# From Prediction to Safe Deployment: A Critical Review of AI-Based Traffic Management, Operational Resilience, and Governance in Mixed-Traffic Urban Networks

Sarthak Bibave, Anand Jadhav, Abhishek Agale, and Aditya Kalyankar
*Department of Master of Computer Applications, Dr. D. Y. Patil Institute of Master of Computer Applications and Management, Akurdi, Pune, India*
Guide: Prof. Rahul Chaudhari

***Abstract*—**Artificial Intelligence (AI) has demonstrated substantial potential for traffic signal optimisation, congestion prediction, and incident detection in simulation environments. However, a persistent gap separates algorithmic performance from dependable real-world operation, particularly in mixed-traffic urban networks where motorcycles, three-wheelers, pedestrians, and non-motorised vehicles share limited road space. This paper does not propose a new AI algorithm or claim experimental validation. Instead, it presents a critical review of the deployment-readiness gap in AI-based traffic management and proposes a six-dimensional framework for assessing whether a system should be permitted to operate on public roads. The framework integrates: (1) AI monitoring, prediction, and signal optimisation; (2) mixed-traffic and pedestrian safety; (3) sensor and communication reliability; (4) cybersecurity and privacy; (5) human supervision and safe fallback; and (6) measurable evaluation criteria. Each dimension is anchored in existing literature and governance standards, notably the NIST AI Risk Management Framework. The review finds that existing surveys acknowledge the simulation-to-deployment gap but do not operationalise deployment readiness as an assessment condition. By positioning the framework as a diagnostic tool complementary to Technology Readiness Level assessments, this paper provides city authorities, traffic engineers, and system developers with a structured basis for judging when an AI traffic management system is operationally safe, rather than merely algorithmically promising.

***Index Terms*—**AI traffic management, deployment readiness, mixed traffic, pedestrian safety, NIST AI RMF, operational resilience, intelligent transportation systems.

## I. INTRODUCTION

Artificial Intelligence has become a dominant research focus in traffic management. Machine learning models forecast short-term traffic flow; computer vision estimates queue lengths; and reinforcement learning optimises signal phase durations. The volume of published work is substantial and accelerating. Recent surveys have synthesised research on large language models for traffic signal control, reinforcement learning for multi-intersection control, and hybrid AI approaches.

These surveys are valuable because they map the algorithmic landscape, compare simulation performance, and identify technical challenges. However, deployment conditions are often treated as a residual category—acknowledged in future-work discussions but not operationalised as admission criteria. A survey of reinforcement-learning-based multi-intersection control identifies a gap between algorithmic development and deployable real-time control systems, including limited treatment of safety and reliability, calibration, heterogeneous traffic modelling, and coordination mechanisms.

This paper addresses that omission. Its contribution is not another taxonomy of AI techniques. It is a deployment-readiness framework for mixed-traffic urban networks: a structured set of conditions that should be satisfied before an AI traffic management system transitions from simulation or pilot operation to public-road deployment. The framework is organised around six dimensions, each derived from gaps identified in the reviewed literature and anchored in recognised governance standards.

The scope is deliberately limited. This paper does not claim that AI traffic management is inherently unsafe or ready for widespread adoption. It argues that asking whether a system is accurate is insufficient. The necessary question is: under what conditions, and with what verification, should a system be permitted to control real traffic?

## II. THE DEPLOYMENT GAP IN EXISTING AI TRAFFIC MANAGEMENT RESEARCH

### *A. Coverage of Existing Surveys*

The current generation of AI traffic management surveys is methodologically sophisticated. Recent work proposes taxonomies for large-language-model-augmented reinforcement learning, LLM-centric reasoning agents, and knowledge-enhanced hybrid systems. Other reviews compare hybrid AI methods and catalogue research on multi-intersection reinforcement learning. Collectively, these studies establish that AI methods can outperform fixed-time control under simulated conditions, identify promising architectural directions, and create a shared vocabulary for comparing approaches.

