MEENAKSHI COLLEGE OF ENGINEERING

(RUN BY MEENAKSHI AMMAL TRUST)

No.12.Vembuli Amman Koil Street, West K.K. Nagar,Chennai – 600 078

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

ScamX

Context-Aware AI-Driven Mobile Security Architecture for Dynamic Fraudulent Message Analysis and Mitigation

Guided by,
Aswathy M P Assistant Professor (CSE)

Team Members:
1. Achyuth Sudharshan N (311421104002)
2. Alina Muskhan M I (311421104003)
3. Neha R (311421104056)
4. Bavithran V M (311421104010)

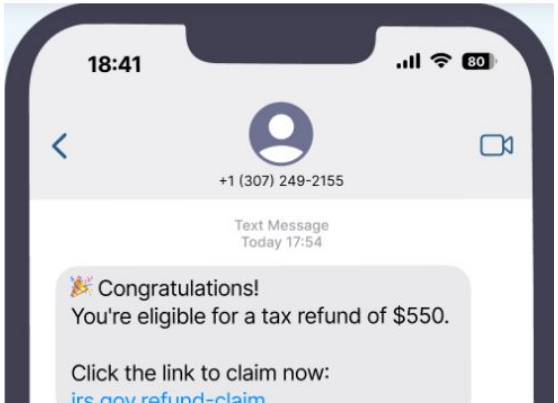
TABLE OF CONTENTS

OBJECTIVE

- Detect fraudulent/spam messages in SMS, WhatsApp, and Telegram notifications.
- Work partially online without storing user data for privacy.
- Leverage an optimized Mobile-BERT model for efficient and accurate real-time scam detection.
- Help users identify scam messages before engaging with them.
- Provide a user-friendly and intuitive interface for seamless fraud detection.
- Enable real-time message scanning without noticeable delay in notifications.

PROBLEM STATEMENT

Fraudulent SMS and scam messages, including phishing attempts and OTP frauds, pose significant financial risks to users. Traditional spam filters struggle to detect evolving scam tactics, rely on internet connectivity, and require high computational resources, raising privacy and security concerns. Currently, no AI-powered, cross-platform fraud detection application exists that efficiently identifies fraudulent messages across SMS and messaging platforms while minimizing resource consumption.



