

Reinforcement learning-based optimal control of wearable alarms for consistent roadway workers' reactions to traffic hazards

Daniel Lu, Semiha Ergan, and Kaan Ozbay

This is the author's version of a work that has been accepted for publication in the Journal of Transportation Safety & Security. The final version can be found published online on January 9, 2025, <https://www.tandfonline.com/doi/full/10.1080/19439962.2024.2449119>.

Lu, D., Ergan, S., & Ozbay, K. (2025). Reinforcement learning-based optimal control of wearable alarms for consistent roadway workers' reactions to traffic hazards. Journal of Transportation Safety & Security, 17(7), 757–781.
<https://doi.org/10.1080/19439962.2024.2449119>

Reinforcement learning-based optimal control of wearable alarms for consistent roadway workers' reactions to traffic hazards

Daniel Lu^a, Semiha Ergan^{a*}, and Kaan Ozbay^a

^aDepartment of Civil and Urban Engineering, New York University, Brooklyn, United States

* semiha@nyu.edu (Corresponding author)

Reinforcement learning-based optimal control of wearable alarms for consistent roadway workers' reactions to traffic hazards

Recent innovations in roadway construction include work zone intrusion alert (WZIA) systems that detect traffic hazards (e.g., speeding or intruding vehicles) and raise alarms (e.g., sounds, lights) using preset attributes (e.g., volume, duration) to warn human workers-on-foot. Designing alarms raised by wearable warning devices (e.g., smartwatch) for roadway workers remains an emerging area of transportation safety research. As roadway work zones begin to adopt these novel technologies, issues relating to the alarms may persist. Different individuals' alarm preferences and alarm fatigue towards repeated exposure to constant alarm attributes can lead to decreases in worker vigilance and responsiveness to traffic hazard. Reinforcement learning (RL)-based controls can adjust alarm attributes in real-time to counteract these issues, ensuring consistent worker reactions. This study proposes an RL-based approach to train alarm control agents using different reward functions that prioritize different worker reactions (e.g., body movement, head turn). Results show that a reward function with equal weight for each type of worker reaction produces an alarm agent that ensures consistent and safe worker reactions to traffic hazards. Findings also inform the future development of RL-based alarms (i.e., fine-tuning rewards) to counteract the lack of safe worker reactions observed in real-world work zones.

Keywords: roadway worker safety; alarm fatigue; reinforcement learning; virtual reality; wearable warning device

Subject classification codes: ITS, Human Factor, Simulation, Traffic Injury

1. Introduction

The U.S. Department of Transportation estimates that construction workers account for about 20% of pedestrian deaths in crashes occurring in roadway work zones (ARTBA, 2024).

Current work zone safety measures include deploying traffic control devices (e.g., barrels, signage) in standardized site layouts to alert drivers approaching work zones (Federal Highway Administration, 2022). Recently developed work zone intrusion alert (WZIA) systems use sensors (e.g., radar, pneumatic tubes) to detect traffic hazards (e.g., speeding or

intruding vehicles) breaching work zone boundaries and trigger alarms (e.g., loudspeakers, flashing lights) to warn roadway workers, but have found that workers are less likely to notice and react quickly the farther they stand from the WZIA system's stationary alarm source (Thapa & Mishra, 2021).

To address these limitations, wearable devices (e.g., smartwatches, safety vests) raise alarm patterns using audio, visual, or haptic modalities. *Alarm patterns* (i.e., signal profile) are defined by attributes like modality (i.e., human senses engaged), duration (i.e., continuous active period), repetitions, and pauses (i.e., inactive period). Workers may react to an alarm pattern by moving, turning their heads, or pressing a device button (Gambatese et al., 2017; Qin et al., 2022; Thapa & Mishra, 2021). Transportation safety research is exploring how to optimize these alarm attributes to enhance worker responses to traffic hazards (Lordianto et al., 2024; Sabeti et al., 2024; Yang & Roofigari-Esfahan, 2023).

As these wearable warning devices and optimized alarm attributes become implemented in roadway work zones, human factors and occupational safety challenges may persist regarding the consistency of workers' alarm reactions. Different workers may have varied individual preferences for specific alarm patterns (Haghighi et al., 2020; T. Li et al., 2022; Papachristos et al., 2020; Seifi et al., 2014). As observed in various human worker fields (e.g., nursing, general construction), repeated exposure to the same patterns can lead to alarm fatigue, reducing vigilance toward safety hazards (Blackmon & Gramopadhye, 1995; Camci et al., 2020; Chae & Kang, 2021; Lee et al., 2019). This phenomenon suggests that constantly raising the same alarm pattern can even cause workers to become less vigilant and responsive (i.e., desensitize) towards traffic hazards over time. Therefore, there is a need to adapt the alarm patterns according to each individual roadway worker's response over the course of their construction workday, ensuring their consistent reactions for their life safety (i.e., prevent injury or death).

State-of-the art algorithms, such as reinforcement learning (RL), have proven to be robust approaches to optimizing control systems in dynamic (i.e., nonstationary) environments, such as video games and robotic controls, without requiring prior domain knowledge (Mnih et al., 2013; Silver et al., 2016). Given the uncertainties in worker preferences and alarm fatigue, controlling alarm patterns is a non-stationary problem (i.e., fixed rules for actions may not perform well), making RL well-suited to address the future challenges of wearable alarms in roadway worker safety. RL-based alarm controls can potentially ensure workers react more consistently to both alarm patterns and the traffic hazards that trigger them.

Implementing RL-based alarm control in wearable devices for roadway work zones faces key challenges, especially in developing the control system without endangering workers. Limited work zone safety data leave critical components, like the RL agent *state* and *reward function*, undefined. Yet the reward function significantly impacts how RL-based alarms promote specific worker reactions to traffic hazards. A basic approach would be to equally reward all types of worker reactions (e.g., body movement= +1, head turn= +1). But ideally, rewards should reflect the relative importance of specific reaction based on empirical evidence relating to worker safety risks. With such detailed safety data unavailable, different RL reward functions that prioritize each type of reaction should be evaluated to confirm if roadway workers perform the corresponding type of reaction more frequently. This sensitivity analysis can then guide how the reward function should be fine-tuned once more data on real-world worker reactions to traffic hazard alarms are observed.

Towards deploying RL-based alarms to ensure roadway worker safety from traffic hazards, this study leverages wearable alarm user studies with human workers on an integrated virtual reality (VR) and traffic simulation platform to provide a data-driven environment for training RL alarm agents. Results analyze how different RL reward

functions produce agents that raise different alarm patterns at different frequencies, which ultimately impacts the frequency of different types of worker reactions (e.g., body move, head turn). The main contribution of this paper is demonstrating a methodology for training an RL agent controlling alarm patterns to encourage specific types of worker reactions based on the agent's specific reward function definition prioritizing those reactions. This methodology informs the future engineering process (i.e., fine tuning reward weights) of RL-based alarm control systems for promoting safe and consistent roadway worker reactions to traffic hazards.

2. Background

Reinforcement learning (RL) is a machine learning paradigm distinct from supervised and unsupervised learning, training agents through interactions with dynamic environments and delayed reward (Mnih et al., 2013; Silver et al., 2016; Sutton & Barto, 2022). Studies focusing on RL with human feedback (RL-HF) optimize actions around uncertainties in human behaviour and has notably improved chatbot responses (e.g., ChatGPT, Claude) with feedback from human user ratings (e.g., thumbs up/down) (Lambert et al., 2023). While RL-HF has not yet been applied to alarm controls for human workers, some studies have trained RL agents to convey safety risks through haptic feedback in VR games for human players in VR games (Brenneis et al., 2022; Pilarski et al., 2022).

A key challenge in RL-HF is the need for large-scale human training data, either addressed by standalone *offline* simulations of humans in RL training environment (i.e., self-play) (Silver et al., 2016) or *online* training with real human users (Pilarski et al., 2022). Offline RL is typically more practical (i.e., only run computer simulations) but requires a realistic representation of human behaviour for training. Online RL, on the other hand, can have significant costs and safety risks (e.g., healthcare, autonomous vehicles) (Figueiredo

Prudencio et al., 2024). Recognizing this tradeoff, this study trains an RL alarm agent *offline* in a data-driven environment with a machine learning-based *worker model* replicating how human workers behaved in a lab-controlled VR-traffic simulation platform. This *offline* data-driven environment approach can serve as an effective “pre-training” of RL alarm agents before they continue learning *online* with real human workers (i.e., offline-to-online RL).

