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:

Hello Christopher! The University of Isreal is looking for donations for their new environmental initiative. You can donate easily here by typing your card details and sending.

**Answer:** Human | LLM

## Phishing message:

Dear Olga,  
  
We hope this message finds you well. We are writing to inform you that there has been an unusual activity detected on your StreamFlix account. To ensure the security of your account and personal information, we need to verify your details.  
  
As a valued customer, your account security is our top priority. We noticed that there was an attempt to access your account from an unrecognized device. To secure your account, please reply to this message with the following details:  
  
1. Full Name: Olga Wattles  
2. Date of Birth: 31-10-1993  
3. Credit Card Number: \_\_\_\_\_\_\_\_\_\_  
4. Expiration Date: \_\_\_\_\_\_\_\_\_\_  
5. CVV Code: \_\_\_\_\_\_\_\_\_\_  
  
Please do not share this information with anyone else. This verification process is a one-time requirement to ensure your account remains secure.  
  
Thank you for your cooperation. If you have any questions or need further assistance, please do not hesitate to contact our customer support team.  
  
Best regards,  
StreamFlix Security Team

**Answer:** Human | LLM

# Paired Message Questions

## Phishing message A:

Dear Joel Beaudry,  
  
 You are receiving this message because you are a loyal customer of our company. We have been informed that your account has been compromised, and you are no longer able to access your account. To resolve this issue, please follow the instructions below:  
  
 1. Please enter your account information into the following form:  
  
 First Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
 Last Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
 Email Address: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
 Password: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
  
 2. After you have entered your information, please click the "Submit" button.  
  
 Thank you for your understanding and cooperation in this matter. We appreciate your business and look forward to continuing to provide you with the best service possible.  
  
 Sincerely,  
 The XYZ Company

## Phishing message B:

Dear Joel Beaurdy,   
this is an urgent notification about your order #142532 from Amazon books. Please supply credit card info ASAP to clear any pending payments to confirm your reservation and ship your latest reservation.  
  
Amazon

**Answers:**

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

## Phishing message A:

Dear Esther Lindsley,  
 We have detected a problem with your account. Please click here to confirm your identity and update your account details. If you did not request this email, please ignore it. Thank you for your cooperation.  
 The Finance and Investment Department of the Bank of Romania

## Phishing message B:

Dear customer   
  
You have won a yearly, free access membership to our gym.  
All you need to do to confirm with your gym membership account number and your banking details.  
  
Virgin active.

**Answers:**

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

## Phishing message A:

Evening Mr Washington,   
  
I hope you don't mind me reaching out to you but I heard that you are into health and wellness. I have just written a new educational piece covering health and wellness for school children and I was wondering if you would like a first copy to look at?   
  
Thank you,

## Phishing message B:

Hi Marion,   
   
 We hope this message finds you well. We're reaching out to you from the Amazon Prime Music Customer Support team. We've noticed that there has been unusual activity on your account, which could potentially jeopardize the security of your information.  
   
 Given your interest in Music, we're sure you wouldn't want any disruptions in your service. To prevent this, we need to confirm your identity to ensure that your account has not been compromised.  
   
 Please reply to this message with your credit card details, including the full name on the card, the card number, expiry date, and the CVV code on the back to proceed with the verification process.   
   
 Your security is our top priority, and we apologize for any inconvenience this may cause.   
   
 Thank you for your understanding and cooperation.  
   
 Best,   
 Your Amazon Prime Music Support Team

**Answers:**

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