

Received 14 February 2024, accepted 19 February 2024, date of publication 26 February 2024, date of current version 6 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3370469

RESEARCH ARTICLE

Autonomous and Human-Driven Vehicles Interacting in a Roundabout: A Quantitative and Qualitative Evaluation

LAURA FERRAROTTI¹, MASSIMILIANO LUCA^{1,2}, GABRIELE SANTIN^{1,3}, GIORGIO PREVIATI^{1,4},
GIANPIERO MASTINU^{1,4}, MASSIMILIANO GOBBI^{1,4}, ELENA CAMPI^{1,4}, LORENZO UCCELLO^{1,4},
ANTONINO ALBANESE^{1,5}, PRAVEEN ZALAYA⁶, ALESSANDRO ROCCASALVA⁶,
AND BRUNO LEPRI¹

¹Mobile and Social Computing Laboratory, Fondazione Bruno Kessler, 38123 Trento, Italy

²Faculty of Computer Science, Free University of Bolzano, 39100 Bolzano, Italy

³Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, 30123 Venice, Italy

⁴Department of Mechanical Engineering, Politecnico di Milano, 20133 Milan, Italy

⁵Italtel S.p.A., 20153 Milan, Italy

⁶Centro Ricerche Fiat, Turin, 10043 Orbassano, Italy

Corresponding author: Massimiliano Luca (mluca@fbk.eu)

This work was supported by the EU H2020 Research and Innovation Programme AI@EDGE (<https://aiatedge.eu/>) under Grant 101015922. The work of Gabriele Santin and Bruno Lepri was supported by the PNRR ICSC National Research Centre for High-Performance Computing, Big Data, and Quantum Computing under the NRRP MUR Program funded by the NextGenerationEU under Grant CN00000013.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethical Committee of the Polytechnic University of Milan under Approval No. 16/2023.

ABSTRACT Optimizing traffic dynamics in an evolving transportation landscape is crucial, particularly in scenarios where autonomous vehicles (AVs) with varying levels of autonomy coexist with human-driven cars. While optimizing Reinforcement Learning (RL) policies for such scenarios is becoming more and more common, little has been said about realistic evaluations of such trained policies. This paper presents an evaluation of the effects of AVs penetration among human drivers in a roundabout scenario, considering both quantitative and qualitative aspects. In particular, we learn a policy to minimize traffic jams (i.e., minimize the time to cross the scenario) and to minimize pollution in a roundabout in Milan, Italy. Through empirical analysis, we demonstrate that the presence of AVs can reduce time and pollution levels. Furthermore, we qualitatively evaluate the learned policy using a cutting-edge cockpit to assess its performance in near-real-world conditions. To gauge the practicality and acceptability of the policy, we conduct evaluations with human participants using the simulator, focusing on a range of metrics like traffic smoothness and safety perception. In general, our findings show that human-driven vehicles benefit from optimizing AVs dynamics. Also, participants in the study highlight that the scenario with 80% AVs is perceived as safer than the scenario with 20%. The same result is obtained for traffic smoothness perception.

INDEX TERMS Transportation, autonomous vehicles, urban mobility, reinforcement learning.

I. INTRODUCTION

Modern society grapples with a large amount of societal challenges. Among the most pressing is the constant increase

The associate editor coordinating the review of this manuscript and approving it for publication was Jiachen Yang¹.

in levels of pollution and related traffic congestions [1], both of which threaten the sustainability and livability of our urban environments [2], [3]. At the same time, at the forefront of the technological revolution are autonomous vehicles (AVs). These vehicles, driven by advanced algorithms and sensors, promise to redefine how we perceive transportation [4],

[5], [6]. They can potentially alleviate some of the most persistent issues of modern urban transportation, from increasing safety [7], [8], [9], [10], [11] to optimizing traffic flow [12], [13], [14], [15]. However, as with any nascent technology, the real-world implementation of AVs is laden with challenges. Among the most significant are the safety, and prohibitive costs associated with testing and validating the efficiency of these vehicles in real conditions. One aspect that is fundamental to address these challenges is the evaluation of driving scenarios in which AVs are present together with human drivers, in order to assess not only the performance attained by the AVs, but the behavior of the whole hybrid multi-agent system of drivers, together with the perception of driving comfort and safety experienced by the human drivers involved. In our paper, in order to conduct such analysis conjugating both safety and realistic conditions, we bypass the logistical and financial constraints of real-world testing by harnessing the power of state-of-the-art simulation tools. Our primary focus is on a small-scale yet intricately complex scenario: a roundabout in Milan, Italy. Utilizing Simulation of Urban MObility (SUMO) [16], a cutting-edge traffic simulator, we create a realistic environment where both AVs and human-driven vehicles (HVs) coexist, navigating the roundabout under realistic traffic loads. Also, we bridge the gap between static simulation and real-world experience by integrating SUMO with VI-WorldSim¹ [17]. This user-friendly, fully integrated graphic environment not only accelerates vehicle development offline but also facilitates a more immersive experience on driving simulators. To enhance the realism of our study, we leverage a high-fidelity cockpit that replicates real-world driving conditions installed at the DrisMi Lab at the Polytechnic University of Milan,² enabling us to evaluate, to the best of our knowledge for the first time, the scenario from a qualitative perspective. In our simulation, HVs adhere to realistic dynamics as simulated by SUMO, while AVs actions are dictated by a policy learned via Reinforcement Learning (RL).

In general, the integration of microscopic traffic simulators with driving simulators is not a novel concept [18], [19]. However, previous studies have presented certain challenges, which our paper addresses and resolves [20], [21], [22], [23]. Specifically, our approach ensures *synchronized* co-simulation in *real-time*, utilizing *the same road network* in both simulators and achieving an impressively low delay of 5 ms.

Our findings, as detailed in this paper, shed light on the multifaceted impact of AV integration. While the benefits of AVs in reducing pollution and alleviating congestion are evident, interestingly enough, the presence of AVs also augments the efficiency of human-driven vehicles. As we show in the Tables in Section V-B, as the AV penetration rate increases, the ripple effects are felt across the entire traffic

ecosystem, offering insights into a future where harmonious coexistence between AVs and HVs might redefine urban transportation. More precisely, we evaluate the learned policy and the interaction between AVs and HVs qualitatively using traffic smoothness perception and safety perception as metrics. The evaluation is carried by surveying the perception of individuals who use the cockpit installed at the DrisMi Lab. While a number of participants felt a difference between the presence of 20% and 80% of AVs on the streets, most of the people preferred the scenario with an 80% penetration rate of AVs both in terms of safety and smoothness as highlighted in the Tables of Section V-B.

The advances presented in our study can be summarized as follows:

- We propose a framework that consists of the integration of three realistic simulators: SUMO, VI-WorldSim, and a cockpit. This framework enables quantitative and qualitative evaluation of AVs and HVs interactions.
- We employ RL to learn AVs behaviors in a real-world scenario (i.e., a roundabout in Milan, Italy) with realistic traffic loads.
- We measure the reduction of crossing time (up to $-10,72\%$ for AVs and $-8,52\%$ for HVs), emissions (up to $-38,98\%$ for AVs and $-39,13\%$ for HVs) and consumption (up to $-35,82\%$ for AVs and $-35,15\%$ for HVs) as a quantitative metric for policy evaluation, and traffic smoothness perception and safety perception as qualitative metrics.

The rest of the paper is structured as follows. In Section II, we briefly introduce SUMO, VI-WorldSim, the cockpit and RL concepts as background notions. We then conduct a literature review of works that study mixed-traffic scenarios, and of works leveraging RL for AV and traffic simulation in Section III. In Section IV, we first discuss how we turn SUMO into an RL environment, the algorithm we use, namely Proximal Policy Optimization (PPO) [24], how the test environment (i.e., the roundabout in Milan) is designed, and how the data for policy fine-tuning and traffic loads are collected. In Section V, we present both the quantitative and qualitative results, and we comment on some of the findings. In Section VI, we summarize the paper and present some interesting future directions to follow.

II. BACKGROUND

A. SIMULATIONS IN SUMO

Simulations are extensively used in a variety of fields such as urban planning [25], transportation [26], [27], [28], robotics [29], epidemiology [30], gaming [31], and others. Experiments in the real world are often costly, dangerous, and infeasible. Thus, simulators provide a solution for evaluating hypotheses and methodologies *in silico* where certain aspects of the behavior faithfully mirror the real world.

SUMO [16] is a state-of-the-art multi-agent simulator for transportation systems that reproduce realistic behaviors of drivers. In SUMO, it is possible to deploy multiple agents that use different transportation means (e.g., cars, public

¹<https://www.vi-grade.com/en/products/vi-worldsim/>

²<https://www.drismi.polimi.it/drismi/>

transit, bicycles) to reach different goals. The agents can move within a street network that defines the environment in which they can operate. Interestingly enough, SUMO is also a *microscopic* traffic simulator, i.e., each agent is modeled as an individual based on separate and different car-following and lane-changing models.