2.1 RL Applications in Construction and Traffic Safety

Previous RL studies in transportation safety have addressed traffic control and autonomous vehicles (Du et al., 2023; Lichtlé et al., 2023), while construction safety studies have focused on heavy equipment (e.g., cranes) and robotics (e.g., drones) (Amani & Akhavian, 2024; Asghari et al., 2022; Cai et al., 2023). Although RL is widely applied in construction for energy efficiency (e.g., building systems) and labor productivity (e.g., robotic bricklaying), studies addressing human worker safety are rare (Asghari et al., 2022). When research studies do consider human safety, RL typically optimizes the movement of hazards, such as preventing vehicle-pedestrian collisions (Lichtlé et al., 2023) or avoiding worker-equipment accidents (Cai et al., 2023; Wang et al., 2024).

However, RL applications for controlling alarms for human workers’ safety remain unexplored. The closest related construction worker safety study used VR to simulate physical labor tasks (e.g., lifting) and proposed using RL to provide visual warnings (e.g., color-coded figures, text) to improve worker posture (Akanmu et al., 2020). Akanmu et al. did not implement or test the performance of RL algorithms, citing a need for significant data from construction worker behaviour. Recognizing these challenges, this study collects extensive data (N=255) on different alarm reactions from human workers (n=22) in a simulated roadway work zone (i.e., VR-based). A machine learning *worker model* based on that collected data enables this study to run numerous episodes (i.e., bootstrapping) for

training RL alarm agents offline. This approach is an effective compromise between the ethical challenges of training RL in real-world work zones and time-consuming human subject testing within controlled lab settings.

2.2 Action-Value RL Methods

While advanced RL algorithms are widely applied in video games and transportation systems, this study investigates the feasibility of RL with the *contextual multi-armed bandit* for roadway worker safety alarm controls. The contextual bandit is a widely used form of associative *action-value* methods in RL (Sutton & Barto, 2022). In its most generic form, an action-value RL agent learns to estimate the expected reward, r_t , that results from taking each possible action, a_t , given the environment state, \mathbf{s}_t , at timestep, t . The agent then chooses the greedy (i.e., optimal) action based on which action has the highest expected reward, as presented in equation (1) (Sutton & Barto, 2022).

$$a_t = \underset{a}{\operatorname{argmax}}(\mathbb{E}[r_t | \mathbf{s}_t, a_t]) \quad (1)$$

The term “multi-armed bandit” refers to an agent that tries to play multiple slot machines with unknown payouts (reward) and learns via trial-and-error which slot machine arm levers (actions) to pull to earn the highest payout (i.e., maximize reward). A simple multi-armed bandit (no context) focuses only on the immediate reward, r_t , observed after an action, a_t , and ignores the state, \mathbf{s}_t . Contextual bandits extend this approach by also associating the immediate reward to both the action, a_t , and state, \mathbf{s}_t (i.e., context). For example, in this study, the contextual bandit will select among 12 possible alarm patterns ($1 \leq a_t \leq 12$) based on a state vector, \mathbf{s}_t , of numerical features representing aspects of a roadway worker’s safety, such as the safety (0 or 1) of their work zone position and safety (0

or 1) of their gaze direction (i.e., where they are looking), resulting in a vector like [1, 0] for *safe* position and *unsafe* gaze direction.

The contextual bandit with an upper confidence bound and linear payoff model (LinUCB) estimates rewards based a computed ceiling (upper bound) on expected rewards for all actions. Equation (2) below illustrates the LinUCB bandit's greedy action, where $\mathbf{x}_{t,a}$ represents the context that the bandit considers when selecting an action, a_t .

$$a_t = \underset{a}{\operatorname{argmax}} (\mathbf{x}_{t,a}^\top \hat{\boldsymbol{\theta}}_a + \alpha \sqrt{\mathbf{x}_{t,a}^\top \mathbf{A}_a^{-1} \mathbf{x}_{t,a}}) \quad (2)$$

In simple cases, the context term, $\mathbf{x}_{t,a}$ in equation (2), can be the same as the state, \mathbf{s}_t , in equation (1). For example, earlier works have proven contextual bandits can excel in recommending news article links (actions) based on user clicks (reward) (L. Li et al., 2010; Tang et al., 2014). For their context, $\mathbf{x}_{t,a}$, those earlier works defined a feature vector incorporating displayed article topics (e.g., editor tags) to consider when recommending alternative articles. Details on how the other parameters in Equation 2 can be found in earlier publications (L. Li et al., 2010; Sutton & Barto, 2022).

Although less advanced than deep RL methods (e.g., DQN, PPO), contextual bandits have been proven effective in selecting discrete actions (e.g., web article links) based on human behavior (e.g., user clicks). This aligns with this study's goal of adapting alarm patterns to changing worker reactions (e.g., body movements, head turns). A prior study by the research team found contextual bandits more interpretable than deep RL for alarm control (Lu et al., 2024). Given the novelty of RL in roadway worker safety, defining the RL alarm agent's reward function is a critical challenge for the transportation industry. The study examines the impact of different reward functions using only the contextual bandit as the RL alarm agent in question.

3. Research Method

The study proposes a methodology consisting of two major phases (Figure 1). Phase 1 collects data on how human roadway workers behave in roadway work zones and react to wearable alarm patterns triggered by traffic hazards (e.g., speeding or intruding vehicles). This phase leverages an integrated VR-traffic microsimulation platform to safely collect data on workers in otherwise life-threatening scenarios (see Section 3.1). The collected data is then used to develop a worker model replicating how humans behaved in the work zone and reacted to different alarm patterns. Phase 2 involves training RL agents (i.e., contextual multi-arm bandit) to optimize the selection of alarm patterns for roadway worker safety and conducting sensitivity analysis of various RL reward functions on the frequency of different types of worker reactions (body move, head turn) (see Section 3.2). With the use of statistical and machine learning models replicating the human worker alarm reactions, the roadway worker model constitutes as part of the data-driven environment for training the RL alarm agent (see Appendix for details).

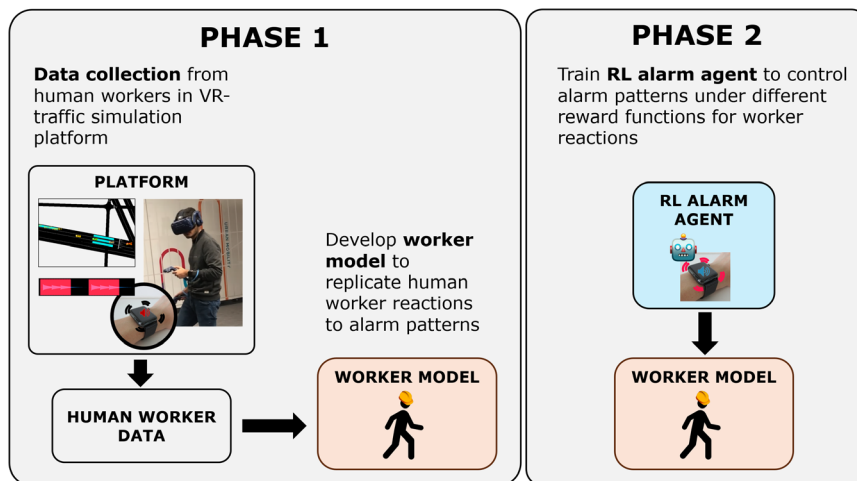


Figure 1. Overview of research methodology for training RL alarm agent.

3.1 Data collection on human workers with wearable alarms in immersive VR-traffic microsimulation platform

Phase 1 of this study leverages an integrated VR-traffic microsimulation platform developed

in prior work by the research team, utilizing Simulation of Urban Mobility (SUMO) simulations to synchronize realistic traffic flows around roadway work zones modelled in Unity VR (Ergan et al., 2022). Given practical and ethical challenges in data collection on human workers reactions in real traffic hazard scenarios (i.e., testing workers reacting to real speeding cars), this platform enables the safe testing of alarms on human workers within an immersive high-fidelity virtual environment resembling roadway work zones. A prior pilot study by the research team conducted wearable alarm user studies on the platform and devised methods of analysing worker reactions (body move, head turn, watch press) to alarm patterns with varied attributes (Qin et al., 2022). Building on this preliminary analysis of alarm attributes, this study collects data from human workers on the VR-traffic platform to develop sophisticated controls that vary those alarm attributes based on worker reactions. While this study uses VR-based human worker data, ongoing real-world tests have shown no statistically significant differences in response times, suggesting the data collected here for training RL agents is representative of natural human reactions (Zhang et al., 2024).

