# SavviDriver: Model-based Framework for Game-based Testing of Autonomous Vehicles in Diverse Multi-agent Traffic Scenarios

Kenneth H. Chan[†], Sol Zilberman[†], Betty H.C. Cheng[†]

Department of Computer Science and Engineering, Michigan State University, 428 S Shaw Ln, East Lansing, 48824, MI, USA.

*Corresponding author(s). E-mail(s): chengb@msu.edu;
Contributing authors: chanken1@msu.edu; zilberm4@msu.edu;
[†]These authors contributed equally to this work.

**Abstract**

Autonomous Vehicles (AVs) must operate safely in the face of uncertainty, including those induced by human behaviors (i.e., external human drivers). Specifically, AVs must exhibit safe responses when encountering previously unseen behaviors from human drivers with different driving styles. For example, aggressive drivers may cut off other vehicles to merge into a lane, or distracted drivers may fail to respond to changing road conditions. A key challenge is how to assess the onboard AV decision-making capabilities to detect and mitigate those potentially unsafe scenarios due to one or more external human-operated vehicles. We observe that AVs and other vehicles on the roadway may share common functional objectives (e.g., to navigate to a given target destination), but otherwise may be motivated by different non-functional objectives, such as safety, minimizing transport time, minimizing fuel consumption, etc. This paper introduces a modular and composable model- and game-based testing framework to enable an AV developer to operationally assess the robustness of an AV in response to human-based uncertainty. Specifically, this work uses goal models to declaratively specify functional and non-functional objectives of vehicles (both the AV under study and those representing external human-operated vehicles) to inform the game-based testing environment that incorporates real-world traffic infrastructure data. We demonstrate the model-based capabilities of our game-based testing approach on a number of scenarios based on real-world traffic accident data involving human drivers.

## 1 Introduction

Increasingly, Autonomous Vehicles (AVs) are being deployed alongside human-driven vehicles (i.e., mixed traffic environments [1]), where they must safely interact with other road users with different behaviors, objectives, and driving styles. To ensure operational correctness and prevent failures, developers must understand how AVs respond to different sources of uncertainty, including the potentially unpredictable behaviors of human drivers. Uncertainty arises when a system encounters an event it cannot handle, often due to incomplete information (i.e., epistemic doubt [2]) and/or unpredictable phenomena in its operating environment (i.e., aleatoric uncertainty [3, 4]). During design-time testing and run-time decision-making, *human-based uncertainty* for AVs in mixed-traffic environments results from epistemic doubt about other drivers' intentions, such as motivations, behaviors, and future decisions. However, significant safety concerns and high resource costs limit "in the field" testing of safety-critical systems in real-world settings [5, 6], where human-based uncertainty poses significant challenges [7, 8, 9]. Thus, there is growing interest in the use of human-like driver models and simulations to test AVs in representative operating contexts, including elements that pose uncertainty, such as mixed traffic interactions [10]. This paper proposes a modular, goal model-based framework that harnesses the non-cooperative gaming paradigm [11] to support online testing and exploration of the behavior of interacting agent vehicles, including AVs and human-operated vehicles, in order to assess the robustness of AVs in response to human-based uncertainty.

Recently, several state-of-the-art *online learning-based testing* approaches [7, 12, 13, 14, 15] have assessed AV responses to uncertainty posed by interactions with other vehicles. *Adversarial* online learning-based testing approaches leverage knowledge about the objectives of the AV to discover operating contexts that cause explicit AV failures (e.g., collision) [12, 14]. Other online learning-based testing approaches involve game-based testing, where game-theoretic formulations of traffic scenarios are used to assess AV behaviors. *Cooperative* game-based testing approaches have been explored to assess the ability of the AV under study to coordinate with other vehicles, where traffic participants share common objectives and exhibit mutually-beneficial behavior(s) [13]. However, these approaches do not reflect the state of practice AV deployment, as vehicles on the road generally do not share goals nor collaborate with each other (due to constraints on human and machine communications). As a means to capture an AV's environment when deployed in real-world settings, we have previously proposed *non-cooperative* game-based testing to explore interactions between vehicles that act independently and only consider their own selfish objectives [7]. Specifically, these different vehicle interactions become a source of uncertainty that can be used to assess the robustness of the AV under study. Reinforcement Learning (RL) [16] has been used in online learning-based testing to operationally implement various types of external vehicle behavior [7, 17, 18] (i.e., RL is used to approximate optimal strategies for players in a traffic-based game, which can then be used as external vehicle controllers whose behavior can be used to "test" the AV system under study [19]). We observe, however, that game-based testing techniques commonly rely on trial-and-error approaches to determine game objectives (i.e., specification of an RL reward function) and are often tightly coupled to specific traffic scenarios. Additionally, these techniques lack modularity and require low-level manual code changes and/or brute-force modifications to explore any changes to operational contexts, including variations on the number and types of drivers, driving strategies, road scene changes, etc. [12, 20]. As such, existing state-of-the-art game-based testing approaches lack systematic or application-level-based support for reconfiguring specifications of vehicles (e.g., speed, driving properties), roadway infrastructure (e.g., highway, merge lanes, intersections) to explore multiple operating contexts to assess AV robustness in the face of human-based uncertainty.

This work introduces SavviDriver (**S**afe **A**utonomous **V**ehicle-to-Human-Controlled **V**ehicle **I**nteractions), a modular and composable goal model-based framework for game-based testing of AVs in diverse multi-agent traffic scenarios. Several key insights are foundational to our approach. First, goal models can be used to declaratively specify and manage game-based testing of multi-agent interactions with respect to functional and non-functional objectives. Second, by refining goal models down to the level of individual requirements that can be assessed for satisfaction/satisficement in terms of utility functions [21], we can use the utility functions to specify the reward structure for an RL-based implementation of a game-based testing framework. Finally, by taking a "separation of concerns" approach [22, 23] to goal modeling, we can separate the context-dependent functional goals from context-independent non-functional goals. This strategy allows developers to *decompose* top-level goals into a library of reusable components (i.e., subtrees) corresponding to different driving styles (e.g., aggressive driving non-functional subtree). Then a developer can *compose* different combinations and structures of subtrees from the library to facilitate the rapid reconfiguration of multi-agent game-based testing with heterogeneous/homogeneous agents.

SavviDriver supports the assessment of AV robustness in response to uncertainty posed by one or more human-based agents through game-based testing, where the discovered uncertainty can then be used to improve AV robustness (e.g., revise requirements, update vehicle behaviors, etc.). Rather than explicitly modeling uncertainty, the proposed non-cooperative game theory approach organically discovers/addresses two sources of uncertainty based on their impact on the observable behavior of the AV: (1) the uncertainty posed by previously unseen external vehicle behaviors, and (2) the uncertainty posed by the unknown/unexpected responses of the AV under study when interacting with the external vehicles in various operating contexts. Consider a developer who is interested in exploring the interaction(s) between the AV under development and different types of external human driver models to discover unexpected outcomes. First, a developer defines the participating game actors (road users such as AV, human-driven vehicles, etc.) and the environmental context (e.g., merging traffic scenario) in which they interact. A developer may declaratively specify different types of human-based driving styles (e.g., aggressive, distracted, timid, etc.) based on real-world traffic data [24, 25, 26], as well as intended AV behavior(s) (e.g., safety, comfort, etc.). Next, SavviDriver uses a variation of the KAOS goal modeling language [27] to explicitly define the functional and non-functional objectives of the actors in the game. SavviDriver supports the reuse of behavior-based templates for goal model subtrees corresponding to different human-based driving styles (e.g., an 'aggressive' driving subtree can be used in different mixed-traffic scenarios). We associate utility functions [28] with leaf-level goals to assess their satisfaction [29]. The utility functions can be used to inform the objective function for the corresponding agent in multi-agent games. As the agents for the AV and human-operated vehicles pursue their respective goals while operating in the traffic scenarios under study (e.g., lane merging), unexpected and unsafe behavior may be exhibited by one or more of the agents due to their unanticipated interaction(s). The discovered behavior of both AV and human-operated vehicles can be used to assess and potentially improve the robustness of the AV and inform changes to goal models to explore additional mixed-traffic interactions, rather than editing low-level simulation code.

To demonstrate how SavviDriver can be used to (1) discover unexpected interactions between AVs and external agents in the environment and (2) improve the robustness of the AV with respect to human-based uncertainty, we have applied it to several use cases that capture real-world traffic accident data [25, 30]. The remainder of this paper is organized as follows. Section 2 provides background information and overviews enabling technologies. Section 3 describes the SavviDriver framework, including its aggregate elements and their application to a running example. Section 4 presents the results of our illustrative use cases. Section 5 discusses our results. Section 6 overviews related work. Section 7 considers the threats to validity. Finally, Section 8 concludes this paper and discusses future work.

## 2 Background

This section describes background topics and enabling technologies used in SavviDriver. First, we provide an overview of existing assurance challenges for AVs and simulation-based testing. Next, we discuss existing state-of-the-art approaches involving game theory and RL to discover uncertainty for AVs. Finally, we describe the goal modeling language and utility functions used in this work.

### 2.1 Challenges for Autonomous Vehicles

Advances in machine learning have improved the capabilities of AVs. Recently, a number of different companies have deployed a range of AVs in real-world settings, including Tesla's Autopilot [31], Waymo's Taxi services [32], and ADASTEC's autonomous bus services [33]. To avoid accidents and protect their occupants, deployed AVs must demonstrate safe behavior in addition to satisfying operational constraints. Based on Waymo's 2023 safety report, AVs have the potential to reduce the frequency and severity of traffic accidents in limited operating contexts [34]. However, various uncertainty factors, including those introduced by machine learning and humans, have been shown to cause unexpected behaviors in AVs [35]. For example, human drivers exhibit a range of driving styles, including aggressive, distracted, or even malicious behaviors (e.g., brake-checking, tailgating, etc.) [36]. Vehicles with these driving styles have been associated with unsafe driving maneuvers on the road, such as aggressively cutting off another vehicle or drifting out of a lane due to distractions [37, 38]. In order to ensure AVs are capable of reacting safely when encountering these unexpected maneuvers, AVs must be exposed to a wide range of human-based behaviors during training and testing [35].

## 2.2 Simulation-based Testing

Simulation-based testing of AVs is a type of *online testing* where the system under study is embedded and executed in a specific operating context [6]. In simulation-based testing, the operating context is represented by a virtual environment where a simulation engine manages corresponding physics, graphics, and interactions. High-fidelity simulation platforms (e.g., CARLA [39], BeamNG [40]) focus on realistic rendering and are best suited for evaluating sensor suites for AVs (e.g., computer vision models, LIDAR sensors, etc.). However, high-fidelity simulators are computationally expensive and may not scale to tasks where a large number of executions are required (e.g., RL training, search-based testing). Lightweight simulators (e.g., BARK [41], HighwayEnv [42], and our in-house simulator TinyRoad [7]) are better suited for evaluating high-level behaviors of AVs by replacing expensive realistic rendering tasks with simplified 2D models. Importantly, both types of simulation platforms implement physics engines such that vehicle actuator inputs accurately update the state of the vehicle and the operating context.

As this work focuses on discovering and evaluating high-level behavioral characteristics of vehicles, we use our in-house TinyRoad simulation platform. TinyRoad uses a standard 2D kinematic bicycle model for vehicle physics. Vehicles are represented by 2D geometry corresponding to an input vehicle type and dimensions. The roadway is represented by a graphical map of the environment, where colored lines denote roadway elements (e.g., black lines for road boundaries, yellow lines for lane boundaries, etc.). Each vehicle is associated with a controller that takes as input the state of the environment (i.e., observation $o_t$) and outputs a tuple for driving actions (i.e., action $a_t$ representing throttle and steering values), at time step $t$. A monitoring module records the state of the environment and vehicles at each time step $t$, where the monitored properties are used for reward calculation and collision detection (e.g., distance between vehicle and other vehicle/road boundary within some threshold).

## 2.3 Game Theory for Uncertainty Exploration

In order to ensure safe behaviors, a developer may explore different techniques to test the AV before deployment. For example, a number of existing researchers have proposed the use of assurance cases and argument structures to prove the safety of AVs [43, 44, 45]. While these works provide a systematic and traceable structure to demonstrate and verify to stakeholders that the system under development is acceptably safe or satisfies some safety constraints, they do not address/propose testing techniques to discover the limitations of the system. Other existing works have explored human-based testing in a simulation environment [46, 47, 48]. However, human-based testing faces the challenges of testing bias and incurs expensive development time and effort [49]. One promising approach to model human behavior is through game theory that enables the modeling and analysis of strategic interactions between *rational decision-makers* [50]. In game theory, a game involves a number of agents that interact with each other in an environment. Agents seek to achieve individual and/or shared objective(s). In *cooperative* game theory, agents can collaborate to maximize collective rewards [50]. However, as AVs are deployed in real-world traffic and interact with traditional human-operated vehicles, it is not feasible to assume collaboration. *Non-cooperative* game theory [11] better captures the relationship between road users than cooperative game theory, as drivers are mainly concerned with their individual objectives and do not (typically) maliciously interact with other road users [7]. Recently, several state-of-the-art gaming approaches have used *RL* to operationalize non-cooperative games between vehicles in traffic simulations to test AVs before deployment [7, 12, 14]. Chan *et al.* [7] proposed the combination of non-cooperative game theory and RL to discover previously unseen or undesirable behaviors for AVs in two-player games. Wachi *et al.* [14] and Hao *et al.* [12] demonstrate different ways that multi-agent games and adversarial RL [51] can be used to discover failure cases for rule-based vehicles in simulation. However, existing game-based testing techniques largely use ad hoc and/or hard-coded techniques to generate traffic scenarios, driving styles, and RL agents, where any changes (e.g., different roadways, the number/types of drivers, or reward structures) may require significant low-level development effort.

## 2.4 Reinforcement Learning

RL is a learning-based approach, where agents learn to perform actions in an environment that maximizes a reward function. Figure 1 shows a high-level overview of RL, where an agent performs some action $a_t \in \mathcal{A}$ affecting the environment at each timestep $t$. The updated state of the observable environment $o_t \in \mathcal{O}$ and the reward associated with the state are then returned to the agent. The reward function informs the behavior(s) of the agent, motivating it to achieve a given task. For example, an agent whose objective is to complete a racing game may have a reward function that motivates it to complete the race as fast as possible, stay on the track, and avoid collisions [52]. Specifically, an agent in RL learns a *policy* $\mathcal{F}_\theta : \mathcal{E} \rightarrow \mathcal{A}$ that represents a mapping between states of the environment and *optimal* actions in each of those states.[1] Learning a policy involves a “trial-and-error” process, where agents make iterative improvements to the current policy by interacting with an environment over a given number of trials. During each trial (game), an agent may discover one or more actions that may improve the overall reward. An optimization algorithm informs the updates to the policy after each trial. Recent advances in Deep Reinforcement Learning (DRL) have demonstrated that Deep Neural Networks (DNNs) can be used to approximate the optimal policy for a given agent [53]. DRL is often used for complex tasks such as driving that may require agents to optimize high-dimensional non-linear objective functions in dynamic environments [54].

In the context of RL-based behavior generation, a challenge is how to design a reward function that accurately rewards and motivates a desired behavior. Existing approaches often use ad-hoc development approaches to design those rewards, resulting in reward functions that are often difficult to manage or interpret and may be

[1]We use the variable $\mathcal{F}$ to represent a policy (as opposed to the commonly used variable $\pi$ in RL literature) to avoid confusion with the payoff function $\pi$ in game theory.

overly complex [55]. In addition, traditional approaches require "trial-and-error" low-level code changes to explore how different reward configurations result in different behaviors. In contrast, SavviDriver provides a systematic means to structure the reward function with support for updates via high-level goal model changes.

