SafeRide: AI-Powered Driver Safety Analysis and SOS Alert System for Passenger Protection in Ride-Hailing Services

Keith Dias
Computer Engineering Dept.
FCRIT, Navi Mumbai, India
keithdias07@gmail.com

Dylan Braganza
Computer Engineering Dept.
FCRIT, Navi Mumbai, India
dylan.brag06@gmail.com

Parth Deshmukh
Computer Engineering Dept.
FCRIT, Navi Mumbai, India
Parthdekhmukh111@gmail.com

Aris Arun
Computer Engineering Dept.
FCRIT, Navi Mumbai, India
arisarun52@gmail.com

Shirin Matwankar
Assistant Professor
Computer Engineering Dept.
FCRIT, Navi Mumbai, India

Abstract—**The rapid growth of ride-hailing platforms has improved urban mobility but raised concerns about passenger safety. Existing systems rely on basic ratings and unstructured reviews that fail to provide actionable risk insights. SafeRide is a Flutter- and Firebase-based Android application designed to address this issue by offering a data-driven driver safety assessment before rides. It uses the OpenAI GPT-4o-mini API to analyze driver reviews and classify incidents into severity levels. These insights are combined with six performance metrics through a weighted model to generate a Final Safety Score (0–100). The app also includes a real-time SOS feature that sends GPS-linked alerts to emergency contacts via Twilio SMS. Testing on simulated Firestore data shows reliable classification, scoring, and alert functionality, enhancing pre-ride safety awareness.**

Index Terms—**Flutter, Firebase Firestore, GPT-4o-mini, Twilio API, Passenger Safety, NLP, Safety Score, SOS Alert**

1 Introduction

1.1 Background

Ride-hailing services have become an indispensable component of modern urban transportation infrastructure. Platforms such as Uber and Ola operate millions of trips daily across India and globally, providing passengers with convenient, on-demand mobility. Despite their scale, these platforms offer limited built-in safety mechanisms beyond a basic star rating and an optional comment field. These limited feedback mechanisms fail to surface nuanced behavioural patterns — such as consistent speeding, inappropriate conduct, or repeated panic-button activations — that are far more predictive of future safety risk than a single numerical rating. The absence of systematic, AI-powered analysis of accumulated review data means that valuable safety intelligence embedded within passenger feedback remains largely untapped.

1.2 Motivation

The motivation behind SafeRide stems from two converging realities. First, incidents of rash driving, verbal harassment, and physical assault in ride-hailing contexts continue to be reported worldwide, disproportionately affecting solo travellers and late-night commuters. Second, the maturity of large language models now makes it feasible to perform semantic risk classification at scale with minimal cost and low latency. Traditional manual review of driver feedback is impractical; GPT-powered batch analysis resolves this challenge efficiently. Current systems lack a proactive safety check before boarding. — as opposed to reactive post-incident reporting — represents a design gap that this project directly addresses. The integration of real-time GPS-based SOS functionality further extends protection into the ride itself, offering passengers a credible emergency recourse independent of the booking platform.

1.3 Aim and Objectives

The primary aim of SafeRide is to build a passenger-centric mobile application that transforms unstructured ride history and textual feedback into a transparent, trustworthy safety signal. The specific objectives of the project are as follows:

- To design and implement an AI-powered driver safety scoring system that aggregates six weighted performance parameters into a single composite score.
- To integrate the GPT-4o-mini API for semantic analysis of accumulated passenger reviews, enabling automatic classification of incidents by severity level.
- To implement a real-time SOS emergency alert mechanism that captures GPS coordinates and dispatches an SMS to a registered emergency contact via the Twilio API.
- To develop a complete user-facing mobile application in Flutter with Firebase Authentication and Cloud Firestore, following Clean Architecture principles.
- To validate the system's correctness and reliability using simulated Firestore data representing diverse driver behavioural profiles.

2 Literature Review

Research in the domain of ride-hailing safety, emergency response systems, and AI-driven feedback analysis provides a strong foundation for the design choices made in SafeRide. Chaudhry et al. [1] conducted a comprehensive examination of passenger insecurity in ride-sharing ecosystems, identifying weak regulatory oversight, inadequate driver vetting, and deficient in-app feedback mechanisms as the primary contributors to harassment and assault incidents. Their findings underscore the need for an intelligent, automated layer of safety intelligence that supplements existing platform-level controls.