Under IRB IRB-FY2020-3946, the research team invited human participants to wearable alarm user studies on the VR-traffic simulation platform to capture data on how workers react to randomized alarm patterns while conducting roadway work zone construction activities (Figure 2). Each participant in the interactive VR simulation performed a typical polyvinylidene difluoride (PVDF) traffic sensor installation on an urban highway (e.g., saw cutting pavement, installing cable inside saw-cut). During these tasks, SUMO randomly generates traffic hazards (e.g., speeding or intruding vehicles) to trigger one of twelve distinct alarm patterns (see Figure 3) with different modalities, durations, and repetitions on a wearable warning device (e.g., smartwatch). Data collected from both how human workers *behaved* (i.e., while doing sensor installation without alarms) and *reacted* to

the alarms form the basis of a data-driven roadway worker model used later to train the RL alarm agent in this study (see Table 1).

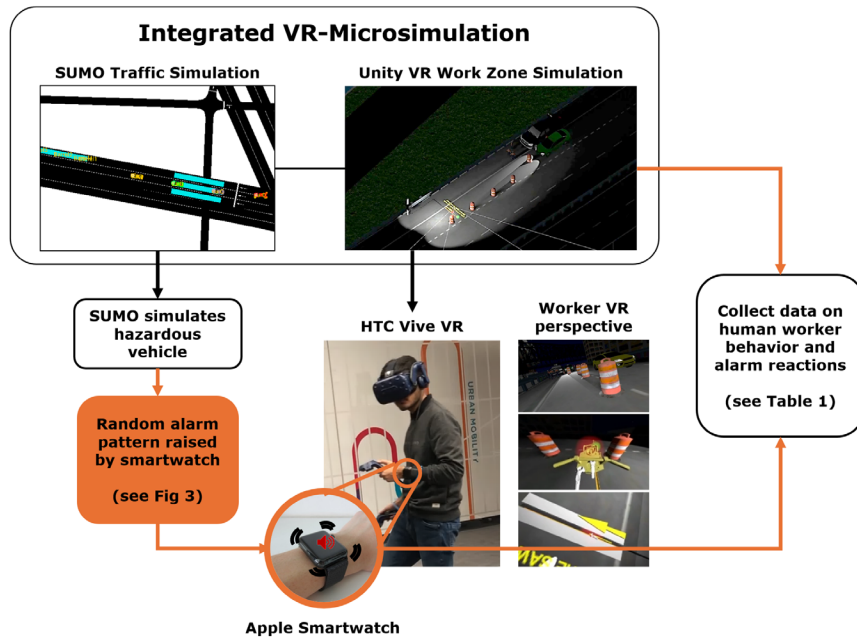


Figure 2. Integrated VR-traffic microsimulation platform for collecting data (see Table 1) on how human roadway workers react to wearable alarm patterns received (see Figure 3).

The twelve alarm patterns encompass three different modalities (haptics only, sound only, and sound and haptics combined), two different durations (60ms and 350 ms) and two repetitions (once or twice). These specific patterns were defined with those attributes after earlier pilot study tested a wider range of alarm attributes showing that pause period had minimal impact on workers' reactions compared to other attributes (Qin et al., 2022). Alarm patterns were also kept to under 1 second in total notification period, providing workers enough time to react to the traffic hazards that trigger the alarm (AASHTO, 2018; Luo et al., 2016).

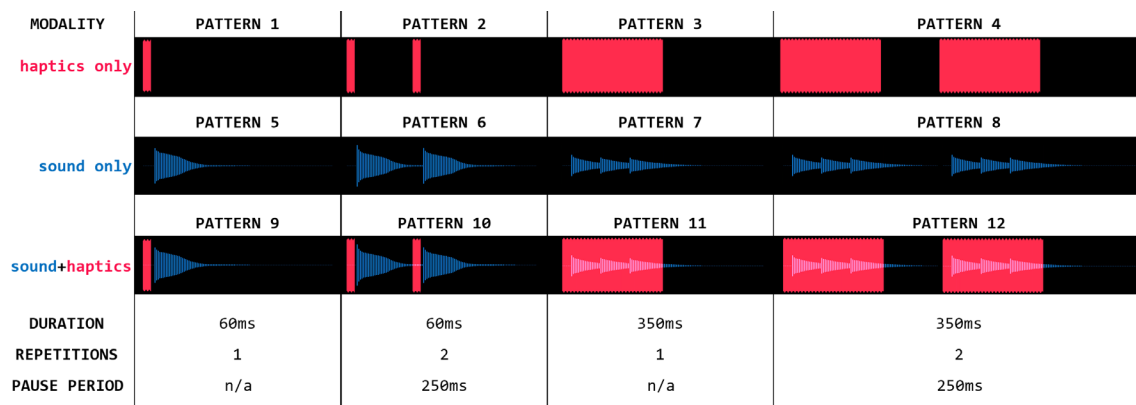
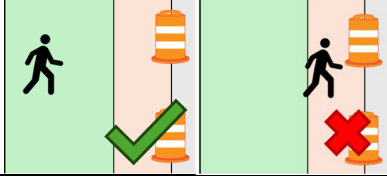
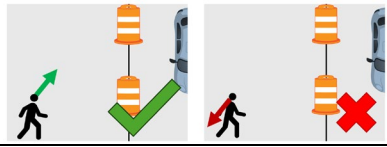
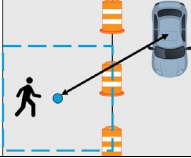

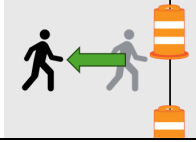



Figure 3. Alarm patterns tested on human roadway workers in wearable alarm user studies.

A total of 22 people participated in the wearable alarm user studies, with ages ranging from 22 to 52, 18 (82%) male and 4 (18%) female. Eight (36%) of the participants had prior construction experience (mean = 6.6 years), including in roadway work zones. No participants reported hearing or vision problems prior to data collection. The number of alarms experienced per participant varied based on how quickly they completed the construction tasks and the randomized traffic hazards generated by SUMO. Each participant typically reacted to twelve alarm patterns in their one-hour user study session allotted by IRB. Overall, a total of **N=255** different human worker reactions to randomized alarm patterns were captured in these user studies.

Table 1 illustrates the how the raw data collected from user studies is processed for the roadway worker model and its implications for RL agent training in Phase 2 (see Section 3.2). Specifically, Table 1 identifies how processed data is used for 1) modelling worker *behaviour* (e.g., safe/unsafe work zone position) to then provide for the RL agent's input *state*, and 2) modelling worker *reactions* to alarms (e.g., head turn) to then provide for the RL agent's reward function.

Table 1. Data collected and processed from human roadway workers in wearable alarm user studies for roadway worker model and RL alarm agent.

Raw Data	Processed Data	Worker Model Component	RL Component
Worker's VR headset position (x,y,z)	Safe/unsafe work zone position 	Behaviour	State
Worker's VR headset orientation (quaternions)	Safe/unsafe worker gaze direction 	Behaviour	State
Vehicle positions in Unity environment	Minimum distance between vehicles and work zone 	Behaviour	State
Worker's progress in sensor installation procedure	Percent progress of construction (%) 	Behaviour	Terminate episode loop
Alarm pattern	Alarm pattern and previous alarm pattern worker experienced	Reaction	Reward
Worker's VR headset position (x,y,z)	Worker moves body to evade traffic hazard after alarm pattern 	Reaction	Reward
Worker's VR headset orientation (quaternions)	Worker turned their head after alarm pattern 	Reaction	Reward
Smartwatch/VR controller button press	Watch press response time after alarm pattern	Reaction	Reward

The following describes how data was processed from raw data collected during user studies into a data format for use in RL training environment (state, reward):

- Safe/unsafe work zone position:** The worker's position is recorded during user studies as x,y,z coordinates in the Unity VR environment. The participants generally moved around 2.7x2.7m square area in the lab space. The resulting work zone in VR environment was around 3.8m in width, spanning the diagonal (i.e., hypotenuse) of that square area. Given this space, the worker's position in the work zone was then labelled as unsafe (data value=0) if they were in the outer third (~1.3m) span near traffic and safe (data value=1) otherwise. While traffic worker safety standards do not have recommended distances workers should stand from the traffic edge of the work zone (i.e., lateral buffer spacing), pedestrian safety studies have assumed similar distances between humans and vehicles as being high-risk (Zuo et al., 2020).
- Safe/unsafe worker gaze direction:** The worker's head orientation was recorded as quaternions per Unity's convention for reporting object rotations (Unity Technologies, 2023). These quaternions are converted into a 3D vector representing the worker's gaze direction (i.e., normal vector of the worker's facial plane). The worker's gaze direction was considered safe if they were facing towards the traffic side and unsafe if not.
- Minimum distance between vehicles and work zone:** The distance between vehicles and the work zone in Unity was recorded at every timestep. The minimum distance between the vehicles and work zone was then used for the roadway worker model development.
- Percent progress of construction:** Workers in the VR simulation had to complete eight steps in the traffic sensor installation procedure. Their progression through these steps was recorded and processed as a percent completion.
- Alarm pattern:** All alarm pattern IDs (see Figure 2) a participant experienced are recorded. The sequence of alarm patterns can play a role in alarm fatigue especially

with the same pattern persistently raised throughout a workday. Worker reactions to each alarm pattern are analysed within a 3 second timeframe after the alarm, consistent with expected reaction times in traffic and construction safety standards (AASHTO, 2018; Luo et al., 2016).