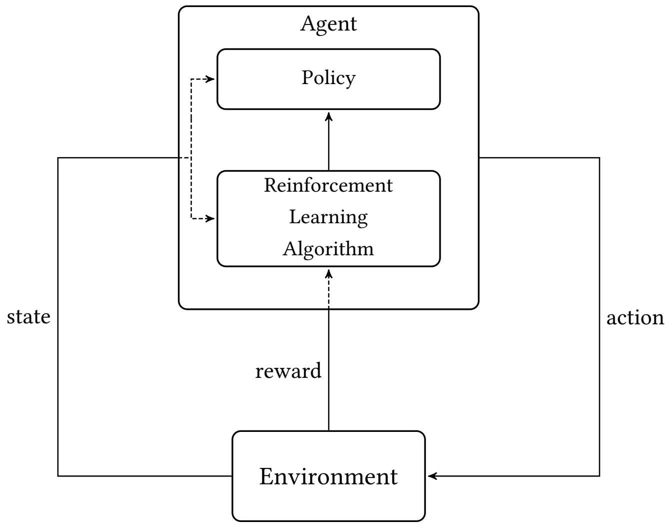


**Fig. 1**: A high-level overview of RL [56, 57].

### 2.4.1 Proximal Policy Optimization

In order to support the discovery of driving strategies that can be implemented by external vehicle controllers, SavviDriver uses the Proximal Policy Optimization (PPO) DRL technique, a well-established learning approach for control problems [16]. PPO is a first-order policy gradient method that learns the parameters $\theta$ of a stochastic policy function $\mathcal{F}_\theta$. The goal of the PPO training is to learn a policy $\mathcal{F}_\theta$ that maximizes the probability of selecting actions that increase cumulative rewards. PPO has been shown to achieve state-of-the-art performance in continuous high-dimensional action spaces and non-stationary environments while maintaining high data-efficiency, a critical property for driving control tasks [16, 19]. Specifically, our work uses an *actor-critic-style* clipped implementation of the PPO algorithm that uses two neural networks working in tandem to approximate an optimal policy [58]. An actor-critic architecture supports stable training by combining the smooth convergence properties of policy gradient methods (actor) while reducing the variance that may result from updating parameters based on diverse training episodes (critic). For an input state $o_t$, the *actor* network approximates the policy $\mathcal{F}_\theta(a_t|o_t)$ (e.g., navigating to a destination using driving actions such as acceleration and braking), while the *critic* network approximates the expected reward value $V(o_t)$. An example of a reward value is the minimal time to reach the destination. Parameter updates at each training step are *clipped* to the range $[1-\epsilon, 1+\epsilon]$ to promote better stability, where $\epsilon$ is a developer-provided hyperparameter [16]. A conservative clipping value incentivizes policy updates to stay within the developer-specified range, preventing destructive updates that may undo previously learned behavior, cause training instability, or even lead to policy collapse [59]. During training, the trainable parameters $\theta$ are updated using stochastic gradient ascent to maximize an approximation of the expected cumulative reward (Expression 1 captures the objective formally),

$$L^{CLIP+VP+S}(\theta) = \hat{\mathbb{E}}[L_t^{CLIP}(\theta) - c_1 L_T^{VF}(\theta) + c_2 S[\mathcal{F}_\theta|(o_t)]] \tag{1}$$

where $L_t^{CLIP}$ is the clipped surrogate objective, $L^{VF}$ is the squared error of state values, and $S[\mathcal{F}_\theta|(o_t)]$ is the entropy bonus [16]. The entropy bonus ensures sufficient exploration by applying perturbations to the optimization objective, encouraging more diverse actions to be selected during training [60]. Additionally, the entropy bonus has been shown to be helpful for complex tasks and may prevent early convergence to suboptimal policies [61]. Finally, PPO's value function *VF* includes a *discount factor* hyperparameter $\gamma$. The discount factor $\gamma$ enables developers to control exploration by adjusting the proportionate weight of rewards with respect to a time horizon [16], such that earlier rewards are prioritized over later ones. A high discount factor (e.g., 0.99) encourages exploration by assigning a relatively low discount for future actions, thereby incentivizing the agent to explore states near the time horizon [62]. For example, an AV learning to merge onto a highway may find that reducing its speed to zero until the end of an episode allows it to avoid penalties from future collisions. However, this behavior prevents the successful completion of the objective (i.e., navigating to a destination) and is therefore suboptimal. By applying perturbations to the reward signal, the vehicle has a greater chance of exiting the suboptimal stable state by attempting new behaviors (e.g., merge maneuver).

## 2.5 Goal-Oriented Requirements Engineering

Van Lamsweerde *et al.* introduced KAOS, a goal-oriented approach to requirements engineering, where a high-level goal or objective is decomposed into subgoals. Each subgoal is then recursively decomposed until a low-level leaf goal is reached [27]. At the leaf level, KAOS goals are considered requirements that are discharged to *system components* for satisfaction or satisficement (i.e., degree of satisfaction) [27].[2] While KAOS does not distinguish functional and non-functional goals explicitly, subgoal decomposition strategies can be used to specify functional goals that reflect non-functional properties. Effectively, our KAOS goal model captures the non-functional properties of driving styles in terms of the functional goals/subgoals. For example, the non-functional subgoal, '`drive safely`', can be refined down to requirements: '`driving at the speed limit`' and '`maintaining a minimal trailing distance`'. Those requirements can then be handled by the system components '`speed controller`' and '`radar`', respectively.

[2]We use the term "system components" instead of "agents" in the KAOS goal model to avoid confusion with the use of the term "agent" for AV and non-ego vehicles.

## 2.6 Utility Functions

Utility functions can be used to map system and environmental attributes to quantitative values that establish a degree of system goal satisfaction [29, 63, 64, 65]. Expression (2) shows the general structure of a utility function, where $u$ represents a scalar utility value between $[0, 1]$, and $v$ represents a measured property obtained from the system and its environment (e.g., vehicle speed, distance to target destination) [29].

$$u = f(v) \tag{2}$$

Previously, it has been shown that *utility functions* can be used to annotate KAOS leaf-level goals to measure their satisficement [29, 64, 66]. Specifically, leaf-level goals are annotated with utility functions that specify how the corresponding system component can be used to measure a quantifiable attribute of the leaf-level goal. For example, the requirement '`driving at the speed limit`' can be quantified by the utility function shown in Expression (3), where $s$ corresponds to the speed limit and $v$ corresponds to the agent's current velocity. The maximum value (i.e., *utility*) occurs when the agent is driving exactly at the speed limit (i.e., $s - x = 0$).

$$f(x) = \begin{cases} 0 & \text{if } v < .95s, v > 1.05s \\ |\frac{s-v}{0.05s}| & \text{if } .95s \leq v \leq 1.05s \end{cases} \tag{3}$$

# 3 Methodology

This section describes the SavviDriver framework, including details of its modeling and analysis processes. Table 1 provides a high-level glossary of the terms and their definitions used in this work. Figure 2 shows an overview of the SavviDriver framework, where circles depict processes, parallel lines depict persistent data stores, labeled arrows depict data flow between entities, and rectangles depict external entities. The SavviDriver framework takes as input real-world traffic data and the Ego under study. First, SavviDriver facilitates the development of goal models for agent objectives that reflect real-world driver behaviors. Next, SavviDriver configures the reward functions and simulated traffic scenario corresponding to the goal models. Then SavviDriver's game-based testing process automatically trains RL-based Non-ego Vehicle (NeV) controllers that exhibit behaviors specified by their respective goal models, where NeV behaviors pose a source of uncertainty for the input Ego under study. Finally, SavviDriver supports the assessment of Ego robustness with respect to the discovered human-based uncertainty exhibited by RL-based NeV agents, which can be used to improve Ego robustness. We next describe the elements of SavviDriver in detail.

### Data

Historical data, documentation, and domain expert input regarding driving data and traffic accidents are all used by the developer to identify and specify different types of driver intentions. For example, we use data collected by key federal US-based agencies to identify driver behaviors that frequently resulted in undesirable traffic outcomes (e.g., United States National Highway Traffic Safety Administration (NHTSA) [24, 25, 67], United States Department of Transportation [26], AAA Foundation for Traffic Safety [68, 69]). These agencies are responsible for investigating traffic incidents and publishing data reports that include the types of maneuvers leading up to accidents, the statistics about characteristics and frequencies of accidents, and the details for the road environments where the accidents occurred. Developers and/or domain experts may use the traffic data to identify common high-level driver behaviors to further explore with SavviDriver at design-time. As such, SavviDriver can be used to explore human-based uncertainty found in real-world traffic scenarios [24, 25, 26], as well as other possible detrimental scenarios. SavviDriver generates RL driving models that are informed by real-world high-level human behavior characteristics that have been documented as contributing factors to traffic accidents [24, 25, 26], often due to the unpredictability/lack of knowledge about the driver intentions/maneuvers (e.g., an aggressive driver may quickly decide to cut off a vehicle without any indication or safety precautions). Specifically, SavviDriver automatically explores manifestations of human-based uncertainty by executing the RL driving models in simulation and analyzing their dynamic interactions with the Ego under study to uncover uncertainty posed by unexpected NeV maneuvers and corresponding undesirable Ego responses. Those unexpected maneuvers/interactions constitute a form of *epistemic uncertainty* as an Ego may not be able to predict and handle interactions with the drivers exhibiting those maneuvers.

During pre-processing, domain experts identify through manual or automated techniques (e.g., accident analysis procedures [26]) three key types of information, in addition to the AV software under study:

- **Scenario Definition**: Concrete parameters of the identified traffic scenario that describe the structural attributes (e.g., lane geometry, topography, etc.), regulatory attributes (e.g., speed limit, direction of the lane) of the traffic infrastructure, and the participating vehicles (e.g., type, initial position, orientation, and velocity of each vehicle).
- **Scenario Objective:** The functional objective of each agent participating in a given scenario, as identified by domain experts. For a given scenario, each agent's scenario objective may capture scenario-specific maneuvers an agent should perform (e.g., merge into right lane), with respect to their initial state in the traffic infrastructure.
- **Driver Types**: The driving style (e.g., aggressive, cautious, etc.) and observed behaviors (e.g., speeding, swerving, etc.) of each agent.
- **Black-box Ego Configuration**: The AV software under study that maps environment inputs to actions. For example, the AV configuration can specify a traditional software-based ADAS system, an RL-based driving model, or any other simulated AV software system.

Running Example: We illustrate the SavviDriver framework using a running example of a merge scenario, motivated by NHTSA's 2022 accident reports [24, 70] that

**Table 1**: Overview of the terminologies used in this paper.

| **Term** | **Definition** |
|---|---|
| ITS | An Intelligent Transportation System (ITS) comprising one or more intelligent vehicle(s) (e.g., AVs) as well as one or more human-operated vehicle(s). |
| Agent | Entity with a concrete role (e.g., vehicle) that has goals and is of interest in the ITS under study.[†] |
| AV | An autonomous vehicle whose behavior is optimized by one or more machine learning component(s). |
| Ego | The autonomous vehicle under study. |
| NeV | Human-operated vehicle that is part of the operational context for the AV in the ITS. |
| Functional goal | Goals that describe functions that the system should perform (e.g., arrive at destination). |
| Non-functional goal | Goals that describe desired properties (e.g., defensive driving) of *how* a system satisfies its functional goals. |
| Scenario Definition | Concrete parameters of the identified traffic scenario, including infrastructure data and initial agent states. |
| Scenario Objective | The individual functional objectives of each agent in a given scenario. |
| Driver Type | The driving style (e.g., aggressive, cautious, etc.) and observed behaviors (e.g., speeding, swerving, etc.) of each agent. |

[†] The term *agent* is common to the disparate techniques used in this work (e.g., goal modeling, reinforcement learning, game theory). We use the definition presented in this table whenever the term *agent* is used.

list lane merging as a frequent cause of accidents. Figure 3 provides a 2D illustration of the Intelligent Transportation System (ITS) under study. The scenario under study is a two-lane road that converges into a single lane. The participating agents include the blue Ego (the default AV under study) driving in the right lane and an orange aggressive NeV that must merge into the Ego's lane. During game-based testing, both vehicles are instantiated in a simulated traffic scenario. During each timestep of the simulation, each vehicle takes as input the current state of the environment and uses its respective controller to perform a driving action. The blue Ego uses the default AV software under study to take actions, while the orange NeV uses the SavviDriver-generated RL-based policy to take actions, thereby exhibiting behaviors that may exploit safety properties of the default blue Ego. Therefore, each game-based testing scenario comprises the SavviDriver-trained NeV controller, a corresponding traffic scenario, and the blue Ego software under study. We view uncertainty from the perspective of an AV developer. In our running example, the blue Ego may be uncertain about the intentions and future behaviors that may be exhibited by the orange NeV as

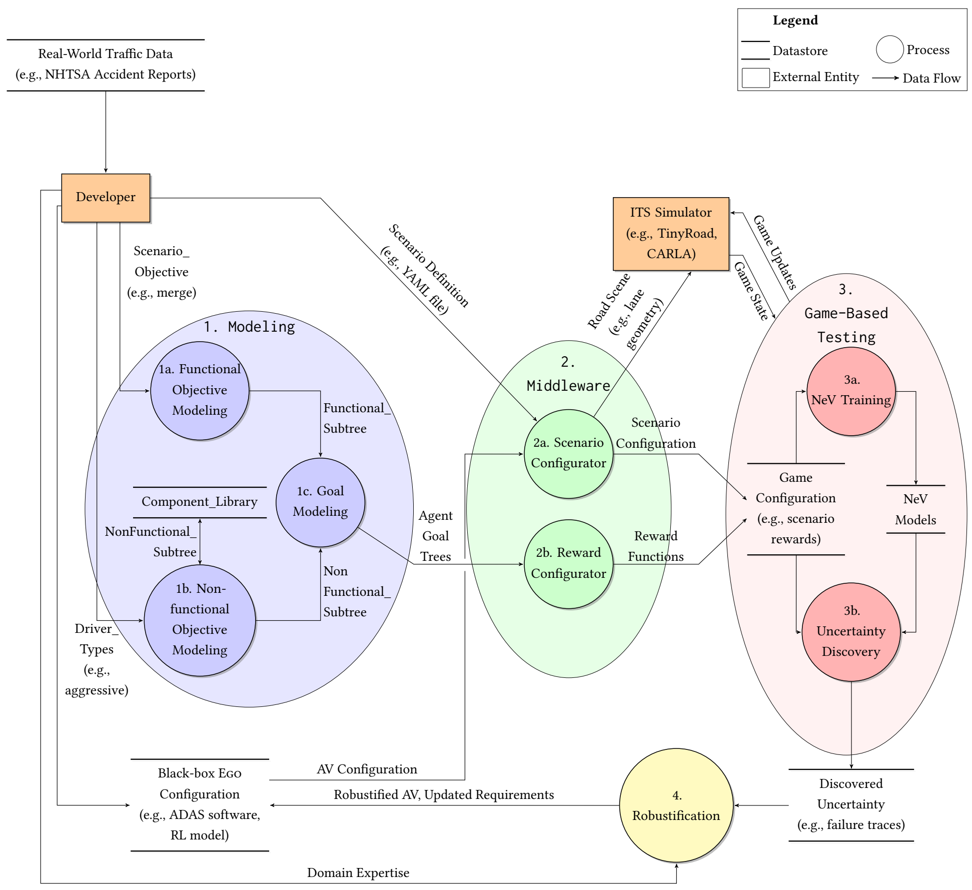

Legend
Datastore
Process
External Entity
Data Flow
Real-World Traffic Data (e.g., NHTSA Accident Reports)
Developer
Scenario_ Objective (e.g., merge)
Scenario Definition (e.g., YAML file)
ITS Simulator (e.g., TinyRoad, CARLA)
Game Updates
Game State
Road Scene (e.g., lane geometry)
1. Modeling
1a. Functional Objective Modeling
Functional_ Subtree
1c. Goal Modeling
Component_Library
NonFunctional_ Subtree
1b. Non-functional Objective Modeling
Non Functional_ Subtree
Driver_ Types (e.g., aggressive)
Agent Goal Trees
2. Middleware
2a. Scenario Configurator
2b. Reward Configurator
Scenario Configuration
Reward Functions
3. Game-Based Testing
3a. NeV Training
Game Configuration (e.g., scenario rewards)
NeV Models
3b. Uncertainty Discovery
Discovered Uncertainty (e.g., failure traces)
4. Robustification
Black-box Ego Configuration (e.g., ADAS software, RL model)
AV Configuration
Robustified AV, Updated Requirements
Domain Expertise

**Fig. 2**: Overview of SavviDriver. Purple processes represent manual modeling steps done by a `Developer`. Green processes represent an automated middleware compilation of `Developer` information. Red processes represent game-based testing components. Finally, the yellow process represents the AV robustification step of the framework, which can incorporate domain knowledge from the `Developer`.

the orange NeV pursues its selfish human-based objectives. Subsequently, the developer may be uncertain about how the blue Ego may respond to unexpected behaviors exhibited by the orange NeV. To this end, the objective of SavviDriver is to reduce the epistemic uncertainty that is associated with unexpected behaviors of the orange NeV by generating those behaviors during the testing phase, as a means to assess the blue Ego's responses in the presence of those behaviors, and then robustifying the blue Ego to the discovered uncertainty.

