2025
https://fupress.org/journal

Human-in-the-Loop Optimization of Urban Traffic Control through Explainable AI Interfaces

Angela Mensah

Lecturer, Dept. of Computer Science, University of Cape Town, Cape Town, South Africa
* **Corresponding Author:** amensah@uct.ac.za

ARTICLE INFO

Received: 10 Feb 2025
Accepted: 27 Apr 2025

ABSTRACT

Urban traffic congestion presents significant challenges to modern cities, necessitating advanced management strategies. Artificial Intelligence (AI) has demonstrated potential in optimizing traffic signal control; however, conventional AI models often operate as "black boxes," limiting operator trust and system accountability. This study proposes a Human-in-the-Loop (HITL) optimization framework enhanced with Explainable AI (XAI) interfaces to improve transparency, decision quality, and operator trust in urban traffic control. A simulation-based experiment was conducted using deep reinforcement learning for signal optimization, integrated with SHAP-based and attention-visualization explanations for human operators. Performance metrics, including average vehicle delay, queue length, decision accuracy, operator trust, and interpretability scores, were collected and analyzed. Results indicate that the HITL-XAI system significantly outperformed the AI-only baseline, reducing vehicle delay by 30% and queue lengths by 36%, while increasing operator trust and interpretability scores by more than 50%. Findings highlight the critical role of human-AI collaboration in complex, safety-critical urban environments. By combining optimization efficiency with real-time interpretability, the proposed framework addresses key limitations of conventional AI systems and offers a scalable, transparent solution for modern Intelligent Transportation Systems (ITS). The study further provides insights into future deployments of HITL-XAI models in real-world urban traffic networks.

Keywords: Human-in-the-Loop (HITL), Explainable AI (XAI), Urban Traffic Control, Deep Reinforcement Learning, Intelligent Transportation Systems (ITS).

INTRODUCTION

In recent years, Intelligent Transportation Systems (ITS) have become a cornerstone of modern urban mobility management. The primary goal of ITS is to enhance traffic flow efficiency, minimize delays, and optimize the overall performance of transportation infrastructure. Within this context, Artificial Intelligence (AI) and Machine Learning (ML) techniques have gained significant traction, offering data-driven solutions for traffic prediction, congestion management, and adaptive signal control.

AI-based systems utilize data from cameras, sensors, and video feeds to make real-time decisions in traffic regulation and monitoring. For example, AI models can detect vehicles, pedestrians, and anomalies, thereby facilitating automated control of traffic signals and road networks. Adaptive signal control systems, which dynamically adjust signal timings based on real-time conditions, have demonstrated measurable improvements. Studies have shown that reinforcement learning (RL)-based signal control can reduce vehicle delay times by up to 47%, thereby enhancing urban mobility and reducing fuel consumption.

However, despite these advancements, certain challenges persist. Automated systems often face difficulties in responding to unexpected changes such as accidents, environmental fluctuations, or unstructured road conditions. Moreover, in safety-critical environments like urban traffic control, complete automation may not always be desirable. A lack of human oversight can lead to unanticipated outcomes, especially in scenarios that demand ethical reasoning or contextual understanding. Hence, there is an emerging need to balance algorithmic

intelligence with human judgment and situational awareness.

Problem Statement

While AI-driven traffic control systems have achieved remarkable success in efficiency and responsiveness, they frequently suffer from a lack of transparency and interpretability. Many models operate as "black boxes," providing output decisions without any clear explanation of how those decisions were made. For traffic operators, this opacity creates a major challenge: it becomes difficult to understand, trust, or appropriately override AI recommendations.

In real-world applications, an operator may need to question or adjust an AI-driven signal control decision. However, without a clear rationale or interpretive feedback, such actions can feel arbitrary or risky. The absence of explainability undermines operator confidence and increases the cognitive burden on human decision-makers. Consequently, there is a pressing need for a collaborative human–AI framework, where humans remain in the loop and AI systems provide interpretable, transparent justifications for their actions. Such collaboration can enhance trust, accountability, and overall system safety.

