# Robots for Safer Pedestrian Crossing on Two-Lane Roads

Andy Luis Flores Comeca*, Nina Masarykova†, Matej Halinkovic†
Marek Galinski†, Pavel Laskov‡, Alexey Vinel*§
* Karlsruhe Institute of Technology, Kaiserstraße 89, 76133 Karlsruhe, Germany
† Slovak University of Technology, Ilkovičova 2, 842 16 Bratislava, Slovakia
‡ University of Liechtenstein, Fürst-Franz-Josef-Strasse, 9490 Vaduz, Liechtenstein
§ Halmstad University, Kristian IV:s väg 3, 301 18 Halmstad, Sweden
Email: *alexey.vinel@kit.edu*

***Abstract*—Robots are becoming increasingly integrated into various industries, including healthcare, and even homes. However, their application and potential in traffic management remain relatively unexplored. This paper investigates the potential of the humanoid social robot ARI to enhance pedestrian safety at two-lane crossings. Positioned on a pedestrian refuge island, ARI utilizes advanced pedestrian and vehicle detection models running on an NVIDIA Jetson AGX Orin, enabling real-time traffic assessment. Thanks to this ability to evaluate the states and intentions of both vehicles and pedestrians, ARI can use gestures resembling human traffic police and voice messages to effectively communicate with both pedestrians and drivers, facilitating safer crossings. To ensure comprehensive traffic monitoring, ARI requires 360-degree environmental awareness and depth information through additional hardware as input to a robust decision-making system to handle diverse traffic scenarios. The robot follows a structured approach: after detecting pedestrians willing to cross makes sure that all vehicles stopped and signal pedestrians to cross the first lane. Using depth information and object tracking algorithm, ARI evaluates that pedestrians crossed the first lane, interacts with them at the pedestrian refuge island, and guides them safely across the second lane. This method improves pedestrian safety by providing clear, dynamic guidance at each stage of the crossing process. The robot-managed crossing assistance system was evaluated in real-world, controlled, traffic conditions.**

***Index Terms*—Social robots, intelligent transportation systems, vulnerable road users, pedestrian crossing, advanced driver assistance systems, road safety.**

## I. Introduction

Robotic technology is increasingly being utilized across various sectors, including industry, healthcare, and household applications. However, its integration into intelligent transportation systems remains relatively unexplored. The implementation of robotics in road traffic management presents significant potential for enhancing safety and improving the overall efficiency of urban transportation networks. Such robots may be able to obtain a comprehensive situational awareness thanks to a diverse set of onboard cameras and sensors, develop a safe and an efficient coordination strategy in real-time by means of artificial intelligence, and act as a reliable and thoughtful helper thanks to a variety of interaction modalities.

The mobility of a robot can further amplify these features, which can eventually bring about 'coordination skills' that surpass the capabilities of the most advanced traffic lights and the most experienced police officers. The robot's integration with existing traffic management systems allows for an effective operation as a part of a larger traffic coordination network. Finally, the humanoid outlook and the social interaction capabilities make the robot an attractive solution for improved safety of especially vulnerable user groups, for instance, children.

Although the deployment of a robot-policeman for intelligent intersection management or steering autonomous driverless vehicles may appear rather futuristic, our paper aims to make the first step and to explore the feasibility of deployment of social robots for a use-case that may have an immediate practical impact: pedestrian crossing assistance. We identify significant potential for social robots in this role, as they can simultaneously communicate with both vehicle drivers and pedestrians while presenting traffic signals in ways that are intuitive to humans, such as replicating gestures used by traffic officers or emulating light patterns from traffic signals.

The importance of such a mediator becomes even more pronounced during disruptions or failures in standard traffic regulation infrastructure, where the robot could assist or temporarily replace human officers. Moreover, the robot could dynamically manage traffic by adapting regulations in real time, such as temporarily closing a pedestrian crossing to enable an emergency vehicle to pass swiftly and safely. Furthermore, in complex traffic scenarios involving a mixture of automated and human participants, social robots could play a critical role in maintaining traffic safety and optimizing traffic flow.