- **Worker moves body after alarm pattern:** After a worker receives an alarm pattern, their VR head position may quickly change to indicate that they have moved their entire bodies in response. If this rate of change is observed in the raw time series data within 3 seconds after the alarm pattern, that particular reaction to that alarm pattern instance was labelled safe (data value=1), unsafe otherwise (data value=0).
- **Worker turned their head after alarm pattern:** After a worker receives an alarm pattern, their VR head orientation may quickly change to indicate that they have turned their heads in response. If this rate of change is observed in the raw time series data within 3 seconds after the alarm pattern, that particular reaction to that alarm pattern instance was labelled safe (data value=1), unsafe otherwise (data value=0).
- **Watch press response time:** All participants were instructed by research staff to press the watch screen or VR controller button once they noticed the alarm pattern (hear/feel). If the timestamp of this smartwatch response is observed in the raw time series data within 3 seconds after the alarm pattern, that particular reaction to that alarm pattern instance was labelled safe (data value=1), unsafe otherwise (data value=0).

As listed in Table 1, each category of processed data informs a data-driven model of how human roadway workers *behave* (i.e., every timestep) and *react* (i.e., times when alarm active) to specific alarm patterns, generating respective values for the *state* and *reward* inputs of an RL agent. This roadway worker model enables this study to run numerous offline training episodes for RL alarm agents (i.e., bootstrapping), allowing the RL alarm agent to try

raising any sequence of alarm patterns to after multiple traffic hazards approach the work zone. Details on how the worker *behaviour* and *reactions* are modelled using statistical (e.g., log normal distributions) and machine learning methods (e.g., transformers) is provided in the Appendix. Details on how the worker model's generated values for *behaviour* and *reaction* feed into the RL agent's inputs for *state* and *reward* are explained in the next section.

3.2 Training RL alarm control agents and evaluating reward function variations on roadway worker safety

The RL paradigm consists of an *agent* learning to select *actions* based on an observed *state* from the *environment* to maximize a *reward*. The agent learns over the course of many episodes, which end when the environment reaches a *terminal state*. Applying RL algorithms to the problem of roadway worker alarms requires that these components need to be numerically defined. Figure 4 illustrates the typical episode loop for training the RL alarm agent within a data-driven simulation environment.

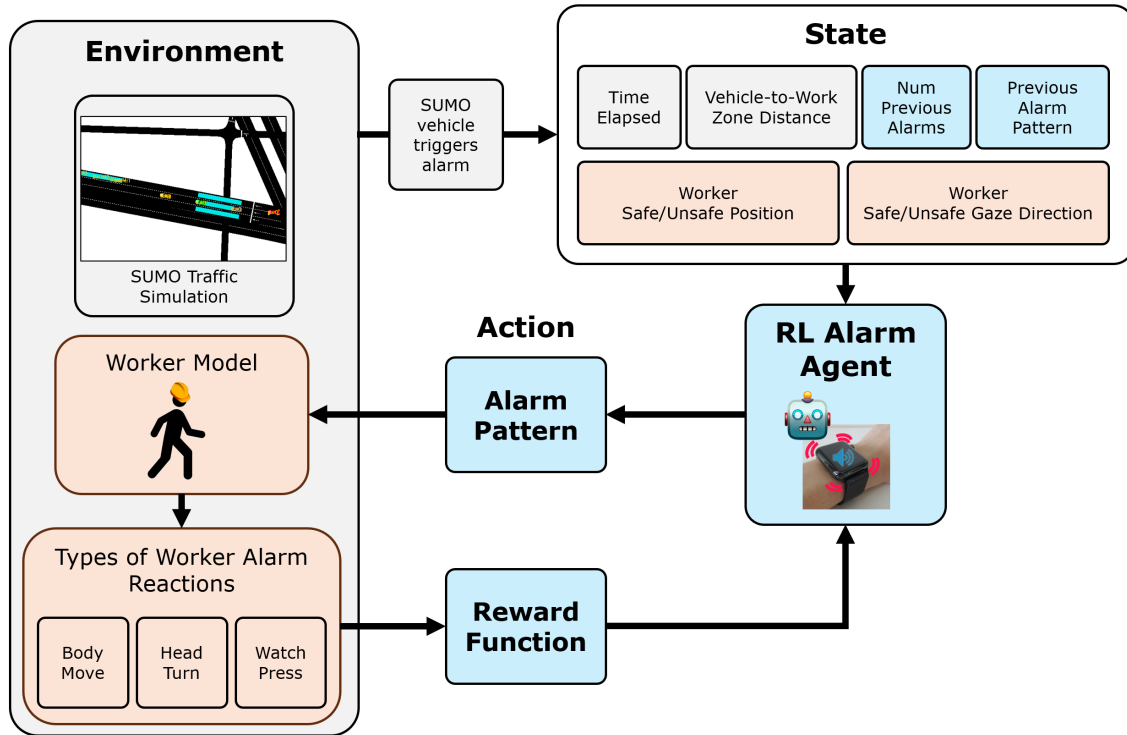


Figure 4. Definitions of RL agent's environment, state, action, and reward function.

The following paragraphs detail the environment, state, actions, and reward components of RL alarm agent:

Environment: The RL alarm agent's external environment includes both the traffic vehicles' movements computed by SUMO traffic simulation and a *roadway worker model* representing how human workers behave and react to different alarm patterns. These two components of the environment generate the input values for the RL alarm agent's *state* and *reward*. The environment's episode loop continues to run at every SUMO traffic simulation timestep (0.1s) and *terminates* when the roadway worker model predicts a 100% construction progress, indicating the worker has completed all construction tasks (i.e., sensor installation).

State: At every timestep, SUMO calculates traffic vehicle locations and the minimum vehicle-to-work zone distance is computed. The roadway worker *behaviour* model recursively generates values for percent worker construction progress and whether it has a safe or unsafe work zone position and gaze direction. Specific details on how these aspects of worker behaviour are modelled can be found in the Appendix, under Figure A1. When

SUMO generates a traffic hazard (e.g., speeding or intruding vehicle) a vector of numerical features (e.g., [20.1, 1, 0, 140.2, 0, 0]) is given to the RL alarm agent as its state. These vector features include *time elapsed* in the episode [seconds], the worker's *safe/unsafe position* and *gaze direction* [0 or 1], vehicle to work zone distance [meters], *number of previous alarms* worker experienced, and the most recent *previous alarm pattern* it raised [1-12, 0 if none]. The last two parameters can be important inputs for the RL agent to account for potential alarm fatigue effects to be investigated in future studies.

RL Agent and Actions: Based on this state, the RL agent selects among the 12 possible alarm patterns (Figure 3) to raise to warn the worker of the traffic hazard. Given this study's choice of LinUCB contextual bandit, the state vector is set equal to the context, $\mathbf{x}_{t,a}$ in equation (2), for the RL agent to estimate an expected reward for each possible alarm pattern. Per Equation (1), the alarm pattern with the highest expected reward, a_t , is selected and raised to the roadway worker model.

Reward: Once the RL alarm agent selects an alarm pattern, the roadway worker *reaction* model predicts three types of binary worker alarm reactions [0 or 1]: body move, head turn, and watch press. Specific details on how the model predicts worker reaction can be found in the Appendix, under Figure A2. The reward signal fed back to the RL alarm agent is computed based on the predicted reactions and the specific reward function definition. Four possible definitions of reward functions are considered in this study, each with their own weights (i.e., priority) assigned to the three types of worker alarm reactions. These sets of reward weights are listed in Table 2. To the best of authors' knowledge, no large work zone safety dataset exists relating specific worker alarm reactions to fatal injuries that can justify a particular set of reward weights. This study therefore explores the impact of different reward function weights on how an RL alarm agent may select different alarm patterns and the consequent differences in worker reactions. The maximum possible reward is +10 as an

experiment control across testing different reward functions. In all four reward functions, the RL agent receives a -10 penalty if the worker model does not react at all to signal ineffective alarm patterns and indicate potential alarm fatigue.

Table 2. Definitions for four different sets of reward weights for each of the three types of worker reactions.