## 3.1 Step 1. Modeling

This step describes how agents are modeled by the `Developer` in SavviDriver; this process is used to individually develop respective goal models to describe Ego and NeV(s) behaviors. Modeling Process 1 overviews the `Developer`'s process for goal modeling promoted by SavviDriver. We use a variation of the KAOS goal modeling language to elaborate the top-level functional objective of each agent (e.g., '`arrive at destination`'). A `Developer` uses KAOS to create an AND/OR goal model to declaratively specify at a high-level *what* an agent's objectives are and the driving style (i.e., non-functional requirements) they should exhibit when satisfying those objectives. We next describe the elements of the goal model-based decomposition strategy, which is used to model both the behavior of the Ego and NeV(s).

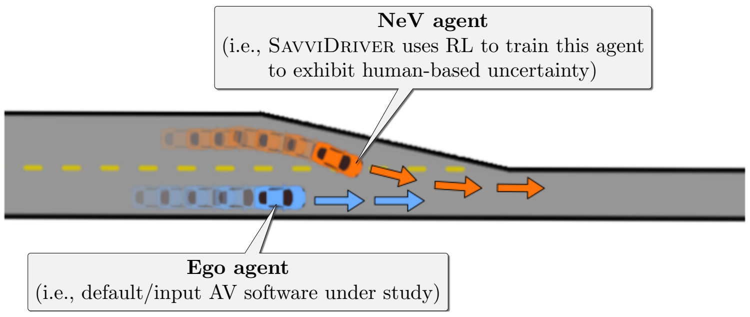

**NeV agent**
(i.e., SavviDriver uses RL to train this agent to exhibit human-based uncertainty)
**Ego agent**
(i.e., default/input AV software under study)

**Fig. 3**: Merge scenario sample showing the Ego (blue) and NeV (orange) in the TinyRoads simulator [7].

For each agent, the `Developer` refines and decomposes the agent's top-level objective into two distinct types of subtrees: functional subtrees and non-functional subtrees. Figure 4 shows an abstract overview of goal decomposition in SavviDriver, where the top-level goal is shown in orange, *KAOS Goals* are depicted by parallelograms, and goal decomposition is represented by refinement arrows. The blue subtree in Figure 4 shows **Step 1a** of SavviDriver, where the `Developer` defines the *context-dependent functional objectives* that correspond to a `Scenario_Objective`. For example, a vehicle may have a `Scenario_Objective` of merging into the right lane of the roadway if its initial position is in a merging lane (see orange vehicle, Figure 3). The green subtree in Figure 4 shows **Step 1b** of SavviDriver, where the `Developer` defines the *context-independent non-functional objectives* corresponding to a `Driver_Type` that captures the driving style(s) a vehicle may exhibit while completing its functional objectives. For example, a vehicle may exhibit aggressive or distracted driving styles while attempting to merge into the right lane of the roadway. Finally, in **Step 1c**, the `Developer` composes the two subtrees from **Step 1a** and **Step 1b** to form the overall KAOS goal tree for the agent.

SavviDriver's goal modeling approach facilitates and promotes modularity and reusability of agent objectives. Specifically, high-level agent objects can be decomposed into a library of reusable *components* corresponding to different driving styles, represented by context-independent subtrees. Figure 5 shows how those components can be reused and composed to explore different mixed-traffic scenarios. The first row (I.) shows the merge scenario with subtree $F_1$ and its children $F_{1.1}$ and $F_{1.2}$ with two different non-functional subtree decompositions: aggressive and distracted. The second row (II.) shows the extraction of the non-functional subtree of each merge scenario and their addition to the `Component Library`. The third row (III.) shows that the subtrees from the `Component Library` can be reused in a different scenario (i.e., left-turn). As such, SavviDriver supports the interchange and reuse of subtrees and scenarios to discover uncertainty (e.g., unexpected interactions between an Ego and different combinations of NeVs in diverse mixed-traffic scenarios).

**Modeling Process: 1** Step 1 of SavviDriver (performed by developer).

```
 1: /* Step 1a. Functional Objective Modeling */
 2: ▷ Decompose Scenario_Objective down to requirements to define Functional_Subtree.
 3: ▷ Annotate leaf-level requirements with utility functions.
 4:
 5: /* Step 1b. Non-functional Objective Modeling */
 6: ▷ If Driver_Type exists in Component_Library.
 7:       ▷ Retrieve NonFunctional_Subtree from existing Component_Library.
 8: ▷ Else
 9:       ▷ Decompose Driver_Type down to requirements to define NonFunctional_Subtree.
10:       ▷ Annotate leaf-level requirements with utility functions.
11:       ▷ Add NonFunctional_Subtree to Component_Library.
12:
13: /* Step 1c. Goal Modeling */
14: ▷ AND-compose Functional_Subtree and NonFunctional_Subtree and attach to top-level goal.
15: ▷ Return agent goal model.
```

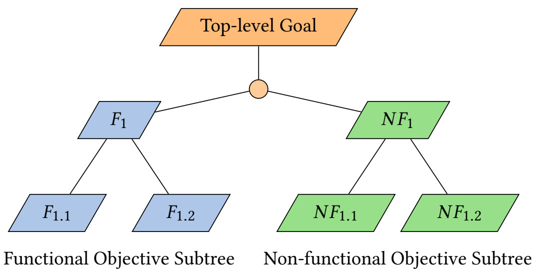


**Fig. 4**: Abstract representation of the KAOS goal tree for a vehicle in a scenario. The functional subtree decomposition is shown in blue, while the non-functional subtree decomposition is shown in green.

Consider the running example in Figure 3 involving the merge lane. We create goal models comprising their respective objectives for both the Ego and the NeV following the process described in **Step 1. Modeling**. Figure 6a and Figure 6b show sample KAOS goal models for the Ego and Aggressive-NeV, respectively. *KAOS Goals* are depicted by parallelograms. Goal decomposition is represented by refinement

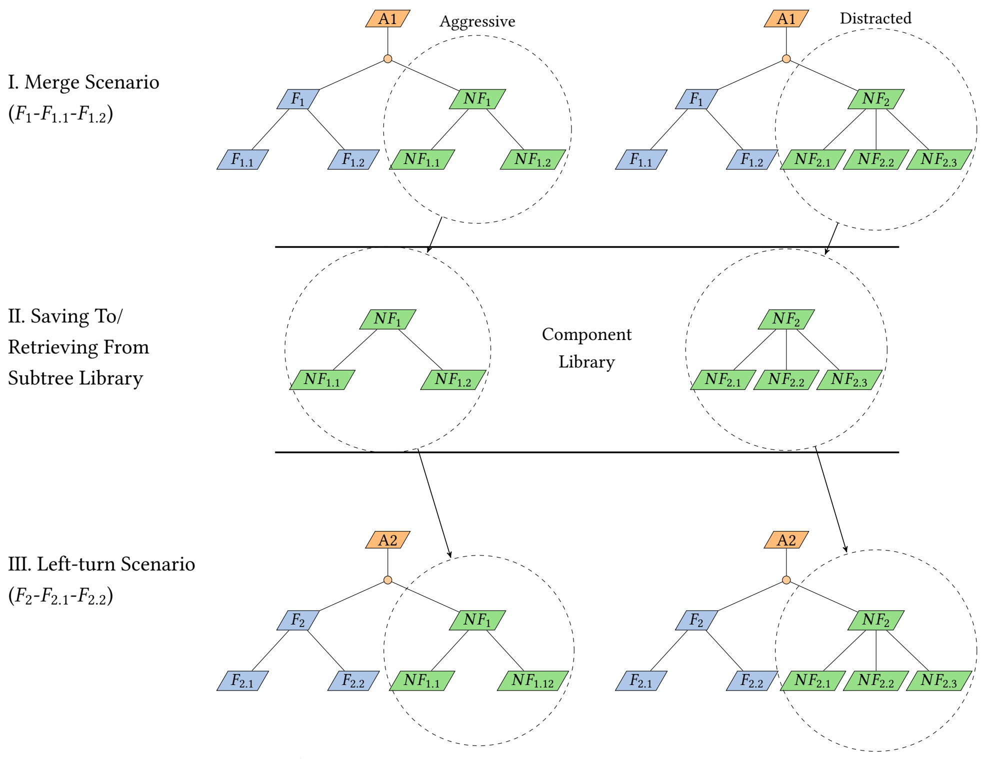
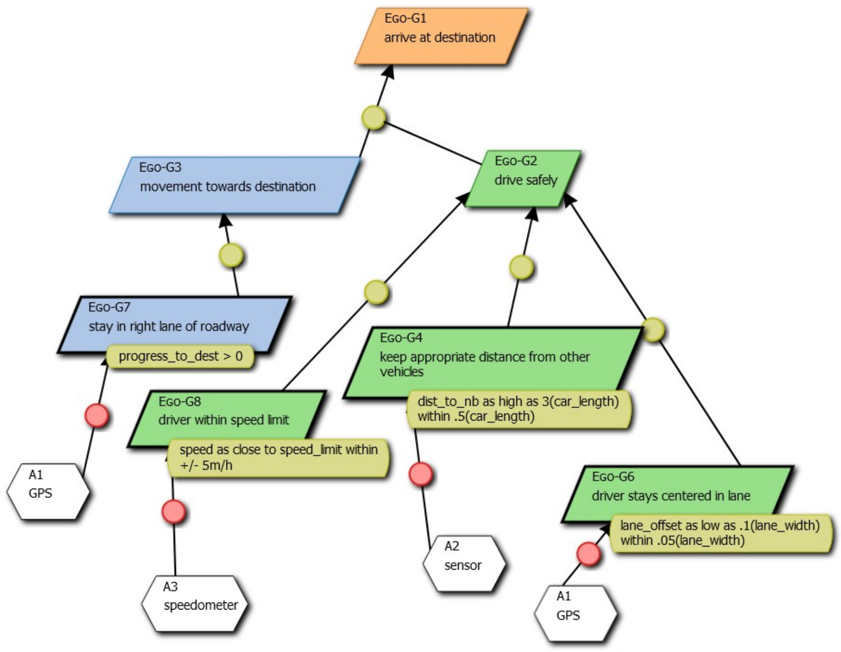


**Fig. 5**: Example of how modular subtrees can be reused in SavviDriver. The first row (I.) shows the merge scenario with two different non-functional objectives (i.e., aggressive ($NF_1 - NF_{1.1} - NF_{1.2}$) and distracted ($NF_2 - NF_{2.1} - NF_{2.2} - NF_{2.3}$)). The second row (II.) shows the extraction of the subtrees from the merge scenario to the `Component Library`. Finally, the third row (III.) shows the left-turn scenario, reusing the non-functional objectives from the merge scenario with a new functional objective.

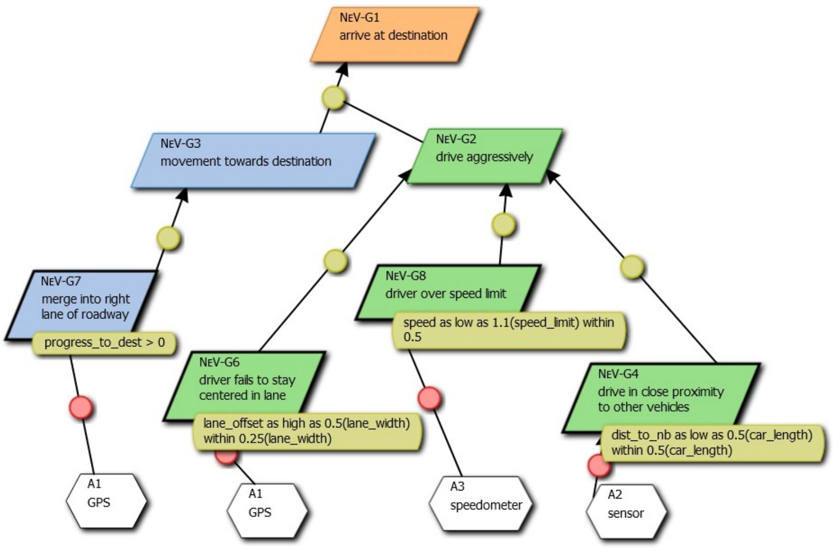
arrows. The `Developer` also specifies utility functions to be associated with leaf-level goals (represented by yellow ellipses), which are also used to specify objective functions for the driver model for each agent [29]. Blue parallelograms denote functional goals, while green parallelograms denote non-functional goals. The utility functions are discharged to system components, represented by white hexagons. We next describe the development of models for each agent in our running example in turn.

### *KAOS Goal Model for the AV (Figure 6a)*

For the Ego, the top level goal Ego-G1 '`arrive at destination`' is `AND-` decomposed into the context-dependent functional goal Ego-G2 '`move towards destination`' (**Step 1a**) and context-independent non-functional goal Ego-G3 '`drive safely`' (**Step 1b**). The functional goal Ego-G2 is decomposed into ITS-specific goal Ego-G7 '`stay in right lane of roadway`'. The safety goal Ego-G3 is further decomposed into subgoals. Associated with each leaf-level goal (i.e., requirement) is a utility function


EGO-G1
arrive at destination
EGO-G3
movement towards destination
EGO-G2
drive safely
EGO-G7
stay in right lane of roadway
progress_to_dest > 0
EGO-G4
keep appropriate distance from other vehicles
EGO-G8
driver within speed limit
dist_to_nb as high as 3(car_length) within .5(car_length)
speed as close to speed_limit within +/- 5m/h
A1
GPS
EGO-G6
driver stays centered in lane
lane_offset as low as .1(lane_width) within .05(lane_width)
A2
sensor
A3
speedometer
A1
GPS

(a) KAOS goal model for the EGO.


NEV-G1
arrive at destination
NEV-G3
movement towards destination
NEV-G2
drive aggressively
NEV-G7
merge into right lane of roadway
NEV-G8
driver over speed limit
progress_to_dest > 0
speed as low as 1.1(speed_limit) within 0.5
NEV-G6
driver fails to stay centered in lane
NEV-G4
drive in close proximity to other vehicles
lane_offset as high as 0.5(lane_width) within 0.25(lane_width)
dist_to_nb as low as 0.5(car_length) within 0.5(car_length)
A1
GPS
A1
GPS
A3
speedometer
A2
sensor

(b) KAOS goal model for AGGRESSIVE-NEV.

**Fig. 6**: Overview of KAOS goal models for the EGO and the AGGRESSIVE-NEV.

that quantifies its satisficement (i.e., degree of satisfaction [27]). For example, the non-functional goal Ego-G4 '`keep appropriate distance from other vehicles`' is formally captured by the utility function in Expression (4), where the Ego vehicle seeks to keep its current distance (`dist_to_nb`) as high as 3 car lengths. The Ego vehicle is additionally rewarded if it exceeds the minimum of 3 car lengths, with a max reward at 3.5 car lengths. Finally, **Step 1c** composes subtrees rooted at goal Ego-G2 and goal Ego-G3 via KAOS `AND-` relation and adds them as children to the top-level goal Ego-G1.

$$f(\texttt{dist\_to\_nb}) = \begin{cases} 0 & \text{if } \texttt{dist\_to\_nb} \leq 3 \cdot \texttt{car\_length} \\ 1 & \text{if } \texttt{dist\_to\_nb} \geq= 3.5 \cdot \texttt{car\_length} \\ \frac{\texttt{dist\_to\_nb} - 3.0 \cdot \texttt{car\_length}}{0.5 \cdot \texttt{car\_length}} & \text{if } 3 \cdot \texttt{car\_length} \leq \texttt{dist\_to\_nb} \leq 3.5 \cdot \texttt{car\_length} \end{cases} \tag{4}$$

***KAOS Goal Model for Aggressive NeV (Figure 6b)***

For the Aggressive-NeV, the top level goal NeV-G1 '`arrive at destination`' is `AND-`decomposed into the context-dependent functional goal NeV-G4 '`move towards destination`' (**Step 1a**) and context-independent non-functional goal NeV-G2 '`get to destination quickly`' (**Step 1b**). The functional goal NeV-G4 is next decomposed into ITS-specific goal NeV-G7 '`merge into right lane of roadway`'. The agent-specific goal NeV-G2 is then refined into a lower-level agent-specific goal NeV-G3 '`drive aggressively`'. The aggressive driving goal NeV-G3 is further `OR-`decomposed into quantified subgoals including speeding, swerving, and distance to vehicles. Finally, **Step 1c** composes subtrees rooted at goal NeV-G4 and goal NeV-G2 via KAOS `AND-` relation and adds them as children to the top-level goal NeV-G1 '`arrive at destination`'.