The use of robotics in traffic management and pedestrian safety has been explored in various contexts, providing a foundation for this study. Gong et al. [1] introduced a life-size traffic police officer robot *IWI*. The robot is equipped with cameras, capable of streaming real-time video to multiple end devices and communicating with traffic participants via hand gestures. The use of gestures is remotely controlled by a human. A similar concept was later adapted by Zhao et al. [2] and Ghaffar [3]. Robotic police officers are intended to interact


Fig. 1: ARI robot traffic management sequence: (a) detects pedestrian; (b) monitors and stops oncoming vehicles; (c) signals pedestrian to cross; (d) uses rear camera to detect and (e) halt vehicles in the second lane to ensure pedestrian safety; and (f) lets vehicles pass once the pedestrain has completed the crossing.

mostly with drivers; however, multiple solutions with robots that interact with pedestrians were proposed. Kumaran et al. [4] introduced a small robotic object mounted on a vehicle dashboard that communicates the intentions of autonomous vehicles near pedestrian crossings. Two simple gestures are used to give pedestrians the signal to cross or not cross the road. A study by Kulhandjian [5] proposed a simple robot based on the Pioneer 3-AT robot with a traffic light-inspired LED matrix mounted on a pole. The pole is also equipped with a camera and LiDAR sensor to assist pedestrians crossing the road. The YOLO [6] machine learning algorithm is used for efficient object detection and classification of pedestrians, cyclists, and vehicles. Red and green LED lights are used to signal both vehicles and pedestrians to stop or go. The robot actively moves to the center of the road when pedestrians are detected, stopping the cars.

Deployment of autonomous robots in outdoor environments is particularly challenging due to the high levels of environmental variability and unpredictability. Factors such as changing lighting conditions, adverse weather, and temporary obstructions can significantly affect their performance. Robotic platforms such as PAL Robotics' ARI, Boston Dynamics' Spot, KUKA's industrial robots, and Clearpath Robotics' OTTO Motors excel in specific applications such as inspection, industrial automation, and indoor logistics. However, they lack the situational awareness, adaptability, and system integration capabilities necessary for dynamic outdoor traffic environments. These limitations present a unique opportunity for the development of robotic solutions that are specifically tailored for traffic management.

To address this gap, we use the ARI humanoid robot as a foundational platform to develop a system capable of meeting traffic management requirements. The primary limitations of ARI in this context are:

1) Situational Awareness: ARI, like many other robots, is designed for controlled indoor environments and lacks the sensor fusion and data processing capabilities required for real-time situational awareness in outdoor traffic scenarios. Improving these capabilities is critical.
2) Adaptability: ARI's current adaptive learning algorithms are insufficient for handling rapidly changing traffic conditions. Real-time learning and scenario-based adaptability need to be integrated.
3) Integration: ARI's limited ability to interact with existing traffic management systems restricts its functionality in broader traffic orchestration networks. Enhancing communication protocols and coordination mechanisms is essential for optimizing traffic flow and safety.

In this study, we focus on the first item: enhancing ARI's situational awareness to enable its effective deployment in dynamic traffic environments. Our work involves upgrading ARI's sensors and perception algorithms to enable it to assist pedestrians with road crossing and improve his capabilities for traffic management. These enhancements aim to ensure reliable robot performance in outdoor scenarios, laying the groundwork for further advancements in robotic traffic management.

Recently, we have already introduced ARI in a virtual reality environment to gather feedback from participants, and evaluated the feasibility of deploying it for managing traffic at pedestrian crossings [7]. In the current work we not only bring our experiments from the virtual reality to the real-world, but also consider more elaborated pedestrian crossing assistance logic.

The paper is organized as follows. Section II introduces the ARI robot hardware, the methodology is explained in Section III, and the experimental results in Section IV. Finally, concluding discussions and outlook comprise Section V.

## II. Hardware Description

For this work, we selected the ARI robot, a 165 cm tall humanoid platform designed to support natural and expressive human-robot interaction across a wide range of applications. ARI is equipped with a comprehensive suite of sensors, actuators, and computing resources, making it a suitable candidate for dynamic and interactive environments such as traffic management scenarios involving pedestrians and vehicles.

