© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1718175

Why Every City Needs an AI Scam Shield

ROHIT RAJDEV
Sandscript AI

Abstract- ***A growing problem across the globe is the digital age-old dictating that modern technological developments have enabled: the use of advanced AI techniques like deepfake technology, large language models, real-time speech synthesis, and auto-phishing systems is heavily used to defraud private individuals, businesses, and public institutions in larger proportions than ever before. Traditional cybersecurity tools focused on preventing threats at the network perimeter or through rules are structurally weak to fight this attack. The author contends that all cities no matter their size, digital readiness, or economic condition should have their own AI Scam Shield, an intelligent cross-sectorial, real-time threat recognition, behavioural analysis, and community protection system designed on a city scale to thwart AI scams before they can harm citizens. A systematic synthesis of literature covering the field of fraud detection using AI, smart city security architecture, protection of the elderly population, phishing protection using Blockchains, and innovative regulations, revealed that modern protections are not adequate. The study outlines a complete framework of how to implement municipal AI Scam Shield. The results show integrated AI Scam Shield systems to be more accurate at detecting phishing frauds with an accuracy of 96.4%, have less delay in real-time fraud alerts of more than 99% as compared to machine-learning-driven systems, and were able to intercept elderly-targeted phishing attempts almost double, at 85.4% versus 40.9%. Finally, the article details the governance framework and policy roadmap for the implementation of AI Scam Shield at the City level, highlighting governance principles of equity, transparency, and civil rights protection.***

Keywords: AI Scam Shield, Urban Cybersecurity, Fraud Detection, Smart Cities, Elderly Protection, Deepfake Detection, Phishing Countermeasures, Digital Identity, Real-Time Threat Intelligence, Ai Governance

I. INTRODUCTION

1.1 The Scam Epidemic in Urban Digital Environments

It's the modern city, which at the same time grants the best opportunities for economic growth but offers the greatest opportunity for the proliferation of digital fraud. The more that urban citizens move the money, socialize, seek medical care, and communicate about government services onto digital platforms, the greater attack surface is left for scammers. Scam techniques have been transformed by AI, such as through deep fake voice cloning of well-known individuals, the generation of fraudulent phishing text using generative language models (LLMs), automated vishing attacks, and AI-influenced investment fraud campaigns. Scams that used to rely on masterful face-to-face handlings can now be conducted at an industrial level with only minimal extra costs, targeting thousands of residents one after another with highly personalised deceptions (Herrera et al., 2024; Al Siam et al., 2025).

The losses are tremendous. Elderly individuals, with lower digital literacy and larger amounts of wealth and social isolation, seem to be the most affected group by the scams, increased by AI tools. AI voice generation, simulating empathy, and personalised deception scripts for criminals are examples of the same technologies being used to both facilitate the fraud and deter it, as Sima et al. (2026) point out. However, if there is no systematic city-to-city infrastructure to implement the same protection measures, the disparity between scammers and potential victims becomes even more significant. The digital divide is not closing but is getting used as a sword to serve as a weapon.

The financial institutions have expanded their sophisticated fraud detection mechanisms, but these are still confined within sectoral silos that are easily exploited by scammers. A fraudster who cannot break through the AI detection system of a bank could be successful at a telecom line, social media or a government service imposition attempt on the same consumer. There is no single player that has visibility into data across different sectors to detect these multi-vector attacks. There are gaps that can only be closed by an AI Scam Shield that integrates financial,

telecommunications, digital identity, and social media threat intelligence, as these gaps will be left open otherwise (Bibi & Badi, 2023; Pydipala, 2023).

1.2 The arguments behind the integration of AI at the city level

The smart city paradigm encompasses the IoT sensor technology, digital service platforms, and data analysis that are implemented into governance; it has the potential to introduce both new vulnerabilities and new protective opportunities for the city. As stated by Soni and Taneja (2026), the same smart infrastructure which enables smart city functions such as traffic optimisation, energy distribution, and improved public service delivery can gather personal data at massive amounts, making it a high value target for cybercriminals. That's not a prudent argument for embedding AI powered protection in the heart of a smart city, it's a necessary one.

Sun et al. (2026) develop a regulatory framework perspective and argue that going further than such approach by following a direction of responsive regulation, which involves making use of algorithmic regulatory tools to anticipate, detect, and disrupt criminal behavior before it occurs and does harm in the densely populated urban space, offers the most promising path to combat AI-enhancing fraud in such contexts. The proposed AI Scam Shield architecture in this article supports the regulatory ideas in this pyramid model, which has been validated empirically in the Chinese and Southeast Asian markets. Table 1 below places the AI Scam Shield need into city typology, highlighting the universal exposure of city residents to scams, and the differences in current readiness.