Dyer [2] highlighted how the informal nature of ride-booking relationships creates inherent safety asymmetries. Loukaitou-Sideris [3] approached the problem through a gender-safety lens, demonstrating that fear of victimisation measurably reduces transport utilisation among women and advocating for design-level interventions that communicate safety signals prior to boarding. These studies collectively motivate the pre-boarding safety check that is central to SafeRide's value proposition.

On the emergency response side, Javed and Luo [4] proposed an SOS system combining voice recognition with GPS positioning and automated messaging, demonstrating that a single-interaction trigger can significantly reduce emergency response latency. SafeRide adapts this design principle by replacing voice activation with a deliberate three-second hold gesture that minimises accidental triggering while retaining immediacy. Miller and Loukakos [5] identified data fragmentation across transit modes as a barrier to effective intelligent transport integration, reinforcing SafeRide's architectural decision to consolidate all driver, ride, and review data within a single Firestore data model.

Jaydarifard, Yigitcanlar, and Paz [6] conducted a recent empirical analysis of risk factors in taxi and ride-hailing contexts, identifying behavioural patterns—including speeding frequency and repeated complaint types—as the strongest predictors of incident risk. This directly informed the logarithmic incident risk formula adopted in SafeRide's scoring engine. The existing commercial systems—Uber [7], Ola [8], bSafe [9], and My Safetipin [10]—each address subsets of this problem space without offering a unified, AI-driven pre-boarding assessment combined with an integrated SOS mechanism.

3 Problem Statement

Despite the widespread adoption of ride-hailing services, passengers currently lack access to a comprehensive, AI-driven pre-boarding safety assessment. Existing platforms aggregate driver ratings into a single scalar value that neither captures behavioural severity nor detects recurring risk patterns across a driver's history. Passenger reviews — which often contain critical safety-relevant information about reckless driving, inappropriate conduct, or emergency situations — are stored but not semantically analysed. Furthermore, no major ride-hailing application offers an integrated, GPS-linked SOS mechanism that operates independently of the booking platform, leaving passengers without a reliable emergency recourse during the ride. SafeRide addresses this three-fold gap: the absence of intelligent review analysis, the absence of a multi-parameter quantitative safety score, and the absence of an in-ride emergency alert system.

4 Proposed System Architecture

SafeRide is structured across four distinct functional layers that interact in a clearly defined data flow, ensuring separation of concerns and enabling independent modification of any layer without affecting the others.

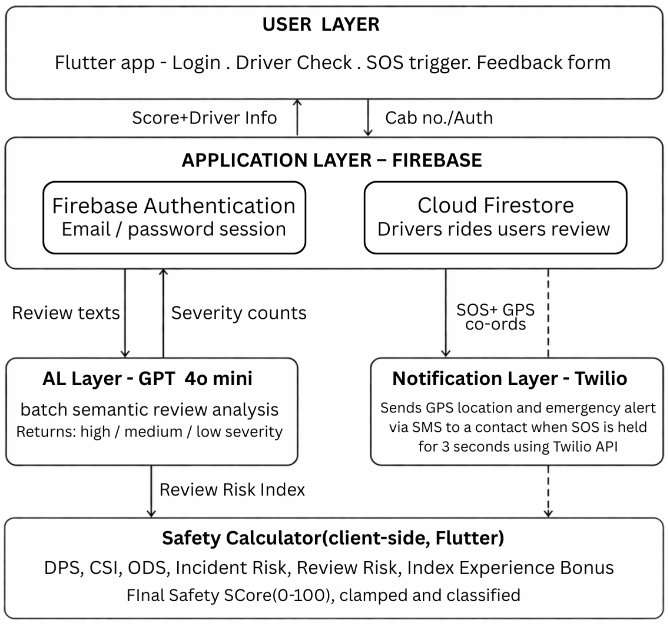

Fig. 1. SafeRide System Architecture

4.1 User Layer

The User Layer comprises the Flutter-based mobile application running on an Android device. This layer is responsible for all user-facing interactions: account registration and login, cab number entry, display of the computed safety score and its component breakdown, SOS gesture detection, and post-ride feedback submission. All business logic is kept strictly outside this layer; the UI consumes state exposed by Riverpod providers and renders it without direct knowledge of backend operations.

