

# DEVELOPMENT OF A REAL-TIME AND RETROSPECTIVE INTELLIGENT OBJECT DETECTION SYSTEM FROM CCTV FOR SECURITY ENHANCEMENT IN LOCAL GOVERNMENT ORGANIZATIONS

CHANVIT PHROMPHANCHAI<sup>1</sup>, SUWUT TUMTHONG<sup>2\*</sup>

<sup>1,2</sup>Faculty of Science and Technology, Rajamangala University of Technology Suvarnabhumi, Bangkok, Thailand

E-mail: <sup>1</sup>166491432007-st@rmutsb.ac.th, <sup>2\*</sup>Suwut.t@rmutsb.ac.th

## ABSTRACT

This research develops an AI-RTRO system to address the limitations of traditional CCTV in Bang Yai Municipality's Smart Safety Infrastructure. The system improves real-time situational awareness and enhances retrospective evidence retrieval for local government. Managing large-scale CCTV data and making rapid decisions are key operational challenges. The system integrates the AI Agent Core framework and IPOF model for object detection, tracking, and retrieval from both real-time and archived video. Deep learning algorithms (YOLO, DeepSORT/ByteTrack, ResNet, Faiss) are used, with Grad-CAM providing transparency. Performance is evaluated in real-world conditions with mAP, IDF1, and latency. Results show accurate detection, stable tracking, and low-latency operation. Metadata management and feature vector indexing improve the performance of retrospective retrieval. The study confirms that integrating AI Agent Core and IPOF improves system performance, retrieval efficiency, decision transparency, and continuous learning. The framework was rated highly by 15 experts, with a mean of 4.74 (S.D. = 0.55), indicating readiness for practical use in local government surveillance.

**Keywords:** *Smart City, Object Detection, Time-Indexed Retrieval, CCTV Analytics, AI Agent Core*

## 1. INTRODUCTION

Smart City development is an important policy that aims to improve citizens' quality of life, increase safety, and enhance the efficiency of sustainable urban management. This is particularly true for the “Smart Safety Infrastructure” dimension, which is the core mechanism of Smart City Thailand [1], [2]. It utilizes digital technology and artificial intelligence to monitor, prevent, and respond to incidents in urban areas in a timely manner. Local agencies, such as Bang Yai Municipality in Nonthaburi Province, have installed closed-circuit television (CCTV) systems to monitor and maintain community safety. However, traditional CCTV systems face limitations due to high video data volumes and rely on manual inspection, leading to delays, the inability to detect incidents in real time, and inconvenient evidence retrieval. To mitigate these limitations, this research proposes an AI-RTRO (Artificial Intelligence – Real-Time & Retrospective Object Detection) system, an intelligent system for “detecting” and “tracking”

objects both in real time and retrospectively from CCTV video. It is designed to be consistent with the AI-Agent Core structure and the IPOF (Input–Process–Output–Feedback) framework, creating a continuous cycle of perception–decision–action–learning based on real data from municipal areas.

AI-RTRO's architecture is divided into two functionalities. First, Data Management & Integration (Input) takes in both real-time streams and 10 - minute recorded clips with necessary metadata (date, time, camera ID, and location) for processing. Then, the AI Agent Core Techniques (Process) is the heart of the system. It uses various deep learning techniques, including YOLO for real-time object detection [3], [4], DeepSORT/ByteTrack for continuous cross-frame multi-object tracking, Tracking [5], [6], supplemented by ResNet for deep object feature extraction [7] and Faiss for vector indexing for fast retrospective image retrieval [8]. Explainable AI also integrates Grad-CAM to explain the model's decisions by highlighting multiple sources of

anomalies in video [9], and provides key performance indicators (KPIs), including mAP for detection accuracy [10], and IDF1 for tracking quality [11], to systematically evaluate and control quality. The processing results are presented through an Output & User Interface in the form of a real-time dashboard, automatic alerts, and a video history search system with Smart Recommendation to support operational decision-making by municipal officials. The system also designs a Feedback System as an Active Learning Loop to bring back difficult cases or model uncertainties, thereby improving the dataset and training new models [12]. It references metrics such as mAP and IDF1 as criteria for deployment approval [13]. In addition, the Environment Test evaluates the system's operation in a real municipal environment, assessing detection accuracy and the availability of camera equipment (e.g., signal loss or maintenance) to ensure the system remains stable before full deployment in the field [14]. AI-RTRO is an intelligent system framework that integrates CCTV data with modern AI techniques under the IPOF process. This enables detection, tracking, retrieval, interpretation, and continuous learning. This solution addresses the "smart security infrastructure" of modern cities and supports faster, more accurate, and sustainable management of Bang Yai Municipality.

To address these limitations, this study proposes an integrated AI-RTRO framework. The framework systematically integrates real-time object detection, time-indexed retrospective retrieval, and intelligent decision support within a unified AI Agent Core architecture. Unlike conventional CCTV analytics that focus solely on detection accuracy, the proposed approach emphasizes end-to-end system intelligence. It does so by integrating the IPOF (Input–Process–Output–Feedback) model to enable continuous learning and operational adaptability. The framework is designed to achieve high detection and tracking performance with efficient retrospective retrieval. It also aims for transparent decision-making through Explainable AI, and sustained performance improvement via active learning mechanisms. Consequently, this research aims to strengthen smart surveillance capabilities for local government organizations. It seeks to transform CCTV systems from passive monitoring tools into intelligent, adaptive, and decision-oriented infrastructures.

The research hypothesis consists of:

H1: The AI-RTRO system can achieve high real-time object detection and tracking performance, as reflected by mAP, IDF1, and low latency, suitable

for deployment in real municipal CCTV environments.

H2: The integration of Time-Indexed Retrieval using metadata and feature vector indexing significantly improves the efficiency and accuracy of retrospective object search in CCTV video archives.

H3: The application of Explainable AI techniques within the AI-RTRO system enhances decision transparency and user trust, supporting operational decision-making in local government organizations.

H4: The integration of the AI Agent Core architecture with the IPOF model enables continuous learning and system adaptability, and the proposed conceptual framework will be evaluated by experts as highly appropriate for Smart City and Smart Safety Infrastructure applications.

## 2. THEORETICAL FRAMEWORK

### 2.1 Smart City

The concept of smart cities has evolved from traditional physical city management to a system driven by data, information technology, and digital innovation, aiming to improve the quality of life, the efficiency of public services, and the sustainability of cities. For city officials, implementing smart city technologies presents clear value. These systems help optimize resource use, streamline service delivery, and enable more informed, timely decision-making. Research indicates that smart cities are a "continuous development process" that systematically integrates technology with infrastructure [15]. Internationally, smart cities often rely on key technologies such as IoT, big data, cloud computing, and AI to support city decision-making, while also addressing cybersecurity and data privacy challenges. IoT sensor systems and infrastructure are identified as the core of real-time environmental monitoring and urban management [16]. Integrating IoT with AI and analytics can enhance a city's ability to efficiently manage traffic, energy, and infrastructure [17].

In the Thai context, the Thailand 4.0 policy and the role of the Digital Economy Promotion Agency (DEPA) have led to the adoption of the smart city concept, with a focus on people as the center of continuous learning and adaptation in urban areas [2]. Meanwhile, smart city surveillance systems have become a key component of public safety, with AI and cloud technologies enabling accurate incident detection, object tracking, and historical data retrieval [18], as well as new approaches such as SaSTL, which analyze massive amounts of data to support real-time security incident monitoring. Overall, the literature suggests that smart cities are digital ecosystems that must

balance efficiency, sustainability, and citizens' rights, while addressing policy gaps, digital divides, and local contexts that require further development [19].

Table 1: Synthesis of 2.1 Smart City workflow

Component	Content	Ref.
Smart Governance & Policy Framework	Establishing policies and regulatory frameworks for smart cities to enhance the quality of life, safety, and sustainability of cities, emphasizing the use of data and digital technology to support decision-making by government agencies.	[1], [2], [15]
Smart Infrastructure & IoT Integration	City infrastructure uses IoT sensors, cloud, and big data to enable real-time surveillance, detection, and event management, improving efficiency, safety, and decision-making for applications such as environmental, traffic, and public safety monitoring systems.	[16], [17], [18]
Smart Safety & Intelligent Surveillance	A data-driven decision-making system that collects, analyzes, and interprets real-time city data to improve the efficiency of public services and reduce management gaps.	[18], [19], [20]
Data-Driven City Management	A data-driven decision-making system that collects, analyzes, and interprets real-time city data to improve the efficiency of public services and reduce management gaps.	[15], [17]
AI-Enhanced Public Services	Using AI to enhance government services such as risk assessment, emergency response, traffic management, anomaly detection, and situation reporting to support city administrators.	[18], [20]
Cybersecurity & Data Privacy	Considering cybersecurity and citizens' data rights, which are key challenges for smart cities that require balancing efficiency and data security.	[16]
Sustainable & Citizen-Centric Development	Smart cities must be designed with a people-centric approach, focusing on sustainability, the environment, the economy, and society, and reducing digital inequality.	[2], [15]

According to the table, the Smart City concept aims to improve citizens' quality of life and safety by integrating digital technologies and big data to support effective policy decisions [1], [2]. Furthermore, infrastructures such as IoT, Cloud, and AI allow cities to better monitor, detect, and respond

to real-time events [15], [16]. Moreover, intelligent image and video analytics can enhance safety surveillance across the city [17], [18]. However, smart cities must address cybersecurity and data privacy to build citizen trust [16]. Finally, development should follow a people-centric approach, focusing on economic, social, and environmental sustainability [15].

