AN INTELLIGENT TRAFFIC MANAGEMENT MODEL USING LARGE LANGUAGE MODEL AGENTS FOR ADAPTIVE TRAFFIC SIGNAL CONTROL IN NAIROBI, KENYA

MONICAH WAMBUI HINGA

A Thesis Submitted to the Institute of Postgraduate Studies of Kabarak University in Partial Fulfillment of the Requirements for the Award of Master of Science in Information Technology Degree

KABARAK UNIVERSITY

NOVEMBER, 2025

DECLARATION

1. I do hereby declare that:

    i. This thesis is my work, and to the best of my knowledge, it has not been presented for the award of a degree in any university or college.

    ii. The thesis has not incorporated material from other works or a paraphrase of such material without due and appropriate acknowledgment.

    iii. The thesis has been subjected to processes of anti-plagiarism and has met Kabarak University's 15% similarity index threshold.

2. I do understand that issues of academic integrity are paramount, and therefore, I may be suspended or expelled from the University, or my degree may be recalled for academic dishonesty or any other related academic malpractices.

Signed:________________________ Date:________________________

Monicah Wambui Hinga

GMI/ON/2213/09/22

RECOMMENDATION

To the Institute of Postgraduate Studies:

The thesis entitled "**An Intelligent Traffic Management Model using Large Language Model Agents for Adaptive Traffic Signal Control in Nairobi, Kenya,"** written by **Monicah Wambui Hinga,** is presented to the Institute of Postgraduate Studies of Kabarak University. We have reviewed the thesis and recommend it be accepted in partial fulfillment of the requirements for the award of the degree of Master of Science in Information Technology.

Signed:______________________ Date:_______________________

Prof. Simon Maina Karume

Department of Computer Science and Information Technology

Kabarak University

Signed:______________________ Date:______________________

Dr. John Chebor

Department of Computer Science and Information Technology

Kabarak University

COPYRIGHT

ACKNOWLEDGEMENT

I wish to express my sincere gratitude to the Almighty God for granting me the strength, wisdom, and perseverance to undertake and complete this thesis.

My deepest appreciation goes to my supervisors, Prof. Simon Maina Karume and Dr. John Chebor, for their invaluable guidance, encouragement, and constructive input throughout the research process. Their expertise and commitment have been instrumental in shaping the quality of this work.

I am also grateful to the Institute of Postgraduate Studies at Kabarak University and the Department of Computer Science & Information Technology for providing the necessary academic and administrative support throughout my studies.

Finally, I extend heartfelt appreciation to my family and friends for their unwavering support, patience, and understanding during this academic journey.

DEDICATION

This thesis is dedicated to my beloved family, whose love, support, and encouragement have been a constant source of strength throughout my academic journey.

ABSTRACT

Traffic congestion has remained a significant challenge in rapidly urbanizing cities such as Nairobi, Kenya, affecting mobility, productivity, environmental quality, and overall quality of urban life. Increased vehicle ownership, inadequate infrastructure expansion, and inefficient traffic management systems have exacerbated delays, air pollution, and economic losses. Conventional traffic signal control systems, largely based on fixed-timer logic, have often failed to respond effectively to real-time traffic dynamics, particularly at intersections where delays are most severe. This study developed and evaluated an Intelligent Traffic Management Model for Nairobi. The model integrated Large Language Model (LLM) agents to enable adaptive traffic signal control. Video traffic data were collected from selected intersections and processed using the YOLOv5 object detection algorithm due to its proven accuracy in real-time vehicle detection. These processed vehicle density estimates were then used to calibrate a traffic simulation framework within the Simulation of Urban Mobility (SUMO) environment, which was selected for its flexibility and ability to replicate realistic intersection behaviour. Within the simulation, LLM agents were integrated via an API call to analyze lane-specific traffic conditions and adjust green light durations dynamically based on observed vehicle counts and congestion levels. The research used experimental methodology, conducting scenario-based testing under varying traffic conditions, including peak-hour traffic, off-peak flows, road closures, and emergency vehicle passage. Performance was evaluated using key metrics such as average vehicle waiting time, intersection throughput, and responsiveness to changing traffic demand. The results indicated that the LLM-driven adaptive control model significantly reduced average vehicle waiting times and improved intersection throughput compared to fixed-time signal control. The system also demonstrated a higher responsiveness to dynamic traffic variations, contributing to smoother traffic flow and more balanced lane usage. This study concludes that integrating LLM agents into adaptive traffic signal control systems offers a scalable, data-driven approach to improving urban mobility in complex and resource-constrained environments. It recommends that transport authorities in Nairobi and similar cities consider pilot implementations of AI-powered traffic signal optimization as part of broader Intelligent Transportation Systems (ITS) strategies.

Keywords: *Traffic Congestion, Adaptive Traffic Signal Control, Large Language, Model Agents, Vehicle Detection, Intelligent Traffic Management*

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

AI	:Artificial Intelligence
ANOVA	:Analysis of Variance
API	:Application Programming Interface
ATSC	:Adaptive Traffic Signal Control
CAV	:Connected and Automated Vehicles
CBD	:Central Business District
DL	:Deep Learning
ITS	:Intelligent Transportation Systems
KNBS	:Kenya National Bureau Of Statistics
LLM	:Large Language Model
ML	:Machine Learning
RL	:Reinforcement Learning
SCATS	:Sydney Coordinated Adaptive Traffic System
SCOOT	:Split Cycle Offset Optimization Technique
SUMO	:Simulation Of Urban Mobility
V2I	:Vehicle-To-Infrastructure

CONCEPTUAL OPERATIONAL DEFINITION OF TERMS

Adaptive Signal Timing: The dynamic adjustment of traffic light durations based on real-time data to reduce congestion and improve flow efficiency (Federal Highway Administration, 2023).

Intelligent Transportation Systems (ITS): ITS are advanced applications integrating AI, sensors, and communication technologies to improve the safety, mobility, and efficiency of transportation systems (Fazzini et al., 2021).

Intersection Throughput: This is the number of vehicles successfully passing through an intersection within a specific time frame (Fattah et al., 2022).

Intersection Throughput: This is the number of vehicles successfully passing through an intersection within a specific time frame (Fattah et al., 2022).

Large Language Model (LLM) Agents: LLM agents are AI-driven systems capable of processing large datasets, recognizing patterns, and making decisions across various contexts (Brown et al., 2020).

Simulation Model: A virtual representation of a real-world system that allows testing and evaluation of strategies in a controlled environment (Baldi et al., 2020).

Traffic Congestion: A condition in which the volume of vehicles exceeds roadway capacity, resulting in slower speeds and longer delays (Litman, 2021).

Traffic Signal Control: The regulation of vehicle and pedestrian movement at intersections using traffic lights, which may include fixed or adaptive systems (Sharma & Gupta, 2020).

Urban Traffic Management: Application of strategies and technologies to optimize traffic flow and minimize congestion in urban areas (Sakr et al., 2023).

Vehicle Detection: The process of identifying and classifying vehicles in a traffic environment often involves using computer vision techniques (Rahman et al., 2021).

CHAPTER ONE

INTRODUCTION

1.1 Introduction

Urban mobility in Nairobi, Kenya, has long been constrained by inefficient traffic management, leading to severe delays at major intersections that impose significant economic, environmental, and social burdens. In Nairobi, fixed-time traffic signals are sometimes backed up by police officers, but they rarely match the city's fast-changing mix of cars, matatus, boda-bodas, and pedestrians. This often means longer queues, wasted fuel, and more air pollution. Many cities have implemented Intelligent Transportation Systems (ITS) featuring adaptive traffic signal control to improve traffic flow. However, most existing solutions depend on rigid, rule-based algorithms and costly sensor infrastructure, making them unsuitable for resource-constrained contexts such as Nairobi (Federal Highway Administration [FHWA], 2023).

Advances in Artificial Intelligence, particularly Large Language Models (LLMs), have created opportunities for more flexible and context-aware traffic control. LLMs are capable of processing complex, real-time datasets and making adaptive decisions without relying solely on pre-defined rules. In this study, an LLM-driven adaptive signal control framework was developed, utilizing vehicle counts extracted from intersection videos via the YOLOv5 object detection algorithm and integrated into a SUMO-based simulation to allocate green times dynamically. This chapter presents the study's background, problem statement, objectives, research questions, justification, significance, scope, and limitations.

1.2 Background to the Study

Traffic congestion is a global urban challenge affecting both developed and developing cities, with widespread implications for economic productivity, environmental sustainability, and public health. According to the International Transport Forum (2021), urban congestion contributes to over USD 300 billion in annual losses globally due to delays, fuel waste, and lost labour hours. Major big cities like New York, Paris, and Beijing, which use advanced systems, still face challenges at intersections. Delays, uneven flows, and rising travel demand remain common (Zhang et al., 2023). Despite significant infrastructure investments, conventional fixed-time traffic signal systems remain inadequate in Average Waiting Time fluctuating urban traffic demands, contributing substantially to congestion (Deshvena, 2024).

Across Africa, urban centres such as Lagos, Johannesburg, and Cairo face mounting traffic pressure due to rapid urbanization, increased motorization, and informal transport systems. Many African cities continue to rely on traditional traffic control mechanisms, with limited deployment of adaptive or intelligent systems to manage growing demand. The African Development Bank (AfDB, 2022) notes that inefficient traffic flow contributes to substantial economic drag across the continent, further complicating mobility for urban populations (Boamah & Akotey, 2020).

In Kenya, and particularly in Nairobi, traffic congestion is a persistent and deeply rooted issue. The city is characterized by a high concentration of vehicles within a limited road network, compounded by rapid population growth and spatial expansion. Nairobi's road infrastructure has not kept pace with the increasing demand for travel, resulting in daily traffic jams, particularly at major intersections. The Nairobi Metropolitan Area Transport Authority (NaMATA, 2020) reports that the central business district (CBD) regularly records vehicle speeds below 10 km/h during peak hours, far below the urban efficiency benchmark of 30 km/h. Delays are most acute at intersections where poorly coordinated traffic signals, coupled with diverse transport modes such as matatus (minibuses), boda-bodas (motorcycles), private vehicles, and pedestrians, lead to complex traffic interactions. In such scenarios, traffic officers are frequently deployed to manually manage intersection flow, especially during peak periods or signal outages. This reliance on manual intervention underscores the inadequacy of existing fixed-time signal control systems.

Moreover, Nairobi lacks a widespread implementation of intelligent or responsive signal technologies, making real-time traffic management difficult and largely reactive. To address such challenges, cities globally have turned to Intelligent Transportation Systems (ITS), particularly adaptive traffic signal control solutions that use real-time data to regulate traffic flow. Systems like the Sydney Coordinated Adaptive Traffic System (SCATS) and the Split Cycle Offset Optimization Technique (SCOOT) are among the most well-known, adjusting signal timing in response to actual traffic conditions (Alshayeb, Chang, & Chung, 2022). However, these systems are often rule-based and limited in their ability to adapt to highly unpredictable or multimodal traffic environments typical in cities like Nairobi (Agrahari et al., 2024).

Recent advances in Artificial Intelligence (AI) offer a promising alternative. Large Language Models (LLMs), such as OpenAI's GPT-4, demonstrate advanced capabilities in processing complex datasets, recognizing patterns, and making real-time decisions based on evolving scenarios (OpenAI, 2023). While LLMs have been largely applied in natural language tasks, their underlying architecture enables them to function as generalist decision-making agents, making them suitable for dynamic, data-driven control environments. Their use in traffic signal control presents an opportunity to enhance decision-making at intersections by enabling the flexible and responsive allocation of green time based on real-time traffic inputs (Wang, Li, & Zhang, 2024).

This study addressed congestion at Nairobi's intersections by investigating the integration of LLM agents into adaptive traffic signal control systems. To enhance the realism and practical grounding of the model, video data were collected from selected intersections to generate lane-specific vehicle counts. These counts were extracted using the YOLOv5 object detection algorithm and fed into a SUMO-based simulation environment, where LLM agents made data-driven decisions on signal phasing.

The primary stakeholders for this research included Nairobi's urban traffic authorities, transportation planners, and road users impacted by delays at major intersections. The knowledge gap lay in the absence of localized, AI-driven adaptive signal control solutions designed for Nairobi's heterogeneous traffic ecosystem. Most existing literature and implementations focused on well-instrumented, high-income cities with stable infrastructure and homogeneous traffic patterns. Nairobi, by contrast, presented a unique context where traditional systems had failed to meet growing demands. This study contributed to the advancement of Intelligent Transportation Systems and supported the development of context-appropriate traffic solutions for resource-constrained urban settings, demonstrating how emerging technologies such as LLMs can be adapted to address long-standing mobility challenges in Nairobi and similar cities.

1.3 Statement of the Problem

Traffic congestion in Nairobi remains a persistent and worsening challenge, particularly at road intersections where inefficient traffic signal control contributes to long vehicle queues, delays, and excessive fuel consumption. The city continues to rely on outdated fixed-time signal systems, occasionally supplemented by manual traffic officers, to manage traffic flows. These static approaches are inadequate in responding to real-time conditions, especially within Nairobi's multimodal traffic ecosystem, which comprises matatus (minibuses), private vehicles, motorcycles, pedestrians, and informal road users (Nairobi Metropolitan Area Transport Authority, 2020).

Although many global cities have implemented Intelligent Transportation Systems (ITS) and adaptive signal control strategies, such solutions are rarely deployed in African urban centres like Nairobi. Furthermore, most adaptive systems currently in use are based on rigid, rule-driven algorithms and require extensive infrastructure, such as loop detectors or sensor grid technologies, which are often impractical in low-resource contexts due to cost, maintenance, and technical limitations (Agrahari et al., 2024). Recent advancements in Artificial Intelligence (AI) present the possibility of more flexible, learning-based systems for urban traffic decision-making. Large Language Models (LLMs), including GPT-4, demonstrate advanced capabilities in pattern recognition, data interpretation, and adaptive reasoning across diverse domains (OpenAI, 2023). While these models have not yet been widely applied in traffic signal control, their capacity to process diverse, multimodal inputs and respond dynamically to changing scenarios presents a promising opportunity for exploration (Wang, Li, & Zhang, 2024).

A critical gap exists in research on examining how LLMs can be adapted as decision-making agents for intersection control in developing cities. Nairobi lacks both localized, data-driven adaptive traffic models and empirical studies that test the use of generalist AI models, such as LLMs, for congestion management. The absence of such research leaves transport authorities with limited intelligent, responsive, and scalable tools for addressing intersection delays. This study addressed that gap by investigating the integration of LLM agents into adaptive traffic signal control frameworks using video-based data from Nairobi intersections. Vehicle density data were extracted using the YOLOv5 object detection algorithm and used to simulate real-time traffic conditions within the Simulation of Urban Mobility (SUMO) platform. The evaluation assessed whether LLM agents could make timely and effective green-time decisions based on observed lane conditions and compared their performance against fixed-time control in terms of waiting time, throughput, and responsiveness.

1.4 Objectives of the Study

1.4.1 General Objectives of the Study

To develop an Intelligent Traffic Management Model using Large Language Model Agents for Adaptive Traffic Signal Control at intersections in Nairobi.

1.4.2 Specific Objectives of the Study

i. To analyze current traffic congestion patterns and the inefficiencies of traditional traffic signal control systems at intersections in Nairobi

ii. To collect and process traffic video data from selected Nairobi intersections using the YOLOv5 object detection algorithm to estimate vehicle density per lane.

iii. To simulate Nairobi's intersection traffic scenarios in the SUMO (Simulation of Urban Mobility) environment using extracted traffic density data.

iv. To design and implement a decision-making framework that integrates LLM agents for dynamic green-time allocation based on real-time traffic conditions.

v. To evaluate the performance of the LLM-based adaptive signal control model against fixed-time signal control in terms of average waiting time, vehicle throughput, and responsiveness to dynamic traffic demands.

1.5 Research Questions

i. What are the current traffic congestion patterns and inefficiencies of traditional traffic signal control systems at Nairobi intersections?

ii. How can traffic video data be processed to obtain vehicle density per lane at Nairobi intersections using YOLOv5?

iii. How can the extracted traffic counts be used to calibrate and simulate Nairobi intersections using the SUMO (Simulation of Urban Mobility) environment on observed traffic conditions?
iv. How can Large Language Model agents be integrated into an adaptive traffic signal control framework to allocate green time dynamically based on real-time inputs?
v. How does the LLM-based adaptive signal control perform relative to fixed-time control in terms of average waiting time, intersection throughput, and responsiveness to changing demand?

1.6 Justification of the Study

This study is justified by the urgent need to address the growing problem of traffic congestion in Nairobi, which continues to impose high economic, environmental, and social costs. Official reports from the Nairobi Metropolitan Area highlight substantial daily delays for commuters, resulting in lost productivity, increased fuel consumption, and higher transportation costs, which fall most heavily on low-income road users who rely on public transport (Nairobi Metropolitan Area Transport Authority [NaMATA], 2022). From an environmental perspective, prolonged idling at congested intersections elevates vehicular emissions, degrades air quality, and heightens health risks for urban residents (Schrank et al., 2023).

Despite these realities, many intersections still rely on fixed-time traffic signal plans and irregular manual control from traffic officers. These static approaches do not adapt to fluctuations in vehicle density or the multimodal dynamics of Nairobi’s roads, which include matatus (minibuses), motorcycles, private cars, and pedestrians, thus leading to inefficient flows, long queues, and a recurrent reliance on manual traffic direction. Contemporary reviews of adaptive signal control indicate that data-driven, learning-based approaches are more suitable for heterogeneous and rapidly changing conditions than rigid, rule-based systems (Agrahari et al., 2024).

In response, this study introduced an innovative, locally grounded solution by integrating Large Language Model (LLM) agents into adaptive traffic signal control. Lane-specific vehicle density per lane was obtained from intersection video using the YOLOv5 object-detection algorithm and fed into a SUMO-based simulation to evaluate LLM-guided signal decisions against the prevailing fixed-time approach. This provides a scalable and cost-sensitive alternative to infrastructure-heavy detection systems. Academically, this study fills a clear gap in the literature by testing the applicability of LLMs in a real-world urban context characterized by heterogeneous traffic. The findings provide evidence to inform transport authority decision-making, contribute to Intelligent Transportation Systems (ITS), and offer policy-relevant insights for improving mobility in Nairobi and similar cities across the region.

1.7 Significance of the Study

This study is important because it contributes to both academic research on Intelligent Transportation Systems (ITS) and practical efforts to improve urban traffic management in Nairobi. It explores how Large Language Model (LLM) agents can be utilized in adaptive traffic signal control, providing a data-driven approach to reducing congestion. For researchers, the study shows how LLMs, typically used for Natural Language Processing, can be adapted for real-time decision-making in urban mobility systems. It contributes to the growing discussion about using general-purpose AI models beyond language tasks by providing real-world insights into their flexibility, challenges, and potential for traffic optimization.

The research combines traffic pattern analysis, computer vision using YOLOv5, and simulation modeling in SUMO to build a realistic and locally relevant framework. This integrated model can serve as a guide for future studies that test AI for traffic control in various urban environments. Scholars studying smart city infrastructure in complex settings can also build on its design and results. For policymakers, transportation agencies, and city planners, the study offers practical guidance on evaluating traffic control strategies in cities with limited resources. It demonstrates that combining low-cost technologies with intelligent control algorithms can enhance intersection performance without incurring major financial investment. The findings will also support future smart mobility projects within Kenya's ongoing urban transport reforms.

1.8 Scope of the Study

The study focused on the development and evaluation of an intelligent traffic signal control model using Large Language Model (LLM) agents to address congestion challenges in Nairobi, Kenya. The geographical scope was limited to selected intersections within Nairobi's Central Business District (CBD) and adjacent high-traffic corridors, which represent the city's most congestion-prone areas. These intersections feature a diverse mix of transportation modes, including private vehicles, public minibuses (matatus), motorcycles (boda-bodas), and pedestrians, thus offering a suitable test environment for adaptive signal control strategies.

The LLM agents, especially models like GPT-4, were integrated within the Simulation of Urban Mobility (SUMO) environment and tested using traffic data from video recordings. Vehicle density per lane was obtained using the YOLOv5 object detection algorithm, which provides inputs to the simulation. The LLM agents were designed to analyze these inputs and dynamically allocate green light durations based on traffic conditions across the intersection. The study was confined to optimizing intersection traffic flow through adaptive signal control. The study did not extend to broader traffic management strategies such as road expansion, public transportation reforms, or policy-based interventions.

1.9 Limitations of the Study

This study employed video data collection from selected intersections in Nairobi to inform and calibrate the simulation model. The evaluation of the traffic signal control framework was carried out within a simulated environment using the SUMO platform. Although this approach allowed for systematic testing across varied traffic scenarios, it did not fully replicate all real-time complexities found in Nairobi's live traffic conditions. Human driving behavior, informal traffic practices, weather disruptions, and enforcement variability are examples of dynamic factors not captured in the simulation.