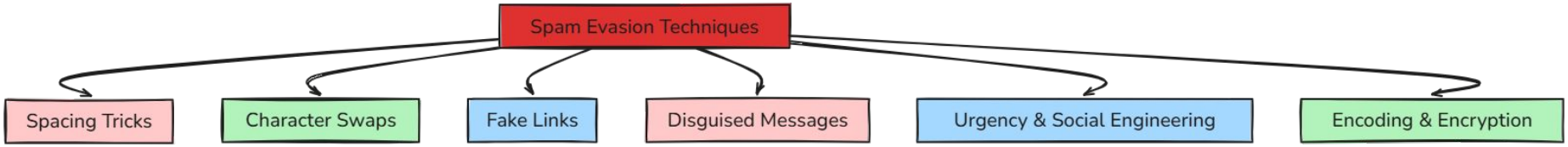
ABSTRACT

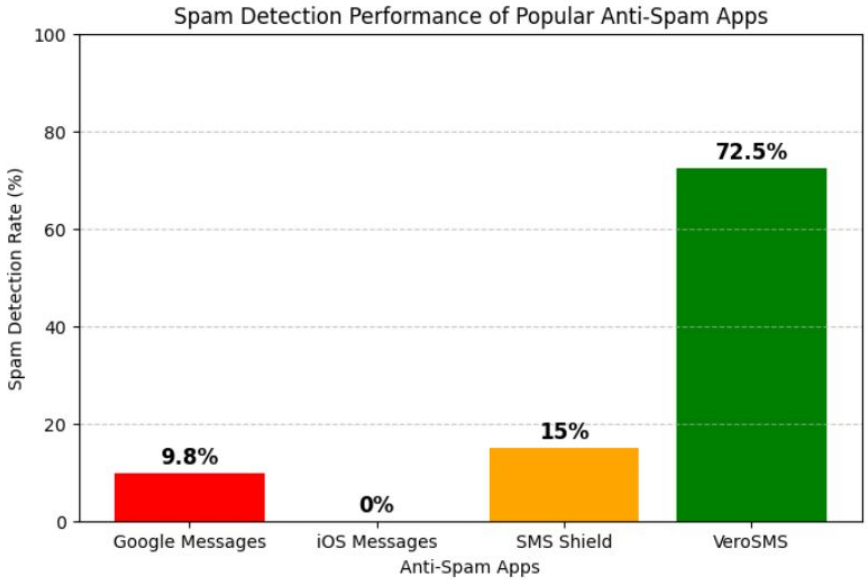
ScamX is a real-time scam detection system designed to identify fraudulent messages across SMS, WhatsApp, and Telegram. It combines a lightweight Mobile-BERT model for efficient, accurate message classification with a secure Flask backend that supports partial online processing via HTTPS—ensuring user privacy and low-latency performance. The system features an intuitive, cross-platform Flutter interface that enables real-time message scanning without noticeable delays. By minimizing false positives and negatives, ScamX builds user trust while helping individuals recognize scams before engaging, making it a practical, privacy-focused solution to modern digital fraud.

Existing System

- **Rule-Based Filters** – Used by Google Messages, Truecaller, etc.
- **Cloud Spam Detection** – Services like Hiya, RoboKiller need internet and access user data
- **Bank & Telecom Filters** – Built-in phishing detection systems are inconsistent and often ineffective.
- **Heavy AI Models** – BERT, GPT-based detection systems.
- **SMS-Only Focus** – Most tools work only on SMS.


LIMITATIONS OF TRADITIONAL SCAM DETECTION
EASILY BYPASSED
Fixed rules can't keep up with evolving scams
PRIVACY RISKS
Cloud tools collect user data and intertenet
LOW ACCURACY
Bank and telecom filters often miss real threats
LIMITED APP COVERAGE
No scam detection on WhatsApp or Telegram
NOT MOBILE-FRIENDLY
Advanced AI models need high computing power