ARI's perception system includes a 2D YD-LIDAR TG15 for planar obstacle detection, and two RGB-D cameras located on its head and torso. These cameras offer a sensing range of 0.3–3 meters, a frame rate of 30 fps, and a field of view of 87° horizontally and 58° vertically, enabling reliable depth perception and object recognition. To support real-time data processing and high-performance computation, ARI integrates an Intel i9 CPU with 32 GB of RAM alongside a Jetson AGX Orin module.

In terms of actuation, ARI features a two-degree-of-freedom (DoF) head for expressive movements, a differential-drive mobile base for navigation (2 DoF), and two articulated arms, each with five DoF. These arms can perform a variety of gestures to convey intuitive signals and foster more engaging interactions with pedestrians and drivers. Communication is further supported through a 10.1-inch touchscreen, integrated stereo speakers, and microphone arrays for voice interaction.

To better support our traffic management use case, we extended ARI's perception capabilities by integrating an additional RGB-D camera at the rear of its head. This enhancement enables the robot to monitor two lane roads scenarios, to monitors vehicles approaching from both sides.

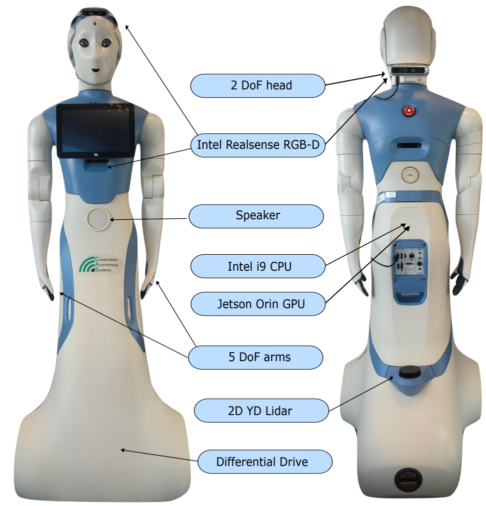

Fig. 2: ARI Robot hardware description

## III. Methodology

The solution is proposed to handle specific traffic scenario. In this work, ARI robot is deployed in a two-lane road environment (see Fig 1). This setup presents significant challenges since the robot must interact simultaneously with vehicles and pedestrians across two lanes. Even for a human traffic controller, managing two lanes requires constant vigilance, hands gestures and verbal commands to ensure safety.

The primary goal of the system is to ensure the safety of vulnerable pedestrians during peak hours, with a particular focus on children commuting to and from school. The designed sequence aims to handle one direction of the flow of the pedestrians efficiently while minimizing vehicle stoppage time.

To address this, ARI is equipped with a front head camera and an additional rear head camera, enabling it to monitor cars or pedestrians in both lanes. Pedestrian monitoring is further enhanced by enabling ARI to move its head, thus amplifying its field of view and allowing complete perception of the crossing area.

### A. Perception module

In order to control the behaviour of the robot and changes of its state, custom perception module is proposed. The module consists of two main blocks: pedestrian perception and vehicle perception. To detect the objects from RGB camera frames, the *medium* version of YOLOv12 [11], [12] is used, consisting of 20.2M parameters. The model pretrained on the COCO dataset [9], consisting of 80 classes, achieves mean average precision (mAP) of 52.5%. In the proposed system, we used it to perceive people, bicycles, cars, motorcycles, buses and trucks, where the model reaches mAP of 0.84, 0.65, 0.73, 0.79, 0.90, and 0.633 respectively. To achieve real-time inference (25 FPS) on Jetson Orin GPU, the model is compiled using TensorRT.

*1) Pedestrian intention prediction:* The crossing sequence begins with the detection of a pedestrian and the identification of his intention to cross. Used pedestrian intention prediction is based on the processing of separate camera frames. A pedestrian is detected by the YOLO model. The prediction from current frame is matched with the historical predictions using the Hungarian algorithm [10], obtaining a trajectory of the object. The pedestrian is considered as having intention to cross once stationary at the border of the pedestrian crossing for X seconds.

*2) Pedestrian distance evaluation:* Once the pedestrian intention is detected and the pedestrian enters the crossing, the RGB image is enriched with depth information from the depth camera. Distance from the sensor is evaluated for each detected pedestrian by taking the median value of the depth pixels from the corresponding bounding box. During our experiments, this was determined to be sufficiently accurate as distant background pixels do not contain any values and thus do not affect the calculation. This approach allows us to compute the distance of a pedestrian efficiently without introducing a significant level of noise, especially compared to the inherent noise of the depth sensor.

