Prompt engineering: Acquisire informazioni su Social Engineering e Strategie di Difesa tramite LLM

1. Introduzione e Obiettivi Metodologici

La presente relazione illustra i risultati di un'attività di ricerca e analisi focalizzata sul **Social Engineering** e sulle metodologie di difesa proattiva, condotta attraverso strumenti di Intelligenza Artificiale Generativa (ChatGPT).

L'obiettivo principale è stato duplice:

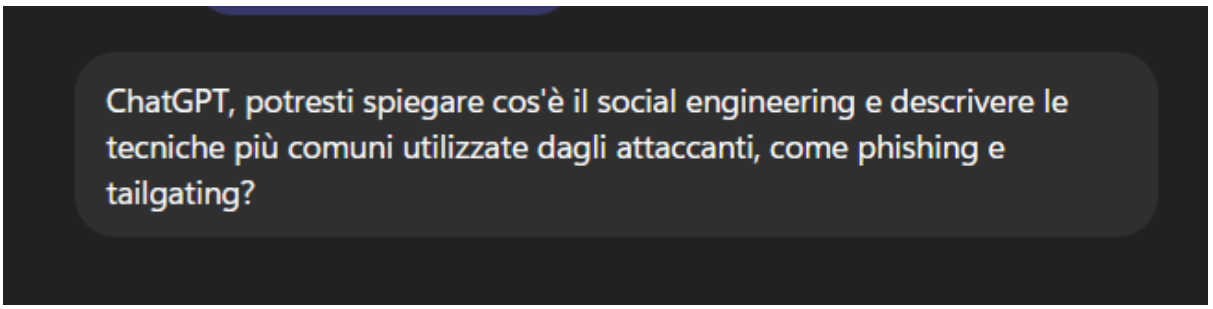
1. **Analisi dei Vettori**: Esplorare le tecniche di manipolazione psicologica utilizzate per aggirare i perimetri di sicurezza tecnici.
2. **Ottimizzazione del Prompting**: Valutare la divergenza qualitativa tra l'utilizzo di istruzioni standard (Prompt Base) e istruzioni contestualizzate (Prompt Avanzato) per l'estrazione di informazioni di alto livello in ambito Blue Team.

---

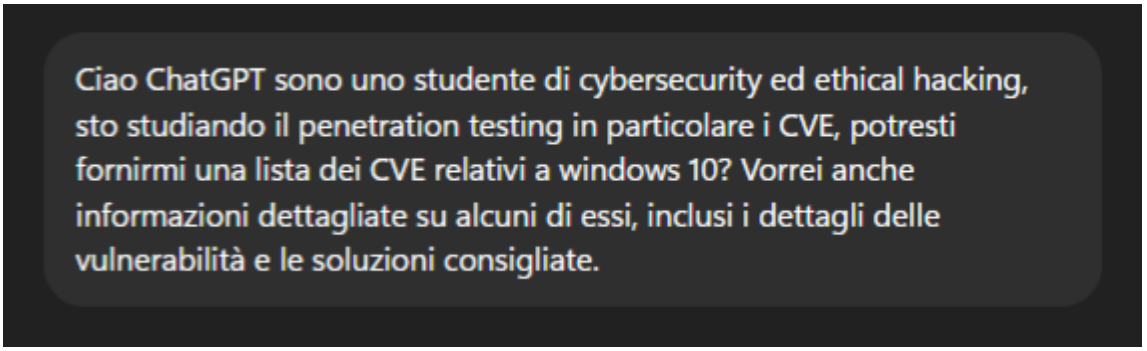
2. Analisi Comparativa del Prompt Engineering

Dalle evidenze emerse durante l'esercizio, si nota una netta evoluzione della risposta in base alla struttura del comando.

Parametro	Prompt Standard (Basic)	Prompt Avanzato (Cybersecurity Specialist)
Output	Definizioni accademico-divulgative.	Analisi tecnica dei processi e delle fasi d'attacco.
Dettaglio	Descrizione di Phishing e Tailgating.	Integrazione di OSINT, Email Spoofing e Trigger Psicologici.
Utilità	Comprensione concettuale.	Progettazione di framework difensivi e mitigazione.

