ijamrsd.com
ISSN: 2583-6404
Journal of Advanced Multidisciplinary Research Studies and Development
Peer Reviewed Referred Journal
Bi-monthly
Open Access

CITIZEN AI SHIELD FRAMEWORK (CASF): AN AI–CYBERSECURITY COLLABORATION TO PREVENT UNINTENTIONAL CYBER CRIMES BY CITIZENS

Dr. Atul Suresh Patil
Associate Professor, Department of Computer Science, JET's ZulalBhilajirao Patil College, Dhule.
Email: atulpatil12@gmail.com

Abstract
The rapid expansion of digital technologies has transformed everyday life. Citizens now depend on online platforms for communication, financial transactions, education, public services, and entertainment. However, increased digital participation has also resulted in higher exposure to cyber risks. Many individuals unknowingly engage in activities that may lead to cyber offences, such as sharing unlawful content, clicking phishing links, participating in fraudulent schemes, or violating data protection norms. These actions are often not malicious but arise from lack of awareness, misleading interfaces, or sophisticated social engineering techniques. Traditional cybersecurity systems primarily focus on detecting external attackers, blocking malware, or protecting enterprise networks. Very limited attention has been given to preventing user-side mistakes that may result in accidental cyber violations. Recent reports indicate a steady rise in cyber incidents in India [4][5], emphasizing the need for preventive, citizen-centric cybersecurity mechanisms. This paper proposes the Citizen AI Shield Framework (CASF), an AI-enabled preventive cybersecurity model designed to assist users before risky actions are completed. The framework integrates local AI-based activity monitoring, behavioral risk analysis, intent prediction, and real-time educational alerts. CASF aligns with Indian cyber laws including the Information Technology Act, 2000 [1], the Digital Personal Data Protection Act, 2023 [2], and CERT-In guidelines [3]. The proposed model focuses on proactive prevention rather than reactive punishment. CASF aims to strengthen digital responsibility, reduce accidental cyber violations, and promote safer online behavior among citizens.

Keywords: Artificial Intelligence, Cybersecurity, Citizen Protection, Preventive Alerts, IT Act 2000, Digital Safety.

► *Corresponding Author: Dr. Atul Suresh Patil*

I. Introduction

Digital transformation has reshaped economic systems, governance structures, and social interactions. In India, digital banking, Unified Payments Interface (UPI), e-commerce platforms, online education, and social networking applications have significantly increased internet usage. While these services offer convenience and economic growth, they also expose citizens to new forms of cyber risks.

Government reports have highlighted significant growth in cyber fraud and digital deception cases [5]. The National Crime Records Bureau has also documented rising cybercrime registrations

across various categories such as online fraud, identity theft, and cyber harassment [4]. These statistics indicate that cybercrime is no longer limited to organized hacking groups; it increasingly involves interactions with ordinary citizens.

Most cybersecurity solutions are designed to protect infrastructure, organizations, and servers. Firewalls, intrusion detection systems, encryption tools, and anti-malware solutions are effective against external threats. However, a large portion of cyber incidents involve human interaction, including clicking suspicious links, sharing misinformation, using pirated software, or disclosing sensitive personal information [6][7].

In many cases, citizens unintentionally violate legal provisions due to lack of awareness or confusion. For example, forwarding defamatory content or sharing copyrighted material may lead to legal consequences under existing laws [1]. Similarly, careless sharing of personal data may violate data protection requirements under the Digital Personal Data Protection Act [2].

Therefore, strengthening user awareness and preventive support mechanisms is equally important as strengthening network defenses. The objective of this study is to design a preventive AI-based framework that supports citizens in avoiding accidental cyber offences while aligning with Indian legal frameworks [1][2].

II. Legal and Policy Context in India

Any preventive cybersecurity framework must operate within national legal and policy boundaries. India has established multiple laws and institutions governing digital activities.

The **Information Technology Act, 2000** defines various cyber offences such as unauthorized access, identity theft, data manipulation, publishing obscene content, and electronic fraud [1]. The Act also outlines penalties and legal procedures. Citizens who unknowingly share unlawful content or engage with malicious platforms may inadvertently violate certain provisions.

The **Digital Personal Data Protection Act, 2023** establishes obligations for responsible handling of personal data [2]. It emphasizes consent, purpose limitation, and secure data processing. Improper sharing or misuse of personal information can lead to financial penalties under this Act.

Furthermore, **CERT-In (Indian Computer Emergency Response Team)** plays a central role in cybersecurity governance [3]. It issues advisories, coordinates incident responses, and provides guidelines for cyber risk mitigation. Any preventive system must align with CERT-In recommendations to ensure national-level consistency.

Together, these frameworks emphasize responsible digital behavior and highlight the importance of educating users about legal consequences.

III. Need for a Preventive AI-Based Framework

Cybersecurity incidents frequently involve human error. Studies indicate that user interaction remains a significant factor in data breaches and fraud cases [6]. Many users lack sufficient digital literacy, making them vulnerable to phishing attacks, misinformation campaigns, malware downloads, and online scams.

