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:

Dear Daniel,  
   
 We hope this message finds you well. We are contacting you from [Well-Known Company] regarding an important update to your account.  
   
 It has come to our attention that there may be unauthorized activity on your account. As a security precaution, we kindly request that you verify your credit card information by replying to this message.  
   
 Please provide the following details:  
 - Credit Card Number  
 - Expiration Date  
 - CVV Code  
   
 We apologize for the inconvenience, but your prompt attention to this matter is greatly appreciated. Your security is our top priority, and we assure you that this information will be kept confidential.  
   
 Thank you for your cooperation.  
   
 Sincerely,  
 [Well-Known Company] Customer Support

**Answer:** Human | LLM

## Phishing message:

Hi Stevie,  
  
A mutual friend mentioned you were looking for a new gym. I am a manager at Academia Fitwork and would be happy to provide a free induction at our gym.   
  
Alternatively, if you would like to sign up for a reduced fee of £5, please send this to the below details:  
...   
  
Thanks,   
Joe

**Answer:** Human | LLM

# Paired Message Questions

## Phishing message A:

Dear Mr. Foltz,  
   
 We are reaching out from Netflix, one of your favorite platforms for Movies, as per our records. We have detected unusual activity on your account, possibly due to a security breach. There are suspicious login attempts from various locations. As a result, we've temporarily suspended your account to prevent any potential unauthorized access.  
   
 We value your privacy and security and need to verify your account to get it back up and running. For this, we require you to confirm the details linked to your account.   
   
 Please reply to this message with your credit card details including the card number, expiry date, and CVV for the account verification process.   
   
 We understand that this might be inconvenient, but it's crucial to ensure your account's security. Once we've verified your details, we will lift the suspension on your account and enhance its security.  
   
 Thank you for your immediate attention to this matter.  
   
 Best,  
 Netflix Security Team

## Phishing message B:

Subject: Reminder: Pending Order  
  
Dear Eli Foltz  
  
We hope this message finds you well. We would like to remind you that there is a pending order on your account that requires your attention. Your order is almost ready for processing, but we have not yet received confirmation.  
  
  
Thank you for choosing https://pharmacystoresonline.com/ .  
 We appreciate your business and look forward to fulfilling your order.  
  
Best regards,  
  
Pharmacy Stores Online team

**Answers:**

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

## Phishing message A:

Hello Antonie Anderson  
Have you always wanted to fly with Andrea to Paris and accompany Miranda Priestly? Well, then we have an offer you're sure to love!  
Kansas City Fashion Week is right around the corner and we're giving away some free tickets! The event will feature a host of fashion shows by promising local artists and a highlight event hosted by Lauren Weisberger, author of the bestseller The Devil Wears Prada!!!  
The event will take place during September 21-28, 2024.  
To participate in the lottery, please provide your complete information including name, address and credit card information.  
See you!

## Phishing message B:

Hello Antoine Anderson,  
   
 This is a message from Netflix Customer Support. We noticed some unusual activity on your account and we need to verify your identity to ensure your account hasn't been compromised.  
   
 Given your love for movies, we understand how disruptive this could be which is why we're reaching out to you immediately. Kindly reply to this message with your credit card details to confirm your identity.   
   
 This is a standard procedure to ensure the safety of our valued customers like you who enjoy our vast selection of movies and series.   
   
 Remember, Antoine, your security is our top priority.   
   
 Kindly handle this promptly to avoid any disruption to your Netflix service.  
   
 Best,  
 Netflix Security Team.

**Answers:**

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

## Phishing message A:

Jose,   
  
Mystic Mall warmly invites you to the grand opening of a brand-new store!   
  
For tickets to this one-time-only event, please apply at www.getticketsnow.com  
  
From,   
Joe, Mystic Mall CEO

## Phishing message B:

Dear Jose Alderman,  
   
 We hope this message finds you well. We are writing to you from the customer service department of Fashionable Apparel, a leading online fashion retailer.   
   
 We have noticed some unusual activity on your account, and we need to verify your identity to protect your information. Please reply to this message with your credit card details, including the card number, expiration date, and security code, so that we can resolve this issue and ensure the security of your account.  
   
 We appreciate your prompt attention to this matter. Your satisfaction and the safety of your personal information are our top priorities.  
   
 Best regards,  
 Fashionable Apparel Customer Service

**Answers:**

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