### *B. What Existing Surveys Leave Open*

Three exclusions recur. First, deployment conditions are described but not assessed. Practical guidelines can inform a development team, but they do not tell a municipal authority whether a particular system is safe to activate. Second, mixed-traffic and pedestrian dimensions remain under-specified. Reported benefits depend on intersection geometry, traffic patterns, and system sophistication, while pedestrian waiting time and conflict measures are not consistently reported alongside vehicle-efficiency measures. Third, human oversight and failure modes are often treated as design preferences rather than structural requirements.

Deployment readiness reverses the sequence in which safety is sometimes considered. Rather than developing an algorithm first and adding safety considerations later, the proposed approach treats safety, reliability, oversight, and governance as conditions that must be verified before operational deployment.

### C. The Missing Diagnostic

The Technology-Policy Readiness Matrix (TPRM) provides a useful parallel by coupling Technology Readiness Levels (TRLs) with Policy Readiness Levels (PRLs). It highlights situations in which technical maturity advances faster than institutional and policy readiness. However, policy readiness does not establish whether a specific controller fails safely under sensor loss or communication interruption. The framework proposed here addresses this operational layer between technical performance in simulation and policy permission to deploy.

## III. SIX-DIMENSIONAL DEPLOYMENT-READINESS FRAMEWORK

The framework consists of six dimensions. For each dimension, the central question is not whether the issue has been considered, but whether it has been verified to an acceptable standard. The dimensions are concurrent conditions; all should be satisfied before full operational deployment.

### A. AI Monitoring, Prediction, and Signal Optimisation

Readiness question: Has the system demonstrated performance under conditions that resemble the target deployment environment, rather than only on a standard simulation benchmark?

The reviewed evidence supports conditional optimism. Hybrid AI approaches and adaptive signal-control systems report reductions in vehicle waiting time, fuel consumption, and emissions under specified simulation conditions. However, these figures depend on model assumptions and evaluation settings. Evaluation practices can be fragmented and may rely on loosely defined measures without sufficient emphasis on measurable intersection-level indicators.

The readiness criterion is therefore a validity condition rather than a universal performance threshold: the system must be evaluated on a network topology, traffic-demand profile, and sensor configuration that correspond to the intended deployment site. In the NIST AI Risk Management Framework (AI RMF), MEASURE 2.5 addresses demonstration of validity and reliability. For traffic management, validity requires evidence that the simulation or pilot environment is representative of the intended operating context.

### B. Mixed-Traffic and Pedestrian Safety

Readiness question: Has the system been evaluated against safety indicators specific to pedestrians, cyclists, and non-motorised road users in the local mixed-traffic context?

This dimension is particularly consequential for Indian and comparable urban contexts. Static theoretical values from design manuals may be inadequate under highly heterogeneous traffic streams. Although adaptive signal control may reduce pedestrian–vehicle conflicts, reported effectiveness varies with intersection geometry and system sophistication. Few studies consistently report pedestrian waiting time or conflict metrics alongside vehicle-efficiency metrics.

Pedestrian safety must therefore be measured, not assumed. A system that reduces vehicle delay while increasing pedestrian risk has not achieved readiness; it has displaced risk onto a more vulnerable road-user group. Evaluation should include pedestrian waiting time, conflict frequency, and compliance with minimum safe crossing intervals, disaggregated by intersection and time of day.

### C. Sensor and Communication Reliability

Readiness question: Has the system demonstrated safe behaviour under sensor failure, communication interruption, and data degradation?

AI traffic systems depend on continuous data. Camera lenses can become dirty, loop detectors can fail, and communication links can experience latency. Sensor failure should be treated as a routine operational condition rather than an exceptional edge case. NIST AI RMF MEASURE 2.6 and MEASURE 2.7 address evaluation of misuse, security, and resilience.

For traffic management, resilience means that the system transitions to a conservative, known-safe state—such as fixed-time control with appropriate clearance intervals—instead of continuing to operate on stale or corrupted data. Readiness requires documented evidence that fallback behaviour has been stress-tested under defined failure scenarios.

### D. Cybersecurity and Privacy

Readiness question: Has the system undergone threat modelling, and are data-minimisation and retention controls verified?

A review of security and privacy in intelligent transportation systems documents an expanding attack surface that includes wireless exploitation, sensor vulnerabilities, and software flaws. Traffic infrastructure is safety-critical: a compromised signal controller can create a physical hazard. NIST AI RMF MEASURE 2.7 addresses security and resilience evaluation, while MEASURE 2.10 addresses examination and documentation of privacy risk.