## 2.2 Object Detection

Object detection is a key component of modern computer vision. Because it can simultaneously identify the location (localization) and type (classification) of objects, it has been widely applied in smart surveillance systems, smart cities, autonomous vehicles, and real-time video analysis. Although early architectures such as Faster R-CNN laid the foundation for accuracy, since 2020, research has evolved towards one-stage architectures that emphasize speed and support more edge device processing, such as YOLOv4–v10, which can efficiently perform real-time tasks under resource constraints [20], [21], [22], [23]. The new generations, YOLOv7–YOLOv10, have also made significant advancements, featuring backbone, head, spatial–channel optimization, and the NMS-Free Detector concept, which improves speed and accuracy. This is especially true in safety and industrial applications where high stability is required. Meanwhile, Transformer-based approaches such as DETR, Deformable DETR, and DINO have transformed the end-to-end sensing architecture, enabling the system to learn context across the entire image, reducing reliance on NMS and increasing the ability to handle complex environments [3]. CCTV Analytics applications face challenges like variable lighting, occlusion, low resolution, and congestion. To address these, multi-object tracking techniques such as ByteTrack and OC-SORT have been integrated. This improves continuity and reduces tracking errors, allowing reliable analysis of real events even in adverse environments [6], [24]. These advances are crucial for AI-RTRO (Artificial Intelligence – Real-Time and Retrospective Object Detection) systems. Such systems require speed, accuracy, and time-indexed retrieval to aid smart cities, security agencies, and local governments in managing vast CCTV data for timely, effective policy decisions.

Table 2: Synthesis of 2.2 Object Detection workflow

Component	Content	Ref.
Modern Detection Architectures	Next-generation object detection architectures, such as YOLOv4–YOLOv10 and Faster R-CNN, aim to achieve high speed and accuracy, supporting complex environments.	[21], [22], [23], [24]
Real-Time Detection Optimization	Real-time detection enhancements via redesigned Backbone, Neck, and Head with NMS-Free techniques and Edge processing.	[3]
Transformer-Based Detection	The use of DETR, Deformable DETR, and DINO architectures to learn deep image context reduces the limitations of traditional NMS.	[3]
Multi-Object Tracking Integration	Integrating detection with tracking, such as DeepSORT and OC-SORT, to solve ID-Switching issues and improve tracking continuity.	[5], [6], [25]
Feature Extraction for Retrieval	Using CNNs such as ResNet to extract object features for retrospective retrieval and object classification..	[7]
Vector-Based Retrieval & Indexing	Using vector databases such as FAISS to quickly search for objects in retrospect, supporting high-volume CCTV systems.	[8]
Explainability in Object Detection	Using Grad-CAM to rationalize model decisions increases the transparency and reliability of object detection systems.	[9]

Object detection has evolved from traditional architectures to high-speed models such as YOLOv4–YOLOv10 (You Only Look Once, versions 4-10), achieving high accuracy and performance in real-world tasks [21], [22], [23], [24]. Building on this progress, newer models like DETR (Detection Transformers) and DINO (DETR with Improved DeNoising training) enable the system to learn deeper image context while optimizing in real time. These models leverage a newly designed Backbone–Head structure and employ a technique called NMS-Free (removing Non-Maximum Suppression post-processing) [3]. Furthermore, integration with tracking algorithms such as DeepSORT (Deep Simple Online and Realtime Tracking) and OC-SORT (Orientation and Center-based SORT) reduces ID-Switching issues and improves object tracking stability [5], [6], [25]. Additionally, the system also applies ResNet (Residual Network) and FAISS (Facebook AI Similarity Search) for fast historical data extraction and retrieval, while Grad-CAM (Gradient-weighted

Class Activation Mapping) enhances the transparency of model decisions [7],[9].

### 2.3 Time-Indexed Retrieval

The concept of Time-Indexed Retrieval has become increasingly important in the era of big data, focusing on queries that must consider “temporal characteristics” such as location, duration, sequence, and event relationships. This has led to the development of specialized index structures and algorithms, such as RD-Index, which supports Range–Duration Queries through a time–duration space structure, significantly improving the efficiency of queries in temporal data. Furthermore, TELII was developed to store events and temporal relationships in an inverted index, enabling faster and more accurate queries in complex databases such as EHRs [25], [26], [27]. In addition, research on temporal data validation and distributed index structures for large-scale time series data has also enhanced the reliability and support for industrial-scale processing. Furthermore, the application of the Time-Indexed Retrieval concept encompasses both temporal and retrospective retrieval (Time-Travel Retrieval) and the management of non-uniform time-series data. Deep learning techniques are integrated to support Content-Based Time Series Retrieval, as well as the use of document statistics across time points in a Hybrid Time-Travel Retrieval system to ensure reproducible retrospective results even with continuous database updates [28]. Furthermore, Durable Top-K Queries and sweep-line algorithms have been proposed to address the ranking problem in temporal-inaccurate data [29], [30]. Overall, the aforementioned frameworks, methods, and tools form the foundation of modern temporal retrieval systems that require high accuracy, speed, and reliability in many domains, including intelligent CCTV systems that require detailed time-course and event correlation retrieval.

Table 3: Synthesis of 2.3 Time-Indexed Retrieval workflow

Component	Content	Ref.
Temporal Query Framework	A search structure that supports temporal queries such as range, duration, sequence, and event relationships to search for data that changes over time.	[26], [27]
Specialized Index Structures	Specialized index structures such as RD-Index for Range–Duration Queries, and TELII for storing events with temporal relationships as an inverted index to speed up searches.	[26], [27], [28]

Component	Content	Ref.
Distributed & Scalable Time-Series Indexing	Massively distributed indexing of large industrial datasets enables the system to efficiently handle large amounts of data.	[27]
Hybrid Time-Travel Retrieval	The Hybrid Time-Travel Retrieval concept combines document statistics and time-based data to provide accurate retrospective searches even when the database is constantly updated, enabling reproducible retrospective results.	[29]
Temporal Ranking & Top-K Durability	Time-tolerant event ranking techniques, such as Durable Top-K Queries and Sweep-Line Algorithms, address ranking issues in imprecise time data.	[30], [31]
Content-Based Temporal Retrieval using Deep Learning	Integrate Deep Learning for Content-Based Retrieval of Temporal Information, such as CNN-based feature extraction for Time Series or Video.	[26], [29]
Time-Indexed Retrieval for CCTV Analytics	Application of time-based search concepts to retrospective object retrieval in CCTV, such as using ResNet vectors and vector databases like Faiss to search for objects at the second level.	[8]
Reliability & Validation of Temporal Data	Validation of time data, including temporal data validation, and handling non-uniform time series to ensure high reliability of the search system.	[28]

The concept of Time-Indexed Retrieval focuses on managing and retrieving data with inherent temporal characteristics by utilizing specialized index structures such as RD-Index and TELII, which support efficient time-range queries, duration-based searches, and event-sequence retrievals [26], [27], [28]. Furthermore, to enhance query speed and scalability for large-scale data, the system has evolved toward distributed indexing [27]. Additionally, Hybrid Time-Travel Retrieval enables accurate historical queries even when the database is continuously updated [29]. Recent studies emphasize not only Durable Top-K event ranking and techniques for resolving temporal inconsistencies using the Sweep-Line method [30], [31], but also the integration of Deep Learning approaches to extract semantic features that improve temporal retrieval accuracy [26], [29].