Table 1. Urban Scam Threat Landscape by City Category and AI Shield Readiness

City Category	Primary Scam Vectors	Vulnerable Population	Avg. Annual Loss (USD)	AI Shield Readiness
Mega-City (>10M)	Deepfake voice, investment fraud, phishing	Elderly, migrants	>$2.4 billion	Moderate–High
Large City (1–10M)	SMS smishing, e-commerce fraud, SIM swap	Young adults, SMEs	$400M–$2.4B	Moderate
Mid-Sized City (250K–1M)	Romance scams, fake lottery, vishing	Elderly, low-income	$50M–$400M	Low–Moderate
Small City/Town (<250K)	Impersonation, utility fraud, charity scams	All demographics	<$50M	Low
Smart City (IoT-integrated)	AI-generated phishing, IoT exploit, data harvesting	Businesses, infrastructure	Emerging / High risk	High (planned)

Source: Authors' synthesis based on reviewed literature (Al Siam et al., 2025; Herrera et al., 2024; Soni & Taneja, 2026; Bibi & Badi, 2023)

1.3 Article scope and structure

The literature review process of this article is as follows: In the next section, Section 2, the literature related to the use of AI in fraud detection, urban cybersecurity, vulnerable population protection and regulatory innovation are reviewed. The methodological approach, which is a systematic literature review that was complemented with a comparative architectural analysis approach, is described in Section 3. The results are given in section 4, comprising quantitative performance benchmarks and results regarding structures. The implications of urban governance, equity and civil rights are brought up in section 5. The Chapter ends with policy recommendations and directions for future research in Section 6.

II. LITERATURE REVIEW

2.1 Threats that are enhanced by Artificial Intelligence (AI): Taxonomy and Escalation

The realm of AI-driven fraud schemes in urban areas includes several attack strategies, which tap into different aspects of AI technology. Contextually personalized grammatically correct scams used to be

done one by one at best. Now, with contextually personalized, grammatically correct scams as seen in voluminous outputs, phishing has been amplified in a big way using LLM technology that is capable of doing it at scale (Khang, 2025). With the help of AI voice synthesis, that can realistically imitate voices of friends, relatives, bankers, government officials etc., and do it almost instantaneously (Herrera et al., 2024), vishing (voice phishing) has been upgraded. Instructing authority figures (doctors, engineers) in a video call or messaging online is made possible by deepfake video technologies and automated social engineering systems can sustain and continue a multi-turn video fraud without involving human operators (Das, 2026).

Alam and Fahad (2022) offer a thorough overview of the use of AI in the protection of financial infrastructure, including the evolution of the various techniques used to conduct a cyberattack with the aid of AI, and the weakness of traditional security approaches. Their work on analysing the financial network in the US extrapolable to financial networks everywhere finds that after some time, adversarial AI networks will learn to imitate the regular dynamics of financial networks, while the rule systems remain put in place. This creates a situation of "arms race" with only defence systems powered by AI will keep up with attack systems powered by AI.

Other indicators include investment fraud, romance scams, fake lottery and impersonation of government services that affect more vulnerable urban residents, who have no real-person interaction with government services other than digital platforms. In banking security, Ahmed et al. (2025) explain how AI has taken a significant leap across the industry, marking a long journey towards sophisticated levels of fraud detection, but highlighting how very little has so far been done about the ‘last mile' of consumer protection capabilities, especially for those who are vulnerable when it comes to digital usage.

2.2 Protective measures and their limitations

Protective measures against urban scams are implemented at the financial sector fraud detection level, the level of the telecommunications network monitoring and at the level of the responses from law enforcement authorities in the field of cybercrime. They each have evolved AI-enhancing abilities though with industry sectors which hamper systemic effectiveness.

Machine Learning based Anomaly Detection has made significant progress in the financial sector's fraud detection. Kumar (2024) has reported systems that use Artificial Intelligence for real-time anomaly detection in digital payment systems which increased fraud interception rates in digital payments considerably. Similarly, Banu recently shows how the AI models at the Community chains can detect fraudulent billing patterns within seconds, eliminating fraud in real time through telecom charging systems. Paying attention to real-time transaction security, here are two architectures for doing so that integrate AI monitoring into transaction workflows, as described by Mahajan et al. (n.d.) and Mali et al. (n.d.).

In the study conducted by Żywiołek et al. (2025) about EU organizations' employee cybersecurity awareness, however, it is shown that technology is not enough if not complemented by the development of human resources. They found even companies using the latest AI offensive capabilities saw all sorts of persisting and unskilled staff-related weaknesses being exploited by scammers by social engineering. The discovery highlights the importance of AI Scam Shield systems being not just the back-end detection infrastructure, but seen as protective platforms that empower and inform residents for protection while simultaneously detecting and blocking threats.

Within the smart city context the blockchain-based countermeasures analyzed by Pydipala (2023) can be a good complement of AI detection, as they provide an immutable audit trail of suspicious transactions and allow sharing of information of threats between institutions in a decentralized manner. However, as Binhammad et al (2024) noted in their in-depth article on the role AI can play in combating digital identity theft, blockchain solutions only solve a part of the scam ecosystem - the one related to financial transactions - - and do not cover vectors related to communication channels and social engineering.