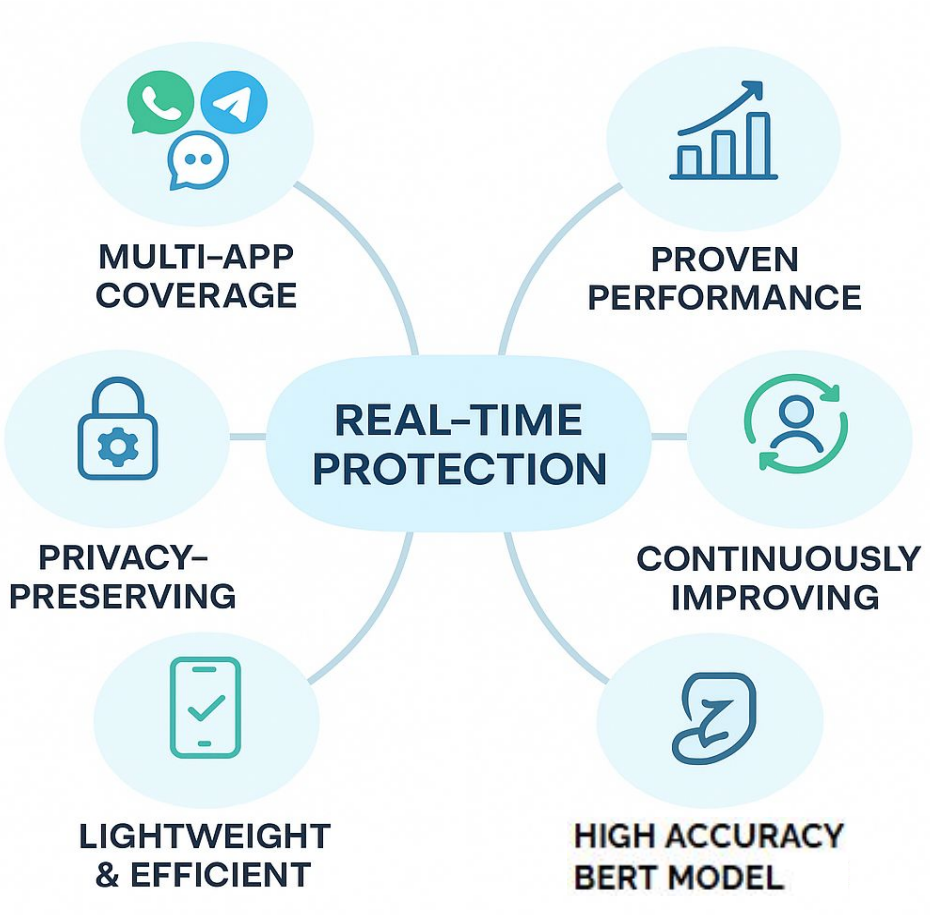


	class_label	message	length
0	ham	Go until jurong point, crazy.. Available only ...	111
1	ham	Ok lar... Joking wif u oni...	29
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...	155
3	ham	U dun say so early hor... U c already then say...	49
4	ham	Nah I don't think he goes to usf, he lives aro...	61

Proposed System - ScamX

- **Mobile Application** designed for real-time scam detection on Android devices.
- **Comprehensive Coverage** of SMS, WhatsApp, and Telegram message interception.
- **Hybrid AI Architecture** utilizing Mobile-BERT backend and a lightweight Flask API for efficient and context-aware classification.
- **Adaptive Learning Mechanism** leveraging user feedback to continuously refine a personalized scam database.
- **Secure Cloud Synchronization** with MongoDB for persistent and accessible message history across sessions.

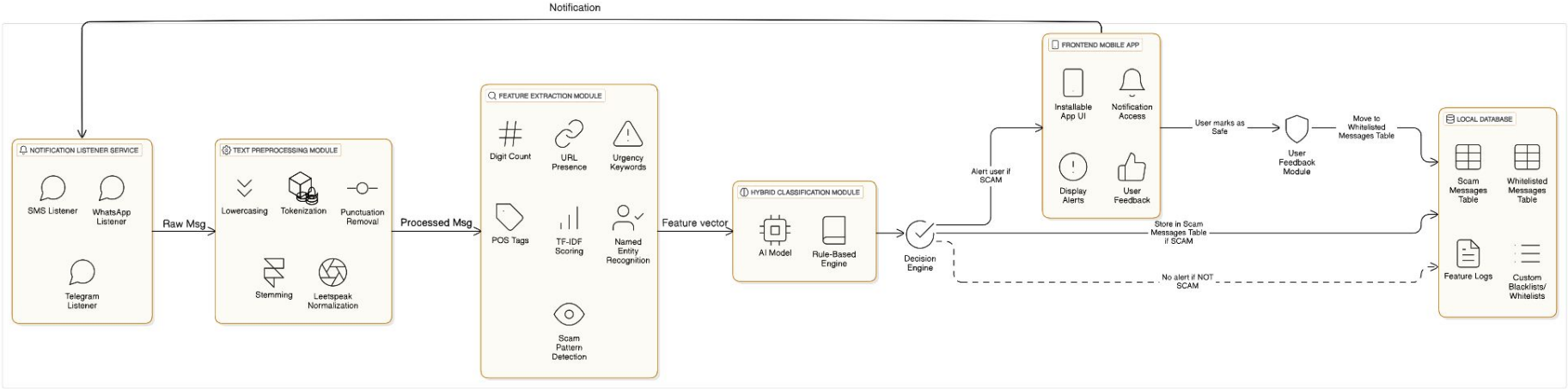
Advantages


MULTI-APP COVERAGE
PROVEN PERFORMANCE
REAL-TIME PROTECTION
PRIVACY-PRESERVING
CONTINUOUSLY IMPROVING
LIGHTWEIGHT & EFFICIENT
HIGH ACCURACY BERT MODEL