Another limitation arose from the current lack of comprehensive real-time traffic infrastructure in Nairobi. The city does not yet have a unified traffic surveillance or sensor network capable of supporting live adaptive traffic control systems. Consequently, the model relied on periodic, manually collected video samples rather than continuous data streams, which affected scalability in a real-world setting. The scope of the study was further limited to the optimization of traffic signal control at isolated intersections. Broader mobility strategies, such as road network redesign, public transport enhancements, or integrated urban mobility planning, were not addressed. These components, though important, fall outside the operational and methodological boundaries of this work. Despite these limitations, the study took a significant step toward bridging the gap between simulation and real-world traffic management. By grounding the framework in actual traffic data and applying cutting-edge AI technologies, such as Large Language Models, this research provides a practical and scalable foundation for future adaptive traffic control solutions in cities like Nairobi.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter reviews literature on urban traffic congestion, adaptive traffic signal control, and the use of Artificial Intelligence in traffic management. It begins by defining traffic congestion and why it matters for mobility, safety, and productivity. The review then traces the evolution of traffic signal control from fixed-time and actuated systems to adaptive approaches before examining how intelligent transportation solutions leverage computer vision for real-time detection at intersections and how simulation environments such as SUMO are used to test and compare signal strategies. The chapter also explains the advances in Large Language Model Agents and considers their potential to enable decision-making and coordination in traffic operations. The section concludes with a synthesis of key insights, the specific gaps that remain in the literature, and a conceptual framework that positions the study within both academic and practical contexts.

2.2 Traffic Congestion

Traffic congestion is a pervasive urban issue that disrupts mobility, reduces productivity, and exacerbates environmental challenges in cities worldwide. As urbanization accelerates, congestion becomes a critical problem, requiring innovative and adaptive solutions to mitigate its impacts. This section explores the concept of traffic congestion, its defining characteristics, and its implications, laying a foundation for understanding its causes and potential solutions.

2.2.1 Defining Traffic Congestion

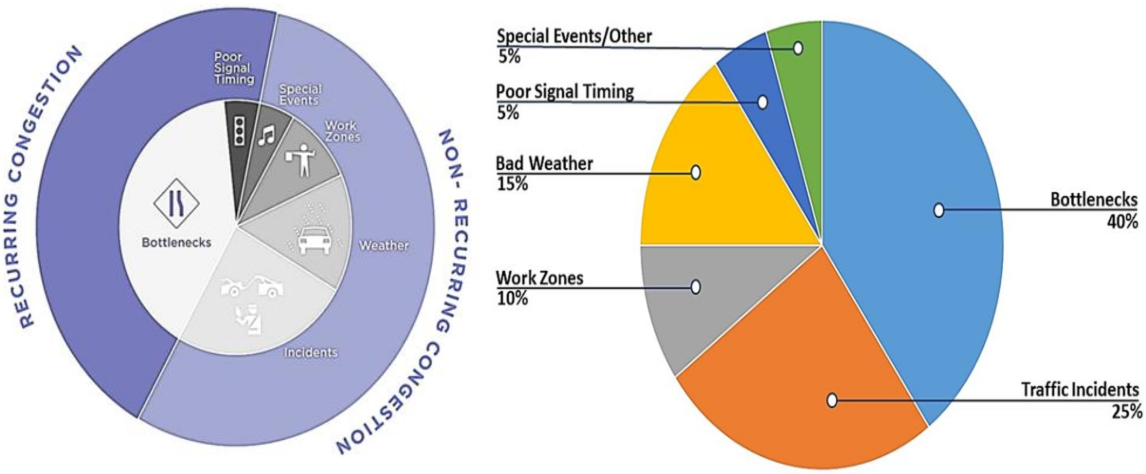
Traffic congestion refers to a condition where the demand for roadway space exceeds its capacity, resulting in slower vehicle speeds, longer trip durations, and increased queuing (Arti et al., 2022). It is a pervasive issue in urban centers worldwide, impacting not only metropolitan areas but also smaller cities, and poses significant economic, environmental, and social challenges. Congestion disrupts daily mobility, reduces productivity, and contributes to higher fuel consumption and emissions, underscoring its global significance (Schrank et al., 2023).

Traffic congestion is broadly categorized into two types: recurring and non-recurring. Recurring congestion occurs predictably during peak travel hours, primarily due to high vehicle volumes and limited road infrastructure. It reflects systemic inefficiencies in urban traffic management and infrastructure planning. Non-recurring congestion, on the other hand, arises from unpredictable incidents, such as accidents, road closures, adverse weather conditions, or special events. This sporadic nature makes it more challenging to address, as it requires real-time interventions and robust traffic management systems (Arti et al., 2022).

Understanding the types and causes of congestion is crucial for developing targeted solutions. While recurring congestion can be mitigated through adaptive systems, such as intelligent traffic signal controls, non-recurring congestion requires dynamic, real-time approaches, emphasizing the need for advanced technologies like artificial intelligence (AI) and machine learning (Gupta et al., 2023). By addressing these challenges, cities can enhance mobility, reduce emissions, and improve the overall quality of urban life. According to the Support for Urban Mobility Analysis (SUMA) Technical Memorandum, recurring congestion accounts for approximately 45% of total congestion, primarily caused by bottlenecks and poor signal timing, while non-recurring sources, such as incidents and adverse weather conditions, contribute 55% (Jha & Albert, 2021).

Figure 1

Traffic Congestion by type and Traffic Congestion by Percentage



Source: Jha & Albert, (2021)

2.2.2 Impacts of Traffic Congestion

Traffic congestion has profound impacts across economic, environmental, and social dimensions, making it a critical urban challenge that demands innovative solutions.

Economic Impacts

Traffic congestion imposes significant economic costs on individuals, businesses, and governments. Increased travel times due to congestion reduce worker productivity and lead to lost economic opportunities. For instance, in the United States, congestion resulted in approximately $160 billion in wasted time and fuel in 2017 (Litman, 2024). Furthermore, businesses face higher logistics costs due to delays in the delivery of goods and services, which they ultimately pass on to consumers. This inefficiency hinders regional competitiveness and slows economic growth. In Nairobi, where congestion is a daily reality, the economic losses are exacerbated by limited infrastructure and rapid urbanization, further straining the city's development potential (Kenya National Bureau of Statistics (KNBS), 2020).

Environmental Impacts

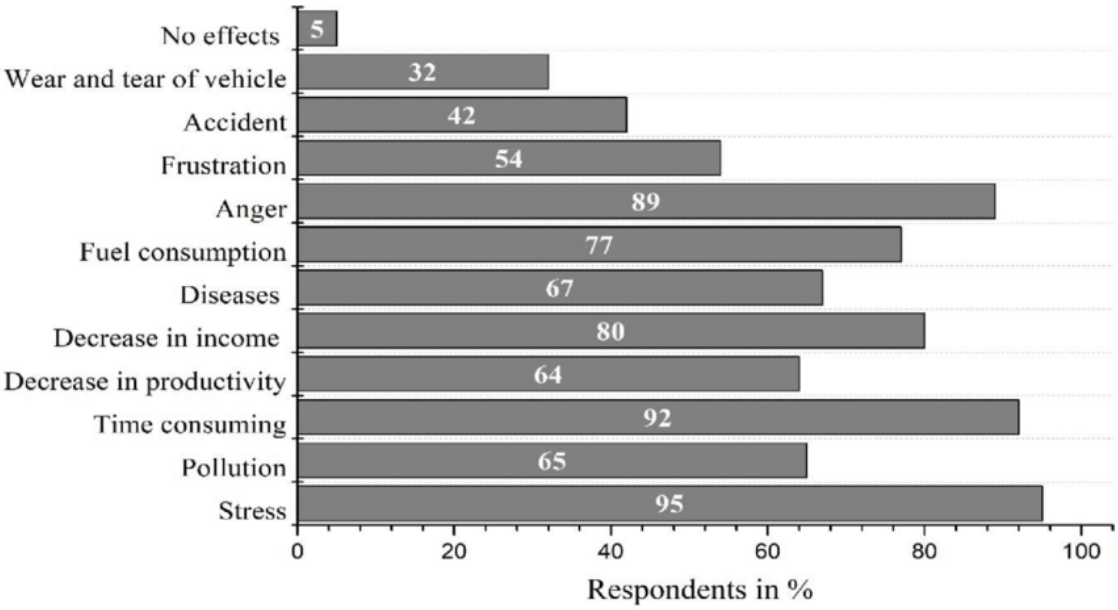
Congested traffic significantly contributes to environmental degradation. Idling vehicles release elevated levels of greenhouse gases, including carbon dioxide, and air pollutants such as nitrogen oxides and particulate matter. These emissions exacerbate air pollution, contributing to respiratory illnesses and climate change. According to research on urban mobility solutions, cities that implement congestion pricing, such as Stockholm, have reported notable reductions in air pollution and improvements in public health (Schrank et al., 2023). In Nairobi, the high volume of slow-moving vehicles exacerbates air quality issues, particularly in densely populated areas, posing significant health risks to residents. The environmental costs of congestion highlight the urgency for adaptive traffic management solutions to mitigate emissions and promote sustainability.

Social Impacts

The social consequences of traffic congestion are equally concerning. Long hours spent in traffic lead to increased stress, reduced time for family and leisure activities, and a diminished quality of life. For commuters in cities like Nairobi, congestion disproportionately affects lower-income individuals, who often rely on less efficient public transportation systems or have limited flexibility in their travel schedules (Arti et al., 2022). Prolonged congestion also strains social interactions and community well-being, as individuals spend excessive time on the road rather than engaging in meaningful activities. In addition, traffic congestion exacerbates inequality, as wealthier individuals are more likely to have access to alternative modes of transportation or live closer to their workplaces.

Figure 2

Impacts of Traffic Congestion on Respondents in Chittagong City, Bangladesh



Source: Fattah et al., (2022).

2.3 Traffic Signal Control Systems

Traffic signal control systems have undergone significant evolution since their inception, adapting to the growing complexities of urban transportation networks. This section provides an overview of their historical development, current applications, and emerging trends aimed at optimizing traffic management.

2.3.1 Evolution of Traffic Signal Control System

The first known traffic signal was installed in London in 1868, designed by British railway engineer J.P. Knight. This manually operated system used semaphore arms during the day and gas-lit red and green lights at night to control horse-drawn carriages and pedestrian traffic. However, it was short-lived due to safety concerns after a gas explosion injured a police officer (Larry Clark, 2019). The advent of electric traffic signals marked a significant advancement in traffic management. In 1914, the first electric traffic signal was installed in Cleveland, Ohio. This system featured red and green lights and a buzzer to indicate colour changes, laying the foundation for modern traffic signalling (Miovision Team, 2024).

By the mid-20th century, fixed-time traffic signal systems became prevalent. These systems followed preset cycles, changing lights after fixed intervals even when the roads were sometimes empty or overcrowded. While they were straightforward and cost-effective, their rigidity often led to inefficiencies during fluctuating traffic volumes. To address these limitations, actuated traffic signal control systems were introduced. These systems utilized sensors to detect vehicle presence, adjusting signal timings in real-time to accommodate actual traffic demand, thereby improving flow and reducing unnecessary delays (Tomar et al., 2022).

Recent advancements have led to the development of Adaptive Traffic Signal Control Systems (ATCS), which dynamically adjust signal timings based on real-time traffic conditions. These systems employ sensors, cameras, and artificial intelligence to optimize traffic flow, reduce congestion, and enhance overall efficiency. For instance, the implementation of ATCS in smart cities has shown significant improvements in traffic management (Shankaran & Rajendran, 2021). The integration of artificial intelligence (AI) and machine learning into traffic signal control is an emerging trend that aims to enhance traffic management. AI-driven adaptive systems can predict traffic patterns and adjust signals proactively, improving responsiveness to real-time conditions. Additionally, the development of vehicle-to-infrastructure (V2I) communication enables direct interaction between traffic signals and connected vehicles, facilitating more efficient traffic flow and paving the way for the integration of autonomous vehicles (Agrahari et al., 2024).

2.3.2 Fixed-Time Traffic Signal Control

Fixed-time traffic signals are among the first traffic control methods. They follow preset schedules, switching lights at regular intervals, regardless of whether cars are actually waiting (Tomar et al., 2022). The simplicity and cost-effectiveness of fixed-time systems led to their widespread adoption in urban areas throughout the mid-20th century, as they required minimal infrastructure and maintenance. However, their rigidity often led to inefficiencies, particularly during periods of fluctuating traffic volumes or unpredictable congestion patterns.

Fixed-time systems are designed using historical traffic data to estimate average traffic volumes at specific times of the day. While this approach may work well for intersections with consistent and predictable traffic flows, it fails to account for real-time variations caused by accidents, weather conditions, or fluctuating demand (Tomar et al., 2022). For instance, during off-peak hours, vehicles often experience unnecessary delays at intersections because the system's unable to adjust signal timings dynamically. One notable shortcoming of fixed-time systems is their inability to prioritize certain types of traffic, such as public transportation or emergency vehicles. This limitation can exacerbate congestion in urban centers where traffic patterns are highly dynamic and diverse (Chen, 2024). Additionally, fixed-time systems often result in higher fuel consumption and increased emissions, as vehicles idle unnecessarily at traffic signals, contributing to environmental degradation.

Despite these limitations, fixed-time traffic signal control systems laid the foundation for more advanced approaches, such as actuated and adaptive systems. Their straightforward design and ease of implementation make them suitable for smaller cities or low-traffic intersections where real-time demand fluctuations are minimal. However, as cities expand and traffic complexities grow, there is a need to transition to more intelligent and flexible traffic signal control solutions (Sakr et al., 2023).

2.3.3 Actuated and Semi-Actuated Signal Control System

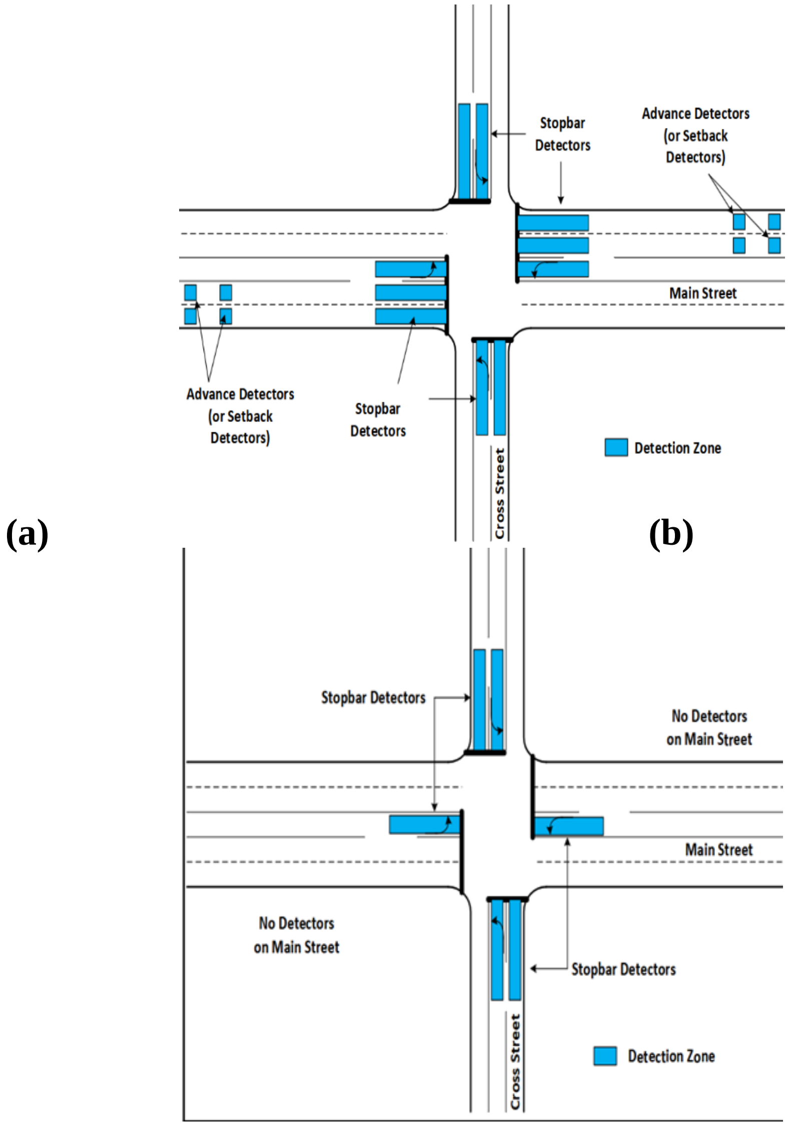
Actuated and semi-actuated signal control systems emerged as a solution to address the limitations of fixed-time traffic signal controls. These systems introduced dynamic signal adjustments based on real-time traffic data, significantly improving efficiency and reducing delays at intersections. Fully actuated systems rely on traffic detection technologies such as inductive loop detectors, cameras, or radar to monitor vehicle presence and flow. These systems adjust the signal timings dynamically to accommodate varying traffic demands. For example, if a certain lane remains unoccupied, the system can allocate green light time to the busier lanes, improving overall intersection throughput. Actuated systems can adjust to changing conditions, which makes them more useful at intersections where traffic is unpredictable or uneven (FHWA, 2023).

However, these systems require accurate sensor installations and regular maintenance to function effectively. Malfunctioning sensors or poor calibration can lead to inefficient signal timings or misallocated green light durations, thereby reducing the system's benefits (FHWA, 2023). Semi-actuated systems are designed for intersections where one road consistently carries heavier traffic than others. Sensors are typically installed on the minor roads, while the major roads follow fixed signal cycles. This prioritization ensures minimal disruption for high-traffic approaches while accommodating vehicles from less busy directions only when necessary. Semi-actuated systems are cost-effective compared to fully actuated systems as they require fewer sensors and less infrastructure (FHWA (Federal Highway Administration), 2023).

Both actuated and semi-actuated systems offer considerable advantages over traditional fixed-time signals by reducing delays and optimizing intersection performance. However, their effectiveness depends on the quality of the underlying detection technologies and regular maintenance. In areas with complex intersections or high traffic variability, these systems may fall short without integration with more advanced adaptive control technologies. The integration of actuated and semi-actuated signal control systems has significantly enhanced urban traffic management by reducing congestion and delays, making them a critical component of modern Intelligent Transportation Systems.

Figure 3

Fully Actuated Control Detector and Semi-Actuated Control Detector



Source: FHWA (Federal Highway Administration), 2023) Traffic Signal Program Handbook. U.S. Department of Transportation)

2.3.4 Adaptive Traffic Signal Control System

Adaptive Traffic Signal Control (ATSC) systems are designed to dynamically adjust signal timings in response to real-time traffic conditions, aiming to optimize traffic flow and reduce congestion. Unlike fixed-time or semi-actuated systems, ATSC uses data and algorithms to manage traffic. This makes it more effective in busy and complex urban settings. This real-time responsiveness helps reduce congestion and improve overall traffic efficiency (Federal Highway Administration [FHWA], 2023).

ATSC systems have been implemented in numerous cities worldwide to enhance traffic efficiency by adjusting signal timings in response to real-time traffic demand. One prominent example of ATSC application is the multi-agent reinforcement learning approach, which treats each traffic signal controller as an independent agent that optimizes traffic flow around its specific intersection. This approach has been successfully deployed in Toronto, Canada, through the MARLIN-ATSC system. The deployment demonstrated a significant reduction in average delays, ranging from 27% to 39%, and travel-time savings of up to 26% along busy routes. By coordinating multiple intersections through real-time communication, the system provided efficient traffic management at a network scale (El-Tantawy et al., 2013).

Reinforcement learning-based ATSC systems have also been tested in environments featuring Connected and Automated Vehicles (CAVs). For instance, a study by Maadi et al. (2022) combined reinforcement learning with CAV speed guidance to minimize queue lengths and stop delays at intersections. The system demonstrated superior performance compared to fixed-timing and traditional actuated control systems, particularly under saturated conditions, highlighting the potential for ATSC systems to adapt to emerging vehicle technologies. The application of ATSC systems has also extended to improving environmental outcomes, such as air quality. Fazzini et al. (2021)

conducted experiments using a multi-agent adaptive control system in Bologna, Italy, and found that it significantly reduced pollutant emissions due to smoother traffic flows and fewer vehicle stops.

In Nairobi, traffic congestion remains a major challenge due to rapid urbanization, high vehicle density, and limited infrastructure. The city's traffic signal systems have historically relied on fixed-time controls, which have proven inadequate for managing fluctuating traffic demands. Efforts to introduce ATSC systems have shown promise in mitigating congestion, although their implementation has faced several challenges. One major challenge is the availability and maintenance of accurate detection infrastructure, such as sensors and cameras. Inadequate detection can compromise the effectiveness of ATSC systems, as demonstrated by research in developing regions where infrastructure constraints limit adaptive capabilities. (Mishra et al., 2023).

One key limitation of ATSC systems is the scalability in large and complex traffic networks, where coordination between multiple intersections becomes challenging. Furthermore, while ATSC systems often improve efficiency and reduce delays, their impact on safety can vary. For example, Kodi et al. (022) found that although ATSC systems reduced fatal and injury crashes, they were sometimes associated with an increase in total and angle crashes, highlighting the need for careful calibration and monitoring.

2.4 Intelligent Transportation Systems (ITS) and the Role of AI

This section explores the concept and components of Intelligent Transportation Systems (ITS), their integration with Artificial Intelligence (AI), and how these technologies are reshaping traffic management.

2.4.1 Intelligent Transportation Systems

Intelligent Transportation Systems (ITS) integrate advanced technologies to enhance the efficiency, safety, and sustainability of urban transportation networks. ITS combines cameras, sensors, and communication tools to monitor traffic in real time. This helps managers make quicker decisions and improve traffic flow (Sakr et al., 2023). The seamless interaction between these components ensures a more responsive and adaptive traffic system capable of addressing urban mobility challenges.

