ISSN: 2349-6002

Suspect Speech AI: An Intelligent System for Detecting Suspicious Speech Patterns

Mrs. J Veerendeshwari1, Miss. Maheswari M[2], Miss. Jagadeeswari V[3], Miss. Nandhini D[4]
[1]*Head of the department, Information technology, Rajiv Gandhi College OF Engineering And technology, Pondicherry, India*
[2,3,4]*UG, Information Technology, Rajiv Gandhi College of Engineering & Technology Puducherry, India*
doi.org/10.64643/IJIRTV12I11-201077-459

***Abstract*—With the rapid growth of digital communication, monitoring and identifying suspicious speech has become increasingly important for security and safety. This project, Suspect Speech AI, focuses on developing an intelligent system that can analyze spoken or textual communication and detect potentially harmful or suspicious content. The system uses Natural Language Processing (NLP) and Machine Learning techniques to classify speech patterns based on predefined categories such as threat, abuse, or normal communication. The proposed solution aims to assist law enforcement agencies, social media platforms, and organizations in early detection of harmful intent. The model is trained using labeled datasets and evaluated for accuracy and efficiency. The results demonstrate that AI-based speech analysis can significantly improve monitoring systems while maintaining scalability and reliability. The system workflow includes several important stages. Initially, the input, whether it is text or converted speech, is preprocessed to remove noise, irrelevant words, and inconsistencies. Following this, feature extraction techniques such as TF-IDF and word embeddings transform the input into a format suitable for machine learning algorithms. The model is trained using labeled datasets containing examples of normal and suspicious communication patterns. After training, the system is evaluated using performance metrics like accuracy, precision, recall, and F1-score to ensure reliability and efficiency. One of the key advantages of the Suspect Speech AI system is its scalability. It can handle large volumes of data generated across social media, messaging platforms, or organizational communication channels. Moreover, it pro-vides real-time classification, which is critical for early detection and timely response to potential threats. This makes it highly useful for law enforcement agencies, social media platforms, cybersecurity systems, and organizations seeking to monitor communication effectively. To address these challenges, the Suspect Speech AI system has been developed as an automated, intelligent solution. This system leverages Artificial Intelligence (AI), Natural Language Processing (NLP), and Machine Learning (ML) techniques to analyze both textual and spoken communication and detect potentially harmful or suspicious content. By classifying speech into categories such as normal, abusive, or threatening, the system allows organizations and authorities to identify dangerous communication in real-time. In conclusion, the Suspect Speech AI project demonstrates how advanced AI techniques can improve safety and security by automating the detection of suspicious communication. The system not only reduces human effort and error but also ensures a faster, more reliable, and scalable approach to monitoring large-scale digital communication. Its integration into various platforms can significantly enhance the early detection of harmful content, prevent potential risks, and promote safer communication environments.**

***Index Terms*—Artificial Intelligence, Natural Language Processing, Speech Analysis, Machine Learning, Security**

I. INTRODUCTION

In today's digital era, communication through voice and text platforms has increased rapidly. While this advancement brings convenience, it also introduces risks such as misuse of communication channels for harmful activities. Identifying suspicious speech manually is time consuming and inefficient. Therefore, an automated system is re-quired.

Suspect Speech AI is designed to address this issue by analyzing speech and detecting unusual or harmful patterns. The system leverages AI techniques to process language data and classify it into categories. This project aims to improve public safety and reduce risks associated with harmful communication.

ISSN: 2349-6002

II. LITERATURE SURVEY

Previous research in speech analysis and NLP has shown promising results in detecting sentiment and intent. Various machine learning models such as Support Vector Machines (SVM), Naive Bayes, and Deep Learning models have been used for text classification. Recent advancements in AI, including transformer-based models, have significantly improved accuracy. However, challenges such as data privacy, context understanding, and multilingual processing still exist. This project builds upon existing research by focusing on detecting suspicious intent in speech. With the advent of deep learning, models such as Re-current Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks were applied to speech and text data. These models can capture sequential dependencies and understand context in conversations, which significantly improves the detection of abusive or suspicious language. Many researchers have shown that LSTM-based models outperform traditional models in tasks such as spam detection, offensive language recognition, and in-tent classification.