## 2.4 CCTV Analytics

CCTV analytics has evolved from rule-based detection to AI and deep learning-based systems. These systems improve the accuracy of object detection, tracking, and real-time behavior recognition. This evolution meets the needs of smart cities and modern security systems. Modern architectures such as YOLOv4-YOLOv11, Vision Transformers[31], [32], and multi-dimensional models play a key role in enhancing these capabilities. Time-indexed retrieval has also made retrospective image searches more efficient. It uses metadata and feature vectors, which support more accurate local government applications. Research directions are developing in the areas of accuracy, speed, and retrospective event retrieval, including edge device functionality and integration with smart city systems, such as safety and traffic. Furthermore, there are developments in anomaly detection systems, Re-ID techniques, and Multi-Camera Tracking that support cross-area object tracking[33], while vector databases such as FAISS and Milvus speed up Time-Indexed Search. CCTV Analytics is thus becoming an intelligent platform that integrates AI, Big Data, and Edge Computing, which serves as a key foundation for PhD research on efficient real-time detection and retrospective retrieval [34], [35], [36], [37], [38].

Table 4: Synthesis of 2.4 CCTV Analytics workflow

Component	Content	Ref.
AI-Based Object Detection	Using next-generation object detection models, such as YOLOv4, YOLOv11, and Vision Transformers, to analyze CCTV images in real time increases accuracy and supports analysis in complex environments.	[32], [33], [37]
Multi-Object & Multi-Camera Tracking	Using tracking techniques such as DeepSORT, OC-SORT, and Multi-Camera Tracking, the system can continuously track multiple objects, supporting high-congestion areas and cross-camera movement.	[34]
Time-Indexed & Retrospective Retrieval	Historical event searches using metadata, feature vectors, and vector databases such as FAISS enable fast, accurate searches for historical objects within seconds.	[35]
Anomaly & Behavior Recognition	AI-powered detection of abnormal behaviors or risky situations, such as unusual movements, unusual crowding, or suspicious objects, to support proactive surveillance.	[32], [33], [39]

CCTV Analytics technology has advanced exponentially through the use of modern object detection models, such as YOLOv4–YOLOv11 and Vision Transformers, to improve accuracy and support real-time operations [32], [33], [54]. Tracking systems such as DeepSORT and OC-SORT enable continuous object tracking even in highly congested environments. Time-Indexed Retrieval and vector databases such as FAISS enable fast and accurate object retrieval within seconds [35]. AI has also been applied to detect abnormal behaviors and risky situations to enhance city safety [32], [39]. Finally, integrating CCTV Analytics with Smart City systems enables cities to respond to incidents quickly and manage safety more efficiently.

### 2.5 AI Agents

The concept of "AI Agents" has evolved dramatically. Traditionally, agents were defined as entities with basic capabilities: perceiving, reasoning, and acting to accomplish tasks in dynamic environments [39], [40], [41]. Advances in large language models (LLMs) have extended these capabilities beyond the basics. AI agents can now become "goal-oriented" autonomous systems, performing granular planning, invoking external tools, learning from experience, and flexibly adapting to real-world contexts [42]. Recent research distinguishes "Vertical AI Agents," which emphasize domain-specific knowledge integrated with LLMs, enabling architectures that support memory, specialized tools, and expert-level cognitive skills [43]. Importantly, a semantic distinction exists between "AI Agents," which focus on individual problem-solving, and "Agentic AI," which refers to systems of multiple agents coordinating and communicating to achieve system-level goals [44]. Architecturally, modern AI agents must incorporate core capabilities such as planning, reasoning, tool use, and reflection to efficiently handle complex tasks [45]. Meanwhile, research on multi-agent collaboration indicates that role partitioning into planners, auditors, or doers, along with effective communication mechanisms, is crucial for the success of collaborative systems in data analysis and decision-making [46]. Security and governance issues remain crucial, particularly in systems related to finance, security, and digital infrastructure, where governance measures such as real-time verification and agent identification are required [47]. A comprehensive review of the literature suggests that AI agents have evolved from a basic concept into architectures that support both single- and multi-agent collaboration, with careful

considerations of trust, security, and ethics [48], [49], [50].

Table 5: Synthesis of AI Agents workflow

Component	Content	Ref.
Foundational AI Agent Capabilities	The basic capabilities of AI Agents include Perception, Reasoning, and Action to operate in ever-changing environments.	[40], [41], [42]
Goal-Oriented & Autonomous Decision-Making	Modern AI agents can plan at a detailed level, learn from experience, and make automated decisions based on defined goals.	[43], [46]
Vertical AI Agents & Domain-Specific Knowledge	AI Agents work together in a multi-agent fashion, with roles such as Planner–Auditor–Executor, to enhance system-level data analysis and decision-making.	[44]
Multi-Agent Collaboration (Agentic AI)	XAI techniques can be classified into two groups: general methods (e.g., LIME, SHAP) and model-specific methods (e.g., Grad-CAM).	[45], [47], [51]
Explainability, Trust & Ethical Supervision	Designing AI Agents to be transparent, auditable, and secure with Explainable AI, real-time auditing, and ethical governance mechanisms.	[38], [39], [48]
Security & Risk Management in AI Agents	Research has identified security risks such as data breaches and tool misuse, requiring countermeasures and agent security architectures.	[46], [40]
Adaptive Learning & Continuous Improvement	AI Agents adapt and learn from new situations through reinforcement learning and feedback, maintaining performance in real-world environments.	[49], [50]

AI agents have evolved from systems that can perceive, reason, and act on their own to architectures that are transparent, auditable, and have strict security oversight [38], [39], [41], [48]. Building on this foundation, advances in LLMs enable agents to have goal-directed capabilities, plan hierarchies, use external tools, and continuously learn from experience [43], [46], [50]. Furthermore, vertical AI agents are also being developed that are embedded with specialized knowledge, enabling more accurate, context-specific decision-making [44], [49]. In addition, agentic AI concepts support the collaboration of multiple agents through specialized roles, such as planner–auditor–executor, to enhance system-level thinking and decision-making [45], [47], [51]. Overall, research aims to create AI agents that are intelligent, secure,

transparent, and responsive to practical applications in modern data analytics and digital infrastructures [40], [42].

### 3. METHODOLOGY

This study adopts a systematic, mixed-methods research methodology that integrates theoretical analysis, system development, experimental evaluation, and expert validation. The research process begins with a comprehensive review of relevant theories and state-of-the-art techniques in object detection, intelligent agents, and CCTV analytics to establish a solid conceptual foundation. Subsequently, a prototype system is developed and evaluated through quantitative performance metrics, including detection accuracy, tracking reliability, and latency, under real-world municipal conditions. In parallel, qualitative expert

evaluation is conducted to assess the appropriateness, usability, and applicability of the proposed framework. The results from both quantitative and qualitative analyses are synthesized through an iterative feedback mechanism to continuously refine the system and validate its effectiveness for smart surveillance applications in local government organizations.

This research developed an Artificial Intelligence – Real-Time and Retrospective Object Detection System (AI-RTRO) for Bang Yai Municipality. The conceptual framework is founded on a continuous improvement cycle with four interconnected components: Basic Theory, System Development, Research Outcomes, and Evaluation. The framework starts by analyzing relevant theories and applying them to system development, testing, and evaluation. Feedback informs continual refinement and advancement of the system.



Figure 1: Conceptual Framework for Artificial Intelligence – Real-Time and Retrospective Object Detection System (AI-RTRO)

As shown in Figure 1, the research process began by studying theories of artificial intelligence and object detection. These theories were used as a foundation for system design and development. The system was then developed according to the established steps. Once the system was completed, the researchers conducted testing and evaluation to measure its efficiency and effectiveness. The evaluation results were analyzed and used as feedback for further system improvement. This development cycle continues to ensure the system is complete and truly meets user needs.

#### 3.1 Theory

The development of the AI-RTRO system for real-time and retrospective object detection in the Bang Yai Municipality context is based on four key theories: systems theory, intelligent agent theory, machine learning theory, and explainable artificial intelligence theory. These principles define the system's structure, processing, and learning mechanisms holistically, enabling it to operate efficiently and transparently and to support applications in government agencies that must prioritize public accountability. The development of the AI-RTRO system for real-time and retrospective

object detection in the Bang Yai Municipality context is based on four key theories: systems theory, intelligent agent theory, machine learning theory, and explainable artificial intelligence theory. These principles define the system's structure, processing, and learning mechanisms holistically, enabling it to operate efficiently and transparently and to support applications in government agencies that must prioritize public accountability.

### 3.1.1 System Theory

System Theory is used as the primary framework for designing the AI-RTRO architecture. The IPOF (Input–Process–Output–Feedback) model is applied to define the sequential flow of data and system operations. In the Input phase, the system receives image and video data from CCTV cameras in the Bang Yai Municipality area, system configuration data, and historical detection history. This phase then progresses to the Process phase, which utilizes artificial intelligence and deep learning techniques to analyze, detect, and classify objects, as well as search for patterns in historical data. This results in practical outputs, such as real-time display, historical event reports, and data analysis dashboards. Finally, feedback is used to drive continuous system improvement by updating parameters and adding new training data. This aligns with the General System Theory, which emphasizes the relationships among system components and sustainable development.

