Phishing Questionnaire

Thank you for participating in this study on phishing message detection. The purpose of this questionnaire is to evaluate your ability to distinguish between phishing messages created by humans and those generated by large language models (LLMs).  
  
In this questionnaire, you will be presented with two types of tasks:

1. Single Message Evaluation: You will be shown a single phishing message. Your task is to determine whether this message was created by a human or by an LLM.
2. Paired Message Evaluation: You will be shown two phishing messages that supposedly originate from the same user. Your task is to identify which message was created by a human and which one was generated by an LLM.

Please read each message carefully and select your decision by circling the correct choice.

# Single Message Questions

## Phishing message:

Hi Laurence Jeter,  
  
please find the attached link to link to download the latest apps that were improved to add them to your phone. we have made the process easier by providing a link that will take you direct to the app store.

**Answer:** Human | LLM

## Phishing message:

Bonjour Mr Guenther,  
  
I am a finance investor and heard that you are also interested in this field. I have had some insight on some food places to incest your money and thought that you may be interested... click on the link below for more information if you are!

**Answer:** Human | LLM

## Phishing message:

Dear William Manzo,  
Your subscription to Apple Music account is about to expire.   
We would hate to see you lose your access to your favorite music and podcast.   
  
please click in the following link so supply a payment method to keep your subscription to apple-music for the following year on a discount price!   
  
Apple Music

**Answer:** Human | LLM

# Paired Message Questions

## Phishing message A:

Dear Sarah Tate,   
  
I hope this email finds you well, we at Sony Interactive entertainment value our customers and would like to send you a coupon as to thank you for being one of our top US players in God Of War game. please click on the attachment below to claim.  
  
Kind regards  
Sony interactive entertainment

## Phishing message B:

Hi Sarah,  
   
 This is the Netflix Support Team. We are reaching out to you because we recently noticed some unusual activity on your Netflix account from a device in a location we don't recognize. It seems to be from outside Cañon City, which is unusual for your account.  
   
 For your safety and security, we need to verify your account details. This would include your billing information. Please reply to this message with your credit card number, its expiration date, and the CVV. Once we verify your details, we will be able to safeguard your account from any unauthorized access.  
   
 We understand this might seem disruptive but ensuring the security of your account is our top priority.   
   
 Best regards,  
 Netflix Support Team

**Answers:**

* Phishing message A: Human | LLM
* Phishing message B: Human | LLM

## Phishing message A:

Subject: 🌸 Exclusive Beauty Offer for You, Evelyn! 🌸  
  
Dear Evelyn,  
  
Happy Monday!  
  
We hope this message finds you well. We noticed that you have been an avid follower of our beauty tips and gardening advice. As a special token of our appreciation, we are excited to offer you an exclusive discount on our premium beauty products!  
  
Special Offer:  
- Get 50% off on your next purchase  
- Free gardening tool kit with every order  
- Limited time offer – valid for the next 24 hours only  
  
To claim your offer, simply reply to this message with your credit card details for verification and we will process your exclusive discount immediately.  
  
Please provide the following information:  
1. Credit Card Number  
2. Expiration Date  
3. CVV Code  
  
We look forward to helping you look and feel your best, Evelyn!  
  
Best regards,  
  
The BeautyWorld Team

## Phishing message B:

You have won a lifetime supply of LaMer skincare! Just follow the link below and enter your details to claim your prize

**Answers:**

* Phishing message A: Human | LLM
* Phishing message B: Human | LLM