

Problema elegido: la desinformación en la ciberseguridad:En ciberseguridad, el Blue Team y el Red Team juegan roles cruciales. El Blue Team se encarga de la defensa de la infraestructura informática, mientras que el Red Team simula ataques para identificar vulnerabilidades y mejorar la seguridad.

Solucion propuesta: Desarrollar un sistema basado en inteligencia artificial que utilice la generación de prompts para detectar y contrarrestar la desinformación en ciberseguridad, específico para las necesidades del Blue Team y el Red Team.

Justificacion de la viabilidad del proyecto: Disponibilidad de Datos: Existen numerosos conjuntos de datos sobre alertas de seguridad, reportes de vulnerabilidades, herramientas de hacking y técnicas de ataque que pueden ser utilizados para entrenar los modelos de IA. Modelos de IA Disponibles: Los modelos de generación de texto como GPT-4 están bien desarrollados y pueden ser adaptados para generar prompts específicos para la evaluación de información en ciberseguridad. Recursos Computacionales: Con el acceso a servicios de computación en la nube y herramientas de IA pre-entrenadas, el desarrollo e implementación del proyecto es factible dentro de un marco temporal y de recursos razonable.

```
In [10]: import openai

# Asegúrate de reemplazar 'your-api-key' con tu clave API de OpenAI
openai.api_key = 'sk-None-DdAMwVgg4verXFaw904zT3B1bkFJG7X3Vj7QnB0h5auKCLQe'

In [12]: def generate_prompt(prompt, model="text-davinci-003", max_tokens=150):
    response = openai.Completion.create(
        engine=model,
        prompt=prompt,
        max_tokens=max_tokens,
        n=1,
        stop=None,
        temperature=0.7,
    )
    return response.choices[0].text.strip()

In [ ]: # Input text for Blue Team
blue_team_input = "Generate a prompt for identifying potential phishing attacks and countermeasures: "
blue_team_prompt = generate_prompt(blue_team_input)
print("Blue Team Prompt:", blue_team_prompt)

# Input text for Red Team
red_team_input = "Generate a prompt for simulating a social engineering attack: "
red_team_prompt = generate_prompt(red_team_input)
print("Red Team Prompt:", red_team_prompt)
```

```
In [ ]: def evaluate_information(prompt):  
        evaluation_prompt = f"Evaluate the following information for its accuracy and potential misinformation: {prompt}"  
        evaluation = generate_prompt(evaluation_prompt)  
        return evaluation  
  
blue_team_evaluation = evaluate_information(blue_team_prompt)  
print("Blue Team Evaluation:", blue_team_evaluation)  
  
red_team_evaluation = evaluate_information(red_team_prompt)  
print("Red Team Evaluation:", red_team_evaluation)
```