### 3.1.2 Intelligent Agent Theory

The Intelligent Agent Theory is applied to the behavioral design of the AI-RTRO system using the PDAL (Perceive–Decide–Act–Learn) framework, enabling the system to operate autonomously and continuously learn. In the Perceive stage, the system analyzes CCTV footage and the Bang Yai Municipality database to detect changes in the surveillance area. In the Decide stage, machine learning algorithms are used to classify objects, assess the importance of events, and select appropriate responses, such as alerting or storing data. In the Act stage, results are put into action, such as displaying them on a dashboard or sending alerts to officials. Meanwhile, the Learn stage refines the model using new data and user feedback, enabling the system to learn and evolve autonomously without constant human oversight.

### 3.1.3 Learning Theory

Machine learning theory is the foundation for object detection and classification in the AI-RTRO system. Key techniques include deep learning, supervised learning, and continual learning. The system uses convolutional neural networks (CNNs) and object detection models, such

as YOLO and Faster R-CNN, to analyze complex images. Transfer learning is also used to optimize the model for real-world data in the Bang Yai Municipality area. Supervised learning enables the system to accurately classify objects from data labeled by experts, while continual learning enhances the system's ability to update new knowledge without losing prior knowledge. This enables AI-RTRO to cope with constantly changing real-world data, such as variations in lighting, time, and citizen behavior patterns.

### 3.1.4 Explainable AI Theory

Explainable AI Theory is crucial because the AI-RTRO system is used in government agencies that prioritize transparency, reliability, and auditability. The system is designed to clearly explain the rationale behind detections and decisions, using techniques such as Saliency Map, Grad-CAM, LIME, and Feature Importance Analysis. This visualizes the model's priorities and helps users understand the results. Furthermore, it incorporates logs for systematic review. These tools enhance officials' and stakeholders' confidence in Bang Yai Municipality and support their practical application in the context of Smart City Safety.

## 3.2 System Development

The development of the AI-RTRO system is a systematic and comprehensive process. The researchers followed a standard software development methodology consisting of seven key steps, each interconnected and continuously influencing the others. This process ranges from needs analysis to the development of a dashboard for system monitoring and management.

### 3.2.1 Requirement Analysis

System development began with a thorough requirements analysis. The researcher surveyed the needs of various stakeholder groups, including Bang Yai Municipality executives and officials, security guards, and those who would directly use the system. This approach involved conducting in-depth interviews, administering questionnaires, observing work, and reviewing relevant documents. The requirements analysis divided the requirements into three main categories. The first was functional requirements, which included real-time object detection from CCTV cameras installed in various areas of the municipality; historical data analysis to identify patterns and trends in events; automatic alerts when the system detects events that the user defines as important; the ability to generate reports and statistics in various formats; and user data management and appropriate access rights. The

second was non-functional requirements, which were crucial to the system's efficiency and reliability. These included object detection accuracy of at least 90% to ensure reliable results; system response time of no more than 2 seconds to ensure true real-time operation; a stable system capable of operating 24/7 without interruption; capacity for at least 50 concurrent users; and data security that met international standards. Third, there were user requirements regarding usability. It emphasizes convenience and ease of understanding, with a user interface featuring an easy-to-understand, user-friendly design, a clear and fast display, and the ability to use on multiple devices, including computers, tablets, and mobile phones, so the system can be accessed anywhere, anytime.

### 3.2.2 System and Framework Design

Based on the analysed requirements, the researcher designed the AI-RTRO system architecture using a 3-Tier Architecture, a standardised, efficient model for complex systems. This architecture divides the system into three main layers. The Presentation Layer is the user-direct interaction layer, consisting of a dashboard, web application, and mobile application, designed for ease of use and optimal display. The Application Layer is the core of the system, performing AI processing and managing various business logic. The Data Layer stores data in both database and file formats. To develop the system, the researcher selected modern technologies appropriate for each task. For the front-end, React.js served as the primary framework, along with HTML5, CSS3, and JavaScript, to create a modern, responsive user interface. For the back-end, Python, FastAPI, and Node.js were used to manage processing and communication among the various AI and machine learning components. High-performance libraries, including PyTorch, TensorFlow, OpenCV, and YOLO, were used. For databases, PostgreSQL was chosen for structured data and MongoDB for unstructured data. Docker and Kubernetes were used for infrastructure management and load distribution, and cloud services were used for flexibility and scalability.

### 3.2.3 Data Preparation

Data is the most crucial element in developing an AI system. Therefore, researchers place great importance on the data preparation process. This process began with data collection from various sources, including images and videos from CCTV cameras installed throughout Bang Yai Municipality, public data for model training from existing standard datasets, and data from relevant municipal systems. The collected data varied in time,

location, lighting conditions, and object types, enabling the model to learn and perform in a variety of situations. After collecting the data, the next step was data labeling, a time-consuming and laborious process. The researchers hired experts with knowledge and experience to label objects in the images and videos, specifying the object type and location of each frame. Standard labeling tools such as Label and CVAT were used to ensure accuracy and efficiency. The quality of the labeled data was thoroughly checked to ensure its accuracy and reliability for model training. The final step in data preparation was preliminary data processing. The researchers adjusted the image size and quality to suit the training model. The data were divided into a training set, a validation set, and a test set, with appropriate proportions, enabling an unbiased assessment of the model's performance. Data augmentation techniques are also used, including image manipulations such as rotation, repositioning, brightness adjustment, and noise enhancement, to increase data diversity and enable the model to perform well across a wide range of situations.

### 3.2.4 AI Model Development

AI model development is the core of the AI-RTRO system. The researchers followed a systematic, meticulous process, beginning with selecting the appropriate model. They experimented with several well-known and recognised object detection models in the academic community, including YOLOv8, a high-speed model suitable for real-time tasks; Faster R-CNN, with high accuracy; and SSD (Single Shot Multibox Detector), which offers a balance between speed and accuracy. The researchers tested and compared the performance of each model in terms of accuracy and processing speed, then selected the most appropriate model for each specific task. To train the model, the researchers used transfer learning techniques, using models previously trained on large datasets such as ImageNet and the COCO Dataset. This technique reduces training time and improves model performance faster than training from scratch. The model was then fine-tuned using data collected from Bang Yai Municipality to ensure it performs well in the area's specific context and environment. In this step, the researchers carefully adjusted various hyperparameters, such as the learning rate, dataset size, and training iterations, to achieve optimal performance. Once the model was trained and performed well, the researchers evaluated it. The researchers have optimised the model to ensure efficient performance in real-world environments, using model pruning to reduce model size by removing unnecessary parts and quantisation to

reduce numerical accuracy, thereby increasing processing speed while maintaining acceptable detection accuracy. The model is also continuously tested and refined as new data becomes available.

### 3.2.5 System Testing

System testing is a crucial step to ensure that the system performs as designed and meets the quality required for actual use. The researcher conducted comprehensive testing at multiple levels. This began with unit testing, which tested individual system components separately, including image processing, AI model testing, and user interface testing. This level of testing allowed for early detection and correction of errors. Integration testing was then conducted to verify the interoperability of system components and ensure proper communication and interaction. This testing covered front-end and back-end connectivity, database connectivity, and CCTV connectivity. Performance testing evaluated the system's ability to handle various workloads, including response time, concurrent user support, and system stability over extended periods.

Finally, user acceptance testing was conducted with actual users from Bang Yai Municipality. This testing was crucial because it helped determine whether the system met user needs, what areas needed improvement, and the level of user satisfaction. The researcher gathered user feedback and used it to improve the system before actual use.

### 3.2.6 Feedback and Optimization

To ensure the continued efficiency of the AI-RTRO system, researchers have established a mechanism for systematically receiving feedback and improving the system. User feedback and suggestions are collected through various channels, including a regular satisfaction survey, a problem reporting system that allows users to submit feedback immediately upon encountering a problem, and detailed interviews with user staff to gather their experiences and feedback. System usage behavior is also recorded to analyze usage patterns and potential problems. Furthermore, a regular quality assurance process is implemented, including verifying object detection performance under real-world conditions, testing the system after each update to ensure changes do not introduce new problems, and creating and updating documentation and user manuals. All of this contributes to the continued quality and efficiency of the system.