*3) Vehicle staticity evaluation:* To ensure safety of the crossing, a pedestrian is only allowed to cross once the robot makes sure that incoming vehicles stopped in front of the crossing or there are no incoming vehicles. Similar approach to III-A1 is used to determine staticity of the vehicle. The main focus is on the first approaching vehicle, as other vehicles behind can still be moving without posing a danger to pedestrians.
The vehicles that are parked might influence the evaluation of the safety and have to be excluded from consideration. Once the robot finishes the transition to appropriate pose, a reference frame is taken. Given a current RGB frame $I_t$ and a reference frame $I_{\text{ref}}$, a raw pixel-wise difference map is computed as:

$$D(x,y) = \sum_{c=1}^{3} \left| I_t(x,y,c) - I_{\text{ref}}(x,y,c) \right|, \tag{1}$$

where $(x,y)$ denotes the spatial coordinates of a pixel, and $c \in \{1,2,3\}$ indexes the RGB channels.

To standardize the difference values across frames and enable thresholding, the difference map $D(x,y)$ is normalised as:

$$D_{\text{norm}}(x,y) = \frac{D(x,y)}{\max(D)} \cdot 255. \tag{2}$$

A threshold is applied on the normalized difference map using a predefined value $\tau$, setting pixels with values greater than $\tau$ to 255 and the rest to 0, yielding a binary motion mask.

For each detection bounding box $B$, the motion change ratio is defined as the proportion of pixels within the box that exhibit significant change:

$$\text{ChangeRatio}(B) = \frac{1}{|B|} \sum_{(x,y)\in B} \mathbf{1}\left[D_{\text{norm}}(x,y) > \tau\right], \tag{3}$$

where $|B|$ is the number of pixels in the box, and $\mathbf{1}[\cdot]$ is the indicator function.

Only those bounding boxes for which the motion change ratio exceeds a fixed threshold (e.g., 25%) are retained.

## B. Overview of the crossing sequence

Fig. 3 describes the system behaviour, where the transition between states of the robot is controlled by the perception module. In each state, the robot uses certain gestures to communicate with vehicles or voice messages to communicate with pedestrians.