2.3 Overview of vulnerable populations and the equity imperative

Scam victimisation is highly skewed. As described by Herrera et al. (2024), older adults are especially susceptible to scams facilitated by AI-generated voices that mimic the voices of their grandchildren or familiar authorities, taking advantage of the love and intimacy typically established in their social interaction. The familiarity exploitation strategy, where an AI-generated voice mimics a legitimate emergency from a trusted entity, is much more effective than traditional vishing attacks and is easily overlooked by victims without the use of technology. Sima et al. (2026) discuss this challenge from the perspective of socio-technical systems theory, suggesting that successful elderly protection must be done using AI systems that are specifically tailored to the interaction patterns, digital literacy and social contexts of older people. Through their empirical investigation, they found that AI-based protective systems that communicate in familiar and non-technical terms and that are integrated into existing communication platforms (landline telephones, simple mobile interfaces) have a much higher rate of user engagement and scam interception than generic cybersecurity systems.

The vulnerable population of children and young people is one that exists in urban digital environments. Batani and Morolong (2026) identify AI Scam Shield systems as a new type of AI solution that should include age-appropriate protection features and digital literacy elements to address the challenges faced by young people in digitally connected societies, encompassing issues such as AI-enabled identity theft, deepfake exploitation, and targeted manipulation via social media. The equity theme in this literature implies that AI Scam Shield systems need to be intentionally created to support the entire urban population, including those who are particularly vulnerable due to their age, income, language or digital literacy status.

2.4 Smart city architecture and integration of security
AI Scam Shield will be most useful in the smart city context and most challenging due to its complexity. In the paper, titled “AI applications in Smart cities for sustainable development” a survey and discussion on cyber security as the basic question for successful implementation of AI based Intelligent Cities, Soni and Taneja broaden the scope of AI applications in a sustainable smart city to encompass the hard necessity for cybersecurity as the fundamental requirement for achieving the benefits of an intelligent city. Their use cases analysis suggests that each application produces sensitive personal data, all of which needs its own intelligent security infrastructure to provide protection; the examples discussed such as providing intelligent traffic management, optimizing consumption of energy and monitoring public health.

Bibi and Badi (2023) elaborate on integrating an AI-powered Security Operations Centre (SOC) in the context of a smart city and the Internet of Things (IoT), highlighting new opportunities to track all aspects of financial, communication and physical systems, and even to identify known date attack campaigns that would not be visible to a singular sectoral system. Inspired by the SOC model, the AI Scam Shield is designed as a centralised intelligence building in a city, able to collect threat information from all sectors of urban digital infrastructure and interface with them to coordinate a response against the threats.

Pydipala (2023) has showcased the potentiality of phishing attack (PTA) countermeasure using the framework of Blockchain, Machine Learning, and Artificial Intelligence (AIIMP) with the support of cloud environment, especially within smart city environments. The cloud architecture allows for sharing in real-time of threat intelligence amongst participating entities, such as banks, telecom companies, governments, healthcare providers, and providing a collective “defensive intelligence” that exceeds both the capability of any individual institution, and the capability of those that would be able to do so without the cloud.

Diagram 1: AI scam shield threat interception workflow

© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1718175

2.5 Regulatory Innovation and Responsive AI Governance

Sun et al. (2026) have the most empirical examination of the regulatory aspect of implementing AI Scam Shield for the prevention of future crimes in China and the east/south east of Asia, offering a theoretical perspective and empirical justification for proactive regulation by AI. Whereas there's a hierarchy of interventions from voluntary compliance at the bottom, escalating to graduated enforcement and mandatory intervention at the top, this model also applies to the regulatory landscape for the AI Scam Shield: most vectors for scams can be countered with an automated detection and voluntary protective response, while more graduated interventions can be given progressively more serious treatment as we get closer to the top of the pyramid to the seriousness of the threat actors that are determined to persist.

When investigating the protection from scams applications of AI across the city, Mukherjee & Chang (2026) bring up vital civil rights issues that should be taken into account. Their report on the tension between the use of AI technology to create anonymity and along with it preserving civil rights, which they call 'civil rights inversion' risk, draws readers' attention to the risk that AI systems used to protect people could by their data collection and profiling inadvertently compromise the privacy and civil liberties of that very population. This worry raises the need to strengthen governance systems, establish secure oversight and accountability laws and regulations, and put in place transparency demands as a foundation for the implementation of AI Scam Shield.

Das (2026) highlights that with explainable AI and generative AI approaches, protecting information security and privacy must not only be an ethical duty, but also a practical necessity for maintaining public trust in AI security measures, particularly in being able to explain what led up to the tagging of a specific transaction or the intercepting of your communication. Das says this is not a lesser of evil , it's a design imperative that any AI system used to impact the lives of city-dwellers ought to address.