SUMO's workflow to simulate realistic traffic is organized as follows. First, the road network is defined to match the real world, and road loads (e.g., a realistic number of agents using a specific street) are specified. Second, the agents are executed in the road network through high-fidelity simulations and scored according to a cost function that measures if certain goals have been reached. Next, as in this work we use a learned Reinforcement Learning policy trained in collaboration among all the AVs involved, SUMO re-plans, re-executes, and re-scores the actions taken by the agents using a co-evolutionary algorithm until nobody can unilaterally improve their trips. At that point, the system has an equilibrium, and we can inspect the individuals' typical behaviors.

B. VI-WorldSim

VI-WorldSim is an innovative software solution designed to facilitate creating and testing lifelike driving scenarios. These scenarios include traffic flow, pedestrians, weather conditions, and sensor feedback. By utilizing VI-WorldSim, individuals can immerse themselves in many realistic environments, from bustling urban streets to expansive highways and specialized test sites. This versatility allows for an in-depth evaluation of how a vehicle model responds to many situations, ensuring that every potential challenge is addressed during development. Also, the realism of the simulations combined with the integration with realistic cockpits can be used to evaluate scenarios qualitatively safely and relatively cheaply. An example of what a VI-WorldSim simulation looks like can be seen in Figure 1 B.

C. VI-WorldSim AND SUMO INTERACTION

The traffic scenario is generated thanks to a co-simulation between SUMO, which is used as a traffic engine, and VI-WorldSim, which simulates the vehicle's motion driven by the human in the loop and allows a graphical representation of the traffic scenario. SUMO is in charge of simulating all virtual vehicles in the roundabout, comprising HVs and AVs. In particular, SUMO receives the data about the car driven by the human in the loop (ego-vehicle) from the driving simulator through VI-WorldSim. The current position of all simulated vehicles is fed to VI-WorldSim which is in charge of all the graphical environment of the driving simulator, of the interface with the human in the loop and of the simulation of the motion of the ego-vehicle according to the request of the real human driver. All these simulations are performed in real-time and the corresponding data are stored in a real-time database. Additional information about the communication schema between SUMO and VI-WorldSim can be found in [17]

A) The Simulator at DrisMi Lab - PoliMI



B) The integration between SUMO and VI-WorldSim



FIGURE 1. A) The cockpit was installed at DrisMi laboratory of the polytechnic university of Milan. B) An example of the output of VI-WorldSim after the integration with SUMO.

TABLE 1. A summary of the physical characteristics of the driving simulator (cockpit).

Physical quantity	Values
Platform Size	6 meters x 6 meters
Visual System (H)	270°
Visual System (V)	90°
Degree of freedom	9
Longitudinal Acc. of the Base	1.5 G-force
Lateral Acc. of the Base	1.5 G-force
Vertical Acc. of the Cockpit	2.5 G-force
Longitudinal Travel	4.2 meters
Lateral Travel	4.2 meters
Vertical Travel	± 298 millimeters
Yaw Angle	± 62°
Roll Angle	± 15°
Pitch Angle	± 15°

D. THE COCKPIT

The human in the loop drives the vehicle from the cockpit of the driving simulator, which moves accordingly to the simulated motion of the ego vehicle. The cockpit can be seen in Figure 1 A. The cockpit is equipped with a telematic box, which is connected to the Controller Area Network bus of the cockpit, reads the dynamic data of the car, and transmits them to a 5G radio platform. The 5G radio platform transmits the data to and from an edge server and is controlled by a Next Unit of Computing (NUC) where the services (e.g., learned policy) are installed. Finally, the edge server hosts the RL infrastructure, including the policy, which is used to control the connected automated vehicles simulated by SUMO. The characteristics of the cockpit are summarized in Table 1.

E. REINFORCEMENT LEARNING

Reinforcement Learning (RL) constitutes a branch within the domains of Artificial Intelligence and Machine Learning that draws its initial inspiration from the Pavlovian paradigm of conditioning, wherein organisms adapt their behavioral choices based on the gratifications (rewards) and aversions (punishments) received. Analogously, in the classic RL setup an artificial agent (sometimes referred to as an actor) is engaged in dynamic interactions with its environment, thereby selecting at discrete time intervals its course of action. RL algorithms aim at training the agent, enabling it to independently attain optimal behaviors within a designated environment, in alignment with contextually defined objectives

inherent to the specific problem at hand. The remaining part of this section summarises the fundamental RL definitions (see for a more detailed introduction [32], [33], [34]). The RL problem can be mathematically formalized as a Markov Decision Process, by considering a tuple (S, A, p, r) such that S and A are the sets of possible agent' states and actions, respectively, while $p : S \times A \times S \rightarrow [0, 1]$ indicates the probability to transition from state s to state s' by acting a , i.e. $p(s, a, s') = \mathbb{P}(s_{t+1} = s' | s_t = s, a_t = a)$. The reward function $r : S \times A \rightarrow \mathbb{R}_0^+$ maps a state-action couple with its immediate, intrinsic desirability in relation to the agent's task at hand. In this context, the agent's decision-making process is modeled as a policy function $\pi : S \times A \rightarrow [0, 1]$, such that $\pi(s, a)$ represents the probability of acting a while in state s , determining *de facto* the agent's behavior. It is possible to define a value function $V_\pi : S \rightarrow \mathbb{R}_0^+$ associated with any policy π , that represents the expectation with respect to p of the cumulative discounted rewards obtained by π over time from each state, considering $\gamma \in (0, 1]$ as discounting factor penalizing rewards obtained further on in time, i.e.,

$$V_\pi(s) = \mathbb{E}_{\substack{s_{t+1} \sim p_t \\ a_t \sim \pi_t}} \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \mid s_0 = s \right],$$

where we define per each step t the distributions $p_t = p(s_t, a_t, \cdot)$ and $\pi_t = \pi(s_t)$. It is sometimes useful to consider the action-value function $Q_\pi : S \times A \rightarrow \mathbb{R}_0^+$ associated to π , where $Q_\pi(s, a)$ indicates the expected cumulative discounted rewards obtained by following policy π over an infinite horizon, starting from state s and performing as a first action a , i.e.

$$Q_\pi(s, a) = \mathbb{E}_{\substack{s_{t+1} \sim p_t \\ a_t \sim \pi_t}} \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \mid s_0 = s, a_0 = a \right],$$

keeping in mind that the following property holds

$$Q_\pi(s, a) = \mathbb{E}_{s' \sim p(s, a, \cdot)} \left[r(s, a) + \gamma V_\pi(s') \right].$$

The agent's overall goal is to learn a policy π^* that maximizes the expected long-term desirability of each state, i.e., such that for each $s \in S$

$$\pi^*(s) = \arg \max_{\pi} V_\pi(s). \quad (1)$$

In many cases it is common to model r as a function linking states and actions with an associated cost (instead of a reward), without loss of generality, and substitute the maximization with a minimization in (1).

III. LITERATURE REVIEW

This section reviews significant contributions in the literature regarding the study of mixed-traffic scenarios and the design of RL policies for AVs. In this work we do not focus on the development of new RL algorithms for AVs training. On the contrary, in order to realistically study the interaction and coexistence of AVs and HVs, we find more sensible to employ

established and, hence, more reliable RL techniques, closer to the deployment phase, than extremely recent methods. Hence we introduce the related literature not from the algorithmic point of view, but summarizing some of the most common problem design choices.

A. AVs-HVs INTERACTION

Predictions suggest a gradual market integration of AVs, with estimates ranging from 24% to 87% market share by 2045 [35], [36]. Most of the studies about the role of AV integration explore how such vehicles will lead to a fundamental transformation in mobility [37], [38], [39], [40], [41]. A number of research papers have delved into the advantages of AVs, and in particular Connected AVs (CAVs), concluding that they can notably decrease traffic incidents, alleviate congestion, lower fuel use, and offer essential mobility options [37], [38], [39], [42], [43], [44], [45], [46], [47]. Researchers highlighted that even a small penetration of AVs may significantly impact, for example, the number of parking requirements and the number of accidents [38], [42]. Other researchers also investigated how street features like Variable Speed Limits and road capacities impact urban scenarios under different levels of AV integration [40], [41]. Focusing on the interaction between HVs and AVs, in [48] researchers propose considerations on the design of AVs that can safely and intuitively interact with other traffic participants, based on common human interaction strategies. A driving simulator experiment was conducted in [49] on 34 individuals, to investigate the behavior of HVs exposed to different road design of lanes dedicated to AVs on motorways. The coexistence and interaction of AVs and HVs was investigated as well in [50], through a field test conducted on 18 participants, focusing on gap acceptance, car-following, and overtaking behaviors, showing that drivers interacting with recognizable AVs adopt smaller critical gaps, and after overtaking, merge closer in front of those. The work in [51] is devoted to understanding HVs behavior in mixed traffic at un-signalized priority T-intersections, through a driving simulator experiment on 95 human drivers, whose findings suggest that human drivers change their gap acceptance behavior in mixed traffic depending on AVs recognizability and driving style. Finally, the authors of [52] aimed at quantifying the behavioral changes caused by human drivers following either an AV or an HV, and their impact on safety, fuel consumption and pollution, by analyzing data from a field experiment on 9 drivers. Their work shows that, when human drivers follow AVs, lower driving volatility in terms of speed and acceleration can be achieved, and consequently more stable traffic flow behavior, lower crash risk, less fuel consumption and emission production are experienced. However, to the best of our knowledge, there is no study that evaluates how the penetration rate of AVs may impact the human driver's perception of traffic smoothness, safety and comfort in a scenario in which AVs and HVs coexist. Such evaluation is the main goal of our study.