LITERATURE REVIEW

Author(s)	Year	Title	Methodology / Model Used	Key Findings / Contributions	Limitations	How ScamX Solves It
Almeida et al.	2011	SMS Spam Collection Dataset	Public SMS spam dataset	Provided a standard dataset for evaluating spam detection models.	No real-time or cross-platform capability.	ScamX uses real-time, multi-platform scanning (SMS, WhatsApp, Telegram).
Huang et al.	2019	MobileBERT: A Compact BERT for Mobile Devices	MobileBERT (Transformer-based)	Efficient inference on mobile devices with near BERT-level performance.	Doesn't address message context or spam-specific use cases.	ScamX customizes MobileBERT for contextual scam detection in notifications.
Sah & Kumar	2021	Spam Detection in WhatsApp Using ML	Naive Bayes, SVM	Feasibility of spam detection in WhatsApp messages.	Only supports WhatsApp, lacks scalability and real-time alerts.	ScamX supports SMS, WhatsApp, and Telegram with real-time notification scanning.
Singh & Sharma	2022	Hybrid Spam Detection Models	On-device + cloud ML	Hybrid design balances offline speed with online accuracy.	No implementation or privacy measures.	ScamX uses partial online processing with secure HTTPS and private on-device inference.
Salman, Ikram & Kaafar	2024	Investigating Evasive Techniques in SMS Spam Filtering	Comparative analysis of ML models	Explored how evasion weakens traditional filters.	Traditional models struggle against new scam techniques.	ScamX adapts with feedback loop & real-time model updates to stay effective.

BLOCK DIAGRAM



MODULES

Scam Detection Engine

- **Model Used:** MobileBERT (JeswinMS4/scam-alert-mobile-bert)
- **Functionality:** Classifies messages as **Dark** (scam) or **NotDark** (safe)
- **Tech Stack:** Utilizes Hugging Face's text-classification pipeline
- **Optimization:** Designed for mobile devices, lightweight and efficient, with no extra training required

The model was fine-tuned on a labeled scam dataset using:

- **Loss Minimization**
- **Backpropagation**
- **Label Supervision** (Dark vs. NotDark)

STORAGE

Smart message storage for user safety and learning

- Uses Google MongoDB, a cloud-based NoSQL database.
- Stores classified messages in two collections: scam messages and whitelisted messages.
- Ensures real-time synchronization between the app and backend.
- Provides organizational clarity for easier data retrieval.
- Supports historical review of scam messages across sessions.
- Enables scalability with efficient data handling and Ensures persistent storage, secure access control, and a smooth user experience.

MESSAGE HISTORY

Easy access to your message classification history

- Displays only scam (Dark) messages to keep the interface focused and clutter-free.
- Users can filter messages by platform – SMS, WhatsApp, or Telegram.
- Makes it easier to trace which scam came from where, improving clarity.
- Helps users stay aware of repeated scams or suspicious trends over time.

ALERT SYSTEM MODULE

Warn users instantly about potential scam messages

- If a message is classified as Dark with ≥ 75% confidence, an alert is sent from the app as a notification or popup externally.
- The scam message will be stored in the Message History section along with others.
- Prevents users from clicking or replying to risky messages.
- Boosts real-time scam awareness and protects users from threats.

USER FEEDBACK

Manual Correction and Continuous Learning

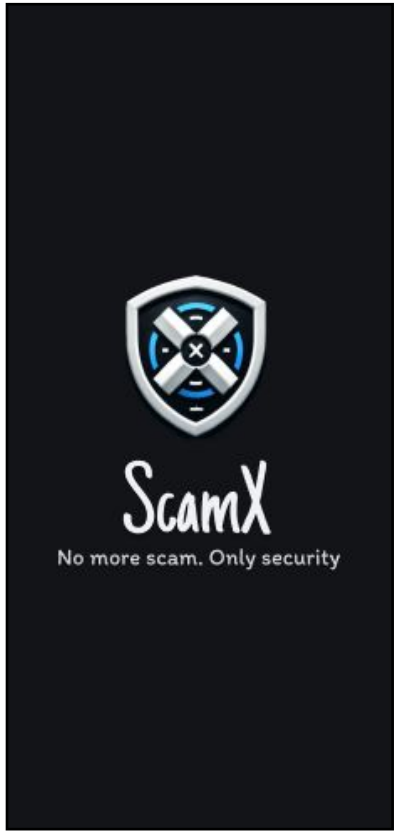
- Users can manually mark false positives as Safe in the app.
- This feature helps in correcting any misclassifications made by the AI model (false positives/negatives).
- Facilitates user-driven scam reporting via an integrated Google Form interface.
- Collected reports are archived in a Google Sheet to inform future model enhancements.

