# Detecting Spoofed Emails with Explainable AI

**Abstract**
The uptick in email internet-based communication has raised concerns about message origination security and authenticity of the message origins. Email sender and message spoofing can be both legitimate and illegitimate. While advertising agencies legitimately send on behalf of organizations, attackers can impersonate legitimate senders with malicious intent to gain trust and carry out malicious activities.
Detecting the difference between legitimate and spoofed emails is a challenging task even experienced users can struggle with. Machine learning models have shown higher accuracy in differentiating legitimate email content from spoofed emails. Researchers have also explored natural language processing and machine learning to scrutinize email content, headers, domains, and other features to detect phishing attempts. These approaches, driven by AI, have the potential to perform well than the traditional heuristic and blacklisting methods. However, considering the evolving nature of phishing techniques and the ability to forge email content, that can pose significant security challenges.
One way for implementing explainable AI is through a web extension or plugin, this will allow the content to be scanned through different email platforms and easily accessible. As AI-enabled phishing evolves, so must AI-enabled phishing detection. It is essential to develop solutions that accurately identify spoofed emails and explain their decisions. This explainability can further enhance user trust, facilitate the investigation of suspicious emails, and contribute to the continuous improvement of detection systems. And by understanding the reasoning behind AI decisions or predictions, security professionals and also the end-users can help in getting a deeper insight into the detection process, leading to  more effective mitigation strategies and a more secure emailing system.