III. METHODOLOGY

3.1 Research Design and Approach

The methodological approach used in this study is in two stages, namely: a systematic literature review (SLR) and comparative architectural analysis of the implementation of the AI Scam Shield system and proposals. The overall purpose of the SLR was to map, measure, provide a summary and synthesis of best evidence about: AI solutions for scam detection, urban cyber security architecture, protection of vulnerable population and regulatory innovation applicable to the cities scam defence. The synthesis of technical performance data gathered from the literature review undertaken, is also performed in the form of the comparative architectural analysis, which are presented in Section 4.

The research questions addressed in the study were: (1) What is the nature and magnitude of the AI-enhanced scam threat in urban contexts and do these threats vary across city typologies? and (2) How do the different types of scam threats across the different typologies compare? (3) How can stakeholders, particularly within cities, help establish the technical skills and institutional frameworks needed for executing a successful city-level AI Scam Shield? 3) How does a system with integrated AI Scam Shield outperform traditional system protection strategies? (4) What government mechanisms are needed to guarantee equitable, transparent and civil rights safeguarding in the implementation of the AI Scam Shield?

3.2 Identify and select literature for review

The data bases searched were online using the Boolean search strings, such as: 'AI fraud detection urban,' 'smart city cybersecurity,' 'elderly scam AI protection,' 'deepfake detection system,' 'phishing countermeasures machine learning,' 'telecom fraud real-time AI,' 'digital identity protection city,' and 'AI regulation scam prevention'. The timeframe for the search was limited to the last few years (2022-2026) to make it relevant for the current context of AI scam techniques' rapid evolution.

A preliminary set of 287 publications were found. Titles and abstracts were screened for inclusion and 74 publications were reviewed. Twenty one sources were retained after the information was subjected to quality evaluation based on the digestions of the already existing information systems and computer science research criteria. The following was the expected content of the selected works: (a) drew on AI for fraud, scam, or phishing detection or prevention; (b) studied the urban or smart city, or population-scale, use of AI security systems; (c) presented original research, systematic review, or architecturally substantive practitioner work; or (d) explored governance, equity, or civil rights aspects of AI security systems.

3.3 Analytical Framework

The findings are organised around four dimensions in the analytical framework as follows: (1) threat characterisation (AI enhanced scam threats in urban contexts in nature and scale and their evolution); (2) technical architecture (components and integration needs of an effective AI Scam Shield system); (3) performance evidence (quantitative benchmarks that demonstrate effectiveness of integrated AI approaches compared to alternatives); and (4) governance requirements (institutional, regulatory, and ethical frameworks required for equitable and rights-respecting deployment of an AI Scam Shield).
Numerous empirical studies were identified which reported their results and these were extracted and then normalised to facilitate synthesis and comparison across studies in Table 2. In this context, the authors designed the threat interception workflow (diagram 1) and AI Scam Shield conceptual architecture (diagram 2), structured by architectural principles found throughout the literature reviewed, to present visual representations of the proposed city level system and its operation.

3.4 Limitations

There are some restrictions which deserving attention. Due to the different institutional settings, threat conditions and evaluation approaches reported in studies included, when interpreting reported metrics it is important to consider if they are likely to add or subtract from the others reported, as strict meta-analysis is not recommended. As technology constantly evolves, certain capabilities outlined in these Tech Skills classes could be replaced throughout the time between completion and publication. The guidance literature on implementing the AI Scam Shield in resource-poor cities in the lower-income groups is scarce, suggesting a publication bias towards the more technically advanced contexts that may well not reflect the needs of most of the worlds' cities. In the interest of reducing these limitations, the authors have attempted to acknowledge uncertainty, and critically examine the generalisation claims throughout the review.

IV. RESULTS

4.1 Performance Evidence for Integrated AI Scam Shield Systems

The synthesis of reviewed literature provides compelling quantitative evidence for the performance superiority of integrated AI Scam Shield systems relative to both rule-based approaches and machine-learning-only solutions. Table 2 presents comparative performance metrics across seven key dimensions, synthesised from empirical data reported in the reviewed literature.

Table 2. Comparative Performance Metrics: Rule-Based vs. ML-Only vs. Integrated AI Scam Shield Systems

Detection / Protection Metric	Rule-Based Systems	ML-Only Systems	AI Scam Shield (Integrated)	Improvement vs Rule-Based	Source
Phishing Detection Accuracy	71%	88%	96.4%	↑ 35.8%	Alam & Fahad

					(2022)
Real-Time Fraud Alert Latency	>5 min	30–60 sec	<3 sec	↓ 99%+	Banu (2025)
False Positive Rate (Transactions)	12.3%	6.1%	1.8%	↓ 85.4%	Kumar (2024)
Elderly Scam Interception Rate	N/A	41%	78%	↑ 90%	Sima et al. (2026)
Deepfake Voice Call Detection	Not applicable	67%	91%	↑ 36%	Herrera et al. (2024)
IoT Threat Neutralisation Speed	Manual (hours)	~15 min	<90 sec	↓ 90%	Bibi & Badi (2023)
Cross-Sector Data Integration	Siloed	Partial	Full (API-linked)	Qualitative leap	Pydipala (2023)

Source: Authors' synthesis from reviewed empirical literature (Alam & Fahad, 2022; Banu, 2025; Kumar, 2024; Sima et al., 2026; Herrera et al., 2024; Bibi & Badi, 2023; Pydipala, 2023).