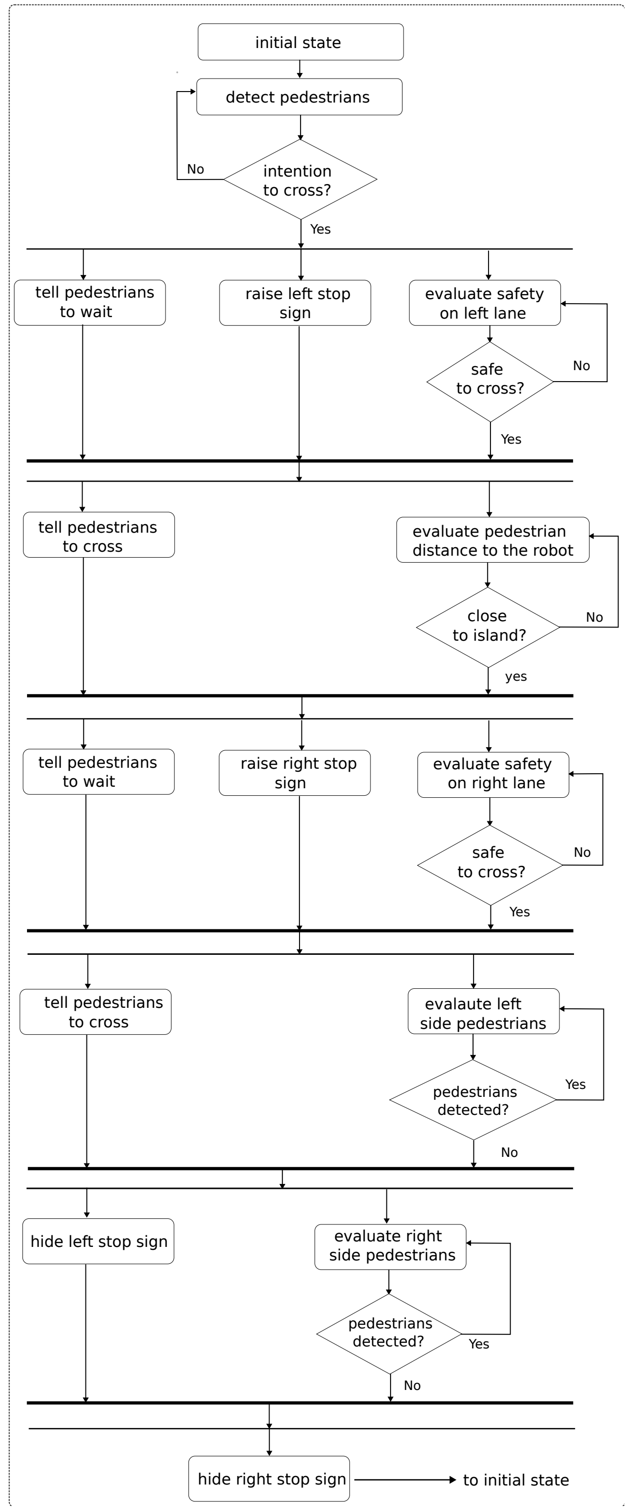


Fig. 3: Diagram of ARI's decision making during the crossing sequence

The robot begins in its *initial state*, facing the left crossing lane (Fig. 1a). RGB frames from the front camera are processed by *pedestrian intention prediction*. If a pedestrian's intention to cross is predicted, ARI blocks the left cars by raising its left arm and issues a verbal command *"WAIT"*, asking the pedestrians to wait. The robot then moves its head to face the approaching vehicles in the left lane to verify whether any vehicles are approaching and if they stop in front of the zebra using *vehicle staticity evaluation* (Fig. 1b). Only once the lane is clear or the vehicle has stopped, ARI issues a voice command *"YOU CAN CROSS"*, allowing the pedestrians to cross the lane.

The robot rotates its head back to pedestrians and monitors their position while they are crossing the lane by measuring the distance to the robot using *pedestrian distance evaluation* (Fig. 1c). Once pedestrians are close to the the center island, ARI instructs them via voice command *"WAIT"* to wait at the island before crossing the right lane. At this point, the head of the robot is rotated to face the vehicles in the right lane with the rear camera. The robot stops cars approaching from the right lane and ensures that no vehicle is approaching or that approaching vehicles have stopped using the *vehicle staticity evaluation* (Fig. 1d).

In the next phase (Fig. 1e), pedestrians cross the right lane. Once ARI verifies the right lane is safe, it instructs the pedestrians via voice message *"CLEAR THE ISLAND"* to leave the island. The robot rotates his head to monitor the whole zebra crossing with the front and rear camera. Using *pedestrian intention prediction*, robot monitors the crossing process to ensure no new pedestrians are arriving from the left. If no new pedestrians in predefined proximity are detected, ARI rotates its left arm, indicating that the vehicles in the left lane can continue by rotating the "STOP" sign in his hand away from vehicles and showing it to potentially approaching pedestrians on the left pavement.

Simultaneously, ARI verifies that all pedestrians have successfully crossed to the right side, utilizing depth information from the rear camera. Finally, the robot lowers its right arm to allow traffic in the right lane to resume and returns to the initial state (Fig. 1f), ready to detect the next pedestrian crossing intention.

## IV. Experimental Results

We evaluated our solution in a controlled environment where ARI is strategically located in the center of a crossing island, oriented parallel to both lanes, as shown in Fig. 4. This placement allows it to effectively manage interactions between pedestrians and vehicles on both sides of the crossing.

The aim of our experimental setup was to evaluate the technical viability of our proof of concept. As such, participants were informed about our goals and functioning of the system as to not endanger their safety in case of failure. We conducted the experiment with a series of different scenarios with the goal of determining the robustness of the proposed system under various conditions. In total, we executed 18 different experiments which can be grouped into one of 3 broader categories:

- **Default** - basic setup for assisting a single pedestrian with crossing the road. Vehicle drivers were instructed to always immediately obey the robot's commands.