B. RL DESIGN FOR AVs

In the context of RL policies training for AVs, literature can be found detailing the most common choices related to the design of the RL problem to be solved, in particular regarding states, actions and rewards design for AVs training. It is frequent to consider a states space that includes the position, heading, velocity of the vehicle, and the presence of obstacles within the sensor's view, possibly employing a Cartesian or Polar occupancy grid centered around the vehicle. This grid is often enhanced with lane-related details such as lane number, path curvature, historical and predicted trajectory of the vehicle, longitudinal measures such as time-to-collision, and broader scene-related information such as traffic regulations and the locations of traffic signals [53], [54]. The action space instead is by definition related to the actuators present on the vehicle and devised to the vehicle control task. Multiple actuators come into play, both continuous (as for instance steering angle, throttle, and brake) and discrete (i.e., gear changes). A framework incorporating temporal abstractions, such as options (sub-policies that extend primitive actions over multiple time steps, to be chosen instead of low-level actions), can simplify action selection [55]. Concerning the design of suitable reward functions for RL agents in the context of autonomous driving, researchers follow a variety of approaches, in order to tackle the many different sub-skills that all together characterize the general goal of autonomously driving a vehicle. Examples include measures such as distance traveled towards a destination [56], speed of the vehicle [56], [57], [58], maintaining the ego vehicle at a standstill [59], avoiding collisions with other road users or scene objects [56], [57], adherence to sidewalk rules [56], staying in the lane, ensuring comfort and stability while avoiding extreme acceleration, braking, or steering [58], [59], and adherence to traffic regulations [57].

IV. METHODOLOGY

This section describes all the steps and experimental setups adopted to carry out the study. First, we describe how SUMO can be turned into a realistic RL environment using Python libraries like Flow [60], Ray RLLib [61] and OpenAI Gym [31].

A. TURNING SUMO INTO A REALISTIC REINFORCEMENT LEARNING-BASED ENVIRONMENT

In this work, we transform SUMO, a high-fidelity multi-agent transportation simulator, into a realistic Reinforcement Learning environment to optimize and evaluate policies for AVs. The ultimate goal is to use SUMO to learn policies in which AVs take optimal actions to reduce emissions and to minimize the time to cross a real roundabout in Milan, Italy.

To transform SUMO into a Reinforcement Learning environment, we integrate Flow [60], an open-source Python package that can be used to create a communication layer between SUMO and Ray RLLib [61]. Remarkably, Flow can be used to investigate the so-called mixed autonomy scenarios

where only a portion of the deployed cars are AVs, and the others are controlled by car following models (CFMs), a set of ordinary differential equations that realistically mimic basic traffic dynamics on single-lane roads. This ability to study mixed autonomy scenarios is paramount for policymakers and traffic engineers as it represents a more realistic short-term scenario. In Flow, SUMO is connected with Ray RLLib and OpenAI Gym [31] through TraCI, a package to control communication over network protocols. The environment in which agents operate consists of a realistic network representing a physical road layout (e.g., speed limits, lanes, length, shape). The actors are the deployed cars. Some of them (marked with "rl_agent") are controlled by a learned policy and make decisions according to a specific goal they have to minimize. Other cars (marked with "human_agent") base their decision on pre-defined driver models.

Flow allows to rely on *observer* functions to map the set of states S to the observations O . This permits to tailor the information provided to the controller, choosing a subset of the SUMO states of the vehicle. We structure the observation vector $o_t^n \in O$ available at instant t to vehicle n so that it contains the last measured value of position x_t^n and acceleration \ddot{x}_t^n of the vehicle. Moreover, we include in o_t^n information obtained from the estimated position and acceleration of its front (F), back (B), left (L) and right (R) neighbors in the scenario, indicated respectively as $\{x_t^k, \ddot{x}_t^k\}_k$ with $k = F, B, L, R$, if such vehicles are present at instant t around vehicle n . The values associated with the neighbors that are missing at time t are replaced by an opportune placeholder. The described components of the observations vector are divided by quantities that are characteristic of the chosen scenario, such as the scenario's dimension x_{\max} and the maximum acceleration \ddot{x}_{\max} . Summarizing, we consider as observation o_t^n the vector

$$o_t^n = \begin{bmatrix} x_t^n/x_{\max} \\ \ddot{x}_t^n/\ddot{x}_{\max} \\ x_t^F/x_{\max} \\ (\ddot{x}_t^F - \ddot{x}_t^n)/\ddot{x}_{\max} \\ x_t^B/x_{\max} \\ (\ddot{x}_t^B - \ddot{x}_t^n)/\ddot{x}_{\max} \\ x_t^L/x_{\max} \\ (\ddot{x}_t^L - \ddot{x}_t^n)/\ddot{x}_{\max} \\ x_t^R/x_{\max} \\ (\ddot{x}_t^R - \ddot{x}_t^n)/\ddot{x}_{\max} \end{bmatrix}. \quad (2)$$

The action $a_t^n \in A$ decided by agent n at time t , instead, consists of the next acceleration value \ddot{x}_{t+1}^n to actuate, and of a discrete decision $c_{t+1}^n \in \{0, 1\}$, corresponding to changing the line/maintaining the current one, that is

$$a_t^n = \begin{bmatrix} \ddot{x}_{t+1}^n \\ c_{t+1}^n \end{bmatrix} \quad (3)$$

Our work aims at leveraging the presented infrastructure to allow AVs to learn via RL techniques the optimal control law to cross a roundabout as fast as possible. In this sense, we design a stage-cost function $r : \mathcal{O} \times \mathcal{A} \rightarrow \mathbb{R}_{\geq 0}$ measuring

the deviation d of the AVs velocity \dot{x}_t^n from a user-defined desired velocity v . In order to penalize the early termination of roll-outs due to collisions or other failures, we subtract such deviation from the peak allowable deviation $d_{\max}(v)$. Finally, to ensure non-negativity, the cost is then bounded below by 0, i.e.,

$$r_v(o_t^n, a_t^n) = \max(0, \|d(\dot{x}_t^n, v) - d_{\max}(v)\|). \quad (4)$$

Simultaneously, we are interested in reducing polluting emissions. Hence, we leverage the previously described velocity-based stage-cost with an analogous one, punishing the deviation from a target level of pollution P , i.e., $d(p_t^n, P)$, where the pollution levels p_t^n are measured in function of the actuated decisions a_t^n , and P is estimated by considering a single vehicle in the scenario, running at constant velocity. The two components of the cost are summed and normalized, and then assigned as a stage-cost to the AV. We use the well-known Proximal Policy Optimization [24] algorithm (see Section IV-B for more details) to learn a policy allowing the AVs to decide at each instant which action to take, based on the agent's individual observations' vector $o_t^n \in O$. All the AVs involved in the roundabout scenario collaborate in training a central policy by providing their simulated experience in order to update the policy parameters. Moreover, during training and deployment, when interrogated, the policy exploits both the local observations of the agent taking the decision and information related to the position and acceleration of other vehicles in the roundabout that are received through a central communication scheme.

Differently from the state-of-the-art studies, we evaluate the learned policy quantitatively and qualitatively, analyzing results obtained through tests carried out at the DriSMi Laboratory of Polytechnic University of Milan. This laboratory has a high-fidelity last-generation cable-driven driving simulator (see Figure 1 and [62], [63]). The traffic scenario is generated thanks to a co-simulation between SUMO (traffic engine), and VI-WorldSim,³ that simulates the motion of the vehicle driven by the human in the loop and allows to have a graphical representation of the traffic scenario. The combination of SUMO and VI-WorldSim allows us to evaluate the learned policy in terms of the traffic smoothness and safety perception of the passengers (i.e., real humans who agreed to participate in the experiments in our case). Further details about the evaluation procedure are shared in Sections V-A and V-B.