Aim and Objectives of the Study

Aim

To develop and analyze a human-in-the-loop optimization framework that leverages explainable AI (XAI) interfaces for improving transparency, trust, and decision-making effectiveness in urban traffic control systems.

Objectives

To review existing AI-based traffic control systems and identify their interpretability gaps.

To design a simulation-based Human-in-the-Loop (HITL) framework that integrates XAI tools.

To evaluate the system's performance under various traffic and human feedback conditions.

To assess improvements in transparency, decision quality, and operator trust.

Significance of the Study

This study holds both theoretical and practical significance. Theoretically, it contributes to the growing body of research on human–AI collaboration in critical infrastructure domains. It emphasizes that AI should not merely automate processes but should also complement human expertise by providing interpretable insights and feedback. Practically, the research proposes an optimization model that bridges the gap between human decision-making and AI automation, leading to safer, more transparent, and accountable traffic management.

Furthermore, the proposed framework aligns with the vision of smart cities, where technology serves human needs through collaboration rather than replacement. By ensuring that traffic control systems remain explainable and responsive to human feedback, this study promotes responsible AI deployment that fosters public trust and enhances operational reliability.

Ultimately, this research contributes toward creating urban traffic systems that are not only efficient but also transparent, understandable, and ethically aligned with human oversight—an essential step toward achieving sustainable and trustworthy smart city infrastructures.

LITERATURE REVIEW

AI and Optimization in Urban Traffic Control

The integration of Artificial Intelligence (AI) into urban traffic management has significantly advanced optimization capabilities in signal control and congestion reduction. Modern Intelligent Transportation Systems (ITS) increasingly utilize Reinforcement Learning (RL) and Deep Q-Networks (DQN) to achieve adaptive control mechanisms that respond to dynamic traffic conditions. These models learn from environmental interactions, adjusting traffic signal phases to minimize waiting times and maximize throughput.

According to Saadi, Aouada, and Ottersten (2025), RL-based methods outperform traditional fixed-time or actuated systems by enabling real-time learning from fluctuating traffic data. Similarly, Cai, Zhang, and Li (2024) demonstrated that deep reinforcement learning algorithms can enhance signal coordination efficiency by up to 30%, reducing both congestion and emissions in dense urban areas.

Automation not only enhances traffic flow efficiency but also reduces operator workload. Wu, Prabowo, and Zhang (2023) emphasized that automated systems can make data-driven decisions at millisecond speeds,

responding rapidly to congestion events or accidents. However, despite these benefits, the “black-box” nature of AI models often limits interpretability. Operators are frequently unable to understand why specific control actions were chosen, which raises concerns regarding safety, accountability, and reliability in critical traffic systems.

Explainable AI (XAI) in Intelligent Transportation Systems

As AI becomes integral to ITS, the demand for Explainable Artificial Intelligence (XAI) has grown, emphasizing the need for transparency and interpretability in decision-making. XAI aims to make the decision logic of AI systems comprehensible to human users without compromising performance. According to Yang, Liu, and Zhao (2023), XAI techniques such as SHapley Additive exPlanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME), and attention visualization have been effective in providing post-hoc interpretability across several AI domains.

In transportation, explainable systems allow operators to understand how signal control decisions are derived from sensor inputs and learned policies. Degas, Vidosavljevic, and Petrovic (2022) noted that explainability is crucial for increasing human trust in automated systems, particularly when high-stakes decisions—such as traffic rerouting or emergency priority allocation—are involved. However, incorporating XAI into real-time environments poses significant computational challenges. Generating meaningful explanations must occur within milliseconds to maintain synchronization with traffic signal operations, which remains an open research issue (Yang et al., 2023).

Moreover, explainability introduces trade-offs between performance and interpretability. For instance, more transparent models (like decision trees) are less efficient for complex urban environments, while deep neural networks offer superior optimization at the cost of transparency. This balance highlights the need for hybrid approaches that retain both interpretability and adaptability.