	<i>Reward Weight Definition</i>			
Type of worker alarm reaction	<i>I</i>	<i>II</i>	<i>III</i>	<i>IV</i>
Body move	+3.33	+6	+3	+1
Head turn	+3.33	+3	+6	+3
Watch press	+3.33	+1	+1	+6
<i>Max Sum</i>	<i>+10</i>	<i>+10</i>	<i>+10</i>	<i>+10</i>
<i>No reaction</i>	<i>-10</i>	<i>-10</i>	<i>-10</i>	<i>-10</i>

Four separate contextual bandit RL agents are trained under the four different reward functions. This study uses an ϵ -greedy algorithm, a typical method which forces the agent to choose a random action with a probability of ϵ at the outset of training [15]. To encourage exploration, ϵ is set equal to 1 (i.e., 100% random) for the first 120 alarm patterns each RL agent tries during training. The probability ϵ then decays exponentially to 0.1, shifting each RL agent to gradually follow its policy (i.e., greedy action) over 100 training episodes. To evaluate the effects of the different reward functions, the four RL agents undergo 100 test episodes where they always choose greedy actions without updating their policy. Test results can reveal the potential impact on worker reaction due to different alarm control policies that the contextual bandits learned during training under different reward functions.

4. Results

Table 3 shows the descriptive statistics on reward earned by RL agents after each alarm, under different reward function weights in training and test episodes. RL agents trained in each of Table 2's reward function definitions (I-IV) achieved a higher reward on average

after each alarm in test episodes compared to training episodes, indicating effective RL training. All subsequent results are specific to test episodes only.

Table 3. Statistics on rewards after each alarm raised by RL agents during training and test episodes.

	I		II		III		IV	
	train	test	train	test	train	test	train	test
n	229	238	241	232	256	252	248	242
μ	7.18	9.44	8.49	9.54	8.43	9.28	6.61	8.54
std	3.97	1.25	2.77	1.07	3.13	2.27	4.03	2.81

Figure 5 displays the frequency of alarm patterns raised during test episodes by different RL agents trained under different reward function weights. Trained under equal reward weighted function (I), the RL agent only raises alarm patterns 12, 4, 1, 10, and 6 during test episodes. Alarm patterns 1 and 4 are of the “haptics only” modality. Alarm patterns 10 and 12 are of “sound and haptics” modality and both have a continuous duration of 350ms. Under all reward weights, the RL agent raises alarm pattern 12 most frequently. In terms of modality, all reward weights result in an RL agent raising “haptics only” (1-4) and “sound and haptics” modality alarm patterns (9-12) more frequently than "sound only" alarm patterns (5-8). Looking at specific alarm patterns, it appears that shifting the reward weights away from watch press reactions (function I vs II and III) results in an RL agent that raises alarm pattern 4, 10 and 12 less frequently. This can denote that RL agents may have recognized that alarm pattern 4, 10 often results in a watch press reaction, but not a head turn or body move reaction from the worker.

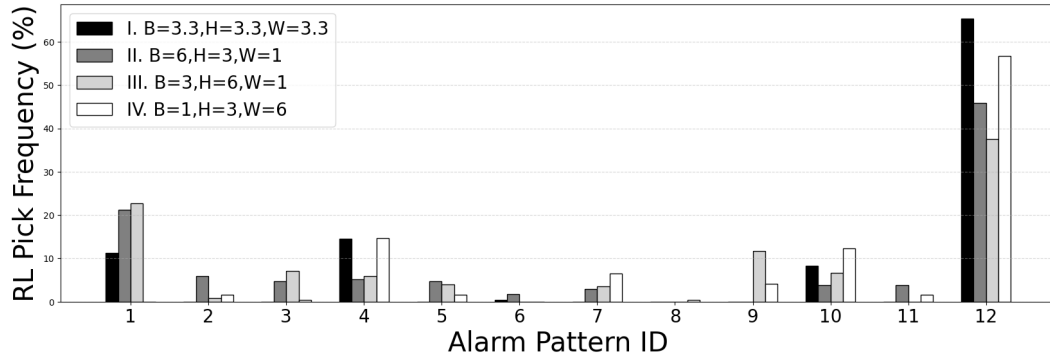


Figure 5. Frequency of alarm patterns raised during test episodes by different RL agents trained under different reward function weights (I-IV).

Table 4 details the distribution of rewards the different RL agents achieve after each alarm pattern they raise during test episodes. Under the equal-weighted reward function (I), the RL agent achieves either a reward of +10 or +6.67 after each alarm, indicating the worker reacts to each alarm in either all three, or two out of three possible ways. The RL agent never gets a penalty of -10, meaning none of its alarms result in no worker reactions. This suggests that workers are reacting more consistently to this particular RL agent's chosen alarm patterns. With a weight shift to prioritize body move reactions from +3.3 to +6, reward function II results in an RL agent achieving a reward of +10 or +9 most often, achieving a higher mean reward compared to reward function I. Rewards indicate that the worker either reacted to each alarm in all 3 possible ways (+10) or only moves their body and turns their head (+6 + 3). With a similar shift towards head turn reactions, reward function III results in an RL agent achieving the reward of +9 more often than that of reward function II, but also sometimes gets a -10 penalty, indicating that the workers occasionally do not react at all. When the reward function IV prioritizes watch press reactions, the RL agent achieves a reward of +10 and +4 most often but gets a rare -10 penalty. This indicates the worker either reaction to each alarm in all 3 possible ways (+10) or only moves their body and turns their head (+3+1). Note that while reward function IV's distribution in Table 4 has changed compared with reward functions II and III, the implications on worker reactions are still

similar. The difference in reward distribution is more of a result of the changing of reward weight assignments.

Table 4. Detailed statistics on the distribution of reward per RL alarm pattern during test episodes under different reward function weights (I-IV).

Reward Statistics			Reward Bin Counts (%)						
	μ	std	10	[9,10)	[6,7)	[4,5)	[3,4)	[1,2)	-10
I	9.4	1.3	198 (83.2%)		40 (16.8%)				
II	9.5	1.1	158 (68.1%)	68 (29.3%)	1 (0.4%)	1 (0.4%)	4 (1.7%)		
III	9.3	2.3	145 (57.5%)	100 (39.7%)	3 (1.2%)			1 (0.4%)	3 (1.2%)
IV	8.5	2.8	186 (76.9%)			52 (21.5%)	3 (1.2%)		1 (0.4%)

Figure 6 illustrates how differently workers reacted to the alarm patterns raised by RL agents trained under different reward function weights. Looking at each possible worker alarm reaction individually, Figure 6 shows that workers move their bodies and turn their heads for nearly all RL alarm patterns, regardless of reward function weights. As a result of decreasing the watch press reward weight, workers press the watch screen less often to alarm patterns raised by RL agents trained under reward functions II and III, compared with reward weight I and IV.

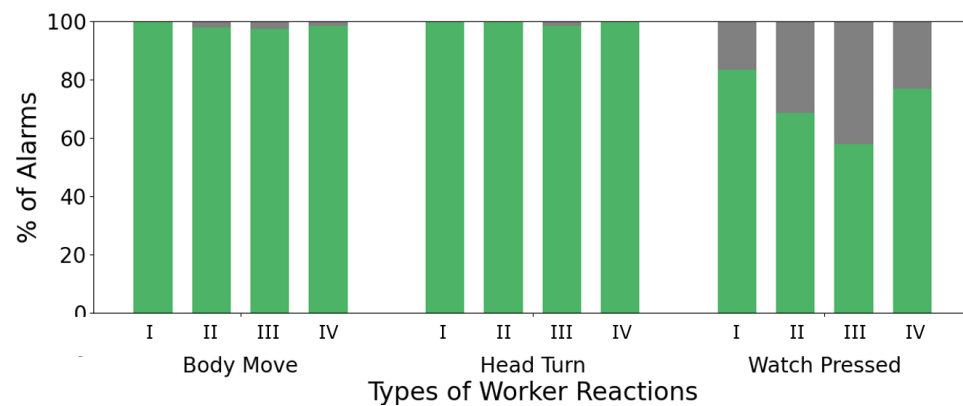


Figure 6. How workers reacted to the alarm patterns raised by RL agents trained under different reward function weights (I-IV).

This result is unsurprising given Figure 5 showing generally similar distributions of alarm patterns overall and especially with respect to alarm pattern 12 (sound and haptics modality combined, duration of 350ms, repeated twice). These results are also consistent with the data collected from human VR user studies used to train the worker model (see Appendix Figure A3), which show more variation in watch press reactions to different alarm patterns than body movement and head turn reactions. Overall, Figure 6 illustrates how different reward weighted functions for training RL alarm agents may lead to specific worker behaviours being promoted or demoted.

5. Discussion

Findings from this study demonstrate that training RL alarm agents with reward functions of different weights for different types of roadway worker reactions can change the variety of alarm patterns raised by each agent. However, under all reward weights, the “sound and haptics” alarm pattern with the greatest continuous duration and number of repetitions (alarm pattern 12) is the dominant RL agent selection. Sound only modality alarms are least often raised by the RL agent. These potential alarm control policies are in line with recommendations from other roadway safety studies finding workers preferring alarms with combined modalities and struggling to hear sound alarms amidst noisy work zone environments (Abdallah et al., 2024; Gambatese et al., 2017).