## 3.2 Step 2. Middleware

This section overviews how SavviDriver's `Middleware` collates and processes information from the agent goal models and scenario specifications to coordinate the game-based testing of AVs. Specifically, the `Middleware` uses two different configurators, a `Scenario Configurator` (**Step 2a**) and a `Reward Configurator` (**Step 2b**), to automatically realize mixed-traffic interactions as a *game* between an Ego and one or more NeV(s), where driver behaviors are informed by agent-specific goal models. We next describe each configurator in turn.

### 2a. Scenario Configurator

We use game theory to provide the structure of the mixed-traffic scenario to support requirements-based exploration of human-based uncertainty. In our framework, a traffic scenario is captured as a game that comprises the following components: (1) the vehicles (i.e., players) of the game, (2) the actions available to each vehicle (e.g., steering, throttle), (3) the information structure (i.e., what a driver knows at a given step of the game), and (4) the objective function each driver seeks to maximize.

To this end, we formally define the mixed-traffic scenario as a non-zero-sum, non-cooperative game [11]. We view traffic scenarios as non-zero-sum games as we do not enforce any constraints on the sum of player payoff functions. In contrast, in zero-sum games, a player's advantage necessarily results in an equivalent loss for the other player(s) (e.g., when a player takes a bigger slice of the pie, the other players have less pie available to consume). A zero-sum game is a special case of a non-zero-sum game, where the total payoff is equal to zero [11, 50]. Let player set $\mathcal{I}$ represent $N$ interacting players, where each player $p_i, p_i \in \mathcal{I}$ corresponds to a vehicle in the game (see Expression (5)).

$$\mathcal{I} = \{p_1, p_2, \dots, p_N\}. \quad (5)$$

The gaming setup captures the number and type of agents, including the Ego and one or more NeVs, as well as the operating context in which they interact, such as a highway merge scenario or left-turn scenario. An `ITS Simulator` provides an environment that enables games to be executed to discover uncertainty. The `Scenario Configurator` (**Step 2a**) takes as input a `Scenario Definition` file to initialize the `ITS Simulator` and provide the `Scenario Configuration` as the operating context for the `Game Configuration`. A sample `Scenario Definition` file is shown in Listing 1. This information captures the road layout, the number and types of agents in play, and their initial states (e.g., position, velocity).

## 2b. Reward Configurator

We next describe the automatic generation of each player's reward. The KAOS model serves as the "middleware" between the gaming setup and the ITS simulation, enabling developers to formally realize high-level developer input in terms of low-level functional goals instead of directly editing simulation code. In SavviDriver, non-cooperative game theory is used to support simulation-based testing to discover previously unseen traffic interactions between the Ego and one or more NeV(s). RL is used to realize the non-cooperative games, where gaming interactions are captured via RL reward functions. However, designing reward functions is challenging, especially for complex tasks with multiple competing objectives such as mixed-traffic scenarios [71]. Moreover, RL-based techniques often rely on ad-hoc development approaches that result in fragile/unstable reward functions; it may be unclear how modifying the reward function impacts observable agent behavior [55]. To this end, the `Reward Configurator` (**Step 2b**) "compiles" high-level agent objectives defined in goal models into RL reward functions, thereby bridging the gap between the previous modeling step and the non-cooperative ITS game, to discover uncertain behaviors for a given traffic scenario.

Listing 1: Sample `Scenario Definition` file for the merge scenario.

```
map:
  filename: merge_road
vehicles:
  - name: Safety-Ego
    position: [427, 750]
    velocity: 40
  - name: Aggressive-NeV
    position: [370, 750]
    velocity: 40
```

Recall that a scenario comprises a set of players $p_i$ with corresponding KAOS goal models $k_i$. A strategy space $\mathcal{S}_i$ represents potential strategies $s_i$ that player $i$ can use for decision-making [72, 73]. In RL, a strategy $s_i$ may be represented by a DNN that maps the current state of the operating environment observed by player $p_i$ (e.g., vehicle locations, velocities, trajectories, etc.) to an action (e.g., actuator inputs such as braking, steering, etc.). Each player $p_i$ seeks to find the "best" strategy $s_i^*$ based on an individual reward (payoff) function $\pi_i$. Each player's DNN can be trained via RL to approximate the best strategy $s_i^*$ [7, 16], which is formally specified in Expression (6):

$$s_i^* = \underset{s_i \in \mathcal{S}_i}{\operatorname{argmax}}\, \pi_i(s_i, s_{-i}^*), \forall s_i \in \mathcal{S}_i, \forall i \in \{1, \ldots, N\}, \tag{6}$$

where $s_{-i}^*$ denotes the best strategies of all other players $\{p_1, ...p_{i-1}, p_{i+1}, ...p_N\}$. Thus, $\pi_i(s_i, s_{-i}^*)$ denotes the reward obtained by player $p_i$ when using strategy $s_i$ in the context of other players using strategies $s_{-i}^*$, respectively [73, 74].

For each vehicle agent (i.e., player $p_i$) in the scenario, SavviDriver automatically extracts utility functions from the associated KAOS goal, $k_i$, to structure its respective reward function $\pi_i$ (see Expression (7)). The reward function maps the utility functions to specific observable properties of the simulation platform, which is then used to guide and constrain the behavior of vehicles in the ITS simulation. Specifically, each KAOS goal model $k_i$ contains a number $M_i$ of utility functions $f_{ij}$ associated with leaf-level goals, where $1 \leq j \leq M_i$. Each $f_{ij}$ is a real-valued function (e.g., see Expression (4)) that maps a specific requirement of player $p_i$'s goal model $k_i$ to a satisficement value [29]. Each utility function associated with the leaf-level goals is assigned a weight value $\alpha_{ij}, \alpha_{ij} \in \mathbb{R}$, denoting a developer-specified weight (e.g., to indicate the importance or impact) of the given goal. The reward function $\pi_i$ directly guides and constrains the behavior of vehicles in the ITS simulation.

$$\pi_i(s_i) = \sum_{j=1}^{M_i} \alpha_{ij} \cdot f_{ij}(p_i). \tag{7}$$

## 3.3 Step 3. Game-Based Testing

This section overviews how SavviDriver integrates and operationalizes the KAOS goal models and their utility functions into a game to explore human-based uncertainty in an ITS simulation. An `ITS Simulator` is used to keep track of the operating environment's state for a given gaming process (see Figure 2, *ITS Simulator*). For each timestep of a game, the following happens. First, agents use their respective strategies to select actions (e.g., braking, throttle, steering, etc.) which are then provided as input to the `ITS Simulator`. Second, the `ITS Simulator` executes the collective actions from all the participating agents. The `ITS Simulator` then updates the environment (e.g., vehicle positions, velocities, etc.) and returns a new ITS state to the respective gaming process. The sequence of each agent's actions and resulting environment states inform the RL learning process and may be archived (e.g., simulation traces) to support further analysis processes. We next overview how SavviDriver uses game-based testing in simulation to discover human-based uncertainty potentially impacting an Ego.

### Step 3a. NeV Training

First, SavviDriver trains the NeV(s) to complete driving tasks while exhibiting a specific driving style in the presence of the Ego (provided as input by a developer). The NeV(s) are in training mode, where the DNN is actively learning new behavior in response to its environment. The goal models from **Step 1b** inform the reward structure of each NeV. Thus, the NeV(s) learn to complete their respective functional goal specified in the corresponding KAOS goal model that reflects a particular driving style (e.g., aggressive, distracted, timid, etc.). For example, an aggressive NeV may exploit the safe driving behavior of an Ego (i.e., maintains a conservative following distance) by cutting in front of the Ego in order to satisfy a NeV non-functional goal to get to its destination quickly.

### Step 3b. AV Uncertainty Discovery

Next, SavviDriver assesses the ability of the Ego to operate appropriately in the presence of the trained NeV(s) (trained in **Step 3a**). The NeV(s) in *execution mode* choose strategies learned during training (**Step 3a**) according to their reward functions; they do not learn new behavior in response to agent interactions. During this step, the Ego may exhibit previously unseen or unexpected behavior as it responds to the other "human-operated" vehicles exhibiting different driving styles. For example, when the aggressive NeV cuts off the Ego, the (unexpected) Ego response may be to swerve to avoid a collision, which may lead to another collision. The result of this step is a collection of *failure traces*, where *failure* is defined as a scenario execution where the Ego did not reach its destination, for example, due to a collision with an NeV or a road obstacle. Failure traces are sequences of agent actions and corresponding environment states (i.e., trace data), indexed by timesteps, for a given execution of the mixed traffic scenario where a failure occurs. Developers can use failure traces to visualize and replay different executions of mixed traffic scenarios to better understand the interactions and environment states that led to interesting behavior or unexpected Ego failures.

## 3.4 Step 4. AV Robustification

Similar to simulation-based testing [75, 76, 77], SavviDriver's generated information (i.e., failure traces from **Step 3b**) can be used by developers to revise the system to prevent and/or mitigate the discovered unexpected Ego behavior. The information can be used in several ways, including updating system requirements, modifying system implementation, or used as training data for learning-enabled system components. We next elaborate on these approaches to robustification.

***Robustification of Software-Based AVs***

For traditional software-based Egos (e.g., Intelligent Driving Model (IDM) [78]), developers can use failure traces to identify goal violations (e.g., collisions, failure to drive towards destination, etc.) and determine how to revise the behavior of the Ego. For example, developers may observe similar failures (i.e., collisions) during merge maneuvers when another vehicle is in close proximity to the Ego, resulting in low rewards. Developers can then use this collection of similar driving patterns/scenarios to specify goals that represent a "cautious" operating mode that increases the desired inter-vehicle distance to avoid collisions during merge maneuvers.

***Robustification of Learning-Based AVs***

For learning-based Egos, SavviDriver supports robustification by retraining RL-based Egos in the context of the discovered uncertainty (i.e., failure traces from **Step 3b**). During retraining, the Ego is in training mode and the NeV(s) are in execution mode. The Ego agent learns to complete the goals specified in its KAOS goal model in the presence of uncertainty posed by the NeV(s) that have learned to exhibit specific driving styles specified by their corresponding goal models. The result of this step is an automatically retrained Ego that can move robustly and safely, completing the given task in the presence of human-based uncertainty exhibited by the NeV.

## 4 Demonstration

To demonstrate the use of the SavviDriver framework to discover unexpected human-based behavior(s) and assess Ego responses, we have applied it in several proof-of-concept use cases based on real-world traffic accident data [30, 79, 80, 81]. We present results for the discovered mixed traffic interactions that cause Ego failure, and then illustrate how developers can use these results to improve an Ego's robustness in response to the human-based uncertainty.

***Experiment Setup***

Our demonstration is intended as a proof-of-concept study to illustrate the modeling, uncertainty discovery, and robustness analysis capabilities of our framework. To this end, we leverage existing tooling for our experiments. First, we use our custom in-house simulation package, TinyRoad [7], to provide a light-weight 2D traffic environment. TinyRoad uses a bicycle-kinematics physics model [82] and simple 2D graphics that require less computational overhead when compared to more sophisticated and computationally expensive simulation platforms (e.g., CARLA [39], Gazebo [83]). Thus, the selected simulation platform supports quick training and evaluation of RL agents while preserving the key behavioral characteristics of Ego and NeV interactions.

While SavviDriver can use any RL algorithm that takes as input the environmental state and outputs a discrete action for the vehicle, we use PPO [16] and the OpenAI Gym [84] libraries to implement RL training and agents within the simulation as a proof-of-concept demonstration for SavviDriver. We implement a custom environment using the Stable Baselines3 library. The observation space comprises sensor inputs; in this work, we use image-based observations of size $\mathbb{R}^{128*128*3}$. The images are processed by a Convolutional Neural Network (CNN), and the resulting vector is then passed to a multi-layer perceptron (MLP) that outputs a distribution over the action space. The action space is a vector in $\{x_1, x_2 \; : x_1, x_2 \in [-1, 1]\}$ where the first component represents throttle values and the second component represents steering values. Table 2 overviews the hyperparameters used in PPO training. We started with hyperparameter values commonly used in other DRL settings [16] and empirically tuned them for our application.

**Table 2**: Overview of RL training parameters.

| **Hyperparameters** | **Value** | **Hyperparameters** | **Value** |
|---|---|---|---|
| Total timesteps | $3 \times 10^6$ | Clip $\epsilon$ | 0.1 |
| Early convergence | True | Entropy Coefficient | 0.02 |
| Optimizer | Adam | Learning Rate | $1 \times 10^{-4}$ |
| Discount Factor $\gamma$ | 0.99 | Replay Buffer Size | 2048 |

***Evaluation Metrics***

We use two well-established and commonly-used vehicle safety indicators to assess the performance of the Ego: the Average Failure Rate (AFR) [85] and the average Time Exposed Time-to-collision (TET) [86, 87]. The safety metrics are defined over a set of episodes $E$, where $|E| = 100$ in each use case. An episode is an execution of the mixed traffic simulation from an initial state to a terminal state (i.e., the Ego reaches its destination or a collision occurs).