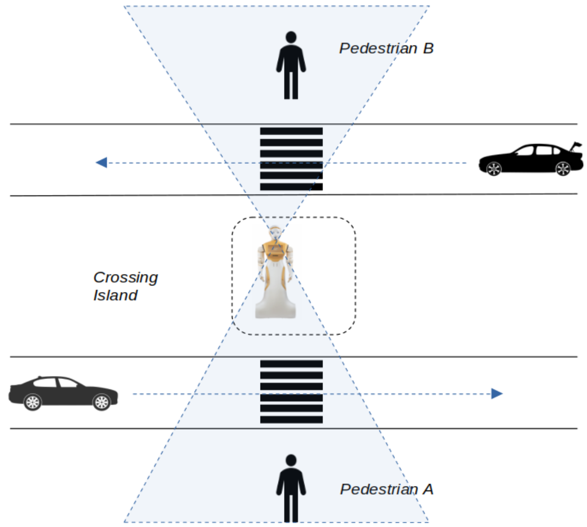


Fig. 4: ARI robot in the middle of the crossing island facing pedestrian and vehicle in both lanes

- **Pedestrian perception robustness validation** - multiple pedestrians were attempting to cross the road independently of each other. Vehicle drivers were instructed to always immediately obey the robot's commands.
- **Vehicle perception robustness validation** - varying number of vehicles. Drivers could ignore the robot's commands.

We measured the success rates of individual states of the solution as described in Section III-B as well as the total success rate of the system as whole. The results can be seen in Table I.

TABLE I: Success Rates by State

| State | Success Rate |
|---|---|
| Pedestrian intention detection | 94.44% |
| Pedestrian distance evaluation (close) | 88.89% |
| Vehicle staticity evaluation (front) | 100% |
| Vehicle staticity evaluation (back) | 94.44% |
| Pedestrian distance evaluation (no pedestrians) | 100% |
| Total Success | 83.33% |

Overall, the proposed system achieves a reasonably high total success rate of 83.33%, which was measured as a ratio of how many complete transitions of the system were executed without failing any of the intermediate steps. If one of the intermediate perception steps failed, the transition was marked as a failure for this measure even if all others executed perfectly.

The proposed system was robust when it comes to different behaviours of pedestrians and vehicles. In the case of pedestrians, ARI was able to correctly handle situations with multiple approaching pedestrians, checking if the crossing is empty before letting vehicles pass and not stopping vehicles when encountering pedestrians with no intention to cross. When it comes to vehicles, the proposed system was able to correctly determine the safety of pedestrians in almost every case. It was robust against situations where one or more drivers ignored its commands, only letting pedestrians cross when vehicles were actually stopped or there were no more vehicles present. It also correctly interprets relevant vehicles for the crossing and is robust to moving vehicles that do not pose a threat to the pedestrian in the crossing. The system was evaluated to work reliably with vehicles up to 50 meters away and moving up to 40 km/h. The only encountered failure case was caused by a vehicle already stopped at the crossing when recording the reference frame for *Vehicle staticity evaluation*, where the pedestrian was let pass because the system incorrectly determined that there were no vehicles instead of detecting that there was a stopped vehicle.

The largest source of errors for our system was the lack of precise depth information stemming from reliance on RGB-D sensors which were sensitive to changes in lighting conditions. As the system operates within certain distance margins to guarantee the safety of pedestrians, accurate distance measurements are needed for making correct and timely decisions.

## V. Conclusion

This work demonstrates the feasibility and effectiveness of deploying a humanoid social robot—ARI—for enhancing pedestrian safety at unsignalized two-lane road crossings. By integrating a robust perception system, real-time object detection with YOLOv12, and depth-based distance estimation, ARI effectively assesses traffic scenarios and communicates with both pedestrians and drivers using gestures and voice commands. The structured behavior sequence ensures controlled, dynamic interaction between traffic participants, significantly improving the safety and clarity of the crossing process.

The experimental evaluation conducted in controlled environments validates the robustness of the proposed system under diverse pedestrian and vehicle behaviors. Achieving an overall success rate of 83.33%, the system demonstrated high reliability in critical perception tasks and decision-making processes. Importantly, it maintained safety even when drivers ignored commands, a crucial feature for real-world deployment.

