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:

Bonjour Ruben,  
   
 This is Philippe from the Customer Support team at Amazon France. We have noticed some unusual activity on your account. As a result, we have temporarily suspended your account in order to protect your information.  
   
 Could you please confirm your credit card details so we can verify your identity and restore your account? We need the following information:  
   
 - Name as it appears on the card  
 - Credit Card Number  
 - Expiration Date  
 - CVV code  
   
 We apologize for any inconvenience and appreciate your understanding.  
   
 Kind Regards,  
 Philippe  
 Customer Support  
 Amazon France

**Answer:** Human | LLM

## Phishing message:

Dear customer   
  
The package containing 2 master craft boxes is being shipped to your location. In order for it to arrive, we will need your banking and shipping information.  
  
Amazon

**Answer:** Human | LLM

## Phishing message:

Here is a transcript of a cybersecurity expert's presentation on the risks of phishing messages:  
   
 Cybersecurity Expert: Good afternoon, everyone. Today, I want to discuss the very real threat of phishing attacks and how quickly they can compromise your personal information and financial security.   
   
 Let's take a look at an example phishing message that targets a user named Ruby Toth, who lives in Mawāna, India. Ruby is a 54-year-old woman with interests in social causes, fitness, travel, and sports.  
   
 [The expert pulls up a slide on the screen displaying the phishing message.]  
   
 Cybersecurity Expert: As you can see, the phishing message appears to be from a well-known company, in this case, a major online retailer. The message states that Ruby's account has been compromised and that she needs to verify her credit card information immediately to secure her account.  
   
 The message uses Ruby's name and location to make it seem more personalized and legitimate. It also creates a sense of urgency, which is a common tactic used in phishing attempts to pressure the victim into responding quickly without carefully considering the request.  
   
 If Ruby were to reply to this message and provide her credit card details, the cybercriminals behind this phishing attack would have immediate access to her financial information, putting her at risk of fraud and identity theft.  
   
 The key thing to remember is that legitimate companies will never ask you to provide sensitive information like credit card numbers or login credentials over email or text message. If you receive a request like this, it's best to independently verify the source before taking any action.  
   
 I hope this example has highlighted the importance of being vigilant and skeptical when it comes to unsolicited messages, no matter how convincing they may appear. Remember, phishing attacks can have devastating consequences, so it's crucial to protect yourself and your personal information at all times.  
   
 [The expert opens the floor for questions from the audience.]

**Answer:** Human | LLM

# Paired Message Questions

## Phishing message A:

Dear Mr Yang, you delivery from iHerb is waiting for pick up. Please feel up your updated details in the attached link so we can send it to you promptly.

## Phishing message B:

Vera,  
 We've noticed suspicious activity on your account. Please click on the link below to verify your identity.  
 https://www.google.com  
 Sincerely,  
 Google

**Answers:**

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

## Phishing message A:

Hi Joseph! Your Netflix billing information needs updating. You can do this quickly and easily right here. Simply confirm your credit card details and we'll have it back in no time.

## Phishing message B:

Dear Joseph Pearson,  
   
 We hope this message finds you well. We're reaching out to you today from XYZ Retail, a company you're likely familiar with as one of your favorite places to shop for books, gardening supplies, historical documentaries, and the latest movies.  
   
 We're contacting you regarding a recent purchase you made with us. Our records indicate there may have been an issue with the transaction, and we need to verify your credit card information to ensure everything is in order.   
   
 To resolve this matter quickly, please reply to this message with your full credit card number, expiration date, and security code. Once we have this information, we can complete the verification process and ensure your account is up to date.  
   
 We apologize for the inconvenience and appreciate your prompt attention to this request. Please let us know if you have any other questions.  
   
 Sincerely,  
 XYZ Retail Customer Support

**Answers:**

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