives did not hinder task completion. Upon attack detection, our defense mechanism improved navigation success rates to 64.5%–93.9% (Table I). Overall, as shown in Figure 6, detection and defense raised average success from 8.0% to 75.3%, demonstrating strong mitigation performance.

c) Object selection with different VLMs: We evaluated five VLMs in their accuracy of object selection for environment modification. The VLMs include Gemini 2.5 Pro, Gemini 2.0 Flash, GPT-4o-mini-high, GPT-4o, and Grok 3. The prompting strategies include zero-shot, CoT, and few-shot. Each data point is an average of fifteen trials, including five trials for each of the three navigation tasks. Table II shows the results. The accuracy is 100% and 93% for objects identified by GPT o4-mini-high and Gemini 2.5, respectively, with all three prompt strategies. For GPT-4o, the accuracy is 87% with few-shot prompt, but lower with zero-shot and CoT. Gemini 2.0 Flash and Grok 3 have low performance.

Scenario	Detection No Attack (TN)	Detection With Attack (TP)	Defense
SM-1	86.7%	100%	70.6%
SM-2	96.7%	100%	64.5%
SM-3	90.0%	100%	93.9%
Lobby-1	83.3%	96.7%	85.3%
Lobby-2	100%	93.3%	75.0%
Average	91.3%	98.0%	77.9%

TABLE I
DETECTION AND DEFENSE SUCCESS RATE.

VLM	Zero-shot	CoT	Few-shot
GPT o4-mini-high	100%	100%	100%
Gemini 2.5 Pro	93%	93%	93%
GPT-4o	60%	73%	87%
Gemini 2.0 Flash	40%	20%	7%
Grok 3	0%	27%	0%

TABLE II
COMPARITHION OF VLMS

V. CONCLUSION

In this paper, we studied the adversarial robustness of mobile robot navigation systems and demonstrated how subtle environmental modifications can result in navigation failures. We have also developed detection and defense mechanisms to mitigate attacks. AARONS is the first VLM-based physical attack on mobile robot navigation systems through making changes to the robot's working environments. Our experiments, conducted in both simulated and real-world environments, demonstrate the effectiveness of our attacks as well as our detection and defense mechanisms.

References

[1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[2] Naif Wasel Alharthi and Martim Brandão. Physical and digital adversarial attacks on grasp quality networks. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1907–1902. IEEE, 2024.

[3] Gianluca Bianchin, Yin-Chen Liu, and Fabio Pasqualetti. Secure navigation of robots in adversarial environments. *IEEE Control Systems Letters*, 4(1):1–6, 2019.

[4] Adith Boloor, Xin He, Christopher Gill, Yevgeniy Vorobeychik, and Xuan Zhang. Simple physical adversarial examples against end-to-end autonomous driving models. In *2019 IEEE International Conference on Embedded Software and Systems (ICESS)*, pages 1–7. IEEE, 2019.

[5] Yulong Cao, Chaowei Xiao, Benjamin Cyr, Yimeng Zhou, Won Park, Sara Rampazzi, Qi Alfred Chen, Kevin Fu, and Z Morley Mao. Adversarial sensor attack on lidar-based perception in autonomous driving. In *Proceedings of the 2019 ACM SIGSAC conference on computer and communications security*, pages 2267–2281, 2019.

[6] Drew Davidson, Hao Wu, Rob Jellinek, Vikas Singh, and Thomas Ristenpart. Controlling {UAVs} with sensor input spoofing attacks. In *10th USENIX workshop on offensive technologies (WOOT 16)*, 2016.

[7] Salvador Dominguez, Gaëtan Garcia, Vincent Frémont, and Arnaud Hamon. An experimental evaluation of robustness and precision for long-term lidar-based localization in highly changing environments. *arXiv preprint arXiv:2003.07726*, 2020.

[8] Wojciech Dudek and Wojciech Szynkiewicz. Cyber-security for mobile service robots–challenges for cyber-physical system safety. *Journal of Telecommunications and Information Technology*, (2):29–36, 2019.