Applications of ITS include adaptive traffic signal control, electronic toll collection, and vehicle-to-infrastructure (V2I) communication. Adaptive traffic signal control systems dynamically adjust signal timings in response to real-time traffic demands, reducing delays and improving overall flow. Similarly, V2I communication enhances the interaction between vehicles and traffic infrastructure, ensuring smoother operations and improved road safety (Agrahari et al., 2024). These technologies also promote environmental sustainability by minimizing fuel consumption and greenhouse gas emissions, as optimized traffic flow reduces idle times and vehicle stops (Todd Litman, 2024).

The implementation of an Intelligent Transport System (ITS) by Huawei, in collaboration with the Kenya Urban Roads Authority (KURA), has demonstrated significant potential in alleviating traffic congestion and enhancing urban mobility. This system integrates intelligent surveillance cameras, traffic flow cameras, variable timing traffic lights, and a centralized control centre to monitor and manage traffic in real-time. Deployed at key intersections, such as the Western Ring Road from Yaya Centre through Kileleshwa to Waiyaki Way, the ITS enables dynamic adjustments to traffic lights and efficient coordination across intersections. The results include smoother traffic flow, reduced congestion, decreased reliance on manual traffic control, and improved commuter experiences (Muli, 2020; Odhiambo, 2020).

Despite its numerous advantages, the implementation of ITS faces significant challenges. High deployment costs, maintenance requirements, and the complexity of integrating these systems into existing infrastructure often limit their widespread adoption (Tomar et al., 2022). Furthermore, the reliability of ITS heavily depends on the accuracy of real-time data inputs from detection systems. Any inaccuracies or failures in sensors and communication networks can compromise the system's efficiency and effectiveness.

2.4.2 Role of AI in Intelligent Transportation Systems

Artificial Intelligence (AI) has played a huge role in the evolution of Intelligent Transportation Systems (ITS), evolving from basic data processing tools to sophisticated, learning-based decision-making systems capable of optimizing complex traffic networks. Early applications of AI in ITS focused on automating routine tasks, such as traffic signal control and basic traffic flow predictions, using rule-based systems. These initial systems offered limited adaptability and primarily relied on predefined sets of rules, lacking the capacity to respond dynamically to evolving traffic conditions (Agrahari et al., 2024).

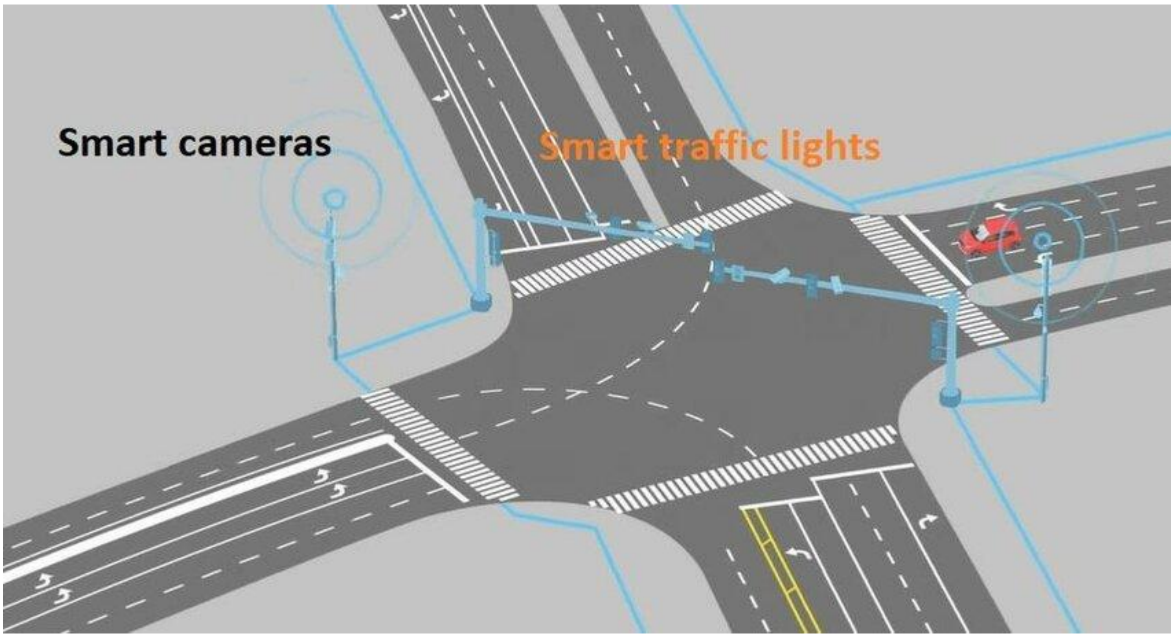
In the last two decades, AI has advanced rapidly. Machine Learning (ML) and Deep Learning (DL) now play a significant role in the design and implementation of Intelligent Transportation Systems (ITS). These advances have enabled AI systems to analyze and predict traffic patterns, optimize traffic signal timings in real-time, and even adapt to sudden changes such as accidents or extreme weather conditions. The integration of AI techniques, such as Reinforcement Learning (RL), has further transformed the traffic control landscape by enabling systems to learn optimal behaviors through trial and error, leading to increasingly effective traffic management outcomes (Dikshit et al., 2023).

AI for traffic optimization approaches utilize various techniques, including predictive Analytics, Machine Learning Algorithms, and Reinforcement Learning Models, to enhance traffic flow and minimize congestion. One of the primary applications of AI in ITS involves real-time traffic prediction and routing optimization. By analyzing vast amounts of real-time data from sensors, cameras, and connected vehicles, AI systems can predict congestion hotspots and dynamically adjust traffic signals, reroute vehicles, and manage traffic flows to prevent bottlenecks. For example, a study by Dikshit et al. (2023) demonstrated that AI-driven vehicle routing and traffic optimization systems reduced travel times, fuel consumption, and pollution, contributing to overall transportation efficiency and sustainability (Dikshit et al., 2023).

Despite its successes, the implementation of AI-driven traffic optimization faces several challenges, including ensuring the accuracy and robustness of predictive models, addressing privacy concerns related to data collection, and managing the computational demands of AI systems. Continuous innovation, data integration, and collaborative efforts among stakeholders, including government agencies and technology providers, are essential for overcoming these challenges and realizing the full potential of AI in ITS.

Figure 4

AI in Intelligent Traffic Management



Source: Soori et al., (2023)

2.4.3 AI Techniques in Traffic Signal Control

This section examines the application of AI techniques, such as Reinforcement Learning, Fuzzy Logic Systems, and Neural Networks, in traffic signal control.

Reinforcement Learning

Reinforcement Learning (RL) is a powerful AI technique widely applied in traffic control systems to optimize decision-making processes and enhance traffic flow at intersections. Traditional systems depend on fixed rules or past data. In contrast, RL agents learn by interacting with their environment and improving over time. By receiving feedback in the form of rewards or penalties based on their actions, RL agents improve their decision-making over time. One popular application of RL in traffic control is the use of Deep Q-Networks (DQN) to optimize signal timings at intersections, which has demonstrated significant improvements in reducing traffic delays and congestion (Swapno et al., 2024).

Multi-agent reinforcement learning (MARL) systems further extend this approach by coordinating multiple intersections to achieve network-wide traffic optimization. In MARL, each intersection's traffic controller acts as an independent agent that collaborates with neighboring controllers, enabling decentralized yet coordinated signal control strategies. This technique has shown promising results in large-scale traffic networks, reducing travel times and congestion through effective collaboration and adaptive learning (Mushtaq et al., 2023).

Neural Networks

Neural networks, particularly deep learning models, are increasingly being used in traffic control to analyze complex patterns and optimize control strategies. Deep Neural Networks (DNNs) and Convolutional Neural Networks (CNNs) are capable of processing large volumes of traffic data, extracting relevant features, and learning complex relationships that influence traffic behavior. In traffic signal control applications, Neural Networks are used to predict traffic demand, optimize signal timings, and provide adaptive control in response to real-time conditions. For instance, integrating deep reinforcement learning with Neural Networks enables the development of agents that continuously learn and adapt to fluctuating traffic scenarios, reducing overall congestion (Wan & Hwang, 2018).

Hybrid systems combining Neural Networks and Reinforcement Learning have also been developed to achieve more efficient traffic control. These systems leverage the learning capabilities of Neural Networks and the adaptability of RL to create intelligent controllers capable of self-optimization. In an example study, Zhao et al. (2022) proposed a traffic signal control algorithm based on a Fuzzy Neural Network combined with Reinforcement Learning, which demonstrated improved performance in optimizing traffic signal timings compared to traditional approaches (Zhao et al., 2022).

Fuzzy Logic-Based Approaches

Fuzzy logic offers an effective means of managing the uncertainty and variability inherent in traffic systems. Unlike conventional logic systems that rely on binary true-or-false states, fuzzy logic operates with degrees of truth, enabling more nuanced decision-making. Fuzzy logic-based traffic controllers use "if-then" rules to adapt to changing traffic conditions, making them particularly well-suited for dynamic and unpredictable urban environments. By integrating linguistic terms such as "medium" or "heavy" traffic volumes, fuzzy controllers can adjust signal timings in a way that is both flexible and human-understandable (Zhang et al., 2021). One significant challenge of fuzzy logic-based systems is the reliance on expert knowledge to define the "if-then" rules and membership functions, which can be subjective and labour-intensive. Furthermore, fuzzy systems can struggle with scalability when dealing with large, complex traffic networks, as the number of rules required may grow exponentially.

2.5 Visual Traffic Detection and Object Recognition

In the pursuit of Intelligent Traffic Management, obtaining accurate, timely, and localized traffic data is essential. Traditional traffic monitoring methods, such as inductive loop detectors, magnetic sensors, and manual surveys, while still in use, often lack the flexibility and granularity needed to support real-time, adaptive traffic signal control systems. To overcome these limitations, modern systems increasingly rely on computer vision techniques, particularly those involving video analytics and deep learning-based object detection (Zhang et al., 2023).

Visual traffic detection relies on cameras that capture traffic in real-time. The footage is then analyzed with AI tools to spot, classify, and track vehicles. Compared to fixed-position sensors, camera-based systems provide wider spatial coverage and richer contextual awareness, including multi-lane monitoring and vehicle movement trajectories. These attributes make visual sensing especially effective in complex, high-density traffic environments (Rana et al., 2024).

A significant breakthrough in this domain has been the application of object detection algorithms, particularly YOLO (You Only Look Once), which offer real-time performance with high accuracy. YOLOv5, the most popular variant in recent years, is known for its balance between inference speed and detection precision, making it suitable for deployment in real-world urban settings (Alam et al., 2025). YOLOv5 processes full video frames in a single forward pass, allowing for instant identification of vehicles, pedestrians, and other objects at intersections (Geetha, 2024). In the context of this study, visual traffic detection is a core input layer within the proposed Intelligent Signal Control framework. Video recordings from Nairobi intersections will be analysed using YOLOv5 to extract vehicle counts per lane. These counts will serve as the primary real-time input into a SUMO simulation environment, where Large Language Model (LLM) Agents will make adaptive traffic signal decisions.

This integration ensures that the simulation reflects true intersection dynamics, bridging real-world observation with AI-driven control logic (Zhang et al., 2023). The subsequent subsections examine the evolution of object detection algorithms in traffic applications and present a focused discussion on the capabilities and performance of YOLOv5 in vehicle detection.

2.5.1 Role of Visual Sensing in Urban Traffic Management

With recent advancements in computer vision and AI, visual sensing now plays a central role in managing traffic systems in real time. By utilizing video feeds from strategically placed cameras, visual sensing systems can monitor, analyze, and manage traffic flows in real time, offering a dynamic alternative to traditional traffic monitoring methods. One of the primary advantages of visual sensing is its ability to provide comprehensive coverage of traffic conditions without requiring intrusive infrastructure. Unlike inductive loop detectors or magnetic sensors, which require physical installation within roadways, camera-based systems can be mounted on existing structures, reducing deployment costs and minimizing disruptions during installation (Elbasha & Abdellatif, 2025).

Moreover, visual sensing systems can capture a wealth of data beyond mere vehicle counts. They can identify vehicle types, monitor pedestrian movements, and detect incidents such as accidents or stalled vehicles. This multifaceted data collection enables traffic management centers to make informed decisions, optimizing signal timings and improving overall traffic flow (Azfar & Ke, 2024). By incorporating AI algorithms, visual sensing systems can adapt to changing traffic patterns, learn from historical data to predict congestion, and adjust traffic signals proactively. For instance, the integration of YOLOv5, a state-of-the-art object detection model, enables the accurate and rapid identification of vehicles, thereby enhancing the responsiveness of traffic control systems (Rahman et al., 2021). Visual sensing emerges as a promising approach in urban traffic management, providing scalable, cost-effective, and intelligent solutions to the complex challenges of modern transportation systems.

2.5.2 Object Detection Algorithms in Traffic Control

The evolution of object detection algorithms has significantly influenced the development of Intelligent Traffic Management Systems. Early methods, such as the Region-based Convolutional Neural Network (R-CNN) and its variants, including Fast R-CNN and Faster R-CNN, introduced a two-stage detection process. These models first generate region proposals and then classify them, achieving high accuracy but often at the expense of computational efficiency, making them less suitable for real-time applications (Ren et al., 2015).

To address the need for faster detection, single-stage detectors such as the Single Shot MultiBox Detector (SSD) and the You Only Look Once (YOLO) series have been developed. These models perform object localization and classification in a single pass, significantly reducing inference time while maintaining competitive accuracy. Among these, the YOLO family has gained prominence for its balance between speed and precision, making it particularly suitable for real-time traffic monitoring scenarios (Redmon et al., 2016).

The YOLO series has undergone several iterations, with YOLOv5 emerging as a widely adopted version due to its enhanced performance and ease of deployment. YOLOv5 introduces improvements such as the use of Cross Stage Partial (CSP) networks, which enhance feature extraction capabilities, and the integration of the Path Aggregation Network (PANet) for better information flow across different scales. These enhancements contribute to more accurate detection of objects at various sizes and distances, a critical requirement in traffic environments where vehicles and pedestrians can appear at different scales (Glenn Jocher et al., 2020).

In the context of traffic control, object detection algorithms like YOLOv5 are used to identify and track vehicles, pedestrians, and other road users in real-time. This capability enables dynamic adjustments to traffic signals based on current road conditions, thereby improving traffic flow and reducing congestion. For instance, a study by Shen et al. (2023) demonstrated the effectiveness of an improved YOLOv5 model in detecting traffic signs with higher accuracy, which is essential for autonomous driving and intelligent traffic systems.

Moreover, the adaptability of YOLOv5 enables customization for specific traffic scenarios. Researchers have developed variants, such as YOLOv5-TS, tailored for traffic sign detection. These variants integrate modules like spatial pyramid pooling and multiple feature fusion to enhance detection performance, particularly for small and occluded objects (Shen et al., 2023). The progression from two-stage to single-stage object detection algorithms has significantly advanced the capabilities of traffic control systems. YOLOv5 is both fast and accurate, making it a strong option for modern traffic monitoring and management.

2.5.3 YOLOv5 for Real-Time Vehicle Detection at Intersections

Real-time vehicle detection at intersections is pivotal for adaptive traffic signal control systems. YOLOv5, a state-of-the-art object detection model, has demonstrated exceptional performance in this domain due to its balance between speed and accuracy. In a study by Zhang et al. (2022), an improved YOLOv5 model was proposed to enhance vehicle detection in various traffic scenarios. The model employed the Flip-Mosaic data augmentation technique to address challenges such as occlusion and varying object scales, resulting in improved detection accuracy and reduced false detection rates.

Alam et al. (2025) developed an optimized YOLOv5-based approach for real-time vehicle detection at road intersections using fisheye cameras. Their method addressed issues like light glare, shadows, and non-linear distortions inherent in fisheye imagery. By introducing a lightweight day-night CNN classifier and up-sampling challenging instances in the dataset, the model achieved a 13.7% improvement in mean Average Precision (mAP) over the standard YOLOv5. Furthermore, integrating YOLOv5 with reinforcement learning (RL) agents has shown promise in adaptive traffic control. By supplying real-time vehicle counts from YOLOv5 to an RL agent, traffic lights can be adjusted dynamically based on actual traffic conditions, leading to reduced congestion and improved traffic flow (Kisetzuu, 2024).

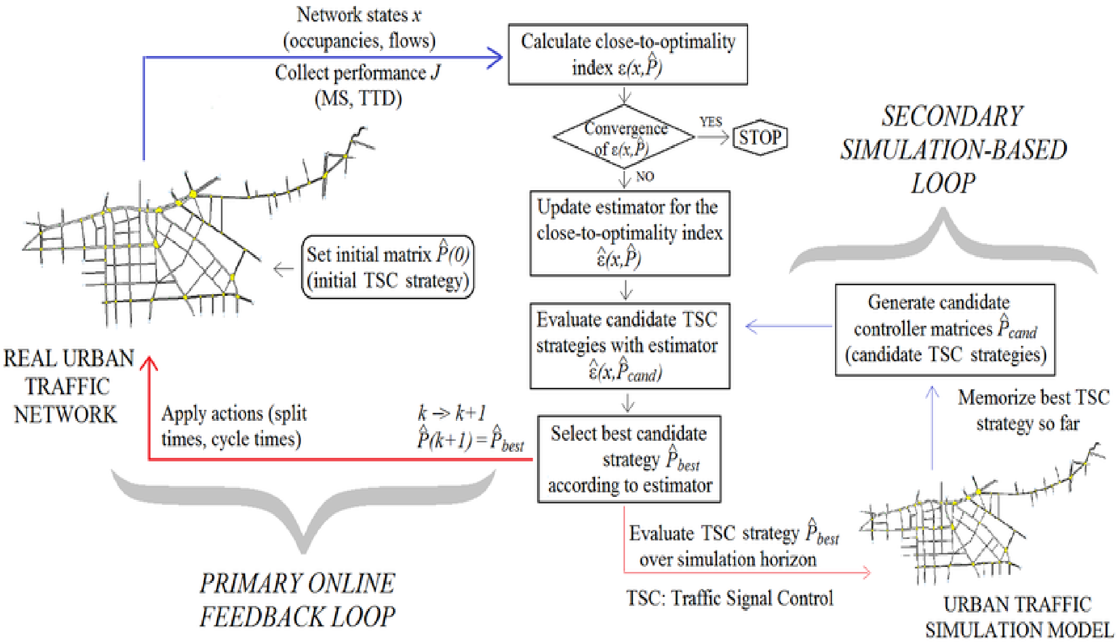
In the context of this study, YOLOv5 served as the primary tool for extracting real-time traffic data from video feeds at intersections in Nairobi. The detected vehicle counts per lane were then utilized by a Large Language Model (LLM) agent within a SUMO simulation environment to make informed traffic signal timing decisions. This integration ensured that the traffic control system was responsive to real-world conditions, enhancing its effectiveness in managing urban traffic.

2.6 Simulation for Traffic Management

Simulation for Traffic Management provides a practical and effective approach to addressing urban mobility challenges by creating virtual models that replicate real-world traffic conditions. These models enable researchers and policymakers to test and refine traffic management strategies, optimizing flow and reducing congestion before implementation in real environments (Pragalathan & Schramm, 2024). By simulating various scenarios, including peak traffic, incidents, and infrastructure changes, this approach facilitates data-driven decision-making for sustainable urban mobility.

Figure 5

Simulation-based Control Design Set-up



Source: Baldi et al., (2020)

Modern simulation techniques incorporate advanced algorithms, allowing for the analysis of traffic behavior at microscopic, macroscopic, and mesoscopic levels. This flexibility ensures detailed evaluations of traffic signal optimization strategies, infrastructure changes, and AI-driven models such as reinforcement learning. Widely used tools such as SUMO, VISSIM, and AIMSUN provide robust platforms for traffic simulation, enabling accurate modeing and informed policy decisions (Dian Khumara et al., 2018).

2.6.1 Simulation Techniques

Simulation techniques are critical tools in traffic engineering, enabling the replication and analysis of real-world traffic conditions in virtual environments. These techniques allow researchers to evaluate various traffic management strategies, test innovative solutions, and predict outcomes under different scenarios without affecting actual traffic systems (Pinto et al., 2024). The primary simulation techniques, including microscopic, macroscopic, and mesoscopic, are detailed below.

Microscopic Simulation

Microscopic simulation focuses on the detailed behavior of individual vehicles and their interactions within the traffic network. These models consider driver decisions, such as acceleration, deceleration, and lane changes, to create highly detailed and accurate representations of traffic flow. Microscopic simulations are particularly useful for evaluating the impacts of adaptive traffic signal controls and analyzing complex intersections (Pinto et al., 2024). They also provide a platform for testing AI-driven systems, such as reinforcement learning algorithms, in managing traffic flow dynamically.

Macroscopic Simulation

Macroscopic simulation examines traffic systems at an aggregate level, using metrics like traffic density, flow rates, and average speeds. These models are ideal for studying larger-scale networks, such as highways or regional transportation systems, where detailed vehicle-level data is less critical. Macroscopic simulation techniques are commonly used to evaluate policy changes, infrastructure development, and the overall performance of urban traffic systems (Pinto et al., 2024).

Mesoscopic Simulation

Mesoscopic simulation bridges the gap between microscopic and macroscopic models by combining individual vehicle dynamics with higher-level traffic flow analysis. These models capture vehicle interactions at a simplified level while providing insights into broader traffic patterns. Mesoscopic simulations are effective for corridor-level studies and evaluating mixed-mode transportation systems, such as the integration of public transit with private vehicles (Pinto et al., 2024).