In our studies, a *failure* is defined as an episode where the Ego under study does not reach the destination, for example, due to a collision between the Ego under study and an NeV or a road obstacle. Expression (8) defines an indicator function $\delta_{\texttt{AFR}}$ that evaluates to 1 if a failure occurs, and 0 otherwise. The AFR metric is defined as the average number of collisions observed over episode set $E$ (see Expression (9)).

$$\delta_{\texttt{AFR}}(e) = \begin{cases} 0 & \text{if Ego}.\textit{state} \neq \text{reach\_dest} \\ 1 & \text{if Ego}.\textit{state} = \text{reach\_dest} \end{cases} \tag{8}$$

$$AFR(E) = 100 \times \frac{1}{|E|} \sum_{e \in E} \delta_{\texttt{AFR}}(e) \tag{9}$$

The Time-to-Collision (TTC) metric [87] is defined as the time remaining until a collision occurs, where collisions are predicted using the linear projection of each vehicle's future position [85]. Expression (10) defines the TTC value for an AV belonging to vehicle set $\mathcal{I}$ at a given timestep $t$. The distance formula $d$ calculates the predicted Euclidean distance between the position $pos_{\text{AV}}(t+\tilde{t})$ of the AV, and the position $pos_j(t+\tilde{t})$ of vehicle $p_j, p_j \in \mathcal{I} - AV$ at future time $t+\tilde{t}$. When no collision is predicted, the TTC value is set to $\infty$ [85].

$$\begin{aligned} TTC(\textsc{Ego}, \mathcal{I} - \textsc{Ego}, t) = \min\,(\{\{\tilde{t} \geq 0 \mid d(pos_{\text{AV}}(t+\tilde{t}), \\ pos_j(t+\tilde{t})) = 0\} \cup \{\infty\}\}), \forall j \in \mathcal{I} - \textsc{Ego}. \end{aligned} \tag{10}$$

Next, Expression (11) defines an indicator function $\delta_{\texttt{TET}}$ that evaluates to 1 if an input value $ttc$ is less than or equal to the given time threshold (seconds) $\tau, \tau \in \mathbb{R}$, and 0 otherwise. Finally, the TET metric is defined as the duration the Ego was exposed to a TTC value below some threshold $\tau$ during episode $e$'s time interval $\{0, t_e\}$, averaged over episode set $E$ (see Expression (12)) [85, 86].

$$\delta_{\texttt{TET}}(ttc) = \begin{cases} 0 & \text{if } ttc > \tau \\ 1 & \text{if } ttc \leq \tau \end{cases} \tag{11}$$

$$TET(E) = \frac{1}{|E|} \sum_{e \in E} \left( \frac{1}{t_e} \sum_{t=0}^{t_e} \delta_{\texttt{TET}}(TTC(\textsc{Ego}, \mathcal{I} - \textsc{Ego}, t)) \right) \tag{12}$$

## 4.1 Use Case: Left Turn

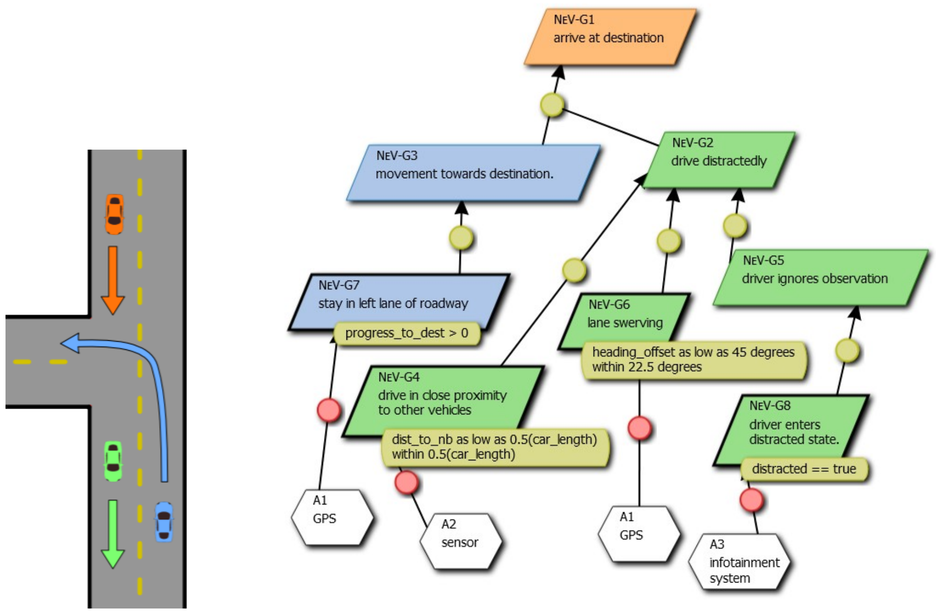
Our first use case explores an uncontrolled left turn intersection. According to the NHTSA [88], a left-turn maneuver is one of the most frequent pre-crash events, especially at intersections without controlled signals [79, 80]. Figure 7a provides an overview of the left-turn use case considered, where the blue Ego seeks to make a left turn and the orange Aggressive-NeV and green Distracted-NeV are traveling in the left lane (towards the blue Ego). Figure 7b provides the corresponding goal model for the green Distracted-NeV.

***Data***

This section describes the key information that is used to define the scenario for this use case. The *scenario* is an unprotected left-turn intersection that comprises a two-lane road with a perpendicular road connecting to the left lane. Therefore, a vehicle in the right lane attempting to make a left turn must cross a lane with oncoming traffic. The *agents* participating in this scenario are an aggressive driver (orange), a distracted driver (green), and the AV (blue). The AV under study uses a traditional IDM for decision-making. The *road layout* is modeled based on real-world map data, and we provide the road geometry as input to SavviDriver.

***Step 1. Modeling***

This section describes the development of goal models (**Step 1. Modeling**) for our left-turn use case. We explore two different NeVs, an Aggressive-NeV and a Distracted-NeV operating in a multi-agent setting. For the Aggressive-NeV, we reference the component library to reuse the non-functional subtree of the Aggressive-NeV defined in our merge running example (see Section 3). The previously developed non-functional subtree is based on NHTSA's traffic data that identifies three leading causes of accidents related to aggressive drivers: speeding, proximity to other road users, and inability to center in a lane [24, 25, 26]. We update the functional objective of the Aggressive-NeV to reflect the functional context-dependent goal: NeV-G7 '`stay in right lane of roadway`'. This process demonstrates how SavviDriver facilitates the reusability of goal models by using previously defined agent goal(s) in a new scenario. Figure 7b shows the KAOS model for the Distracted-NeV, where the top-level goal of NeV-G1 '`arrive at destination`' is decomposed into subgoals such as NeV-G8 '`driver enters distracted state`', NeV-G6 '`lane swerving`', and NeV-G4 '`drive in close proximity to other vehicles`'.



NeV-G1
arrive at destination
NeV-G3
movement towards destination.
NeV-G2
drive distractedly
NeV-G5
driver ignores observation
NeV-G7
stay in left lane of roadway
progress_to_dest > 0
NeV-G6
lane swerving
heading_offset as low as 45 degrees within 22.5 degrees
NeV-G4
drive in close proximity to other vehicles
dist_to_nb as low as 0.5(car_length) within 0.5(car_length)
NeV-G8
driver enters distracted state.
distracted == true
A1
GPS
A2
sensor
A1
GPS
A3
infotainment system

(a) Graphical depiction of left-turn use case.

(b) KAOS goal model for the green Distracted-NeV.

**Fig. 7**: Overview of left-turn use case and green Distracted-NeV goal model. The left-turn use case involves three vehicles. The blue Ego attempts to make a left turn, while the orange Aggressive-NeV and green Distracted-NeV are driving in the left lane.

***Step 2. Middleware***

This section describes how SavviDriver processes goal models and scenario information to automatically initialize the gaming setup and simulator for the left-turn use case. First, **Step 2a** takes as input the scenario definition file that captures key context-dependent information for this use case and instantiates the simulator. Next, the reward function for each agent is automatically compiled from the aggregate utility functions in their respective goal models (**Step 2b**). Listing 2 provides a pseudocode example of the automatically generated reward function for the Distracted-NeV. The utility function associated with goal NeV-G4 rewards the Distracted-NeV for driving as close as possible to other vehicles, with a maximum reward assigned when its distance to the closest vehicle is less than half of its length. The utility function associated with NeV-G6 rewards the Distracted-NeV for swerving and is parameterized by the current heading offset from the lane center. The maximum reward is achieved when the current offset is greater than 45 degrees. The utility function associated with NeV-G8 rewards the Distracted-NeV for entering a distracted state and is parameterized by a boolean value that indicates if the driver is distracted. The NeV-G8 utility function returns a reward of 1 if the driver is distracted and 0 otherwise. The NeV-G7 rewards the Distracted-NeV for making progress between timesteps. The NeV-G7 utility function returns a reward of 1 if the driver makes progress and 0 otherwise. Each utility function evaluates to a real-valued number (i.e., level of satisficement). Using Expression (7), the reward is computed as the linear weighted sum of the satisficement values multiplied by developer-specified weights. Since the Aggressive-NeV has the same non-functional objectives as the merge running example, we regenerate the Aggressive-NeV's reward structure to reflect the new context-dependent functional objective for the left-turn use case. We describe the results for the left-turn use case next.

***Step 3. Game-Based Testing***

Figure 8 illustrates a failure scenario observed during the uncertainty discovery process (**Step 3b**) for the left-turn use case. The blue Ego's initial attempt to make a left-turn maneuver (see Ⓐ, Figure 8) is aborted due to the swerving behavior of the green Distracted-NeV. Then the blue Ego attempts another left-turn maneuver in close proximity to the orange Aggressive-NeV, causing it to enter the Ego's lane (see Ⓑ, Figure 8). The orange Aggressive-NeV's behavior causes the blue Ego to accelerate and collide with the road barrier (see Ⓒ, Figure 8), therefore resulting in an increased failure rate when the two NeVs are in close proximity to the Ego's lane. Table 3 provides quantitative results for the left-turn use case. In the absence of the NeVs, the Ego is able to complete the left turn with a failure rate of approximately 0.00%. However, when exposed to the distracted and aggressive NeVs, the Ego's failure rate increases to 12.00%. Additionally, the observed TET values in Table 3 indicate that the Ego was exposed for the highest average duration of time to critical TTC values $\tau$, $\forall \tau \in \{1, 2, 3\}$ in the presence of the NeVs. TTC values below a specific time threshold (e.g., $\tau = 3$ seconds [86]) potentially indicate that a vehicle would not have enough time to perform a collision-avoiding maneuver, such as applying the brakes to stop the vehicle. We observe higher average TET values indicating the Ego was in a near-collision state when exposed to both NeVs.

Listing 2: Pseudocode for the Distracted-NeV's reward function corresponding to its KAOS goal model.

```
def reward(p):
    # utility function for G4
    #   p.nb := distance to nearest neighbor
    #   p.s  := car length of NeV
    fi_0 = (0 if p.nb >= 0.5*p.s else
        (1 if p.nb < 0.5*p.s else
        (0.5*p.s-p.nb)/0.5))

    # utility function for G6
    #   p.hof  := NeV's heading offset from current lane
    fi_1 = (0 if p.hof <= 22.5 else
        (1 if p.hof >= 45 else (45-p.hof)/22.5))

    # utility function for G8
    #   p.d := boolean indicating if NeV is distracted
    fi_2 = (0 if not p.d else 1)

    # utility function for G7
    #   p.prog := percent increase in route completion from last update
    fi_3 = (1 if p.prog > 0 else 0)

    # weights - default is equal weights
    alpha = [1, 1, 1, 1]

    # Reward function (see Expression (4))
    uf = [fi_0, fi_1, fi_2, fi_3]
    reward = 0
    for j in range(|uf|):
        reward += alpha[j]*uf[j]
    return reward
```

***Step 4. Robustification (of Software-based Ego)***

The discovered uncertainty and corresponding failure traces are used to inform a reconfiguration of the Ego's IDM, a rule-based controller. Specifically, during the replay and analysis of the failure traces, we observed that the Ego did not properly wait for an opening to make a left turn, attempting the turning maneuver in close proximity to other vehicles. We confirmed our visual observations through further examination of the failure trace data, specifically the distance between the vehicles leading up to a failure. Informed by our analysis, we developed a KAOS functional subtree for 'cautious' mode [89] for the Ego that implemented a blocking routine until the intersection is clear with respect to a given distance threshold before attempting to perform the left-turn maneuver. After reconfiguration in **Step 4. Robustification**, the robustified Ego is better able to wait for openings in the intersection, thereby reducing its failure rate from 12.00% down to 1.00%. This use case demonstrated how SavviDriver was able to reuse existing goal models, extend them with different numbers and types of drivers, apply them to a new ITS use case to discover unexpected human behavior with the corresponding unexpected Ego responses, and then use this information to reconfigure the Ego, resulting in a reduced failure rate. Importantly, no low-level simulation code was modified to explore this robustification process.

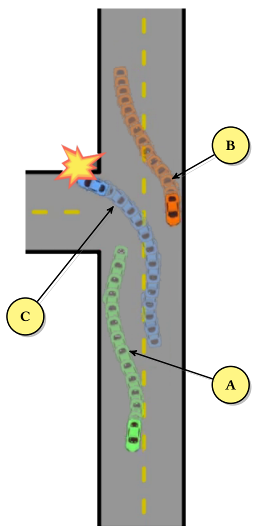


**Fig. 8**: Left turn use case overview. The green Distracted-NeV's sporadic movement led to a preemptive turn in the blue Ego. The Ego then speeds up to avoid collision with the orange Aggressive-NeV, resulting in a crash of the Ego.

**Table 3**: Evaluation of *AFR* and *TET* metrics for the left-turn case study, where red color indicates the worst value for each metric. A higher AFR indicates a higher average collision rate and a higher TET indicates higher average time spent in a near-collision state (i.e., predicted time-to-collision less than $\tau$ seconds).

| **Setting** | **AFR** | **TET** | | |
|---|---|---|---|---|
| | | $\tau = 1$s | $\tau = 2$s | $\tau = 3$s |
| Base AV | 0.00% | 0.00% | 0.00% | 0.50% |
| Uncertainty Discovery | **12.00%** | **2.90%** | **3.80%** | **4.90%** |
| Robustified AV | 1.00% | 0.10% | 0.80% | 2.70% |

## 4.2 Use Case: Weave Lane

Our next use case explores a weave lane highway merge. Figure 9 provides a graphical depiction of the weave lane highway merge use case, where the blue Ego is trailed by the orange NeV. Both vehicles must exit the on-ramp and merge onto the highway. A weave lane is a challenging highway merge scenario, as vehicles entering the highway often start at a low speed due to the entrance curve speed limit and can cause automated vehicles to fault [30]. For example, a 2020 Waymo Ego exhibited unexpected behavior during a highway merge scenario where the Ego failed to complete the merge maneuver [81]. We next detail how SavviDriver is used to model and explore specific human-based behaviors that may lead to vehicle failures in the weave lane scenario.

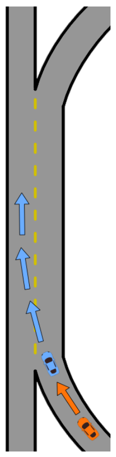


**Fig. 9**: A scenario capturing the weave lane use case in TinyRoad. Both the Ego and Aggressive-NeV are traveling from an on-ramp and attempt to merge onto the highway.

***Data***

Using historical data and/or domain expertise, we identify three types of information required for SavviDriver and capture them in a scenario definition. The *scenario* is a weave lane that comprises a highway on-ramp that merges onto the main lane of the highway. The *agents* participating in this scenario are the Aggressive-NeV and the Ego. In contrast to the left-turn use case, this use case explores an Ego that uses an RL-based controller for decision-making that was trained to satisfy the requirements captured in the goal model shown in Figure 6a [7]. The *road layout* is modeled based on real-world map data, and we provide the road geometry as input to SavviDriver.

***Step 1. Modeling***

This section describes the development of goal models for the weave lane use case. Since the non-functional objectives describe *how* an agent should achieve a task, they can often be reused between different driving scenarios. In this use case, we reuse the non-functional subtrees saved in our `Component Library` for both the Aggressive-NeV and Ego. For both agents, we update their functional objective to reflect the context-dependent top-level function goal: '`merge onto highway`'. This process illustrates how a developer may leverage the `Component Library` during **Step 1. Modeling** to reuse existing context-independent non-functional subtrees and apply them in a new scenario to discover additional sources of human-based uncertainty.

***Step 2. Middleware***

In this step, SavviDriver processes goal models and scenario information to automatically initialize the gaming setup and simulator for the weave lane use case. **Step 2a** processes the scenario definition file and instantiates the simulator. Next, the goal models are automatically processed to generate the reward structure for the game-based testing process (**Step 2b**). Specifically, utility functions associated with leaf-level goals are aggregated into a linear weighted sum that represents the reward for each agent in the game. Notably, SavviDriver automatically generates a new reward structure for the weave lane use case through model-level updates rather than low-level simulation code changes.

***Step 3. Game-Based Testing***

Figure 10 illustrates an Ego failure observed during the uncertainty discovery phase. In the presence of the NeV, the Ego attempted an early merge maneuver onto the highway to maintain a safe distance away from the Aggressive-NeV. However, the Ego failed to account for the trajectory of the NeV, resulting in an increased failure rate when both vehicles merged in close proximity to each other. Table 4 shows a comparison between the average Ego AFR and TET over 100 random episodes. In this use case, we use an AV that has been trained with an RL as a baseline. The baseline Ego is able to complete the highway merge in the presence of external rule-following traffic with a failure rate of only 2.0%. However, when exposed to the aggressive NeV (**Step 3b**), the Ego's failure rate increases to 34.0%. Additionally, the observed TET values (see Table 4) indicate that the Ego was exposed for the highest average duration of time to critical TTC values (i.e., $\tau$, $\forall\tau \in \{1, 2, 3\}$) in the presence of the aggressive NeV. Higher average TET values indicate the Ego was more frequently exposed to near-collision scenarios, which may contribute to the increased failure rate. For example, we observe that the Ego was in a near-collision state (within $\tau = 3$ seconds) for 12.60% of the episodes when exposed to the aggressive NeV.

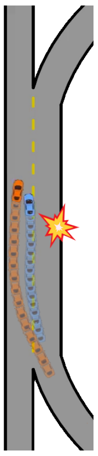


**Fig. 10**: Demonstration of a failure discovered by SavviDriver, where the orange Aggressive-NeV completes its merge before the blue Ego and speeds up, causing a collision as the Ego tries to merge.