Human-in-the-Loop (HITL) Approaches in Decision Systems

The Human-in-the-Loop (HITL) paradigm emphasizes human oversight in AI decision-making, ensuring systems remain accountable, interpretable, and ethically aligned. HITL models integrate human judgment directly into the learning or control loop, allowing for continuous feedback, correction, and supervision. Wu et al. (2021) outlined that HITL approaches enhance safety and reliability in domains where automation errors could lead to catastrophic consequences.

Applications of HITL are prominent in sectors like aviation, healthcare, and transportation, where decisions must be transparent and adaptable to context. Tsiakas and Groll (2022) proposed that human-in-the-loop frameworks contribute to trustworthy sociotechnical systems, promoting cooperation between humans and machines rather than full automation. In transportation, operators can review and adjust AI-driven control suggestions, creating a balance between human expertise and algorithmic efficiency.

The cognitive benefits of HITL systems are equally significant. They reduce operator stress by offering interpretable AI insights while maintaining human situational awareness. However, challenges remain—particularly the human cognitive overload caused by frequent AI feedback or poorly designed interfaces. Zanzotto (2019) highlighted that achieving optimal human-AI interaction requires interfaces that simplify complex model outputs into intuitive visual explanations.

Gaps in Existing Literature

While existing research demonstrates progress in both optimization and explainability, several gaps remain. Firstly, most studies focus on optimization performance without addressing interpretability or human feedback integration (Saadi et al., 2025; Cai et al., 2024). Few frameworks have effectively combined reinforcement learning with explainable mechanisms to support transparent human decision-making in real-time control environments.

Secondly, empirical studies validating operator trust improvement are limited. Although XAI frameworks enhance interpretability, there is insufficient experimental evidence linking transparency to actual behavioral outcomes among traffic operators (Yang et al., 2023). Finally, the challenge of maintaining real-time interpretability under high-load traffic scenarios is underexplored. Computational delays in generating explanations may reduce system responsiveness, potentially undermining optimization performance.

METHODOLOGY

Research Design

This research adopts an experimental simulation-based design to evaluate the impact of integrating

Explainable AI (XAI) within a Human-in-the-Loop (HITL) optimization framework for urban traffic control. The primary focus is to examine how human feedback and AI transparency collectively improve decision accuracy, trust, and system performance under dynamic traffic conditions.

The experiment is designed around a controlled simulation environment representing an urban intersection equipped with multiple signal phases and sensor inputs. The simulation includes both AI-only (baseline) and HITL-XAI-enhanced configurations to enable comparative performance evaluation. This design allows systematic testing of how real-time human feedback, supported by XAI explanations, influences system optimization outcomes.

Simulation Environment and Data Collection

The simulation uses synthetic traffic flow data modeled on real-world conditions—such as vehicle arrival rates, signal timings, and pedestrian crossing frequencies—based on datasets from typical urban traffic management systems.

Key data sources include vehicle count sensors, adaptive signal controllers, and environmental parameters like weather and time-of-day variations.

The simulation platform is implemented using SUMO (Simulation of Urban Mobility) integrated with a reinforcement learning controller developed in Python (TensorFlow). The reinforcement learning model is trained to minimize vehicle delay and queue length, while the human-in-the-loop component provides corrective feedback during decision cycles.

Each simulation run consists of:

Baseline (AI-only) phase: System operates purely on reinforcement learning (DQN-based) control without human feedback.

HITL-XAI phase: System presents interpretable visual explanations (e.g., SHAP plots) of its control decisions to a human operator, who may provide approval, rejection, or modified suggestions.

Traffic conditions such as low, medium, and high congestion levels are tested to evaluate adaptability and robustness.

Experimental Procedure

The experiment follows five key stages:

Initialization:

Traffic data and signal parameters are loaded into the simulation model. The reinforcement learning agent initializes with random policy weights.

Training Phase:

The AI model undergoes 500 episodes of training, optimizing its policy through the reward function:

$$R_t = -(\alpha \times D_t + \beta \times Q_t + \gamma \times E_t) \tag{1}$$

Where:

R_t is the reward at time t

D_t is the average vehicle delay at time t

Q_t is the queue length at time t

E_t is the emission level at time t,

α, β, and γ are the relative importance weights for each of these metrics. These coefficients determine the trade-off between optimizing the delay, queue length, and emissions in the system.

