# The Real Deal: A Review of Challenges and Opportunities in Moving Reinforcement Learning-Based Traffic Signal Control Systems Towards Reality

Rex Chen, Fei Fang and Norman Sadeh

*Institute of Software Research, School of Computer Science, Carnegie Mellon University*

## Abstract

Traffic signal control (TSC) is a high-stakes domain that is growing in importance as traffic volume grows globally. An increasing number of works are applying reinforcement learning (RL) to TSC; RL can draw on an abundance of traffic data to improve signalling efficiency. However, RL-based signal controllers have never been deployed. In this work, we provide the first review of challenges that must be addressed before RL can be deployed for TSC. We focus on four challenges involving (1) uncertainty in detection, (2) reliability of communications, (3) compliance and interpretability, and (4) heterogeneous road users. We show that the literature on RL-based TSC has made some progress towards addressing each challenge. However, more work should take a systems thinking approach that considers the impacts of other pipeline components on RL.

## Keywords

Traffic signal control, Reinforcement learning, Intelligent transportation system, System deployment, Review

## 1. Introduction

As the traffic volume of metropolitan areas continues to grow worldwide, gridlock is becoming an increasingly prevalent concern. According to the *2021 Urban Mobility Report* [1], gridlock led to over 4 billion hours in travel delay and $100+ million in congestion costs across the United States in 2021. This not only impacts commercial productivity but also has environmental consequences. One important mechanism for alleviating gridlock is improving the timing of traffic signals [2]. Historically, most jurisdictions have used fixed timing plans based on traffic models, which assume fixed values of factors such as lane volumes and arrival rates [3]. To minimize implementation burden, traditional traffic signal control (TSC) either uses one fixed plan throughout the entire day, or rotates through several plans depending on the time of the day. However, fixed plans cannot respond in real time to changes in traffic demand [3, 4].

Large traffic volumes also offer an abundance of data that can be used for real-time optimization of signal timing plans. Many deployed systems combine logic-triggered state changes with data-driven searches over sets of schedules [3]. However, an increasing number of approaches

traverse larger search spaces using optimization and scheduling algorithms [5]. Among these approaches, reinforcement learning (RL) has yielded significant improvements over fixed and actuated TSC algorithms in simulations [6]. RL allows systems to learn from the consequences of their decisions, which enables them to achieve continuous self-improvement. Deployments of RL algorithms have achieved success in a variety of complex domains involving human interaction, such as card games [7], real-time strategy games [8], and other applications in transportation such as dispatching for ride-hailing services [9].

However, to our knowledge, RL-based TSC algorithms have never been deployed. This is in spite of the fact that papers introducing novel algorithms in this area commonly list real-world deployment as a goal for future work [10]. We believe that this discrepancy has arisen due to a focus on methodological contributions, instead of on a holistic systems thinking approach based on the data-to-deployment pipeline [11]. If RL-based signal controllers are to achieve success in deployment, domain experts in TSC and in RL must have a shared view of the problem. We take a step towards bridging the gap between research and deployment by providing the first review of challenges that may arise from end-to-end deployments of RL-based TSC, which we intend to provide a common basis of collaboration between research in TSC and RL.

We begin by describing our review methodology in Section 1.1. Then, we provide a high-level review of the fields of TSC and RL in Section 2. Next, we explore four engineering challenges. For each of these challenges, we will provide a review of (1) how these challenges are significant concerns for the state of the art in RL-based TSC; (2) what practical considerations relevant to these challenges have arisen in deployments of non-RL TSC systems; and (3) what progress has been made in the RL-based TSC literature towards solving these challenges.

- **Uncertainty in detection.** (Section 3) Typically, RL-based TSC algorithms learn based on metrics such as queue length or travel time. These require accurate vehicle detection technologies, which may not always be available in the field. Strategies to deal with detector uncertainty and failure are a prerequisite of deployment.

- **Reliability of communications.** (Section 4) Some decentralization is necessary for RL-based TSC. Coordination between intersections is important for optimizing network-level metrics, yet most work in RL-based TSC has not considered the practicalities of dealing with failure and latency in inter-intersection communications.

- **Compliance and interpretability.** (Section 5) Jurisdictions will not have confidence in RL-based signal controllers without assurances about compliance to standards (e.g., minimum green time) and safety requirements. The interpretability of models is important for ensuring that signalling plans can be audited and adjusted by stakeholders.

- **Heterogeneous road users.** (Section 6) Most simulations for RL-based TSC assume that all cars are the same size and have the same free-flow speed. However, cars share the road with pedestrians, buses, emergency vehicles, and other road users. Algorithms must detect and respond to the needs of different road users in a safe, equitable manner.

Finally, we end with concluding thoughts and suggestions for future work in Section 7.

## 1.1. Methodology

To obtain an overview of the domain of RL-based TSC, we conducted a targeted search on Google Scholar with the keywords "traffic signal"/"traffic light", "reinforcement learning", and "review"/"survey". We identified the four challenges addressed in the following sections through these reviews. From here, we conducted snowball sampling based on their citations to locate papers in the RL literature that discuss these challenges. For RL papers, we focused on those published after 2015, since this field has rapidly evolved over the past several years. We also performed additional targeted Google Scholar searches to find literature which describes non-RL deployments of TSC, by searching the keywords "traffic signal"/"traffic light" and "adaptive" in conjunction with the following keywords:

- For **Section 3**, "uncertainty", "noise", "sensing error", "accuracy'.

- For **Section 4**, "coordination", "communication", "closed loop", "message", "NTCIP".

- For **Section 5**, "compliance", "safety", "accountability", "interpretability"/"explainability".

- For **Section 6**, "pedestrian"/"leading pedestrian interval", "cyclist", "transit", "emergency vehicle", "priority", "preempt".

## 2. Related work

### 2.1. Traffic signal control

Traffic signal control (TSC) aims to allocate green time at an intersection to traffic moving in different directions. Every approach (roadway entering the intersection) is split into lanes for forward, left-turn, and (possibly) right turn movements (which may be assumed to always be permissible) [12, 13]. For efficiency, pairs of compatible movements are often arranged into phases and signalled simultaneously [10, 14, 15]. The task is to find some division of green time between phases for each intersection in a road network, which maximizes metrics such as the throughput of the network. We refer the reader to [16] for details of the problem formulation.

Different approaches to dividing green time include choosing phase durations or phase sequences, or fixing a phase sequence within a cycle and choosing the length of the cycle or the proportions of each phase within the cycle [10, 12, 15]. Three main types of algorithmic approaches exist. In fixed-time control, which has historically been a popular strategy [3], a small number of fixed plans are optimized based on past traffic data under the assumption of uniform demand. In actuated control, detector inputs (such as vehicle presence data from loop detectors) are used in conjunction with a fixed set of logical rules. Finally, adaptive control uses more complex prediction and optimization algorithms to control signalling plans [12, 16].

### 2.2. Reinforcement learning

One emerging approach to adaptive control has been reinforcement learning (RL). RL is a sequential decision-making paradigm wherein agents learn how to act through trial-and-error interactions with an environment. The goal of RL is to learn policies, which describe how agents

should act given the state of the environment. Early work in reinforcement learning during the 1980s and 1990s, which included the seminal $Q$-learning algorithm [17], relied on tabular enumeration of environment states and agent actions. RL remained relatively difficult to scale until the emergence of methods based on function approximation in the 2010s, specifically the use of neural networks for deep RL [18]. Since then, the popularity and complexity of RL has experienced explosive growth. Deep RL has also found novel applications in practical domains such as robotics, natural language processing, finance, and healthcare [19]. Transportation has been one of the most significant applications of deep RL, with tasks including autonomous driving [20], vehicle dispatching [9] and routing [21], and traffic signal control (see Section 2.3). We refer the reader to [22] for an in-depth review of the history of reinforcement learning.