Privacy risks include the possibility that camera footage and location data reveal individual movement patterns. Readiness requires auditable controls: defined retention periods, verified de-identification or blurring procedures, role-based access controls tested against unauthorised access, and an incident-response plan coordinated with traffic operations.

### E. Human Supervision and Safe Fallback

Readiness question: Can a qualified operator understand, override, and safely terminate the AI system's control within a defined time?

Human oversight is frequently identified as essential, but it must be an operational capability rather than merely a design feature. A system can be installed and activated before the operating agency has the capacity to supervise it. The operator interface should display the current system state, confidence levels, and active constraints. Override procedures must be tested under realistic conditions, and transition to manual or fixed-time control must be achievable within a period consistent with traffic-safety requirements.

### *F. Measurable Evaluation Criteria*

Readiness question: Does the evaluation plan include negative conditions—criteria under which the system would be judged unsuccessful even if vehicle metrics improve?

Traffic management is a public service. A system that reduces average vehicle delay while worsening pedestrian safety, creating downstream queue spillback, or degrading service in less-monitored neighbourhoods has not succeeded. Evaluation should include distributional and safety conditions alongside efficiency metrics.

Disaggregated reporting should include travel time and delay by road-user type; queue length by approach and intersection; pedestrian waiting time and conflict counts; and performance under peak, off-peak, and incident conditions. The criterion is not a single aggregate improvement figure, but a pattern of results that does not conceal harm to particular groups or operating conditions.

## IV. APPLICATION OF THE FRAMEWORK: A HYPOTHETICAL ASSESSMENT

To illustrate the framework as a diagnostic rather than a checklist, consider a hypothetical deployment of an AI signal controller at a mixed-traffic intersection in Pune. The system uses reinforcement learning and has been tested in a standard microsimulation environment.

Dimension 1—AI performance: The simulation uses a calibrated network with local vehicle counts. The validity condition is partially satisfied. However, the simulation assumes lane-based discipline, while the deployment site exhibits weak lane discipline. Readiness is conditional.

Dimension 2—pedestrian safety: The simulation reports vehicle delay but not pedestrian waiting time or conflict frequency. No local pedestrian-volume data are incorporated. Readiness is not satisfied.

Dimension 3—reliability: The design specifies fallback to fixed-time control, but no stress testing under sensor failure has been conducted. Readiness is not satisfied.

Dimension 4—security and privacy: Camera data are processed locally, but no threat model is documented and retention periods are undefined. Readiness is not satisfied.

Dimension 5—human supervision: The operator dashboard displays signal state but not confidence levels or active constraints. An override procedure exists but has not been tested. Readiness is conditional.

Dimension 6—evaluation: The evaluation plan includes vehicle delay and throughput but excludes pedestrian metrics and downstream queue spillback. Readiness is not satisfied.

The overall assessment is that the system is not deployment-ready. It may be suitable for a supervised pilot with augmented pedestrian monitoring and formal fallback testing, but activation as the primary controller would be premature. This conclusion concerns operational conditions, not the intrinsic quality of the algorithm.

## V. DISCUSSION

### *A. Contribution Relative to Existing Surveys*

The framework does not compete with algorithmic surveys on coverage or taxonomic sophistication. Its contribution is structural: it converts deployment considerations from general challenges into assessment conditions. Existing surveys identify the simulation-to-deployment gap and catalogue technical and governance challenges. This paper provides a structure for determining whether a specific system at a specific site has addressed those challenges sufficiently to proceed.

### *B. Complementarity with TPRM*

The Technology-Policy Readiness Matrix addresses whether a jurisdiction's regulatory and institutional capacity matches its technical procurement. The proposed deployment-readiness framework addresses whether a specific system at a specific site meets operational safety and governance conditions. The two diagnostics are complementary, not substitutable.

### *C. Limitations and Future Work*

The framework is conceptual. It has not been applied to a live deployment or validated through expert elicitation. Its readiness criteria are qualitative: they define what must be demonstrated, not universal thresholds at which the demonstration is sufficient. This is deliberate because thresholds are context-dependent and should be set by the operating authority in consultation with traffic engineers. Future work should develop operational scoring rubrics and test inter-rater reliability across assessors and sites.