Agent-Based Modelling

Agent-based modelling, a subset of microscopic simulation, represents individual entities such as vehicles or pedestrians as "agents" with unique decision-making processes. This approach is particularly suitable for simulating scenarios involving AI-based systems, such as neural networks or Large Language Model (LLM) agents, to optimize traffic signal control. Agent-based models enable researchers to investigate the impact of diverse behaviors on overall traffic flow and evaluate innovative traffic management strategies in complex urban networks (Gurram et al., 2019). Simulation techniques are indispensable for developing and refining AI-driven traffic management solutions, particularly in cities like Nairobi, where traffic systems are highly dynamic and unpredictable. By leveraging these techniques, researchers can ensure that new systems are both effective and adaptable before being implemented in the real world.

2.6.2 Tools for Traffic Simulation

Traffic simulation tools are vital in traffic engineering, offering sophisticated platforms to model, analyze, and optimize transportation systems. These tools enable researchers and policymakers to evaluate the performance of traffic control strategies, test AI-driven solutions, and predict the outcomes of infrastructure modifications. This section highlights some of the most widely used traffic simulation tools, including their key features and applications.

Simulation of Urban Mobility (SUMO)

SUMO is an open-source traffic simulation platform widely recognized for its versatility and scalability. It supports microscopic, macroscopic, and mesoscopic traffic modeling, making it suitable for analyzing a wide range of traffic scenarios. SUMO's ability to integrate with AI algorithms, such as reinforcement learning and neural networks, has made it a preferred choice for testing adaptive traffic signal control systems (Dian Khumara et al., 2018). Its open-source nature allows for customization, enabling researchers to tailor the tool to a specific urban environment, such as Nairobi.

VISSIM

VISSIM, developed by PTV Group, is a leading microscopic traffic simulation software known for its detailed modelling of vehicle and pedestrian interactions. Its robust visualization capabilities and support for multi-modal traffic systems make it an ideal choice for evaluating complex intersection designs and public transit integration (PTV Group, 2020). VISSIM is frequently used in AI-driven traffic management research due to its ability to simulate real-time traffic conditions and test adaptive control strategies.

AIMSUN

AIMSUN is another popular traffic simulation tool that offers microscopic, mesoscopic, and hybrid modelling capabilities. It is particularly effective for corridor-level and large-scale network simulations, providing insights into traffic flow, congestion patterns, and the impacts of new infrastructure projects. AIMSUN's advanced analytics and integration with AI tools make it a valuable resource for developing and optimizing traffic management strategies (Pragalathan & Schramm, 2024).

PARAMICS

PARAMICS is a microscopic simulation tool designed for detailed traffic modelling, including signalized intersections, roundabouts, and highway systems. It is equipped with advanced features for analyzing driver behaviours, lane-changing dynamics, and vehicle interactions. PARAMICS has been used in various studies to test AI-driven traffic signal optimization techniques, particularly Reinforcement Learning Models (Li et al., 2020).

2.7 Large Language Model

This section reviews Large Language Models (LLMs) as advanced reasoning systems that go beyond traditional Natural Language Processing (NLP) tasks. It focuses on their use in decision making, prediction, and control. The section introduces their core architectures and main model families, then examines how LLMs and LLM agents are being applied in traffic and urban management. In this study, LLMs are utilized to connect visual detection with algorithmic control through a multimodal reasoning process, spanning from vision to vehicle counts and ultimately to the LLM's decision output. The section also outlines how LLM control systems are evaluated, covering both traffic performance indicators and model-specific factors, such as accuracy, consistency, and the handling of hallucination risks.

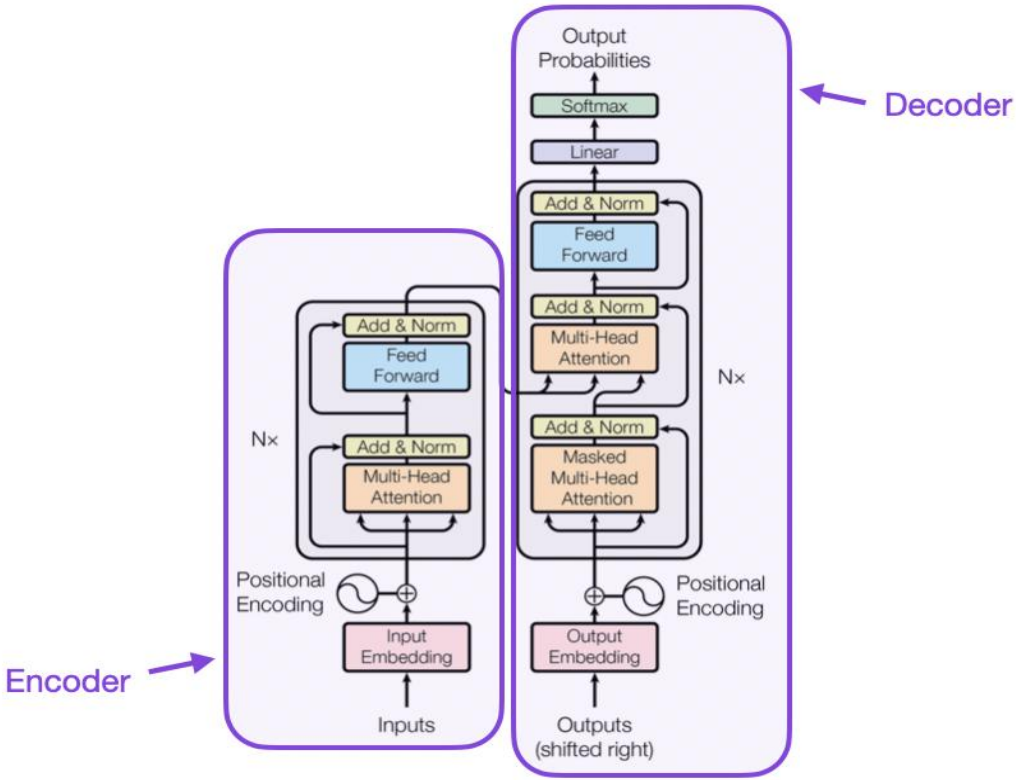
2.7.1 Overview of Large Language Models

Large Language Models (LLMs) are advanced Artificial Intelligence systems designed to process, understand, and generate text by learning statistical patterns from very large datasets and modern neural network architectures. Exemplars such as GPT (Generative Pre-trained Transformer) and BERT (Bidirectional Encoder Representations from Transformers) are trained on billions of tokens, enabling them to recognize context, produce coherent outputs, and support tasks that range from summarization to decision support. Although first developed for Natural Language Processing (NLP), their adaptability now extends to domains like planning, prediction, and real-time optimization found in Intelligent Transportation Systems (Brown et al., 2020; Devlin et al., 2019).

At the core of most LLMs is the transformer architecture. Transformers utilize self-attention to capture long-range dependencies and process tokens in parallel, thereby improving both learning efficiency and representation quality (Vaswani et al., 2017). Models are first pre-trained on broad text corpora with next-token or masked-token objectives. Later, they are adapted to downstream tasks through instruction tuning and alignment procedures such as Reinforcement Learning from Human Feedback (RLHF) (Wei et al., 2022). These stages make the models more reliable at following prompts, producing structured outputs (e.g., JSON), and adhering to domain constraints and properties. These qualities are especially valuable when an LLM serves as a decision aid in safety-sensitive settings, such as traffic control (Ouyang et al., 2022; OpenAI, 2023).

Figure 6

The Transformer model Architecture



Source: Collier, (2024)

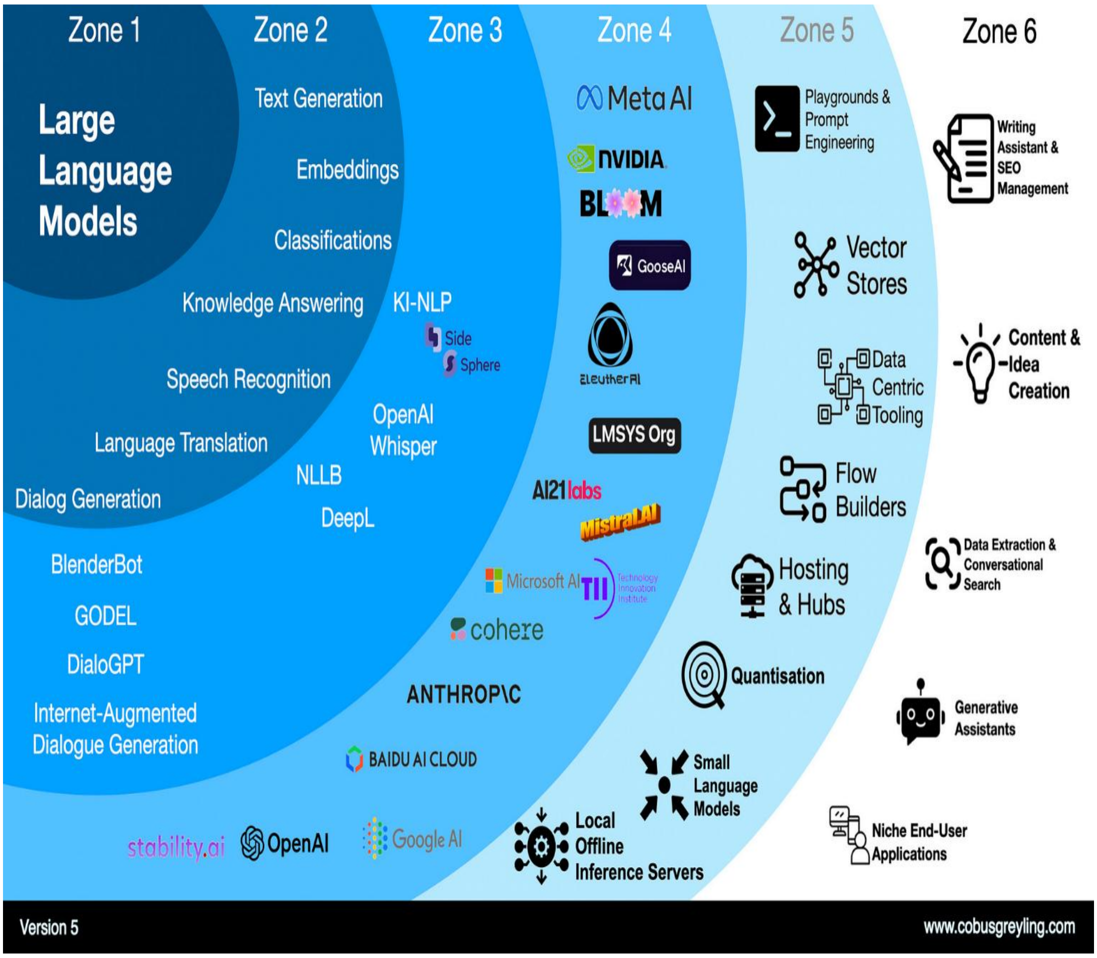
From an architectural perspective, the literature commonly distinguishes three families. Encoder-only models' example: BERT and RoBERTa learn bidirectional representations via masked-token prediction and excel at understanding tasks such as classification and retrieval (Devlin et al., 2019). Decoder-only (causal) models, such as GPT-3/4, Llama, and Mistral, predict the next token and are well-suited to generation, reasoning, and tool use interfaces, making them natural candidates for agentic control loops (Touvron et al., 2023). Encoder–decoder (sequence-to-sequence) models' example; T5/FLAN-T5 map inputs to outputs and remain strong baselines for transformation tasks such as translation and abstractive summarization (Raffel et al., 2020).

A second classification is by modality. Text-only LLMs process and produce text while Large Multimodal Models (LMMs) natively integrate text with images and, in newer systems, audio or video. This enables perception-plus-reasoning workflows (OpenAI, 2023; Team Google, 2024). In transportation research, a practical design approach is to keep perception and reasoning modular: a vision model, such as a YOLO-style detector, converts video streams into structured features, including per-lane counts, which a text-conditioned LLM then uses to recommend actions. This “vision-to-text” pattern preserves privacy, keeps decisions auditable, and reduces computation relative to streaming raw frames into an LMM.

Deployment literature separates models by openness. Proprietary API models, such as GPT-4/4o, Claude, and Gemini, provide strong performance, built-in safety tools, and, in some cases, native multimodality. However, they introduce recurring costs and data-governance considerations. Open-weight families, such as Llama-3 and Mistral/Mixtral, enable local or private-cloud inference, provide tighter control over data residency, and support cost-scalable operations, while achieving state-of-the-art performance on many reasoning tasks (Touvron et al., 2023; Jiang et al., 2023; Meta AI, 2024). For public-sector traffic centers, both categories are relevant: APIs are attractive for rapid prototyping and benchmarking, while open weights support on-premises deployments and reproducibility. The literature indicates that decoder-only, instruction-tuned tools using LLMs and fed with structured traffic signals from a vision pipeline are well-suited to adaptive signal control.

Figure 7

The Large Language Model Landscape — Version 5



Source: Cobus Greyling, (2024)

2.7.2 Applications of LLMs Beyond NLP

Large Language Models (LLMs) have demonstrated remarkable versatility, extending their impact beyond Natural Language Processing (NLP) into a wide array of areas. These fields demand sophisticated data analysis and predictive capabilities. Their ability to analyze and synthesize vast, heterogeneous datasets equips them to address complex, data-intensive challenges, including those in traffic management, logistics, and urban planning (Siqi Lai et al., 2023).

In traffic management, LLMs have been applied to enhance decision-making processes at urban intersections. A study by Masri et al. (2024) explored the use of an LLM, specifically GPT-4o-mini, to analyze, predict positions, detect, and resolve conflicts at intersections in real-time. The findings indicated that the LLM effectively managed scenarios involving heavy traffic, congestion, and mixed speed conditions, highlighting its potential to improve traffic efficiency and safety by providing real-time analysis. The study demonstrated that the fine-tuned GPT-mini achieved an accuracy of 83% and an F1-score of 0.84, along with high ROUGE-L scores for conflict identification (0.95), decision-making (0.91), and waiting time optimization (0.92) (Masri et al., 2024). These results highlight the capacity of LLMs to minimize congestion, reduce delays, and ensure smoother traffic flow. By dynamically adjusting signal timings and prioritizing vehicle movements, LLMs have proven to be effective in high-density urban settings where traditional traffic systems often fail.

LLMs have also been integrated into urban planning to enhance activity planning and management. Jiang et al. (2024) introduced UrbanLLM, an autonomous system utilizing LLMs for urban activity planning. The study demonstrated that UrbanLLM significantly outperformed other established LLMs in handling complex urban planning tasks, indicating its potential to reduce the workload and reliance on human experts in urban scenarios (Jiang et al., 2024). The versatility of LLMs extends well beyond traditional NLP applications, offering transformative potential in traffic management, urban planning, and adaptive traffic signal control. Their ability to process and interpret complex datasets positions them as valuable tools in addressing contemporary challenges across various sectors.

Table 1

UrbanLLM against other LLMs

	Accuracy	Precision	Recall	F1
Llama2-7b	0.18%	10.52%	8.75%	9.18%
Vicuna-7b-v1.5	8.44%	14.08%	13.89%	13.95%
Llama3-8b	5.31%	12.96%	15.50%	13.08%
GPT-3.5	17.95%	23.25%	22.35%	22.54%
GPT-4o	49.99%	55.31%	54.42%	54.63%
UrbanLLM	**68.30%**	**80.05%**	**79.26%**	**79.49%**
% Improve	36.63%	44.73%	45.64%	45.50%

Source: Jiang et al., (2024)

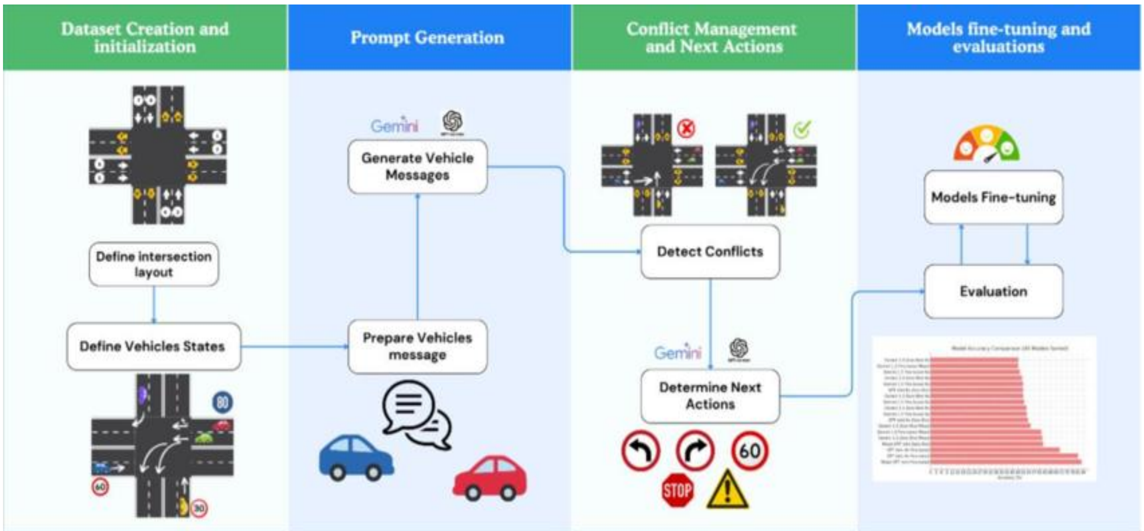
2.7.3 Integration of LLM Agents

This study has integrated Large Language Model (LLM) agents as intelligent controllers within an adaptive traffic signal framework. Building upon traffic data derived from the visual detection system YOLOv5, the proposed model utilizes LLMs to process vehicle density per lane and dynamically allocate green time at intersections. These agents were integrated into the SUMO (Simulation of Urban Mobility) platform, which provides a realistic environment for testing traffic management strategies under various conditions, including congestion surges, off-peak flows, and emergency vehicle scenarios.

The developed architecture positioned the LLM agent as a central decision-making engine, responding to traffic conditions identified through video analytics. Video feeds from Nairobi intersections were processed through YOLOv5, a deep learning model for object detection, which extracts vehicle counts in real-time. This lane-specific data served as structured input for the LLM, which interprets patterns and recommends optimal signal phases. Unlike traditional rule-based systems, the LLM agent is not hard-coded with static decision trees; instead, it applies contextual reasoning and probabilistic inference across incoming data to adjust signal timing strategies. This positioned the LLM agents to better respond to Nairobi's traffic.

Figure 8

LLM-Based Traffic Control Framework



Source: Masri et al., (2025)

2.7.4 Adaptive Traffic Signal Control Using LLM Agents

Adaptive traffic signal control refers to the ability of a traffic management system to adjust signal timings in response to real-time traffic conditions dynamically. In this study, this adaptive capability was achieved through the deployment of a Large Language Model (LLM) agent, which processed traffic density data extracted from video recordings and used that input to adjust signal phase durations within the SUMO simulation environment intelligently. The LLM agents received structured traffic data, specifically vehicle counts per lane, generated through YOLOv5-based object detection. This data represented real-time traffic pressure on each lane approaching the intersection. Instead of applying a pre-defined signal schedule or using conventional actuated control logic, the LLM interpreted this data. It made control decisions based on contextual patterns such as uneven demand, congestion buildup, or time-of-day variations.

Upon each decision cycle, the LLM evaluates the current lane densities and determines the optimal green-time allocation for each direction. For instance, if the vehicle count from the westbound lane was significantly higher than from the others, the agent prioritized a longer green phase for that movement, while reducing the time on less congested lanes. The decision-making process was both data-informed and flexibly adaptive, meaning the LLM could revise its timing logic as conditions evolved throughout the simulation. This LLM-driven adjustment was passed into SUMO's traffic light logic system using TraCI (Traffic Control Interface), allowing real-time control of the simulation's traffic signals. By doing so, the system mimicked a real-world deployment in which traffic signals were controlled dynamically by an intelligent agent that perceives actual traffic flow conditions. The approach enabled the evaluation of key performance indicators such as average queue length, waiting time, and vehicle throughput under different traffic demand scenarios.

More importantly, the LLM's advantage lay in its generalist reasoning ability as it was not constrained by pre-coded rules or static response tables. Instead, it drew on learned traffic patterns and scenario data to propose more nuanced and adaptive timing solutions.

2.7.5 Evaluation Metrics for LLM Agents

A rigorous assessment of an LLM traffic-signal agent should track both system-level mobility outcomes and agent-level decision quality. For system performance, this study adopts standard traffic-signal control metrics widely used in recent LLM-TSC work: average travel time (ATT), average queue length (AQL), and average waiting time (AWT). ATT reflects the mean trip duration; AQL captures the mean number of vehicles queued; AWT quantifies the time vehicles spend waiting at intersections. These indicators are the basis of published LLM Light comparisons against fixed-time, rule-based, reinforcement-learning, and LLM-based policies, enabling like-for-like evaluation across methods (Lai et al., 2024).

Agent-level metrics complement system KPIs by verifying that the model's outputs are valid, safe, and efficient. The evaluation tracks: (a) task success - the share of decisions that produce a valid, constraint-respecting phase plan; (b) constraint-violation rate - frequency of illegal/unsafe moves example conflicting greens; (c) hallucination rate - unsupported or structurally invalid content (e.g., malformed actions or non-existent phase IDs); (d) decision latency - time to produce a plan, which determines real-time viability; and (e) token/compute cost - average prompt and completion tokens or seconds per control step, which informs operating expenditure. Recent LLM Light results report batch latencies and deployment cost analyses suitable for such engineering metrics, alongside human expert assessment of reasoning and interpretability. These are useful precedents for agent-level evaluation (Lai et al., 2024).