Explainability Integration:

During the HITL phase, the XAI module generates interpretive outputs using SHAP (Shapley Additive Explanations) and attention heatmaps that display how sensor data influence AI decisions. These visualizations are shared via a dashboard interface for operator review.

Human Feedback Loop:

The operator reviews the AI's proposed action and explanation. Based on interpretability, the human may accept or override the decision. Each override is recorded as feedback for the model's retraining.

Evaluation and Comparison:

The two setups—AI-only and HITL-XAI—are compared across performance metrics such as:

Average delay per vehicle (seconds)

Queue length (vehicles)

Decision accuracy (%)

Operator trust rating (Likert scale 1–5)

Interpretability score (subjective human rating)

Variables and Parameters

Table 1. Variables used in the HITL-XAI optimization experiment.

Variable Type	Variable Name	Description / Measurement
Independent	AI configuration	Two setups: AI-only vs. HITL-XAI
Dependent	Traffic flow efficiency	Average vehicle delay and queue length
Dependent	Operator trust	Mean Likert score from operator feedback
Control	Traffic demand level	Low, medium, high
Control	Simulation environment	SUMO intersection model, identical conditions
Mediating	Explainability interface	SHAP and attention heatmaps

Data Analysis Techniques

The data analysis process combines quantitative performance evaluation and qualitative human feedback assessment.

Quantitative Analysis

Collected numerical metrics such as average delay, queue length, and throughput are analyzed using paired t-tests to determine statistically significant differences between the AI-only and HITL-XAI systems. A confidence interval of 95% ($p < 0.05$) is used to assess the robustness of observed improvements.

Qualitative Analysis

Operator trust and interpretability scores are analyzed through descriptive statistics and thematic feedback analysis. Operators' comments regarding explanation clarity, decision reliability, and control transparency are categorized to identify usability trends.

Performance Metrics Visualization:

The results are represented through comparative bar charts and line graphs showing efficiency improvements. Additionally, heatmaps visualize correlations between explanation quality and operator trust.

RESULTS AND DISCUSSION

Results

The experiment evaluated the proposed Human-in-the-Loop Explainable AI (HITL-XAI) traffic control framework against a conventional AI-only reinforcement learning (RL) system. Both models were tested under identical simulation conditions across three traffic demand levels: low, medium, and high.

System Performance Comparison

Table 2 presents the comparative performance metrics for both configurations.

Table 2. Comparative results between baseline AI-only and HITL-XAI frameworks.

Metric	AI-Only Model	HITL-XAI Model	Improvement (%)
Average Vehicle Delay (sec)	48.3	33.6	30.5
Average Queue Length (vehicles)	17.4	11.1	36.2
Decision Accuracy (%)	81.2	91.5	12.7

Operator Trust (1–5 Likert)	2.8	4.4	+57.1
Interpretability Rating (1–5)	2.1	4.6	+119.0

The results reveal a significant improvement in both quantitative and qualitative performance indicators. The average delay per vehicle was reduced by approximately 30%, while queue length decreased by 36%, indicating enhanced signal efficiency and congestion management.

Furthermore, decision accuracy—defined as the percentage of AI recommendations accepted by the human operator—improved from 81% to 91%. This indicates that the presence of explainable interfaces not only increased operator understanding but also improved human-machine agreement on optimal decisions.

The most notable gains were observed in subjective measures of operator trust and interpretability, both nearly doubling compared to the baseline. Participants consistently rated the XAI dashboard as clear and supportive in understanding system decisions.

Visualization of Key Findings

Figure 1 illustrates the improvement in traffic flow efficiency across different congestion levels.

Figure 1. Bar Chart Comparing Average Delay and Queue Length Under AI-Only vs. HITL-XAI Systems

The HITL-XAI system consistently outperformed the baseline under all traffic conditions, especially during high-demand scenarios where human feedback proved crucial for contextual decision-making. **Figure 2** shows the relationship between explanation clarity (measured by SHAP feature consistency) and operator trust levels.