Traditional security tools operate reactively. Antivirus software detects malicious files after download. Firewalls block unauthorized network access. Spam filters identify suspicious emails. However, these systems rarely provide contextual legal awareness or preventive education.

For example:

- A user may download pirated software without realizing it violates copyright laws.
- A citizen may forward harmful content without verifying its authenticity.

- A user may share Aadhaar or banking details publicly without understanding privacy implications.

A preventive system that provides real-time, understandable alerts before action completion can reduce accidental violations significantly. Artificial Intelligence offers the capability to analyzebehavior patterns, assess risk probability, and deliver intelligent guidance.

Therefore, a citizen-centered AI-driven preventive approach is necessary to complement traditional cybersecurity defenses.

IV. Proposed Framework: Citizen AI Shield Framework (CASF)

The **Citizen AI Shield Framework (CASF)** is conceptualized as a layered AI-powered preventive mechanism integrated into user devices and digital platforms.

The framework emphasizes:

- Proactive risk detection
- Legal-awareness-based alerts
- Behavioral improvement
- Privacy preservation

A. Core Architecture Components

1. User Device Layer

This layer includes smartphones, tablets, laptops, and desktop systems used by citizens. CASF operates either as:

- A lightweight background services
- A browser extension
- An integrated mobile application module
- Or embedded functionality within banking or social media applications

The framework is designed to function seamlessly without disrupting normal user activity.

2. Activity Monitoring Layer (Local AI)

This layer performs metadata-based monitoring using on-device AI models. It analyzes:

- Website authenticity and domain reputation
- File signatures and hash values
- Application permission requests
- Behavioural anomalies
- Message forwarding patterns

Importantly, processing occurs locally to comply with data protection principles under the DPDP Act [2]. Personal content is not uploaded to external servers without explicit consent.

3. Risk Analysis & Intent Prediction Layer

Machine learning models evaluate:

- Risk level (Low, Medium, High)
- Probability of fraudulent engagement
- Legal implications under IT Act provisions [1]
- Cyber threat patterns aligned with CERT-In advisories [3]

Intent prediction helps distinguish between accidental and repeated risky actions. For instance, repeated attempts to access known phishing domains may trigger stronger alerts.

4. Preventive Guidance & Alert System

This component generates real-time, user-friendly messages such as:

- “This link appears to be phishing.”

- "Sharing this content may violate IT Act provisions."
- "This file may infringe copyright."
- "This action may expose personal data."

Alerts are educational rather than punitive. The system encourages safe alternatives instead of simply blocking actions.

5. Compliance & Learning Module

This module provides:

- Micro-learning tips
- Short legal awareness summaries
- Behavioral feedback dashboards
- Personalized digital safety suggestions

Over time, the system adapts to user behavior, reducing unnecessary alerts and improving digital responsibility.

V. Architecture Diagram

The CASF architecture consists of interconnected layers beginning from the User Device to Government Support Systems.

The architecture includes:

- User Device
- Local AI Monitoring
- Risk & Intent Analysis
- Preventive Alert Engine
- Trusted Verification Services (CERT-In feeds [3])
- Government & Helpdesk Integration

This layered structure ensures both independence (local AI) and institutional alignment (government advisory systems).