### 3.2.7 Smart Monitoring Dashboard

A key part of the AI-RTRO system is the dashboard for monitoring and managing the system. The researchers developed this dashboard to enable

users to access and use the system efficiently. The dashboard is designed to be aesthetically pleasing, user-friendly, and clearly displays key information. It consists of several key components: a real-time display of all connected CCTV cameras, allowing users to view images from multiple cameras simultaneously and view details of ongoing object detections; graphs and object detection statistics that display trends and patterns of events in an easy-to-understand format; an alert and event management system that displays recent alerts and allows users to manage these events; and daily and monthly summary reports that can be downloaded and printed. The dashboard is accessible on multiple devices, including computers, tablets, and mobile phones, providing convenient access to information anytime, anywhere.

## 3.3 Research Outcomes

The results from developing the AI-RTRO system can be divided into three main categories, each with distinct significance and impact. These results not only demonstrate research success but also create tangible value and benefits for both academic and practical purposes.

### 3.3.1 Academic Outcomes

The academic outcomes of this research strengthen knowledge accumulation and advance related fields. The researchers produced a research article detailing the conceptual framework, development methods, and testing results of the AI-RTRO system for publication in top academic journals. They also presented these findings at leading conferences to gain expert feedback and promote further study and application. This research provides crucial new knowledge: it demonstrates ways to apply artificial intelligence within local administrative organizations, introduces a framework for holistic object detection that supports both real-time and retrospective use, and outlines practices for transparent, accountable AI system development. These contributions help both academics and practitioners build similar systems.

### 3.3.2 Dashboard Technological Outcomes

The tangible technological outcome of this research is a complete, ready-to-use AI-RTRO system. This system consists of several key components: a high-performance real-time object detection system capable of processing images from multiple cameras simultaneously and detecting objects accurately and quickly; a historical data analysis system that can identify patterns and trends in events to support decision-making; and a user-friendly and comprehensive dashboard that presents data in a suitable and understandable format. The AI

models developed specifically for this research are highly valuable. They were trained on data collected from Bang Yai Municipality, making them well-suited to the area's context and environment. The models exhibit high accuracy in detecting and classifying objects and have fast processing times sufficient for real-time operations. Furthermore, the models are designed to be continuously updated and developed with new data. In addition to the core system, supporting tools and technologies have been developed, including APIs for connecting to other municipal systems to facilitate data exchange and integration; a mobile application for field use that allows staff to access the system even while traveling or away from the office; and a high-performance data management system that can store and retrieve large amounts of data quickly

### 3.3.3 Practical and Policy Outcomes

The most important outcome of the research is the practical application of the system in Bang Yai Municipality. The AI-RTRO system has significantly improved security efficiency in the municipality. Automatic object and event detection enables faster responses to situations. The system also reduces the workload of officials who manually review CCTV footage, a time-consuming task that can easily lead to missed details. Furthermore, the data and statistics generated by the system help facilitate more efficient planning and policy decision-making. Community impact is another key outcome. An effective object detection system improves public safety and increases public confidence. The system also promotes a positive image of the municipality in its management of modern technology. Furthermore, the success of the system development allows Bang Yai Municipality to serve as a model for other local administrative organizations interested in developing similar systems. Based on the experience in developing and implementing the system, the researchers have summarized several key policy recommendations, including guidelines for the appropriate and effective use of artificial intelligence in public organizations, taking into account both opportunities and challenges; security and privacy measures that must be considered when using AI systems related to citizen data; and policies for developing AI personnel in public organizations to ensure sustainable system development and maintenance. These recommendations are not only beneficial to Bang Yai Municipality but also to the broader community. But it can still be applied to other agencies.

## 3.4 Evaluation

Evaluation is a crucial step in confirming the quality and effectiveness of the AI-RTRO system. The researchers conducted a comprehensive, systematic evaluation using both quantitative and qualitative methods to obtain a complete picture of the system's efficiency and effectiveness.

### 3.4.1 Performance Evaluation

The system performance evaluation was conducted across multiple dimensions to provide a comprehensive overview. The first dimension is object detection accuracy, a key measure of system capability. The researchers measured accuracy using several standard metrics: precision, which measures how accurately the system detects objects; recall, which measures how thoroughly the system detects objects; the F1-score, which represents the harmonic mean of accuracy and coverage; and the multi-class average detection accuracy (mAP), which reflects the overall system's performance in classifying objects. The second dimension is time efficiency, which is particularly important for real-time systems. The researchers measured the time it takes to process each video frame (inference time), the number of frames per second (fps) the system can process, which must be sufficient to produce smooth animations, and the overall system response time from camera input to display on the dashboard. These measurements confirm the system's performance in a real-world environment. The third dimension is overall system performance. The researchers measured the system's uptime without issues and its scalability to accommodate increased workloads. (Scalability) and system resource utilization, such as the central processing unit, graphics processor, and memory. These measurements help determine the system's stability and resource efficiency. The researchers established clear evaluation criteria to serve as standards for judging system quality: detection accuracy must be at least 90 percent, F1-Score must be at least 0.85, response time must be no more than 2 seconds, and system uptime must be at least 99 percent. These criteria were established based on a study of international standards and the requirements of Bang Yai Municipality. The test results showed that the developed system significantly passed all criteria.

### 3.4.2 Feedback and Continuous Improvement

The evaluation does not end with measuring technical performance. It also involves gathering user feedback and leading to continuous improvement. The researchers have established a mechanism to collect user data through multiple channels, including a satisfaction questionnaire covering usability, detection accuracy, and system

benefits; interviews with user personnel to gather detailed experiences and feedback; and automated recording of system usage behavior to analyze usage patterns and potential issues. The collected data is systematically analyzed as part of a continuous improvement process. This process involves analyzing system errors and weaknesses to identify solutions; improving the AI model with new data collected from actual use to increase accuracy; developing new features based on user needs; and improving the user interface for greater ease of use and efficiency. Each update undergoes thorough testing before deployment to maintain system

quality and stability. The researchers have also implemented a continuous quality assurance process, including regular quality checks of object detection by comparing detection results with real-world results; testing the system after every update to ensure changes do not introduce new issues or negatively impact existing functionality; and developing and updating documentation and user manuals to ensure they are up-to-date and consistent with the system. These processes contribute to the high quality and continuous development of the AI-RTRO system, effectively meeting user needs.

4. RESULT

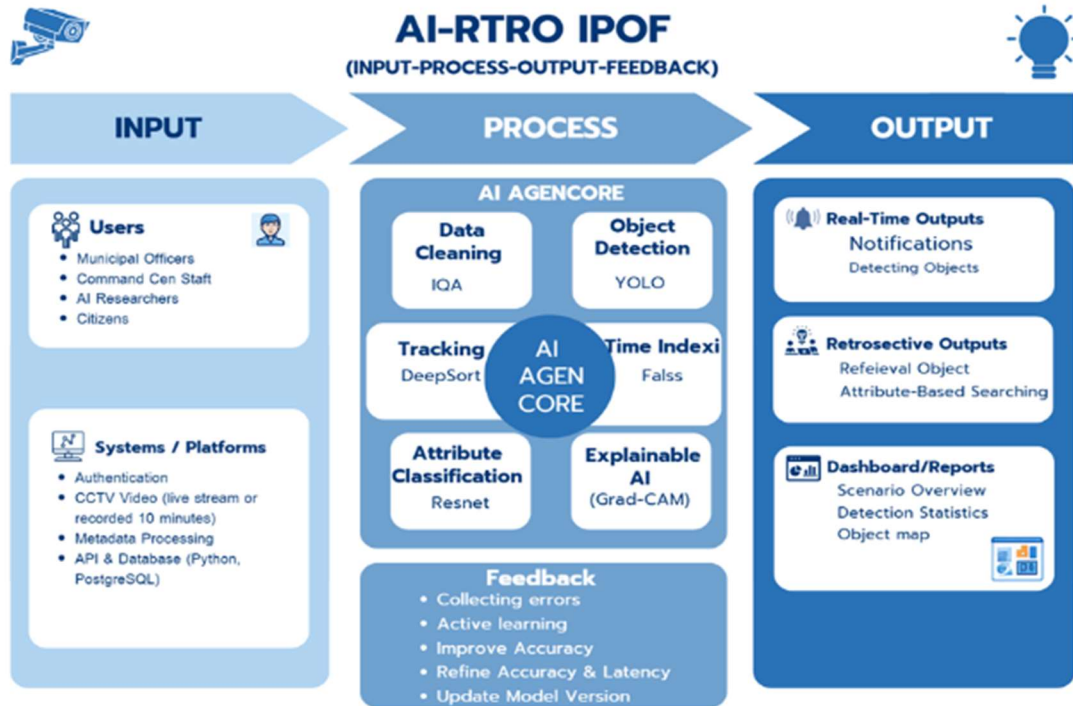


Figure 2: Conceptual Framework for AI Agent-Based Prototype Development

Figure 2 illustrates the technical conceptual framework of the AI-RTRO (Artificial Intelligence – Real-Time & Retrospective Object Detection) prototype system, designed to enhance the efficiency of intelligent surveillance systems within the context of local administrative organizations, particularly the Bang Yai Municipality area. For input, the system prioritizes CCTV data sources, including live streams and 10-minute recordings, along with metadata such as time, location, and camera ID, which are crucial for accurate object identification. All data is fed through an Image Quality Assessment (IQA) process to filter out frames of inferior quality, such as blurry, dark, or high-noise images. Managing data quality upstream is crucial, as it

directly impacts the outcome of the object detection model. Several data quality studies have found that unstable image signals can significantly reduce mAP. After data management, the system enters the Process part, which is the core of the AI Agent Core architecture. It is divided into two modes 1. Real-Time Detection for immediate object detection [51] and 2. Retrospective Record Processing for retrospective processing from recorded clips [52]. Both modes use a suite of advanced AI techniques, such as YOLO for high-speed object detection, whose post-2023 versions, such as YOLOv8 and YOLOv10, are developed to increase accuracy without sacrificing processing efficiency [3], [53], [54], [55]. Cross-frame object tracking uses