B. PROXIMAL POLICY OPTIMIZATION

To learn the policy parameters, we employ the Proximal Policy Optimization (PPO) method [24]. PPO is a prominent policy-optimization algorithm, deriving from the Trust Region Policy Optimization (TRPO) algorithm [64], and improving it from the flexibility and computational complexity point of view. Both PPO and TRPO are designed to stably and efficiently optimize decision policies, focusing

on refining policies by iteratively adjusting their parameters while limiting the extent of these updates in order to maintain stability during learning. This is realized by PPO within a dual-step process: policy evaluation and policy improvement. During policy evaluation, data are gathered by executing the current policy within the environment. Subsequently, the advantages

$$A_{\pi_\theta}(s, a) = Q_{\pi_\theta}(s, a) - V_{\pi_\theta}(s)$$

of actions taken with respect to expected returns are computed. The advantages serve as a measure of how favorable the chosen actions were, with respect to expected outcomes. Following policy evaluation, policy improvement is performed through several epochs of optimization. In each epoch, PPO computes surrogate objectives that quantify the change in the policy's performance with respect to the previously considered set of policy parameters θ_{old} , guided by the advantage values, i.e.,

$$L_\theta(s, a) = \min \left(\frac{\pi_\theta(a|s)}{\pi_{\theta_{\text{old}}}(a|s)} A_{\pi_\theta}(s, a), \quad C A_{\pi_\theta}(s, a) \right),$$

with

$$C = \text{clip} \left(\frac{\pi_\theta(a|s)}{\pi_{\theta_{\text{old}}}(a|s)}, \quad 1 - \epsilon, \quad 1 + \epsilon \right).$$

These surrogate objectives facilitate the optimization of the policy in a manner that promotes positive action shifts while maintaining a threshold ϵ on the magnitude of policy updates. This threshold, referred to as the "clip parameter," curbs policy updates from straying too far from the original distribution, ensuring a measure of stability and preventing drastic policy shifts that could lead to instability. PPO's distinctive feature lies in its capacity to strike a balance between exploiting the advantages of updated policies and maintaining a controlled adjustment process. By constraining policy updates and employing the surrogate objective, PPO achieves stable and incremental policy improvements, contributing to efficient and reliable RL.

These characteristics, together with the performance attained by PPO in many benchmark examples and applications [65], motivated us to choose PPO as the learning algorithm for AVs policy optimization. Moreover, in order to conduct an objective evaluation of the effects of deploying AVs on realistic scenarios in the presence of human drivers, technologies that are both well-performing and well established, as for instance PPO, should be preferred to algorithms that, although cutting-edge, are not at maturity, and hence are less commonly deployed in real applications.

C. ROUNDABOUT DESIGN AND DATA COLLECTION

To operate in an environment that is as realistic as possible, we designed a roundabout inspired by a real-world one in Milan, Italy. In Figure 2 A), we can see how the roundabout looks in the real world, while in panel B),

³<https://www.vi-grade.com/en/products/vi-worldsim/>

we can see the SUMO's roundabout. It is a four-leg mini-roundabout, showing medium-high traffic and, therefore, being a challenging environment for the AV policy. Moreover, it has some important details, which make this particular scenario of general interest, specifically:

- every leg has pedestrian crosswalks immediately before the entrance of vehicles inside the circulatory roadway.
- Two of the legs are central arteries of the city, greatly increasing traffic on the roundabout.
- The roundabout has a standard configuration widely distributed in European urban areas [66] with significant flows.

A calibration procedure was conducted to replicate the number of vehicles approaching the intersection and their positions during the simulation. Firstly, measurements were taken for the maximum queue length, upstream and downstream flows for each leg, considering road vehicles, pedestrians, and bicycles on the actual roundabout. This process was repeated for six consecutive time slots, each lasting 10 minutes.

Subsequently, the results of these measurements were compared with simulations conducted in SUMO to calibrate the seven most relevant parameters that define the traffic conditions in the considered scenario. The parameters under consideration include the distance at which a pedestrian is considered, the minimum time interval to cross the path of another vehicle when entering the roundabout, the time before a driver enters the roundabout even if obstructing the way of an incoming vehicle, the maximum acceleration and deceleration of the vehicles, the time interval between vehicles, and the drivers' reaction time. For each simulation run, a cost function was constructed to compare the mean and maximum queue lengths between the measured and simulated data. The calibrated parameters are those that minimize the differences between the two sets of data. Specifically, Table 2 lists the SUMO parameters detailed below and their values:

- **jmCrossingGap**: minimum distance between the vehicle and the pedestrian that is heading toward the point of conflict of its trajectory with that of the vehicle;
- **jmTimegapMinor**: minimum time interval for a vehicle to enter an intersection where it does not have the right-of-way, before a vehicle with right-of-way;
- **impatience**: driver's intent to obstruct a vehicle with the right of way;
- **accel**: maximum acceleration for the selected vehicle type;
- **decel**: maximum deceleration for the selected vehicle type;
- **tau**: minimum time interval between consecutive vehicles;
- **actionStepLength**: driver reaction time.

As can be noted from 2 B), although present in the real roundabout, in the final network there are no restricted lanes or pedestrian crosswalks. Such elements have been removed, after the calibration process, as they are not within the scope

TABLE 2. SUMO calibrated parameters.

Parameter	Value
jmCrossingGap	1,3545
jmTimegapMinor	1,7792
impatience	0,1182
accel	1,7634
decel	4,2939
tau	1,3472
actionStepLength	0,505



FIGURE 2. A) The selected roundabout in Milan observed from OpenStreetMap; and in B), the corresponding design of the roundabout in SUMO. Also, realistic traffic loads for fine-tuning have been measured through field measurements.

of the AI@EDGE project. Correction coefficients, taken from the literature, have been used to take these elements into account in the modified model. For the same reason, only cars are considered; other vehicles and pedestrians are considered via equivalent coefficients [67].

V. RESULTS AND DISCUSSION

A. QUANTITATIVE EVALUATION

In our work, we focus on evaluating the efficacy of the learned policy in reducing the time AVs need to cross the roundabout and emission and fuel consumption. Concerning the former, given a simulation \mathcal{S} of 3600 seconds (i.e., 1 hour) with a given percentage of AVs and a total of n vehicles v_1, v_2, \dots, v_n , we associate to each vehicle v_i its entering time $t_{v_i}^{\text{in}}$ and exiting time $t_{v_i}^{\text{out}}$. The average time needed by the cars in the simulation \mathcal{S} to cross the roundabout is then computed as

$$\mu(\mathcal{S}) = \frac{1}{n} \sum_{i=1}^n (t_{v_i}^{\text{out}} - t_{v_i}^{\text{in}}). \quad (5)$$

Note that by sampling $V = \{v_1, v_2, \dots, v_n\}$ it is possible to estimate the average time for AVs and HVs. As a reminder, the simulation is carried out to have a total of 1540 vehicles passing through the roundabout over one hour as emerged from the field measurements. Time is measured in seconds. In Table 3 and in the right panel of Figure 3, we can see the results when having a different amount of AVs leveraging the learned policy.

TABLE 3. Given a certain penetration rate of AVs (column %AVs), we measure the average time that AVs and HVs need to cross the roundabout as in Eq.(5). The measurements are in seconds. We observe that as the number of AVs (column #AVs) increases, both AVs and HVs, on average, need less time to cross the roundabout, highlighting how HVs may also benefit from the optimization and diffusion of AVs.

%AVs	μ AVs (s)	μ HVs (s)	#AVs	#HVs
0	-	17.94	0	1540
10	17.15	17.32	154	1386
20	17.02	17.17	308	1232
30	16.92	17.12	462	1078
40	16.76	17.08	616	924
50	16.58	16.99	770	770
60	16.33	16.75	924	616
70	16.19	16.66	1078	462
80	15.98	16.63	1232	308
90	15.74	16.41	1386	154
100	15.31	-	1540	0

At 0% AV penetration, only HVs are present. The average time taken by HVs to cross the roundabout stands at 17.94 seconds. As we introduce AVs into the system, there's a slight improvement in the average crossing times for both types of vehicles. At 10% AV penetration, the average crossing time for AVs is 17.15 seconds. The HVs also experience a marginal decrease in their average time, clocking in at 17.32 seconds. As the percentage of AVs increases from 10% to 50%, there is a gradual and consistent decrease in the average crossing times for both AVs and HVs. At the midpoint, with 50% AVs and 50% HVs, AVs take an average of 16.58 seconds, while HVs take a slightly longer 16.99 seconds. Beyond the 50% mark, as AVs become more dominant, their average crossing times continue to decrease, reaching 15.31 seconds at 100% penetration. Interestingly, the HVs also benefit from the increased AV presence. At 90% AV penetration, when only a small fraction of HVs remain, their average crossing time reduces to 16.41 seconds - the minimum reached value. Throughout the progression, AVs consistently demonstrate a trend of reduced average crossing times with their increased presence. HVs also benefit from the introduction of AVs. The data suggests that as the percentage of AVs in traffic increases, the roundabout navigation becomes more efficient for both vehicle types. This could be attributed to the predictability and coordination of AVs, which seems to not only benefit their kind but also aid in optimizing the flow for human drivers.

Another evaluation metric that the policy should optimize is fuel consumption and pollution. Both information can be extracted from each vehicle's property and computed by SUMO. To give an idea of how the model is performing, instead of using the values provided by SUMO we report a more interpretable score normalized between 0 and 1. To this end, for each scenario \mathcal{S} , we take the worst-performing vehicle (score_{\max}) and the best-performing vehicle (score_{\min}) as normalizing factors. We then performed a min-max normalization, ending up with a score between 0 and 1 for each vehicle, with 0 indicating lower emission or lower fuel

TABLE 4. Given a certain penetration rate of AVs (column %AVs), we measure the consumption score and emission score of both AVs and HVs, following Eq.6. Similarly to what happened with the crossing time, we observe that as the number of AVs (column #AVs) increases, both AVs and HVs, on average, consume less fuel and reduce their emissions.