**Table 4**: Evaluation of *AFR* and *TET* metrics for the weave lane case study, where red color indicates the worst value for each metric. A higher AFR indicates a higher average collision rate and a higher TET indicates higher average time spent in a near-collision state (i.e., predicted time-to-collision less than $\tau$ seconds).

| **Setting** | **AFR** | **TET** | | |
|---|---|---|---|---|
| | | $\tau = 1s$ | $\tau = 2s$ | $\tau = 3s$ |
| Base AV | 2.00% | 4.20% | 7.50% | 7.90% |
| Uncertainty Discovery | **34.00%** | **8.60%** | **11.40%** | **12.60%** |
| Robustified AV | 18.00% | 5.00% | 7.10% | 8.00% |

## *Step 4. Robustification (of RL-based Ego)*

**Step 4. Robustification** robustifies the Ego vehicle by retraining its RL-based controller in the presence of the discovered uncertainty. After retraining, the Ego learned to reduce its speed during the highway merge maneuver, thereby reducing its failure rate to 18.0% (down from 34.0%). Specifically, after retraining, the Ego learned to delay the merging maneuver to allow the aggressive driver to pass at a safe distance. This demonstration also illustrated how SavviDriver was able to reuse composable components of existing behavior models and apply them to a new use case to discover dangerous human behavior and unexpected Ego responses. Then SavviDriver used this information to retrain the Ego, resulting in a reduced failure rate.

## 4.3 Use Case: Adaptive Cruise Control

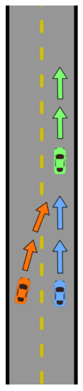
Our final use case explores a merge scenario involving an AV equipped with a traditional software implementation of an Adaptive Cruise Control (ACC) system, an Advanced Driver Assistant System (ADAS) feature found in most modern vehicles. The ACC system uses onboard sensors, such as radar and lidar, to maintain a safe trailing distance between itself and a *target* vehicle. The goal of this use case is to illustrate how SavviDriver may be used to assist in the engineering, assessment, and manual reconfiguration of ADASs (i.e., ACC) by developers. Figure 11a shows the road and vehicle configuration for the ACC scenario. An orange NeV seeks to merge into the lane of the blue Ego AV before reaching the end of the road. The AV seeks to maintain a safe trailing distance to a green target vehicle traveling at a constant speed.

***Data***

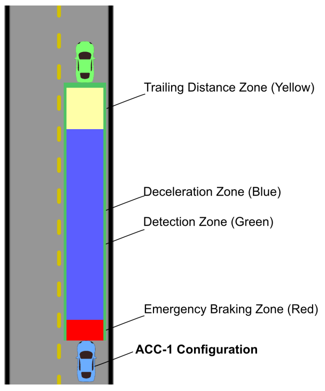
This section describes the key information that is used to define the scenario for this use case. The scenario is a highway road that comprises two lanes traveling in the same direction. The agents participating in this scenario are an aggressive driver, a target vehicle, and the AV. The road layout is modeled based on real-world map data, and we provide the road geometry as input to SavviDriver.

***Step 1. Modeling***

The ACC-1 software is based on real-world Original Equipment Manufacturers' (OEM) ACC implementations [90, 91]. Figure 12 shows a sample goal model corresponding to the ACC system [21]. The initial ACC configuration (ACC-1) is shown in Figure 11b and implemented as follows. If no target vehicle is detected in the detection zone (i.e., green rectangle), then ACC-1 accelerates to maintain the set speed. If a target vehicle is detected, then the ego vehicle constantly decelerates over a large distance (i.e., blue rectangle) and maintains the same speed as the target vehicle in the trailing distance zone (i.e., yellow rectangle). Additionally, the ACC-1 controller is equipped with an Emergency Braking System (EBS). The EBS applies a hard brake (i.e., maximum braking force) when a target vehicle is detected within the emergency braking zone of the ego vehicle (i.e., red rectangle). For the Aggressive-NeV, we reuse the goal model described in our running merge example in Section 3 (see Figure 6b). This agent's functional goal is to merge into the Ego's lane before they reach the end of the road.



(a) Overview of the ACC use case.



(b) Graphical depiction contrasting the ACC-1 system.

**Fig. 11**: Overview of the ACC use case and the ACC-1 system. Subfigure (a) shows the ACC use case that involves three vehicles. The green NeV target vehicle travels straight at a constant speed. The blue EGO vehicle uses a rule-based ACC controller to maintain a fixed trailing distance to the vehicle in front. The orange AGGRESSIVE-NEV seeks to merge into the right lane. For Subfigure (b), the yellow box shows the trailing distance the ego vehicle aims to maintain. The green box denotes the distance when the ego vehicle detects the green non-ego target vehicle. The blue box indicates the deceleration zone. Finally, the red box denotes the emergency braking zone, where the ego vehicle applies the maximum braking force if an object is detected within the zone.

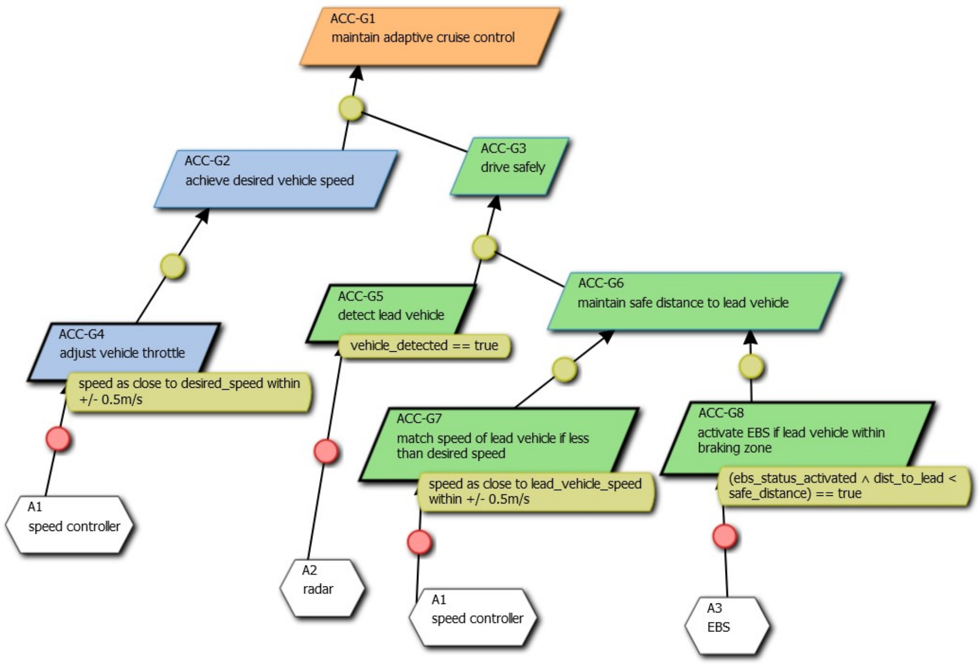

ACC-G1
maintain adaptive cruise control
ACC-G2
achieve desired vehicle speed
ACC-G3
drive safely
ACC-G4
adjust vehicle throttle
speed as close to desired_speed within +/- 0.5m/s
ACC-G5
detect lead vehicle
vehicle_detected == true
ACC-G6
maintain safe distance to lead vehicle
ACC-G7
match speed of lead vehicle if less than desired speed
speed as close to lead_vehicle_speed within +/- 0.5m/s
ACC-G8
activate EBS if lead vehicle within braking zone
(ebs_status_activated ∧ dist_to_lead < safe_distance) == true
A1
speed controller
A2
radar
A1
speed controller
A3
EBS

**Fig. 12**: KAOS model for ACC subsystem.

***Step 2. Middleware***

Similar to the previous use cases, SavviDriver automatically instantiates the simulation environment, gaming setup, and compiles the reward function for each agent using the aggregate utility functions from their respective goal models. We next describe the game-based testing and discovered uncertainty for the ACC use case.

***Step 3. Game-Based Testing***

Figure 13a illustrates a critical scenario observed during uncertainty discovery. The orange Aggressive-NeV suddenly merges in front of the blue Ego, causing the ACC system to activate its EBS. Table 5 shows a comparison between the average duration(s) of EBS activation over 100 random episodes. In the baseline scenario, the Ego does not activate its EBS as it follows the target vehicle. However, when exposed to the Aggressive-NeV, the average EBS duration increases from 0.00 seconds to 3.52 seconds. This increase indicates the ACC experienced more near-collision events in the presence of the Aggressive-NeV.

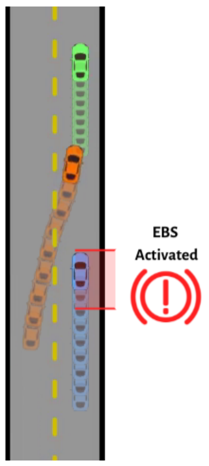
***Step 4. Robustification (of ADAS subsystem)***

Observing the unexpected maneuver from the Aggressive-NeV and response from the Ego configured with ACC-1, a developer can engineer or reconfigure the ACC system, yielding an alternative ACC-2 controller to mitigate the unsafe scenario in **Step 4. Robustification**. Figure 13b shows an overview of the ACC-2 configuration. In contrast to the gradual deceleration when a target vehicle is detected by the ACC-1 controller (see Figure 11b), the ACC-2 controller's behavior is revised to maintain its speed until it reaches a short deceleration zone before applying the brakes to match the speed of the target vehicle (see Figure 13b). After robustification, the average duration of EBS activation over 100 episodes decreases to 0.16 seconds in the presence of the Aggressive-NeV. As such, this use case highlights how a developer may use SavviDriver to identify problems in an initial configuration of an ACC controller, engineer a mitigating solution, and confirm the solution addresses previously identified unexpected behaviors. The reconfiguration of the ACC system demonstrated in this use case captures a real-world use case when ACC systems were initially deployed in the early 2000s by OEMs.[3] First iterations of ACC systems applied a similar configuration as our ACC-1 system. As OEMs discovered that aggressive drivers may cut off the ADAS-equipped vehicles, they changed the behavior of their ACC models to apply the brakes as they enter the "short deceleration zone" of the target vehicle (i.e., ACC-2). This example shows that by applying SavviDriver, developers can identify unsafe situations during testing and modify the behavior of their system to mitigate similar situations.

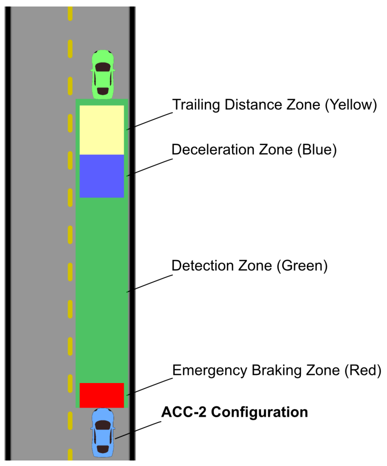
## 5 Discussion

This section discusses our results, including key findings and envisioned usage scenarios.

[3]This information is based on data provided by our industrial collaborators.



(a) Example trajectory of the orange Aggressive-NeV who abruptly merged in front of the blue Ego, activating the EBS of the Ego.



(b) Graphical depiction contrasting the ACC-2 system. This configuration maintains a smaller trailing distance compared to ACC-1.

**Fig. 13**: Overview of (a) sample uncertainty discovered by SavviDriver and (b) the corresponding robustified Ego configuration.

**Table 5**: Evaluation of ACC use case, where red color indicates the worst value for the EBS duration metric. A higher EBS duration indicates a higher average time spent in a near-collision state.

| **Setting** | **EBS Duration (s)** |
|---|---:|
| Base AV (ACC-1) | 0.00 |
| Uncertainty Discovery | **3.52** |
| Robustified AV (ACC-2) | 0.16 |

## 5.1 Synergizing Game Theory, Reinforcement Learning, and Goal Modeling

This paper has shown that game theory and RL can be synergistically combined for uncertainty exploration in AV testing. While existing research has explored cooperative game theory to assess how AVs and human-driven vehicles may cooperate to improve a common goal (e.g., improve traffic flow), these studies do not reflect the challenges of AVs when deployed, where they must successfully navigate scenarios with other human-driven vehicles that cannot communicate or share intentions. Other work has explored adversarial agents, where the objective is to minimize the reward of the AV under study. SavviDriver leverages non-cooperative game theory to model a traffic scenario, where players of the game (i.e., different vehicles) do not explicitly communicate, form coalitions, nor share high-level rewards. By defining high-level functional objectives (i.e., the objective associated with the scenario) and the non-functional objectives (i.e., how the vehicles shall achieve those objectives), our approach discovers maneuvers that can lead to undesirable outcomes.

To the best of our knowledge, SavviDriver is the first to integrate goal-based modeling with game theory and RL in order to address modularity and reuse in the discovery of human-based uncertainty. Existing research has explored the modeling of environmental uncertainty (e.g., raindrops, lighting conditions, etc.), where developers may have explicit *a priori* models (e.g., physics-based solutions, simulations, etc.) for capturing and generating samples of uncertainty [92]. However, there is not an explicit modeling approach that captures the broad range of epistemic uncertainty posed by interactions with different types of external human drivers. Instead, a key contribution of this work is the use of game theory and RL to automatically generate manifestations of human-based uncertainty based on declarative specifications of driver behaviors. SavviDriver's goal models provide developers with a means to systematically decompose objectives into sub-objectives, each of which can be further refined down to the level of requirements. Those goal models comprise modular subtrees that can then be reused and/or recombined with existing goal models to form new external vehicle objectives that produce novel human-based behaviors and unexpected responses from the AV. For example, in the Weave Lane use case, we demonstrate modularity by combining the context-independent non-functional subtrees for aggressive behavior with a new functional subtree corresponding to the '`merge onto highway`' top-level function objective. During the design phase, automotive developers can archive a collection of goal models corresponding to human-based driving behaviors, and *reuse* them to explore the impact of human-based uncertainty in a broad range of traffic scenarios/operating contexts. For example, in the ACC use case, we demonstrated reuse of the Aggressive-NeV goal model from our running example in a merge scenario involving a new type of Ego (ACC-1 and ACC-2 systems). By decoupling functional and non-functional objectives via reusable and modular goal models, developers can explore different dimensions of human-based uncertainty through the manipulation of high-level models rather than low-level code. We quantify the ability to generate uncertainty by assessing the Ego's behavior when interacting with external vehicles controlled by SavviDriver's automatically-discovered NeV behavior models. An Ego failure denotes epistemic doubt (due to uncertainty) in the Ego software under study as it is not able to handle the previously unseen contexts. We showed how developers can leverage those samples of epistemic uncertainty to reduce epistemic gaps in the AV's behavior through robustification techniques, including the reconfiguration of AV software/requirements.

## 5.2 Experimental Findings and Applications

The results of the SavviDriver framework (i.e., failure traces corresponding to human-based uncertainty) can be used in several ways by stakeholders, including AV robustification to human-based uncertainty, revision/addition of requirements, and may even inform regulatory changes. For each use case, the input Ego software initially appeared to have safe behavior with low failure rates. We demonstrated that SavviDriver can be used to train RL-based NeV agents that caused undesirable behaviors in the input Egos, leading to a significant increase in failure rates when exposed to the discovered uncertainty. Finally, we showed several different robustification approaches across our use cases, where the result of the robustification step was an Ego that could mitigate uncertainty, thereby reducing its failure rates when interacting with the discovered NeV behaviors. For learning-based AVs, those failure traces may be used as supplemental training data to improve the robustness of the AV to the discovered uncertainty. For example, in the Weave Lane use case (see Section 4.2), we demonstrated how retraining the learning-based Ego in the presence of the aggressive NeV improved the Ego's performance, reducing the AFR by 16.00%, and reducing the TET by approximately 4.00%. In the Left Turn and ACC use cases (see Section 4.1 and Section 4.3, respectively), we demonstrated how the discovered failure traces could be used by developers to revise the Ego's requirements (and concrete implementation), resulting in reduced AFR and EBS duration, respectively. The discovered failure traces may also be used by regulatory agencies and/or civil engineers to improve roadway safety. For example, in the Left Turn use case, the discovered human-based uncertainty may inform the addition of traffic lights/stop signs to the uncontrolled intersection to prevent similar failures in the real world.