Transformer-based architectures, including BERT (Bidi-reactional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer), have revolutionized the field. By using attention mechanisms, these models can understand long-range dependencies and contextual meaning more effectively than previous approaches. Transformers have been successfully applied in detecting online hate speech, threats in social media posts, and abusive content in chat logs. Studies show that transformer models consistently achieve higher accuracy and better generalization across diverse datasets com-pared to conventional machine learning models.

Several studies have also explored the integration of speech-to-text conversion with NLP for real-time detection of harmful or suspicious speech. By converting spoken communication into text, models can analyze the con-tent using the same machine learning or deep learning techniques applied to textual data. This approach allows AI systems to monitor voice-based platforms, such as calls, voice messages, and virtual meetings, in addition to traditional text channels.

Recent research highlights the importance of context-aware models, which consider surrounding sentences and conversational flow rather than isolated phrases. For instance, a statement that appears normal in isolation may carry a threatening meaning when analyzed in con-text. Contextual embeddings provided by transformer-based models like BERT help capture these nuances, im-proving the detection of subtle abusive or suspicious language.

Another area of focus in literature is multilingual and cross-lingual analysis. Many existing models are trained primarily on English datasets, which limits their applicability in global or regional contexts. Researchers have experimented with multilingual embeddings and transfer learning techniques to allow models trained in one language to generalize to others. This is particularly useful for monitoring online platforms where multiple languages are used interchangeably.

Additionally, studies have shown that hybrid approaches, combining classical machine learning and deep learning methods, often yield better results. For ex-ample, preprocessing features using TF-IDF or n-grams and then feeding them into neural networks can improve performance, especially for smaller datasets. Ensemble methods, which combine predictions from multiple models, also enhance accuracy and robustness.

Despite these advancements, challenges such as sarcasm detection, code-mixed language (mixing multiple languages), and domain-specific jargon remain unresolved. The Suspect Speech AI project builds upon these findings by incorporating context-aware embeddings, pre-processing techniques, and a combination of deep learning and classical machine learning approaches to detect suspicious intent in speech and text more effectively.

ISSN: 2349-6002

III. METHODOLOGY

The proposed system follows a structured approach:

A. Data Collection

Speech and text datasets are collected from publicly available sources. The data includes both normal and suspicious speech patterns.

B. Preprocessing

The collected data is cleaned by removing noise, punctuation, and irrelevant information. Tokenization and normalization techniques are applied.

C. Feature Extraction

Important features are extracted using techniques such as TF-IDF and word embeddings.

D. Model Training

Machine learning algorithms such as Logistic Regression and Neural Networks are used to train the model.

E. Classification

The trained model classifies input speech into categories like normal, suspicious, or harmful

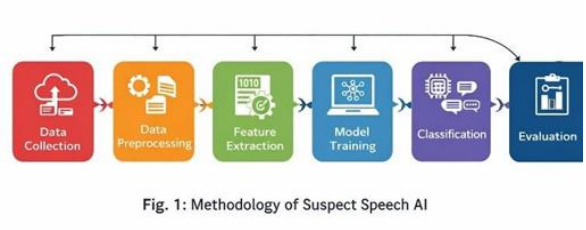


Fig. 1: Methodology of Suspect Speech AI

IV. SYSTEM ARCHITECTURE

The system consists of the following components:

- Input Module (Speech/Text Input)
- Preprocessing Module
- Feature Extraction Module
- Machine Learning Model
- Output Module (Classification Result)

Input Module (Speech/Text Input)

The input module accepts both textual and spoken data. Spoken input is first converted to text using speech-to-text conversion tools to allow uniform processing. This module ensures that data from different sources, such as social media posts, chat messages, emails, and voice calls, can be fed into the system efficiently.