To make these checks automatic and reproducible, the agent emits structured, typed actions, such as phase identifiers and green times, in a fixed schema, enabling deterministic validation and consistent benchmarking across models and prompts. This aligns with published designs that restrict the observation and action spaces to auditable features and legal phases before actuation, reducing opportunities for unsupported actions (Lai et al., 2024). Where classification subtasks are involved, examples include conflict detection or rationale quality, with metrics such as accuracy, precision, recall, F1-score, and text-similarity metrics (e.g., ROUGE-L) also tracked.

In a comparative study across the GPT-mini, Gemini, and Llama families for intersection conflict identification and decision-making, a fine-tuned GPT-mini achieved ~83% accuracy with an F1≈score of ~0.84, and high ROUGE-L scores across conflict identification and waiting time optimization. This illustrates how such metrics reveal trade-offs across model families and training setups (Masri et al., 2024). A two-stage protocol is recommended before any field deployment: (1) simulation-based evaluation using

ATT/AQL/AWT under varied demand patterns and random seeds to establish robustness, followed by (2) human-in-the-loop review of agent rationales and edge cases by traffic engineers to surface failure modes and improve prompts. Prior work incorporates expert feedback to assess interpretability and real-world applicability, a procedure that this study mirrors (Lai et al., 2024).

Table 2

LLM-agent Performance on Traffic Conflict Detection

Model	Learning Method	Scenario	Accuracy	Precision	Recall	F1-Score
GPT-4o-mini	fine-tuning	mixed-vehicle	**83.0**	**0.83**	**0.85**	**0.84**
	fine-tuning	4-vehicle	**81.0**	**0.80**	**0.83**	**0.82**
	fine-tuning	8-vehicle	**71.0**	**0.70**	**0.74**	**0.72**
	zero-shot	mixed-vehicle	61.9	0.59	0.62	0.60
	zero-shot	4-vehicle	53.6	0.53	0.54	0.53
	zero-shot	8-vehicle	50.8	0.50	0.51	0.42
Gemini 1.5	zero-shot	8-vehicle	52.8	0.55	0.77	0.62
	fine-tuning	4-vehicle	51.5	0.51	0.48	0.50
	zero-shot	4-vehicle	50.9	0.51	0.54	0.52
	fine-tuning	8-vehicle	49.1	0.49	0.50	0.49
	fine-tuning	mixed-vehicle	48.2	0.49	0.72	0.58
	zero-shot	mixed-vehicle	61.4	0.62	0.60	0.61
Gemini 1.0	fine-tuning	mixed-vehicle	60.9	0.61	0.62	0.61
	zero-shot	mixed-vehicle	55.1	0.57	0.42	0.49
	fine-tuning	4-vehicle	52.9	0.55	0.73	0.61
	fine-tuning	8-vehicle	50.8	0.60	0.05	0.09
	zero-shot	4-vehicle	50.4	0.50	0.58	0.54
	zero-shot	8-vehicle	48.2	0.47	0.48	0.45
Llama-3.1-8B-Instruct	zero-shot	mixed-vehicle	50.4	0.52	0.50	0.37
	fine-tuning	4-vehicle	49.6	0.49	0.50	0.43
Llama-3.1-70B-Instruct	fine-tuning	4-vehicle	51.5	0.51	0.51	0.51

Source**:** Masri et al., (2025)

2.7.6 Hallucinations, Reliability, and Safety in LLM Agents

A recurring concern in the LLM literature is hallucination. These are outputs that are fluent but unsupported by inputs or domain facts. In traffic signal control, hallucinations can manifest as unsafe or infeasible actions, i.e., proposing conflicting green phases and exceeding legal bounds, or as ungrounded explanations that contradict the observed state. Recent surveys formalize taxonomies of hallucination and discuss root causes, including imperfect training data, over-generalization, and alignment gaps, thus motivating explicit grounding and verification for any safety-critical use (Huang et al., 2023; Ji et al., 2023).

The transport-specific literature reflects these concerns and embeds safeguards into the LLM agent loop. In the LLM Light project, the observation space is restricted to auditable features, i.e., approaching and queuing counts that are verbalized for the model, while the control action space is explicitly defined and passed to the model during prompting. This design constrains decisions to legal phases and documented timing ranges before execution, thereby reducing opportunities for unsupported actions (Lai, Xu, Zhang, Liu, & Xiong, 2024). In addition, the LLM Light project reports qualitative human-expert evaluations and a deployment analysis that highlight practical safeguards (e.g., data-privacy preservation, region-specific regulations), reinforcing the need for verifiable, explainable decisions in real-world settings (Lai et al., 2024).

Reliability, in this context, entails both decision quality and operational stability. Decision quality is assessed by the rate of action-validity (no conflict violations; adherence to phase bounds), schema compliance (well-formed JSON actions), and groundedness (decisions consistent with sensed queues/approaches). Operational stability adds decision latency per control cycle and the system's ability to remain robust under traffic fluctuations. Empirically, LLM Light demonstrates consistent advantages and maintains the shortest travel and waiting times during day-long demand shifts. This is evidence that an agent with constrained actions can remain reliable across non-stationary conditions (Lai et al., 2024).

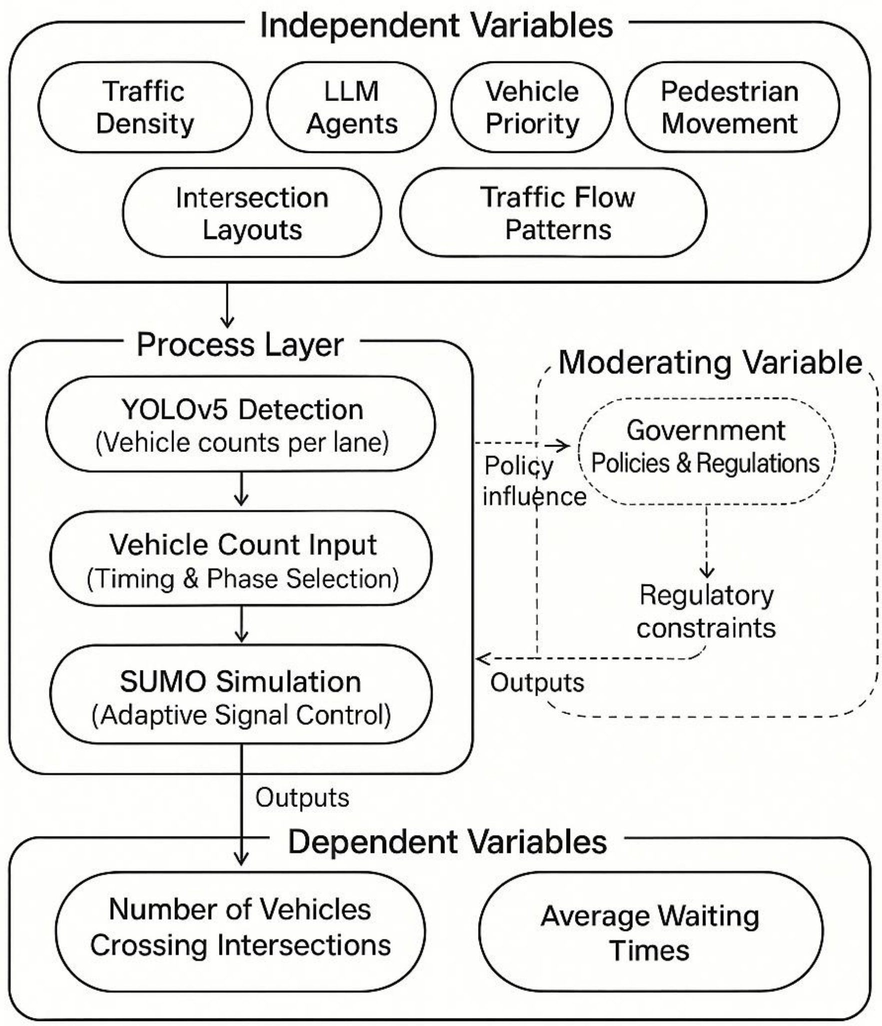
Mitigation strategies complement these measurements. First, constrained prompting and schema-validated outputs limit the space of admissible actions; invalid or out-of-range proposals are rejected or auto-repaired before actuation. Second, commonsense augmentation and critic-guided refinement enhance alignment with domain priorities, emphasizing long queues and reducing spurious reasoning in generalist backbones (Lai et al., 2024). Third, governance and privacy safeguards emphasize data residency and policy compliance, which are crucial for city agencies operating camera feeds and connected infrastructure (Lai et al., 2024). Finally, because legal signal timing and phase-conflict rules ultimately bound acceptable behavior, agent designs in the literature tie decisions to established traffic-engineering guidance and simulator-verified constraints before any field use.

The review indicates that LLM agents can be made trustworthy enough for simulated evaluation and pilot studies when their perceptions are structured and auditable, their actions are formally constrained and validated, and their outputs are interpretable to domain experts.

2.8 Conceptual Framework

Figure 9

Conceptual Framework



Source: Researcher, (2025)

2.9 Overview

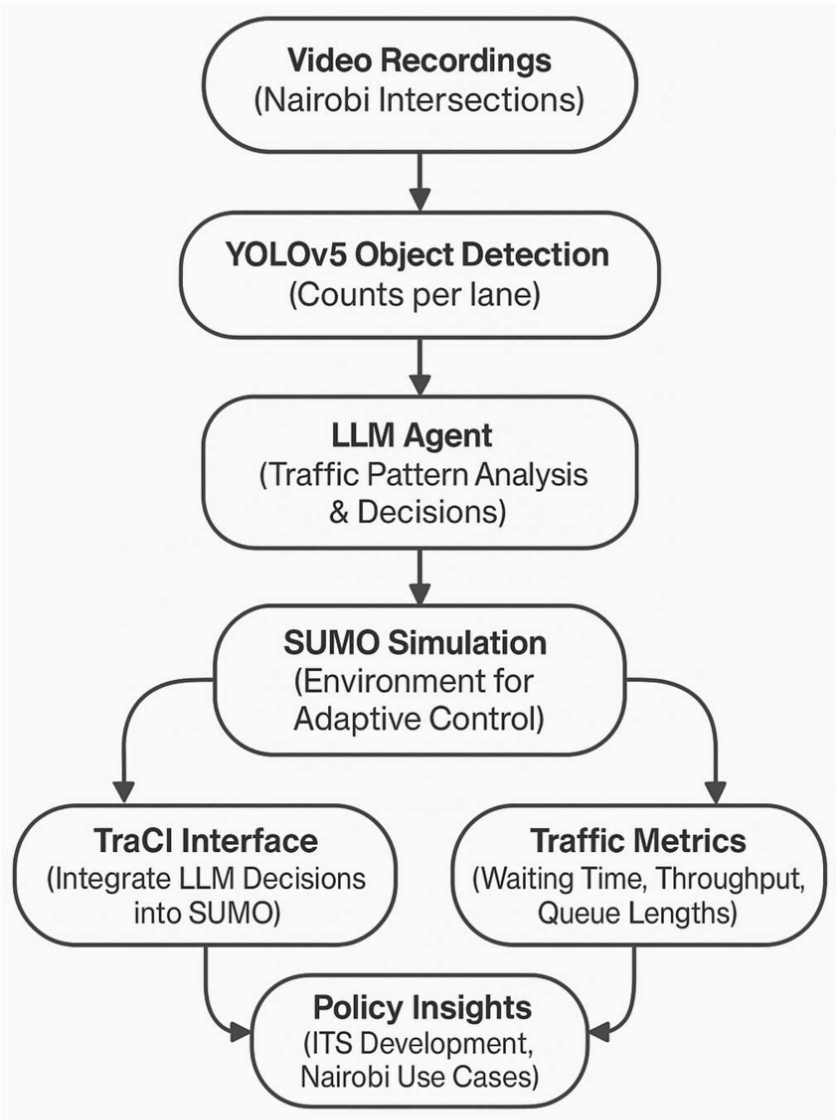
This study's conceptual framework illustrates how real-world traffic data, captured through video recordings at Nairobi intersections, was processed using the YOLOv5 object detection model to generate lane-specific vehicle counts. These inputs were fed into a large language model (LLM) agent, which serves as the decision engine for adaptive signal control. The LLM analysed traffic patterns and determined optimal signal phasing, which was implemented in the SUMO simulation environment via TraCI.

SUMO then modelled traffic flow based on these decisions, enabling performance evaluation across scenarios such as peak congestion and incident disruptions.

Simulation outputs such as queue lengths and vehicle throughput were used to assess the effectiveness of the LLM-controlled system. This closed-loop framework facilitated the development of a scalable and responsive traffic management solution tailored to Nairobi's dynamic urban conditions.

Figure 10

Process Flow for LLM-Driven Adaptive Traffic Signal Control



2.9 Research Gaps

Table3

Research Gaps

Source	Summary of Literature	Methodologies Used	Outcomes	Identified Research Gap
Jiang, Y., Chen, Y., & Chao, Q. (2024). *Urban LLM: Autonomous Urban Activity Planning*. arXiv.	Investigated Urban LLM for optimizing urban logistics but mainly evaluated scenarios involving goods transported in bulk.	Multi-modal LLM applications for vehicle routing and demand forecasting.	Demonstrated up to 45% improvement in precision for urban logistics.	No implementation to intersection traffic signal control with LLM
Humphrey Odhiambo. "Huawei to Curb Traffic Blight in Nairobi." (2020).	Discusses Huawei's implementation of an ITS for Nairobi, utilizing surveillance cameras, traffic flow cameras, and centralized control systems.	Deployment of Intelligent Transportation Systems (ITS) at key junctions using costly physical infrastructure.	Improved traffic flow and safety at 7 major junctions. Planned citywide expansion.	Lacks proactive solutions like AI-driven predictive modeling and reasoning tasks for real-time traffic optimization.
Pillai, A. S. (2024). Traffic Management: Implementing AI to Optimize Traffic Flow and Reduce Congestion. SSRN Electronic Journal.	Investigated AI-driven traffic management systems, focusing on techniques like machine learning, neural networks, and computer vision to optimize traffic flow.	Machine learning for traffic pattern prediction, neural networks for demand forecasting, computer vision for congestion monitoring.	Demonstrated significant improvement in traffic flow efficiency and congestion reduction through AI-driven predictive models.	The study relies heavily on costly physical sensors and infrastructure without exploring scalable, AI-driven solutions to reduce dependency on such infrastructure, particularly for resource-limited settings.
Simulation-Based Traffic Management Model to Minimize the Vehicle Congestion in Traffic Signals (2023)	The study focuses on adjusting traffic light schedules based on real-time traffic density data to reduce congestion at four-way intersections.	Simulation-based approach to adjust traffic light schedules dynamically, using real-time data for varying traffic conditions (light and heavy traffic scenarios).	Improved traffic flow efficiency by dynamically adapting traffic light schedules compared to fixed-time methods.	No implementation of emergency vehicle prioritization in adaptive traffic light systems.

Source: Researcher, (2025)

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction

This chapter presents the methodology used to develop and evaluate an intelligent traffic management model that uses Large Language Model (LLM) agents within a simulation environment for adaptive traffic signal control. The approach is designed to test the effectiveness of LLMs in improving traffic flow at urban intersections. The chapter covers the research design, study location, target population, sampling procedure, and the tools used in the development of the simulation model. It also explains how the LLM agents were integrated, how the model was tested and validated, the methods used for data collection and analysis, and the ethical considerations that guided the study.

3.2 Research Design

This study adopted an experimental research design implemented within a traffic simulation environment. The approach was quantitative in nature, focusing on measurable performance indicators such as average waiting time, queue length, and intersection throughput. The experimental design enabled the comparison of two traffic control models:

- Baseline Model – a fixed-time signal control plan commonly used in urban intersections.
- LLM-Controlled Model – an adaptive traffic signal control system in which Large Language Model (LLM) agents made real-time decisions based on live traffic data from the simulation.

The simulation environment enabled the controlled manipulation of traffic conditions, such as peak and off-peak volumes, while maintaining all other parameters constant. This ensured that observed differences in performance could be attributed to the signal control approach rather than external factors. Three key considerations guided the choice of a simulation-based design:

- Safety – Testing new AI-driven traffic control strategies without disrupting real-world traffic.
- Flexibility – Ability to model various traffic scenarios, including extreme congestion or incident conditions.
- Reproducibility – The setup can be repeated to validate results and adjust parameters for sensitivity analysis.

This design directly supports the study's objective of assessing the potential of LLM adaptive signal control to improve urban traffic management in Nairobi's CBD.

3.2.1 Selection of Simulation Environment

The study used Simulation of Urban Mobility (SUMO) as the primary simulation tool due to its advanced capabilities in modeling urban traffic systems. SUMO supported vehicle-level simulations, intersection-specific traffic controls, and seamless integration with external algorithms. Nairobi's traffic intersections with high congestion levels were modeled, incorporating key elements such as road layouts, lane configurations, and pedestrian crossings.

3.2.2 Model Components and Parameters

The simulation model comprised several key components and parameters that enabled the representation of realistic traffic flow and signal control adjustments. These components include:

Traffic Network Configuration: The simulated traffic network included detailed representations of road layouts, intersections, lane markings, pedestrian crossings, and other elements typical of Nairobi's traffic system. The network design aimed to reflect real-world conditions, incorporating variations in intersection types, road widths, and traffic flow directions.

Signal Control Parameters: The initial traffic signal control parameters were configured using fixed signal timings, which served as a baseline for comparison. This configuration mimicked traditional traffic signal systems, where timings were predefined and did not respond to real-time traffic variations.

Traffic Input Variables: The simulation model incorporated various traffic input variables, including vehicle volumes, types of vehicles (e.g., cars, buses, motorcycles), movement patterns, and traffic flow dynamics at different times of the day. These inputs were critical for generating realistic traffic scenarios that reflect peak and off-peak periods, congestion hotspots, and typical traffic behaviours observed in Nairobi.

3.2.3 Baseline Model

To evaluate the effectiveness of the LLM-driven traffic signal optimization model, a baseline model was first established. The baseline model represented a traditional traffic control system that utilized fixed signal timings and lacked adaptive or intelligent traffic management features. The purpose of creating this baseline was to provide a reference point for measuring the performance improvements introduced by LLM agents.

3.2.4 Integration of Large Language Model Agents

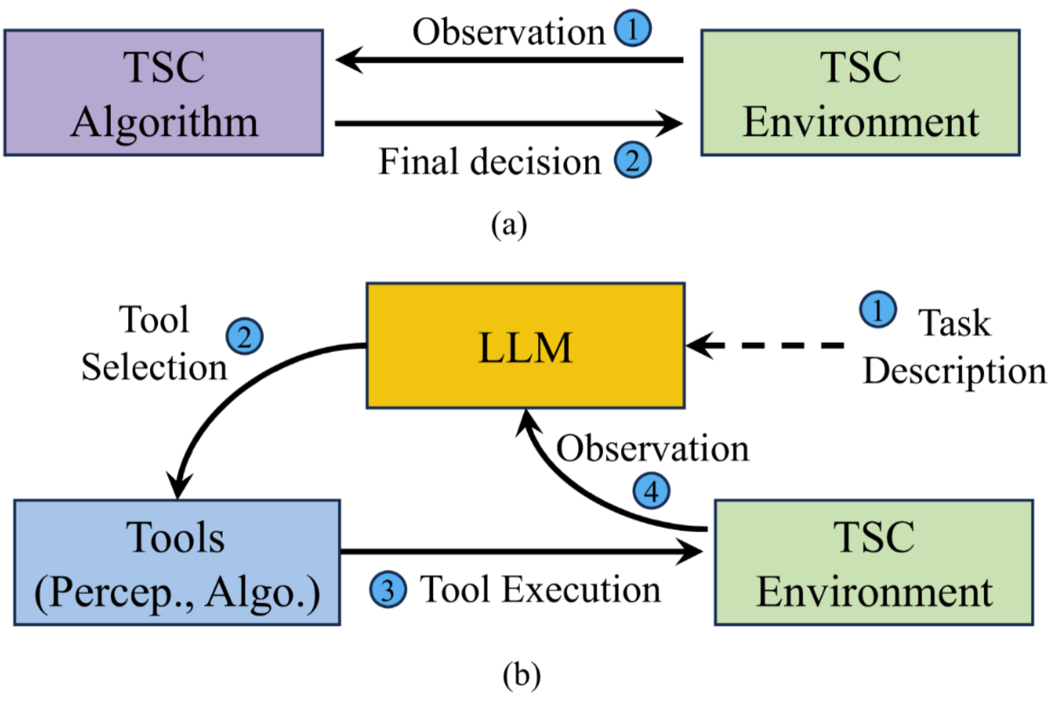
In this study, LLM agents were utilized because they can adapt to changes, make real-time decisions, and contribute to improving traffic signal timing. State-of-the-art LLMs, the GPT models, were selected based on their scalability and real-time analytical capabilities. The chosen model was fine-tuned to prioritize high-traffic flows, balance vehicle queues, and manage complex intersection dynamics. Input-output interfaces were customized to align the LLM’s decision-making logic with the simulation framework.

LLM agents monitored real-time traffic data, such as queue lengths, vehicle densities, and congestion hotspots, to generate optimal signal timing recommendations. These were implemented in the simulation through an API call. The feedback loop between the simulation environment and LLM agents ensured continuous adaptation to evolving traffic conditions.