SavviDriver provides developers with a means to model driver behavior and discover human-based uncertainty in simulation. While we used our custom in-house simulator (i.e., TinyRoad) as a proof-of-concept, SavviDriver is not directly coupled to a specific simulation platform. Automotive developers and domain experts can leverage high-fidelity in-house tooling to instantiate specific scenarios and use SavviDriver to discover potential faults in the design of both the vehicle software and/or roadway infrastructure. Using high-fidelity simulators, developers can explore variations of vehicle objectives (i.e., leveraging SavviDriver's modular goal models), vehicle types (e.g., truck, motorcycle, SUV, etc.), and/or operating contexts (e.g., highway ramp merge) to explore different dimensions of human-based uncertainty. By assessing the AV's interaction with different types of drivers in a range of road configurations, SavviDriver can identify edge cases and other undesirable interactions before deployment, thereby facilitating revisions to the requirements, addition of new requirements, or updates to the system under study.

## 6 Related Work

This section overviews related work relevant to SavviDriver. First, we discuss related work with game theory and RL. Next, we discuss relevant research for game-based testing of AVs. Finally, we overview other model-based approaches to RL, uncertainty discovery, or AVs.

A number of researchers have proposed game theory or RL-based approaches to train AV logic or improve the robustness of AVs against uncertainty. Liniger *et al.* [52] proposed a non-cooperative game theory approach to train agents for autonomous racing games, but uses the same reward objective for all agents (i.e., agents compete towards the same goal). Cao *et al.* [17] proposed an imitation learning approach to learn different types of driving models for an AV, and uses RL to learn to swap between the different learned policies. Li *et al.* [93] proposed a cooperative game-theoretic approach to model driver and vehicle interactions, but do not consider human-based uncertainty using a non-cooperative setting and have strictly defined reward functions for RL. Gupta *et al.* [18] introduced a framework to train two dueling agents in an RL setting, training a resilient ego vehicle against a (possibly adversarial) non-ego agent. Zhou *et al.* [94] proposed the UBRL framework, identifying potentially unreliable decisions of an RL agent, but does not account for human-induced uncertainty. Chan *et al.* [7] demonstrated that RL and non-cooperative game theory can be combined to discover undesirable AV behaviors. However, existing works largely use ad hoc development approaches to structure game objectives and/or traffic scenarios. Our work uses goal models to explicitly capture the functional/non-functional objectives of different driving styles and provides a reusable approach to explore human-based uncertainty for AVs through gaming and RL.

Game-based testing has been used to assess and/or improve AVs with respect to uncertainty posed by other drivers. Liu *et al.* [15] propose a multi-agent RL framework to train AVs in the presence of other vehicles, where different vehicle behavior is realized by adjusting the proportion between cooperative and selfish rewards. Hao *et al.* [12] use adversarial games and RL to test AV robustness to vehicles that are trained to cause AV failures (e.g., collisions) while exhibiting behaviors that mimic real-world driving action distributions (i.e., naturalistic priors). Wachi *et al.* [14] combine multi-agent games and adversarial RL [51] to discover failure cases for rule-based vehicles in simulation. Ma *et al.* [13] use level-k game theory [95] and RL to assess and improve an IDM decision-making mechanism (i.e., lane changing) in the presence of vehicles with manually defined reward functions that can exhibit cooperative or competitive behavior. However, game-based testing with RL approaches are often tightly coupled to specific types of vehicle models or a single dimension of human-based uncertainty and use ad-hoc development approaches for exploring different traffic scenarios. In contrast, our work provides a modular framework that uses goal models to declaratively specify human-based driving styles and supports the composition of these models to explore multiple dimensions of human-based uncertainty (i.e., different types and/or combinations of driving styles characterized by functional/non-functional objectives).

Other related work has explored testing AV behaviors in simulation. Dosovitskiy *et al.* [39], Son *et al.* [46], and Zhou *et al.* [47] proposed a number of simulation environments that can be used for AV testing. Birchler *et al.* [96], Zheng *et al.* [97], Zhong *et al.* [98] proposed several methods to generate test cases for an AV in simulation, but do not consider uncertainty from human-induced behaviors. Gambi *et al.* [48] combined procedural content generation and search-based testing to explore AV failure in various automatically generated traffic scenario settings. Humeniuk *et al.* [99] synthesized EC with RL to generate different scenarios to test autonomous systems, where RL is used to seed the EC search. Stocco *et al.* [100] introduced ThirdEye, a white-box AV failure predictor using machine learning that periodically monitors the AV's reliability by identifying instances where the AV may be unconfident. While addressing robustness, they do not address human-induced uncertainty nor support a model-based approach. Fremont *et al.* [101] proposed Scenic, a probabilistic modeling language for specifying and generating a distribution of simulation environments for testing AVs. However, their approach focuses on modeling the environment and does not address means to capture non-functional objectives of agents in the environment.

Other researchers have proposed model-based approaches for AVs or RL. Langford *et al.* [29] introduced MoDALAS, demonstrating a model-based approach to manage and verify the run-time assurance of machine learning components using KAOS goal models and utility functions. However, their approach does not consider human-based non-functional objectives when addressing run-time assurance. Liaskos *et al.* [102] proposed extending the i* framework to model a formal specification for Markov decision processes. Rudolph *et al.* [103] proposed a method to identify and consider strategies of RL agents in the presence of other players for self-adaptation. Jiang *et al.* [104] proposed the THGC model, which uses prior domain knowledge to group agents in a multi-agent RL setting. Bruggner *et al.* [105] proposed a model-based approach to AV simulation, but models the different components of individual autonomous agents (i.e., perception, planning, and control). Leung *et al.* [106] introduced goal modeling for RL agents, where policies are represented as goals. However, their work does not use models to capture the high-level objectives of the agents and their interactions. Schwan *et al.* [55] proposed a goal specification language to formalize a reward function for a given RL task (e.g., drive to destination). In contrast, our approach uses KAOS-inspired goal models to specify RL rewards as a means to train agents that exhibit specific driving styles, including those that are human-based.

Finally, researchers have also proposed techniques that consider the role of humans during the development and deployment of cyber-physical systems (e.g., AVs). Cleland-Huang *et al.* [107] and Camara *et al.* [108] focused on humans as an *internal* and *collaborative* contributor to provide decision-making support and other self-adaptive actions (e.g., MAPE-K actions [109]). Orthogonally, we focus on assessing/improving the robustness of AVs with respect to *external* human-based uncertainty in a *non-cooperative* setting. Gavidia-Calderon *et al.* [110] focused on the detection of different types of humans in the adaptive system's environment, which then informs system adaptations. Our objective, in contrast, is to discover human-based uncertainty-induced *edge cases* that are detrimental to the safe operation of AVs.

## 7 Threats to Validity

This paper described an agent-based goal modeling approach to uncertainty discovery for AVs using non-cooperative game theory and RL. There may be possible deviations between agent behaviors observed in simulation and reality (i.e., "reality gap" [111]). Additionally, agents in simulation environments take synchronous actions, while real traffic interactions are asynchronous, which may contribute to the reality gap. The relevance of the discovered behavior is limited by how well the goal models capture the intended driving styles of vehicles in mixed-traffic interactions. The goal models used in this work were designed by domain experts in the field of automotive assurance, and variations to the KAOS goal models may yield different experimental results [112]. Finally, repeated experiments may lead to or discover different types of agent behaviors or uncertainty, as RL uses stochastic processes to train agents. Additionally, RL algorithms may be sensitive to hyperparameter changes. In our studies, we used established defaults for the hyperparameters [16], with minor changes (e.g., number of training steps) to account for domain-specific/computational resource limits. To ensure the feasibility of the approach, each use case shows the result of an average of 100 episodes for comparison, similar to validation studies in other existing game-based testing approaches [12, 13].

## 8 Conclusion

This paper introduced the SavviDriver framework for using goal modeling to provide a systematic process in the RL-based realization of multi-agent game-based testing to discover unexpected behaviors. SavviDriver is a game-based testing framework that assesses the ability of the AV system to safely interact with various types of human drivers, who may exhibit a range of non-functional driving styles as they achieve their functional goals. Our framework promotes the discovery of interesting and/or unexpected behavior for both the Ego and the NeV(s), which can then be used to assess and improve the robustness of the AV with respect to multiple dimensions of human-based uncertainty.

We applied SavviDriver to three use cases to demonstrate the systematic refinement of high-level human driving styles and objectives into functional and non-functional goals, based on real-world traffic accident data [24, 67]. We implemented the multi-agent game-based testing using RL, based on the reusable goal models, and then demonstrated our framework's ability to discover undesirable behavior in the AV under study. These failures are often caused by selfish, yet non-malicious objectives of the non-ego vehicles. We show that these use cases can discover failures in common challenging road scenes based on existing data from NHTSA and other traffic reports. For the ACC system use case, we were able to use SavviDriver to discover undesirable behaviors, where the failure traces led to a similar set of changes recommended by the OEM during the early deployment stages of the ACC systems [90, 91]

Several directions are being considered for future work. For example, we will explore the training of multiple RL agents with additional driving styles (e.g., timid, intoxicated, etc.) and their composition(s) within a given agent to discover unique strategies of agents and potentially reduce training time. Additional studies may explore how evolutionary computation can be used to explore different gradients and/or combinations of human driving styles (i.e., different levels of aggressiveness or speeding). Other research may investigate the use of procedural content generation to automatically create traffic scenarios for study, including those with various environmental conditions (e.g., slippery road surfaces, road obstacles, potholes, etc.), city-scale traffic environments, and large-scale intelligent transportation systems.

## Acknowledgments

We greatly appreciate contributions from Nick Polanco on our preliminary work. We also thank Joshua E. Siegel and Shaunak D. Bopardikar for the insightful discussions and detailed feedback on our framework. This work was supported in part by funding provided by Michigan State University. Finally, we gratefully acknowledge the detailed feedback provided by the reviewers on earlier versions of this paper.

## References

[1] Chen, B., Sun, D., Zhou, J., Wong, W. & Ding, Z. A future intelligent traffic system with mixed autonomous vehicles and human-driven vehicles. *Information Sciences* **529**, 59–72 (2020).

[2] Rushby, J. Logic and Epistemology in Safety Cases. *Lecture Notes in Computer Science* 1–7 (2013).

[3] Hüllermeier, E. & Waegeman, W. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning* **110**, 457–506 (2021).

[4] Ramirez, A. J., Jensen, A. C. & Cheng, B.H.C.. *A taxonomy of uncertainty for dynamically adaptive systems. 2012 7th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS)*, 99–108 (IEEE, Zurich, Switzerland, 2012).

[5] Wachenfeld, W. & Winner, H. The release of autonomous vehicles. *Autonomous Driving: Technical, Legal and Social Aspects* 425–449 (2016).

[6] Haq, F. U., Shin, D., Nejati, S. & Briand, L. Can Offline Testing of Deep Neural Networks Replace Their Online Testing? *Empirical Software Engineering* **26**, 90 (2021).

[7] Chan, K. H., Zilberman, S., Polanco, N., Siegel, J. E. & Cheng, B.H.C.. *SafeDriveRL: Combining Non-cooperative Game Theory with Reinforcement Learning to Explore and Mitigate Human-based Uncertainty for Autonomous Vehicles. Proceedings of the 19th International Conference on Adaptive and Self-Managing Systems (SEAMS 2024)*, 214–220 (2024). Short Paper.

[8] Yang, K. *et al.* Uncertainties in onboard algorithms for autonomous vehicles: Challenges, mitigation, and perspectives. *IEEE Transactions on Intelligent Transportation Systems* **24**, 8963–8987 (2023).

[9] Burton, S. & Herd, B. Addressing uncertainty in the safety assurance of machine-learning. *Frontiers in Computer Science* **5**, 1132580 (2023).

[10] Chao, Q. *et al.* A survey on visual traffic simulation: Models, evaluations, and applications in autonomous driving **39**, 287–308 (2020).

[11] Nash, J. Non-cooperative games. *Annals of Mathematics* **54**, 286–295 (1951). URL http://www.jstor.org/stable/1969529.

[12] Hao, K. *et al.* Adversarial Safety-Critical Scenario Generation using Naturalistic Human Driving Priors. *IEEE Transactions on Intelligent Vehicles* 1–16 (2023).

[13] Ma, Y. *et al.* Evolving testing scenario generation and intelligence evaluation for automated vehicles. *Transportation Research Part C: Emerging Technologies* **163**, 104620 (2024).

[14] Wachi, A. Failure-scenario maker for rule-based agent using multi-agent adversarial reinforcement learning and its application to autonomous driving (2019).

[15] Liu, W. *et al.* Learning to model diverse driving behaviors in highly interactive

autonomous driving scenarios with multiagent reinforcement learning. *IEEE Systems Journal* (2025).

[16] Schulman, J., Wolski, F., Dhariwal, P., Radford, A. & Klimov, O. Proximal Policy Optimization Algorithms (2017). arxiv:arXiv:1707.06347.

[17] Cao, Z. *et al. Reinforcement learning based control of imitative policies for near-accident driving. Robotics: Science and Systems* (2020).

[18] Gupta, P., Coleman, D. & Siegel, J. E. Towards physically adversarial intelligent networks (pains) for safer self-driving. *IEEE Control Systems Letters* **7**, 1063–1068 (2022).

[19] Folkers, A., Rick, M. & Büskens, C. *Controlling an Autonomous Vehicle with Deep Reinforcement Learning. 2019 IEEE Intelligent Vehicles Symposium (IV)*, 2025–2031 (2019). arxiv:1909.12153.

[20] IEEE. *A review of reward functions for reinforcement learning in the context of autonomous driving.*

[21] Ramirez, A. J., Jensen, A. C., Cheng, B.H.C.. & Knoester, D. B. *Automatically exploring how uncertainty impacts goal satisfaction. Proceedings of International Conference on Automated Software Engineering (ASE 2011)*, 568–571 (2011).

[22] Dijkstra, E. W. On the role of scientific thought. *Selected writings on computing: a personal perspective* 60–66 (1982).

[23] Parnas, D. L. On the criteria to be used in decomposing systems into modules. *Communications of the ACM* **15**, 1053–1058 (1972).

[24] Media, NHTSA. Driving behaviors reported for drivers and motorcycle operators involved in fatal crashes. https://www.iii.org/fact-statistic/facts-statistics-aggressive-driving (2021).

[25] Media, NHTSA. Nhtsa estimates for 2022 show roadway fatalities remain flat after two years of dramatic increases. https://www.nhtsa.gov/press-releases/traffic-crash-death-estimates-2022 (2023).

[26] Richard, C. M., Campbell, J. L., Brown, J. L. *et al.* Task analysis of intersection driving scenarios: Information processing bottlenecks. Tech. Rep., Turner-Fairbank Highway Research Center (2006).

[27] van Lamsweerde, A. *Goal-oriented requirements engineering: A guided tour. Proceedings Fifth IEEE International Symposium on Requirements Engineering*, 249–262 (2001).

[28] Walsh, W. E., Tesauro, G., Kephart, J. O. & Das, R. *Utility functions in autonomic systems. International Conference on Autonomic Computing, 2004. Proceedings.*, 70–77 (IEEE, 2004).

[29] Langford, M. A., Chan, K. H., Fleck, J. E., McKinley, P. K. & Cheng, B.H.C.. *Modalas: Model-driven assurance for learning-enabled autonomous systems. 2021 ACM/IEEE 24th International Conference on Model Driven Engineering Languages and Systems (MODELS)*, 182–193 (IEEE, 2021).

[30] Thwarted on the On-ramp: Waymo Driverless Car Doesn't Feel the Urge to Merge. https://www.thetruthaboutcars.com/2018/05/thwarted-ramp-waymo-driverless-car-doesnt-feel-urge-merge/ (2018).

[31] Bradley, R. Tesla Autopilot. https://www.technologyreview.com/technology/tesla-autopilot/.

[32] Freedman, I. G., Kim, E. & Muennig, P. A. Autonomous vehicles are cost-effective when used as taxis. *Injury epidemiology* **5**, 1–8 (2018).

[33] Us, Y. Overcoming deployment hurdles: Adastec's approach to automated bus integration (2023). URL https://www.adastec.com/post/overcoming-deployment-hurdles.