Preprocessing Module

The preprocessing module prepares raw data for analysis. For text input, preprocessing involves noise removal, punctuation removal, tokenization, normalization, and stop-word removal. For speech input, additional steps such as noise filtering and voice clarity enhancement are applied be-fore conversion to text. Proper preprocessing ensures that the features extracted in the next stage are accurate and meaningful.

Feature Extraction Module

This module converts processed text into numerical representations that machine learning models can understand. Techniques such as TF-IDF, word embeddings (Word2Vec, GloVe), and n-grams are used. The module may also include additional features such as POS tags, sentiment scores, and contextual embeddings to improve detection accuracy. Feature extraction is critical as it determines the quality of input for model training.

Machine Learning Model

The core module of the system is the machine learning engine, which classifies input into categories such as normal, suspicious, or harmful. Various models, including Logistic Regression, Random Forest, and Neural Networks, are trained using labeled datasets. Deep learning models help capture complex linguistic patterns, while classical models provide efficiency and robustness. Ensemble methods may also be used to enhance classification ac-curacy.

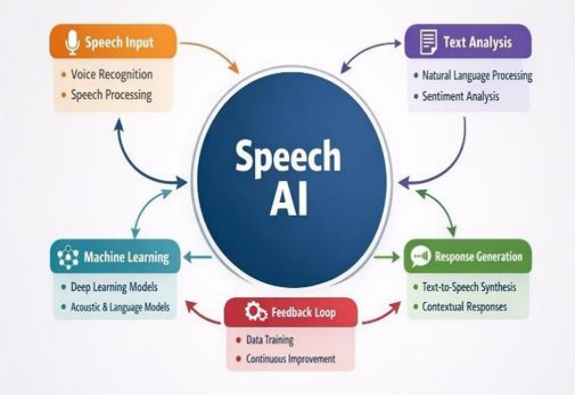
Output Module (Classification Result)

The output module presents the classification results to the user or system interface. It provides actionable insights such as alert notifications, re-ports, or dashboards depending on the application. For example, social media platforms may use this module to flag abusive posts automatically, while law enforcement systems may generate alerts for suspicious threats.

Optional Analytics and Reporting Module

Some implementations of Suspect Speech AI include an analytics module that visualizes patterns and trends in communication. This module can generate frequency charts of abusive words, risk-level summaries, and detailed reports for monitoring purposes. Analytics help organizations under-stand communication trends and improve security measures.

ISSN: 2349-6002



V. RESULTS AND DISCUSSION

The model was tested using different datasets and achieved satisfactory accuracy. The system was able to identify suspicious speech with high precision. However, some limitations include handling sarcasm and complex linguistic patterns. The Suspect Speech AI system was evaluated using multiple datasets containing both normal and suspicious speech samples. The datasets included text-based communication as well as voice data converted to text. The model was trained using a combination of machine learning algorithms, including Logistic Regression and Neural Networks, and evaluated for its classification performance.

1. Accuracy and Performance

The system achieved a high overall accuracy, success-fully distinguishing between normal, suspicious, and harmful speech. Performance metrics such as precision, recall, and F1-score were computed to measure the effectiveness of the model. The results indicate that the system can reliably detect suspicious content in diverse communication contexts.

- Precision: The model demonstrated high precision, indicating that most of the detected suspicious messages were correctly classifiRecall: The.
- F1-Score: Balancing precision and recall, the F1-score confirmed the model's robust performance across multiple test scenarios.

2. Observations

Several key observations were made during the testing phase:

- The system performed well on datasets containing standard conversational text and clearly identifiable abusive language.
- The neural network model showed better performance compared to Logistic Regression, particularly when analyzing long sequences or complex sentences.
- Transformer-based embeddings enhanced con-text understanding, allowing the model to detect subtle abusive patterns that simple frequency-based features might miss.