With reward functions that prioritize watch press reactions less (weight definitions II and III), Figure 5 shows that the RL alarm agent learns policies to choose alarm patterns 4, 10, and 12 less frequently. Alongside trends in Figure 6, results indicate that when the RL agent raises those alarm patterns less frequently, roadway workers press the watch screen less often. While pressing a wearable warning device screen or button may not be the most important life-saving form of worker reaction to a traffic hazard, these trends demonstrate the

potential downstream impact of changes in RL reward functions on which alarm patterns are raised and consequent worker reactions for their safety.

Results in Table 4 point towards the equal weighted reward function I being the best for producing an RL alarm agent that ensures a roadway worker consistent safe reaction to traffic hazards. Unlike reward functions III and IV, the RL agent under reward function I achieves the reward +10 most often and never gets a penalty of -10. Therefore, workers appear to consistently respond with all 3 possible reactions and never get desensitized to the alarms under reward function I (i.e., no alarm fatigue). In terms of roadway safety implications, this suggests that reward function I trains the most effective RL agent to raise alarm patterns that prompt workers to consistently react to traffic hazards. As RL-based alarms become widely implemented in work zones, public safety agencies should expand data collection to include specific worker reactions and fatal injury rates. These statistics can help develop a domain-informed RL reward function before real-world work zone implementation.

6. Conclusions

Future studies will investigate how to implement online RL alarm agents in both controlled lab settings and real-world work zones, with human participant demographics more representative of US roadway workers. Given this study's method using only a multi-arm contextual bandit, future work will also investigate other RL algorithms (e.g., DQN, PPO). Studies will also examine the sequences of alarm patterns and their effects on workers' alarm fatigue in their reactions. Overall, this work demonstrates the feasibility of using RL algorithms to control alarms for roadway worker safety from traffic hazards. Using a VR-traffic platform, this study collects data on worker reactions to alarm patterns to train an RL alarm control agent. The study defines the RL alarm agent's state, actions, and different

reward functions within a data-driven environment and explores four reward functions prioritizing different worker reactions. Assigning equal reward weights to all reaction types leads to more consistent and safe worker reactions to traffic hazards, while reducing reward weights for specific reactions (e.g., watch press) decreases their frequency. These findings can guide transportation safety practitioners in adopting RL alarm agents and fine-tuning their reward functions to promote safe and consistent worker reactions in real-world roadway work zones.

Acknowledgements

The authors would like to acknowledge the funding agencies for this project: C2SMARTER funding under grant number (69A3551747119) and a 50% cost-share by New York University. Special thanks to Dr. Fan Zuo for his technical advice on simulating traffic near roadway work zones and Bill Xu Zhou for assisting as a machine learning developer.

Declaration of interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- AASHTO. (2018). *A Policy on Geometric Design of Highways and Streets* (7th ed.). American Association of State Highway and Transportation Officials.
- Abdallah, A., Ibrahim, A., Russell-Vernon, C., Nnaji, C., Gambatese, J., & Shober, J. (2024). Advancing safety in short-term utility work zones: Assessing the role of work zone intrusion alert technologies (WZIATs). *Transportation Research Interdisciplinary Perspectives*, 26, 101133. <https://doi.org/10.1016/j.trip.2024.101133>

- Akanmu, A. A., Olayiwola, J., Ogunseiju, O., & McFeeters, D. (2020). Cyber-physical postural training system for construction workers. *Automation in Construction*, 117, 103272. <https://doi.org/10.1016/j.autcon.2020.103272>
- Amani, M., & Akhavian, R. (2024). *Ergonomic Optimization in Worker-Robot Bimanual Object Handover: Implementing REBA Using Reinforcement Learning in Virtual Reality*.
- ARTBA. (2024). *Work Zone Data*. <https://workzonesafety.org/work-zone-data/>
- Asghari, V., Wang, Y., Biglari, A. J., Hsu, S.-C., & Tang, P. (2022). Reinforcement Learning in Construction Engineering and Management: A Review. *Journal of Construction Engineering and Management*, 148(11). [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0002386](https://doi.org/10.1061/(ASCE)CO.1943-7862.0002386)
- Blackmon, R. B., & Gramopadhye, A. K. (1995). Improving Construction Safety by Providing Positive Feedback on Backup Alarms. *Journal of Construction Engineering and Management*, 121(2), 166–171. [https://doi.org/10.1061/\(ASCE\)0733-9364\(1995\)121:2\(166\)](https://doi.org/10.1061/(ASCE)0733-9364(1995)121:2(166))
- Brenneis, D. J. A., Parker, A. S., Johanson, M. B., Butcher, A., Davoodi, E., Acker, L., Botvinick, M. M., Modayil, J., White, A., & Pilarski, P. M. (2022). Assessing Human Interaction in Virtual Reality With Continually Learning Prediction Agents Based on Reinforcement Learning Algorithms: A Pilot Study. In F. Cruz, C. Hayes, F. L. da Silva, & F. P. Santos (Eds.), *Proc. of the Adaptive and Learning Agents Workshop (ALA 2022)*,.
- Cai, J., Du, A., Liang, X., & Li, S. (2023). Prediction-Based Path Planning for Safe and Efficient Human–Robot Collaboration in Construction via Deep Reinforcement Learning. *Journal of Computing in Civil Engineering*, 37(1). [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0001056](https://doi.org/10.1061/(ASCE)CP.1943-5487.0001056)

- Camci, A., Abbott, P., & Rooney, D. (2020). Investigating Alarm Fatigue with AlarmVR: A Virtual ICU for Clinical Alarm Research. *Forum Acusticum*, 2973–2978.
- Chae, J., & Kang, Y. (2021). Designing an Experiment to Measure the Alert Fatigue of Different Alarm Sounds Using the Physiological Signals. In C. Feng, T. Linner, & I. Brilakis (Eds.), *38th International Symposium on Automation and Robotics in Construction* (pp. 545–552). <https://doi.org/10.22260/ISARC2021/0074>
- Du, W., Dash, A., Li, J., Wei, H., & Wang, G. (2023). Safety in Traffic Management Systems: A Comprehensive Survey. *Designs*, 7(4), 100. <https://doi.org/10.3390/designs7040100>
- Federal Highway Administration. (2022). *Manual on Uniform Traffic Control Devices (MUTCD)* (Revision 3 July 2022). U.S. Department of Transportation. https://mutcd.fhwa.dot.gov/pdfs/2009r1r2r3/pdf_index.htm
- Gambatese, J. A., Lee, H. W., & Nnaji, C. A. (2017). *Work zone intrusion alert technologies: assessment and practical guidance*. <https://rosap.ntl.bts.gov/view/dot/32574>
- Haghighi, N., Vladis, N., Liu, Y., & Satyanarayan, A. (2020). *The Effectiveness of Haptic Properties Under Cognitive Load: An Exploratory Study*.
- Lambert, N., Gilbert, T. K., & Zick, T. (2023). *The History and Risks of Reinforcement Learning and Human Feedback*.
- Lee, Y.-C., Cherng, F.-Y., King, J.-T., & Lin, W.-C. (2019). To Repeat or Not to Repeat?: Redesigning Repeating Auditory Alarms Based on EEG Analysis. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–10. <https://doi.org/10.1145/3290605.3300743>
- Li, L., Chu, W., Langford, J., & Schapire, R. E. (2010). A contextual-bandit approach to personalized news article recommendation. In M. Rappa & P. Jones (Eds.), *Proceedings*

of the 19th international conference on World wide web (pp. 661–670). ACM.

<https://doi.org/10.1145/1772690.1772758>

Li, T., Haines, J. K., De Eguino, M. F. R., Hong, J. I., & Nichols, J. (2022). Alert Now or Never: Understanding and Predicting Notification Preferences of Smartphone Users. *ACM Transactions on Computer-Human Interaction*, 29(5), 1–33.

<https://doi.org/10.1145/3478868>

Lichtlé, N., Jang, K., Shah, A., Vinitzky, E., Lee, J. W., & Bayen, A. M. (2023). Traffic Smoothing Controllers for Autonomous Vehicles Using Deep Reinforcement Learning and Real-World Trajectory Data. *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, 4346–4351.

<https://doi.org/10.1109/ITSC57777.2023.10421828>

Lordianto, B., Lu, D., Ho, W., & Ergun, S. (2024). Hapti-met: A Construction Helmet with Directional Haptic Feedback for Roadway Worker Safety. In B. Riveiro & P. Arias (Eds.), *31st International Workshop on Intelligent Computing in Engineering* (pp. 453–462).