The body of work that we review in this paper can be seen as a parallel to work in RL for robotics that attempts to close the gap between simulations and reality. RL methods, especially deep RL methods, require an abundance of data to learn from environmental interactions. Due to the cost of real-world data collection, simulators are often employed instead to generate large quantities of interactions. However, simulators can never perfectly emulate reality. This problem, which is referred to as the reality gap [23], has been addressed by the sim-to-real literature. Some sim-to-real methods employ randomization in sensors and controllers to learn robust policies (domain randomization); some explicitly model the reality gap and try to unify the feature spaces of the source and target environments (domain adaptation); some train policies to generalize across different tasks (meta-RL); some attempt to learn from demonstrations of behaviour in target environments (imitation learning); and others attempt to improve simulators. We refer the reader to [24, 25] for surveys of these methods. In this work, we draw parallels between some of these methods and developments in RL-based TSC. However, at the same time, TSC involves unique challenges that are usually not present in robotics. Environments in robotics where sim-to-real methods have been applied (see [24]) are usually highly controlled with well-defined objectives (e.g., [26]) and minimal interaction with other agents. However, TSC may be affected by varying environmental conditions and large numbers of road users.

## 2.3. Related reviews

Various reviews of applications of RL in TSC have been published. While each of the following reviews captures distinct aspects of the field that are highly relevant to our work, none of them have focused on the key issue of practical engineering challenges that present barriers to deployment, and — crucially — how to solve them instead of leaving them as open problems.

[27], [28], and [29] provided brief syntheses of early RL-based TSC methods in reviews of applications of AI in transportation. [30] and [31] were the first to take a systematic approach to reviewing RL-based TSC algorithms; the former performed the first experimental comparison of RL algorithms with a synthetic network, while the latter addressed data sources such as models of road networks and vehicle arrivals. Both reviewed state, action, and reward formulations. These reviews considered traditional algorithms in RL such as Q-learning and SARSA.

With the increasing popularity of deep learning to address challenges of scalability in RL, [4, 32] (the latter a follow-up to [31]) both reviewed deep RL methods for TSC and provided recommendations for designing novel deep RL-based TSC algorithms. [4] focused on choosing state, action, and reward representations, with some discussion of data processing, but did not

consider downstream challenges in deployment. [32] provided a broad overview of various algorithm and architecture designs with less of a focus on practicalities.

Both [15, 33] reviewed alternative state, action, and reward formulations among deep RL-based TSC algorithms, as well as options for inter-agent coordination and simulation-based evaluation. They outlined, but did not investigate, challenges to deployment. [15] further compared deep RL-based algorithms to traditional actuated and adaptive methods. Likewise, as part of a wider review on deep RL for intelligent traffic systems, [34] reviewed problem formulations and the history of algorithmic developments for RL-based TSC. Finally, [10] performed a highly systematic overview of the past 26 years of research in this domain that provides quantitative support for some of the patterns that we identify.

## 3. Uncertainty in detection

### 3.1. Significance of challenges

States are described in inputs to RL-based TSC algorithms using abstracted features. These include vehicles' queue lengths, positions, and speeds [10]. Many works take for granted that these state features are readily available [35]. As reported by [10], 67% of surveyed papers did not envision any specific data sources. Even in papers where potential data sources were specified, it is unclear how robust the methods would be to detector noise or failure. For instance, among algorithms that use vehicle positions as state features, [36, 37, 38, 39] all used the simulator SUMO to obtain noiseless images of single-intersection toy networks; [40] extended this approach with a 3D simulator for images from the perspectives of traffic cameras; and [41] used simulated traffic in SUMO based on flow rates from traffic camera footage. Each of these methods provides a sanitized representation that may not necessarily be representative of real-world conditions. Furthermore, the loss of information to noise may cause state aliasing [42], which hinders the generalizability of learned policies to different demand scenarios [43].

### 3.2. Lessons from deployments

Types of instruments for traffic sensing include intrusive detectors (installed into the road surface) and non-intrusive detectors (mounted above the road surface) [44, 45]. Among intrusive detectors, loop detectors are relatively inexpensive, accurate, and robust to weather and time of day, but they are also highly vulnerable to wear and tear [46]. When they fail, loop detectors are being increasingly replaced by non-intrusive detectors such as video-based and radar detection systems [44], which can be flexibly reconfigured to detect different road segments and vehicle types. However, the accuracy of these systems degrades in inclement weather, and video detectors are also inaccurate at night and on high-speed roads [45, 47]. RL-based signal controllers must be designed with these limitations in mind; learning ensembles of models [48] to capture the strengths of different detectors may improve robustness. Although data about speed and position from connected vehicles can be useful, penetration remains low, so they must be integrated with traditional detector data. [49] showed in simulations that connected vehicle data could improve adaptive control even with limited penetration. Furthermore, agencies may configure their detectors differently. To account for uncertainty in vehicle stopping positions,

for instance, the size of the detection zone behind the stop bar may vary [50]; detectors may also report data at different frequencies [51]. Thus, verifying the mapping from real detector data to abstract state representations is an important task for RL-based TSC.

Agencies often address problems in detection by modifying their detection setup [44] or by configuring parameters such as passage time (i.e., the amount of time that a phase is extended for upon actuation) [45]. [5] explicitly addressed error in queue length detection for their adaptive controller SURTRAC. To mitigate underestimation, they used heuristics based on differences in vehicle counts reported by advance and stop bar detectors [52]. They considered overestimation acceptable, as it provides the algorithm with buffer time; similarly, [53] found that moderate queue length overestimation significantly improves the performance of adaptive control.

### 3.3. Progress toward solutions

Two lines of work within RL-based TSC have the potential to address detection uncertainty.

First, various authors have investigated the effects of reducing the dimensionality of the state space. In particular, [3] showed that complex image representations of intersection state achieve inferior performance compared to a simple representation containing only vehicle counts and phases. [54] reached similar conclusions with a state representation based on queue length. Both papers also provided optimality results that connected these formulations to traditional methods in TSC. Meanwhile, [43, 55] investigated the effects of switching to coarser state representations with a single algorithm. [55] found that occupancy and speed data (e.g., from loop detectors) yielded near-identical performance to high-fidelity position data (e.g., from cameras). However, the experiments of [43] suggested that coarser state discretizations harm generalization across sudden shifts in traffic flow. Regardless, simpler state representations could facilitate identification and debugging of issues caused by detection uncertainty.

Second, other work has attempted to imbue RL-based TSC algorithms with robustness to detection uncertainty. Several methods are analogous to domain randomization in the sim-to-real literature [26, 56]. The approach of [57] is closest to the sim-to-real literature: they randomize weather and lighting conditions in their traffic simulator and train policies based on the resulting images. [58] applied Dropout to neural network units to prevent overfitting and thus to learn robust policies. They evaluated their algorithm with a simulation of probabilistic detector failure. As is done in adversarial machine learning, [59] injected Gaussian noise into queue length observations, and validated their approach with simulations where trucks cause vehicle count overestimation. Meanwhile, to handle miscalibrated measurements, [35] combined next state prediction with imitation learning from a real traffic controller (SCOOTS), [60] used autoencoders to denoise input data, and [61] evaluated the effects of lane-blocking incidents and detector noise on performance. Finally, in a growing body of work that uses connected vehicle data for RL, [62] was the first to explicitly address partial observability by adding the phase duration into the state space to learn its indirect impact on delay.

Overall, these methods are helpful approaches for improving the robustness of RL-based TSC to detection uncertainty. However, they should be designed and tuned to address the challenges of specific deployments, leveraging past knowledge to identify and address potential causes of detector noise or failure. It may also help to model partial observability as part of the problem.