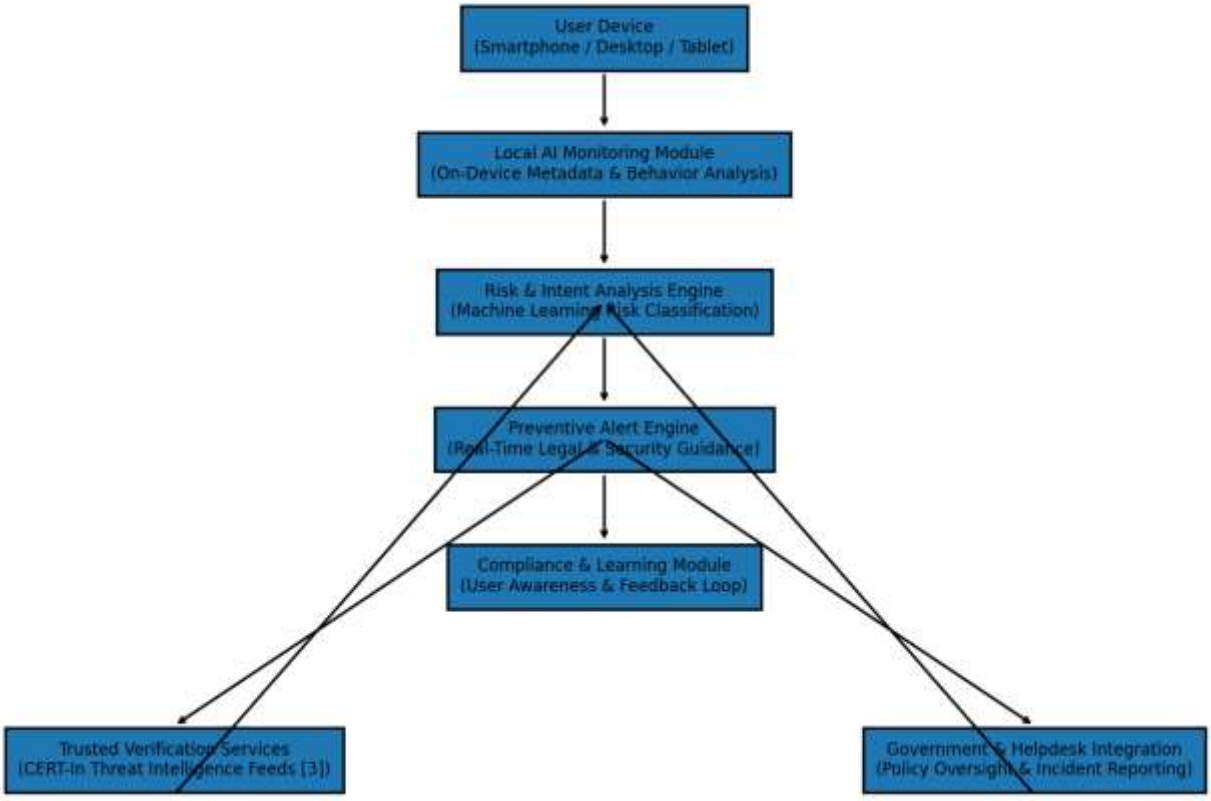

Figure: CASF architecture

VI. Real-Life Application Scenarios

1. Phishing Prevention

When a suspicious link is detected, CASF cross-checks domain reputation using threat intelligence aligned with CERT-In advisories [3]. Before the user enters credentials, a warning appears explaining potential risks. This prevents identity theft and financial fraud.

2. Copyright Violation Alert

If a user attempts to upload or distribute pirated content, CASF flags the action and informs about possible legal risks under IT Act provisions [1]. The alert may include educational information about intellectual property rights.

3. Data Privacy Risk

If sensitive personal information is shared publicly, CASF warns about potential non-compliance with the Digital Personal Data Protection Act [2]. This reduces privacy violations and identity misuse.

4. Misinformation Forwarding

AI-based natural language processing identifies potentially harmful misinformation. The system suggests verification before forwarding and provides links to trusted sources.

VII. Societal Benefits

1. Reduction in Accidental Cyber Offences

Preventive alerts reduce violations caused by ignorance or confusion. This decreases legal disputes and protects citizens from unintended consequences.

2. Improved Digital Literacy
Continuous micro-learning enhances user awareness about cyber laws, safe browsing, and responsible data handling.
3. Support to Law Enforcement
With fewer accidental violations, enforcement agencies can focus on organized cybercriminal networks [4].
4. Stronger Digital Economy
Greater trust in digital systems encourages safe online banking, e-commerce growth, and adoption of digital governance platforms.

VIII. Privacy and Ethical Considerations

CASF prioritizes privacy through:
- Local-device AI processing
- Minimal data collection
- Transparent consent mechanisms
- Encrypted metadata analysis

These principles align with DPDP requirements [2] and responsible AI guidelines [9]. The system is designed to assist users, not monitor them intrusively.

IX. Implementation Challenges

Despite its advantages, CASF faces challenges:
- False positives leading to alert fatigue
- User resistance due to privacy concerns
- Integration complexity across platforms
- Keeping legal mappings updated as laws evolve [1][2]

Continuous model training, regulatory updates, and public awareness campaigns are essential.

X. Future Research Directions

Future research may explore:
- Federated learning for enhanced privacy
- Multilingual alert systems
- Behavioral psychology integration
- Child and senior citizen specific modules
- Government-citizen feedback integration

XI. Conclusion

The Citizen AI Shield Framework (CASF) presents a preventive AI–cybersecurity collaboration model focused on protecting citizens from unintentional cybercrimes. By combining local AI monitoring, intent prediction, and legal awareness alerts, CASF promotes responsible internet usage.
Aligned with the Information Technology Act, 2000 [1], Digital Personal Data Protection Act, 2023 [2], and CERT-In guidelines [3], the framework provides a structured and legally compliant pathway toward safer digital participation in India.
When implemented ethically and transparently, CASF can reduce cybercrime risks, improve digital literacy, strengthen citizen confidence, and contribute to a secure digital society.

References

1. Government of India, *Information Technology Act, 2000 (Updated Version)*.
2. Government of India, *Digital Personal Data Protection Act, 2023*, Ministry of Electronics & IT.
3. CERT-In, *Cyber Security Guidelines and Advisories*, Government of India.
4. National Crime Records Bureau (NCRB), *Crime in India Report*, Latest Edition.
5. Press Information Bureau (PIB), Government of India, Cyber Fraud Statistics Release.
6. Verizon, *Data Breach Investigations Report*, 2023.
7. IBM Security, *Cost of a Data Breach Report*, 2023.
8. S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, Pearson.
9. OECD, *AI Principles and Responsible AI Guidelines*, 2021.