**Research Proposal**

**Working Title**

**AI-Powered Detection of Vishing Attacks: Voice and Language Analysis for Real-Time Scam Prevention**

**Research Area and Context**

The recent large-scale increase in phishing attacks poses a significant threat to elderly populations in Western countries such as the United Kingdom, New Zealand, and the United States. Due to their lack of knowledge about modern technological developments, individuals within the elderly population are often targeted by scammers who utilise sophisticated social engineering techniques. Telephonic scam calls (known as voice phishing, or vishing) are one of the most frequently occurring types of phishing attack. Typically, the attack starts with a scammer cold calling a set of random phone numbers, pretending to represent a governmental body or a reputed organisation and tricking the victims into transferring huge amounts of money from their savings to crypto accounts or foreign bank accounts. Other common forms of phishing attack are pop-up phishing, social media phishing and email phishing. Despite several fraud detection methods being available, scammers are still able to successfully manipulate their victims. Some of today’s prevalent vishing attacks include scams relating to Authority impersonation, Family-in-distress pretext, AI-driven robocalls / automated agents, Urgency/time pressure and Social-proof & distraction tactics(Triantafyllopoulos *et al.*, 2025).

These scams usually leave victims facing significant financial losses and damage to their quality of life. With the increasing number of scams, large amounts of money are being taken away from these countries’ economies. Particularly, in the United Kingdom, the UK economy is estimated to be losing approximately £27 billion annually because of these cyber scams (Ali *et al.*, 2019). UK businesses are hit the most at a cost of approximately £21 billion, followed by the government and individuals, with losses of approximately £3 billion each(Ali *et al.*, 2019). In New Zeland, As of Q1 2023, NZD 5.8 million of direct financial loss to cyber incidents was reported by CERT NZ—a 66% rise compared to the previous quarter(Pope, 2023).In Q1 2024, the losses rose further to NZD 6.6 million(Pope, 2023).

Despite modern technological advancements, current measures for scam prevention are ineffective, and this area is largely overlooked (Sun *et al.*, 2024). The current most effective approach to tackle phishing is manual reporting of incidents. This alerts security teams, and phishing websites can be taken down quickly. However, the number of people reporting fraud is, unfortunately, minimal.For instance, in their study on company-targetd phishing email attacks, Sun et al. (2024) conducted an experiment and found that only 7% of employees reported phishing attempts.

Several studies have been conducted in the past decade about the human factors affecting scam vulnerability and how different categories of people react to scams(Gallo *et al.*, 2024). Many studies suggest that psychological factors relating to victims are used to exploit their cognitive vulnerabilities(Gallo *et al.*, 2024). Authority (the tendency to obey people who hold governmental or higher positions), Liking (the tendency to approve of people who are acquiescent), Scarcity (assigning greater value to opportunities that are limited), Consistency (behaving in ways that are consistent with past decisions), Social proof (blindly following what others say), Reciprocity (the desire to repay others when favours are received) are a few such vulnerabilities (Gallo *et al.*, 2024).

In this research, I would like to conduct an in-depth study into telephone scams and develop an AI-based Android application that can identify and stop possible vishing attacks by monitoring all calls to a mobile phone, disconnecting the call and notifying the user about the attack. For further investigation, Natural Language Processing (NLP) to convert speech to text and Support Vector Machines(SVM) to extract emerging tactics.

Currently, there are a few AI-based mobile apps available.

**TrueCaller:** This has alarge scam and telemarketing phone number dataset, and it works by blocking calls from these numbers. It supports call recording and transcription, but it does not support call disconnection during vishing (TrueCaller, 2025).

**Hiya:** This detects scam calls in real time based on an AI model and flags them. It does not offer live disconnection of ongoing calls (The Telephony Co., 2025).

**RealCall:** This offers a bot-answering feature, which answers the call first before passing it on to the user and engaging or hanging up on suspected spam calls. However, this approach could block real users or result in unnecessary delays for calls that are urgent (Wikipedia, 2025).

**CallApp:** This offers caller ID and community-based call blocking similar to TrueCaller. It blocks calls from malicious numbers, but it cannot disconnect mid-call (Wikipedia, 2025).

**Aim**

To investigate how AI-based voice and language analysis can support the real-time detection of telephone scams (vishing), and evaluate the effectiveness of android app that can disconnect and notify targeted user.

**Objectives**

1. Identify and tabulate the latest trends in vishing attacks and look for repeated patterns, locations etc.
2. Build a machine learning (ML) model and train it on public scam datasets.
3. Develop a mechanism powered by AI that can monitor and disconnect incoming calls from ML-identified scammers.
4. Run this scheme through simulated calls and repeat until the desired accuracy is achieved.
5. Establish a notification framework that warns the targeted user about the situation in real time.

**Research Questions**

1. What linguistic and acoustic features distinguish vishing calls from legitimate calls?
2. What are the trade-offs between false positives (blocking genuine calls) and false negatives (missed scams)?
3. How accurately can supervised ML/NLP models detect scam calls using publicly available and simulated datasets?
4. How to reduce vishing with the help of Artificial Intelligence and Machine Learning?
5. What design features (e.g., notifications, explanations, user control) make AI scam-detection tools acceptable and usable?
6. To what extent an AI-powered app can get to identify and disconnect fraud calls real time and notify the target user?

**Methodological Framework**

This research will address the importance of quantitative methodologies and adopt an experimental approach to tackle the issue.

**Data Collection**

**Prototype system:** An AI-powered mechanism that can identify and block scammers during vishing and notify the target about the attack.

