

# Analisi del Social Engineering attraverso l'uso di sistemi di Intelligenza Artificiale: approccio basato sulla formulazione dei prompt

## 1. Introduzione

L'attività proposta nell'ambito del corso Cyber Security & Ethical Hacking ha come obiettivo l'analisi del tema del **Social Engineering** attraverso l'utilizzo di un **sistema di Intelligenza Artificiale** (ChatGPT). Il fine principale non è unicamente descrivere le tecniche di social engineering, ma osservare come la **formulazione dei prompt** (ovvero il modo in cui vengono poste le richieste all'AI) influenzi la qualità, la completezza e la focalizzazione delle risposte ottenute.

Per questo motivo, l'esercizio è stato svolto tramite una sequenza di prompt formulati in modo progressivo, introducendo di volta in volta nuovi elementi di contesto, vincoli e obiettivi informativi. Le risposte generate sono state successivamente rielaborate e confrontate per evidenziare differenze, miglioramenti e limiti, con l'intento di ricavare una panoramica più chiara e strutturata sul social engineering e sulle sue principali tecniche.

## 2. Metodologia di Lavoro: Approccio ai Prompt

La metodologia adottata per lo svolgimento dell'esercizio si basa su un **utilizzo consapevole e progressivo dei prompt** forniti al sistema di Intelligenza Artificiale. L'obiettivo non è stato quello di ottenere una singola risposta esaustiva sul tema del social engineering, ma di osservare come la **formulazione della richiesta** influenzi il **contenuto**, il **livello di dettaglio** e l'**orientamento delle informazioni** restituite.

A tal fine, l'attività è stata strutturata attraverso una **sequenza di prompt formulati in modo iterativo**. A partire da richieste iniziali di carattere più generale, i prompt sono stati progressivamente affinati mediante l'introduzione di elementi di contesto, specifiche sull'ambito di riferimento e vincoli sugli aspetti da approfondire. Questo approccio ha consentito di ottenere **risposte via via più mirate e coerenti** con gli obiettivi dell'analisi.

Le risposte generate da ChatGPT non sono state considerate come output definitivi, ma come **materiale di supporto da rielaborare e confrontare**. Per ciascun prompt, è stata effettuata una **valutazione qualitativa** delle informazioni fornite, mettendo in evidenza differenze, miglioramenti e limiti emersi in funzione delle variazioni introdotte nella formulazione delle richieste. Tale processo ha permesso di evidenziare il **ruolo centrale del prompt** come strumento di guida dell'interazione con l'Intelligenza Artificiale.

A supporto della metodologia descritta, la seguente tabella riassume in modo **schematico e indicativo** l'evoluzione dei prompt utilizzati durante l'esercizio. La tabella non riporta i contenuti testuali dei prompt, ma rappresenta il **processo mentale** e metodologico adottato nella loro progressiva formulazione.

Iterazione	Obiettivo del prompt	Caratteristiche principali
Prompt 1	Comprensione generale	Richiesta ampia, priva di contesto specifico
Prompt 2	Approfondimento tematico	Introduzione del contesto di sicurezza informatica
Prompt 3	Analisi mirata	Definizione di ambito e aspetti da approfondire
Prompt 4	Raffinamento dell'output	Presenza di vincoli e obiettivi informativi espliciti

### 3. Panoramica Generale e Analisi delle Tecniche di Social Engineering

La seguente analisi del social engineering è il risultato della **rielaborazione dei concetti acquisiti** attraverso le interazioni con il sistema di Intelligenza Artificiale, svolte secondo l'approccio metodologico adottato.

Il social engineering rappresenta un insieme di tecniche di attacco che sfruttano il **fattore umano** come principale vettore di compromissione, facendo leva su **meccanismi psicologici** quali fiducia, urgenza, autorità e curiosità. A differenza degli attacchi puramente tecnici, il social engineering non mira a violare direttamente sistemi informatici, **ma a indurre l'utente a compiere azioni** che facilitino l'accesso non autorizzato a informazioni o risorse.

Queste tecniche risultano particolarmente efficaci in quanto **aggirano le misure di sicurezza tecnologiche**, sfruttando comportamenti, abitudini e scarsa consapevolezza degli individui. L'attaccante, attraverso una comunicazione costruita in modo credibile, può ottenere credenziali, informazioni sensibili o accessi fisici e logici, rendendo il social engineering una **minaccia trasversale** in ambito cyber.