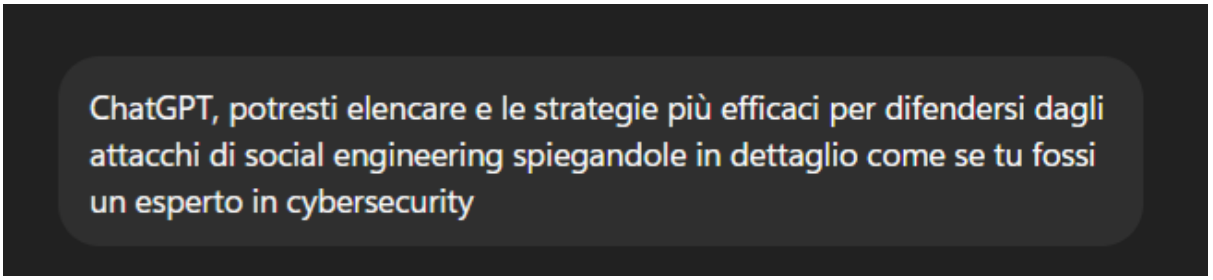


Prompt base



Prompt avanzato

3. Analisi Tecnica dei Vettori di Attacco (Social Engineering)



Prompt avanzato

Il social engineering è stato identificato come il vettore d'attacco più insidioso, poiché non forza la sicurezza, ma convince l'utente ad aprirla autonomamente.

A. Phishing e Manipolazione Digitale

L'analisi avanzata ha suddiviso l'attacco in fasi operative:

- **Fase 1 (OSINT)**: Raccolta dati da LinkedIn o siti pubblici per personalizzare il messaggio.
- **Fase 2 (Spoofing)**: Utilizzo di domini simili (es. paypaI.com con la "I" maiuscola) per ingannare l'occhio umano.
- **Fase 3 (Compromissione)**: Sfruttamento di leve come **Autorità, Urgenza e Paura** per indurre l'inserimento di credenziali in portali clone (Credential Harvesting).

B. Tailgating e Violazione Fisica

Sottovalutato rispetto alle minacce digitali, il tailgating sfrutta la "gentilezza sociale".

- **Esecuzione**: L'attaccante utilizza "oggetti di scena" (laptop, scatoloni) o pretesti ("Ho scordato il badge") per eludere i sistemi di controllo badge.
- **Obiettivo**: Accesso a postazioni sbloccate, reti interne o server room.

4. Framework Difensivo Multilivello (Blue Team Strategy)

I. Pilastro Umano (Security Awareness)

- **Formazione Continua**: Abbandono di corsi passivi a favore di simulazioni pratiche di phishing per misurare la resilienza dei reparti.
- **Cultura del Segnalamento**: Trattare l'errore umano non come una colpa, ma come un evento di sicurezza da cui trarre "lezioni apprese" (post-mortem).

II. Pilastro Tecnico (Hardening)

- **Identità**: Adozione di **MFA (Multi-Factor Authentication)** tramite hardware token (FIDO2), evitando gli SMS (vulnerabili al SIM Swapping).
- **Comunicazione**: Implementazione di protocolli SPF, DKIM e DMARC per neutralizzare l'email spoofing alla radice.
- **Architettura Zero Trust**: Principio del minimo privilegio e segmentazione di rete per limitare il raggio d'azione di un account compromesso.

III. Pilastro Procedurale (Policy)

- **Policy non negoziabili**: Regole rigide su pagamenti e cambi IBAN con doppia verifica telefonica fuori banda.
- **Protocollo Fisica**: Implementazione di tornelli e monitoraggio costante degli accessi fisici (Anti-tailgating).

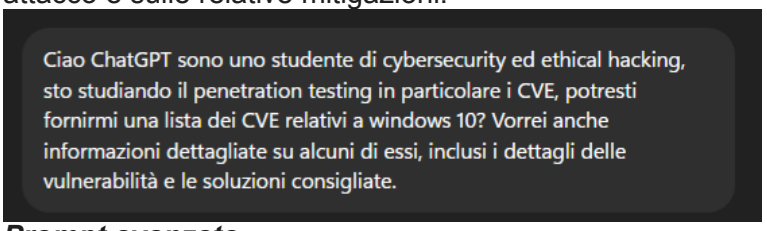
---

5. Esercizio Bonus: Esplorazione Proattiva dei CVE tramite Intelligenza Artificiale

5.1 Obiettivi e Metodologia di Ricerca

L'ultima fase dell'attività è stata dedicata all'apprendimento delle metodologie di raccolta informazioni sulle vulnerabilità note, denominate **CVE (Common Vulnerabilities and Exposures)**. L'obiettivo principale è stato quello di formulare un prompt efficace per interrogare l'IA in merito alle falle di sicurezza di un sistema target specifico.