# 4. Reliability of communications

## 4.1. Significance of challenges

Some level of controller decentralization is often applied in RL-based TSC, because the computational cost of RL may be prohibitive when the state and action space dimensionalities are high. At the same time, to ensure that controllers take the traffic conditions of other intersections into account for signalling decisions, a growing number of works have implemented mechanisms for inter-intersection coordination [33]. Typical approaches involve sharing states [63, 64, 65, 66, 67, 68], actions [69], or hidden state representations from neural networks [70, 71] between controllers for neighbouring intersections. While much of this work has focused on designing neural network architectures to leverage shared information (such as graph neural networks [66, 67, 70, 71]), less attention has been devoted to the mechanisms by which information must be exchanged in the first place. If there are inconsistencies in the availability of communication infrastructure and detectors between intersections (see also Section 3), it is unclear how they may affect the performance of RL-based TSC.

## 4.2. Lessons from deployments

In practice, signal controllers are commonly deployed as part of closed-loop systems, where control is distributed over three levels. At the top level, traffic management centres (TMCs) make policy-based signalling decisions, often involving dialogue with other stakeholders. These decisions are used to configure field master controllers (FMCs), which are installed on-site and coordinate multiple local intersection controllers (LICs) [72]. Each FMC aggregates traffic conditions reported by connected LICs to make signalling decisions over a small region; FMCs also synchronize the clocks of LICs to ensure that they are coordinated [12, 14]. As 90% of TSC systems in the United States are closed-loop [73], upgrades to adaptive control have largely been implemented within this hierarchical organization [51]. LICs may make some limited decisions based on local traffic conditions, but coordination is still largely delegated to FMCs even in adaptive control [72]. Transitioning to adaptive control has also required agencies to update to Type 2070 or ATC controllers [12], but some controllers in road networks may retain relatively outdated hardware [14]. RL-based signal controllers will likely be deployed into such ecosystems, where control is distributed hierarchically and different intersections have different capabilities for control and/or detection. Thus, algorithms based on techniques for domain adaptation from the sim-to-real literature may be helpful.

Messages are sent between controllers and TMCs using multiple communication media in modern TSC systems [12]. For wired connections, fibre optic cables are increasingly replacing traditional copper wires or coaxial cables. Wireless communication systems implemented using radio or Wi-Fi are also becoming increasingly common [44]. Thus, communication bandwidth is not likely to be a concern, except in jurisdictions where fibre optic infrastructure is not readily available. However, a major issue reported by agencies in [44] was connection reliability: poor signal strength often results in data loss or latency. In terms of data formatting, the NTCIP 1202 standard includes standard object definitions for actuated signal controllers, which has also been used for adaptive systems [73]. Communications for RL would need to fit into this standard, at least until it is updated (as has already been done for connected vehicles) [74].

In SURTRAC, [5] encoded data for communication between neighbouring intersections using JSON messages with standard types.

## 4.3. Progress toward solutions

One line of work in RL-based TSC has sought to learn more compact representations of information. Although bandwidth is not a concern, reducing message dimensionality could still mitigate the impact of communication failures. Several algorithms directly exchange state values of learned policies instead of learning from exchanged state representations. In [75, 76], state values are directly exchanged between neighbours and weighted; [37, 77, 78] leveraged the max-plus algorithm for coordination graphs, which is known to converge to near-optimality even for cyclic graphs [79]. Meanwhile, [80] designed an architecture to exchange information from the previous time step to ensure robustness to latency, and showed that it asymptotically reduces communication relative to neighbour-based approaches by 50%. [81] demonstrated that cumulative rewards can be estimated based only on vehicle counts on inbound approaches.

Some work has also focused on designing RL-based TSC algorithms for hierarchically distributed frameworks of communication and control, which could improve RL's robustness, scalability, and applicability for deployment in closed-loop systems. [82] implemented a two-level architecture where LICs can either act independently or receive joint actions from FMCs based on predictions of the regional traffic state. [63] introduced a feudal RL algorithm, in which "manager" controllers do not directly control the actions of "worker" controllers, but instead set goals that influence their rewards. [83] trained multiple sub-policies that minimize various proxy metrics such as queue length and waiting time, and a high-level controller that adaptively delegates control to sub-policies to minimize the longer-term metric of travel time. However, all of these architectures are conceptual and further work is needed to deploy them.

# 5. Compliance and interpretability

## 5.1. Significance of challenges

At the heart of the fact that RL-based TSC algorithms have not been deployed are the potential regulatory and safety risks that are introduced by RL [15, 34]. The issue of trust and safety for RL is by no means exclusive to the domain of TSC [84, 85, 86], but in this case the stakes are high because contollers must interact with a large number of human users and mistakes may have fatal consequences. For RL-based signal controllers to be trusted, we need to assess — both prospectively or retrospectively — whether their decisions comply with standards and reasonable expectations [87]. However, the proliferation of deep RL algorithms based on complicated state representations runs counter to this goal, as assessment of compliance is not possible if we cannot understand or at least verify their decisions. At the same time, issues of interpretability and safety have rarely been discussed in the literature on RL-based TSC [10] and are more often mentioned as desiderata for future work in reviews [10, 15, 34].

## 5.2. Lessons from deployments

In the real world, regulatory frameworks for traffic signalling are often scattershot. In the United States, the federal *Manual on Uniform Traffic Control Devices* [88] includes standards about the necessity, meaning, and placement of different traffic signals. Many of these standards involve the control of individual movement signals, which would be abstracted away from RL through phase-based action space definitions. However, factors such as yellow change and red clearance intervals are left to "engineering judgement". States may impose further requirements on signal timing plans based on regional transportation policies [14]. In a review of signal timing policies for 15 states, [89] found recommendations for factors such as minimum green, yellow change, and red clearance intervals, as well as when to serve turn movements. Such recommendations should be incorporated into the design of the RL action space, as was done by [5] who treated safety constraints as inputs to SURTRAC. Yet, these recommendations can also be arbitrary and dependent on data (e.g., vehicle and pedestrian clearing times [89]), and algorithmic approaches to stakeholder preference learning [90] may help to find better values.

One common strategy to ensure the safety of signal timing plans is to review common types and causes of crashes in historical data [89]. Naturally, this is a reactive approach that requires crashes to happen in the first place, and crash reports may also be biased by severity or by environmental conditions [14]. Accident modification factors (AMFs) are a popular method of quantitative analysis; they statistically estimate the effectiveness of particular changes to signal timing plans based on their expected reductions in crash rate [91, 92, 93]. We are unaware of any work in RL that estimates or uses AMFs, but they may be a valuable pathway to interpretability. The *Highway Safety Manual* also provides standard crash risk assessment models, but these models often require extensive tuning to local conditions [94, 95, 96].

## 5.3. Progress toward solutions

Some work has enhanced the interpretability of RL-based TSC through algorithm design. [97] focused on learning surrogate policies that are regulatable, i.e. monotonic in state variables, which allows parameters to be viewed as weights. [98] learned human-auditable decision tree surrogates using VIPER, an algorithm that identifies critical states where suboptimality harms future rewards. Closer to the literature on interpretability for machine learning, [99] used SHAP values to analyze how induction loop detections contribute to choices of phases for a controller in a simulated roundabout. They found that advance detectors have higher SHAP values as they are more indicative of congestion. Similarly, [57] used Grad-CAM to generate heatmaps for image-based inputs. Instead of directly interfacing with the simulator, [100] used logical rules based on signal controllers to post-process RL policy outputs for ensuring compliance.

Further work has applied heuristic modifications to RL algorithms to enforce safety. [101] prevented their system from taking actions when pedestrians are detected in crosswalks, and enforced minimum green times for pedestrians. [102] drew on their models of rear-end conflict rates (based on various observable intersection state features [103]) to design a reward formulation that minimizes such conflicts. Similarly, [104] used a binary logistic crash risk model to define crash penalties while also minimizing waiting time. Using a state formulation based on individual signals, [105] regularized the red light duration of signalling plans to mitigate unsafe

behaviour caused by driver frustration with extended red lights. [106] included yellow change intervals in their action space and added a penalty for emergency braking by vehicles.

