Prof. Bhavya V, , 2025, 13:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

Guardian-AI: on-Device AI Security for Sensitive Data

Prof. Bhavya V, Jayanthan P, Kiran K G, M Jathin Reddy, Manu S,
Dept. of Computer Science and Engineering
Dayananda Sagar Academy of Technology and Management

Abstract- This paper introduces Guardian-AI, an innovative Android application designed to enhance on-device security and privacy for sensitive user data such as images, audio, documents, and clipboard text. Leveraging AI-powered content detection, the system performs local processing without any internet connectiv- ity, thereby eliminating network-based threats. Sensitive data is protected through a combination of AES-256 encryption, biomet- ric authentication, and controlled access mechanisms including blurred previews and protected clipboard management. The design prioritizes robust protection against unauthorized access, even in scenarios involving physical device compromise, while maintaining usability and performance on resource-constrained mobile devices. This work outlines the system architecture, key security features, and challenges associated with implementing a fully offline privacy-preserving solution.

Keywords - On-device AI, Data privacy, AES-256 encryp- tion, Android security, Elliptic Curve Cryptography (ECC), Biometric authentication, Sensitive data protection, Offline se- curity, Encrypted storage, Protected clipboard, Whisper AI, Key management.

I. INTRODUCTION

With the increasing reliance on mobile devices for storing and handling personal and sensitive data, concerns about data privacy and security have never been greater. Traditional security mechanisms often depend on cloud-based services, which expose user data to potential network vulnerabilities and third-party access. Furthermore, physical compromise of a device can lead to unauthorized data exposure if proper on- device protections are not in place.

This paper presents Guardian-AI, an innovative Android application designed to provide comprehensive, on-device se- curity for sensitive user data such as images, audio, documents, and clipboard text. By leveraging AI techniques for sensitive content detection and employing robust encryption protocols combined with biometric authentication, Guardian-AI ensures that data remains secure and private without requiring any internet connectivity. This offline-first approach significantly reduces attack surfaces associated with network-based threats. Guardian-AI addresses key challenges of secure storage, controlled access, and real-time protection of sensitive infor- mation on mobile devices, providing users with peace of mind even in scenarios involving physical device compromise.

II. BACKGROUND

This section outlines the foundational technologies and concepts critical to the design of secure, on-device AI systems for sensitive data protection.

Symmetric Encryption and AES

Symmetric encryption uses a single secret key for both encryption and decryption, making it computationally efficient for resource-constrained devices. The Advanced Encryption Standard (AES) is a widely adopted symmetric cipher known for its

security and performance, often employed in mobile and IoT environments for protecting sensitive data [1].

Elliptic Curve Cryptography (ECC)
ECC is an asymmetric cryptographic technique that offers strong security with shorter key lengths compared to tradi- tional methods like RSA. This results in faster computations and reduced storage requirements, making ECC suitable for key management on mobile and embedded devices [2].

Android Platform Security Model
Android employs a multi-layered security architecture that integrates application sandboxing, permission controls, hardware-backed keystore, and biometric authentication. This model balances usability and robust protection against a wide range of threats, providing a reliable foundation for secure app development [3].

Biometric Authentication on Android
Biometric techniques such as fingerprint and facial recog- nition enhance user authentication by leveraging physiolog- ical traits. Android provides standardized biometric APIs to facilitate secure and user-friendly authentication mechanisms, which are vital for protecting sensitive data on devices [4].

On-Device AI and Privacy Preservation
Executing AI inference and learning directly on devices re- duces latency and mitigates privacy risks associated with cloud processing. Techniques such as privacy-preserving neural net- works and compressive models enable efficient and secure AI operations suitable for mobile and edge environments [5]–[7].

III. RELATED WORK

Privacy-Preserving Computation and On-Device AI
Recent advancements in privacy-preserving image process- ing for mobile and edge devices highlight the importance of secure on-device computation. Huang et al. introduced "Find and Dig," a mechanism that preserves privacy in deep neural networks by protecting sensitive regions in images during mobile inference [5]. Complementing this, Yan et al. proposed a compressive privacy-preserving model for edge computing environments that balances privacy and utility by combining differential privacy and compression techniques in deep learning-based services [6].