[9] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231, 1996.

[10] Dieter Fox, Wolfram Burgard, and Sebastian Thrun. The dynamic window approach to collision avoidance. *IEEE Robotics & Automation Magazine*, 4(1):23–33, 1997.

[11] Pinyao Guo, Hunmin Kim, Nurali Virani, Jun Xu, Minghui Zhu, and Peng Liu. Exploiting physical dynamics to detect actuator and sensor attacks in mobile robots. *arXiv preprint arXiv:1708.01834*, 2017.

[12] Ke He, Dan Dongseong Kim, and Muhammad Rizwan Asghar. Adversarial machine learning for network intrusion detection systems: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 25(1):538–566, 2023.

[13] Aleksandar Jokic, Amir Khazraei, Milica Petrovic, Zivana Jakovljevic, and Miroslav Pajic. Cyber-attacks on wheeled mobile robotic systems with visual servoing control. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6342–6348. IEEE, 2023.

[14] Andrew J Kerns, Daniel P Shepard, Jahshan A Bhatti, and Todd E Humphreys. Unmanned aircraft capture and control via gps spoofing. *Journal of field robotics*, 31(4): 617–636, 2014.

[15] Amir Khazraei, Haocheng Meng, and Miroslav Pajic. Stealthy perception-based attacks on unmanned aerial vehicles. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3346–3352. IEEE, 2023.

[16] Kyo Kim, Siddhartha Nalluri, Ashish Kashinath, Yu Wang, Sibin Mohan, Miroslav Pajic, and Bo Li. Security analysis against spoofing attacks for distributed uavs. *Decentralized IoT Systems and Security*, 2020.

[17] Yushan Li, Jianping He, Cailian Chen, and Xinping Guan. Intelligent physical attack against mobile robots with obstacle-avoidance. *IEEE Transactions on Robotics*, 39(1):253–272, 2022.

[18] Amos Matsiko. Overcoming adversaries in multirobot navigation. *Science Robotics*, 9(86):eado2404, 2024.

[19] Minheng Ni, Lei Zhang, Zihan Chen, and Wangmeng Zuo. Don't let your robot be harmful: Responsible robotic manipulation. *arXiv preprint arXiv:2411.18289*, 2024.

[20] Geonhyeok Park and Woojin Chung. Uncertainty-aware lidar-based localization for outdoor mobile robots. *Journal of Field Robotics*, 41(8):2790–2804, 2024.

[21] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, Andrew Y Ng, et al. Ros: an open-source robot operating system. In *ICRA workshop on open source software*, volume 3, page 5. Kobe, 2009.

[22] Sean Quinlan and Oussama Khatib. Elastic bands: Connecting path planning and control. In *[1993] Proceedings IEEE International Conference on Robotics and Automation*, pages 802–807. IEEE, 1993.

[23] Anand Singh Rajawat, Romil Rawat, Rabindra Nath Shaw, and Ankush Ghosh. Cyber physical system fraud analysis by mobile robot. *Machine learning for robotics applications*, pages 47–61, 2021.

[24] Hocheol Shin, Dohyun Kim, Yujin Kwon, and Yongdae Kim. Illusion and dazzle: Adversarial optical channel exploits against lidars for automotive applications. In *International Conference on Cryptographic Hardware and Embedded Systems*, pages 445–467. Springer, 2017.

[25] Gemma Team, Thomas Mesnard, Cassidy Hardin, Robert Dadashi, Surya Bhupatiraju, Shreya Pathak, Laurent Sifre, Morgane Rivière, Mihir Sanjay Kale, Juliette Love, et al. Gemma: Open models based on gemini research and technology. *arXiv preprint arXiv:2403.08295*, 2024.

[26] Sebastian Thrun, Maren Bennewitz, Wolfram Burgard, Armin B Cremers, Frank Dellaert, Dieter Fox, Dirk

Hähnel, Charles Rosenberg, Nicholas Roy, Jamieson Schulte, et al. Minerva: A second-generation museum tour-guide robot. In *Proceedings 1999 IEEE International Conference on Robotics and Automation*, volume 3. IEEE, 1999.