DATASET USED

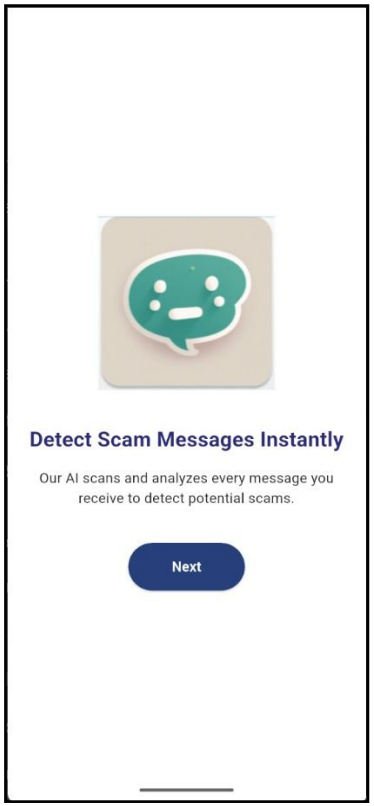
- **Kaggle –** SMS Spam Collection (Public dataset)
- **Custom Dataset 1 –** 3000 messages (SMOTE-balanced)
- **Custom Dataset 2 –** 1000 messages (manually curated)

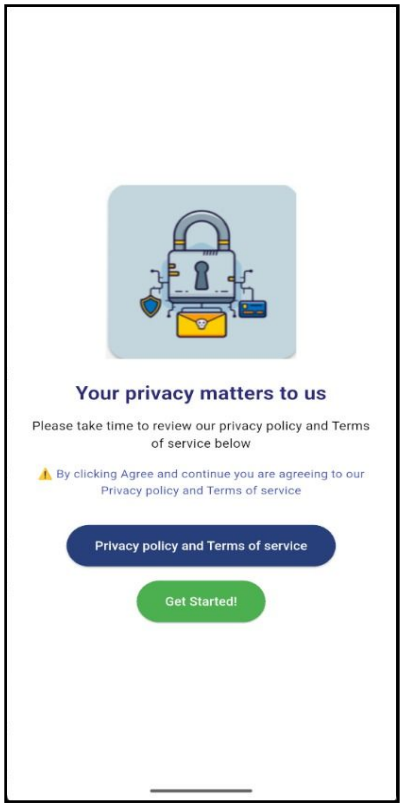
Message	Label	Source
Invest â,¹5000 and get â,¹50,000 in a week. DM now!	Dark	Telegram
Youâ€™ve been selected for a government grant. Apply here!	Dark	WhatsApp
Forward this to 10 people and win a free iPhone!	Dark	WhatsApp
Live session starts at 5 PM. Donâ€™t miss it!	NotDark	Telegram
Youâ€™ve been selected for a government grant. Apply here!	Dark	WhatsApp
Free data recharge! Only valid for today: scamlink.co	Dark	WhatsApp
Reminder: Project deadline is next Monday.	NotDark	Telegram
Happy birthday! ðŸŽ‰ Wish you lots of joy.	NotDark	WhatsApp

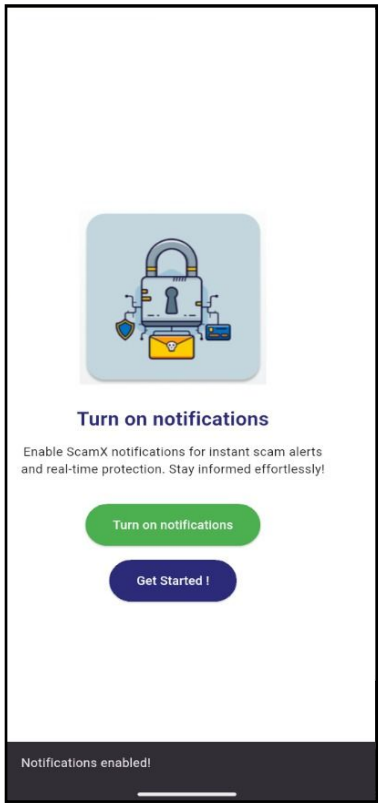
OUTPUT - Splash screen and App launcher Icon