The trend toward independent on-device artificial intelli- gence is gaining traction, with Wu and Hu exploring solutions enabling AI inference directly on mobile and embedded hard- ware. Their work addresses challenges such as latency, privacy, and energy efficiency without reliance on cloud services [7]. Voice recognition, a critical component in mobile security and user interaction, has been comprehensively surveyed by Chandolikar et al., who discussed various voice recognition approaches, their accuracy, computational requirements, and practical applications [8].

Cryptography and Secure Data Transmission
Lightweight symmetric encryption algorithms have been extensively reviewed for suitability in Internet of Things (IoT) and resource-constrained environments. Kumar and Pillai ana- lyzed various lightweight symmetric ciphers focusing on their security strength and efficiency, highlighting their applicability to mobile and edge platforms [1].

Security for mobile applications, especially in sensitive domains such as healthcare, has been enhanced by combining symmetric encryption (AES) with asymmetric key manage- ment techniques like Elliptic Curve Cryptography (ECC). Abdullah et al. proposed such cryptographic combinations to ensure confidentiality and efficient performance on mobile devices [2]. Furthermore, Chidambaranathan et al. developed an AI-based automated detection system for cryptographic algorithms, which improves security monitoring and forensic analysis [9].

Biometric Authentication and Android Security
Biometric authentication techniques on Android devices have been surveyed by Zhang et al., who detailed the strengths, limitations, and practical considerations of fingerprint, facial recognition, and other biometric methods [4].

At the system level, Mayrhofer et al. provided a formal analysis of the Android Platform Security Model, revealing how its multi-layered security architecture balances usability, performance, and threat mitigation on widely deployed devices [3].
Complementing this, Abdullah and Zeebaree reviewed com- mon vulnerabilities in Android mobile applications and pro- posed mitigation methods to assist developers in securing apps against prevalent attack vectors [10].
Together, these works establish a solid foundation for de- signing secure, privacy-aware, and efficient AI-driven applica- tions tailored for mobile and edge platforms.

IV. APPLICATIONS AND USE CASES

The Guardian-AI system is designed to address a wide range of real-world privacy and security scenarios where users handle sensitive data locally on their mobile devices. Below are key applications and practical use cases that demonstrate the relevance and versatility of the proposed solution.

Personal Privacy Management
Guardian-AI empowers individual users to protect their private documents, photos, audio notes, and text data with- out relying on external cloud storage. For instance, images containing passwords, credit cards, or confidential notes are automatically blurred, and can only be viewed after biometric authentication. This is particularly useful for users who prefer to store sensitive screenshots, scanned IDs, or personal notes directly on their phones.

Secure Communication and Redaction
For journalists, lawyers, or healthcare professionals who often record voice memos or receive voice messages con- taining sensitive names, locations, or IDs, Guardian-AI's on- device audio redaction can automatically detect and censor these entities using beeps. The same applies to stored texts or documents, where sensitive segments (e.g., emails, account numbers) are masked with asterisks until proper authentication is provided.

Field Applications in Low-Connectivity Environments
The system is particularly beneficial in environments where internet access is limited or restricted due to security or policy constraints (e.g., military zones, disaster relief zones, remote healthcare camps). Since Guardian-AI operates entirely offline, sensitive data collected in the field (e.g., patient records, incident photos, reports) remains encrypted and inaccessible to unauthorized parties even if the device is physically lost or compromised.

Enterprise and BYOD Security
For organizations that allow employees to use personal mobile devices under BYOD (Bring Your Own Device) poli- cies, Guardian-AI provides an added layer of data governance. Employees can securely store work-related documents, audio notes from meetings, or confidential emails on their devices without the risk of accidental leaks or unauthorized sharing, as access is gated behind encryption and biometrics.

Privacy-Conscious Content Sharing
Users who want to share documents, screenshots, or audio messages can use Guardian-AI to redact or blur sensitive portions before sharing. This ensures that shared content does not unintentionally expose private information. For example, a user can blur account numbers in a bank statement screenshot or mask personal details in a shared voice note.

V. THEMES IDENTIFIED IN LITERATURE SUREVY

Need for On-Device AI
Prior research emphasized the importance of executing AI tasks on-device to eliminate privacy risks from cloud-based processing. This led the team to design Guardian-AI with an offline-first architecture that performs all detection and redaction locally.

