December 2023

Suspicious Call Detection and Mitigation Using Conversational AI

Kolati Mallikarjuna Rao

Bhavikkumar Patel

Suspicious Call Detection and Mitigation Using Conversational AI

ABSTRACT

Spam or scam calls and messages are an annoyance, pose security risks, and can harm users that fall prey to calls that request money transfer or other action. Detection of such calls is difficult as scammers evade detection through tactics such as changing numbers they call from, modifying the call script, etc. While some smartphone applications can detect such calls or text based on caller ID, call origination, etc., such techniques cannot easily adapt to new scams. This disclosure describes the use of a conversational AI model to detect suspicious calls. The conversational AI model is trained on a dataset of spam/scam calls and other calls to detect spam/scam calls. With user permission, when the user receives a call from an unknown number, call content is automatically transcribed and analyzed in real time to determine if the call is likely suspicious. When such a call is detected, alerts are provided to the user to ensure that the user does not share sensitive information. If the user permits, a conversational AI agent based on the conversational AI model can answer the call and conduct a conversation with such a caller without user intervention. The use of conversational AI to detect spam can reduce the number of spam calls. The conversational AI agent can be trained to adapt to new strategies employed by spam callers.

KEYWORDS

- Suspicious call
- Spam call
- Scam call
- Conversational AI
- Auto answering
- Scam detection
- Robocall
- Call screening

BACKGROUND

Suspicious or unsolicited calls or text messages that are fraudulent or deceptive in nature are used by malicious actors and are a problem. These are referred to as spam or scam calls. These pose security risks to the users.

Spam or scam calls are attempts to trick users into providing personal information such as government identifiers (e.g., social security numbers), passwords, financial information (e.g., bank details), or other sensitive data. For example, callers may pose as government officials or service providers (e.g. technical support providers). Callers may employ a threat (e.g., of government action) or an offer to gain the user's trust. In some cases, the user may be asked to pay to a scammer's account.

Detecting spam or scam calls or text messages requires the call recipient to probe the call with questions and understand the conversation flow in real time. Some smartphone applications can detect such calls or text based on one or more of a combination of caller ID, list of numbers previously identified as suspicious, machine learning algorithms that detect such calls, etc. and can alert the user. However, as scammers evolve new tactics to evade detection, the effectiveness of such techniques can over a period of time. For example, scammers may change their numbers, change the script used for the call to evade detection, etc.

DESCRIPTION

This disclosure describes the use of a conversational AI model to detect suspicious calls. The conversational AI model is trained on a dataset of spam/scam calls and other calls to detect spam/scam calls. With user permission, when the user receives a call from an unknown number, call content is automatically transcribed and analyzed in real time to determine if the call is likely suspicious. When such a call is detected, alerts are provided to the user to ensure that the

user does not share sensitive information. If the user permits, a conversational AI agent based on the conversational AI model can answer the call and conduct a conversation with such a caller without user intervention. The use of conversational AI to detect spam can reduce the number of spam calls. The conversational AI agent can be trained to adapt to new strategies employed by spam callers.

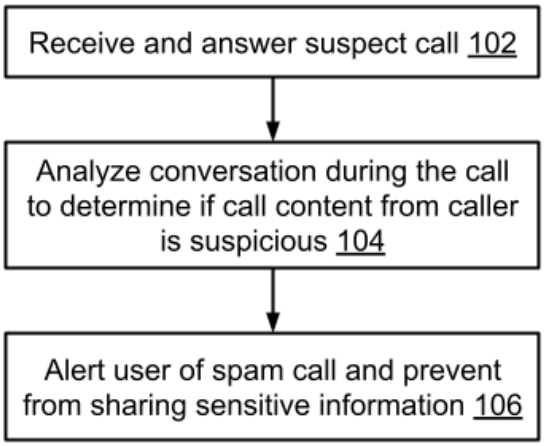

Fig. 1: Suspicious call detection using conversational AI

Fig. 1 shows detection of a suspicious call detection using a conversational AI model. The conversational AI model is trained on a large dataset of spam/scam calls and non-spam calls, e.g., to classify new calls as likely spam or not. When a call is received on a user device from an unknown and possibly suspicious number, the call can be answered (102) by the user or automatically by a conversational AI agent based on user preference settings.

With user permission, the conversation with the caller is analyzed (104) in real time to determine whether the call is suspicious or not. The analysis may include converting audio to text through automatic speech recognition. The converted text may be analyzed by the conversational model to determine if the call content matches spam. The content can help identify that the call matches patterns associated with spam calls.

For example, the call content may match names of government agencies, technical support businesses, or financial institutions (scammers pretending to be agents of such organizations); fake badge numbers or other identifiers (scammers pretending to be someone); , types of threats such as arrest, deportation or fines, offers made such as offer to provide freebies or other benefits, language used, etc. (type of call content that only arises in spam calls); information bank account, credit card, government identifiers, one-time passwords, etc. being requested from the user (type of requests typical of spam calls, but uncommon otherwise); etc. Requests to provide remote access to user devices, requests to install software remotely, etc. may also be detected and users may be alerted to refuse access. If the call is determined to be a suspicious call, the user is alerted (106) about the call being suspicious and asked not to provide any sensitive information to the caller. Additionally, an option may be provided to the user to report the call as suspicious, alert government agencies, etc.

The conversational AI model may also identify suspicious text messages from known or unknown numbers, e.g., messages that request payment through cryptocurrency, gift cards, or transfer to unknown accounts, even messages that appear to be from known contacts. Other examples of text messages that are scam include messages appearing to come from a business (e.g., bank or other financial institution) or government authority (e.g., tax department) but being received from an unofficial number. For example, a warning and suggestion to connect with the person directly can be provided to the user before the user takes action on a text message received.

The described techniques offer several benefits. Trained conversational models that directly handle spam calls and messages can reduce or eliminate user burden of dealing with such calls and messages. Model updates can ensure that detection can continue to be reliable

even when scammers change their strategies. The techniques can be deployed easily on any device such as smartphones, tablets, computers, etc. The techniques can be integrated easily with existing call management systems and are scalable.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user's received calls or messages, a user's contacts, a user's preferences, or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

This disclosure describes the use of a conversational AI model to detect suspicious calls. The conversational AI model is trained on a dataset of spam/scam calls and other calls to detect spam/scam calls. With user permission, when the user receives a call from an unknown number, call content is automatically transcribed and analyzed in real time to determine if the call is likely suspicious. When such a call is detected, alerts are provided to the user to ensure that the user does not share sensitive information. If the user permits, a conversational AI agent based on the conversational AI model can answer the call and conduct a conversation with such a

caller without user intervention. The use of conversational AI to detect spam can reduce the number of spam calls. The conversational AI agent can be trained to adapt to new strategies employed by spam callers.

REFERENCES

1. "Introducing Truecaller Assistant - Powerful AI Call Screening for iOS and Android," available online at https://www.truecaller.com/blog/features/introducing-truecaller-assistant-powerful-ai-call-screening-for-ios-and-android accessed Aug 28, 2023.
2. "How Do Phones Identify Potential Spam Calls?" available online at https://builtin.com/machine-learning/spam-calls, accessed Aug 28, 2023.
3. "Stopping spam calls directly in the network - Ericsson" available online at https://www.ericsson.com/en/blog/2022/9/an-end-to-spam-calls-how-to-stop-a-growing-multi-billion-dollar-industry, accessed Aug 28, 2023.
4. "Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone," available online at https://blog.research.google/2018/05/duplex-ai-system-for-natural-conversation.html?m=1, accessed Sep 5, 2023.