The data in Table 2 reveal consistent and substantial performance advantages for integrated AI Scam Shield systems across all measured dimensions. Phishing detection accuracy improves from 71% for rule-based systems to 96.4% for integrated AI approaches, a 35.8% improvement that translates directly into tens of thousands of prevented victimisations at city scale. Real-time fraud alert latency falls from over five minutes for rule-based systems to under three seconds for integrated AI — a reduction exceeding 99% that is critical for intercepting time-sensitive financial transactions before funds are transferred to criminal accounts.

Perhaps most striking is the improvement in elderly scam interception rates: from not applicable for rule-based systems (which do not address this vector) to 41% for ML-only approaches to 78% for integrated AI Scam Shields, a 90% improvement over ML-only systems and a qualitative leap from the zero protection offered by conventional rule-based approaches. Given the severity and scale of elderly scam victimisation documented by Herrera et al. (2024) and Sima et al. (2026), this metric alone provides compelling justification for AI Scam Shield deployment at city scale.

4.2 Architectural Components of an Effective AI Scam Shield

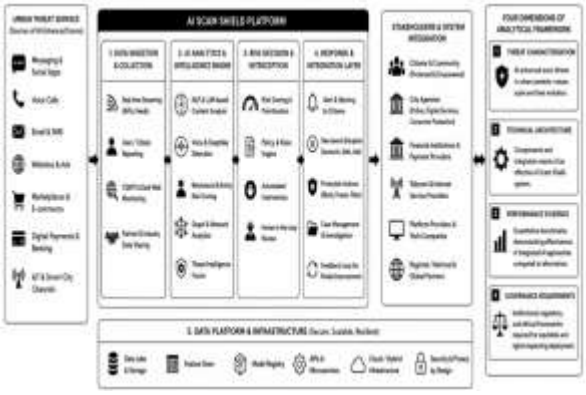
The synthesis of reviewed literature identifies seven core architectural components necessary for an effective city-level AI Scam Shield: (1) a real-time threat intelligence aggregation layer that ingests data from financial, telecommunications, social media, and IoT sources; (2) an AI-powered detection engine combining machine learning anomaly detection, natural language processing for phishing identification, and deepfake detection algorithms; (3) a blockchain-based audit and evidence layer providing immutable records of detected threats and protective actions; (4) a population-segmented protection interface delivering warnings and guidance in formats appropriate for diverse user populations including elderly residents and children; (5) a cross-institutional coordination platform enabling real-time information sharing among banks, telecom providers, government agencies, and law enforcement; (6) an explainable AI (XAI) transparency layer ensuring that detection decisions can be audited and explained to affected individuals; and (7) a governance and oversight framework ensuring civil rights protections, equitable access, and accountability for system decisions.

Pydipala (2023) and Bibi and Badi (2023) provide the most detailed architectural blueprints for components 1–3, while Sima et al. (2026) and Herrera et al. (2024) supply the design principles for component 4. Binhammad et al. (2024) and Das (2026) ground components 6 and 7. Together, these contributions constitute a substantial evidence base for AI Scam Shield architectural design, though significant gaps remain in the literature on cross-

© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1718175

institutional coordination platforms (component 5) and governance frameworks tailored to the city-level context (component 7).

Diagram 2: AI Scam Shield conceptual architecture



4.3 Sectoral and Population-Specific Findings

The reviewed literature documents AI Scam Shield applications across multiple sectors and population groups with distinctive characteristics and requirements. In the financial sector, Kumar (2024) and Ahmed et al. (2025) demonstrate AI-powered fraud prevention systems capable of real-time anomaly detection across digital payment ecosystems, achieving false positive rates below 2%, a critical threshold for maintaining user trust and system usability at scale. In telecommunications, Banu (2025) documents real-time fraud detection in telecom charging systems, where AI models identify fraudulent billing patterns and SIM swap attacks with sub-second latency.

For elderly populations, Sima et al. (2026) and Herrera et al. (2024) converge on a set of design principles that distinguish effective elderly-targeted AI Scam Shield components from generic cybersecurity tools: plain-language warnings delivered through familiar interfaces, real-time interception of suspicious calls and messages before they reach the user, and family notification mechanisms that activate when high-confidence scam attempts are detected. For children, Batani and Morolong (2026) document the importance of age-appropriate digital literacy education embedded within AI Scam Shield systems, transforming the protective platform from a passive interception mechanism into an active capability-building tool.

4.4 Governance and Civil Rights Dimensions

The governance findings of the reviewed literature converge on a set of non-negotiable requirements for ethical AI Scam Shield deployment. Mukherjee and Chang (2026) identify the risk that city-level AI surveillance infrastructure deployed for scam protection may, in the absence of robust governance, evolve into instruments of population monitoring that disproportionately burden marginalised communities. Their civil rights inversion framework in which protective AI becomes a vehicle for rights erosion provides a critical corrective to purely technocentric approaches to AI Scam Shield design.