Selection of Lightweight and Privacy-Preserving Models
Studies on edge AI and compressive models influenced the use of lightweight neural networks for sensitive data detection, ensuring low power usage and fast inference on mobile devices.

Incorporation of AES and ECC
Research on cryptographic algorithms like AES (for sym- metric encryption) and ECC (for secure key management) supported the decision to use AES-256 for file encryption and ECC with the Android Keystore for secure key handling.

Multi-Layered Android Security Utilization
Android's security architecture, as analyzed in prior works, inspired the integration of features like app sandboxing, bio- metric APIs, and hardware-backed keystore, ensuring layered protection against threats.

ADOPTION of Biometric Authentication
Literature on biometric security methods highlighted their reliability and user-friendliness. This led to the integration of fingerprint and facial recognition for secure access to decrypted data.

Support for Multi-Modal Data Redaction: Research on OCR, speech recognition, and NER-based redaction showed the feasibility of sensitive data detection across formats. Guardian-AI thus supports image blurring, audio beeping, and text masking—inspired directly by these studies.

Decision to Avoid Cloud Dependence
Research warning about third-party and network-level data exposure led to the strict decision to avoid all forms of cloud storage, ensuring complete user data sovereignty.

Inspiration for Future Improvements
Literature on user feedback loops, hardware-level security, and real-time AI motivated future development ideas like adap- tive redaction, secure clipboard, and integration into OEM sys- tems.

VI. CONCLUSION AND FUTURE WORK

This survey examined foundational technologies and recent advancements underpinning on-device, AI-assisted privacy systems for mobile platforms. Core components such as AES- 256 encryption, Elliptic Curve Cryptography (ECC), biometric authentication, and Android's secure architecture constitute the primary building blocks for secure mobile computing. Recent studies affirm the growing feasibility of offline AI inference, which enhances privacy while maintaining performance and usability.

The literature further highlights the role of lightweight neural networks in facilitating real-time processing of sensitive multimodal data—including text, images, and audio—when integrated with biometric access controls and secure local storage. Collectively, these developments indicate strong po-tential for constructing privacy-preserving, offline-capable mo- bile applications, particularly in scenarios with limited or no connectivity.

However, a notable gap remains: the absence of a com- prehensive, end-to-end system that cohesively integrates these technologies into a seamless and user-centric solution. Ad- dressing this challenge involves exploring the trade-offs in system complexity, resource constraints, and user experience. Future research should focus on the practical deployment of such unified frameworks, potentially paving the way for robust privacy-first architectures like Guardian-AI. The insights con-solidated in this review provide a foundation for guiding the development and evaluation of next-generation secure mobile systems.

Future Work
Several key areas are identified for future enhancement:
Real-Time Processing: One major direction for fu-ture development involves transitioning from batch-based redaction to fully real-time processing capabilities. This would allow sensitive information in images, audio, or documents to be detected and redacted immediately as the content is being generated, such as during a photo capture, voice recording, or message receipt. Achieving this requires optimizing inference pipelines to operate with minimal latency, integrating with Android's real-time data flow systems (e.g., CameraX, MediaRecorder, or Accessibility APIs), and ensuring continuous on- device operation without compromising user experience or device responsiveness.

Optimized On-Device Models: Replacing large, server- grade neural networks with compact, resource-efficient models is critical for mobile deployment. Techniques such as quantization, pruning, knowledge distillation, and architecture search (AutoML) can be used to compress models without significantly degrading accuracy. These optimizations allow AI components to run efficiently on- device using TensorFlow Lite or ONNX runtimes, and leverage mobile-specific accelerators like DSPs, NPUs, and GPUs via NNAPI. The goal is to maintain high privacy-preserving inference capabilities while signifi- cantly reducing memory usage, CPU load, and power drain.

Secure Clipboard Management: The system clipboard is a frequent source of accidental data leakage in mo- bile applications. A future-ready privacy system should incorporate a protected clipboard layer that automatically encrypts any sensitive data that is copied, tags it with con- tent origin and expiration metadata, and requires biomet- ric or passcode verification for pasting. This functionality must operate system-wide, intercepting clipboard events and maintaining access logs locally for user review. Such a mechanism would ensure that sensitive content—such as passwords, financial data, or personal identifiers—is never exposed unintentionally.