Future work will address on improving the weakness and limitations of the system caused primarily by sensors which proved to be insufficient for executing certain tasks needed for enhancing pedestrian safety. We plan to improve the sensor suite of ARI by integrating a 3D LiDAR puck which will help with making rapid and accurate decisions. Our aim is to also integrate a more advanced pedestrian intention prediction method. Once sufficient improvements are made, the proposed system will be evaluated in a user study with unbiased participants. A separate study is planned to determine the user acceptance and accessibility of the humanoid social robot for increasing pedestrian safety; targeting trust and perceived safety, communication and understandability among school-aged children. This will help us determine the suitability of our proposed solution for protecting a specific group of vulnerable road users.

We also plan to integrate a V2X unit enabling communication with other connected road participants. This can enhance its situational awareness thanks to cooperative perception, greatly increasing perception range. It would also allow the robot to notify vehicles in advance and also limit their speed in critical situations to further improve the safety of vulnerable road users.

## Acknowledgment

This work has been funded as part of the KIT Future Fields project ”V2X4Robot”, ELLIIT Excellence Center and Hilti Foundation. This paper is part of the CulturalRoad project, funded by the European Union under grant agreement No. 101147397. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the EU or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the EU nor the granting authority can be held responsible for them.

## References

[1] L. Gong et al.,“Real-time human-in-the-loop remote control for a life-size traffic police robot with multiple augmented reality aided display terminals,” 2017 2nd International Conference on Advanced Robotics and Mechatronics (ICARM), Hefei and Tai’an, China, 2017, pp. 420-425, doi: 10.1109/ICARM.2017.8273199.

[2] L. Zhao et al., “A Bionic Arm Mechanism Design and Kinematic Analysis of the Humanoid Traffic Police,” 2019 IEEE 9th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), Suzhou, China, 2019, pp. 1606-1611, doi: 10.1109/CYBER46603.2019.9066510.

[3] F. Ghaffar, ”Controlling Traffic with Humanoid Social Robot,” arXiv preprint arXiv:2204.04240, Systems Design Engineering, University of Waterloo, Waterloo, Canada, 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2204.04240.

[4] S.C. Kumaran, A. Oberlender, A. Grishko, B. Megidish, H. Erel, “To Cross or Not-to-Cross: A Robotic Object for Mediating Interactions Between Autonomous Vehicles and Pedestrians,” 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN).

[5] H. Kulhandjian, “Smart Robot Design and Implementation to Assist Pedestrian Road Crossing,” 2024

[6] J. Redmon et al., “You only look once: Unified, real-time object detection, ” 2016 Proceedings of the IEEE conference on computer vision and pattern recognition, doi: 10.1109/CVPR.2016.91

[7] M. Schrapel, M. Bied, B. Bruno, and A. Vinel, ”Experiencing Social Robots for Traffic Guidance using Virtual Reality Videos,” Mensch und Computer 2024 - Workshopband, Gesellschaft für Informatik e.V., 2024, doi: 10.18420/muc2024-mci-demo-318.

[8] M. Stuede, K. Westermann, M. Schappler and S. Spindeldreier, ”Sobi: An Interactive Social Service Robot for Long-Term Autonomy in Open Environments,” 2021 European Conference on Mobile Robots (ECMR), Bonn, Germany, 2021, pp. 1-8, doi: 10.1109/ECMR50962.2021.9568838.

[9] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick (2014). Microsoft coco: Common objects in context. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13 (pp. 740-755). Springer International Publishing.

[10] H.W. Kuhn, “The Hungarian method for the assignment problem” 1955 Naval Research Logistics, 2: 83-97, doi: 10.1002/nav.3800020109

[11] Tian, Yunjie, Ye, Qixiang, & Doermann, David. (2025). YOLOv12: Attention-Centric Real-Time Object Detectors. arXiv preprint arXiv:2502.12524.

[12] Tian, Yunjie, Ye, Qixiang, & Doermann, David. (2025). YOLOv12: Attention-Centric Real-Time Object Detectors [Computer software]. Retrieved from https://github.com/sunsmarterjie/yolov12. Licensed under AGPL-3.0.