Lu, D., Ergun, S., & Ozbay, K. (2024). Reinforcement learning-based optimal control of alert systems to mitigate roadway workers' alarm fatigue. In N. Stamatiadis & R. Souleyrette (Eds.), *2024 Road Safety and Simulation Conference*.

Luo, X., Li, H., Huang, T., & Rose, T. (2016). A field experiment of workers' responses to proximity warnings of static safety hazards on construction sites. *Safety Science*, 84, 216–224. <https://doi.org/10.1016/j.ssci.2015.12.026>

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). *Playing Atari with Deep Reinforcement Learning*.

<https://doi.org/10.48550/arXiv.1312.5602>

- Papachristos, E., Merritt, T. R., Jacobsen, T., & Bagger, J. (2020). Designing Ambient Multisensory Notification Devices: Managing Disruptions in the Home. *19th International Conference on Mobile and Ubiquitous Multimedia*, 59–70.
<https://doi.org/10.1145/3428361.3428400>
- Pilarski, P. M., Butcher, A., Davoodi, E., Johanson, M. B., Brenneis, D. J. A., Parker, A. S. R., Acker, L., Botvinick, M. M., Modayil, J., & White, A. (2022). *The Frost Hollow Experiments: Pavlovian Signalling as a Path to Coordination and Communication Between Agents*.
- Qin, J., Lu, D., & Ergan, S. (2022, October). Towards Increased Situational Awareness at Unstructured Work Zones: Analysis of Worker Behavioral Data Captured in VR-based Micro Traffic Simulations. *Proceedings of the 19th International Conference on Computing in Civil and Building Engineering*.
- Sabeti, S., Ardecani, F. B., & Shoghli, O. (2024). *Augmented Reality Warnings in Roadway Work Zones: Evaluating the Effect of Modality on Worker Reaction Times*.
- Seifi, H., Anthonypillai, C., & MacLean, K. E. (2014). End-user customization of affective tactile messages: A qualitative examination of tool parameters. *2014 IEEE Haptics Symposium (HAPTICS)*, 251–256. <https://doi.org/10.1109/HAPTICS.2014.6775463>
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484–489.
<https://doi.org/10.1038/nature16961>
- Sutton, R., & Barto, A. (2022). *Reinforcement Learning: An Introduction* (F. Bach, Ed.; 2nd ed.). MIT Press.

- Tang, L., Jiang, Y., Li, L., & Li, T. (2014). Ensemble contextual bandits for personalized recommendation. *Proceedings of the 8th ACM Conference on Recommender Systems*, 73–80. <https://doi.org/10.1145/2645710.2645732>
- Thapa, D., & Mishra, S. (2021). Using worker's naturalistic response to determine and analyze work zone crashes in the presence of work zone intrusion alert systems. *Accident Analysis & Prevention*, 156, 106125. <https://doi.org/10.1016/j.aap.2021.106125>
- Unity Technologies. (2023, November 10). *Unity - Manual: Rotation and orientation in Unity*. Unity - Manual. <https://docs.unity3d.com/6000.0/Documentation/Manual/QuaternionAndEulerRotationsInUnity.html>
- Wang, Z., Huang, C., Yao, B., & Li, X. (2024). Integrated reinforcement and imitation learning for tower crane lift path planning. *Automation in Construction*, 165, 105568. <https://doi.org/10.1016/j.autcon.2024.105568>
- Yang, X., & Roofigari-Esfahan, N. (2023). Vibrotactile Alerting to Prevent Accidents in Highway Construction Work Zones: An Exploratory Study. *Sensors*, 23(12), 5651. <https://doi.org/10.3390/s23125651>
- Zhang, S., Ergun, S., & Ozbay, K. (2024). Work Zone Safety: Benchmarking Studies between Virtual Reality-based Traffic Co-simulation Platform and Real Work-Zones. In Z. Wang (Ed.), *ICRA 3rd Workshop on Future of Construction: Lifelong Learning Robots in Changing Construction Sites*.
- Zuo, F., Ozbay, K., Kurcu, A., Gao, J., Yang, H., & Xie, K. (2020). Microscopic simulation based study of pedestrian safety applications at signalized urban crossings in a connected-automated vehicle environment and reinforcement learning based optimization of vehicle decisions. *Advances in Transportation Studies*, 2, 113–126.

Appendix

The roadway worker model generates output data for the RL agent state and reward based on input data. Each piece of model output is explained in terms of relevant inputs and statistical and machine learning modelling techniques.

A.1 Modelling worker behaviour

Worker model always generates an output/predictions in the RL episode's future timestep based on input data in the current timestep/time elapsed. All raw data for the worker behaviour was recorded during the user studies at a timestep of 16.6ms (~60Hz VR headset display framerate) and then interpolated (i.e., downsampled) to 100ms (10Hz) timesteps used in RL training episodes (i.e., shorter real-time training with fewer overall timesteps)

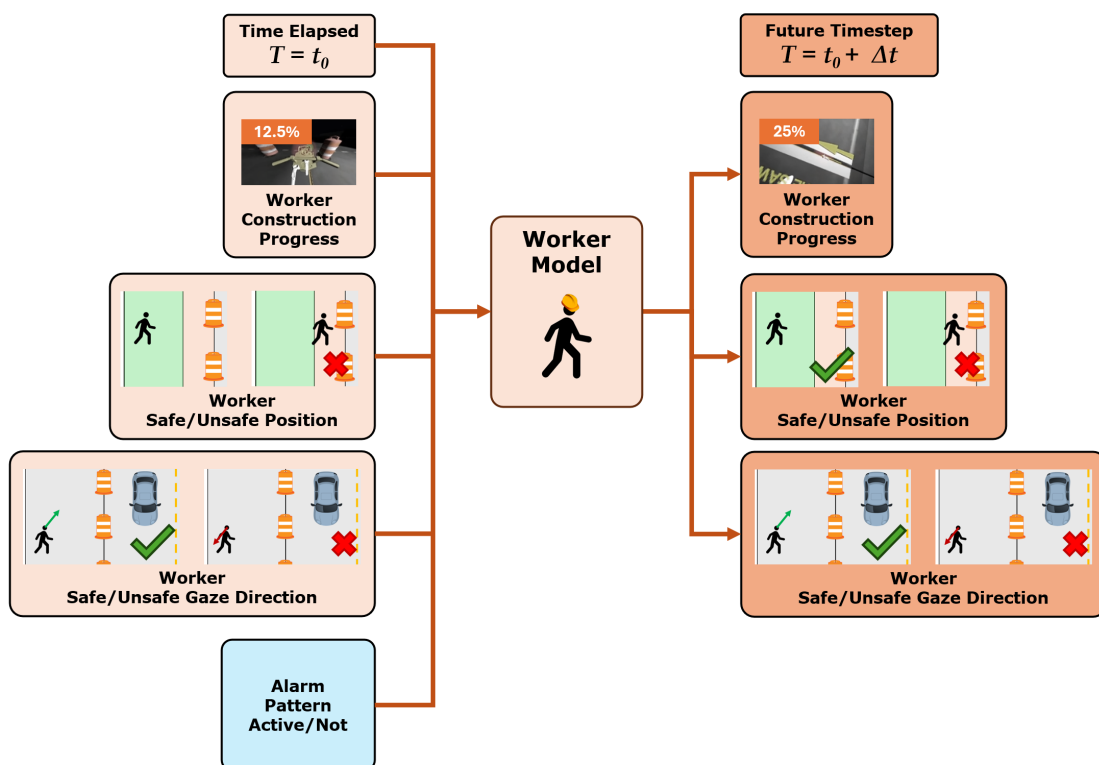


Figure A1. Worker behaviour model for generating states for RL agent (see Figure 4)

Worker behaviour model output data: Worker construction progress (0-100%)

Relevant input data: Episode time elapsed, prior time step construction progress

How output is modelled given input data: The mean and standard deviation times it took human workers to complete each step of the traffic sensor installation was computed. A log normal distribution based on those computed means and standard deviations predicts the time the roadway worker model took to complete each step.

Worker behaviour model output data: Safe/unsafe work zone position (0 or 1)

Relevant input data: Worker construction progress, prior time step safe/unsafe work zone position, RL agent alarm pattern active/not

How output is modelled given input data: The average time period a human worker continuously remained in a safe or unsafe position was computed from the user studies dataset. The roadway worker model position remains safe or unsafe until that time passed in the episode. Then the worker model changes its position safety based on the probability of how often human workers were in safe/unsafe positions in each stage of construction progress during the user studies. This probability also depends on whether the RL alarm agent has activated an alarm pattern or not.