%AVs	AV		HV		#AVs	#HVs
	consumption	emission	consumption	emission		
0	-	-	0.74	0.69	0	1540
10	0.67	0.59	0.69	0.63	154	1386
20	0.61	0.56	0.64	0.58	308	1232
30	0.58	0.51	0.60	0.56	462	1078
40	0.54	0.48	0.58	0.54	616	924
50	0.53	0.46	0.56	0.51	770	770
60	0.51	0.44	0.54	0.48	924	616
70	0.47	0.40	0.50	0.46	1078	462
80	0.46	0.38	0.49	0.44	1232	308
90	0.45	0.37	0.48	0.42	1386	154
100	0.43	0.36	-	-	1540	0

consumption and 1 indicating the worst performance. In particular, the adopted formula is:

$$\mu(\mathcal{S}) = \frac{1}{n} \sum_{i=1}^n \left(\frac{\text{score}(v_i) - \text{score}_{\min}}{\text{score}_{\max} - \text{score}_{\min}} \right). \quad (6)$$

where n is the number of vehicles in simulation \mathcal{S} and the term $\text{score}(v_i)$ in the formula can indicate both the emission or the consumption level of the i -th vehicle v_i . The results can be seen in Table 4 and in the left -emissions- and center -consumption- panel of Figure 3.

When no AVs are present (0% AV penetration), the average fuel consumption score for HVs stands at 0.74, with an emissions score at 0.69. As AVs are introduced into the system at 10% penetration, their fuel consumption and emissions are recorded at 0.67 and 0.59, respectively. Interestingly, even with a modest AV penetration, there's an observable improvement in HV metrics as well. The consumption for HVs drops to 0.69, and their emissions decrease to 0.63. As the percentage of AVs in the system grows from 10% to 50%, there's a consistent improvement in both consumption and emissions for both vehicle types. At 50% AV penetration, AVs record consumption and emission values of 0.53 and 0.46, respectively, while HVs show values of 0.56 and 0.51. Beyond 50% AV penetration, the trend of decreasing consumption and emissions continues for AVs, reaching their lowest at 0.43 and 0.36, respectively at 100% AV penetration. HVs also see continuous improvement. Throughout the entire range, AVs exhibit lower consumption and emission values compared to HVs. The difference becomes more pronounced as the percentage of AVs increases, highlighting the efficiency of autonomous driving systems. In conclusion, not only do AVs demonstrate better fuel efficiency and lower emissions, but their presence also seems to positively influence the performance of human-driven vehicles, similar to travel time reduction. This could be attributed to smoother and more predictable traffic flow, leading to less stop-and-go traffic and more consistent driving speeds learned by the policy.

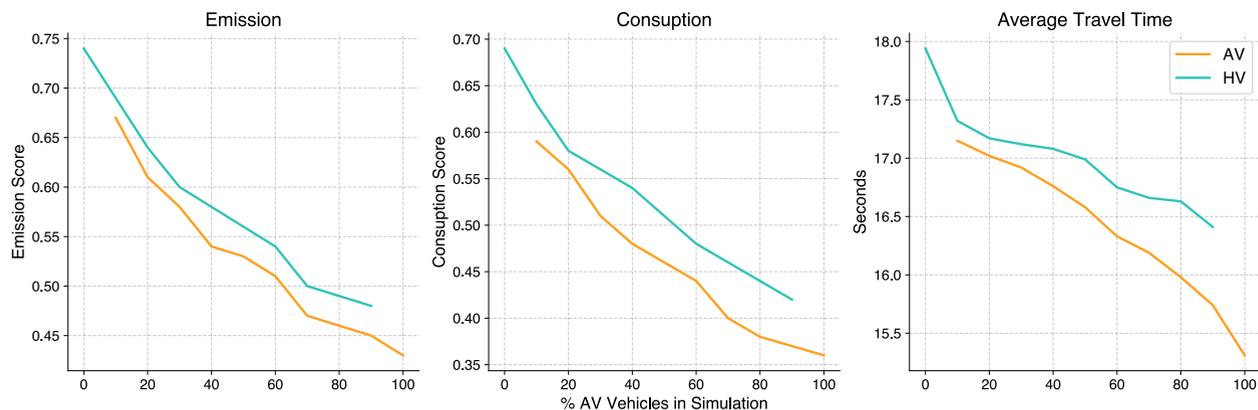


FIGURE 3. Concerning the quantitative analysis, we measured the impact of AV penetration rate on emissions, fuel consumption and average time to cross the scenario. On the left, we have the results related to emissions. In the middle, we show the results for fuel consumption while on the right we have the average crossing time. As we can observe, the best results are obtained for the scenario with 100% of AVs. However, having some AVs (e.g., 10% to 90% penetration rate already provides some benefits. Interestingly enough, also HV benefits from the penetration rate of AVs.

B. QUALITATIVE EVALUATION

As previously mentioned, tests have been conducted at the DriSMi Laboratory of the Polytechnic University of Milan⁴ to qualitatively evaluate indicators of the passengers’ comfort in the described driving scenario, considering different AV penetration levels. The laboratory has a high-fidelity, last-generation driving simulator as described in Section II-D. The integration between SUMO, VI-WorldSim and the simulator itself allows us to evaluate, to the best of our knowledge for the first time, traffic smoothness and safety perception of passengers as qualitative metrics to analyze the passengers’ comfort. To collect the necessary information for our study, a panel of ten participants has been selected for the preliminary tests. The participants were chosen from individuals without previous experience with driving simulators. The panel consists of 5 females and 5 males, aged between 22 and 33 years, with driving experience ranging from 1 to 15 years. Before the test, each participant was given instructions on how to operate the driving simulator and signed an informed consent form. Additionally, each participant spent about ten minutes driving in a simple motorway scenario to become familiar with the driving simulator before the actual test. Individuals were exposed in simulation to experiences in two driving scenarios, characterized by the 20% and 80% of AV penetration rate, respectively. Following the simulated ride, the participants were asked to fill out a quick survey, consisting of three close-ended questions:

- Regarding traffic smoothness, which of the following statements do you agree with the most?
- Regarding safety perception, which of the following statements do you agree with the most?
- Globally, which of the two scenarios did you prefer?

Per each question, the interviewed individuals were asked to select one among five possible alternatives, summarized in the following three tables (Table 5, 6 and 7 associated

TABLE 5. A summary of the answers to the first question of the survey. As we can see, most of the voters say that the scenario with 80% of AVs was smoother with respect to the one with 20% of AVs. One participant did not perceive differences between the two scenarios and 40% of participants preferred the scenario with 20% of AVs, although the majority of them only partially.

Regarding traffic smoothness, which of the following statements do you agree with the most?	Number of answers
Traffic in the scenario with 20% of AVs was definitely smoother than in the scenario with 80% AVs	1
Traffic in the scenario with 20% of AVs was partially smoother than in the scenario with 80% AVs	3
Traffic in the scenario with 20% of AVs was partially less smooth than in the scenario with 80% AVs	4
Traffic in the scenario with 20% of AVs was definitely less smooth than in the scenario with 80% AVs	1
I did not perceive differences	1

with the first, second, and third question, respectively). In total we collect feedback from 10 individuals. While the sample size is limited, a few tentative conclusions can be drawn. In Table 5, most of the respondents (5 out of 10) felt that the traffic was smoother in the scenario with 80% AVs, to varying degrees. A smaller segment (4 out of 10) felt the opposite, indicating that the 20% AV scenario was smoother. Just 1 out of 10 respondents did not perceive any significant difference in traffic smoothness between the two scenarios. Results suggest that as AV penetration increases, the perceived smoothness of traffic might improve. However, divergent opinions underscore the complexity of human perceptions and the subjective nature of such assessments.

Table 6 shows that a combined total of 8 respondents felt that the scenario with 80% AVs was safer to some degree compared to the 20% AV scenario. On the contrary, only 2 out of 10 respondents felt that the 20% AV scenario was safer in any capacity. All the participants perceived

⁴<https://www.drismi.polimi.it/drismi/>

TABLE 6. A summary of the answers to the second question of the survey. According to 80% of the participants, the scenario with 80% of AVs is perceived safer than the scenario with 20% of AVs. Zero participants did not perceive differences between the two scenarios and 20% of participants partially preferred the scenario with 20% of AVs.

Regarding safety perception, which of the following statements do you agree with the most?	Number of answers
Traffic in the scenario with 20% of AVs was definitely safer than in the scenario with 80% AVs	0
Traffic in the scenario with 20% of AVs was partially safer than in the scenario with 80% AVs	2
Traffic in the scenario with 20% of AVs was partially less safe than in the scenario with 80% AVs	2
Traffic in the scenario with 20% of AVs was definitely less safe than in the scenario with 80% AVs	6
I did not perceive differences	0

TABLE 7. A summary of the answers to the third question of the survey. From a general perspective, 70% of participants preferred the scenario with 80% of AVs while 30% preferred the scenario with 20% of AVs, although only partially.

