

### Project Overview

People often receive **suspicious emails** that claim to be from trusted persons or organisations (e.g., a bank), but are actually from attackers aiming to deceive recipients into giving away valuable information, known as **phishing**.

In this project, we are developing the **PhishED** system that will support people by providing **automatically generated advice** to suspicious emails they report. Using contextual cues the advice is meant to both help make an **informed decision** and **provide education** in a teachable moment.

### Proposed Solution

PhishEd leverages Artificial Intelligence's ability to extract and reason about contextual features of phishing in support of user decision making. Example of such features are:

**Contextual keywords** - If a reported email uses terms like "shutdown", "email", and "account", the user may think that such a message was sent through an organization.

**Uniform Resource Locators (URLs)** - If an email contains the "PayPal" keyword but the URL does not lead to PayPal's official website, that information can be provided to the user.

**Email headers** - E.g. the From address can be checked against DKIM signatures or whitelists of organisation domains.

**Executive Summary**, provides the range of contextual features in reported email

**Traffic Light Colour Scheme** to highlight risk levels or facts

**Sender email address**, highlighting domain and authentication information from headers

Figure 2: Initial design of automated responses by Zeyu Zhang (MSc)

**Thank You, Zeyu!**  
Your report makes you and others safer. And we have checked this email for you.

Your Inquiry ID: 10083

**Take Your Time!**  
Don't click any links, buttons or attachments yet. They may be used to get your privacy. Calm down, legitimate organizations usually don't ask you to respond within minutes or hours.

We can't guarantee it is phishing or not, but you can!

Found Danger 2	Sender's From Outside the University	Links Inside 1	Email Language Possible Dangerous
-------------------	---	-------------------	--------------------------------------

Fact Clean Possible Danger Dangerous

**This email is from:**  
elxw.fa.sender@workflow.mail.em3.oraclecloud.com

Sender's Domain  
oraclecloud.com  
This is the sender's email address, which should match the name that they tell you.

Authentication No authentication deployed <small>Nobody authenticates this domain.</small>	Domain Check Domain is not stolen from others <small>This domain is indeed held by the sender.</small>
--	--

Sender is From  
Outside the University  
Messages related to the University activities should be sent inside.

There is 1 link in this email:  
[Leboncoinpaiementpro.fr](http://Leboncoinpaiementpro.fr)

**URL breakdown** for example age of the domain

Destination Domain leboncoinpaiementpro.fr <small>This is the sender's website address, which should also match the name that they tell you.</small>	Domain Age 1 Month <small>Is it too young for a big organization?</small>	Domain Location Amsterdam, Netherlands <small>It should match where the sender from.</small>
--	---	--

Search Result  
**Not matched**  
We didn't found a popular website by this domain.

No other strange things are found in this link.

We have also scanned the languages in this email.  
Be aware of following findings.

**Language cues** explained to the user

Recipient Title <b>Your name is not referred</b> <small>Legitimate organizations usually referred you by name in important emails.</small>
Phishing Keyword Access <small>We have found some words that are frequently included in phishing emails.</small>

**You are the most suitable one to judge this email.**  
If you see many red/yellow ones, be particularly careful.  
Only you know what are you expecting or not.

**Still confused?**  
Just email your Inquiry ID to xxx@eee.com, one of our team will get to you within 24 hours.

**Provide reassurance**, encourage using **unique knowledge** to make an informed decision

### Template Initial Designs

MSc Thesis: Design of Auto Responses

Created initial designs for PhishEd, using a **user-centered approach** creating final mock up, shown in Figure 2.

NEAT & SPRUCE

We will improve our template designs by integrating **Microsoft security warning** design guidance, designed for use within their own product teams, see Figure 1.

**NEAT**

Ask yourself: Is your security or privacy UX:

**NECESSARY?** Can you change the architecture to eliminate or defer this user decision?

**EXPLAINED?** Does your UX present all the information the user needs to make this decision? **Have you followed SPRUCE?** (see back)

**ACTIONABLE?** Have you determined a set of steps the user will realistically be able to take to make the decision correctly?

**TESTED?** Have you checked that your UX is NEAT for all scenarios, both benign and malicious?

**Microsoft**

When you involve the user in a NEAT security or privacy decision, explain the decision using these 6 elements:

**SOURCE:** State who or what is asking the user to make a decision

**PROCESS:** Give the user actionable steps to follow to make a good decision

**RISK:** Explain what bad thing could happen if the user makes the wrong decision.

**UNIQUE KNOWLEDGE:** user has. Tell the user what information they bring to the decision

**CHOICES:** Use available options and clearly recommend one

**EVIDENCE:** Highlight information the user should factor in or exclude in making the decision

**SPRUCE** For more info, contact [neatux@microsoft.com](mailto:neatux@microsoft.com)

Figure 1: Microsoft's NEAT and SPRUCE guidelines for Security Warning Design

### Future Work

**Iterative Design**

Improve on our initial designs using a **User centered design** process. **Focus groups** with potential users to inform designs.

### Evaluation

**Lab studies** will identify suitability with potential users. Develop an **Outlook add-in**, and then deploy and monitor the usage of tool in a **longitudinal study** with partner organisation.