**Datatype system:** There are several publicly available scam call conversation logs on the internet. For example, the Fraud Call Detection Dataset (Kaggle, 2021).

**Testing procedure:** The ML algorithm will need to be trained on a scam dataset and then normal conversation datasets. The model will then be ready to run in simulated live conditions.

**Data Analysis**

**Tools:** For ML processing, the study will use Python’s PyAudio, Pandas, scikit-learn and Flask libraries.

For Android, Android Studio, Android SDK and Gradle Plugin for ML binding can be used.

**Preprocessing:** Different graphs will be plotted using a training dataset (e.g. call frequency, origin patterns, scam pattern types).

**Modelling:** The study will train the model and compare its accuracy with well-known models such as random forest, neural networks, support vector machines etc. and find the most accurate model.

**Evaluation and Validation**

Success will be determined by the AI prototype’s ability to accurately detect vishing calls from a wide range of test calls and end those calls. The performance of the model will be calculated by conducting experiments in which vishing calls are successfully detected and stopped: 85% or more will be considered Excellent; accuracy of 60% or above will be considered Good – in the case of a Good result, more training data will need to be supplied to improve accuracy and the process will need to be repeated. A detection rate of 40% or above will be considered Average. In this case, the model will need to be replaced, a new model will need to be trained, and the full process will need to be repeated. A detection rate below 40% will be considered a Fail. Once the model successfully achieves accuracy of 85% or above, this will be considered a Pass.

**Ethics and Legalities**

* **Data anonymization:** Remove all the sensitive data such as personal identifiers or bank card information from call transcripts before storing.
* **Compliance:** The App only uses open-source software for the app development and use only publicly available datasets to train ML models. Therefore, no ethics approval is required.
* **User control:** The user must be able to manually review and discard any vishing call transcripts to preserve end-user autonomy.

**Rationale and Original Contribution**

I have chosen this research topic because vishing is an issue that affects a wider population; almost every email or mobile user has experienced vishing in their day-to-day life, and every single scamming attempt has a huge impact on our society both financially and mentally. In recent years, this issue has proliferated and been overlooked by the relevant authorities. Even though AI-based mobile apps require a lot of careful planning and coding, it require only a one-time setup with little maintenance.

Finally, this topic is well suited to my assignment because I have previous coding experience in Android development and have access to public datasets to support my research. I have estimated that I will have sufficient time to cover data collection, coding and setup, and to follow my supervisor’s directions etc.

This research will deliver a truly original contribution by manifesting an Android app that can monitor calls in real time and disconnect scammers during calls, store the NLP transcript in cloud for spam classification. Currently, there are mobile apps on the market that can detect scam calls; however, they focus mainly on detecting deep-fake audio or vishing only in post-call analysis. Most scammers change their phone numbers at regular intervals to evade the scam databases used by these apps. Without live monitoring and a disconnection mechanism, it will be difficult to stop vishing at the right time, as most scammers build a sense of urgency and post-call notifications could be overlooked.

**Contribution to Knowledge**

Most of the applications in the Android play store rely on a database where all the spam call numbers are stored. This blocks any fraudulent calls from a known number. This study focuses on real time call blocking from calls including unknown or new numbers. Some other application detects the vishing post-call and sends notification to the user. However, this android-based application capable of disconnecting the scammer during the call and notify user immediately.

By systematically processing and recording the linguistic and acoustic information from scam calls could even lead to explore detailed study of these aspects of vishing which are not explained or explored in this literature.

**References**

Ali, M.A. *et al.* (2019) ‘Consumer-facing technology fraud: Economics, attack methods and potential solutions’, *Future Generation Computer Systems*, 100, pp. 408–427. Available at: https://doi.org/10.1016/J.FUTURE.2019.03.041.

*CallApp - Wikipedia* (2025). Available at: https://en.wikipedia.org/wiki/CallApp (Accessed: 3 August 2025).

*Fraud Call Detection Dataset* (2021). Available at: https://www.kaggle.com/datasets/narayanyadav/fraud-call-india-dataset?select=fraud\_call.file (Accessed: 2 August 2025).

Gallo, L. *et al.* (2024) ‘The human factor in phishing: Collecting and analyzing user behavior when reading emails’, *Computers and Security*, 139. Available at: https://doi.org/10.1016/j.cose.2023.103671.

Pope, R. (no date) *CYBER SECURITY INSIGHTS Q1*.

*RealCall - Wikipedia* (2025). Available at: https://en.wikipedia.org/wiki/RealCall (Accessed: 3 August 2025).

Sun, Z. *et al.* (2024) ‘From Victims to Defenders: An Exploration of the Phishing Attack Reporting Ecosystem’, in *ACM International Conference Proceeding Series*. Association for Computing Machinery, pp. 49–64. Available at: https://doi.org/10.1145/3678890.3678926.

*Top 11 Best Call Screening Apps to Block Spam Calls (2025)* (2025). Available at: https://www.thetelephony.co/stay-alert-stay-safe/best-call-screening-apps (Accessed: 3 August 2025).

Triantafyllopoulos, A. *et al.* (2025) ‘Vishing: Detecting social engineering in spoken communication — A first survey & urgent roadmap to address an emerging societal challenge’, *Computer Speech and Language*, 94. Available at: https://doi.org/10.1016/j.csl.2025.101802.

*Truecaller - Leading Global Caller ID & Call Blocking App* (2025). Available at: https://www.truecaller.com/ (Accessed: 3 August 2025).