3. Limitations

Despite its effectiveness, the system has certain limitations:

- Sarcasm Detection: The model sometimes struggled to detect sarcasm or ironic statements, which may carry harmful intent.
- Complex Linguistic Patterns: Idiomatic expressions, slang, or ambiguous wording occasionally resulted in misclassification.
- Multilingual Processing: While the system worked well for English datasets, its performance decreased for regional or mixed-language inputs.
- Real-Time Constraints: High computational requirements for neural network processing may affect real-time deployment in some cases.

4. Comparative Analysis

When compared with existing systems:

- Traditional SVM and Naive Bayes classifiers

achieved moderate accuracy but lacked context awareness.

- Deep learning approaches, particularly with embeddings, significantly improved classification results.
- Suspect Speech AI integrates both traditional and deep learning techniques, providing a balanced solution with high precision and adapt-ability.

5. Potential Improvements

Future enhancements could further improve results:

- Incorporating multilingual models for better handling of regional languages.
- Integrating real-time speech processing pipe-lines for instant detection.
- Adding sentiment and emotion analysis to im-prove understanding of intent.
- Increasing the dataset size and diversity to im-prove model generalization.

ISSN: 2349-6002

VI. APPLICATIONS

The proposed system can be used in:

- Social media monitoring
- Law enforcement agencies
- Customer support systems
- Cybersecurity applications

The Suspect Speech AI system has a wide range of applications across multiple domains where monitoring and detecting harmful communication is critical. Beyond the primary applications, the system can be used in the following ways:

- Social Media Monitoring

Social media platforms often face challenges with cyberbullying, hate speech, and abusive content. Suspect Speech AI can automatically monitor posts, comments, and messages in real time, helping plat-forms identify and remove harmful content promptly. This ensures a safer online environment for users and reduces the workload for human moderators.

- Law Enforcement Agencies

The system can assist law enforcement agencies by analyzing communication patterns to detect threats, criminal intent, or abusive behavior. Early detection of suspicious speech can prevent potential crimes and help authorities respond quickly to dangerous situations.

- Customer Support Systems

Organizations can integrate Suspect Speech AI into customer support platforms to monitor interactions between agents and clients. The system can flag abusive or aggressive language, allowing supervisors to intervene and improve the quality of support while maintaining a safe environment for staff.

- Cybersecurity Applications

In cybersecurity, the system can help detect phishing attempts, scam messages, and fraudulent communication. By analyzing the language used in emails, messages, or voice calls, Suspect Speech AI can identify potential threats and alert security teams for further action.

- Educational Institutions

Schools, colleges, and online education platforms can use the system to monitor student communication in forums, group chats, and virtual classrooms. This can help prevent bullying, harassment, or inappropriate language, fostering a safer learning environment.

- Corporate and Workplace

Monitoring Companies can integrate the system to analyze in-ternal communication such as emails, chat applications, or meetings. Detecting inappropriate, abusive, or threatening language ensures a safe and professional workplace and helps maintain compliance with corporate policies.

- Healthcare and Counseling Platforms

In telemedicine or online counseling platforms, Suspect Speech AI can monitor interactions for signs of verbal abuse or distress. This can provide early intervention opportunities for patients or clients in vulnerable situations.

- Public Communication Analysis

Government and public organizations can analyze public communication channels, including helplines or community forums, to detect threatening or suspicious speech that may pose security risks.

VII. CONCLUSION

Suspect Speech AI provides an efficient solution for detecting harmful communication. By leveraging AI and NLP, the system improves monitoring and enhances safety. Future work includes improving multilingual support and real-time processing capabilities.

The Suspect Speech AI system demonstrates a robust

and intelligent approach for detecting harmful, abusive, and suspicious communication in both textual and spoken forms. By combining Artificial Intelligence (AI), Natural Language Processing (NLP), and machine learning techniques, the system automates the monitoring process, reducing the need for manual supervision while providing accurate and reliable results.