Globally, which of the two scenarios did you prefer?	Number of answers
I definitely preferred the scenario with 20% AVs with respect to the scenario with 80% AVs	0
I partially preferred the scenario with 20% AVs with respect to the scenario with 80% AVs	3
I partially preferred the scenario with 80% AVs with respect to the scenario with 20% AVs	3
I definitely preferred the scenario with 80% AVs with respect to the scenario with 20% AVs	4
I cannot say which scenario I preferred	0

a difference in safety between the two scenarios. These results suggest an inclination towards perceiving higher AV penetration as safer. Analogously to what was observed while commenting the perception of smoothness in traffic, even when it comes to safety the diversity in opinions highlights that safety perceptions can be subjective and can vary among individuals.

Finally, Table 7 shows that, overall, these results hint at a leaning towards the 80% AV scenario (as shown by the 7 over 10 individuals preferring it either definitely or partially, versus only 3 just partially preferring the 20% AV scenario), suggesting that most participants might perceive benefits in scenarios with higher AV penetration. Naturally, even in this case the variety of responses also emphasizes the subjective nature of such preferences, and individual perceptions can vary based on personal experiences or beliefs. Further research could delve deeper into the reasons behind these preferences, providing a more comprehensive understanding of public sentiment toward AV integration, possibly by benefiting of a larger sample of collected feedback from individuals.

VI. CONCLUSION

This research has brought to light significant insights into how human drivers may perceive in terms of safety and comfort the integration of AVs in complex urban

traffic scenarios, specifically examining a roundabout in Milan, Italy. Our approach leveraged state-of-the-art traffic simulation tools like SUMO, VI-WorldSim, well-established RL methods and a high-fidelity cockpit to understand the dynamics of AVs penetrations alongside HVs.

The findings of our study underscore the substantial benefits that AVs offer in mitigating common urban challenges such as pollution and traffic congestion. By employing established RL techniques, notably the PPO algorithm, we were able to model and analyze the behavior of AVs in realistic traffic settings. The simulation results are promising, showing that AVs enhance their operational efficiency and positively influence the overall traffic flow, benefiting HVs in the process.

Our primary objective included conducting qualitative assessments with human participants. These assessments revealed a notable preference for environments where autonomous vehicles (AVs) are more prevalent, attributing this preference to enhanced safety and smoother traffic flow. Conducting such evaluations is crucial, as AVs and human-operated vehicles (HVs) will increasingly share roads in the future. Notably, previous research has not explored human drivers’ perceptions of AVs and HVs coexisting in a realistic, comprehensive manner, as our proposed framework does. Our findings are in line with quantitative data, indicating a future where AVs and HVs can coexist more seamlessly, leading to safer and more efficient urban environments.

While our study has made strides in understanding the potential of AVs in urban settings, it also opens avenues for future research. Further exploration into alternative reinforcement learning algorithms could provide deeper insights into the optimization of AV behavior. Additionally, expanding the scope of human-in-the-loop evaluations with a larger and more diverse participant pool would be invaluable in enriching our understanding of public perception and acceptance of AVs. Such studies could also explore other aspects of urban mobility impacted by AV integration, including pedestrian safety and public transportation systems.

ACKNOWLEDGMENT

(Laura Ferrarotti and Massimiliano Luca contributed equally to this work.)

REFERENCES

- [1] *The Sustainable Development Goals: Report 2022*, United Nations Dept. Econ. Social Affairs, New York, NY, USA, 2022.
- [2] X. Xie, I. Semanjski, S. Gautama, E. Tsiligianni, N. Deligiannis, R. Rajan, F. Pasveer, and W. Philips. “A review of urban air pollution monitoring and exposure assessment methods,” *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 12, p. 389, Dec. 2017.
- [3] M. Luca, G. M. Campedelli, S. Centellegher, M. Tizzoni, and B. Lepri, “Crime, inequality and public health: A survey of emerging trends in urban data science,” *Frontiers Big Data*, vol. 6, May 2023, Art. no. 1124526.
- [4] E. Thomas, C. McCrudden, Z. Wharton, and A. Behera, “Perception of autonomous vehicles by the modern society: A survey,” *IET Intell. Transp. Syst.*, vol. 14, no. 10, pp. 1228–1239, Oct. 2020.
- [5] M. Chikaraishi, D. Khan, B. Yasuda, and A. Fujiwara, “Risk perception and social acceptability of autonomous vehicles: A case study in Hiroshima, Japan,” *Transp. Policy*, vol. 98, pp. 105–115, Nov. 2020.