While we have reviewed many promising methods that have been developed for the interpretability and safety of RL-based TSC, more work is still needed on determining which of these methods correspond well to stakeholder requirements. Furthermore, there is a substantial literature on <u>safe reinforcement learning</u> using constrained optimization [107, 108, 109], which has hitherto not been applied to TSC; it is likely that such work can provide more rigorous theoretical guarantees about algorithm behaviour. We also believe that, to deal with safety failures ethically, work is needed in algorithmic accountability for RL-based signal controllers.

## 6. Heterogeneous road users

### 6.1. Significance of challenges

Traditional models of traffic flow used for TSC assume, simplistically, that all vehicles are identical [110, 111]. In reality, the assumption of identical or even unimodal traffic is often unrealistic, because many types of vehicles and road users — each with different needs and behavioural patterns — interact with each other on roads. RL algorithms can still implicitly encode these assumptions through simplistic state spaces, since common state variables such as queue length and vehicle position [15] do not account for inter-vehicle variation. Although such state formulations can be helpful for deriving optimality results based on traditional models in TSC [3, 54], it is unclear how these assumptions may impact the performance and safety of RL-based signal controllers in practice, especially because road users such as pedestrians and cyclists may behave non-intuitively. Dedicated simulators developed for RL-based TSC likewise abstract away inter-vehicle variation [112]. [10] found in 160 papers on RL-based TSC that only three accounted for non-private vehicle types, and only one accounted for pedestrians.

### 6.2. Lessons from deployments

In practice, agencies make a variety of adjustments to signalling plans to accommodate different classes of road users other than regular passenger vehicles, including pedestrians, cyclists, transit vehicles, and emergency vehicles [14]. In this section, we focus on current practice in the field for pedestrians and transit/emergency vehicles. When balancing the needs of different road user classes in RL-based signal controllers, stakeholders' requirements should be taken into account; in the US, for instance, agencies' opinions differ on whether preemption for trains should take priority over pedestrians [89].

For pedestrians, the simplest option is for the pedestrian signal to be activated in the direction of the through movement, as is implicitly assumed by many works in RL and made explicit in some (e.g., [113]). However, doing so may cause pedestrians to impede the flow of left-turning and right-turning traffic, which creates safety hazards. In practice, <u>leading pedestrian intervals</u> (LePIs) mitigate this risk by allowing pedestrians to start crossing before cars are permitted to make turns [14]. Alternative phase sequence designs add lagging pedestrian intervals (after turning phases) or phases exclusively for pedestrians. [114] developed a benefit-cost model to assess the safety-delay tradeoffs for LePIs at individual intersections. Beyond safety, additional

work has tried to minimize the delay of pedestrians so that they are treated equitably compared to drivers, as codified by regulations in Germany, the UK, and China [115]. For the deployed SURTRAC system, [116] adaptively set pedestrian walk intervals based on predicted phase lengths to avoid cutting them short, while [117] considered using vehicular volumes and pedestrian actuation frequencies to switch between controller modes. We are unaware of any work in RL that has explicitly included LePIs as part of the action space formulation.

As for handling transit and emergency vehicles, typical strategies include the prioritization and preemption of signals. Prioritization handles requests made by vehicles through vehicle-to-infrastructure (V2I) communications, and may or may not result in adjustments to signalling plans. Meanwhile, preemption (often used for firetrucks or trains) deterministically replaces the signal plan with a predefined routine that favours the preempting vehicle. Typically, signal controllers need multiple cycles after preemption to recover from the interruption [14]. The adaptive SCATS controller natively implements both prioritization and preemption; compared to prior practice, [118] found that SCATS' performance improvements were robust to prioritization, and [119] found that it could reduce recovery time from preemption. These results suggest the potential of implementing prioritization and preemption with RL-based methods; in particular, explicit modelling of recovery from preemption may further improve recovery times. In addition to interactions at intersections, RL-based signal controllers should also consider the effects of transit and emergency vehicles on traffic between intersections. For instance, when buses are stopped on roads, they may block other traffic from passing. As initial steps towards implementing bus prioritization in the SURTRAC system, [120] delayed the allocation of green time in intersections located downstream from stopped buses, and [121] predicted bus dwelling times at stops by leveraging V2I communications.

## 6.3. Progress toward solutions

One paper in RL-based TSC was cited by [10] as explicitly modelling pedestrians: [101] defined the reward using the weighted average of the local intersection's vehicular queue length, neighbouring intersections' vehicular queue lengths, and the local intersection's pedestrian queue length. Beyond this paper, several other works have explicitly considered pedestrians as part of the problem formulation. [122] likewise addressed joint vehicle-pedestrian control at intersections, but made no assumptions about pedestrian detector capabilities. [123] used deep RL to control a signalized crosswalk across a road (with the actions being to set the pedestrian signal to green or red), and found that it outperformed actuation under moderate levels of pedestrian demand in simulations. [61] analyzed the performance of RL-based TSC in the presence of jaywalking pedestrians that cause vehicles to slow.

Several works in RL-based TSC have also considered prioritization and preemption. For prioritization, [57] upweighted buses and emergency vehicles in their throughput-based reward formulation; [124] used a state representation based on the cell transmission traffic model and modelled priority as a binary variable; [125] adopted an implicit approach based on minimizing delay per person instead of per vehicle; [126] and [127] both considered prioritization for trams, with the former's rewards being based on tram schedule adherence and the latter using model predictive control to model driver behaviour; and [128] adaptively altered vehicles' priorities depending on queue length, waiting time, and emergency vehicle presence. For preemption,

[129] learned TSC policies for emergency vehicle routing with rewards that encourage low vehicle density, and [130] used RL to learn policies for notifying connected vehicles to clear out lanes for emergency vehicles to pass.

Lastly, [100] included demand data from the field for multiple types of road users — including pedestrians, cyclists, motorcyclists, trucks, and buses — in their benchmark simulation for RL-based TSC, LemgoRL, which is based on a real road network; they also included pedestrian waiting times in rewards and enforced minimum pedestrian green times. There is a need to connect high-fidelity simulations such as LemgoRL to the various approaches for handling different road user classes that we outlined above, so as to ensure their ecological validity.

## 7. Conclusion

We have reviewed four barriers to the deployment of RL-based controllers for TSC. Each of these barriers has been insufficiently addressed by the majority of new work in RL-based TSC, which has focused on algorithmic contributions. However, TSC algorithms do not exist in a vacuum — they must be trained based on data from detectors, interface with signals through controllers, and control the movements of a variety of road users. Challenges both intrinsic to RL algorithms and in other pipeline components may cascade into failures with significant implications for the efficiency and safety of transportation infrastructure. Based on our literature review, we suggested ways in which further work in RL-based TSC could address these challenges.

Echoing the recommendations of [11], we emphasize the importance of engaging in consultation with agency stakeholders and experts in TSC for RL practitioners. This can break down information silos that would otherwise prevent the recognition of issues during requirements engineering and integration (cf. [131]); we could not have identified these challenges ourselves without engaging with the literature on traditional TSC. Additionally, as we discussed, the practicalities of these challenges — including the availability and configuration of detectors, signalling constraints, and the priorities of different road users — will often vary depending on the statuses of road networks and their responsible agencies. While benchmark simulations based on synthetic networks facilitate evaluation, we advocate for the creation of more simulations like [100] that incorporate realistic domain constraints. RL algorithms that are trained using such benchmarks would likely have better generalizability and robustness in deployments.