Worker behaviour model output data: Safe/unsafe gaze direction (0 or 1)

Relevant input data: Episode time elapsed, prior time step safe/unsafe gaze direction, alarm pattern active/not, RL agent alarm pattern active/not

How output is modelled given input data: The average time period a human worker continuously remained in a safe or unsafe gaze direction was computed from the user studies dataset. The roadway worker model gaze direction remains safe or unsafe until that time period passed in the episode. Then the worker model changes its gaze direction safety based on the probability of how often human workers were in safe/unsafe gaze direction over the

time elapsed in user study simulations. This probability also depends on whether the RL alarm agent has activated an alarm pattern or not.

A.2 Modelling worker reactions

Worker model only generates predictions once RL alarm agent selects an alarm pattern.

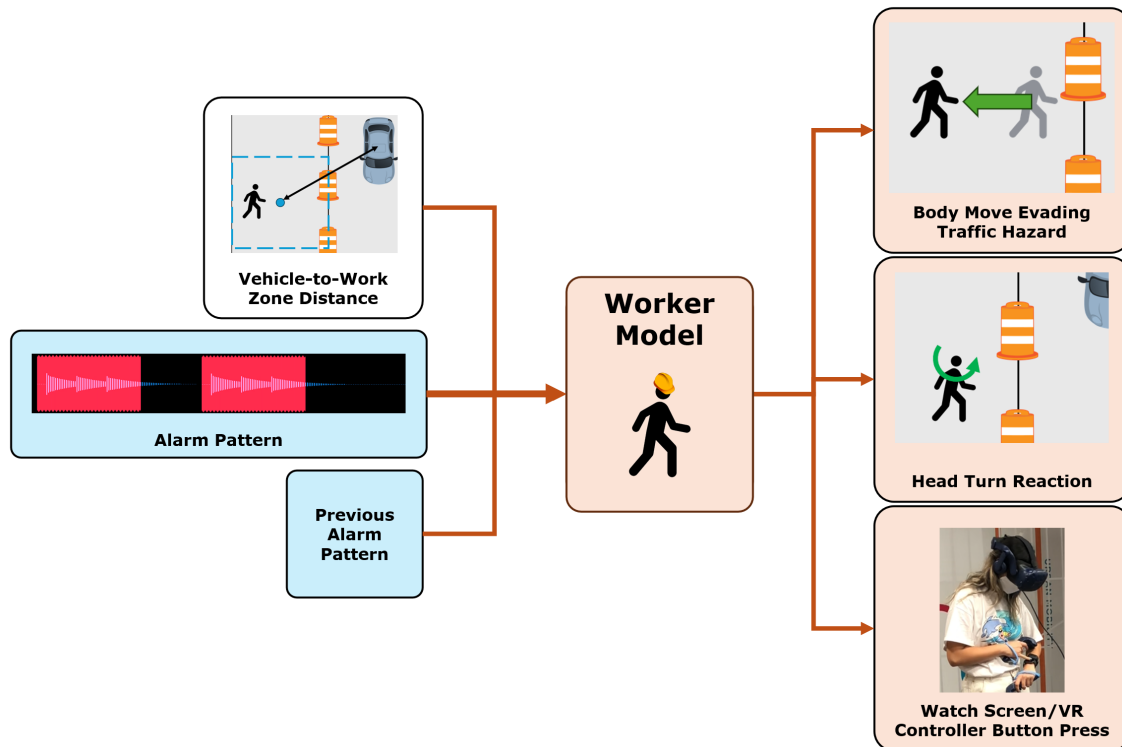


Figure A2. Worker reaction model for computing RL agent reward (see Table 2)

Worker reaction model output data: Head turn reaction (0 or 1)

Relevant input data: Current RL alarm pattern (1-12), previous RL alarm pattern raised (0 if none), minimum distance between work zone and vehicles

How output is modelled given input data: Transformer machine learning model predicts the head turn reaction (1 or 0) based on the above inputs. Accuracy of predictions for the collected user study dataset shown in Figure A3 below.

Worker reaction model output data: Body move evading traffic hazard (0 or 1)

Relevant input data: Current RL alarm pattern (1-12), previous RL alarm pattern raised (0 if none), worker's head turn reaction (1 if reacted, 0 if not)

How output is modelled given input data: Transformer machine learning model predicts the body move reaction (1 or 0) based on the above inputs. Accuracy of predictions for the collected user study dataset shown in Figure A3 below.

Worker reaction model output data: Watch/VR controller button press (0 or 1)

Relevant input data: Current RL alarm pattern (1-12)

How output is modelled given input data: The probability that human workers pressed the watch screen/VR controller button for each alarm pattern was computed. Among the human workers that did press the watch/VR controller button, a mean and standard deviation response time was computed as well. The roadway worker model, given the RL alarm pattern, first computes the probability it will press the watch screen or not. If so, a log normal distribution defined by the mean/std for the RL's chosen alarm pattern predicts a response time. If that response time is within 3 seconds, the worker model's final output for watch press reaction is 1, otherwise 0.

The above methods for modelling each type of worker alarm reaction yielded fairly accurate predictions (blue/dark grey) compared to the reactions observed dataset (green/light grey) collected from human worker wearable alarm user studies (Figure A3). The model's generated worker reactions during RL alarm agent training and testing can be considered realistic given these accuracies.

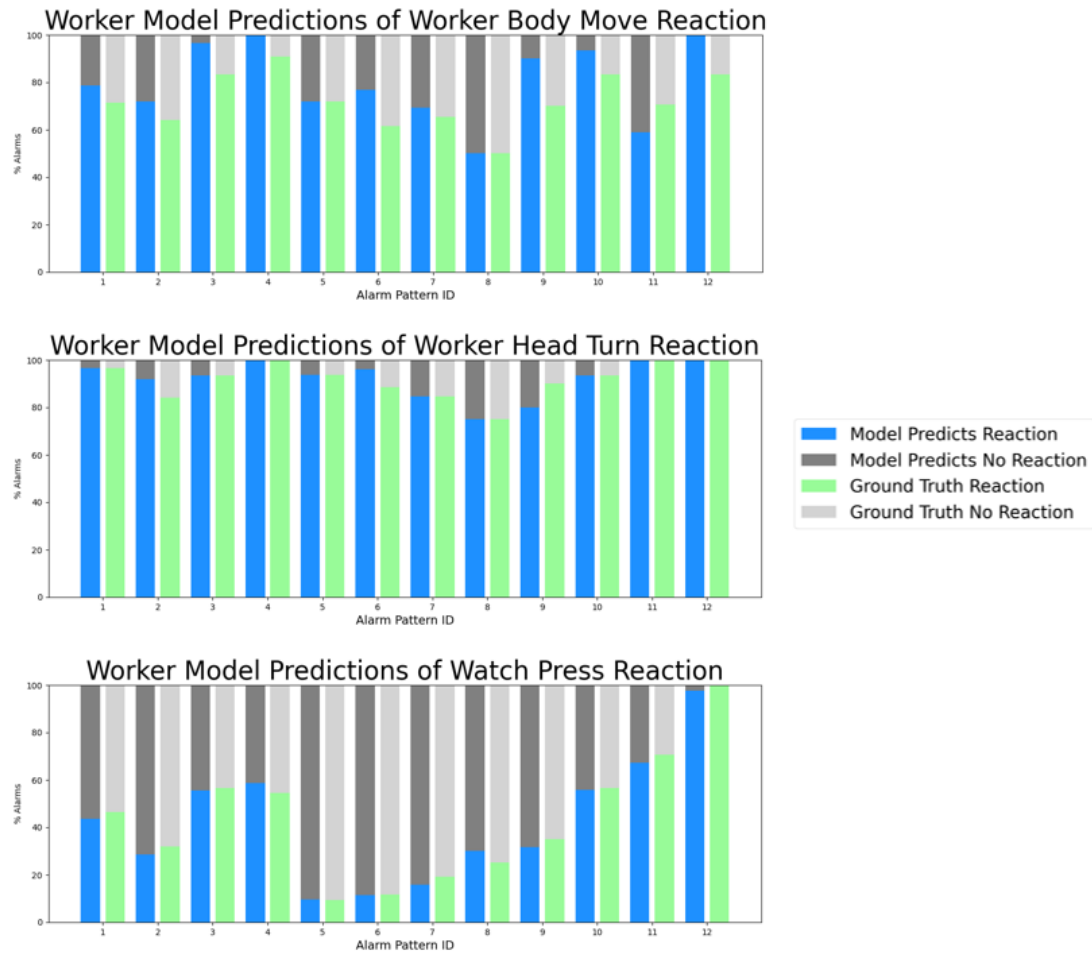


Figure A3. Worker model predictions of reactions to each alarm pattern compared to observed reactions in the wearable alarm user studies.