## VI. CONCLUSION

AI traffic management research does not need only another survey cataloguing algorithms or simulation results. It also needs a structured answer to the question that existing surveys raise but do not resolve: under what conditions should an AI system be permitted to control real traffic, and how can those conditions be verified?

This paper has proposed a six-dimensional deployment-readiness framework addressing AI performance, mixed-traffic and pedestrian safety, sensor reliability, cybersecurity and privacy, human supervision, and measurable evaluation. The framework is anchored in gaps identified in the reviewed literature and in NIST AI RMF assessment requirements. It is positioned as an operational diagnostic complementary to policy-level readiness assessments such as TPRM.

The framework's value lies in preventing deployment decisions based on simulation performance alone, with safety, security, and oversight treated as post-deployment concerns. In mixed-traffic urban networks, where pedestrians, cyclists, and informal transport share constrained road space, the consequences of such decisions can fall disproportionately on vulnerable road users. The question is not

simply whether AI can improve traffic management, but whether the conditions for doing so safely have been established. The proposed framework is a step toward answering that question.

**TABLE I. SUMMARY OF THE SIX-DIMENSIONAL DEPLOYMENT-READINESS FRAMEWORK**

| Dimension | Readiness question | Verification requirement | NIST AI RMF anchor |
|---|---|---|---|
| AI performance | Demonstrated under deployment-like conditions? | Calibrated simulation or pilot using local topology and demand | MEASURE 2.5 |
| Mixed-traffic safety | Are vulnerable-road-user metrics measured? | Disaggregated conflict, waiting-time, and crossing-compliance data | MEASURE 2.5 |
| Sensor reliability | Is safe fallback verified under failure? | Stress tests for sensor loss, communication delay, and data corruption | MEASURE 2.6, 2.7 |
| Cybersecurity and privacy | Is threat modelling complete and are retention controls verified? | Documented threat assessment, audited access controls, defined retention | MEASURE 2.7, 2.10 |
| Human supervision | Are override and termination tested? | Operational test of operator interface and fallback procedure | MEASURE 2.7 |
| Evaluation criteria | Are negative conditions included? | Disaggregated reporting with safety and distributional thresholds | MEASURE 2.5 |

## REFERENCES


[1] Y. Wu, C. Luo, J. Wang, Y. Wang, Y. Jiang, and Z. Yao, "A survey of large language models for urban traffic signal control," Expert Systems with Applications, 2026, doi: 10.1016/j.eswa.2026.129093. [Verify publication status, volume, and article number.]

[2] U. Madububambachu, R. Fatima, A. Sherif, and K. Khalil, "Security and privacy solutions in intelligent transportation systems: A survey," Internet of Things, vol. 34, Art. no. 101812, 2025, doi: 10.1016/j.iot.2025.101812.

[3] M. R. Albous, A. Lacheret, and M. Stephens, "The governance of AI-enabled transport: Bridging Kuwait's tech-policy gap vis-à-vis the UAE and Singapore," Transport Policy, 2026. [Add volume, pages/article number, DOI, and verify bibliographic record.]

[4] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1, U.S. Department of Commerce, Jan. 2023.

[5] "RL-based multi-intersection traffic signal control: A structured survey of trends and open deployment challenges," IEEE Xplore, 2026. [Author names, conference/journal title, pages, DOI, and publication record were not supplied; complete and verify before submission.]

[6] S. Omefe, "Assessment of the effectiveness of adaptive traffic signal control systems in reducing pedestrian-vehicle conflicts at high-risk crossings," GSC Advanced Research and Reviews, vol. 24, no. 1, pp. 140–153, 2025. [Verify title, author details, and publication record.]

[7] A. K. Garasiya, "A systematic review of traffic signal design methodologies at urban intersections under heterogeneous traffic conditions," InSight Bulletin, vol. IV, no. 2, pp. 1–7, 2026. [Verify journal/publisher and publication record.]

[8] "Comparative review of hybrid AI approaches for traffic signal systems," IEEE Xplore, 2026. [Author names, conference title, pages, DOI, and publication record were not supplied; complete and verify before submission.]