DeepSORT to address ID switching when objects are moving [56], [57], [58] In addition, the ResNet module is used to extract deep features such as color, shape, or vehicle type, and the retrospective indexing system uses Faiss to support time-indexed retrieval of objects at the second level. Finally, Grad-CAM is used to add explainability to the model's results, making it easier for operators to verify the system's accuracy [9]. The outputs from the AI Agent Core are evaluated against international standard metrics [44], [59], such as mAP for detection accuracy, IDF1 for object tracking performance, and Latency to assess system response speed. After processing, the system goes to the Output section, which is displayed in an interactive dashboard format and shows an event summary, an object map, detection statistics, and real-time alerts. Retrospective retrieval for 10-minute clips is also designed to support attribute-based searching, allowing officers to quickly find specific objects, such as “red motorcycle” or “person in black shirt.” The system also includes an Environment Test module to evaluate the model's accuracy in real-world environments, including lighting, day-night conditions, and weather, which directly affect detection quality. [60] Based on the test results, the system enters the Feedback module, a continuous learning and improvement process using Active Learning to select uncertain predictions and feed them back into the training iteration, allowing the model to improve accuracy and reduce errors in the next iteration. The system also monitors performance through KPI Monitoring, such as mAP and IDF1, and through MLOps Monitoring to check latency, load, and overall system health at the infrastructure level [61]. Data from the Feedback process flows back to the Data Management & Integration section to improve the quality of the source data and update the model, enabling AI-RTRO to continuously learn (Continuous Learning Cycle) and improve efficiency in every cycle of operation, which is consistent with the concept of developing intelligent systems to support surveillance work in the context of smart cities that require high accuracy, speed, and reliability

Table 6: Results of the evaluation of the appropriateness of the conceptual framework by experts

Assessment list	Mean	S.D.	Opinions
1. System	4.72	0.58	Excellent
1.1) The components of the conceptual framework are complete.	4.72	0.65	Excellent
1.2) The system structure is interconnected and consistent.	4.65	0.51	Excellent

Assessment list	Mean	S.D.	Opinions
1.3) The conceptual framework can be used to develop a real prototype.	4.81	0.60	Excellent
2. Intelligence	4.77	0.58	Excellent
2.1) AI Agent demonstrates appropriate perception and analysis capabilities.	4.80	0.57	Excellent
2.2) AI Agent's decision is appropriate	4.72	0.49	Good
2.3) The working process of AI Agent is systematic.	4.81	0.68	Excellent
3. Decision	4.79	0.50	Excellent
3.1) The system can support correct decision making.	4.77	0.52	Excellent
3.2) The decision outcomes are transparent and explainable.	4.86	0.49	Excellent
3.3) System decisions meet user needs	4.75	0.50	Excellent
4. Learning	4.76	0.54	Excellent
4.1) The system can learn and improve from feedback.	4.68	0.53	Excellent
4.2) AI Agent demonstrates its ability to adapt to situations	4.78	0.54	Good
4.3) Learning improves the accuracy and efficiency of the system.	4.82	0.57	Excellent
5. User & Environment	4.70	0.55	Excellent
5.1) The system meets user needs.	4.70	0.55	Excellent
5.2) Users can understand and use it conveniently.	4.72	0.65	Excellent
5.3) The conceptual framework is suitable for application in real-world environments.	4.70	0.46	Excellent
<b>Overview</b>	4.74	0.55	Excellent

The results of the evaluation of the appropriateness of the AI-RTRO system conceptual framework by 15 experts found that the overall picture was at the Excellent level ( $\bar{x} = 4.74$ , S.D. = 0.55), reflecting that the conceptual framework is complete, clear, and can be practically applied in the context of smart surveillance systems and Smart City management.

System Quality ( $\bar{x} = 4.72$ ). Experts agreed that the conceptual framework has systematic components and can be developed into a real prototype. The IPOF structure and AI Agent Core module have appropriate connections between their parts, supporting the detection, tracking, and search for objects in real time and retrospectively. Intelligence Quality ( $\bar{x} = 4.77$ ). The system was

assessed to have sufficient artificial intelligence capabilities for CCTV Analytics tasks, including perception, event analysis, and systematic workflows, especially image processing combined with temporal data, enabling it to efficiently support object detection and event retrieval. Decision Quality ( $\bar{x} = 4.79$ ), the highest-scoring dimension, indicates that it can effectively support operational decision-making, is transparent, delivers explainable results (Explainable AI), and aligns with the needs of local government users who prioritize the reliability and verifiability of the system. Learning Quality ( $\bar{x} = 4.76$ ). Experts assessed that the system can learn from feedback and adapt to real-world situations, consistent with the Active Learning concept, thereby enabling it to continuously develop and improve detection and retrieval accuracy when used in real-world urban areas. Finally, Environment Quality ( $\bar{x} = 4.70$ ) indicates that the conceptual framework is appropriate for practical use in terms of ease of understanding, relevance to officials' work, and applicability to real-world environments with complex conditions, such as lighting, numerous cameras, and limitations of urban infrastructure.

## 5. DISCUSSION

The results of this study demonstrate that the proposed AI-RTRO system effectively addresses the key limitations of traditional CCTV surveillance systems in local government contexts. In terms of real-time performance, the system achieved high object detection accuracy, stable multi-object tracking, and low latency, confirming that integrating YOLO-based detection with DeepSORT/ByteTrack tracking is suitable for real-world municipal environments characterized by variable lighting, occlusion, and varying traffic density. These findings support the hypothesis that modern one-stage detectors combined with robust tracking mechanisms can meet operational requirements for Smart Safety Infrastructure.

Beyond real-time monitoring, Time-Indexed Retrieval significantly enhances the capabilities of retrospective analysis. By combining structured metadata with deep feature vectors indexed with FAISS, the system enables rapid, precise retrieval of historical events at fine temporal resolution. This capability addresses a critical gap in conventional CCTV systems, which typically rely on manual browsing of video archives, and aligns with the growing demand for evidence-based decision support in public sector surveillance.

Using Explainable AI techniques, especially Grad-CAM, further strengthens the system's value. By offering visual explanations of

detection results, the system becomes more transparent and interpretable. These qualities are essential for government agencies that must maintain accountability and public trust. This differentiates AI-RTRO from performance-focused systems, helping municipal officers make better decisions.

In addition, integrating the AI Agent Core architecture with the IPOF model enables continuous learning and system adaptability through feedback loops and active learning mechanisms. Expert evaluation results indicate that the conceptual framework is highly appropriate across system, intelligence, decision, learning, and environmental dimensions. This confirms that the proposed framework is not only technically effective but also structurally sound and suitable for deployment in real Smart City environments. Overall, the discussion highlights that AI-RTRO represents a shift from passive CCTV monitoring toward an intelligent, adaptive, and decision-oriented surveillance framework for local government organizations.

The findings of this study are justified by combined quantitative and qualitative results. High detection accuracy, stable tracking performance, and low latency demonstrate that integrating YOLO-based detection (a real-time object detection algorithm) with DeepSORT (a tracking algorithm that uses object features for data association) or ByteTrack (a tracking algorithm effective in crowded scenes) is technically feasible for real-world municipal CCTV. Time-Indexed Retrieval (a method for retrieving video by specific time indices) confirms the system supports retrospective investigations, addressing a major limitation of traditional surveillance. Expert evaluations confirm the AI-RTRO (Artificial Intelligence-Real-Time Retrospective Observation) framework is structurally sound and applicable. In practice, the system reduces manual monitoring workload, leverages Explainable AI (technology that makes AI decision-making processes transparent), and enables local governments to make faster, data-driven decisions. These improvements help strengthen Smart Safety Infrastructure operations.