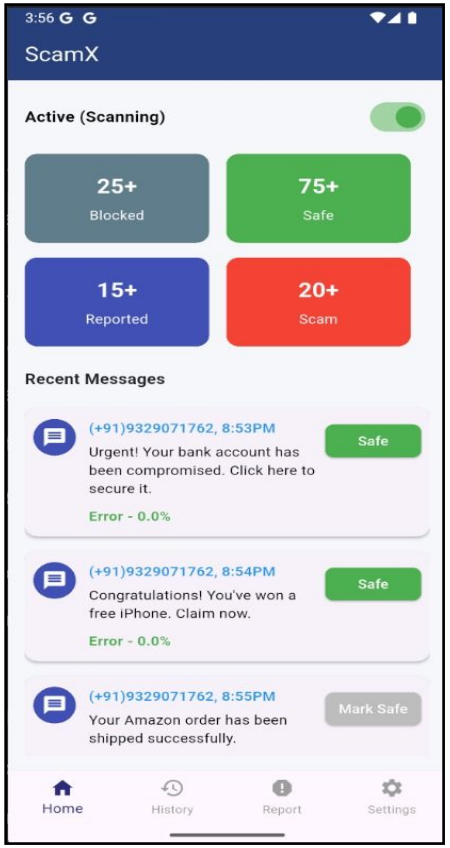
Onboarding screens

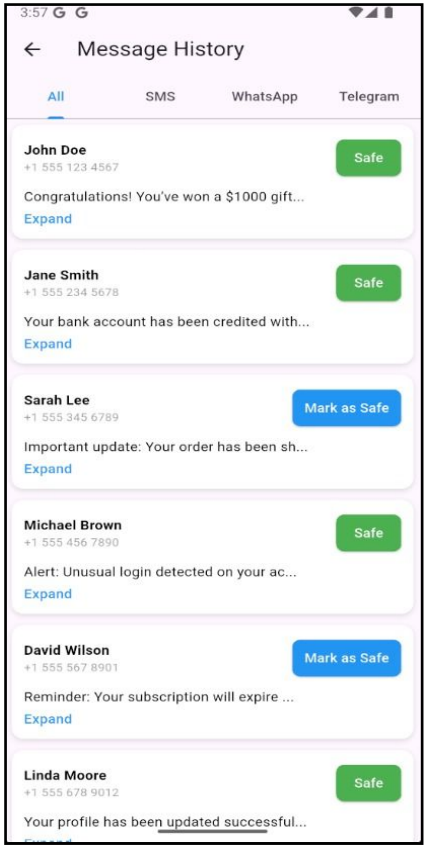


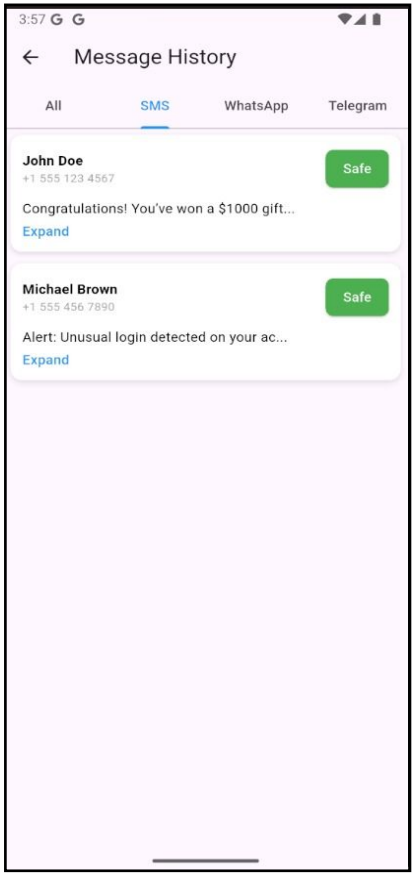