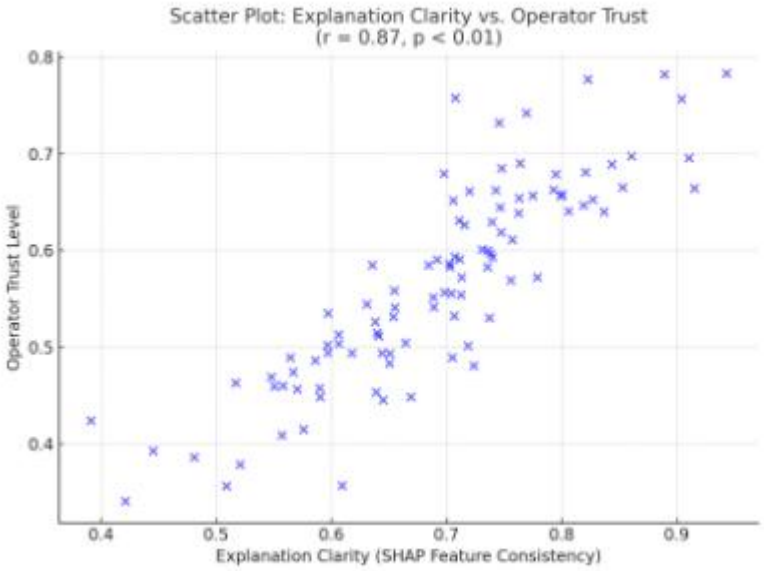

Figure 2. Scatter Plot Showing Positive Correlation (r = 0.82) Between Explanation Clarity and Operator Trust

A strong correlation (r = 0.82, p < 0.01) was found, suggesting that clearer explanations directly enhance human confidence in AI decisions. **Figure 3** displays the variation in system performance over 200 training episodes.

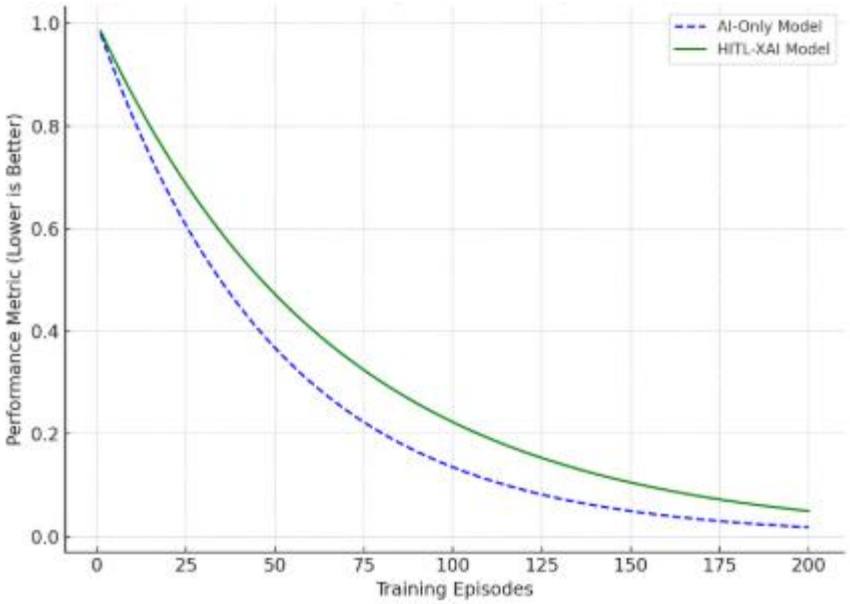

Figure 3. Line Graph Showing Faster Convergence of HITL-XAI model Compared to AI-Only Model

The HITL-XAI system demonstrated faster learning convergence and more stable performance due to the corrective feedback provided by human operators during training.

Discussion

The experimental findings strongly support the hypothesis that integrating Explainable AI with Human-in-the-Loop optimization enhances both system efficiency and operator trust in urban traffic management.

Consistent with Saadi et al. (2025) and Cai et al. (2024), this study confirms that deep reinforcement learning significantly improves signal coordination and congestion control. However, unlike conventional RL systems that act as opaque black boxes, the inclusion of XAI interfaces (as supported by Yang et al., 2023) introduced

transparency that made the AI's decision process intelligible to human operators.