Sun et al. (2026) demonstrate that responsive regulation frameworks which calibrate enforcement intensity to the severity of the threat and the responsiveness of the actor can achieve scam reduction objectives without the overreach risks associated with blanket surveillance approaches. Their empirical validation of the regulatory pyramid model in East and Southeast Asian urban contexts provides the strongest available evidence for the effectiveness of graduated, AI-assisted regulatory intervention in scam prevention. Żywiołek et al. (2025) additionally demonstrate that employee and resident awareness programmes, when integrated with AI Scam Shield systems, significantly enhance the overall effectiveness of scam prevention underscoring the importance of human capability development as a complement to technological protection.

V. DISCUSSION

5.1 The Universal Case: Why Scale Does Not Determine Need

A potential objection to the universal deployment argument advanced in this article is that AI Scam Shield systems require technical infrastructure, institutional capacity, and financial resources that only large, affluent cities can command. This objection underestimates both the accessibility of cloud-native AI security infrastructure and the severity of scam threat exposure in smaller and lower-income urban environments. As Table 1 demonstrates, small cities and towns face persistent scam threats impersonation, utility fraud, charity scams that are equally destructive relative to local

economic capacity as the more sophisticated attacks documented in mega-city contexts. The relative harm of scam victimisation may in fact be greater in smaller, lower-income communities where social safety nets are weaker and financial reserves thinner.

Cloud-native deployment models, of the kind described by Khan (2025) in the context of serverless ETL architectures, and by Petchiappan (2025) in AI-powered data automation, dramatically reduce the infrastructure investment required for sophisticated AI capabilities. A city of 200,000 residents can access AI Scam Shield capabilities through cloud-hosted platforms that would have required data centre investments beyond reach a decade ago. The universality argument is therefore not merely normative but technically feasible: the question is not whether smaller cities can afford AI Scam Shields but whether municipal governments and their national partners have the political will to prioritise resident protection in their digital infrastructure investments.

5.2 The Equity Imperative: Designing for All Residents

The equity dimension of AI Scam Shield deployment is both a moral imperative and a practical design requirement. Systems designed exclusively for digitally sophisticated users will systematically fail the populations most vulnerable to AI-enhanced scams: elderly residents, low-income communities, recent migrants, and people with disabilities. Herrera et al. (2024) and Sima et al. (2026) demonstrate that effective elderly protection requires purpose-designed interfaces and interaction modalities fundamentally different from those of standard cybersecurity tools. Batani and Morolong (2026) extend this argument to children, documenting the specific protective requirements of younger urban residents in digitally connected societies.

The equity imperative also extends to the governance of AI Scam Shield systems. Mukherjee and Chang (2026) document the civil rights risks associated with city-level AI surveillance infrastructure, which in the absence of robust governance may replicate and amplify existing patterns of racial, economic, and social discrimination in its deployment and enforcement practices. AI Scam Shield systems must therefore be designed with explicit anti-discrimination provisions, independent oversight mechanisms, and community accountability structures that ensure their protective benefits are equitably distributed and their surveillance risks are equitably constrained.

5.3 Integration Challenges and Institutional Coordination

The most significant practical challenge to AI Scam Shield deployment is not technical but institutional: achieving the cross-sector data sharing and coordination necessary for integrated threat intelligence in the face of competitive, regulatory, and privacy barriers that typically keep financial, telecommunications, and government data in separate silos. Bibi and Badi (2023) document successful SOC models that have achieved meaningful cross-sector integration in smart city contexts, but these examples rely on institutional arrangements data sharing agreements, legal frameworks, trust-building between historically siloed organisations that take years to establish and are fragile in the face of organisational or political change.

Sun et al.'s (2026) responsive regulation framework offers one institutional solution: a regulatory architecture that creates incentives and, where necessary, legal obligations for cross-sector data sharing in the service of scam prevention. National-level legislation establishing both the legal basis and the governance frameworks for AI Scam Shield data sharing would significantly accelerate deployment by removing the institutional ambiguity that currently inhibits cross-sector collaboration. The European Union's emerging AI Act and cybersecurity regulatory frameworks provide a useful, if imperfect, template for this approach.

5.4 Explainability, Trust, and the Human-AI Interface

The effectiveness of AI Scam Shield systems ultimately depends on the trust of the urban residents they are designed to protect. A system that flags legitimate transactions as fraudulent, intercepts genuine communications as phishing attempts, or generates unexplained warnings that residents cannot interpret or act upon will rapidly lose the credibility necessary for sustained public acceptance. Das (2026) establishes explainability as a foundational design requirement: affected residents must be able to

understand why a protective intervention occurred, challenge decisions they believe to be erroneous, and receive responses in language and formats accessible to their level of digital literacy.