More generally, we uncovered a diversity of work that addresses each challenge, which previous reviews of TSC have not comprehensively surveyed. This suggests that RL-based TSC is closer to deployment than might be suggested by a review of state-of-the-art methods. If future developments focus on combining algorithmic improvements with both real-world considerations and reproducibility techniques to facilitate collaboration [132], we believe that the integration of RL to improve real-world transportation infrastructure is within reach.

## Acknowledgments

# References

[1] D. Schrank, L. Albert, B. Eisele, T. Lomax, 2021 Urban Mobility Report, Technical Report, Texas A&M Transportation Institute, 2021.

[2] S. Chin, O. Franzese, D. Greene, H. Hwang, R. Gibson, Temporary losses of highway capacity and impacts on performance: Phase 2, Technical Report ORNL/TM-2004/209, Oak Ridge National Laboratory, 2004.

[3] G. Zheng, X. Zang, N. Xu, H. Wei, Z. Yu, V. Gayah, K. Xu, Z. Li, Diagnosing reinforcement learning for traffic signal control, arXiv preprint (2019). `arXiv:1905.04716`.

[4] M. Gregurić, M. Vujić, C. Alexopoulos, M. Miletić, Application of deep reinforcement learning in traffic signal control: An overview and impact of open traffic data, Applied Sciences 10 (2020) 4011.

[5] S. Smith, G. Barlow, X.-F. Xie, Z. Rubinstein, Smart urban signal networks: Initial application of the SURTRAC adaptive traffic signal control system, in: Proceedings of the 23rd International Conference on Automated Planning and Scheduling, ICAPS '13, 2013, pp. 434–442.

[6] C. Chen, H. Wei, N. Xu, G. Zheng, M. Yang, Y. Xiong, K. Xu, Z. Li, Toward a thousand lights: Decentralized deep reinforcement learning for large-scale traffic signal control, in: Proceedings of the 34th AAAI Conference on Artificial Intelligence, AAAI '20, 2020, pp. 3414–3421.

[7] N. Brown, T. Sandholm, Superhuman AI for heads-up no-limit poker: Libratus beats top professionals, Science 359 (2017) 418–424.

[8] O. Vinyals, I. Babuschkin, W. M. C. amd Michaël Mathieu, A. Dudzik, J. Chung, D. H. Choi, R. Powell, T. Ewalds, P. Georgiev, J. Oh, D. Horgan, M. Kroiss, I. Danihelka, A. Huang, L. Sifre, T. Cai, J. P. Agapiou, M. Jaderberg, A. S. Vezhnevets, R. Leblond, T. Pohlen, V. Dalibard, D. Budden, Y. Sulsky, J. Molloy, T. L. Paine, C. Gulcehre, Z. Wang, T. Pfaff, Y. Wu, R. Ring, D. Yogatama, D. Wünsch, K. McKinney, O. Smith, T. Schaul, T. Lillicrap, K. Kavukcuoglu, D. Hassabis, C. Apps, D. Silver, Grandmaster level in StarCraft II using multi-agent reinforcement learning, Nature 575 (2019) 350–354.

[9] Z. T. Qin, X. Tang, Y. Jiao, F. Zhang, Z. Xu, H. Zhu, J. Ye, Ride-hailing order dispatching at didi via reinforcement learning, INFORMS Journal on Applied Analytics 50 (2020) 272–286.

[10] M. Noaeen, A. Naik, L. Goodman, J. Crebo, T. Abrar, Z. S. H. Abad, A. L. Bazzan, B. Far, Reinforcement learning in urban network traffic signal control: A systematic literature review, Expert Systems with Applications 199 (2022) 116830.

[11] A. Perrault, F. Fang, A. Sinha, M. Tambe, AI for social impact: Learning and planning in the data-to-deployment pipeline, arXiv preprint (2019). `arXiv:2001.00088`.

[12] R. L. Gordon, W. Tighe, Traffic Control Systems Handbook, Federal Highway Administration, 2005.

[13] G. Zheng, Y. Xiong, X. Zang, J. Feng, H. Wei, H. Zhang, Y. Li, K. Xu, Z. Li, Learning phase competition for traffic signal control, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM '19, 2021, pp. 1963–1972.

[14] P. Koonce, L. Rodegerdts, K. Lee, S. Quayle, S. Beaird, C. Braud, J. Bonneson, P. Tarnoff, T. Urbanik, Traffic Signal Timing Manual, Federal Highway Administration, 2008.

[15] H. Wei, G. Zheng, V. Gayah, Z. Li, A survey on traffic signal control methods, arXiv preprint (2019). `arXiv:1904.08117`.

[16] M. Eom, B.-I. Kim, The traffic signal control problem for intersections: a review, European Transport Research Review 12 (2020) 50.

[17] C. J. C. H. Watkins, P. Dayan, Q-learning, Machine Learning 8 (1992) 279–292.

[18] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, D. Hassabis, Human-level control through deep reinforcement learning, Nature 518 (2015) 529–533.

[19] Y. Li, Deep reinforcement learning, arXiv preprint (2018). `arXiv:1810.06339`.

[20] B. R. Kiran, I. Sobh, V. Talpaert, P. Mannion, A. A. A. Sallab, S. Yogamani, P. Pérez, Deep reinforcement learning for autonomous driving: A survey, IEEE Transactions on Intelligent Transportation Systems (2021).

[21] M. Nazari, A. Oroojlooy, M. Takáč, L. V. Snyder, Reinforcement learning for solving the vehicle routing problem, in: Proceedings of the 32nd International Conference on Neural Information Processing Systems, NIPS '18, 2018, pp. 9861–9871.

[22] R. S. Sutton, A. G. Barto, Early history of reinforcement learning, in: Reinforcement Learning: An Introduction, The MIT Press, 2018, pp. 11–17.

[23] J.-B. Mouret, K. Chatzilygeroudis, 20 years of reality gap: a few thoughts about simulators in evolutionary robotics, in: Proceedings of the 2017 Genetic and Evolutionary Computation Conference Companion, GECCO '17, 2017, pp. 1121–1124.

[24] W. Zhao, J. P. Queralta, T. Westerlund, Sim-to-real transfer in deep reinforcement learning for robotics: a survey, in: Proceedings of the 2020 IEEE Symposium Series on Computational Intelligence, SSCI '20, 2020, pp. 737–744.

[25] K. Dimitropoulos, I. Hatzilygeroudis, K. Chatzilygeroudis, A brief survey of Sim2Real methods for robot learning, in: Proceedings of the 2022 International Conference on Robotics in Alpe-Adria Danube Region, RAAD '22, 2022, pp. 133–140.

[26] M. Andrychowicz, B. Baker, M. Chociej, R. Józefowicz, B. McGrew, J. Pachocki, A. Petron, M. Plappert, G. Powell, A. Ray, J. Schneider, S. Sidor, J. Tobin, P. Welinder, L. Weng, W. Zaremba, Learning dexterous in-hand manipulation, The International Journal of Robotics Research 39 (2020) 3–20.

[27] B. Abdulhai, L. Kattan, Reinforcement learning: Introduction to theory and potential for transport applications, Canadian Journal of Civil Engineering 30 (2003) 981–991.

[28] A. L. C. Bazzan, Opportunities for multiagent systems and multiagent reinforcement learning in traffic control, Autonomous Agents and Multi-Agent Systems 18 (2009) 342–375.

[29] A. L. C. Bazzan, F. Klügl, A review on agent-based technology for traffic and transportation, The Knowledge Engineering Review 29 (2013) 375–403.

[30] P. Mannion, J. Duggan, E. Howley, An experimental review of reinforcement learning algorithms for adaptive traffic signal control, in: Autonomic Road Transport Support Systems, Springer, 2016, pp. 47–66.

[31] K.-L. A. Yau, J. Qadir, H. L. Khoo, M. H. Ling, P. Komisarczuk, A survey on reinforcement learning models and algorithms for traffic signal control, ACM Computing Surveys 50 (2017) 34.