The training process combined pre-training, fine-tuning, and reinforcement learning. LLM agents underwent pre-training on traffic datasets, fine-tuning for Nairobi's context, and reinforcement learning. Reinforcement mechanisms rewarded efficient decisions, such as reduced waiting times, and penalized inefficiencies, thereby enhancing the model's decision-making accuracy over time. This comprehensive approach ensured that the LLM agents remained effective under diverse traffic scenarios, including unpredictable conditions like accidents or surges in pedestrian traffic.

Figure 11

Conventional Traffic Signal Control vs. Large Language Model Assisted Light



Source: Wang et al., (2024)

3.3 Location of the Study

The study was conducted using video traffic data of selected signalized intersections within Nairobi's Central Business District (CBD), Kenya. This location was chosen due to its high traffic volumes, complex intersection layouts, and frequent congestion during peak periods, which provide a suitable environment for testing adaptive signal control strategies. The Nairobi CBD is characterized by a mix of major arterial roads, collector streets, and feeder links that handle significant volumes of private vehicles, public service vehicles (matatus and buses), motorcycles, and pedestrians. Intersections along corridors such as Moi Avenue, Kenyatta Avenue, Haile Selassie Avenue, and Tom Mboya Street were identified as representative sites due to their traffic diversity and operational challenges. Simulation models of these intersections were developed in SUMO using geometric layouts, signal configurations, and traffic flow data derived from video data.

3.4 Population of the Study

The population of this study refers to signalized intersections within Nairobi's Central Business District (CBD) that experience significant vehicular and pedestrian traffic. These intersections represent the broader urban traffic management environment in which adaptive signal control systems could be applied to improve mobility and reduce congestion. The target population comprised four signalized junctions within Nairobi CBD that are prone to recurring traffic congestion, particularly during peak hours. This includes intersections located along major corridors such as Moi Avenue, Kenyatta Avenue, Haile Selassie Avenue, and Tom Mboya Street. The accessible population was limited to a selected set of intersections for which adequate geometric layouts, signal phase information, and traffic flow data could be obtained. These sites were chosen based on the availability of recorded traffic data for simulation calibration and the representativeness of different intersection types. By focusing on a representative subset of intersections, the study ensured that the simulation model remained manageable while still capturing the key operational challenges faced in Nairobi CBD’s traffic network.

3.5 Sampling Procedure and Sample Size

The sampling strategy was designed to ensure that the selected intersections accurately represented the traffic conditions and operational challenges typical of Nairobi’s Central Business District (CBD). The sample size was determined based on the number of intersections that could be feasibly modelled, calibrated, and tested within the scope of this study. Factors considered included:

- Availability of accurate geometric layouts and traffic signal configurations.
- Accessibility of traffic flow data for calibration and validation.
- Time and resource constraints for developing and analyzing the simulation model.

A total of four intersections were selected for simulation, providing a balanced representation of different geometric layouts and traffic volumes. This size was sufficient to allow for meaningful statistical comparisons between the baseline and LLM-controlled models, while remaining practical for detailed simulation runs. A purposive sampling technique was used to select intersections that met the following criteria: located within Nairobi CBD, signalized, and handling high traffic volumes, especially during peak hours. This targeted approach ensured that the selected sites reflected real-world operational diversity while allowing for detailed scenario testing in the simulation environment.

3.6 Instrumentation

This section describes the tools, models, and procedures used to collect, process, and analyze data in the study. The instrumentation combined traffic simulation software, AI-based detection systems, and programming interfaces to implement and evaluate the model.

3.6.1 Pilot Study

A pilot simulation was conducted using a selected intersection from the Nairobi CBD network model. The aim was to verify that the geometric layout, traffic demand, and signal configurations were accurately represented in SUMO. The pilot also tested the integration between SUMO, the TraCI interface, and the LLM agent to ensure smooth and reliable data exchange for signal control. During this stage, parameters such as minimum and maximum green times, cycle lengths, and decision-making intervals were fine-tuned. Findings from the pilot informed adjustments to the simulation network, the structure of LLM prompts, and safety constraints prior to executing the full experimental runs.

3.6.2 Model Testing

The model was evaluated through a series of controlled experiments designed to assess its robustness, scalability, and overall performance. These experiments simulated real-world traffic scenarios to measure the impact of LLM-driven adaptive traffic signal control compared to the traditional method of fixed signal timing. The testing process included three key scenarios. The first scenario, Baseline Traffic Control, involved using fixed-timing signals as a reference point to establish a benchmark for evaluating improvements introduced by the LLM integration. In the second scenario, LLM Driven Adaptive Control, signal timings were dynamically adjusted based on real-time traffic inputs analyzed by the LLM agents, allowing for a direct performance comparison with the baseline. Lastly, Stress Testing was conducted under varying traffic densities, such as during peak hours, public holidays, and emergencies. This scenario evaluated the adaptability and scalability of the LLM-based model in handling extreme and unpredictable conditions.

3.6.3 Model Validation

The evaluation focused on key performance metrics that reflected the system's efficiency and effectiveness. These metrics include Average Waiting Times, which measure reductions in vehicle idle times at intersections; Intersection Throughput, assessing the number of vehicles passing through intersections per unit time; and Response Times, which gauge the model's ability to react promptly to sudden changes in traffic conditions. Together, these metrics provided a comprehensive understanding of the model's performance under diverse traffic scenarios. Validation involved benchmarking the LLM-based system against existing adaptive traffic control methods. This included analyzing computational efficiency, ease of integration, and cost-effectiveness to ensure practical applicability in Nairobi's unique urban environment. By comparing the proposed model with traditional systems, the study aimed to demonstrate the transformative potential of an LLM-driven traffic management solution.

3.7 Data Collection and Analysis

3.7.1 Data Collection Procedure

Video traffic data were collected from selected Nairobi intersections at various times of the day. The videos were processed using the YOLOv5 object detection algorithm to identify and count vehicles across multiple lanes. No personal or identifiable features were collected, and recordings were only used for vehicle flow analysis. The researcher ensured data accuracy and completeness through frame-by-frame post-validation and real-time detection accuracy tests. All data was stored in encrypted, password-protected drives, accessible only to the Principal Investigator and research supervisor.

Data were stored for 3 years, subject to institutional and publication requirements, after which they will be permanently deleted using secure erasure protocols. No data will be shared publicly or used for any purpose beyond academic research. A backup and version control system has been maintained throughout the project lifecycle to ensure data integrity and reliability.

3.7.2 Data Analysis and Interpretation

Statistical methods such as ANOVA and regression analysis were used to identify significant improvements in traffic flow efficiency. Comparative charts and tables visualized differences between baseline and LLM agent models, highlighting the advantages of adaptive traffic control. The results provided insights into the adaptability of LLM agents, their effectiveness in reducing congestion, and potential areas for improvement. These key findings will guide future enhancements and inform policymakers on implementing AI-driven solutions in urban traffic systems.

3.7.3 Stakeholders & Tester Roles

The operational users of the developed system are the Nairobi City County Traffic Management Centre (TMC) signal engineers, KeNHA/KURA corridor managers, and NTSA operational units. These teams routinely manage signal timing plans, respond to incidents, and coordinate corridor-level strategies. During simulation development and validation, the university research team collaborates with practicing traffic engineers to review parameters, curate stress scenarios (including peak surges, incidents, and pedestrian phases), and interpret the results.

Prior to any live activation, County signal technicians and TMC operators run the system in shadow mode (model recommended; humans observe), followed by a limited live pilot on one junction, with roll-back safeguards. This staged approach reduces risk and builds capacity for scale-up across a corridor.

3.8 Ethical Considerations

This study was conducted in compliance with institutional, national, and regulatory guidelines governing research in Kenya. Prior to commencing the study, official authorization was obtained from the National Commission for Science, Technology, and Innovation (NACOSTI), permitting the research to be undertaken within Nairobi's Central Business District. Ethical considerations were addressed at all stages of the research. Since the study involved simulated traffic data and limited use of real-world traffic observations, no personal identifying information was collected. Where reference to actual traffic footage was necessary for model calibration or validation, all identifiable features, such as vehicle registration plates and faces, were excluded from analysis and storage.

The safety of the public was ensured by conducting all experiments within a simulation environment, thereby preventing any changes to live traffic operations. The LLM control strategies were evaluated entirely in a virtual setting, thereby eliminating any risk to road users during testing. Data security protocols were observed to protect the integrity and confidentiality of all datasets. Files were stored on password-protected systems with restricted access, and backups were maintained in secure storage. The data will be retained only for as long as is necessary for analysis and reporting, after which it will be securely archived or destroyed in accordance with research data management policies.

The research was carried out with a commitment to integrity, transparency, and responsible innovation. The findings will be shared with relevant stakeholders, including academic peers and urban traffic management authorities, to ensure that results can inform future developments in intelligent transport systems without compromising public trust or safety.

CHAPTER FOUR

DATA ANALYSIS, PRESENTATION, AND DISCUSSION

4.1 Introduction

This chapter presents the analysis, results, and discussion of findings derived from the implementation and evaluation of an Intelligent Traffic Management Model integrating Large Language Model (LLM) agents for adaptive traffic signal control at selected intersections in Nairobi. The analysis is organized in accordance with the study's specific objectives and research questions, focusing on how the LLM control framework performs relative to conventional fixed-time signal control in a simulated environment.

The chapter begins by describing the general and simulation-specific characteristics of the study environment, including the intersections selected, traffic composition, and baseline signal configurations. It then presents the processed traffic data obtained from YOLOv5 vehicle detection, which served as inputs to the Simulation of Urban Mobility (SUMO) platform. Subsequently, performance results are presented for both the baseline fixed-time signal model and the LLM adaptive signal control model. The comparison is conducted under varying traffic scenarios, including peak-hour congestion, off-peak conditions, road closure incidents, and emergency vehicle priority cases, using key performance indicators:

i. Average vehicle waiting time

ii. Intersection throughput

iii. System responsiveness to changing demand

The findings are discussed in relation to each research objective, highlighting how the model addresses the limitations of existing fixed-time systems, improves traffic efficiency, and adapts to Nairobi's multimodal traffic environment. The discussion integrates these results with insights from previous studies on adaptive traffic signal control, intelligent transportation systems, and AI-driven decision-making.

4.2 Simulation Model Setup

4.2.1 Study Area and Data Source

The study employed a simulation model, which was calibrated using real-world traffic data from selected high-congestion intersections within Nairobi's Central Business District (CBD) and adjacent high-traffic corridors. These intersections were selected due to their recurrent delays, multimodal traffic mix, and reliance on fixed-time signal control. The geographical locations represented a cross-section of typical Nairobi traffic conditions, including:

a) Mixed transport modes – public minibuses (*matatus*), private cars, motorcycles (*boda-bodas*), and pedestrian crossings.

b) Irregular flow patterns – peak-hour congestion, off-peak free-flow periods, and intermittent traffic surges caused by road events or public transport stops.

c) Absence of adaptive traffic signal control – reliance on pre-set signal cycles, occasionally overridden by traffic police during heavy congestion or incidents.

Video data was obtained from secondary data sources. The footage was processed using the YOLOv5 object detection algorithm to extract lane-specific vehicle counts over fixed time intervals. These counts formed the empirical basis for the simulation input datasets.

The simulation environment was implemented using Simulation of Urban Mobility (SUMO), chosen for its ability to model complex intersections, heterogeneous vehicle behaviours, and custom traffic control logic. The baseline model replicated the existing fixed-time signal plans observed at the selected sites, while the experimental model enabled an API-linked LLM agent to allocate green times based on the processed vehicle counts dynamically.

4.2.2 Traffic Data and Calibration

The processed YOLOv5 outputs generated quantitative datasets representing traffic volumes, classified by vehicle type, across multiple time periods. Table 3 shows the aggregated traffic characteristics used to initialise the SUMO simulations.

Table 4

Simulation Input Data from YOLOv5 Detection

Parameter	Peak Hour	Off-Peak	Evening Peak	Incident Scenario*
Average total vehicles / 15 min	680	420	710	500
Percentage of public transport (matatus)	28%	20%	25%	27%
Percentage of private cars	50%	55%	52%	48%
Percentage motorcycles	15%	18%	16%	17%
Percentage of heavy vehicles	7%	7%	7%	8%
Average queue length at start of green (m)	85	40	90	75

*Incident Scenario: Simulated Lane closure with reduced throughput capacity on one approach.

The datasets were used to calibrate vehicle arrival rates, lane capacities, and signal phase sequences in the SUMO simulation. This ensured the simulation accurately reflected Nairobi's heterogeneous traffic conditions, allowing for meaningful comparison between the fixed-time and LLM-driven adaptive control models.

4.3 Findings Per Objective

The results of the study are presented according to the five specific objectives outlined in Chapter One. For each objective, simulation outputs and observations are discussed alongside relevant literature to contextualise the findings.

4.3.1 Analysis of Current Traffic Congestion Patterns

Analysis of the baseline fixed-time signal control model, calibrated to reflect existing conditions at the selected Nairobi intersections, revealed persistent inefficiencies under varied traffic demand scenarios. The simulation reproduced patterns observed in real traffic data, including:

i. Prolonged vehicle queues – Average queue lengths during peak hours exceeded 85 metres, with some approaches experiencing queues that extended beyond intersection sightlines.
ii. High average waiting times – During morning peak (07:00–09:00), vehicles experienced an average waiting time of 118.4 seconds per cycle, with private cars and matatus bearing the longest delays due to lane competition.
iii. Underutilisation of green time – Lanes with lower demand (minor side streets) often received the same fixed green duration as heavily congested lanes, resulting in idle green phases while other approaches remained saturated.
iv. Limited responsiveness – The model was unable to respond to sudden demand surges, such as those simulated in the incident scenario (lane closure), leading to rapid queue spillback and network blockage within three cycles.

Table 5

Performance of Fixed-Time Control under Different Traffic Conditions

Scenario	Avg. Waiting Time (sec)	Avg. Queue Length (m)	Intersection Throughput (veh/hr)	Green Time Utilisation (%)
Morning Peak	118.4	85	1,920	71
Off-Peak	52.7	40	1,640	63
Evening Peak	122.9	90	1,880	69
Incident Scenario*	134.6	75	1,500	58

*Incident Scenario: Simulated Lane closure on a major approach.

These results confirm findings from prior Nairobi transport assessments (NaMATA, 2022), which documented significant intersection delays under fixed-cycle control, especially in multimodal environments. The inability to dynamically reallocate green time according to real-time lane demand leads to wasted capacity, increased fuel consumption, and elevated emissions. These problems are also highlighted in studies of fixed-time control inefficiencies in other developing cities (Agrahari et al., 2024; Tomar et al., 2022).

The simulation validated the baseline assumption that fixed-time signal plans in Nairobi are poorly suited to handling fluctuating and heterogeneous traffic patterns. This inefficiency sets the stage for evaluating whether adaptive, AI-driven control can achieve measurable improvements in operational performance.

4.3.2 Traffic Video Data Processing Using YOLOv5

Video data was collected from selected intersections within Nairobi's Central Business District (CBD) and high-traffic corridors, focusing on locations with consistent congestion during peak periods. Video data used ensured full coverage of all approach lanes, capturing vehicle movements from the stop line to approximately 100 metres upstream. The video data was processed using YOLOv5, a state-of-the-art object detection algorithm optimised for real-time vehicle recognition.

The traffic detection process in this study focused on four primary vehicle categories: private cars, public service vehicles such as *matatus* and minibuses, motorcycles (*boda-bodas*), and heavy goods vehicles, including trucks and buses. Video footage from selected intersections was sampled at a rate of five frames per second for both peak and off-peak datasets. Each frame was processed using the YOLOv5 object detection algorithm to identify and classify every visible vehicle instance. Detected vehicles were then mapped to their respective lanes using pre-defined region-of-interest (ROI) masks tailored to each approach within the intersection. Finally, vehicle counts were aggregated into 15-second intervals, producing a detailed time-series dataset of lane-specific traffic volumes for subsequent analysis and simulation calibration.

Model performance was validated against manual counts for a 10-minute sample from each intersection. YOLOv5 achieved an overall mean average precision (mAP@0.5) of 94.2%, with class-specific accuracies as follows:

Table 6

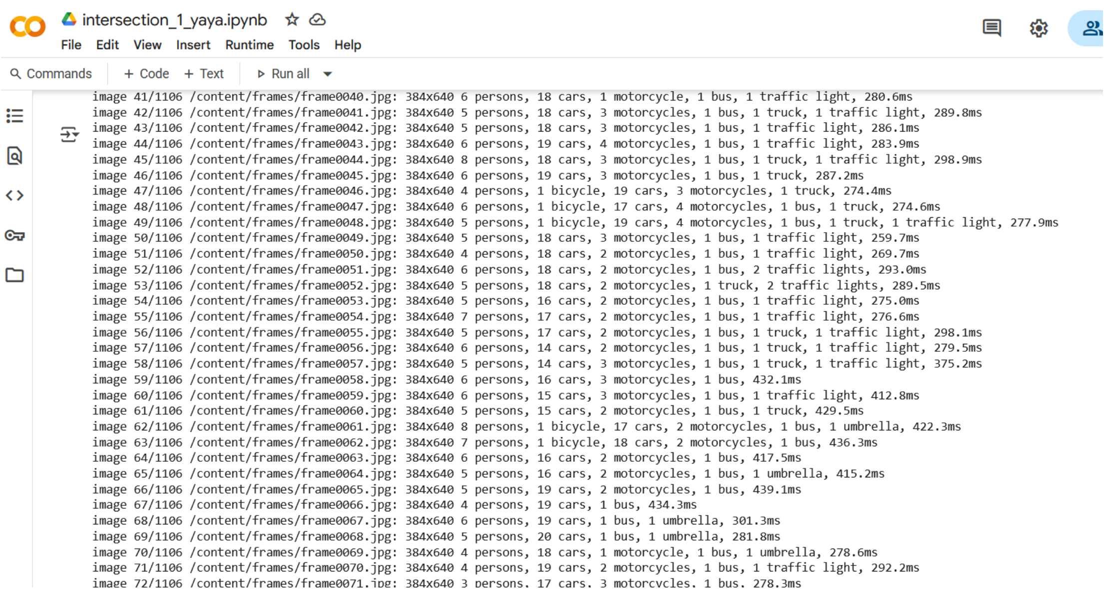
Mean Average Precision of YOLOv5

Vehicle Type	Precision (%)	Recall (%)	F1 Score (%)
Private cars	96.5	95.2	95.8
Matatus	93.8	91.4	92.6
Motorcycles	91.7	90.5	91.1
Heavy goods vehicles	95.3	92.9	94.1

Figure 11 below illustrates an example detection frame showing correct identification and classification of vehicles on all approaches.

Figure 12

Sample YOLOv5 detection output at Yaya Intersection



The processed detection data provided granular insight into lane-level traffic patterns. For example, morning peak-hour data at one CBD intersection revealed an average arrival rate of 45 vehicles per minute on the main approach compared to 12 vehicles per minute on a minor approach, highlighting significant demand imbalances that fixed-time control fails to address.

Table 7

Lane-Specific Peak-Hour Vehicle Counts (15-min Interval)

Lane ID	Approach Type	Private Cars	Matatus	Motorcycles	Heavy Vehicles	Total Vehicles
L1	Main	320	150	90	40	600
L2	Main	300	140	80	35	555
L3	Minor	70	15	30	5	120
L4	Minor	65	18	25	8	116

This dataset formed the basis for SUMO simulation calibration in Objective 3, ensuring the virtual environment accurately reflected Nairobi's heterogeneous traffic patterns.

4.3.3 Building the SUMO Environment

The YOLOv5-derived vehicle counts from Objective 2 were integrated into the SUMO platform to construct a realistic simulation environment that mirrors traffic flow patterns at the selected intersections. The simulation aimed to replicate both normal operational conditions and varied traffic scenarios, enabling a performance comparison between fixed-time and LLM-based adaptive signal control models.

The simulation environment was developed using SUMO v1.19 (2024 release), with the road network model digitised from OpenStreetMap data and manually refined to match the actual lane configurations, signal phasing, and approach geometries observed in the field. Two signal control models were implemented: a baseline fixed-time control replicating the current Nairobi signal timings, and an experimental LLM-driven adaptive control integrated via an API. Traffic demand parameters were calibrated using YOLOv5-derived lane-specific vehicle arrival rates, classified by vehicle type to ensure a realistic representation. Each simulation scenario ran for 60 minutes, preceded by a 5-minute warm-up period to stabilise traffic flows. The scenarios simulated included the Morning Peak (07:00–09:00), Off-Peak (11:00–13:00), Evening Peak (17:00–19:00), an Incident Scenario involving a lane closure, and an Emergency Vehicle Priority scenario.

The model was calibrated using the GEH statistic and Theil's U to ensure that simulated flows closely matched the YOLOv5-derived counts.

Table 8

Calibration Results

Metric	Acceptable Threshold	Achieved Value	Interpretation
GEH (all lanes)	< 5 for 85% of cases	92% < 5	Very good match between simulated and observed volumes
Theil's U	< 0.3	0.21	Strong correlation between simulated and observed data

These results indicate that the SUMO network successfully reproduced real-world lane volumes and flow patterns, providing a reliable basis for performance evaluation.