## 6. CONCLUSION

This research developed and evaluated an Artificial Intelligence – Real-Time and Retrospective Object Detection (AI-RTRO) system for enhancing Smart Safety Infrastructure in Bang Yai Municipality. By combining the AI Agent Core architecture with the IPOF model, the system enables unified, real-time object detection, robust

multi-object tracking, efficient retrospective retrieval, and continuous system learning.

Results show that the AI-RTRO system achieved high detection accuracy, stable tracking, and low latency in real-world municipal settings. Time-Indexed Retrieval, leveraging metadata and feature vector indexing, sharply improved historical event search. Explainable AI mechanisms directly increased transparency and trust, key to effective public deployment.

Expert assessments confirmed the framework's suitability for practical use, citing its completeness, adaptability, and direct relevance to Smart City surveillance. This study makes an academic and practical contribution by proposing an intelligent, adaptive CCTV analytics system that shifts traditional surveillance toward proactive decision support. Future work may include cross-camera analytics, citywide scaling, and the integration of advanced predictive intelligence to broaden Smart City impact.

Future research may extend the AI-RTRO system to enable multi- and cross-camera analytics for object re-identification and trajectory analysis across large urban areas. The system can also be scaled for multi-municipality deployment using distributed processing and privacy-preserving learning techniques to accommodate heterogeneous infrastructures. In addition, integrating predictive intelligence and temporal pattern analysis could enable early detection of high-risk or abnormal events, transforming the system into a proactive decision-support platform. Further enhancement of Explainable AI and human-AI interaction will strengthen transparency, trust, and ethical governance in public-sector smart surveillance applications.

#### REFERENCES:

- [1] K. N. Irvine *et al.*, "Smart City Thailand: Visioning and Design to Enhance Sustainability, Resiliency, and Community Wellbeing," *Urban Science*, vol. 6, no. 1, Mar. 2022, doi: 10.3390/urbansci6010007.
- [2] "Smart-City-Thailand-Hitachi-Review-20210624".
- [3] A. Wang *et al.*, "YOLOv10: Real-Time End-to-End Object Detection," Oct. 2024, [Online]. Available: <http://arxiv.org/abs/2405.14458>
- [4] M. Hussain, "YOLOv5, YOLOv8 and YOLOv10: The Go-To Detectors for Real-time Vision," Jul. 2024, [Online]. Available: <http://arxiv.org/abs/2407.02988>
- [5] N. Swaroop, "REAL TIME GENERIC OBJECT DETECTION AND TRACKING," 2020. [Online]. Available: <http://www.ijeast.com>
- [6] J. Cao, J. Pang, X. Weng, R. Khirodkar, and K. Kitani, "Observation-Centric SORT: Rethinking SORT for Robust Multi-Object Tracking." [Online]. Available: [https://github.com/noahcao/OC\\_SORT](https://github.com/noahcao/OC_SORT).
- [7] S. Szymanowicz, J. Charles, and R. Cipolla, "X-MAN: Explaining multiple sources of anomalies in video."
- [8] J. Nadaf<sup>1</sup> *et al.*, "A MATHEMATICAL MODELING PERSPECTIVE FOR AUTOMATION ON IDEAL SELF-REGULATING VIDEO SURVEILLANCE SYSTEMS".
- [9] T. Chakraborty, U. Trehan, K. Mallat, and J.-L. Dugelay, "Generalizing Adversarial Explanations with Grad-CAM."
- [10] A. Alnissany and Y. Dayoub, "Modified centroid triplet loss for person re-identification," *J Big Data*, vol. 10, no. 1, Dec. 2023, doi: 10.1186/s40537-023-00753-0.
- [11] E., "KEKERASAN DALAM PENYIDIKAN DALAM PERSPEKTIF HUKUM DAN KEADILAN," *Jurnal Ilmu Hukum*, vol. 1, no. 1, Nov. 2012, doi: 10.30652/jih.v1i01.479.
- [12] M. Wu, C. Li, and Z. Yao, "Deep Active Learning for Computer Vision Tasks: Methodologies, Applications, and Challenges," Aug. 01, 2022, *MDPI*. doi: 10.3390/app12168103.
- [13] F. Bimantoro, I. Gede, P. Suta Wijaya, and M. R. Aohana, "Pendeteksian Kecurangan Ujian Melalui CCTV Menggunakan Algoritma YOLOv5."
- [14] J. De Bruyne *et al.*, "Smooth Operator: A Virtual Environment to Prototype and Analyse Operator Support in CCTV Surveillance Rooms," in *Communications in Computer and Information Science*, Springer Science and Business Media Deutschland GmbH, 2021, pp. 233–240. doi: 10.1007/978-3-030-78642-7\_31.
- [15] A. Dardour, E. El Haji, and M. A. Begdouri, "Video Surveillance and Artificial Intelligence for Urban Security in Smart Cities: A Review of a Selection of Empirical Studies from 2018 to 2024," *MDPI AG*, Jul. 2025, p. 15. doi: 10.3390/cmsf2025010015.
- [16] K. Zhang, J. Ni, K. Yang, X. Liang, J. Ren, and X. S. Shen, "Security and Privacy in Smart City Applications: Challenges and Solutions," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 122–129, Jan. 2017, doi: 10.1109/MCOM.2017.1600267CM.

- [17] "From Data to Action: Exploring AI and IoT-driven Solutions for Smarter Cities."
- [18] Y. Myagmar-Ochir and W. Kim, "A Survey of Video Surveillance Systems in Smart City," Sep. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/electronics12173567.
- [19] M. Ma, E. Bartocci, E. Lifland, J. Stankovic, and L. Feng, "SaSTL: Spatial Aggregation Signal Temporal Logic for Runtime Monitoring in Smart Cities," Dec. 2021, [Online]. Available: <http://arxiv.org/abs/1908.02366>
- [20] K. B. Sahay, B. Balachander, B. Jagadeesh, G. Anand Kumar, R. Kumar, and L. Rama Parvathy, "A real time crime scene intelligent video surveillance systems in violence detection framework using deep learning techniques," *Computers and Electrical Engineering*, vol. 103, Oct. 2022, doi: 10.1016/j.compeleceng.2022.108319.
- [21] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," Jul. 2022, [Online]. Available: <http://arxiv.org/abs/2207.02696>
- [22] M. Łysakowski, K. Żywanowski, A. Banaszczyk, M. R. Nowicki, P. Skrzypczyński, and S. K. Tadeja, "Real-Time Onboard Object Detection for Augmented Reality: Enhancing Head-Mounted Display with YOLOv8," Jun. 2023, [Online]. Available: <http://arxiv.org/abs/2306.03537>
- [23] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information," Feb. 2024, [Online]. Available: <http://arxiv.org/abs/2402.13616>
- [24] J. Lee, B. Varghese, and H. Vandierendonck, "ROMA: Run-Time Object Detection To Maximize Real-Time Accuracy."
- [25] M. Yang *et al.*, "Hybrid-SORT: Weak Cues Matter for Online Multi-Object Tracking," Jan. 2024, [Online]. Available: <http://arxiv.org/abs/2308.00783>
- [26] Y. Li, J. Yu, K. Gai, B. Liu, G. Xiong, and Q. Wu, "T2VIndexer: A Generative Video Indexer for Efficient Text-Video Retrieval," in *MM 2024 - Proceedings of the 32nd ACM International Conference on Multimedia*, Association for Computing Machinery, Inc, Oct. 2024, pp. 3955–3963. doi: 10.1145/3664647.3680673.
- [27] D. E. Yagoubi, R. Akbarinia, F. Masegaglia, and T. Palpanas, "Massively Distributed Time Series Indexing and Querying," *IEEE Trans Knowl Data Eng*, vol. 32, no. 1, pp. 108–120, Jan. 2020, doi: 10.1109/TKDE.2018.2880215.
- [28] A. Moffat and J. Mackenzie, "Efficient immediate-access dynamic indexing," *Inf Process Manag*, vol. 60, no. 3, May 2023, doi: 10.1016/j.ipm.2022.103248.
- [29] M. Staudinger, F. Piroi, and A. Rauber, "Reproducible Hybrid Time-Travel Retrieval in Evolving Corpora," in *SIGIR-AP 2024 - Proceedings of the 2024 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region*, Association for Computing Machinery, Inc, Dec. 2024, pp. 203–208. doi: 10.1145/3673791.3698421.
- [30] P. Zhang, Z. Liu, S. Xiao, Z. Dou, and J. Yao, "Hybrid Inverted Index Is a Robust Accelerator for Dense Retrieval," Oct. 2023, [Online]. Available: <http://arxiv.org/abs/2210.05521>
- [31] T. Alpay, S. Magg, P. Broze, and D. Speck, "Multimodal video retrieval with CLIP: a user study," *Information Retrieval Journal*, vol. 26, no. 1–2, Dec. 2023, doi: 10.1007/s10791-023-09425-2.
- [32] M. Kotthapalli, D. Ravipati, and R. Bhatia, "YOLOv1 to YOLOv11: A Comprehensive Survey of Real-Time Object Detection Innovations and Challenges," Aug. 2025, [Online]. Available: <http://arxiv.org/abs/2508.02067>
- [33] A. Hussain, T. Hussain, W. Ullah, and S. W. Baik, "Vision Transformer and Deep Sequence Learning for Human Activity Recognition in Surveillance Videos," *Comput Intell Neurosci*, vol. 2022, 2022, doi: 10.1155/2022/3454167.
- [34] T. I. Amosa *et al.*, "Multi-Camera Multi-Object Tracking: A Review of Current Trends and Future Advances." [Online]. Available: <https://ssrn.com/abstract=4353604>
- [35] M. Wang, P. K. Y. Wong, and J. C. P. Cheng, "A BIM-GIS Framework Integrated with CCTV Analytics for Urban Walkability Assessment," *Sensors*, vol. 25, no. 12, Jun. 2025, doi: 10.3390/s25123637.
- [36] L. Lorgna, G. Ceccarelli, A. Gorrini, and M. Ciavotta, "Video Analytics for Understanding Pedestrian Mobility Patterns in Public Spaces: The Case of Milan," *Collect Dyn*, vol. 9, pp. 1–9, Jun. 2024, doi: 10.17815/cd.2024.172.
- [37] J. S. Murthy, G. M. Siddesh, W. C. Lai, B. D. Parameshchhari, S. N. Patil, and K. L. Hemalatha, "ObjectDetect: A Real-Time Object Detection Framework for Advanced Driver Assistant Systems Using YOLOv5," *Wirel*