- [6] K. Hilgarter and P. Granig, "Public perception of autonomous vehicles: A qualitative study based on interviews after riding an autonomous shuttle," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 72, pp. 226–243, Jul. 2020.
- [7] J. Wang, L. Zhang, Y. Huang, and J. Zhao, "Safety of autonomous vehicles," *J. Adv. Transp.*, vol. 2020, pp. 1–13, Oct. 2020.
- [8] R. Mariani, "An overview of autonomous vehicles safety," in *Proc. IEEE Int. Rel. Phys. Symp. (IRPS)*, Mar. 2018, pp. 6A.1-1–6A.1-6.
- [9] P. Koopman, U. Ferrell, F. Fratrick, and M. Wagner, "A safety standard approach for fully autonomous vehicles," in *Proc. Int. Conf. Comput. Saf., Rel., Secur.* Cham, Switzerland: Springer, 2019, pp. 326–332.
- [10] P. Koopman and M. Wagner, "Autonomous vehicle safety: An interdisciplinary challenge," *IEEE Intell. Transp. Syst. Mag.*, vol. 9, no. 1, pp. 90–96, Spring 2017.
- [11] L. Ye and T. Yamamoto, "Evaluating the impact of connected and autonomous vehicles on traffic safety," *Phys. A, Stat. Mech. Appl.*, vol. 526, Jul. 2019, Art. no. 121009.
- [12] A. Mushtaq, I. U. Haq, M. U. Imtiaz, A. Khan, and O. Shafiq, "Traffic flow management of autonomous vehicles using deep reinforcement learning and smart rerouting," *IEEE Access*, vol. 9, pp. 51005–51019, 2021.
- [13] J. Zambrano-Martinez, C. Calafate, D. Soler, L.-G. Lemus-Zñiga, J.-C. Cano, P. Manzoni, and T. Gayraud, "A centralized route-management solution for autonomous vehicles in urban areas," *Electronics*, vol. 8, no. 7, p. 722, Jun. 2019.
- [14] E. Namazi, J. Li, and C. Lu, "Intelligent intersection management systems considering autonomous vehicles: A systematic literature review," *IEEE Access*, vol. 7, pp. 91946–91965, 2019.
- [15] A. Miglani and N. Kumar, "Deep learning models for traffic flow prediction in autonomous vehicles: A review, solutions, and challenges," *Veh. Commun.*, vol. 20, Dec. 2019, Art. no. 100184.
- [16] D. Krajzewicz, G. Hertkorn, C. Rössel, and P. Wagner, "Sumo (simulation of urban mobility)—An open-source traffic simulation," in *Proc. 4th middle East Symp. Simulation Modeling*, 2002, pp. 183–187.
- [17] G. Prevati and G. Mastinu, "Sumo roundabout simulation with human in the loop," in *Proc. SUMO Conf.*, vol. 4, 2023, pp. 29–40.
- [18] S. Espié and J.-M. Auberlet, "Joint use of driving simulation and traffic simulation for the study of road infrastructures and equipments," in *Proc. Joint Int. Conf. Comput. Decis. Making Civil Building Eng.*, 2006, pp. 1–12.
- [19] I. Vladislavjevic, J. M. Cooper, P. T. Martin, and D. L. Strayer, "Importance of integrating driving and traffic simulations: Case study of impact of cell phone drivers on traffic flow," *Transp. Res. Board*, Washington, DC, USA, Tech. Rep., 2009.
- [20] M. Hasan, D. Perez, Y. Shen, and H. Yang, "Distributed microscopic traffic simulation with human-in-the-loop enabled by virtual reality technologies," *Adv. Eng. Softw.*, vol. 154, Apr. 2021, Art. no. 102985.
- [21] M. Barthauer and A. Hafner, "Testing an adaptive cruise controller with coupled traffic and driving simulations," in *Proc. EPiC Ser. Comput.*, 2019, pp. 48–55.
- [22] M. Barthauer and A. Hafner, "Coupling traffic and driving simulation: Taking advantage of SUMO and SILAB together," *EPiC Ser. Eng.*, vol. 2, pp. 56–66, Jun. 2018.
- [23] J. Kathis, B. Schott, and F. Chucholowski, "Co-simulation of the virtual vehicle in virtual traffic considering tactical driver decisions," in *Proc. EPiC Ser. Comput.*, 2019, pp. 21–28.
- [24] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017, *arXiv:1707.06347*.
- [25] P. Waddell, "UrbanSim: Modeling urban development for land use, transportation, and environmental planning," *J. Amer. Planning Assoc.*, vol. 68, no. 3, pp. 297–314, Sep. 2002.
- [26] J. Barceló, *Fundamentals of Traffic Simulation*, vol. 145. Cham, Switzerland: Springer, 2010.
- [27] M. Pursula, "Simulation of traffic systems-an overview," *J. Geograph. Inf. Decis. Anal.*, vol. 3, no. 1, pp. 1–8, 1999.
- [28] L. Khaidem, M. Luca, F. Yang, A. Anand, B. Lepri, and W. Dong, "Optimizing transportation dynamics at a city-scale using a reinforcement learning framework," *IEEE Access*, vol. 8, pp. 171528–171541, 2020.
- [29] E. Todorov, T. Erez, and Y. Tassa, "MuJoCo: A physics engine for model-based control," in *Proc. IEEE/RSS Int. Conf. Intell. Robots Syst.*, Oct. 2012, pp. 5026–5033.
- [30] S. Eubank, H. Guclu, V. S. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczka, and N. Wang, "Modelling disease outbreaks in realistic urban social networks," *Nature*, vol. 429, no. 6988, pp. 180–184, May 2004.
- [31] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, "OpenAI gym," 2016, *arXiv:1606.01540*.
- [32] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.
- [33] D. Bertsekas, *Reinforcement Learning and Optimal Control*. Belmont, MA, USA: Athena Sci., 2019.
- [34] D. Bertsekas and J. N. Tsitsiklis, *Neuro-Dynamic Programming*. Belmont, MA, USA: Athena Sci., 1996.
- [35] A. Talebpour and H. S. Mahmassani, "Influence of connected and autonomous vehicles on traffic flow stability and throughput," *Transp. Res. C, Emerg. Technol.*, vol. 71, pp. 143–163, Oct. 2016.
- [36] M. S. Rahman and M. Abdel-Aty, "Longitudinal safety evaluation of connected vehicles' platooning on expressways," *Accident Anal. Prevention*, vol. 117, pp. 381–391, Aug. 2018.
- [37] D. Lee and D. J. Hess, "Regulations for on-road testing of connected and automated vehicles: Assessing the potential for global safety harmonization," *Transp. Res. A, Policy Pract.*, vol. 136, pp. 85–98, Jun. 2020.
- [38] H. Yang, Z. Wang, and K. Xie, "Impact of connected vehicles on mitigating secondary crash risk," *Int. J. Transp. Sci. Technol.*, vol. 6, no. 3, pp. 196–207, Sep. 2017.
- [39] A. Olia, H. Abdelgawad, B. Abdulhai, and S. N. Razavi, "Assessing the potential impacts of connected vehicles: Mobility, environmental, and safety perspectives," *J. Intell. Transp. Syst.*, vol. 20, no. 3, pp. 229–243, May 2016.
- [40] G. Perraki, C. Roncoli, I. Papamichail, and M. Papageorgiou, "Evaluation of a model predictive control framework for motorway traffic involving conventional and automated vehicles," *Transp. Res. C, Emerg. Technol.*, vol. 92, pp. 456–471, Jul. 2018.
- [41] H. Liu, X. Kan, S. E. Shladover, X.-Y. Lu, and R. E. Ferlis, "Modeling impacts of cooperative adaptive cruise control on mixed traffic flow in multi-lane freeway facilities," *Transp. Res. C, Emerg. Technol.*, vol. 95, pp. 261–279, Oct. 2018.
- [42] W. Zhang, S. Guhathakurta, J. Fang, and G. Zhang, "Exploring the impact of shared autonomous vehicles on urban parking demand: An agent-based simulation approach," *Sustain. Cities Soc.*, vol. 19, pp. 34–45, Dec. 2015.
- [43] E. F. Z. Santana, G. Covas, F. Duarte, P. Santi, C. Ratti, and F. Kon, "Transitioning to a driverless city: Evaluating a hybrid system for autonomous and non-autonomous vehicles," *Simul. Model. Pract. Theory*, vol. 107, Feb. 2021, Art. no. 102210.
- [44] R. Kala and K. Warwick, "Motion planning of autonomous vehicles in a non-autonomous vehicle environment without speed lanes," *Eng. Appl. Artif. Intell.*, vol. 26, nos. 5–6, pp. 1588–1601, May 2013.
- [45] A. Mahdavian, A. Shojaei, and A. Oloufa, "Assessing the long- and mid-term effects of connected and automated vehicles on highways' traffic flow and capacity," in *Proc. Int. Conf. Sustain. Infrastruct.*, Nov. 2019, pp. 263–273.
- [46] I. Mavromatis, A. Tassi, R. J. Piechocki, and M. Sooriyabandara, "On urban traffic flow benefits of connected and automated vehicles," in *Proc. IEEE 91st Veh. Technol. Conf.*, May 2020, pp. 1–7.
- [47] L. Ye, T. Yamamoto, and T. Morikawa, "Heterogeneous traffic flow dynamics under various penetration rates of connected and autonomous vehicle," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 555–559.
- [48] A. Schieben, M. Wilbrink, C. Kettwich, R. Madigan, T. Louw, and N. Merat, "Designing the interaction of automated vehicles with other traffic participants: Design considerations based on human needs and expectations," *Cognition, Technol. Work*, vol. 21, no. 1, pp. 69–85, Feb. 2019.
- [49] M. Schoenmakers, D. Yang, and H. Farah, "Car-following behavioural adaptation when driving next to automated vehicles on a dedicated lane on motorways: A driving simulator study in The Netherlands," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 78, pp. 119–129, Apr. 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1369847821000176>
- [50] S. Soni, N. Reddy, A. Tsapi, B. van Arem, and H. Farah, "Behavioral adaptations of human drivers interacting with automated vehicles," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 86, pp. 48–64, Apr. 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1369847822000201>

- [51] N. Reddy, S. P. Hoogendoorn, and H. Farah, "How do the recognizability and driving styles of automated vehicles affect human drivers' gap acceptance at T-intersections?" *Transp. Res. F, Traffic Psychol. Behav.*, vol. 90, pp. 451–465, Oct. 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1369847822002157>
- [52] I. Mahdinia, A. Mohammadnazar, R. Arvin, and A. J. Khattak, "Integration of automated vehicles in mixed traffic: Evaluating changes in performance of following human-driven vehicles," *Accident Anal. Prevention*, vol. 152, Mar. 2021, Art. no. 106006. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0001457521000373>
- [53] E. Leurent, "A survey of state-action representations for autonomous driving," HAL, Lyon, France, Tech. Rep., 2018.
- [54] B. Ravi Kiran, I. Sobh, V. Talpaert, P. Mannion, A. A. A. Sallab, S. Yogamani, and P. Pérez, "Deep reinforcement learning for autonomous driving: A survey," 2020, *arXiv:2002.00444*.
- [55] R. S. Sutton, D. Precup, and S. Singh, "Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning," *Artif. Intell.*, vol. 112, nos. 1–2, pp. 181–211, Aug. 1999.
- [56] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proc. Conf. Robot Learn.*, 2017, pp. 1–16.
- [57] C. Li and K. Czarnecki, "Urban driving with multi-objective deep reinforcement learning," 2018, *arXiv:1811.08586*.
- [58] S. Kardell and M. Kuosku, "Autonomous vehicle control via deep reinforcement learning," M.S. thesis, Chalmers Univ. Technol., Göteborg, Sweden, 2017.
- [59] J. Chen, B. Yuan, and M. Tomizuka, "Model-free deep reinforcement learning for urban autonomous driving," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 2765–2771.
- [60] C. Wu, A. R. Kreidieh, K. Parvate, E. Vinitsky, and A. M. Bayen, "Flow: A modular learning framework for mixed autonomy traffic," *IEEE Trans. Robot.*, vol. 38, no. 2, pp. 1270–1286, Apr. 2022.
- [61] E. Liang, "RLlib: Abstractions for distributed reinforcement learning," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 3053–3062.
- [62] G. Previati, G. Mastinu, E. Campi, M. Gobbi, L. Uccello, Á. V. Daniel, A. Albanese, A. Roccasalva, G. Santin, M. Luca, B. Lepri, L. Ferrarotti, and N. di Pietro, "Roundabout traffic: Simulation with automated vehicles, AI, 5G, edge computing and human in the loop," in *Proc. 25th Int. Conf. Adv. Vehicle Technol. (AVT)*, Aug. 2023.
- [63] G. Previati, L. Uccello, G. Mastinu, M. Gobbi, A. Albanese, A. Roccasalva, G. Santin, M. Luca, B. Lepri, and L. Ferrarotti, "Cooperative connected and automated mobility in a roundabout," SAE Tech. Paper, 2024. [Online]. Available: <https://www.sae.org/publications/technical-papers/content/2024-01-2002/>
- [64] J. Schulman, S. Levine, P. Moritz, M. I. Jordan, and P. Abbeel, "Trust region policy optimization," 2015, *arXiv:1502.05477*.
- [65] C. Yu, A. Velu, E. Vinitsky, J. Gao, Y. Wang, A. Bayen, and Y. Wu, "The surprising effectiveness of PPO in cooperative multi-agent games," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 24611–24624.
- [66] A. Pratelli, M. Lupi, C. Pratelli, and A. Farina, "Mini-roundabouts for improving urban accessibility," in *Modelling of the Interaction of the Different Vehicles and Various Transport Modes*. Springer, 2020, pp. 333–382.
- [67] O. Giuffrè, A. Granà, M. L. Tumminello, and A. Sferlazza, "Capacity-based calculation of passenger car equivalents using traffic simulation at double-lane roundabouts," *Simul. Model. Pract. Theory*, vol. 81, pp. 11–30, Feb. 2018.