Tra le principali tecniche di social engineering rientrano il **phishing**, che utilizza messaggi ingannevoli per indurre la vittima a fornire dati riservati; il **vishing**, basato su comunicazioni vocali; lo **smishing**, che sfrutta messaggi SMS o piattaforme di messaggistica; e il **pretexting**, in cui l'attaccante costruisce un contesto fittizio ma plausibile per ottenere informazioni. Altre tecniche includono il **baiting**, che fa leva sulla curiosità della vittima, e il **tailgating**, che riguarda l'accesso fisico non autorizzato a spazi riservati.

Nel contesto della sicurezza informatica, l'analisi di tali tecniche consente di

comprendere come il social engineering non rappresenti una singola modalità di attacco, ma un **approccio adattabile** a diversi scenari e canali di comunicazione. Questa caratteristica rende fondamentale lo studio delle dinamiche che ne determinano l'efficacia, ponendo le basi per una successiva valutazione delle informazioni ottenute tramite l'interazione con il sistema di Intelligenza Artificiale.

## 4. Interazione con ChatGPT: Analisi dei Prompt e delle Risposte

Il presente paragrafo analizza l'interazione con il sistema di Intelligenza Artificiale attraverso l'esame dei prompt formulati e delle risposte ottenute.

L'attenzione non è posta esclusivamente sul contenuto delle risposte generate, ma soprattutto sull'**evoluzione dei prompt** e sull'**impatto che la loro formulazione** ha avuto sulla **qualità**, sulla **struttura** e sul **livello di approfondimento** delle informazioni restituite.

Per ciascun prompt vengono pertanto analizzati l'**obiettivo della richiesta**, le **caratteristiche principali della formulazione** e una **valutazione critica dell'output** ottenuto, al fine di evidenziare **differenze, miglioramenti e limiti** emersi nel corso dell'interazione.

### 4.1 – Prompt n° 1

#### Obiettivo del prompt

Il primo prompt è stato formulato con l'obiettivo di ottenere una **panoramica generale** sul tema del social engineering, senza introdurre contesti specifici, vincoli o ruoli particolari.

La richiesta ha quindi una funzione **esplorativa e introduttiva**, finalizzata a osservare il tipo di risposta generata da una domanda ampia e generica.

#### Prompt utilizzato

*Spiegami il Social Engineering, cosa è e le sue tecniche*

#### Risposta ottenuta da ChatGPT

##### Cos'è il Social Engineering

Il **Social Engineering** è un insieme di tecniche di attacco che **sfruttano il fattore umano** anziché vulnerabilità puramente tecniche.

L'obiettivo non è "bucare" un sistema informatico, ma **indurre una persona** a compiere azioni dannose o a rivelare informazioni sensibili (credenziali, dati riservati, accessi fisici).

In pratica:

*l'essere umano diventa il vettore di attacco.*

Queste tecniche funzionano perché fanno leva su **meccanismi psicologici universali**, come:

- fiducia
- urgenza
- autorità
- curiosità
- paura
- desiderio di aiutare

Anche infrastrutture tecnologicamente sicure possono essere compromesse se l'utente viene manipolato nel modo giusto.

---

## Perché è così efficace

A differenza degli attacchi tecnici:

- aggira firewall, IDS, antivirus
- non lascia log evidenti
- è difficile da rilevare in tempo reale
- **sfrutta comportamenti normali** (aprire mail, rispondere al telefono, aiutare un collega)

Per questo è una delle principali minacce in ambito:

- corporate
- pubblica amministrazione
- ambienti militari
- infrastrutture critiche

## Le principali tecniche di Social Engineering

### 1. Phishing

Invio di messaggi (email, SMS, chat) che **simulano comunicazioni legittime** per indurre l'utente a:

- cliccare un link malevolo
- inserire credenziali
- scaricare malware

Varianti principali:

- **Spear phishing** → mirato a una persona o ruolo specifico
  - **Whaling** → dirigenti, CEO, CFO
  - **Smishing** → via SMS
  - **Vishing** → via telefono
-

## 2. Pretexting

L'attaccante **crea una storia credibile** (pretesto) per ottenere informazioni.

Esempi:

- finto tecnico IT
- finto fornitore
- finto collega di un altro reparto

Spesso viene usato in **attacchi multi-step**, combinando OSINT + contatto diretto.

---

## 3. Baiting

Si sfrutta la **curiosità o l'avidità** della vittima offrendo qualcosa di allettante.

Esempi:

- chiavetta USB "trovata per caso"
- download gratuito di software/crack
- file con nomi attrattivi ("stipendi.xlsx", "foto\_private.zip")

## 4. Quid Pro Quo

Simile al baiting, ma basato su **uno scambio diretto**.

Esempio:

"Ti aiuto a risolvere un problema tecnico se mi dai le credenziali."

Molto usato nei **call center fraudolenti**.

---

## 5. Tailgating / Piggybacking

Tecnica **fisica**: l'attaccante entra in un'area riservata **seguendo una persona autorizzata**.