[27] Sebastian Thrun, Dieter Fox, Wolfram Burgard, and Frank Dellaert. Robust monte carlo localization for mobile robots. *Artificial intelligence*, 128(1-2):99–141, 2001.

[28] Milos Vasic and Aude Billard. Safety issues in human-robot interactions. In *2013 ieee international conference on robotics and automation*, pages 197–204. IEEE, 2013.

[29] Tuan Vuong, Avgoustinos Filippoupolitis, George Loukas, and Diane Gan. Physical indicators of cyber attacks against a rescue robot. In *2014 IEEE International Conference on Pervasive Computing and Communication Workshops*, pages 338–343. IEEE, 2014.

[30] Hui Wei, Hao Tang, Xuemei Jia, Zhixiang Wang, Hanxun Yu, Zhubo Li, Shin'ichi Satoh, Luc Van Gool, and Zheng Wang. Physical adversarial attack meets computer vision: A decade survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[31] Wenxi Wu, Fabio Pierazzi, Yali Du, and Martim Brandão. Characterizing physical adversarial attacks on robot motion planners. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 14319–14325. IEEE, 2024.

[32] xAI. Grok ai, 2024. URL https://grok.com/. Accessed: 2025-04-29.

[33] Xuesu Xiao, Bo Liu, Garrett Warnell, and Peter Stone. Motion planning and control for mobile robot navigation using machine learning: a survey. *Autonomous Robots*, 46(5):569–597, 2022.

[34] Hao Xu, Jinhui Zhang, Zhongqi Sun, and Hongjiu Yang. Event-based wireless tracking control for a wheeled mobile robot against reactive jamming attacks. *IEEE Transactions on Control of Network Systems*, 10(4): 1925–1936, 2023.

[35] Jean-Paul A Yaacoub, Hassan N Noura, Ola Salman, and Ali Chehab. Robotics cyber security: Vulnerabilities, attacks, countermeasures, and recommendations. *International Journal of Information Security*, 21(1):115–158, 2022.

[36] Ziyi Yang, Shreyas S Raman, Ankit Shah, and Stefanie Tellex. Plug in the safety chip: Enforcing constraints for llm-driven robot agents. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 14435–14442. IEEE, 2024.

[37] Huan Yin, Xuecheng Xu, Sha Lu, Xieyuanli Chen, Rong Xiong, Shaojie Shen, Cyrill Stachniss, and Yue Wang. A survey on global lidar localization: Challenges, advances and open problems. *International Journal of Computer Vision*, pages 1–33, 2024.

[38] Wei Emma Zhang, Quan Z Sheng, Ahoud Alhazmi, and Chenliang Li. Adversarial attacks on deep-learning models in natural language processing: A survey. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 11 (3):1–41, 2020.

[39] Xu Zhang, Zhenyuan Yuan, Siyuan Xu, Yang Lu, and Minghui Zhu. Secure perception-driven control of mobile robots using chaotic encryption. *IEEE Transactions on Automatic Control*, 2023.

[40] Shifan Zhu, Xinyu Zhang, Shichun Guo, Jun Li, and Huaping Liu. Lifelong localization in semi-dynamic environment. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 14389–14395. IEEE, 2021.

[41] Yuke Zhu, Roozbeh Mottaghi, Eric Kolve, Joseph J Lim, Abhinav Gupta, Li Fei-Fei, and Ali Farhadi. Target-driven visual navigation in indoor scenes using deep reinforcement learning. In *2017 IEEE international conference on robotics and automation (ICRA)*, pages 3357–3364. IEEE, 2017.

[42] Daniel Zügner, Amir Akbarnejad, and Stephan Günnemann. Adversarial attacks on neural networks for graph data. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2847–2856, 2018.

[43] Daniel Zügner, Oliver Borchert, Amir Akbarnejad, and Stephan Günnemann. Adversarial attacks on graph neural networks: Perturbations and their patterns. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 14 (5):1–31, 2020.