4.2 Application Layer — Firebase

The Application Layer is built on Google Firebase and serves as the central data store and authentication hub. Firebase Authentication manages email-and-password credential verification and session persistence. Cloud Firestore stores all persistent data across two root collections: users (containing profile information and emergency contact details) and drivers (containing driver metadata and a rides subcollection per driver, each ride document holding all quantitative parameters and the GPT analysis result). The Firestore document model enables efficient single-query retrieval of all ride history for a given driver, essential for performing aggregate AI analysis in a single API round-trip.

4.3 AI Layer — OpenAI GPT-4o-mini

The AI Layer performs semantic analysis of passenger reviews. When a driver is queried, all associated review texts are retrieved from Firestore and combined into a single prompt submitted to the GPT-4o-mini API. The model returns counts of high-, medium-, and low-severity incidents, which are used in the Review Risk component of the scoring engine. Batch processing ensures low latency and cost efficiency.

In the current prototype, API calls are made directly from the client application to simplify development within academic constraints. However, this approach is not secure for production, as API keys embedded in mobile applications can be extracted from APK files. In a real-world deployment, a secure backend proxy (e.g., Firebase Cloud Functions or a dedicated server) would be used to handle API requests and protect credentials.

4.4 Notification Layer — Twilio SMS

The Notification Layer is activated exclusively upon SOS trigger. When the passenger holds the SOS button for three continuous seconds, the Geolocator plugin captures the device's current GPS coordinates. These coordinates, along with a pre-composed safety message, are submitted to the Twilio REST API, which dispatches an SMS to the emergency contact number registered in the user's Firestore profile.

4.5 Data Flow

A passenger authenticates via Firebase Auth and enters a cab number. The Driver Repository queries Firestore for the driver document and all associated ride sub-documents. The Safety Calculator applies the six scoring formulas client-side to produce the DPS, CSI, ODS, Incident Risk, Review Risk, and Experience Bonus components, summed into the Final Score and displayed on the Driver Risk Screen. Concurrently, all review texts are forwarded to GPT-4o-mini; the returned severity counts update the Review Risk component in real time. If the passenger triggers SOS, the SOS Provider captures GPS coordinates and dispatches a Twilio SMS to the emergency contact.

5 Methodology

5.1 Data Collection

Driver and ride data are organised within Cloud Firestore. Each driver document stores static attributes — cab number, model, colour, gender, and years of experience — while the rides subcollection holds per-trip records capturing ten quantitative parameters: driving skill rating, behaviour rating, cleanliness rating, on-time pickup rate, acceptance rate, cancellation rate, accident count, speeding reports, panic button usage, and a freeform review text. For development and evaluation, a Node.js seeding script populates Firestore with a representative dataset of diverse behavioural profiles.

5.2 AI Processing — GPT Review Analysis

All review texts associated with a queried driver are extracted from Firestore and concatenated into a single structured prompt. GPT-4o-mini is instructed to perform aggregate semantic analysis and return a JSON object containing three integer counts: high-severity incidents (threats, assault, dangerous driving), medium-severity incidents (reckless overtaking, inappropriate behaviour, speeding complaints), and low-severity incidents (minor rudeness, slight discomfort). This structured output is parsed and stored in a GptAnalysisModel object consumed by the scoring engine.

5.3 Safety Score Calculation

The overall safety score is computed on the client side by combining multiple performance indicators, as summarised in Table 1. Each component captures a distinct aspect of driver behaviour and service quality, and is weighted to reflect its relative importance in determining passenger safety.

For brevity, the table uses abbreviated variables. Here, S, B, and C denote Skill, Behaviour, and Cleanliness respectively. OT represents on-time performance, A refers to ride acceptance rate, and Cxl denotes cancellation rate. The incident-related variables include Acc (accident rate), Spd (speeding rate), and Pan (panic/SOS trigger rate). Review-based inputs are represented as H, M, and L for high-, medium-, and low-severity feedback counts, with T denoting the total number of rides. Y indicates the driver's years of experience.

The Driver Performance Score (DPS), Customer Satisfaction Index (CSI), and Operational Dependability Score (ODS) are positively weighted components that reward consistent and reliable driving behaviour. In contrast, the Incident Risk term introduces a logarithmic penalty based on the frequency of unsafe events, ensuring that repeated violations have a progressively higher impact while avoiding excessive sensitivity to outliers. The Review Risk Index is incorporated indirectly through incident-related measures, while the Experience Bonus provides a mild positive adjustment for experienced drivers.