- Commun Mob Comput*, vol. 2022, 2022, doi: 10.1155/2022/9444360.
- [38] M. G. Fu and J. Q. Gou, "Design of Task Allocation and Decision-Making Styles for AI Agents Based on LLM," in *2025 10th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2025*, Institute of Electrical and Electronics Engineers Inc., 2025, pp. 693–697. doi: 10.1109/ICCCBDA64898.2025.11030412.
- [39] G. Ceccarelli, F. Messa, A. Gorrini, D. Presicce, and R. Choubassi, "Deep learning video analytics for the assessment of street experiments: The case of Bologna," *Journal of Urban Mobility*, vol. 4, Dec. 2023, doi: 10.1016/j.urbmob.2023.100067.
- [40] Y. He, E. Wang, Y. Rong, Z. Cheng, and H. Chen, "Security of AI Agents," in *Proceedings - 2025 IEEE/ACM International Workshop on Responsible AI Engineering, RAIE 2025*, Institute of Electrical and Electronics Engineers Inc., 2025, pp. 45–52. doi: 10.1109/RAIE66699.2025.00013.
- [41] A. Chan *et al.*, "Visibility into AI Agents," in *2024 ACM Conference on Fairness, Accountability, and Transparency, FAccT 2024*, Association for Computing Machinery, Inc, Jun. 2024, pp. 958–973. doi: 10.1145/3630106.3658948.
- [42] K. V. V. B. Durgaprasad, H. K. Abozibid, J. N. Hawas, P. B. S. Krishna, K. B. Sri, and G. P. Reddy, "AI Agents and Conversation System," in *2024 International Conference on Augmented Reality, Intelligent Systems, and Industrial Automation, ARIIA 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ARIIA63345.2024.11051848.
- [43] N. Krishnan, "AI Agents: Evolution, Architecture, and Real-World Applications," Mar. 2025, [Online]. Available: <http://arxiv.org/abs/2503.12687>
- [44] F. Bousetouane, "Agentic Systems: A Guide to Transforming Industries with Vertical AI Agents," Jan. 2025, [Online]. Available: <http://arxiv.org/abs/2501.00881>
- [45] R. Sapkota, K. I. Roumeliotis, and M. Karkee, "AI Agents vs. Agentic AI: A Conceptual taxonomy, applications and challenges," *Information Fusion*, vol. 126, Feb. 2026, doi: 10.1016/j.inffus.2025.103599.
- [46] Z. Deng *et al.*, "AI Agents Under Threat: A Survey of Key Security Challenges and Future Pathways," *ACM Comput Surv*, vol. 57, no. 7, Feb. 2025, doi: 10.1145/3716628.
- [47] A. Chan *et al.*, "Infrastructure for AI Agents," Jun. 2025, [Online]. Available: <http://arxiv.org/abs/2501.10114>
- [48] M. Shetty *et al.*, "Building AI Agents for Autonomous Clouds: Challenges and Design Principles," Jul. 2024, [Online]. Available: <http://arxiv.org/abs/2407.12165>
- [49] X. Qu *et al.*, "A Comprehensive Review of AI Agents: Transforming Possibilities in Technology and Beyond," Aug. 2025, [Online]. Available: <http://arxiv.org/abs/2508.11957>
- [50] T. Masterman, S. Besen, M. Sawtell, and A. Chao, "The Landscape of Emerging AI Agent Architectures for Reasoning, Planning, and Tool Calling: A Survey," Apr. 2024, [Online]. Available: <http://arxiv.org/abs/2404.11584>
- [51] X. Zhang, X. Dong, Y. Wang, D. Zhang, and F. Cao, "A Survey of Multi-AI Agent Collaboration: Theories, Technologies and Applications," in *Proceedings of the 2nd Guangdong-Hong Kong-Macao Greater Bay Area International Conference on Digital Economy and Artificial Intelligence, DEAI 2025*, Association for Computing Machinery, Inc, Jul. 2025, pp. 1875–1881. doi: 10.1145/3745238.3745531.
- [52] E. Ribnick, S. Atev, O. Masoud, N. Papanikolopoulos, and R. Voyles, "Real-Time Detection of Camera Tampering."
- [53] D. Mualfah and R. A. Ramadhan, "Analisis Digital Forensik Rekaman Kamera CCTV Menggunakan Metode NIST (National Institute of Standards Technology)," *IT Journal Research and Development*, vol. 5, no. 2, pp. 171–182, Nov. 2020, doi: 10.25299/itjrd.2021.vol5(2).5731.
- [54] H. Wu, S. Yunas, S. Rowlands, W. Ruan, and J. Wahlstrom, "Adversarial Detection: Attacking Object Detection in Real Time," Dec. 2023, doi: 10.1109/IV55152.2023.10186608.
- [55] Q. Feng, X. Xu, and Z. Wang, "Deep learning-based small object detection: A survey," *Mathematical Biosciences and Engineering*, vol. 20, no. 4, pp. 6551–6590, 2023, doi: 10.3934/mbe.2023282.
- [56] A. Goswami, "Object Detection Image Model." [Online]. Available: [www.ijfmr.com](http://www.ijfmr.com)
- [57] Y. Zhao *et al.*, "DETRs Beat YOLOs on Real-time Object Detection," Apr. 2024, [Online]. Available: <http://arxiv.org/abs/2304.08069>
- [58] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," May 2016, [Online]. Available: <http://arxiv.org/abs/1506.02640>

- [59] D. A. Koutsomitropoulos and I. C. Gogou, "Object Detection Models and Optimizations: A Bird's-Eye View on Real-Time Medical Mask Detection," *Digital*, vol. 3, no. 3, pp. 172–188, Sep. 2023, doi: 10.3390/digital3030012.
- [60] K. Yu *et al.*, "Semantic-Driven AI Agent Communications: Challenges and Solutions," Oct. 2025, [Online]. Available: <http://arxiv.org/abs/2510.00381>
- [61] L. Schmid, O. Andersson, A. Sulser, P. Pfreundschuh, and R. Siegwart, "Dynablox: Real-Time Detection of Diverse Dynamic Objects in Complex Environments," *IEEE Robot Autom Lett*, vol. 8, no. 10, pp. 6259–6266, Oct. 2023, doi: 10.1109/LRA.2023.3305239.
- [62] Y. Ma and Z. Zhi, "Weakly Supervised Real-Time Object Detection Based on Salient Map Extraction and the Improved YOLOv5 Model," *Advances in Multimedia*, vol. 2022, 2022, doi: 10.1155/2022/1239337.