[32] F. Rasheed, K.-L. A. Yau, R. M. Noor, C. Wu, Y.-C. Low, Deep reinforcement learning for traffic signal control: A review, IEEE Access 8 (2020) 208016–208044.

[33] H. Wei, G. Zheng, V. Gayah, Z. Li, Recent advances in reinforcement learning for traffic signal control: A survey of models and evaluation, ACM SIGKDD Explorations Newsletter 22 (2021) 12–18.

[34] A. Haydari, Y. Yilmaz, Deep reinforcement learning for intelligent transportation systems: A survey, IEEE Transactions on Intelligent Transportation Systems 23 (2022) 11–32.

[35] H. Wang, Y. Yuan, X. T. Yang, T. Zhao, Y. Liu, Deep Q learning-based traffic signal control algorithms: Model development and evaluation with field data, Journal of Intelligent Transportation Systems (2022).

[36] W. Genders, S. Razavi, Using a deep reinforcement learning agent for traffic signal control, arXiv preprint (2016). arXiv:1611.01142.

[37] E. van der Pol, F. A. Oliehoek, Coordinated deep reinforcement learners for traffic light control, in: Proceedings of the 30th Conference on Neural Information Processing Systems, NIPS '16, 2016, pp. 1–8.

[38] S. S. Mousavi, M. Schukat, E. Howley, Traffic light control using deep policy-gradient and value-function based reinforcement learning, IET Intelligent Transport Systems 11 (2017) 417–423.

[39] X. Liang, X. Du, G. Wang, Z. Han, A deep reinforcement learning network for traffic light cycle control, IEEE Transactions on Vehicular Technology 68 (2019) 1243–1253.

[40] D. Garg, M. Chli, G. Vogiatzis, Deep reinforcement learning for autonomous traffic light control, in: Proceedings of the 2018 3rd International Conference on Intelligent Transportation Engineering, ICITE '18, 2018, pp. 214–218.

[41] H. Wei, G. Zheng, H. Yao, Z. Li, IntelliLight: A reinforcement learning approach for intelligent traffic light control, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '18, 2018, pp. 2496–2505.

[42] M. T. J. Spaan, N. Vlassis, A point-based pomdp algorithm for robot planning, in: Proceedings of the 2004 IEEE International Conference on Robotics and Automation, ICRA '04, 2004, pp. 2399–2404.

[43] L. N. Alegre, A. L. Bazzan, B. C. da Silva, Quantifying the impact of non-stationarity in reinforcement learning-based traffic signal control, PeerJ Computer Science 7 (2021) e575.

[44] D. Sun, L. Dodoo, A. Rubio, H. K. Penumala, M. Pratt, S. Sunkari, Synthesis study of Texas signal control systems: technical report, Technical Report FHWA/TX-13/0-6670-1, Texas A&M Transportation Institute, 2012.

[45] S. Sunkari, A. Bibeka, N. Chaudhary, K. Balke, Impact of Traffic Signal Controller Settings on the Use of Advanced Detection Devices, Technical Report FHWA/TX-18/0-6934-R1, Texas A&M Transportation Institute, 2019.

[46] D. Gibson, M. K. P. Mills, D. R. Jr., Staying in the loop: The search for improved reliability of traffic sensing systems through smart test instruments, Public Roads 62 (1998).

[47] A. Rhodes, D. M. Bullock, J. R. Sturdevant, Z. T. Clark, Evaluation of Stop Bar Video Detection Accuracy at Signalized Intersections, Technical Report FHWA/IN/JTRP-2005/28, Joint Transportation Research Program, Indiana Department of Transportation and Purdue University, 2005.

[48] K. Lee, M. Laskin, A. Srinivas, P. Abbeel, SUNRISE: A simple unified framework for ensemble learning in deep reinforcement learning, in: Proceedings of the 38th International Conference on Machine Learning, ICML '21, 2021, pp. 6131–6141.

[49] S. M. A. B. A. Islam, M. Tajalli, R. Mohebifard, A. Hajbabaie, Effects of connectivity and traffic observability on an adaptive traffic signal control system, Transportation Research Record 2675 (2021) 800–814.

[50] A. M. T. Emtenan, C. M. Day, Impact of detector configuration on performance measurement and signal operations, Transportation Research Record 2674 (2020) 300–313.

[51] F. Luyanda, D. Gettman, L. Head, S. Shelby, D. Bullock, P. Mirchandani, ACS-Lite algorithmic architecture: Applying adaptive control system technology to closed-loop traffic signal control systems, Transportation Research Record 1856 (2003) 175–184.

[52] X.-F. Xie, G. J. Barlow, S. F. Smith, Z. B. Rubinstein, Accounting for Real-World Uncertainty in Real-Time Adaptive Traffic Control, Technical Report ATCSTR12, Carnegie Mellon University, 2012.

[53] C. Cai, B. Hengst, G. Ye, E. Huang, Y. Wang, C. Aydos, G. Geers, On the performance of adaptive traffic signal control, in: Proceedings of the Second International Workshop on Computational Transportation Science, ICWTS '09, 2009, pp. 37–42.

[54] H. Wei, C. Chen, G. Zheng, K. Wu, V. Gayah, K. Xu, Z. Li, PressLight: Learning max pressure control to coordinate traffic signals in arterial network, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '19, 2019, pp. 1290–1298.

[55] W. Genders, S. Razavi, Evaluating reinforcement learning state representations for adaptive traffic signal control, in: Proceedings of the 9th International Conference on Ambient Systems, Networks and Technologies, ANT '18, 2018, pp. 26–33.

[56] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, P. Abbeel, Domain randomization for transferring deep neural networks from simulation to the real world, in: Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS '17, 2017, pp. 23–30.

[57] D. Garg, M. Chli, G. Vogiatzis, Fully-autonomous, vision-based traffic signal control: from simulation to reality, in: Proceedings of the 21th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '22, 2022, pp. 454–462.

[58] F. Rodrigues, C. L. Azevedo, Towards robust deep reinforcement learning for traffic signal control: Demand surges, incidents and sensor failures, in: Proceedings of the 2019 International Conference on Intelligent Transportation Systems, ITSC '19, 2019, pp. 3559–3566.

[59] K. L. Tan, A. Sharma, S. Sarkar, Robust deep reinforcement learning for traffic signal control, Journal of Big Data Analytics in Transportation 2 (2020) 263–274.

[60] C. Li, F. Yan, Y. Zhou, J. Wu, X. Wang, A regional traffic signal control strategy with deep reinforcement learning, in: Proceedings of the 37th Chinese Control Conference, CCC '18, 2018, pp. 7690—7695.

[61] M. Aslani, S. Seipel, M. S. Mesgari, M. Wiering, Traffic signal optimization through discrete and continuous reinforcement learning with robustness analysis in downtown Tehran, Advanced Engineering Informatics 38 (2018) 639–655.

[62] W. Li, M. Zhao, Y. Fu, K. Ruan, X. Di, CVLight: Decentralized learning for adaptive traffic

signal control with connected vehicles, arXiv preprint (2021). `arXiv:2104.10340`.

[63] J. Ma, F. Wu, Feudal multi-agent deep reinforcement learning for traffic signal control, in: Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '20, 2020, pp. 816–824.

[64] T. Chu, J. Wang, L. Codecà, Z. Li, Multi-agent deep reinforcement learning for large-scale traffic signal control, IEEE Transactions on Intelligent Transportation Systems 21 (2020) 1086–1095.

[65] M. Xu, J. Wu, L. Huang, R. Zhou, T. Wang, D. Hu, Network-wide traffic signal control based on the discovery of critical nodes and deep reinforcement learning, Journal of Intelligent Transportation Systems 24 (2020) 1–10.

[66] M. Wang, L. Wu, J. Li, L. He, Traffic signal control with reinforcement learning based on region-aware cooperative strategy, IEEE Transactions on Intelligent Transportation Systems (2021).