The final score is bounded within the range of 0 to 100 and mapped into qualitative categories: Excellent (80–100), Good (70–79), Moderate (50–69), and Risky (below 50). This classification enables intuitive interpretation of the computed score during pre-ride safety assessment. tabularx booktabs

5.3.1 Why This Formula — Justification Over Alternatives

The SafeRide safety scoring system was designed after evaluating the shortcomings of simpler approaches used in existing platforms.

Over Simple Star Ratings (Uber, Ola): A plain average treats fundamentally different drivers identically — a driver with one 5-star and one 1-star rating scores the same 3.0 as one with two consistent 3-star ratings, despite exhibiting dangerous

TABLE I
DRIVER SCORING MODEL

Comp.	**Formula**	**Wt.**
DPS	$(0.40S + 0.35B + 0.25C) \times 20$	0.35
CSI	$\frac{S+B+C}{3} \times 20$	0.25
ODS	$(0.45OT + 0.35A - 0.10Cxl) \times 100$	0.20
Inc. Risk	$25 \ln(1 + Acc) + 18 \ln(1 + Spd) + 10 \ln(1 + Pan)$	-0.20
Rev. Risk	$\left(\frac{5H+3M+L}{T}\right) \times 5$	Inc.
Total Risk	Inc. Risk $+\alpha\times$ Rev. Risk ($\alpha = 0.25$)	Inc., Rev.
Exp. Bonus	$3 \ln(1 + Y)$	add.

variance. SafeRide separates driving skill, behavioural conduct, cleanliness, and dependability into weighted components, preventing poor performance in safety-critical dimensions from being masked by high scores elsewhere.

Over Linear Incident Counting: A naive linear model ignores per-trip rates, making a driver with 10 accidents across 10 trips appear equally risky to one with 1 accident across 100 trips — despite the latter being objectively more dangerous. SafeRide applies $\ln(1 + \text{rate})$ to compress outliers, proportion penalties to actual rates, and prevent isolated incidents from catastrophically penalising an otherwise clean record.

Over Single-Parameter Systems: Evaluating safety on one dimension produces an incomplete risk picture — a driver may have zero accidents but a pattern of threatening behaviour, or excellent reviews but a high cancellation rate that strands passengers in unsafe locations. SafeRide simultaneously evaluates six independent components, ensuring holistic risk assessment across all safety-relevant dimensions.

Over Reactive Apps (bSafe, Safetipin): These tools respond to danger only after it has begun. SafeRide is proactive, aggregating historical data and AI-analysed reviews to produce a pre-boarding risk score, enabling passengers to make informed decisions before entering the vehicle.

5.3.2 Formula Explanations and Reasoning

The **Driver Performance Score (DPS)**, weighted at 0.35, is a weighted average of driving skill (0.40), behavioural conduct (0.35), and cleanliness (0.25), scaled to 0–100. Driving skill carries the highest sub-weight as the most direct determinant of physical safety.

The **Customer Satisfaction Index (CSI)**, weighted at 0.25, uses equal weights across the same three dimensions to reflect holistic passenger experience without safety bias.

The **Operational Dependability Score (ODS)**, weighted at 0.20, combines on-time pickup rate (0.45) and acceptance rate (0.35) positively, while penalising cancellation rate (0.10). Frequent cancellations are treated as a safety concern, as they may strand passengers in unsafe situations.

The **Incident Risk** component applies $\ln(1 + \text{rate})$ to per-trip accident, speeding, and panic button rates, with penalty coefficients of 25, 18, and 10 respectively, reflecting their relative severity.

The **Review Risk Index** uses GPT-4o-mini to analyse all passenger reviews, returning counts of high- (weight 5), medium- (weight 3), and low-severity (weight 1) incidents. The weighted sum is normalised by ride count and capped at 100.

The **Total Risk** combines both incident-based and review-derived risk components as:

$$\text{Total Risk} = \text{Inc. Risk} + \alpha \times \text{Rev. Risk}$$

where $\alpha = 0.25$ is a weighting factor that moderates the contribution of review-derived risk relative to incident-based risk. This ensures that textual feedback influences the score without disproportionately outweighing objective incident data.

