# Chapter 2: LITERATURE REVIEW



## Introduction

Social engineering attacks have emerged as a significant cybersecurity threat, leveraging psychological manipulation to deceive individuals into divulging confidential information. Traditional security mechanisms, such as firewalls and antivirus software, often fail to detect these attacks due to their reliance on human error rather than system vulnerabilities. Artificial Intelligence (AI) has shown promising capabilities in detecting and mitigating social engineering attacks by analyzing patterns, behaviours, and linguistic cues. This chapter explores existing literature on social engineering attacks, AI-based detection techniques, classification techniques, existing AI GUI tools, machine learning software, and the challenges in implementing effective solutions.



Figure 1: Literature Review Structure Diagram

## Artificial Intelligence Social Engineering Detection Techniques

Artificial Intelligence (AI) has been employed to tackle various types of social engineering attacks, each of which exploits human psychology to deceive individuals or systems. The following subsections describe how AI technologies are applied to detect specific attack types, focusing on the detection methods, models used, and their respective strengths.

### Phishing

Phishing AI detects phishing through Natural Language Processing (NLP), behavioral analytics, and content inspection. Machine learning models such as Support Vector Machines (SVM), Random Forest, and deep learning architectures like LSTM and BERT are used to analyze message content, URLs, sender authenticity, and context to flag phishing attempts.

#### Support Vector Machines (SVM)

Support Vector Machine (SVM) model as a hyper plane in which several classes are represented. So that we may reduce the amount of mistake we make, SVM will create the hyper plane in an iterative fashion. SVM's purpose is to classify datasets such that a maximum marginal hyper plane may be discovered (Jain & Gupta, 2023).



Figure 2:

Based on diagram above, clients either connect directly to the server or through a proxy that is linked to servers through HTTP for message transmission (Anusha, et al, 2017).