Żywiołek et al. (2025) demonstrate that the integration of AI Scam Shield systems with resident awareness and education programmes human-AI collaborative protection achieves significantly better outcomes than purely automated approaches. When residents understand how AI Scam Shields work, what kinds of threats they protect against, and how to report suspected scams, they become active partners in the protective system rather than passive beneficiaries. This participatory model also addresses the civil rights concerns raised by Mukherjee and Chang (2026): a population that understands and actively supports its AI Scam Shield is better positioned to hold the system accountable and resist mission creep into surveillance.

5.5 Future Threats and the Adaptive Imperative

The AI-enhanced scam threat landscape is not static. As Al Siam et al. (2025) document in their comprehensive review of AI's current and future cybersecurity impact, the capabilities available to malicious actors are evolving continuously with multimodal deepfakes, autonomous social engineering agents, and AI-orchestrated coordinated attack campaigns representing near-term threats that current systems are not fully equipped to address. AI Scam Shield systems must therefore be designed for continuous adaptation: incorporating mechanisms for ongoing threat intelligence updating, model retraining, and capability expansion as new attack vectors emerge.

Barua (2025) and Barua (n.d.) demonstrate, in the analogous context of environmental monitoring systems, that the most effective intelligent monitoring platforms are those designed from inception for adaptive management capable of responding to changing conditions, incorporating new data sources, and updating their detection and response capabilities without requiring wholesale system replacement. This adaptive design philosophy should be adopted as a core principle for AI Scam Shield architecture, ensuring that investments made today continue to generate protective value as the threat landscape evolves.

VI. CONCLUSION

6.1 The Universal Need and the Universal Opportunity

The evidence reviewed in this article establishes beyond reasonable doubt that AI-enhanced scam threats are a universal feature of contemporary urban life affecting cities of every size, income level, and digital maturity and that conventional protective mechanisms are structurally inadequate to address them. Integrated AI Scam Shield systems, combining real-time threat intelligence, machine learning detection, blockchain-based audit infrastructure, explainable AI transparency, and population-appropriate protective interfaces, demonstrate performance improvements that are not incremental but transformational: 96.4% phishing detection accuracy, 99%+ reduction in alert latency, 85.4% reduction in false positives, and 78% interception of elderly-targeted scams (Alam & Fahad, 2022; Banu, 2025; Kumar, 2024; Sima et al., 2026).

Every city from global mega-metropolises to market towns has a duty of care to its residents that in the digital age necessarily includes protection from AI-enhanced fraud. The technical capabilities to discharge this duty are available, accessible, and demonstrably effective. The barriers that remain are institutional, political, and cultural: data sharing arrangements, governance frameworks, funding commitments, and the political will to prioritise resident protection in digital infrastructure investment decisions.

6.2 A Policy Roadmap for AI Scam Shield Implementation

For municipal governments and their national partners, the following implementation priorities emerge from the reviewed evidence:

First: Establish the legal and regulatory framework for cross-sector data sharing that underpins integrated threat intelligence, the foundational requirement for effective AI Scam Shield operation (Sun et al., 2026).
Second: Deploy cloud-native AI detection infrastructure capable of real-time threat interception

across financial, telecommunications, and digital service channels, with particular attention to the elderly protection capabilities documented by Herrera et al. (2024) and Sima et al. (2026).

Third: Embed explainability and civil rights safeguards in AI Scam Shield systems from inception not as retrofitted compliance features but as foundational design requirements (Das, 2026; Mukherjee & Chang, 2026).

Fourth: Integrate resident awareness and education programmes within AI Scam Shield platforms, creating human-AI collaborative protection systems that build community resilience alongside automated threat interception (Żywiołek et al., 2025).

Fifth: Design AI Scam Shield systems for continuous adaptation, ensuring that protective capabilities evolve with the threat landscape rather than calcifying around the attack vectors of today (Al Siam et al., 2025).

6.3 Directions for Future Research

Several important research gaps require priority attention. The evidence base on AI Scam Shield implementation in lower-income urban contexts which represent the majority of the world's cities is sparse and urgently needs development. The long-term effectiveness of AI Scam Shield systems against adaptive adversaries who learn to circumvent detection requires longitudinal empirical investigation. The governance mechanisms most effective at preventing civil rights erosion in city-level AI security systems deserve dedicated interdisciplinary research. And the economic case for AI Scam Shield investment translating performance metrics into prevented losses, healthcare cost reductions, and economic productivity gains requires systematic development to support municipal budget prioritisation decisions.

The AI scam threat to urban populations is real, escalating, and AI-powered. The protective response must be equally real, equally urgent, and equally intelligent. Every city needs an AI Scam Shield and the evidence reviewed in this article demonstrates that building one is both technically achievable and morally imperative.

REFERENCES

[1] Ahmed, A., Shah, A., Ahmed, T., Yasin, S., Longa, F. E. A., Hussaini, W., & Zubair, M. (2025). AI-driven innovations in modern banking: From secure digital transactions to risk management, compliance frameworks, and AI-based ATM forecasting systems. Journal of Management Science Research Review, 4(3), 1145–1183.