[67] Z. Zeng, GraphLight: Graph-based reinforcement learning for traffic signal control, in: Proceedings of the 6th International Conference on Computer and Communication Systems, ICCCS '21, 2021, pp. 645–650.

[68] P. Zhou, T. Braud, A. Alhilal, P. Hui, J. Kangasharju, ERL: Edge based reinforcement learning for optimized urban traffic light control, in: Proceedings of the 3rd International Workshop on Smart Edge Computing and Networking, SmartEdge '19, 2019, pp. 849–854.

[69] H. Ge, Y. Song, C. Wu, J. Ren, G. Tan, Cooperative deep Q-learning with Q-value transfer for multi-intersection signal control, IEEE Access 7 (2019) 40797–40809.

[70] H. Wei, N. Xu, H. Zhang, G. Zheng, X. Zang, C. Chen, W. Zhang, Y. Zhu, K. Xu, Z. Li, CoLight: Learning network-level cooperation for traffic signal control, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM '19, 2019, pp. 1913–1922.

[71] T. Nishi, K. Otaki, K. Hayakawa, T. Yoshimura, Traffic signal control based on reinforcement learning with graph convolutional neural nets, in: Proceedings of the 2018 International Conference on Intelligent Transportation Systems, ITSC '18, 2018, pp. 877–883.

[72] H. Chen, J. J. Lu, Comparison of current practical adaptive traffic control systems, in: Proceedings of the 10th International Conference of Chinese Transportation Professionals, ICCTP '10, 2010, pp. 1611–1619.

[73] D. Gettman, S. G. Shelby, L. Head, D. M. Bullock, N. Soyke, Data-driven algorithms for real-time adaptive tuning of offsets in coordinated traffic signal systems, Transportation Research Record 2035 (2007) 1–9.

[74] Z. Huang, E. Leslie, A. Balse, Infrastructure Connectivity Certification Test Procedures for Infrastructure-Based Connected Automated Vehicle Components: Test Procedures, Signal Phase and Timing — NTCIP 1202 v03, Technical Report FHWA-JPO-20-802, Leidos, 2019.

[75] Y. Wang, S. Geng, Q. Li, Intelligent transportation control based on proactive complex event processing, in: Proceedings of the 3rd International Conference on Mechanics and Mechatronics Research, ICMMR '16, 2016, pp. 1–5.

[76] W. Liu, G. Qin, Y. He, F. Jiang, Distributed cooperative reinforcement learning-based traffic signal control that integrates V2X networks' dynamic clustering, IEEE Transactions

on Vehicular Technology 66 (2017) 8667–8681.

[77] S. Yang, B. Yang, H.-S. Wong, Z. Kang, Cooperative traffic signal control using multi-step return and off-policy asynchronous advantage actor-critic graph algorithm, Knowledge-Based Systems 183 (2019) 104855.

[78] T. Chu, J. Wang, Traffic signal control by distributed reinforcement learning with min-sum communication, in: Proceedings of the 2017 American Control Conference, ACC '17, 2017, pp. 5095–5100.

[79] J. R. Kok, N. Vlassis, Using the max-plus algorithm for multiagent decision making in coordination graphs, in: Proceedings of the Fourth Robot Soccer World Cup, RoboCup '05, 2005, pp. 1–12.

[80] D. Xie, Z. Wang, C. Chen, D. Dong, IEDQN: Information exchange DQN with a centralized coordinator for traffic signal control, in: Proceedings of the 2020 International Joint Conference on Neural Networks, IJCNN '20, 2020, pp. 1–8.

[81] Q. Jiang, M. Qin, S. Shi, W. Sun, B. Zheng, Multi-agent reinforcement learning for traffic signal control through universal communication method, arXiv preprint (2022). `arXiv:2204.12190`.

[82] M. Abdoos, A. L. Bazzan, Hierarchical traffic signal optimization using reinforcement learning and traffic prediction with long-short term memory, Expert Systems with Applications 171 (2021) 114580.

[83] B. Xu, Y. Wang, Z. Wang, H. Jia, Z. Lu, Hierarchically and cooperatively learning traffic signal control, in: Proceedings of the 35th AAAI Conference on Artificial Intelligence, AAAI '21, 2021, pp. 1–9.

[84] L. Brunke, M. Greeff, A. W. Hall, Z. Yuan, S. Zhou, J. Panerati, A. P. Schoellig, Safe learning in robotics: From learning-based control to safe reinforcement learning, Annual Review of Control, Robotics, and Autonomous Systems 5 (2022).

[85] J. García, F. Fernández, A comprehensive survey on safe reinforcement learning, Journal of Machine Learning Research 16 (2015) 1437–1480.

[86] C. Yu, J. Liu, S. Nemati, G. Yin, Reinforcement learning in healthcare: A survey, ACM Computing Surveys 55 (2023) 1–36.

[87] F. R. Ward, I. Habli, An assurance case pattern for the interpretability of machine learning in safety-critical systems, in: Proceedings of the 2020 International Conference on Computer Safety, Reliability, and Security, SAFECOMP '20, 2020, pp. 395–407.

[88] DOT, Manual on Uniform Traffic Signal Control Devices, revision 2 ed., US Department of Transportation, 2012.

[89] J. Bonneson, M. Pratt, K. Zimmerman, Development of a Traffic Signal Operations Handbook, Technical Report FHWA/TX-09/0-5629-1, Texas A&M Transportation Institute, 2009.

[90] M. K. Lee, D. Kusbit, A. Kahng, J. T. Kim, X. Yuan, A. Chan, D. See, R. Noothigattu, S. Lee, A. Psomas, A. D. Procaccia, WeBuildAI: Participatory framework for algorithmic governance, Proceedings of the ACM on Human-Computer Interaction 3 (2019) 1–35.

[91] D. Lord, J. A. Bonneson, Role and application of accident modification factors within highway design process, Transportation Research Record 1961 (2006) 65–73.

[92] L. Wu, D. Lord, Y. Zou, Validation of crash modification factors derived from cross-sectional studies with regression models, Transportation Research Record 2514 (2015)

88–96.

[93] J. Ma, M. D. Fontaine, F. Zhou, J. Hu, Estimation of crash modification factors for an adaptive traffic-signal control system, Journal of Transportation Engineering 142 (2016) 04016061.

[94] X. Sun, Y. Li, D. Magri, H. H. Shirazi, Application of *Highway Safety Manual* draft chapter: Louisiana experience, Transportation Research Record 1950 (2006) 55–64.

[95] C. Sun, H. Brown, P. Edara, B. Carlos, K. Nam, Calibration of the *Highway Safety Manual* for Missouri, Technical Report 25-1121-0003-177, Mid-America Transportation Center, 2013.

[96] F. Xie, K. Gladhill, K. K. Dixon, C. M. Monsere, Calibration of *Highway Safety Manual* predictive models for Oregon state highways, Transportation Research Record 2241 (2011) 19–28.

[97] J. Ault, J. P. Hanna, G. Sharon, Learning an interpretable traffic signal control policy, in: Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '20, 2020, pp. 88–96.

[98] V. Jayawardana, A. Landler, C. Wu, Mixed autonomous supervision in traffic signal control, in: Proceedings of the 2021 International Conference on Intelligent Transportation Systems, ITSC '21, 2021, pp. 1767–1773.

[99] S. G. Rizzo, G. Vantini, S. Chawla, Reinforcement learning with explainability for traffic signal control, in: Proceedings of the 2019 International Conference on Intelligent Transportation Systems, ITSC '19, 2019, pp. 3567–3572.

[100] A. Müller, V. Rangras, G. Schnittker, M. Waldmann, M. Friesen, T. Ferfers, L. Schreckenberg, F. Hufen, J. Jasperneite, M. Wiering, Towards real-world deployment of reinforcement learning for traffic signal control, in: Proceedings of the 20th IEEE International Conference on Machine Learning and Applications, ICMLA '21, 2021, pp. 507–514.