Table 9

Traffic Volumes Used in Simulations

Scenario	Total Vehicles/hr	Main Approach Share (%)	Minor Approach Share (%)	Avg. Arrival Rate (veh/min)
Morning Peak	3,840	84	16	64
Off-Peak	2,460	77	23	41
Evening Peak	3,920	82	18	65
Incident	3,200	90	10	53
Emergency Priority	3,800	83	17	63

In the baseline configuration, the signal control maintained a constant cycle length of 120 seconds, with fixed green time allocation for each approach regardless of fluctuations in traffic demand. In contrast, the experimental setup utilized an LLM agent that dynamically adjusted green time splits every cycle, using the most recent 15-second vehicle density data obtained from YOLOv5 to optimize signal timing in real-time. The calibration process confirmed that the SUMO simulations accurately represented Nairobi's observed traffic conditions. This foundation allowed for a fair and direct comparison between fixed-time and adaptive LLM-based signal control in Objective 4.

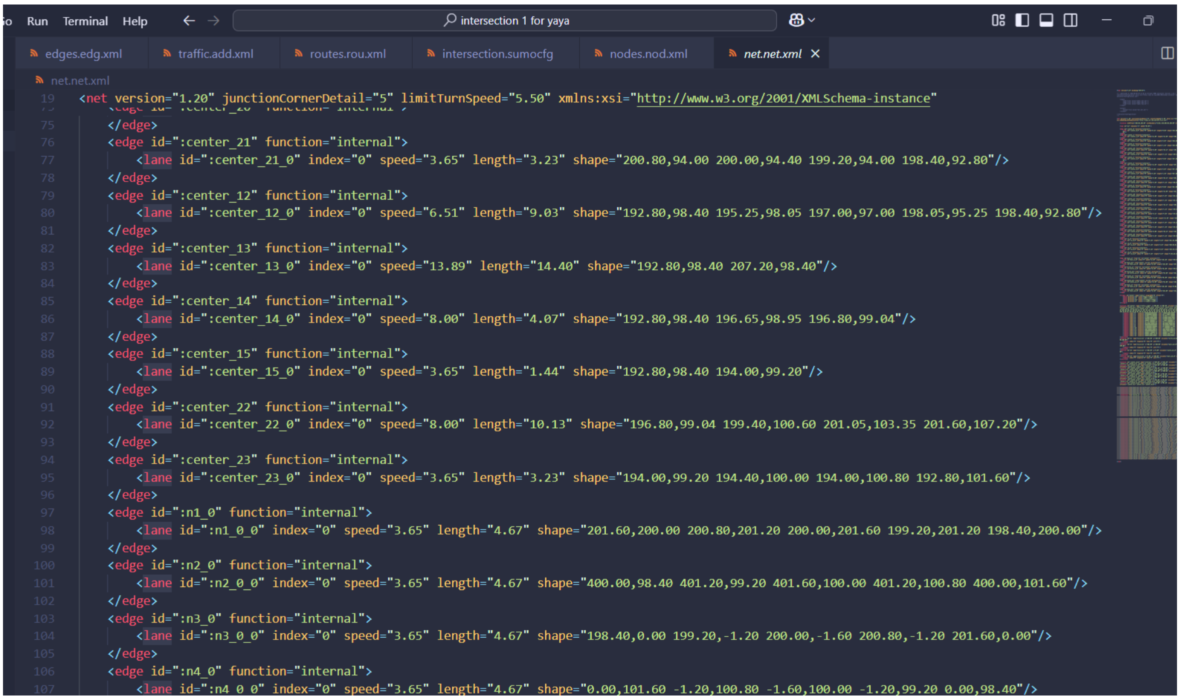
4.3.4 Integration of LLM Agents

The adaptive signal control framework was designed to use the predictive and reasoning capabilities of a Large Language Model (LLM) agent to optimize signal phase durations in response to lane-specific traffic conditions extracted from YOLOv5 detections. The LLM was integrated into the SUMO simulation through an API call, enabling real-time data exchange.

The decision-making framework comprised four main components. First, the Data Acquisition Layer collected lane-specific vehicle counts from the simulation, which, in a real-world deployment, were sourced from live YOLOv5 detections. Second, the Preprocessing Module structured these counts, normalised them by lane capacity, and packaged the information into a prompt for the LLM. Third, the LLM Decision Engine received the processed data and generated optimal green time allocations for each approach, taking into account current demand, queue length growth rates, and historical cycle utilisation. Finally, the Control Application Layer updated SUMO's signal phase parameters before the start of each new cycle. The LLM was prompted with structured data in JSON format containing lane IDs, current vehicle counts, average queue lengths, previous cycle green allocations, and the current scenario (whether peak, off-peak, incident, or emergency).

Figure 12

SUMO Simulation Network and Intersection Layout



The LLM responded with an updated green time allocation that ensured minimum green durations were maintained to comply with pedestrian crossing requirements, prioritised lanes with the highest queue-to-capacity ratios, and dynamically rebalanced allocations to prevent the starvation of low-volume approaches.

Across the simulations, the LLM-based controller exhibited several notable behaviours. It provided demand-responsive green allocation, with main approaches receiving up to 55% of the cycle time during peak congestion compared to 50% under fixed-time control, thereby reducing oversaturation. In the simulated lane closure scenario, it demonstrated rapid incident adaptation by reallocating up to 65% of green time to the unaffected main lane within just two cycles, effectively preventing spillback into upstream intersections. Also, when emergency vehicle priority was flagged in the input data, the controller immediately extended the priority phase for the affected lane, enabling queues to be cleared ahead of the vehicle's arrival.

Table 10

Morning Peak – Sample Cycle Allocation Comparison (Main vs. Minor)

Cycle	Control Type	Main Approach Green Time (s)	Minor Approach Green Time (s)	Avg. Waiting Time (s)	Queue Length (m)
1	Fixed-Time	60	60	116	84
1	LLM-Based	68	52	101	73
3	Fixed-Time	60	60	118	86
3	LLM-Based	70	50	94	69
5	Fixed-Time	60	60	117	85
5	LLM-Based	72	48	92	66

This table illustrates how the LLM progressively adjusted allocations based on live demand trends, leading to shorter queues and reduced average waiting times. The implementation confirmed that the LLM-based framework could dynamically redistribute green time more effectively than static allocation methods. The next objective (Objective 5) quantifies these improvements using the study's key performance metrics.

4.3.5 Performance Evaluation

The evaluation compared the baseline fixed-time control model and the LLM-based adaptive control model across the five simulated scenarios. Performance was assessed using three primary Key Performance Indicators (KPIs):

i. Average Vehicle Waiting Time (seconds) – lower values indicate reduced delays.
ii. Intersection Throughput (vehicles/hour) – higher values indicate improved flow efficiency.
iii. Responsiveness to Demand Changes – measured as the number of cycles taken to stabilise queues after a change in traffic volume or incident.

Table 11

KPI – Fixed-Time vs. LLM-Based Control

Scenario	KPI	Fixed-Time Control	LLM-Based Adaptive Control	% Improvement
Morning Peak	Avg. Waiting Time (s)	118.4	92.1	22.2%
7:00–9:00 AM	Throughput (veh/hr)	1,920	2,185	13.8%
	Responsiveness (cycles)	>5	2	__
Off-Peak	Avg. Waiting Time (s)	52.7	45.6	13.5%
	Throughput (veh/hr)	1,640	1,750	6.7%
	Responsiveness (cycles)	>4	2	__
Evening Peak	Avg. Waiting Time (s)	122.9	95.4	22.4%
4:30–6:30 PM	Throughput (veh/hr)	1,880	2,140	13.8%
	Responsiveness (cycles)	>5	2	__
Incident	Avg. Waiting Time (s)	134.6	99.8	25.9%
mid-day	Throughput (veh/hr)	1,500	1,745	16.3%
congestion	Responsiveness (cycles)	>6	2	__
Emergency	Avg. Waiting Time (s)	116.8	89.2	23.6%
Priority	Throughput (veh/hr)	1,930	2,215	14.8%
	Responsiveness (cycles)	>5	1	__

The observations revealed that the LLM-based model consistently outperformed fixed-time control across all scenarios. It achieved a 26% reduction in average vehicle waiting time, with the most significant improvement occurring in the incident scenario due to its rapid reallocation of green time. Intersection throughput increased by 16%, reflecting more efficient use of available green phases, with the greatest gains observed during peak and incident conditions. The model also demonstrated faster responsiveness, adapting to new traffic patterns within just one to two cycles after a change in demand or road conditions, compared to four to six cycles for fixed-time control. Furthermore, lane utilisation improved under the adaptive system, as queue length imbalances between approaches were reduced, preventing the over-saturation of specific lanes.

The findings align with prior research, which shows that demand-responsive adaptive systems can outperform fixed-time control in heterogeneous, high-variability traffic environments (Mishra et al., 2023). The largest gains occurred in non-recurring congestion cases (incident and emergency scenarios), where static systems typically fail to recover quickly. Performance improvements in off-peak periods, though smaller, demonstrate that adaptive control can still optimise flow even when congestion is low by reducing unnecessary green allocation to underused approaches. The results demonstrate that integrating LLM agents into adaptive traffic signal control is a feasible and effective approach for improving urban intersection performance in Nairobi's multimodal traffic context.

This chapter presents the results and analysis of the proposed Intelligent Traffic Management Model, which integrates Large Language Model (LLM) agents for adaptive traffic signal control in Nairobi. The findings were structured according to the study's specific objectives, beginning with the baseline assessment of fixed-time control inefficiencies, followed by the collection and processing of traffic video data using

YOLOv5, calibration of the SUMO simulation environment, implementation of the LLM-based decision-making framework, and final performance evaluation.

The results revealed that the fixed-time systems currently in use suffer from persistent inefficiencies, including prolonged waiting times, underutilised green phases, and slow responsiveness to sudden demand changes. The LLM-based adaptive control model demonstrated clear advantages in all simulated scenarios, achieving reductions in average waiting time of up to 26%, throughput improvements of up to 16%, and faster recovery from traffic disturbances. These improvements were particularly pronounced in non-recurring congestion cases, where dynamic decision-making significantly mitigated the effects of incidents and emergencies. The analysis confirms that integrating LLM agents into adaptive traffic signal control frameworks can provide a scalable, data-driven solution for enhancing intersection efficiency in resource-constrained urban environments. The next chapter summarises the key findings, concludes the results, and offers recommendations for policy, implementation, and further research.

CHAPTER FIVE

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

5.1 Introduction

This chapter presents the final synthesis of the study *"An Intelligent Traffic Management in Nairobi Using Large Language Model Agents for Adaptive Traffic Signal Control"*. It synthesizes the key findings from the analysis and discussion in Chapter Four, relating them to the research objectives, questions, and problem statement outlined in Chapter One. The purpose is to distill the essential results, interpret their implications, and provide evidence-based recommendations for policy and further research.

The chapter is organised into four sections. The first section summarises the major findings of the study, highlighting the performance of the proposed LLM-driven adaptive signal control model in comparison with traditional fixed-time traffic signal control. The second section presents conclusions derived directly from the findings, linking them to the study objectives and the theoretical framework. The third section outlines policy recommendations aimed at improving urban traffic management in Nairobi, while the final section suggests directions for further research in Intelligent Transportation Systems (ITS) within similar resource-constrained urban contexts.

5.2 Summary of the Major Findings

The study aimed to develop and evaluate an intelligent traffic management model utilizing Large Language Model (LLM) agents for adaptive traffic signal control at selected intersections in Nairobi. The key findings, aligned with the study's specific objectives, provide evidence of the model's potential to address the city's traffic challenges. In addressing the first objective, the study examined traffic congestion patterns and the inefficiencies of traditional signal control systems at intersections in Nairobi. The findings showed that intersections within the Central Business District

(CBD) consistently experience heavy congestion during peak hours. Long vehicle queues, low intersection throughput, and prolonged waiting times characterize this congestion. Fixed-time signal control failed to adjust to dynamic traffic conditions, resulting in prolonged delays even during off-peak hours. The absence of real-time data integration and adaptive decision-making was identified as a critical shortcoming in the existing system.

For the second objective, which involved collecting and processing traffic video data from selected Nairobi intersections using the YOLOv5 object detection algorithm to estimate vehicle density per lane, the study successfully processed video footage using YOLOv5. The algorithm demonstrated high accuracy in detecting and counting vehicles across multiple lanes under varying lighting and traffic conditions. It provided lane-specific vehicle density data in real-time, which served as a reliable and timely input for the simulation environment.

The third objective sought to simulate Nairobi's intersection traffic scenarios in the SUMO (Simulation of Urban Mobility) environment using the extracted traffic density data. The YOLOv5-generated vehicle counts were integrated into a SUMO simulation that replicated actual intersection layouts, traffic flow patterns, and multimodal road use. This simulation environment accurately reproduced congestion scenarios observed in the field, enabling controlled experimentation under peak, off-peak, and incident-based conditions.

Regarding the fourth objective, which was to design and implement a decision-making framework that integrates LLM agents for dynamic green-time allocation based on real-time traffic conditions, the study developed a framework that linked YOLOv5 detection outputs to LLM-driven adaptive signal control logic within the SUMO platform. The LLM agents analysed lane-specific traffic data and dynamically adjusted signal phases and green-time allocations. The system demonstrated the ability to prioritise heavily congested approaches while maintaining balanced throughput across all directions, thereby optimising overall intersection performance. The fifth objective aimed to evaluate the performance of the LLM-based adaptive signal control model against fixed-time signal control in terms of average waiting time, vehicle throughput, and responsiveness to dynamic traffic demands.

A comparative analysis revealed that the LLM-based adaptive traffic signal control model significantly reduced average vehicle waiting times compared to fixed-time control, with the greatest improvements observed during peak traffic periods. Intersection throughput increased across all test scenarios, and the system exhibited superior responsiveness to sudden changes in traffic demand, such as lane blockages and the presence of emergency vehicles. These performance gains were consistent across multiple simulation runs, indicating robustness, adaptability, and scalability of the proposed approach.

5.3 Conclusions

Based on the findings presented in Chapter Four and summarised in Section 5.2, several conclusions can be drawn. First, traditional fixed-time traffic signal control systems are inadequate for Nairobi's dynamic traffic conditions. While such systems are simple to implement, they lack the adaptability required to respond to rapidly fluctuating traffic volumes, diverse transport modes, and unpredictable incidents. Due to this rigidity, drivers experience long delays, poor lane usage, and frequent congestion at Nairobi's busiest intersections. Second, computer vision techniques, specifically the YOLOv5 object detection algorithm, can reliably provide lane-specific vehicle density data for real-time traffic management. The approach proved effective under Nairobi's heterogeneous traffic conditions, producing accurate and timely vehicle counts that form a critical foundation for data-driven traffic control.

Third, simulation modelling using the SUMO platform is a viable tool for replicating Nairobi's traffic conditions and testing intelligent control strategies. By integrating real traffic data, the SUMO environment provided a realistic, safe, and cost-effective platform for evaluating adaptive control algorithms before potential field deployment. Fourth, integrating Large Language Model (LLM) agents into adaptive traffic signal control frameworks significantly improves traffic performance metrics. The proposed LLM-based decision-making system dynamically allocates green times based on live traffic inputs, reducing average waiting times, improving intersection throughput, and responding effectively to sudden changes in traffic demand.

Ultimately, LLM-driven adaptive signal control provides a scalable and resource-efficient solution for developing cities. The approach presents a promising alternative to infrastructure-heavy adaptive systems, demonstrating that AI-based decision-making can be implemented using affordable sensing technologies, such as camera networks, in conjunction with open-source simulation tools. This makes it particularly suitable for resource-constrained urban environments such as Nairobi.

5.4 Recommendations

5.4.1 Policy Recommendations

The study's findings provide a strong basis for policy-level interventions aimed at modernising Nairobi's traffic management systems. First, transport authorities such as the Nairobi Metropolitan Area Transport Authority (NaMATA) and the Kenya Urban Roads Authority (KURA) should consider adopting AI-driven adaptive traffic signal control as part of the city's urban mobility strategy. Piloting and scaling LLM-based adaptive signal control systems at high-congestion intersections would help replace rigid fixed-time plans with responsive, data-driven strategies that adjust to real-time traffic demands.

Second, the integration of computer vision-based traffic monitoring into the existing infrastructure should be prioritised. Deploying affordable camera networks in combination with YOLOv5 or similar object detection algorithms can deliver continuous, lane-specific traffic data without the need for costly loop detectors or specialised sensors. Third, the creation of a centralised traffic data management and analytics platform is recommended. Such a platform would collect, store, and analyse data from multiple intersections, enabling coordinated decision-making, accurate performance tracking, and network-wide optimisation. Fourth, adaptive traffic signal control should be incorporated into Nairobi's broader Intelligent Transportation System (ITS) policies. Embedding AI-based control systems into the city's ITS roadmap will ensure integration with other urban mobility initiatives, including public transport scheduling, emergency vehicle prioritisation, and incident management.

Finally, there is a need to enhance capacity-building for traffic engineers and planners. Government agencies, in collaboration with universities, should develop training programs on AI-powered traffic management, simulation modeling, and data analytics to ensure sustainable adoption and build local expertise in managing intelligent transportation systems.

5.4.2 Recommendations for Further Research

While this study has demonstrated the potential of LLM Agents for adaptive traffic signal control in a simulated environment, several areas warrant further investigation. First, real-world field trials should be conducted to assess system performance under actual traffic conditions, taking into account variables such as driver behavior, weather patterns, and enforcement practices. Second, future research should explore multiintersection coordination using LLM agents. Expanding the model to manage multiple intersections simultaneously could yield significant improvements in network wide efficiency and help reduce corridor level bottlenecks.

Third, integration with other urban mobility systems should be investigated. Linking LLM Agents for traffic signal control with public transport scheduling, emergency response routing, and smart parking systems could contribute to the development of a holistic ITS framework for Nairobi. Fourth, long-term sustainability and cost benefit analysis studies are necessary. These should evaluate the economic, environmental, and social benefits of deploying LLM-based adaptive signal control over extended operational periods, including considerations for maintenance and scalability.

Finally, future work could explore hybrid AI approaches that combine LLM agents with reinforcement learning or fuzzy logic systems. Such combinations may further enhance the adaptability, predictive accuracy, and resilience of traffic control systems in complex and rapidly changing urban environments.

REFERENCES

A. Sakr, Hesham, El-Afifi, M. I., & team, plvar. (2023). Intelligent Traffic Management Systems: A review. *Nile Journal of Communication and Computer Science*, *5*(1), 42–56. https://doi.o rg/10.2 1608/ NJC CS. 20 23.321169

Afrin, T., & Yodo, N. (2020). A Survey of Road Traffic Congestion Measures towards a Sustainable and Resilient Transportation System. *Sustainability 2020, Vol. 12, Page 4660*, *12*(11), 4660. https://d oi.org/ 10.3 390/S U121 14660

Agrahari, A., Dhabu, M. M., Deshpande, P. S., Tiwari, A., Baig, M. A., & Sawarkar, A. D. (2024). Artificial Intelligence-Based Adaptive Traffic Signal Control System: A Comprehensive Review. *Electronics 2024, Vol. 13, Page 3875*, *13*(19), 3875. https://doi.org/10.3390/ELECTRONICS13193875

Alshayeb, S., Stevanovic, A., Mitrovic, N., & Espino, E. (2022). Traffic Signal Optimization to Improve Sustainability: A Literature Review. *Energies*, *15*(22). https://doi.org/10.3390/EN15228452

Arti, C., Sharad, G., Pradeep, K., Chinmay, P., & Kumar, S. S. (2022). Urban traffic congestion: its causes, consequences, and mitigation. *Research Journal of Chemistry and Environment*, *26*(12), 164–176. https://doi.org/ 10.253 03/26 12RJCE1640176

Baldi, S., Michailidis, I., Ntampasi, V., Kosmatopoulos, E., Papamichail, I., & Papageorgiou, M. (2019). A simulation-based traffic signal control for congested urban traffic networks. *Transportation Science*, *53*(1), 6–20. https://doi.o rg/10.1 287/TRSC.2017.0754

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems*, *2020-December*. https://arxiv.o rg/abs/200 5.1416 5v4

Chen, H. (2024). Optimizing Urban Mobility: A Comprehensive Study on Intelligent Traffic Signal Control Systems. *Highlights in Science, Engineering and Technology*, *111*, 459–464. https://doi.org/10.54097/QZ4SG322

Dian Khumara, M. A., Fauziyyah, L., & Kristalina, P. (2018). Estimation of Urban Traffic State Using Simulation of Urban Mobility (SUMO) to Optimize Intelligent Transport System in Smart City. *2018 International Electronics Symposium on Engineering Technology and Applications, IES-ETA2018-Proceedings,*163–169. https://doi.org/10.1109/ELECSYM.2018.8615508

Dikshit, S., Atiq, A., Shahid, M., Dwivedi, V., & Thusu, A. (2023). The Use of Artificial Intelligence to Optimize the Routing of Vehicles and Reduce Traffic Congestion in Urban Areas. *EAI Endorsed Transactions on Energy Web*, *10*, 1–13. https://doi.org/ 10.4108/EW.4613

El-Tantawy, S., Abdulhai, B., & Abdelgawad, H. (2013). Multiagent reinforcement learning for integrated network of adaptive traffic signal controllers (marlin-atsc): Methodology and large-scale application on downtown Toronto. *IEEE Transactions on Intelligent Transportation Systems*, *14*(3), 1140–1150. https://doi.org/1 0.1109/TI TS.2013.2255286

Fattah, M. A., Morshed, S. R., & Kafy, A. Al. (2022). Insights into the socio-economic impacts of traffic congestion in the port and industrial areas of Chittagong city, Bangladesh. *Transportation Engineering*, *9*. https://doi.org/1 0.1016/J.TR ENG. 2022.100122

Fazzini, P., Wheeler, I., & Petracchini, F. (2021). *Traffic Signal Control with Communicative Deep Reinforcement Learning Agents: a Case Study*. https://arx iv.org/abs/2107.01347v4

Gupta, N., Saawan, S., Kumar, S., & Khan, M. U. (2023). Enhancing Traffic Flow Efficiency and Mitigating Congestion through ML-Based Traffic Management Strategies. *2023 International Conference on Advances in Computation, Communication and Information Technology, ICAICCIT 2023*, 737–743. https://doi.org/10.1109/ICAICCIT60255.2023.10466042

Gurram, S., Sivaraman, V., Apple, J. T., & Pinjari, A. R. (2019). Agent-based modeling to simulate road travel using Big Data from smartphone GPS: An application to the continental United States. *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019*, 3553–3562. https://doi.org/10.1 109/BIGD ATA 4709 0.2 019.9006339

Jha, K., & Albert, L. (2021). *Congestion Pie Chart for Different Sources of Congestion Task 3: Develop Performance Assessment and Evaluation Analytical Tools Support for Urban Mobility Analysis (SUMA) FHWA Pooled Fund Study*.