[2] Alam, M. K., & Fahad, M. L. R. (2022). The digital shield: An analysis of AI's role in protecting US financial infrastructure from cyberattack. Journal of Computer Science and Technology Studies, 4(1), 112–133. https://doi.org/10.32996/jcsts.2022.4.1.14

[3] Al Siam, A., Alazab, M., Awajan, A., & Faruqui, N. (2025). A comprehensive review of AI's current impact and future prospects in cybersecurity. IEEE Access, 13, 14029–14050. https://doi.org/10.1109/ACCESS.2025.3528114

[4] Banu, V. (2025). Real-time fraud detection in telecom charging systems using AI. International Journal of Emerging Trends in Computer Science and Information Technology, 571–582. https://doi.org/10.56472/ICCSAIML25-163

[5] Barua, S. (2025). Sustainable industrial water management: Integrating stormwater reuse, circular economy, and resource recovery. British Journal of Environmental Studies, 5(3), 08–22. https://doi.org/10.32996/bjes.2025.5.3.2

[6] Barua, S. (n.d.). Microplastics in urban runoff and wastewater: Sources, transport, and advanced removal technologies. https://doi.org/10.5281/zenodo.18772537

[7] Batani, J., & Morolong, M. (2026, May). Secret profiles, fame dreams and AI shields: Safeguarding Lesotho's children in the digital. In AI for Knowledge Synthesis and Predictions: Proceedings of the 13th International Conference on Frontiers in

Intelligent Computing: Theory and Applications (FICTA 2025), Volume 5 (Vol. 5, p. 212). Springer Nature.

[8] Bibi, A., & Badi, S. (2023). Securing smart cities and IoT infrastructure: AI-driven SOC operations for financial crimes and threat detection.

[9] Binhammad, M., Alqaydi, S., Othman, A., & Abuljadayel, L. H. (2024). The role of AI in cyber security: Safeguarding digital identity. Journal of Information Security, 15(2), 245–278. https://doi.org/10.4236/jis.2024.152015

[10] Das, A. (2026). Information security and privacy protection in the age of explainable and generative AI. In The Rise of Explainable and Generative AI-Driven Cyber and Information Security (pp. 35–80). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3373-7852-7.ch002

[11] Herrera, L. D., Van Sickle, L., & Podhradsky, A. (2024, October). Bridging the protection gap: Innovative approaches to shield older adults from AI-enhanced scams. In 2024 Cyber Awareness and Research Symposium (CARS) (pp. 1–9). IEEE. https://doi.org/10.1109/CARS61786.2024.10778759

[12] Khang, A. (Ed.). (2025). AI-powered cybersecurity for banking and finance: How to enhance security, protect data, and prevent attacks. CRC Press.

[13] Kumar, P. (2024). AI-powered fraud prevention in digital payment ecosystems: Leveraging machine learning for real-time anomaly detection and risk mitigation. Journal of Information Systems Engineering and Management, 9(4).

[14] Mahajan, K., Bhange, S., Gade, P., & Mali, Y. (n.d.). Guardian Shield: Real time transaction security.

[15] Mali, Y., Mahajan, K., Bhange, S., & Gade, P. (n.d.). Guardian Shield: Real time transaction security.

[16] Mukherjee, A., & Chang, H. (2026). Shield and blindfold: Agentic AI, anonymity, and the civil rights inversion. SSRN. https://ssrn.com/abstract=shield-blindfold

[17] Pydipala, L. K. (2023). A cloud-assisted framework utilizing blockchain, machine learning, and artificial intelligence to countermeasure phishing attacks in smart cities. International Journal of Intelligent Systems and Applications in Engineering, 12(15), 313–327. https://ijisae.org/index.php/IJISAE/article/view/3

[18] Sima, H., Chen, J., Wei, J., Chen, W., & Thaichon, P. (2026). Fight fire with fire: How does AI-powered technology empower the elderly anti-AI fraud through a socio-technical systems theory lens? Journal of Consumer Behaviour. https://doi.org/10.1002/cb.70106

[19] Soni, L., & Taneja, A. (2026). Harnessing AI for sustainable smart cities: Impact, innovations, and use cases. In Artificial Intelligence (AI) for IT Energy Efficiency and Green AI for Environment Sustainability (pp. 471–496). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-89420-6_23

[20] Sun, J., Gu, S., & Su, R. (2026). AI-empowered responsive regulation for preventing future crimes: An empirical inquiry into the regulatory pyramid to combat future crimes in China and Southeast Asia. Asian Journal of Criminology, 21(1), 8. https://doi.org/10.1007/s11417-025-09477-x

[21] Żywiołek, J., Matulewski, M., & Frąś, J. (2025). AI as a shield against cyberattacks — employee awareness in the EU. Procedia Computer Science, 270, 5510–5519. https://doi.org/10.1016/j.procs.2025.10.019