LAURA FERRAROTTI received the bachelor's degree in mathematics from the University of Pisa, the master's degree (cum laude) in mathematical engineering from Politecnico di Torino, and the Ph.D. degree in computer science and system engineering from the IMT School for Advanced Studies, Lucca. She was with the Dynamical Systems, Control and Optimization Research Unit, IMT School for Advanced Studies, on algorithms' design for the synthesis of data-driven controllers

based on reinforcement learning and their adaptation to a multi-agent collaborative learning scenario. She is currently a Postdoctoral Researcher with the Mobile and Social Computing Laboratory, Fondazione Bruno Kessler. Her main research interests include cooperative AI, federated learning, reinforcement learning, data-driven control, and socially-inspired multi-agent reinforcement learning.



MASSIMILIANO LUCA is currently pursuing the joint Ph.D. degree with the Faculty of Computer Science, Free University of Bolzano, and Mobile and Social Computing Laboratory, Fondazione Bruno Kessler. During the Ph.D. degree, he had the opportunity to visit and collaborate with institutions like Massachusetts Institute of Technology (MIT), Centro Nazionale delle Ricerche (CNR), and the Institute for Cross-Disciplinary Physics and Complex Systems—Consejo Superior de Investigaciones Científicas (IFISC/CSIC). Also, he is affiliated with the Social Data Science Hub, The University of Edinburgh. Previously, he was a Research Intern with Centro Ricerche Fiat (now Stellantis) and Telefonica Research. His research interests include generalization and geographical transferability capabilities of spatio-temporal deep learning models.



GABRIELE SANTIN received the B.Sc., M.Sc., and Ph.D. degrees in applied mathematics from the University of Padova, Italy. He was a Postdoctoral Researcher with the University of Stuttgart, Germany, and a Researcher with the Mobile and Social Computing Laboratory (MobS), Digital Society Center, Bruno Kessler Foundation, Trento, Italy, and an Associate Member of the Cluster of Excellence Data-Integrated Simulation Science in Stuttgart. He is currently an Assistant Professor with the Ca' Foscari University of Venice. His research interest includes at the intersection of machine learning a computational sciences.



GIORGIO PREVIATI received the Ph.D. degree from Politecnico di Milano, in 2006. He is currently an Associate Professor with Politecnico di Milano. His research interests include machine and vehicle design, with a particular reference to vehicle and vehicle subsystems testing, structural and multidisciplinary optimization, and material modeling. He contributed to the development of a series of test rigs for the measurement of inertia properties and sensors for force and moment measurement. He was with the Research and Development Department of a leading farm tractor industry. In 2010, he was a Visiting Researcher with TU Dresden. He is the author of over 100 peer-reviewed articles, two books, and one patent.



GIANPIERO MASTINU was born in Como, in August 1960. He received the Graduate degree (Hons.) from Politecnico di Milano, in 1984, and the Ph.D. degree in applied mechanics, in 1989. He was appointed as a Full Professor with TU Delft, in 2000. He was a Full Professor with Politecnico di Milano, in 2004, where he is currently a Full Professor in road and off-road vehicles. He is also a Co-Founder of the Laboratory for Safety of Transport Systems (LaST), Politecnico di Milano, and a Coordinator of the Active Safety Division. He is also a Founder of the Driving Simulator Laboratory, Politecnico di Milano. He is also a General Secretary of the Cluster Mobility of Lombardy (Government of Lombardy Region). He is also a Coordinator of the Scientific Board of the National Mobility Cluster of Italy. He is also a member of the EAG of the FIA. He is also the Formerly Director of the Master "Vehicle Engineering" organized by University of Modena and Politecnico di Milano (sponsored by primary automotive companies from Emilia and Lombardy). He is the author of five books and of about 200 articles on international journals. His scientific activity has focused on the design, construction, test of machines, structures, and mechanical systems, with a particular reference to ground vehicles. Topics: conceptual, functional and structural design (optimization methods), experimental mechanics, modeling and simulation of machines, and road and rail vehicle system engineering. He was the President of the XIX IAVSD Symposium in Milan, in August 2005.



MASSIMILIANO GOBBI received the master's degree (Hons.) in mechanical engineering, in 1994, and the Ph.D. degree in applied mechanics, in 1997. In 1998, he was a Visiting Scholar with the Department of Mechanical Engineering, University of California at Berkeley, Berkeley, CA, USA. He is currently a Full Professor with Politecnico di Milano, Milan, Italy. He has authored more than 200 articles. He is the author of seven international patents and three books.



PRAVEEN ZALAYA received the Bachelor of Technology degree in mechanical engineering from Indian Institute of Technology (IIT), in 2019, and the M.Sc. degree in autonomous and connected vehicle engineering from Politecnico di Torino, in 2022. He is currently with Centro Ricerche Fiat (now Stellantis), Turin, Italy, as a Pre-Development Engineer. He is also in the connected vehicle area, driving innovation for various cutting-edge connectivity technologies, such as 5G, C-V2X, and satellite communication. He is concentrating his efforts on developing multiple UC's for V2X applications in a simulated environment, merging the ADAS system with enhanced perception capabilities from the V2X systems. He also works on application SW development for complete E2E prototypes of various connected car services and safety-related tasks.



ELENA CAMPI was born in Genoa, in 1996. She received the B.Sc. degree in mechanical engineering and the M.Sc. degree in ground vehicles from Politecnico di Milano, in 2018 and 2021, respectively. At the moment, she is currently a Fellowship Researcher with DriSMi, Driving Simulator Laboratory, Politecnico di Milano. Her current research interests include traffic simulations with human-in-the-loop and on control algorithms of autonomous and connected vehicles.



ALESSANDRO ROCCASALVA received the Graduate degree in mechanical engineering. He is currently with Centro Ricerche Fiat (now Stellantis), Turin, Italy, as a Testing and Validation Engineer. He has a background in diesel and gasoline testing activities. He performed both development and reliability testing activities (such as engine oil aeration assessment, oil pan level and blow-by circuit assessment @ tilting bench, max power, and piston lubrication assessment). At the moment, he is also with the Advanced Connectivity Team and he is involved in the European project "AI@EDGE" to follow the telematics box module integration with the driving simulator with DriSMi.



LORENZO UCCELLO received the M.Sc. degree in mechanical engineering, ground vehicles, in 2023. He is currently a Researcher with Politecnico di Milano. He is also with DriSMi, the dynamic driving simulator of Politecnico di Milano. His research interests include cooperative and automated driving, spanning diverse areas, ranging from communication protocols and simulation to the implementation, and validation of digital twins. He is exploring the effect due

to the human factor, proactively modifying the simulation environment's performance, and leveraging state-of-the-art sensors like EEG, ECG, and SPR with the purpose of analyzing human-machine interaction.



BRUNO LEPRI received the Ph.D. degree in computer science from the University of Trento. He is currently a Senior Researcher with Fondazione Bruno Kessler, Trento, Italy, where he leads the Mobile and Social Computing Laboratory (MobS). Recently, he has launched the Center for Computational Social Science and Human Dynamics, a joint initiative between Fondazione Bruno Kessler and the University of Trento. Since July 2022, he has been the Chief Scientific Officer of Ipazia, a new company active on AI solutions for financial services and energy management. From 2019 to 2022, he was also the Chief AI Scientist of ManpowerGroup, where he has collaborated with the global innovation team on AI projects for recruitment and HR management. He is also a Senior Researcher with Data-Pop Alliance, the first think-tank on big data and development co-created by the Harvard Humanitarian Initiative, MIT Media Laboratory, Overseas Development Institute, and Flowminder. Finally, he has co-founded Profilio, a startup on AI-driven psychometric analysis. In 2010, he won a Marie Curie Cofund Postdoctoral Fellowship and he has held a three year postdoctoral research position with the MIT Media Laboratory. His research interests include computational social science, personality computing, network science, and machine learning. His research has received attention from several international press outlets and obtained the ten year impact award at MUM 2021, the James Chen Annual Award for the Best 2016 UMUI Paper, and the Best Paper Award at ACM Ubicomp 2014.



ANTONINO ALBANESE is currently the Research and Innovation Manager of Italtel S.p.A., Milan, Italy. He has a long experience in the design, development and testing of telecommunication products related to fixed, mobile, multiservice networks (voice/data/video), distributed on cloud, and virtualized infrastructures. He has management expertise in European collaborative research and innovation projects, being actively involved in projects related to several ICT areas,

including telecommunications, cloud, network function virtualization, service defined networking, the IoT, edge computing, AI, and cybersecurity. His main research interests include high performance computing and HW acceleration technologies that can provide a performance boost in data processing at the edge cloud, improve energy, and cost efficiency.

...