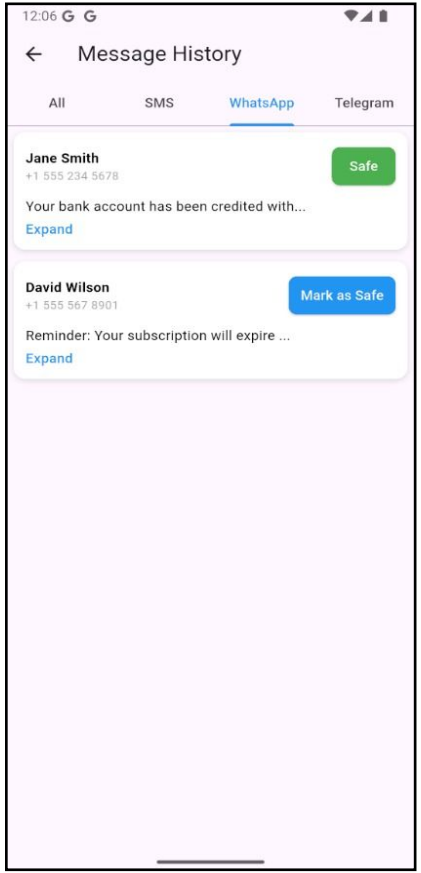


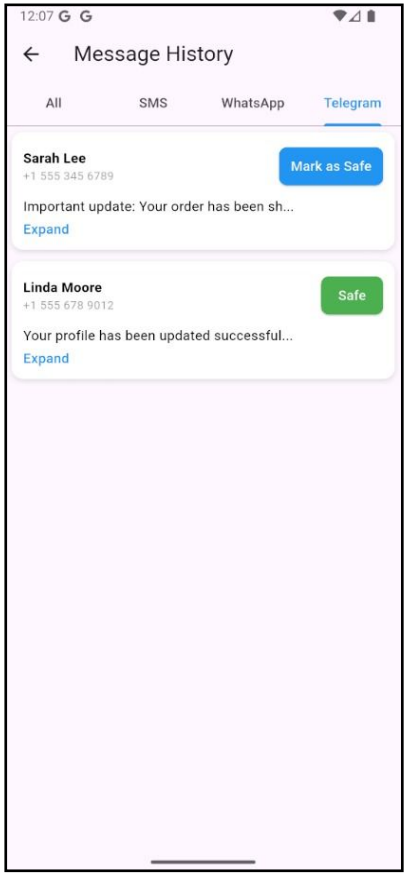
User dashboard and Message history screen