[34] Team, T. W. Waymo significantly outperforms comparable human benchmarks over 7+ million miles of rider-only driving. https://waymo.com/blog/2023/12/waymo-significantly-outperforms-comparable-human-benchmarks-over-7-million.

[35] Koopman, P., Osyk, B. & Weast, J. *Autonomous vehicles meet the physical world: Rss, variability, uncertainty, and proving safety. Computer Safety, Reliability, and Security: 38th International Conference, SAFECOMP 2019, Turku, Finland, September 11–13, 2019, Proceedings 38*, 245–253 (Springer, 2019).

[36] Sagberg, F., Selpi, Bianchi Piccinini, G. F. & Engström, J. A review of research on driving styles and road safety. *Human factors* **57**, 1248–1275 (2015).

[37] Paleti, R., Eluru, N. & Bhat, C. R. Examining the influence of aggressive driving behavior on driver injury severity in traffic crashes. *Accident Analysis & Prevention* **42**, 1839–1854 (2010).

[38] Antić, B., Čabarkapa, M., Čubranić-Dobrodolac, M. & Čičević, S. The influence of aggressive driving behavior and impulsiveness on traffic accidents. *US Transportation Collection* (2018).

[39] Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A. & Koltun, V. *Carla: An open urban driving simulator. Conference on robot learning*, 1–16 (PMLR, 2017).

[40] BeamNG GmbH. BeamNG.tech. URL https://www.beamng.tech/.

[41] Bernhard, J., Esterle, K., Hart, P. & Kessler, T. *Bark: Open behavior benchmarking in multi-agent environments. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2020). URL https://arxiv.org/pdf/2003.02604.pdf.

[42] Leurent, E. An environment for autonomous driving decision-making. https://github.com/eleurent/highway-env (2018).

[43] Menon, C. & Alexander, R. A safety-case approach to the ethics of autonomous vehicles. *Safety and reliability* **39**, 33–58 (2020).

[44] Chowdhury, T., Wassyng, A., Paige, R. F. & Lawford, M. *Systematic evaluation of (safety) assurance cases. International Conference on Computer Safety, Reliability, and Security*, 18–33 (Springer, 2020).

[45] Mohamad, M., Åström, A., Askerdal, Ö., Borg, J. & Scandariato, R. *Security assurance cases for road vehicles: An industry perspective. Proceedings of the 15th International Conference on Availability, Reliability and Security*, 1–6 (2020).

[46] Son, T. D., Bhave, A. & Van der Auweraer, H. *Simulation-based testing framework for autonomous driving development. 2019 IEEE International Conference on Mechatronics (ICM)*, Vol. 1, 576–583 (IEEE, 2019).

[47] Zhou, L. *et al. Garchingsim: An autonomous driving simulator with photorealistic scenes and minimalist workflow. 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, 4227–4232 (IEEE, 2023).

[48] Gambi, A., Mueller, M. & Fraser, G. *Automatically testing self-driving cars with search-based procedural content generation. Proceedings of the 28th ACM SIGSOFT International Symposium on Software Testing and Analysis*, 318–328 (2019).

[49] Leventhal, L. M., Teasley, B. M., Rohlman, D. S. & Instone, K. *Positive test bias in software testing among professionals: A review. Human-Computer Interaction: Third International Conference, EWHCI'93 Moscow, Russia, August 3–7, 1993 Selected Papers 3*, 210–218 (Springer, 1993).

[50] Rapoport, A. *Game theory as a theory of conflict resolution* Vol. 2 (Springer Science & Business Media, 2012).

[51] Pinto, L., Davidson, J., Sukthankar, R. & Gupta, A. *Robust adversarial reinforcement learning. International conference on machine learning*, 2817–2826 (PMLR, 2017).

[52] Liniger, A. & Lygeros, J. A noncooperative game approach to autonomous racing. *IEEE Transactions on Control Systems Technology* **28**, 884–897 (2019).

[53] Arulkumaran, K., Deisenroth, M. P., Brundage, M. & Bharath, A. A. Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine* **34**, 26–38 (2017).

[54] Kiran, B. R. *et al.* Deep reinforcement learning for autonomous driving: A survey. *IEEE transactions on intelligent transportation systems* **23**, 4909–4926 (2021).

[55] Schwan, S., Klös, V. & Glesner, S. *A goal-oriented specification language for reinforcement learning. International Conference on Modeling Decisions for Artificial Intelligence*, 169–180 (Springer, 2023).

[56] Kaelbling, L. P., Littman, M. L. & Moore, A. W. Reinforcement learning: A survey. *Journal of artificial intelligence research* **4**, 237–285 (1996).

[57] Tzorakoleftherakis, E. Reinforcement Learning: A Brief Guide. https://www.mathworks.com/company/technical-articles/reinforcement-learning-a-brief-guide.html (2019).

[58] Konda, V. & Tsitsiklis, J. Actor-critic algorithms. *Advances in neural information processing systems* **12** (1999).

[59] Elsayed, M., Lan, Q., Lyle, C. & Mahmood, A. R. Weight clipping for deep continual and reinforcement learning. *arXiv preprint arXiv:2407.01704* (2024).

[60] Williams, R. J. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning* **8**, 229–256 (1992).

[61] Mnih, V. *et al. Asynchronous methods for deep reinforcement learning. International conference on machine learning*, 1928–1937 (PMLR, 2016).

[62] Sutton, R. S., Barto, A. G. *et al.* Reinforcement learning. *Journal of Cognitive Neuroscience* **11**, 126–134 (1999).

[63] Garlan, D., Cheng, S.-W., Huang, A.-C., Schmerl, B. & Steenkiste, P. Rainbow: Architecture-based self-adaptation with reusable infrastructure. *Computer* **37**, 46–54 (2004).

[64] Ramirez, A. J. & Cheng, B.H.C.. *Automatic derivation of utility functions for monitoring software requirements. International Conference on Model Driven Engineering Languages and Systems*, 501–516 (Springer, 2011).

[65] DeGrandis, P. & Valetto, G. *Elicitation and utilization of application-level utility functions*. *Proceedings of the 6th international conference on Autonomic computing*, 107–116 (2009).

[66] Fredericks, E. M. *Automatically hardening a self-adaptive system against uncertainty*. *Proceedings of the 11th International Symposium on Software Engineering for Adaptive and Self-Managing Systems*, SEAMS '16, 16–27 (Association for Computing Machinery, New York, NY, USA, 2016). URL http://doi.org/10.1145/2897053.2897059.

[67] Media, NHTSA. Distracted driving. https://www.nhtsa.gov/risky-driving/distracted-driving (2021).

[68] Gross, A. Risky business – more than half of all drivers engage in dangerous behavior. https://newsroom.aaa.com/2023/11/risky-business-more-than-half-of-all-drivers-engage-in-dangerous-behavior/ (2023).

[69] Media, A. Likelihood of aggressive driving among u.s. drivers (2019). URL https://exchange.aaa.com/safety/driving-advice/aggressive-driving/.

[70] Mergia, W. Y., Eustace, D., Chimba, D. & Qumsiyeh, M. Exploring factors contributing to injury severity at freeway merging and diverging locations in ohio. *Accident Analysis & Prevention* **55**, 202–210 (2013).

[71] Luo, Y. *et al.* *A Driving Intention Prediction Method for Mixed Traffic Scenarios*. *2022 IEEE 7th International Conference on Intelligent Transportation Engineering (ICITE)*, 296–301 (IEEE, Beijing, China, 2022).

[72] Fujiwara-Greve, T. in *Nash Equilibrium* (ed.Fujiwara-Greve, T.) *Non-Cooperative Game Theory* Monographs in Mathematical Economics, 23–55 (Springer Japan, Tokyo, 2015).

[73] Fudenberg, D. & Tirole, J. Noncooperative game theory for industrial organization: an introduction and overview. *Handbook of industrial Organization* **1**, 259–327 (1989).

[74] Lanctot, M. *et al.* *A unified game-theoretic approach to multiagent reinforcement learning*. *Advances in Neural Information Processing Systems*, Vol. 30 (Curran Associates, Inc., 2017). URL https://proceedings.neurips.cc/paper_files/paper/2017/file/3323fe11e9595c09af38fe67567a9394-Paper.pdf.

[75] Huck, T. P., Ledermann, C. & Kröger, T. *Simulation-based testing for early safety-validation of robot systems*. *2020 IEEE Symposium on Product Compliance Engineering-(SPCE Portland)*, 1–6 (IEEE, 2020).

[76] Koopman, P. & Wagner, M. Toward a framework for highly automated vehicle safety validation. Tech. Rep., SAE Technical Paper (2018).

[77] Abbas, H., O'Kelly, M., Rodionova, A. & Mangharam, R. *Safe at any speed: A simulation-based test harness for autonomous vehicles*. *Cyber Physical Systems. Design, Modeling, and Evaluation: 7th International Workshop, CyPhy 2017, Seoul, South Korea, October 15-20, 2017, Revised Selected Papers 7*, 94–106 (Springer, 2019).

[78] Treiber, M., Hennecke, A. & Helbing, D. Congested traffic states in empirical observations and microscopic simulations. *Physical review E* **62**, 1805 (2000).

[79] Types of Unsignalized Intersections - Unsignalized Intersection Improvement Guide. https://toolkits.ite.org/uiig/types.aspx.

[80] National Highway Traffic Safety Administration. Query of fatality analysis reporting system (fars) web-based encyclopedia. https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars (2014). Accessed: 2014-12-04.

[81] Liberatore, S. Self-driving race car crashes into a wall from the starting line. https://www.dailymail.co.uk/sciencetech/article-8899021/Self-driving-race-car-crashes-straight-wall-starting-line-Roborace.html (2020).

[82] Kong, J., Pfeiffer, M., Schildbach, G. & Borrelli, F. *Kinematic and dynamic vehicle models for autonomous driving control design. 2015 IEEE intelligent vehicles symposium (IV)*, 1094–1099 (IEEE, 2015).

[83] Koenig, N. & Howard, A. *Design and use paradigms for gazebo, an open-source multi-robot simulator. 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, Vol. 3, 2149–2154 (IEEE, 2004).

[84] Brockman, G. *et al.* Openai gym. *arXiv preprint arXiv:1606.01540* (2016).

[85] Westhofen, L. *et al.* Criticality Metrics for Automated Driving: A Review and Suitability Analysis of the State of the Art. *Archives of Computational Methods in Engineering* **30**, 1–35 (2023).

[86] Minderhoud, M. M. & Bovy, P. H. Extended time-to-collision measures for road traffic safety assessment. *Accident Analysis & Prevention* **33**, 89–97 (2001).

[87] Saffarzadeh, M., Nadimi, N., Naseralavi, S. & Mamdoohi, A. R. *A general formulation for time-to-collision safety indicator. Proceedings of the Institution of Civil Engineers-Transport*, Vol. 166, 294–304 (Thomas Telford Ltd, 2013).

[88] Choi, E.-H. Crash Factors in Intersection-Related Crashes: An On-Scene Perspective: (621942011-001) (2010).

[89] Langford, M. A., Zilberman, S. & Cheng, B.H.C.. Anunnaki: A modular framework for developing trusted artificial intelligence. *ACM Trans. Auton. Adapt. Syst.* (2024). URL https://doi-org.proxy2.cl.msu.edu/10.1145/3649453.

[90] Barry, K. Guide to Adaptive Cruise Control. https://www.consumerreports.org/cars/car-safety/guide-to-adaptive-cruise-control-a9154580873/ (2022).

[91] Company, F. M. Adaptive cruise control - setting the adaptive cruise control gap (2021). URL https://www.fordservicecontent.com/Ford_Content/vdirsnet/OwnerManual/Home/Content?variantid=7505&languageCode=en&countryCode=USA&Uid=G2047539&ProcUid=G2045342&userMarket=USA&div=f&vFilteringEnabled=False&buildtype=web.

[92] Langford, M. A. & Cheng, B.H.C.. Enki: A Diversity-driven Approach to Test and Train Robust Learning-enabled Systems. *ACM Transactions on Autonomous and Adaptive Systems* **15**, 1–32 (2020).

[93] Li, N. *et al.* Game theoretic modeling of driver and vehicle interactions for verification and validation of autonomous vehicle control systems. *IEEE Transactions on control systems technology* **26**, 1782–1797 (2017).

[94] Zhou, W., Cao, Z., Deng, N., Jiang, K. & Yang, D. Identify, estimate and bound the uncertainty of reinforcement learning for autonomous driving. *IEEE Transactions on Intelligent Transportation Systems* **24**, 7932–7942 (2023).

[95] Peters, H. *Game theory: A Multi-leveled approach* (Springer, 2015).

[96] Birchler, C., Khatiri, S., Bosshard, B., Gambi, A. & Panichella, S. Machine learning-based test selection for simulation-based testing of self-driving cars software. *Empirical Software Engineering* **28**, 71 (2023).

[97] Zheng, X. *et al.* *Rapid generation of challenging simulation scenarios for autonomous vehicles based on adversarial test. 2020 IEEE International Conference on Mechatronics and Automation (ICMA)*, 1166–1172 (IEEE, 2020).

[98] Zhong, Z., Kaiser, G. & Ray, B. Neural network guided evolutionary fuzzing for finding traffic violations of autonomous vehicles. *IEEE Transactions on Software Engineering* (2022).

[99] Humeniuk, D., Khomh, F. & Antoniol, G. Reinforcement learning informed evolutionary search for autonomous systems testing. *ACM Transactions on Software Engineering and Methodology* **33**, 1–45 (2024).

[100] Stocco, A., Nunes, P. J., d'Amorim, M. & Tonella, P. *Thirdeye: Attention maps for safe autonomous driving systems. Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering*, 1–12 (2022).

[101] Fremont, D. J. *et al.* Scenic: A language for scenario specification and data generation. *Machine Learning* **112**, 3805–3849 (2023).

[102] Liaskos, S., Khan, S. M., Golipour, R. & Mylopoulos, J. *Towards goal-based generation of reinforcement learning domain simulations. iStar*, 22–28 (2022).

[103] Rudolph, S., Tomforde, S. & Hähner, J. On the detection of mutual influences and their consideration in reinforcement learning processes. *arXiv preprint arXiv:1905.04205* (2019).

[104] Jiang, H. *et al.* Multi-agent deep reinforcement learning with type-based hierarchical group communication. *Applied Intelligence* **51**, 5793–5808 (2021).

[105] Bruggner, D., Hegde, A., Acerbo, F. S., Gulati, D. & Son, T. D. *Model in the loop testing and validation of embedded autonomous driving algorithms. 2021 IEEE Intelligent Vehicles Symposium (IV)*, 136–141 (IEEE, 2021).

[106] Leung, J., Shen, Z., Zeng, Z. & Miao, C. *Goal modelling for deep reinforcement learning agents. Machine Learning and Knowledge Discovery in Databases. Research Track: European Conference, ECML PKDD 2021, Bilbao, Spain, September 13–17, 2021, Proceedings, Part I 21*, 271–286 (Springer, 2021).

[107] Cleland-Huang, J. *et al.* Human–machine Teaming with Small Unmanned Aerial Systems in a MAPE-K Environment. *ACM Transactions on Autonomous and Adaptive Systems* **19**, 1–35 (2024).

[108] Cámara, J., Moreno, G. & Garlan, D. Reasoning about Human Participation in Self-Adaptive Systems. *2015 IEEE/ACM 10th International Symposium on Software Engineering for Adaptive and Self-Managing Systems* 146–156 (2015).

[109] Arcaini, P., Riccobene, E. & Scandurra, P. *Modeling and analyzing MAPE-K feedback loops for self-adaptation. 2015 IEEE/ACM 10th International Symposium on Software Engineering for Adaptive and Self-Managing Systems*, 13–23 (IEEE, 2015).

[110] Gavidia-Calderon, C., Kordoni, A., Bennaceur, A., Levine, M. & Nuseibeh, B. The IDEA of Us: An Identity-Aware Architecture for Autonomous Systems. *ACM Transactions on Software Engineering and Methodology* **33**, 1–38 (2024).

[111] Combemale, B. *et al.* A Hitchhiker's Guide to Model-Driven Engineering for Data-Centric Systems. *IEEE Software* **38**, 71–84 (2021).

[112] Franch, X. The i* framework: The way ahead. *2012 Sixth International Conference on Research Challenges in Information Science (RCIS)* 1–3 (2012).