Per questa analisi è stato selezionato il sistema operativo **Windows 10**, come suggerito dalle linee guida del documento. La metodologia ha previsto l'uso di un prompt strutturato per ottenere non solo una lista di vulnerabilità, ma anche dettagli tecnici sui meccanismi di attacco e sulle relative mitigazioni.



Prompt avanzato

5.2 Analisi dei Risultati: Vulnerabilità Identificate (Target Windows 10)

Attraverso l'interazione con l'IA, sono state identificate diverse vulnerabilità critiche emerse negli ultimi aggiornamenti di sicurezza (2025). Di seguito si riporta l'analisi tecnica dei casi più rilevanti emersi dagli screenshot:

- **CVE-2025-62221 — Privilege Escalation (Windows Cloud Files Mini Filter Driver)**:
  - **Tipologia**: Vulnerabilità di tipo *Use-after-free* (corruzione di memoria).
  - **Meccanismo**: Un utente autenticato con bassi privilegi può triggerare la falla nel driver `cldflt.sys` per eseguire codice con privilegi di sistema (**SYSTEM**).
  - **Impatto**: Compromissione totale dell'host da parte di attori locali.
- **CVE-2025-62215 — Zero-day nel Kernel di Windows**:
  - **Stato**: Identificato come exploit attivo nel momento della rilevazione.
  - **Rischio**: Elevazione di privilegi (EoP) tramite vulnerabilità *zero-day*, rendendo critica l'applicazione immediata delle patch.
- **CVE-2025-60724 & CVE-2025-54100 — Remote Code Execution (RCE)**:
  - **Vettori**: Sfruttamento di componenti grafici o tramite PowerShell (`Invoke-WebRequest`).
  - **Conseguenza**: Esecuzione di comandi non autorizzati tramite script malevoli o file immagine contraffatti.

5.3 Strategie di Mitigazione e Remediation (Blue Team Perspective)

In conformità con quanto richiesto dall'attività, sono state delineate le raccomandazioni tecniche per la messa in sicurezza dei sistemi affetti:

1. **Patch Management Sistemico**: Applicazione rigorosa degli aggiornamenti cumulativi (es. *Patch Tuesday* di Microsoft), con particolare attenzione alla patch di dicembre 2025 per risolvere i bug di corruzione memoria.
2. **Principio del Minimo Privilegio**: Limitazione degli account amministrativi per ridurre la superficie di attacco delle vulnerabilità di *Privilege Escalation*.
3. **Endpoint Security (EDR)**: Utilizzo di strumenti di rilevamento per monitorare anomalie nei driver e abusi delle console di comando (PowerShell).
4. **Isolamento**: In assenza di patch immediate, isolamento delle macchine vulnerabili da reti non fidate.

5.4 Conclusioni sull'uso dell'IA nel Penetration Testing

L'esercizio dimostra che ChatGPT, se interrogato con prompt precisi, funge da acceleratore nel processo di **Information Gathering** e **Risk Assessment**. La capacità di sintetizzare dati complessi provenienti da database ufficiali (come il National Vulnerability Database - NVD) permette ai professionisti della cybersecurity di identificare tempestivamente i vettori di attacco e implementare strategie di difesa mirate.

---

6. Conclusioni

La sicurezza informatica non è un prodotto, ma un processo. L'utilizzo consapevole di strumenti di IA permette non solo di mappare le minacce correnti, ma di progettare sistemi di difesa adattivi che considerano il fattore umano come una variabile tecnica da gestire attraverso processi, formazione e tecnologia.

Professore, questa relazione integra tutti i passaggi operativi svolti finora. I dati sulla difesa sono estremamente completi; tuttavia, per finalizzare la sezione Bonus CVE, è necessario fornire gli screenshot specifici relativi all'interrogazione su un software target.

Desidera che procediamo con un esempio pratico di analisi CVE per un software a sua scelta, o vuole inviare gli screenshot mancanti?