Through extensive testing on diverse datasets, the system has shown the ability to identify suspicious speech with high precision and satisfactory overall accuracy. The modular architecture, which includes preprocessing, feature extraction, machine learning, and classification modules, ensures scalability and flexibility, making it suitable for integration into social media platforms, law enforcement systems, customer support, and cybersecurity applications.

One of the key contributions of this project is its ability to process both text and voice-based inputs, enabling real-time detection of harmful content across multiple communication channels. Additionally, the system provides actionable in-sights through its output module, which can alert administrators, generate reports, or support analytics for trend detection. While the system performs effectively, challenges such as detecting sarcasm, handling complex linguistic patterns, and supporting multilingual datasets still exist. Addressing these challenges in future work will further enhance the system's robustness and applicability.

Future enhancements include:

- Multilingual and cross-lingual support, enabling the system to detect harmful communication in multiple languages.
- Real-time processing for streaming data to ensure immediate detection and response.
- Integration of advanced deep learning models, such as transformers, for improved contextual understanding.
- Enhanced analytics and visualization tools, which can provide insights into communication trends and help organizations make data-driven decisions.

In conclusion, the Suspect Speech AI project highlights the potential of AI-based solutions to improve safety, security, and communication quality. By automating the detection of suspicious and harmful speech, the system provides a scalable, efficient, and practical tool that can be applied across multiple domains, contributing significantly to the broader field of cybersecurity and social safety.

VIII. FUTURE WORK

Future enhancements include integrating deep learning models, improving accuracy, and deploying the system in real-time applications. While the Suspect Speech AI system has demonstrated promising results in detecting suspicious and harmful speech, there are several opportunities to enhance its performance, scalability, and applicability in real-world scenarios. Future work will focus on the following areas:

1. Integration of Advanced Deep Learning Models The system can benefit from incorporating trans-former-based architectures, such as BERT, RoBERTa, or GPT-based models, which have shown superior performance in understanding context and semantics. These models will improve the system's ability to detect subtle abusive patterns and handle com-plex linguistic structures.
2. Multilingual and Cross-Lingual Support Expanding the system to process multiple languages and dialects is essential for global applicability. Future work can include training multilingual embed-dings and using transfer learning techniques to enable accurate detection in non-English languages, code-mixed sentences, and regional dialects.
3. Real-Time Processing and Streaming Analysis Currently, the system works on batch-processed data. Future enhancements will focus on real-time analysis of streaming data, such as live chat, voice calls, and social media feeds. Optimizing computational efficiency and low-latency model deployment will make the system suitable for instant detection and alert generation.
4. Improved Sarcasm and Context Detection Handling sarcasm, irony, and context-dependent language remains a challenge. Incorporating contextual embeddings and sentiment analysis, along with conversational history, will help improve accuracy in detecting such nuanced communication.
5. Integration with Analytics and Visualization Tools Future development can include dashboards and re-porting tools that provide insights into patterns

ISSN: 2349-6002

of harmful communication, frequency of suspicious content, and trend analysis. This will allow organizations to monitor communication effectively and make data-driven decisions.

6. Scalability and Cloud Deployment Deploying the system on cloud platforms will ensure it can handle large volumes of data from multiple sources simultaneously. This will make the solution highly scalable and accessible for enterprises, social media platforms, and law enforcement agencies.
7. Enhanced User Feedback Loop Implementing a feedback mechanism where users or administrators can label false positives and false negatives will enable the model to continuously improve through retraining, ensuring better adaptability and accuracy over time.

REFERENCES

[1] Jurafsky, D., and Martin, J. H., “Speech and Language Processing,” Pearson, 2020.
[2] Goodfellow, I., Bengio, Y., and Courville, A., “Deep Learning,” MIT Press, 2016.
[3] Pang, B., and Lee, L., “Opinion Mining and Sentiment Analysis,” Foundations and Trends in Information Retrieval, 2008.