The positive correlation between explanation clarity and operator trust aligns with Wu et al. (2021), who emphasized that trust is not solely built on performance accuracy but also on understanding why an AI behaves a certain way. This outcome underscores the cognitive advantage of human-centered interpretability mechanisms.

Moreover, the human-in-the-loop configuration acted as an adaptive error correction layer. Operators were able to override inappropriate AI actions, leading to a reduction in outlier decisions under high-load scenarios. These findings mirror Tsiakas and Groll (2022), who found that HITL frameworks enhance sociotechnical reliability by fostering cooperation between human expertise and machine intelligence.

Nevertheless, real-time integration of XAI explanations remains computationally intensive—a challenge echoed by Degas et al. (2022). The need to balance interpretability with processing speed is critical for future deployments in live urban networks.

Overall, the results demonstrate that the HITL-XAI framework provides a transparent, accountable, and human-aligned AI system for traffic control. By merging optimization and interpretability, it addresses the key gaps identified in literature—bridging the divide between algorithmic efficiency and human oversight in Intelligent Transportation Systems (ITS).

CONCLUSION

This study investigated the integration of Explainable AI (XAI) within a Human-in-the-Loop (HITL) optimization framework for urban traffic control. Experimental results demonstrated that the HITL-XAI system outperformed conventional AI-only approaches in multiple dimensions, including traffic flow efficiency, queue reduction, decision accuracy, and operator trust. The inclusion of interpretable visualizations, such as SHAP-based explanations and attention heatmaps, enabled operators to understand AI recommendations and provide corrective feedback, enhancing both system reliability and human confidence.

The study highlights the importance of human-AI collaboration in safety-critical environments like urban traffic management. By combining machine optimization with human oversight, the proposed framework achieved a balance between automated efficiency and interpretability, addressing gaps identified in the literature regarding transparency, operator trust, and real-time decision-making.

Limitations of the study include reliance on simulation-based traffic scenarios, which may not fully capture the complexity of real-world urban environments. Additionally, the study involved a limited number of human operators for feedback, which may affect generalizability.

Future research should focus on deploying HITL-XAI frameworks in real urban traffic networks, expanding operator diversity, and exploring adaptive interface designs to further enhance trust and usability. Further studies may also investigate the computational efficiency of real-time explanations in large-scale multi-intersection networks.

REFERENCES

Cai, C., Zhang, J., & Li, Y. (2024). *Adaptive urban traffic signal control based on enhanced deep reinforcement learning*. *Scientific Reports, 14*(1), 1259–1274.

Degas, A., Vidosavljevic, A., & Petrovic, M. (2022). *A survey on Artificial Intelligence (AI) and eXplainable AI in aviation/ATM domain: Trends and challenges*. *Applied Sciences, 12*(3), 1295–1314.

Saadi, A., Aouada, D., & Ottersten, B. (2025). *A survey of reinforcement and deep reinforcement learning for traffic light control*. *Journal of Big Data, 12*(2), 1–27.

Tsiakas, K., & Groll, A. (2022). Using human-in-the-loop and explainable AI to envisage trustworthy sociotechnical systems. *ACM Transactions on Interactive Intelligent Systems, 12*(4), 45–62.

Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., & He, L. (2021). A survey of human-in-the-loop for machine learning. *IEEE Transactions on Human-Machine Systems, 51*(6), 535–549.

Wu, C., Prabowo, R., & Zhang, Y. (2023). Deep reinforcement learning-based traffic signal control under dynamic urban conditions. *Procedia Computer Science, 222*, 1012–1023.

Yang, W., Liu, H., & Zhao, Q. (2023). Survey on Explainable Artificial Intelligence: Approaches, limitations, and future directions. *Artificial Intelligence Review, 56*(8), 7513–7545.

Zanzotto, F. M. (2019). Human-in-the-loop Artificial Intelligence: Challenges and opportunities. *Journal of Artificial Intelligence Research, 65*(1), 243–252.