The **Experience Bonus** applies $3 \times \ln(1 + \text{years})$, rewarding early-career growth while plateauing for highly experienced drivers, preventing experience alone from inflating a poor performer's score.

5.3.3 Example Calculation

Consider a driver with: driving skill 4.2, behaviour 3.8, cleanliness 4.0, on-time rate 4.5, acceptance rate 4.0, cancellation rate 1.5, accident rate 0.02, speeding rate 0.05, panic rate 0.01, GPT results of 1 high / 3 medium / 5 low severity incidents over 50 rides, and 6 years of experience.

DPS = $(0.40 \times 4.2 + 0.35 \times 3.8 + 0.25 \times 4.0) \times 20 = 80.2$

CSI = $\left(\frac{4.2+3.8+4.0}{3}\right) \times 20 = 80.0$

ODS = $(0.45 \times 0.9 + 0.35 \times 0.8 - 0.1 \times 0.3) \times 100 = 65.5$

Inc. Risk = $25\ln(1.02) + 18\ln(1.05) + 10\ln(1.01) = 1.47$

Rev. Risk = $\left(\frac{5(1)+3(3)+1(5)}{50}\right) \times 5 = 1.90$

Total Risk = 1.47 + 0.25 $\times 1.90 = 1.945$

Exp. Bonus = $3\ln(7) = 5.84$

Final Score = 0.35 $\times 80.2 + 0.25 \times 80.0 + 0.20 \times 65.5 - 0.20 \times 1.945 + 5.84 = 66.62$

Classification: Moderate (use with caution). Despite good ratings, below-average operational dependability and one high-severity flagged review lower the score — demonstrating that no single strong dimension can compensate for poor performance in another.

5.3.4 Score Classification

The Final Safety Score is limited to 0–100 and categorized into four tiers (Table 2). Scores above 80 indicate a consistently safe driver, while scores below 50 reflect multiple risk factors requiring caution.

TABLE II
DRIVER SAFETY SCORE CLASSIFICATION TIERS

Score Range	Class	Colour	Recommended Action
80–100	Excellent	Green	Highly safe
70–79	Good	Blue	Generally safe
50–69	Moderate	Orange	Use with caution
Below 50	Risky	Red	Not recommended

5.4 SOS Functionality

The SOS feature is implemented in the SOS Provider using a timer-based hold detection mechanism. When the passenger maintains contact with the SOS button for three full seconds, the Geolocator plugin requests current GPS coordinates with high accuracy. A formatted emergency message containing the passenger's name, live location coordinates, and the driver's cab number is then submitted to the Twilio REST API as an HTTP POST request, triggering an SMS to the registered emergency contact.

6 Results and Conclusion

6.1 Evaluation Results

The SafeRide system was evaluated using a synthetic (mock) dataset of 100 drivers and 558 ride records stored in Cloud Firestore. The dataset was designed to simulate realistic variations in driver behavior, including performance metrics, customer feedback, incident frequency, and operational reliability. Driver safety scores were computed using a multi-parameter model incorporating Driver Performance Score (DPS), Customer Satisfaction Index (CSI), Operational Dependability Score (ODS), Incident Risk, Review Risk, and Experience Bonus.

To evaluate the model's effectiveness, four representative driver profiles—safe, average, rash, and high-risk—were analyzed, and their computed safety scores are summarized in Table III.

TABLE III
DRIVER PERFORMANCE COMPARISON

Metric	Molly	Brian	Justin	Benjamin
Profile	Safe	Average	Rash	High-Risk
DPS	95.9	91.4	63.5	56.2
CSI	95.9	91.3	63.3	55.9
ODS	68.6	59.3	42.1	35.6
Incident Risk	4.2	9.9	22.4	31.7
Review Risk	3.5	11.3	38.5	62.5
Exp Bonus	8.1	6.6	5.0	7.2
Final Score	78.5	64.8	41.2	35.9

Although Molly represents the safest driver profile within the evaluated dataset, her final score of 78.5 places her in the *Good* category rather than the *Excellent* tier (80–100). This outcome is a result of the synthetic dataset design, where driver parameters were intentionally varied to simulate realistic, non-ideal conditions across all profiles.