Robust Key Management: Effective end-to-end privacy requires secure handling of encryption keys across the entire data lifecycle. Future implementations must en- hance Android Keystore integration to support secure generation, binding to biometric credentials, backup and recovery via encrypted vaults, and revocation in case of compromise. The use of hardware-backed keys (e.g., via StrongBox or TrustZone) ensures keys never leave the secure enclave. Additionally, systems should offer user- transparent key rotation and hierarchical key derivation schemes to separate data classes while maintaining min- imal user friction.

Adaptive User Feedback Loop: While automatic redac- tion can reduce user burden, it also introduces the risk of false positives and negatives. A privacy-centric system should incorporate an adaptive feedback loop that allows users to verify, correct, or refine redacted content across modalities. These corrections can be securely stored and used to incrementally update the model or pattern-matching rules locally, without ever leaving the device. This not only increases accuracy over time but also respects user control and trust, establishing a feedback- based personalization mechanism that is compliant with privacy-by-design principles.

Unified User Interface: A fragmented or technically overwhelming UI can be a major barrier to adoption. A future-ready system must consolidate all privacy-related features into a single, cohesive Android interface that presents clear workflows and consistent design language. This includes modules for real-time redaction previews, secure data vaults, access control settings, activity logs,
model accuracy feedback, and customization of privacy rules. Accessibility and usability testing should guide interface design to ensure that security features remain understandable and actionable for non-expert users.

Hardware and OS-Level Integration: For deeper and more seamless privacy, Guardian-AI's concepts should be explored for implementation at the firmware and OS level. Emerging mobile chipsets increasingly include AI cores (e.g., Google's Tensor or Apple's Neural Engine) that support secure, low-power, offline inference. Em- bedding redaction pipelines and data classification tasks into such hardware could allow Guardian-AI to operate invisibly in the background, integrated with the OS's media framework and file system, offering privacy as a native feature without compromising speed, battery life, or user experience.

The Guardian-AI system aims to become a practical, trust- worthy tool for users who demand control over their private information without relying on external services. As develop- ment progresses, our focus remains on maintaining privacy by design, strong encryption, and seamless user experience.

REFERENCES

1. S. Kumar and C. Pillai, "An analysis of light weight symmetric encryp- tion algorithms for secure

data transmission in iot," in 2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS). IEEE, 2024.

2. H. S. Abdullah, O. O. Khalifa, and A. H. A. Hashim, "Enhanced mobile app security for healthcare applications," in 2024 9th International Conference on Mechatronics Engineering (ICOM). IEEE, 2024.
3. R. Mayrhofer, J. Stoep, C. Brubaker, and N. Kralevich, "The android platform security model," ACM Transactions on Privacy and Security, vol. 24, pp. 1–35, 04 2021.
4. X. Zhang, T. He, and X. Xu, "Android-based smartphone authentication system using biometric techniques: A review," in 2019 International Conference on Cyber-Security and Resilience (CRC). IEEE, 2019.
5. H. Huang, H. Zhao, C. Hu, C. Chen, and Y. Li, "Find and dig: A privacy- preserving image processing mechanism in deep neural networks for mobile computation," in 2021 International Joint Conference on Neural Networks (IJCNN). IEEE, 2021.
6. Y. Yan, Q. Pei, and H. Li, "Privacy-preserving compressive model for enhanced deep-learning-based service provision system in edge computing," IEEE Access, vol. 7, pp. 92 921–92 937, 2019.
7. Y. Wu and J. Hu, "Towards independent on-device artificial intelligence," in 2022 IEEE Computer Society Annual Symposium on VLSI (ISVLSI). IEEE, 2022.
8. N. Chandolikar, A. Gawas, C. Joshi, P. Roy, and M. Vishwakarma, "Voice recognition: A comprehensive survey," in 2022 International Mobile and Embedded Technology Conference (MECON). IEEE, 2022.
9. S. Chidambaranathan, M. Santhanaraj, E. Ajitha, S. F. A. S, S. B, and

   T. Kathirvel, "Automated detection of encryption algorithms using ai techniques," in 2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS). IEEE, 2024.
10. H. Abdullah and S. R. M. Zeebaree, "Android mobile applications vul- nerabilities and prevention methods: A review," in 2021 2nd Information Technology To Enhance e-learning and Other Application (IT-ELA). IEEE, 2021.