[101] Y. Liu, L. Liu, W.-P. Chen, Intelligent traffic light control using distributed multi-agent Q learning, in: Proceedings of the 2017 International Conference on Intelligent Transportation Systems, ITSC '17, 2017, pp. 1–8.

[102] M. Essa, T. Sayed, Self-learning adaptive traffic signal control for real-time safety optimization, Accident Analysis & Prevention 146 (2020) 105713.

[103] M. Essa, T. Sayed, Traffic conflict models to evaluate the safety of signalized intersections at the cycle level, Transportation Research Part C: Emerging Technologies 89 (2018) 289–302.

[104] Y. Gong, M. Abdel-Aty, J. Yuan, Q. Cai, Multi-objective reinforcement learning approach for improving safety at intersections with adaptive traffic signal control, Accident Analysis & Prevention 144 (2020) 105655.

[105] L. Liao, J. Liu, X. Wu, F. Zou, J. Pan, Q. Sun, S. E. Li, M. Zhang, Time difference penalized traffic signal timing by LSTM Q-network to balance safety and capacity at intersections, IEEE Access 8 (2020) 80086–80096.

[106] B. Yu, J. Guo, Q. Zhao, J. Li, W. Rao, Smarter and safer traffic signal controlling via deep reinforcement learning, in: Proceedings of the 29th ACM International Conference on Information and Knowledge Management, CIKM '20, 2020, pp. 3345–3348.

[107] S. Bohez, A. Abdolmaleki, M. Neunert, J. Buchli, N. Heess, R. Hadsell, Value constrained model-free continuous control, arXiv preprint (2019). `arXiv:1902.04623`.

[108] D. Ding, K. Zhang, T. Basar, M. Jovanovic, Natural policy gradient primal-dual method for constrained Markov decision processes, in: Proceedings of the 34th International Conference on Neural Information Processing Systems, NeurIPS '20, 2020, pp. 8378–8390.

[109] Z. Liu, Z. Cen, V. Isenbaev, W. Liu, Z. S. Wu, B. Li, D. Zhao, Constrained variational policy optimization for safe reinforcement learning, in: Proceedings of the 39th International Conference on Machine Learning, ICML '22, 2022, pp. 1–9.

[110] D. Branston, H. van Zuylen, Comparison of queue-length models at signalized intersections, Transportation Research 12 (1978) 47–53.

[111] F. Viloria, K. Courage, D. Avery, Comparison of queue-length models at signalized intersections, Transportation Research Record 1710 (2000) 222–230.

[112] H. Zhang, S. Feng, C. Liu, Y. Ding, Y. Zhu, Z. Zhou, W. Zhang, Y. Yu, H. Jin, Z. Li, CityFlow: A multi-agent reinforcement learning environment for large scale city traffic scenario, in: Proceedings of the 2019 World Wide Web Conference, WWW '19, 2019, pp. 3620–3624.

[113] M. Guo, P. Wang, C.-Y. Chan, S. Askary, A reinforcement learning approach for intelligent traffic signal control at urban intersections, in: Proceedings of the 2019 International Conference on Intelligent Transportation Systems, ITSC '19, 2019, pp. 4242–4247.

[114] A. Sharma, E. Smaglik, S. Kothuri, O. Smith, P. Koonce, T. Huang, Leading pedestrian intervals: Treating the decision to implement as a marginal benefit–cost problem, Transportation Research Record 2620 (2017) 96–104.

[115] K. Tang, M. Boltze, Z. Tian, H. Nakamura, Initial comparative analysis of international practice in road traffic signal control, in: Global Practices on Road Traffic Signal Control, Elsevier, 2019, pp. 285–310.

[116] S. Smith, Surtrac for the People: Upgrading the Surtrac Pittsburgh Deployment to incorporate Pedestrian Friendly Extensions and Remote Monitoring Advances, Technical Report 01730614, Mobility21, 2020.

[117] S. Kothuri, A. Kading, E. Smaglik, C. Sobie, Improving Walkability Through Control Strategies at Signalized Intersections, Technical Report NITC-RR-782, National Institute for Transportation and Communities, 2017.

[118] C. Slavin, W. Feng, M. Figliozzi, P. Koonce, Statistical study of the impact of adaptive traffic signal control on traffic and transit performance, Transportation Research Record 2356 (2016) 117–126.

[119] J. Peters, P. O'Brien, J. Pachman, Memorandum: Farmington Road Adaptive Traffic Control Benefits Analysis, Technical Report, DKS Associates, 2011.

[120] A. Mahendran, S. Smith, M. Hebert, X.-F. Xie, Bus Detection for Adaptive Traffic Signal Control, Technical Report, Carnegie Mellon University, 2014.

[121] S. Smith, I. Isukapati, E. Bronstein, C. Igoe, Integrating transit signal priority with adaptive signal control in a connected vehicle environment: Phase 1 Final Report, Technical Report 01675986, Mobility21, 2018.

[122] B. Yin, M. Menendez, A reinforcement learning method for traffic signal control at an isolated intersection with pedestrian flows, in: Proceedings of the 19th COTA International Conference of Transportation Professionals, CICTP '19, 2019, pp. 3123–3135.

[123] Y. Zhang, J. Fricker, Investigating smart traffic signal controllers at signalized crosswalks: A reinforcement learning approach, in: Proceedings of the 7th International Conference on Models and Technologies for Intelligent Transportation Systems, MT-ITS '21, 2021,

pp. 1–6.

[124] P. Chanloha, J. Chinrungrueng, W. Usaha, C. Aswakul, Cell transmission model-based multiagent Q-learning for network-scale signal control with transit priority, The Computer Journal 57 (2014) 451–468.

[125] S. M. A. Shabestray, B. Abdulhai, Multimodal iNtelligent Deep (MiND) traffic signal controller, in: Proceedings of the 2019 International Conference on Intelligent Transportation Systems, ITSC '19, 2019, pp. 4532–4539.

[126] L. Zhang, S. Jiang, Z. Wang, Schedule-driven signal priority control for modern trams using reinforcement learning, in: Proceedings of the 17th COTA International Conference of Transportation Professionals, CICTP '17, 2017, pp. 2122–2132.

[127] G. Guo, Y. Wang, An integrated MPC and deep reinforcement learning approach to trams-priority active signal control, Control Engineering Practice 110 (2021) 104758.

[128] N. Kumar, S. S. Rahman, N. Dhakad, An integrated MPC and deep reinforcement learning approach to trams-priority active signal control, IEEE Transactions on Intelligent Transportation Systems 22 (2021) 4919–4928.

[129] H. Su, Y. D. Zhong, B. Dey, A. Chakraborty, EMVLight: A decentralized reinforcement learning framework for efficient passage of emergency vehicles, in: Proceedings of the 36th AAAI Conference on Artificial Intelligence, AAAI '22, 2022, pp. 1–11.

[130] H. Su, K. Shi, J. Chow, L. Jin, Dynamic queue-jump lane for emergency vehicles under partially connected settings: A multi-agent deep reinforcement learning approach, arXiv preprint (2021). arXiv:2003.01025.

[131] N. Nahar, S. Zhou, G. Lewis, C. Kästner, Collaboration challenges in building ML-enabled systems: Communication, documentation, engineering, and process, in: Proceedings of the 44th International Conference on Software Engineering, ICSE '22, 2022, pp. 1–22.

[132] J. Pineau, P. Vincent-Lamarre, K. Sinha, V. Larivière, A. Beygelzimer, F. d'Alché-Buc, E. Fox, H. Larochelle, Improving reproducibility in machine learning research (a report from the NeurIPS 2019 Reproducibility Program), Journal of Machine Learning Research 22 (2021) 1–20.