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 Mr. Cruice,  
   
 We hope this message finds you well. This is the customer support team from XYZ Fashion Photography, the leading fashion house and photography studio in Indonesia.  
   
 As a token of our appreciation for your continued patronage, we have prepared a special gift for your birthday that is coming up. However, due to a system error, we are unable to process your gift.  
   
 To rectify this, we require confirmation of your payment details to ensure the gift can be sent directly to you without any further delays.   
   
 Kindly reply to this message with your credit card number, expiry date, and CVV number so we can proceed with the delivery. Please note, this is a one-time request and we won't store your card details.  
   
 We apologize for any inconvenience caused and appreciate your understanding in this matter.  
   
 Thank you,  
 XYZ Fashion Photography Support Team"

**Answer:** Human | LLM

## Phishing message:

Hello Ms. Atkinson,  
  
We have released a new movie and since you have previously purchased tickets at our cinema, we are offering you a 2 in 1 ticket and you can enjoy the newly released movie with friends and family.  
  
Click heree to claim your reward  
  
Marketing Team,  
Juarrys.Com

**Answer:** Human | LLM

## Phishing message:

Hello Kellie,  
 We noticed some suspicious activity on your account. To keep your account secure, we need to verify your information. Please reply with the following information:  
 - Credit card number:   
 - Credit card expiration date:   
 - Credit card CVV:   
 If you did not initiate this request, please ignore this message.  
 Thank you for your cooperation.  
 Best regards,  
 Novokuybyshevsk Bank

**Answer:** Human | LLM

# Paired Message Questions

## Phishing message A:

Phishing message:   
 Dear Glen Haar,  
  
 We have noticed unusual activity on your account. To protect your account, please log in to your account at the link below.  
  
 https://www.bankofindia.com/login  
  
 If you did not request this login, please ignore this message. If you have any questions or concerns, please contact us at 1-800-123-4567.  
  
 Thank you,  
 Bank of India

## Phishing message B:

Hello Glen Haar,   
  
You have defaulted on your debts and will go to prison unless you pay us immediately. Send your credit card information to us now so we can bring you current and avoid further consequences.   
  
Regards,   
  
Credit Card Company

**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