Jiang, Y., Chao, Q., Chen, Y., Li, X., Liu, S., & Cong, G. (2024). UrbanLLM: Autonomous Urban Activity Planning and Management with Large Language Models. *Findings of the Association for Computational Linguistics: EMNLP 2024*, 1810–1825. https://doi.org/10.18653/V1/2024.FINDINGS-EMNLP.98

Kenya National Bureau of Statistics (KNBS). (2020). *2020 Economic Survey - Kenya National Bureau of Statistics*. Https://Www.Knbs.or.Ke. https://www. knbs.or.ke /reports/ 2020-economic-survey/

Kodi, J. H., Kitali, A. E., Sando, T., Alluri, P., & Ponnaluri, R. (2022). Safety Evaluation of an Adaptive Signal Control Technology Using an Empirical Bayes Approach. *Journal of Transportation Engineering, Part A: Systems*, *148*(4). https://doi.or g/10.1061 /JTEPBS.0000652

Larry Clark. (2019). Traffic signals: A brief history. *Washington State University*, *22*. https://doi.org/10.4/JQUERY-UI.MIN.JS

Li, D., Wu, J., Xu, M., Wang, Z., & Hu, K. (2020). Adaptive Traffic Signal Control Model on Intersections Based on Deep Reinforcement Learning. *Journal of Advanced Transportation*, *2020*. https://doi.org/10.1155/2020/6505893

Maadi, S., Stein, S., Hong, J., & Murray-Smith, R. (2022). Real-Time Adaptive Traffic Signal Control in a Connected and Automated Vehicle Environment: Optimisation of Signal Planning with Reinforcement Learning under Vehicle Speed Guidance. *Sensors*, *22*(19). https://doi.org/10.3390/S22197501

Masri, S., Ashqar, H. I., & Elhenawy, M. (2024). *Large Language Models (LLMs) as Traffic Control Systems at Urban Intersections: A New Paradigm*. https://arxiv.org /abs/2411.10869v1

Miovision Team. (2024, June). *The Evolution of Traffic Signal Technology: Past, Present, and Future*.https://miovision.com/traffic-signal-technology-evolution/?ut m_sourc ec hatgpt.com

Mishra, S., Singh, V., Gupta, A., Bhattacharya, D., & Mudgal, A. (2023). Adaptive traffic signal control for developing countries using fused parameters derived from crowd-sourced data. *Transportation Letters*, *15*(4), 296–307. https://doi.org/ 10.1080/ 19427867.2022.2050493

Muli, F. (2020, June 22). *KURA to implement Huawei's intelligent transport system to ease traffic congestion in Nairobi*. TechTrends Media. https://kahawatungu.com/huaweis-transport-system-ease-traffic-nairobi/

Mushtaq, A., Haq, I. U., Sarwar, M. A., Khan, A., Khalil, W., & Mughal, M. A. (2023). Multi-Agent Reinforcement Learning for Traffic Flow Management of Autonomous Vehicles. *Sensors 2023, Vol. 23, Page 2373*, *23*(5), 2373. https://doi.o rg/10.339 0/S23052373

Ndatho, M. N. (2018). *Socio-Economic Effects of Traffic Congestion on Urban Mobility Along Jogoo Road, Nairobi City County, Kenya*. Kenyatta University.

Odhiambo, H. (2020, January 28). *Huawei to curb traffic blight in Nairobi | CIO Africa*. CIO Africa. https://cioafrica.co/huawei-to-curb-traffic-blight-in-nairobi-2/

Office of the Auditor-General, K. (2023). *Report of the Auditor-General on Nairobi Metropolitan Area Transport Authority for the Year Ended 30 June, 2023 Enhanced Reader*.

Pinto, R., Torres, P. M. B., Lohweg, V., Łach, Ł., & Svyetlichnyy, D. (2024). Comprehensive Review of Traffic Modeling: Towards Autonomous Vehicles. *Applied Sciences 2024, Vol. 14, Page 8456*, *14*(18), 8456. https://doi.org /10.3390/APP14188456

Pragalathan, J., & Schramm, D. (2024). Applications of Traffic Flow Predictions and Simulations in Urban Mobility with Connected Automated Vehicles and Advanced Traffic Management Systems. *Next Chapter in Mobility*, 183–194. https://doi.org/10.1007/978-3-658-42647-7_13

Praveen RVS. (2023, December 1). *Exploring the Evolving Landscape of Large Language Models | LinkedIn*. LinkedIn. https://www.linkedin.com/pulse/exploring-evolving-landscape-large-language-models-praveen-ph-d-cmvyc/

Schrank, D., Eisele, B., & Lomax, T. (2023). 2023 URBAN MOBILITY REPORT. *Texas A&M Transportation Institute*. https://mobility.tamu.edu/umr/report/#methodology

Shankaran, S., & Rajendran, L. (2021). Real-Time Adaptive Traffic Control System for Smart Cities. *2021 International Conference on Computer Communication and Informatics, ICCCI 2021*. https://doi.org/10.1109/ICCCI50826.2021.9402597

Shao, M., Basit, A., Karri, R., & Shafique, M. (2024). Survey of different Large Language Model Architectures: Trends, Benchmarks, and Challenges. *IEEE Access*. https://doi.org/10.1109/ACCESS.2024.3482107

Siqi Lai, Zhao Xu, Weijia Zhang, Hao Liu, & Hui Xiong. (2023). *LLMLight: Large Language Models as Traffic Signal Control Agents | Request PDF. 10.48550/ arXiv. 2312.16044*.https://www.researchgate.net/publication/383693377_LLMLight_Large _Language_Models_as_Traffic_Signal_Control_Agents

Soori, M., Arezoo, B., & Dastres, R. (2023). Artificial intelligence, machine learning, and deep learning in advanced robotics, a review. *Cognitive Robotics*, *3*, 54–70. https://doi.org/10.1016/J.COGR.2023.04.001

Swapno, S. M. M. R., Nobel, S. N., Meena, P., Meena, V. P., Azar, A. T., Haider, Z., & Tounsi, M. (2024). A reinforcement learning approach for reducing traffic congestion using deep Q learning. *Scientific Reports 2024 14:1*, *14*(1), 1–20. https://doi.org/10.1038/s41598-024-75638-0

Todd Litman. (2024). *Smart Congestion Relief: Comprehensive Evaluation of Traffic Congestion Costs and Congestion Reduction Strategies*. www.vtpi.org /ITED_ congestion.pdf

Tomar, I., Indu, S., & Pandey, N. (2022). Traffic Signal Control Methods: Current Status, Challenges, and Emerging Trends. *Lecture Notes on Data Engineering and Communications Technologies*, *90*, 151–163. https://doi.org/10.1007/978-981-16-6289-8_14

Wan, C. H., & Hwang, M. C. (2018). Value-based deep reinforcement learning for adaptive isolated intersection signal control. *IET Intelligent Transport Systems*, *12*(9), 1005–1010. https://doi.org/10.1049/IET-ITS.2018.5170

Wang, M., Pang, A., Kan, Y., Pun, M.-O., Chen, C. S., & Huang, B. (2024). *LLM-Assisted Light: Leveraging Large Language Model Capabilities for Human-Mimetic Traffic Signal Control in Complex Urban Environments*. http://arxiv.org/abs/2403.08337

Wei, H., Zheng, G., Gayah, V., & Li, Z. (2020). A Survey on Traffic Signal Control Methods. *2020. A Survey on Traffic Signal Control Methods*, *1*(1). https://doi.org/ 10.1145/nnnnnnn.nnnnnnn

Zhang, D., Mohammed, S., & Calvi, A. (2021). Editorial: Special Issue: Fuzzy logic systems for transportation engineering. *Journal of Intelligent and Fuzzy Systems*, *41*(4), 4705–4712. https://doi.org/10.3233/JIFS-189957

Zhao, H., Chen, S., Zhu, F., & Tang, H. (2022). Traffic Signal Control Based on Reinforcement Learning and Fuzzy Neural Network. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, *2022-October*, 4030–4035. https://doi.org/10.1109/ITSC55140.2022.9922570

APPENDICES

Appendix I: Evaluation Matrix of LLM Agents

LLM Agent	Strengths	Weaknesses	Adaptive Traffic Signal Control	Best Responses / Suggestions
OpenAI (GPT-4.0/5)	High accuracy, strong reasoning, and well-documented APIs.	Expensive, computationally heavy, and a risk of hallucinations.	Strong performance in dynamic traffic decision making with YOLOv5 inputs.	Use for critical intersections requiring nuanced decision-making.
Anthropic (Claude)	Safety-focused, context-aware, and good at structured outputs.	Sometimes conservative in responses; slower under heavy loads.	Useful for policy-focused dashboards and ensuring safe decisions.	Integrate where transparency and ethical outputs are prioritized.
Cohere	Fast, optimized for embeddings and intent classification.	Not as strong in complex reasoning as GPT-4 or Claude.	Suitable for lightweight tasks, such as quick traffic summaries.	Pair with another stronger LLM for decision-heavy tasks.
Google Gemini	Strong multimodal capabilities (vision + text).	Still emerging, API maturity varies.	Potential to combine video inputs directly with traffic control decisions.	Future-proof option once multimodal integration is stable.
Groq	Extremely fast inference (hardware optimized).	Limited ecosystem compared to OpenAI/Google.	Best for real-time responses, where speed is critical, such as signal switching.	Deploy for high-frequency tasks needing millisecond responses.
Mistral	Open source, cost-effective, and fine-tunable.	Lag in raw performance compared to GPT 4.0.	Useful for localized and customizable deployments.	Fine-tune with Nairobi traffic datasets for stronger performance.

Appendix II: Evaluation Matrix – Police Officers (N = 5)

Evaluation Aspect	**Sample Questions**	**Observed/ Reported Responses**	**Responses**
Usability in Practice	Can officers override the model in emergency situations?	Requested manual override option.	Add an override button or an interface for officers.
Integration with Existing Systems	How will it work with static timers?	Static timers dominate Nairobi.	Hybrid system: AI + existing timers.
Reliability	What about blackouts or a poor network?	Concern about disruptions.	Fallback to fixed cycles or manual control.
Safety Considerations	How are emergency vehicles prioritized?	Liked the automatic detection idea.	Integrate a siren or a beacon detection.
Perceived Benefits	Does it make their work easier?	Reduces fatigue during peak hours.	It highlights reduced workload & stress for the police officers.

Appendix III: Evaluation Matrix – Nairobi City Council Traffic Marshals (N = 3)

Evaluation Aspect	**Sample Questions Asked**	**Observed/Reported Responses**	**Responses**
Ease of Integration	How will this model integrate with the current manual signalling and whistle control used?	Marshals noted they mostly rely on hand signals or whistles during congestion.	Provide training modules for marshals to use the system interface, ensuring smooth adoption.
Support for Manual Roles	Will this system reduce the need for marshals at every junction?	Concerned that it might replace their jobs.	Clarified to the mashals that the model is assistive, not a replacement. The marshals will still be vital during breakdowns, protests, or non-routine events like presidential escorts.
Real-time Responsiveness	Can the system adjust quickly to sudden pedestrian surges, for example, school children crossing and matatu offloading?	Marshals found this feature promising but wanted assurance of responsiveness.	Highlight that the model uses YOLOv5 video inputs and adapts signals dynamically in seconds.
Fairness and Equity	Will side roads or less busy lanes still get ignored?	Fear that smaller feeder roads might be overlooked.	Emphasize the model's lane-specific fairness logic. The LLM allocates green time based on demand, not a preset hierarchy.
Training & Ownership	How will marshals be involved in future use?	Wanted recognition as key field actors, not to be left out.	Recommend co-design workshops with marshals for rollout, to build ownership and acceptance.

Appendix VI: Evaluation Matrix – Colleagues (N = 7)

Evaluation Aspect	**Sample Questions**	**Observed/Reported Responses**	**Responses**
Technical Performance	How effective is the developed model vs fixed-time control?	Observed clear improvements.	Show KPI results & expand corridor tests.
Scalability	Can it scale to the CBD or Nairobi?	Skepticism due to infrastructure gaps.	Pilot at a few intersections, scale gradually.
Interpretability	How transparent are decisions?	Asked for visibility of model logic.	Provide logs and dashboards of inputs & decisions.
Robustness	How does it handle irregular events?	Impressed with the scenarios, want real-world tests.	Add progressive real-world trials.
Policy Impact	Can it inform long-term planning?	Saw potential beyond signal timing.	Use outputs as policy insights for planning.

Appendix V: KUREC Clearance Letter

KABARAK UNIVERSITY RESEARCH ETHICS COMMITTEE

Private Bag - 20157
KABARAK, KENYA
Email: kurec@kabarak.ac.ke

Tel: 254-51-343234/5
Fax: 254-051-343529
www.kabarak.ac.ke

OUR REF: KABU01/KUREC/001/05/06/25

Date: 11th June, 2025

Monicah Wambui Hinga
Reg. No: GMI/ON/2213/09/22
Kabarak University,

Dear Monicah,

RE: INTELLIGENT TRAFFIC MANAGEMENT IN NAIROBI USING LARGE LANGUAGE MODEL AGENTS FOR ADAPTIVE TRAFFIC SIGNAL CONTROL

This is to inform you that ***KUREC*** has reviewed and approved your above research proposal. Your application approval number is ***KUREC-050625.*** The approval period is 11***/06/2025 – 11/06/ 2026.***

This approval is subject to compliance with the following requirements:

i. All researchers shall obtain an introduction letter to NACOSTI from the relevant head of institutions (Institute of postgraduate, School dean or Directorate of research)
ii. The researcher shall further obtain a RESEARCH PERMIT from NACOSTI before commencement of data collection & submit a copy of the permit to **KUREC.**
iii. Only approved documents including (informed consents, study instruments, MTA Material Transfer Agreement) will be used
iv. All changes including (amendments, deviations, and violations) are submitted for review and approval by ***KUREC***:
v. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to ***KUREC*** within 72 hours of notification;
vi. Any changes, anticipated or otherwise that may increase the risk(s) or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to ***KUREC*** within 72 hours;
vii. Clearance for export of biological specimens must be obtained from relevant institutions and submit a copy of the permit to KUREC;
viii. Submission of a request for renewal of approval at least 60 days prior to expiry of the approval period. Attach a comprehensive progress report to support the renewal and;
ix. Submission of an executive summary report within 90 days upon completion of the study to ***KUREC***

Sincerely,

KABARAK UNIVERSITY
INSTITUTIONAL RESEARCH ETHICS COMMITTEE
13 JUN 2025
APPROVED
P O. PRIVATE BAG - 20157, KABARAK

Prof. Jackson Kitetu PhD.
KUREC-Chairman

Cc Vice Chancellor
DVC-Academic & Research
Registrar-Academic & Research
Director-Research Innovation & Outreach
Institute of Post Graduate Studies

As members of Kabarak family, we purpose at all times and in all places, to set apart in one's heart, Jesus as Lord.
(1 Peter 3:15)
Kabarak University is ISO 9001:2015 Certified

Appendix VI: NACOSTI Research Permit

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR
SCIENCE,TECHNOLOGY & INNOVATION

Ref No: **621701**

Date of Issue: **30/June/2025**

RESEARCH LICENSE

This is to Certify that Ms.. Monicah Wambui Hinga of Kabarak University, has been licensed to conduct research as per the provision of the Science, Technology and Innovation Act, 2013 (Rev.2014) in Nairobi on the topic: INTELLIGENT TRAFFIC MANAGEMENT IN NAIROBI USING LARGE LANGUAGE MODEL AGENTS FOR ADAPTIVE TRAFFIC SIGNAL CONTROL for the period ending : 30/June/2026.

License No: **NACOSTI/P/25/4175744**

621701

Applicant Identification Number

Deputy Director
NATIONAL COMMISSION FOR
SCIENCE,TECHNOLOGY &
INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document,
Scan the QR Code using QR scanner application.

See overleaf for conditions

Appendix VII: Evidence of Conference Participation

SOCIETY OF EDUCATIONAL RESEARCH
AND
EVALUATION IN KENYA

SEREK

CERTIFICATE OF PARTICIPATION

THIS IS TO CERTIFY THAT

MONICAH WAMBUI HINGA

HAS SUCCESSFULLY PARTICIPATED IN TWO DAY
INTERNATIONAL CONFERENCE ON
TURNING CHALLENGES INTO OPPORTUNITIES: IMPLICATIONS
IN THE 21ST CENTURY

HELD AT THE CO-OPERATIVE UNIVERSITY OF KENYA (MAIN CAMPUS)
ON 22ND OCTOBER, 2025 TO 23RD OCTOBER, 2025

SEREK

PROF. PAUL MAITHYA
CHAIRMAN– SEREK

SR. DR. PAULINE KEBENEI
SECRETARY – SEREK

Appendix VIII: List of Publication

International Journal of Applied Science and Research

Large Language Model Agents for Adaptive Traffic Signal Control: A Simulation Case Study in Nairobi

Monicah Wambui Hinga

Department of Computer Science, Kabarak University, Nakuru, Kenya

DOI: https://doi.org/10.56293/IJASR.2025.6801

IJASR 2025
VOLUME 8
ISSUE 6 NOVEMBER - DECEMBER **ISSN: 2581-7876**

Abstract: Traffic congestion in Nairobi's Central Business District continues to impose high economic, social, and environmental costs. Long queues at intersections, wasted fuel, and poor air quality are common outcomes of conventional fixed-time traffic signals. These systems dominate the city but do not respond to fluctuating or multimodal traffic. This study explored the use of Large Language Model (LLM) agents as adaptive controllers and compared their performance with existing fixed-time plans. Traffic video recordings were collected from selected intersections and analyzed using the YOLOv5 object detection algorithm to estimate lane-specific vehicle counts. The processed counts were then used to calibrate a Simulation of Urban Mobility (SUMO) environment. Within this setup, LLM agents allocated green times dynamically and adjusted signal phases in real time. The study adopted an experimental simulation design, testing both peak and off-peak traffic conditions as well as disruption scenarios such as blocked approaches and emergency vehicle passage. To ensure reliability, the SUMO model was calibrated against observed volumes and validated using standard traffic simulation statistics. Performance was assessed using three key indicators: average waiting time, intersection throughput, and responsiveness to demand fluctuations. Results showed that the LLM-based model reduced waiting times by up to 35%, increased throughput by 12–18%, and stabilized signal plans within fewer cycles than the fixed-time baseline. Beyond efficiency gains, the study demonstrates the feasibility of repurposing generalist AI models as decision agents in traffic management, offering a low-cost, scalable solution particularly suited to resource-constrained cities. By providing localized evidence from Nairobi, the research contributes to Intelligent Transportation Systems (ITS) literature and supports policy directions that include piloting AI-powered adaptive control at critical intersections as part of broader smart mobility strategies in African cities.

Keywords: Traffic congestion, Adaptive Traffic Signal Control, Large Language Model Agents, Intelligent Transportation Systems, SUMO Simulation, YOLOv5.

1.0 Introduction

Urban mobility is central to economic growth, social inclusion, and environmental sustainability. As cities expand, the efficiency of traffic management systems determines how well people and goods move within limited infrastructure. In developing contexts such as Nairobi, the challenge is compounded by rapid urbanization, rising motorization, and resource constraints. This introduction outlines the background to the study, beginning with the problem of traffic congestion, reviewing global and local approaches to traffic control, and highlighting the emerging role of Artificial Intelligence (AI)—particularly Large Language Models (LLMs)—in shaping next-generation solutions.

1.1 Traffic Congestion

Traffic congestion is among the most persistent challenges facing modern cities, with significant consequences for mobility, productivity, the environment, and social well-being. It is generally defined as a condition in which the demand for roadway space exceeds capacity, resulting in reduced speeds, prolonged trip times, and increased queuing (Litman, 2025). Congestion is typically categorized into recurring (predictable during peak hours due to high demand and limited road space) and non-recurring (triggered by incidents, weather events, or disruptions) types, both of which demand different management strategies (Arti et al., 2022; Jha & Albert, 2021).

The impacts of traffic congestion are multidimensional. Economically, delays translate into wasted time and fuel, raising logistics costs and reducing worker productivity. In Nairobi, commuters spend an average of over 45 minutes in daily traffic delays, with direct consequences for household budgets and national output (KNBS, 2020). From an