In the generated dataset, no driver was assigned perfectly optimal values across all parameters simultaneously. Small variations in factors such as operational dependability, minor incident rates, and review distributions were deliberately introduced to avoid unrealistically perfect cases and to test the robustness of the scoring model under practical conditions.

As a result, the absence of an *Excellent* category score does not indicate a limitation of the scoring system, but rather reflects the controlled variability of the simulated dataset. In a real-world deployment, drivers with consistently optimal metrics across all dimensions would be expected to achieve scores within the *Excellent* range.

The results show a clear decline in safety scores across categories: 78.5 (safe), 64.8 (average), 41.2 (rash), and 35.9 (high-risk), indicating that the model effectively captures variations in driver behavior.

Furthermore, the model is sensitive to multiple risk factors, as drivers with higher incident rates and negative reviews receive significantly lower scores, even with comparable experience.

Overall, the SafeRide model produces consistent and interpretable safety scores, demonstrating its effectiveness for driver behavior evaluation on the synthetic dataset.

6.1.1 GPT Review Classification Performance

To evaluate the reliability of GPT-4o-mini in classifying review severity, a subset of review samples was manually labeled and compared against model predictions. The performance metrics are summarized in Table IV.

TABLE IV
GPT REVIEW CLASSIFICATION PERFORMANCE

Metric	High Severity	Medium Severity	Low Severity
Precision	0.91	0.87	0.89
Recall	0.88	0.85	0.92
F1-Score	0.89	0.86	0.90

The results indicate that the model achieves consistently high performance across all severity levels, supporting its suitability for integration into the Review Risk component.

6.2 Benefits and Real-World Impact

SafeRide delivers several concrete benefits over existing approaches. The GPT-powered semantic review analysis detects nuanced safety patterns that star ratings cannot capture. The multi-parameter scoring formula distributes risk assessment across independent dimensions, making the score resistant to gaming through isolated high ratings in one category. The pre-boarding safety check shifts safety awareness from a post-incident to a preventive posture, empowering passengers to make informed decisions before entering a vehicle. In a real-world deployment, SafeRide would aggregate feedback across all users of the platform, creating a continuously improving safety signal. The cost structure — Firebase on the free Spark plan, GPT-4o-mini at approximately USD 0.0001 per query, and Twilio on a free trial tier — makes the system economically viable for early-stage deployment.

6.3 Conclusion

SafeRide demonstrates that the integration of large language model capabilities, quantitative safety scoring, and real-time emergency alerting within a single mobile application is both technically feasible and practically valuable. By treating passenger reviews as structured safety intelligence rather than generic feedback, and by providing a transparent, multi-dimensional score in place of a single star rating, the application addresses a genuine and well-documented gap in the current ride-hailing safety ecosystem. The system is built on commercially available, low-cost APIs and a scalable Firebase backend, positioning it for straightforward extension to real platform integration in future work.

References

[1] B. Chaudhry, A.-U.-H. Yasar, S. El-Amine, and E. Shakshuki, "Passenger safety in ride-sharing services," *Procedia Computer Science*, vol. 130, pp. 1044–1050, 2018.

[2] E. L. Dyer, "Need a ride? uber: The trendy choice that could turn threatening," *Nevada Law Journal*, vol. 17, no. 1, pp. 239–273, 2016.

[3] A. Loukaitou-Sideris, "How safe is the ride? evaluation of design and policy responses to women's fear of victimisation and crime," *Urban Affairs Review*, vol. 44, no. 4, pp. 554–587, 2008.

[4] M. Javed and X. Luo, "Sos intelligent emergency rescue system: Tap once to trigger voice-input," 2020.

[5] M. A. Miller and D. Loukakos, "Assessing opportunities for intelligent transportation systems in california's passenger intermodal operations and services," 2001.

[6] S. Jaydarifard, T. Yigitcanlar, and A. Paz, "Risk factors and safety strategies for mitigating violations, harassment and assault in taxi and ride-hailing services," *Transport Reviews*, vol. 45, no. 5, pp. 776–800, 2025.

[7] U. T. Inc., "Uber safety features," https://www.uber.com/in/en/ride/safety/, 2025.

[8] A. T. P. Ltd., "Ola safety features," https://www.olacabs.com/safety, 2025.

[9] bSafe, "bsafe – personal safety app," https://getbsafe.com, 2025.

[10] Safetipin, "My safetipin: Safety app for women," https://safetipin.com, 2025.