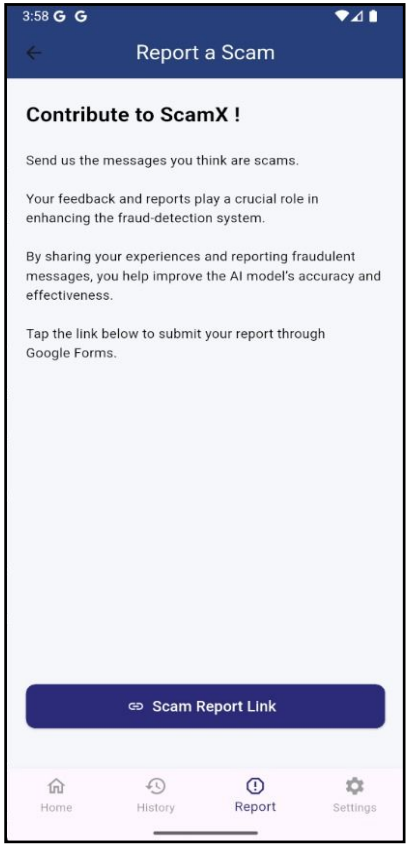


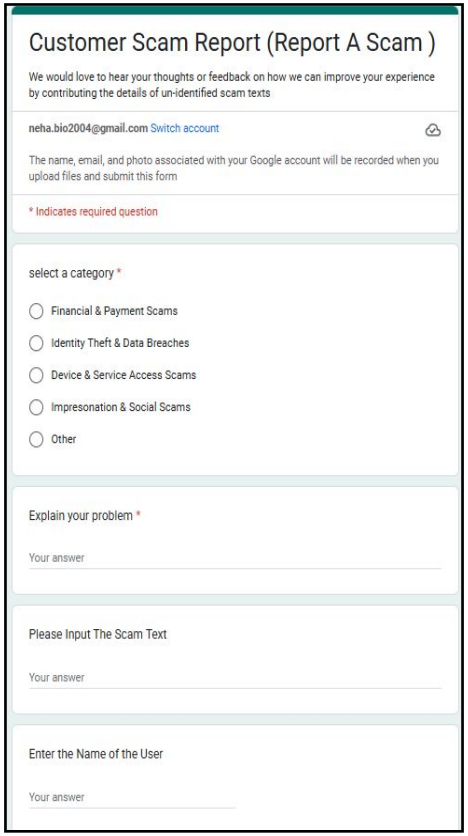


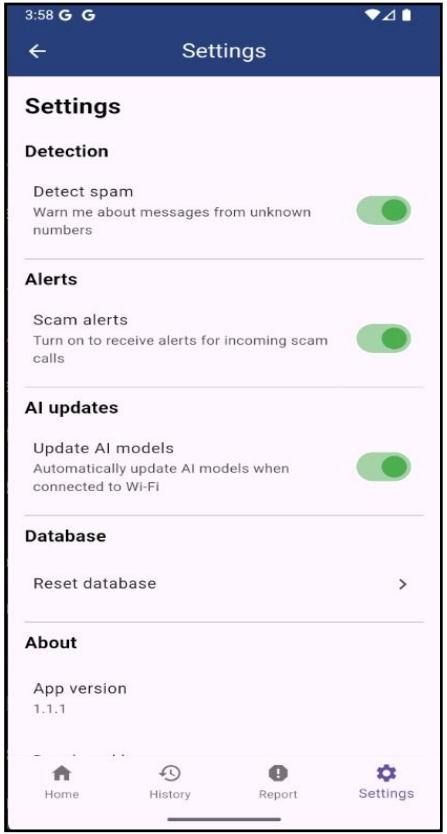




Filtering Messages based on platform







Report a scam screen along with the Google form

ScamX Analysis

Cross-Platform UI

Built with Flutter for a smooth, fast UI on all devices

Real-Time AI Detection

Mobile-BERT ensures fast, smart scam detection

Secure Flask Backend

Deployed on Render with HTTPS & auto scaling

Google MongoDB

Real-time sync + message history for trust

Hybrid Online/Offline

Combines on-device & backend detection

User Feedback Loop

Improves accuracy with user input like "Mark as Safe"

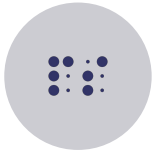
CONCLUSION

- ScamX enables real-time scam detection across SMS, WhatsApp, and Telegram.
- Built with Flutter, Mobile-BERT, and Flask for a fast, secure experience.
- Supports partial online processing to balance privacy and performance.
- Cloud-backed with MongoDB to store history and handle data reliably.

FUTURE ENHANCEMENTS

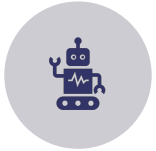
On-device inference:

Enables offline scam detection using lightweight transformer models.



Multilingual support:

Detects scams in regional languages for broader coverage.



Adaptive scam pattern learning:

Learns from user reports to stay updated and context-aware.

In-app analytics dashboard:

Visualizes scam trends and user reporting activity.

Expanded platform compatibility:

Supports more apps and wearables for real-time alerts.

REFERENCES

Dataset: SMS Spam Collection Dataset

AI Model: https://huggingface.co/JeswinMS4/scam-alert-mobile-bert

UI Design: https://www.figma.com/

Hugging Face : https://huggingface.co/docs/transformers/main_classes/pipelines

Google Colab – https://colab.research.google.com/

(SMOTE) – https://imbalanced-learn.org/stable/

Scikit-learn – https://scikit-learn.org/stable/

THANK YOU