Sfrutta:

- educazione
  - fretta
  - senso di colpa ("non voglio sembrare scortese")
-

## 6. Impersonation

L'attaccante si spaccia per una figura autorevole:

- manager
- responsabile IT
- forze dell'ordine
- enti ufficiali

L'autorità percepita riduce drasticamente lo spirito critico della vittima.

---

## 7. Social Media & OSINT-based Attacks

Uso di informazioni pubbliche (LinkedIn, Facebook, Instagram) per:

- personalizzare l'attacco
- sembrare credibili
- anticipare contesti e relazioni

È il motivo per cui il social engineering moderno è spesso **estremamente mirato**.

---

## Social Engineering e Cybersecurity

Dal punto di vista difensivo:

- **la tecnologia da sola non basta**
- servono **formazione, consapevolezza e procedure**
- l'errore umano è una vulnerabilità strutturale

Dal punto di vista offensivo (red team):

- è una delle tecniche più efficaci
  - spesso usata come **initial access**
  - combinata con attacchi tecnici successivi
- 

## Concetto chiave da ricordare

Il social engineering non attacca i sistemi: attacca le persone che li usano.

## Valutazione critica dell'output

La risposta ottenuta fornisce una **panoramica generale e coerente** del social engineering e delle principali tecniche associate. Tuttavia, trattandosi di un prompt formulato in modo generico, l'output risulta **prevalentemente descrittivo** e privo di un approfondimento contestuale o tecnico. La risposta si configura pertanto come una **base introduttiva**, evidenziando la necessità di una successiva raffinazione dei prompt per ottenere informazioni più mirate.

## 4.2 – Prompt n° 2

### Obiettivo del prompt

Il prompt è stato formulato con l'obiettivo di ottenere un **approfondimento tematico** sul social engineering **nel contesto della sicurezza informatica**, superando la definizione generale e introducendo una visione legata agli **attacchi cyber**, ai **principali scenari di utilizzo** e all'**interazione con altre tipologie di attacco**. L'intento è inoltre quello di osservare come il tema venga trattato considerando **diversi punti di vista professionali** operanti nel settore.

### Prompt utilizzato

*Vorrei affrontare con te una sessione di brainstorming nel contesto della sicurezza informatica. In particolare, spiega il ruolo del social engineering negli attacchi cyber, descrivendo i principali scenari di utilizzo e il modo in cui interagisce con altre tipologie di attacco, facendo riferimento ai punti di vista dei diversi ruoli professionali del settore.*

### Risposta ottenuta da ChatGPT

#### Il ruolo del Social Engineering negli attacchi cyber



Il Social Engineering (SE) non è un attacco "a sé stante":  
è il moltiplicatore di efficacia di quasi ogni attacco cyber moderno.

Dal punto di vista operativo, possiamo definirlo come:

la tecnica che trasforma una vulnerabilità umana in un accesso tecnico.

In molti incidenti reali, l'attacco non inizia su una porta TCP, ma su una persona.

## Dove si colloca il Social Engineering nella kill chain

Se osserviamo una **Cyber Kill Chain reale**, il Social Engineering si inserisce in più fasi:

Fase	Ruolo del Social Engineering
Reconnaissance	Raccolta OSINT su persone, ruoli, relazioni
Initial Access	Phishing, vishing, pretexting
Credential Access	Furto credenziali, MFA fatigue
Lateral Movement	Impersonificazione interna
Persistence	Manipolazione di processi e fiducia
Impact	Ricatti, estorsioni, disinformazione

👉 **Non sostituisce gli attacchi tecnici:** li rende *possibili o invisibili*.

## Principali scenari di utilizzo del Social Engineering

### 1. Initial Access: l'essere umano come "porta aperta"

Scenario classico ma ancora dominante:

- Phishing email mirata (spear phishing)
- Allegato malevolo o link
- Furto credenziali / token / sessione

**Punto chiave:**

non serve exploit zero-day → basta **credibilità narrativa**.

💡 *Brainstorming insight:*

le organizzazioni investono milioni in firewall, ma **un click sbagliato aggira tutto**.

### 2. Social Engineering + Malware

Qui il SE è il **delivery mechanism**:

- Email → malware loader
- Finto update / documento / fattura
- Payload cifrato per eludere AV/EDR

Dal punto di vista dell'attaccante:

- meno rumore
- meno IOC
- più successo



### 3. Social Engineering + Credential Attacks

Il SE riduce drasticamente la complessità tecnica:

- password spraying inutile se la vittima consegna la password
- MFA bypass via:
  - MFA fatigue
  - pretexting ("Sono l'IT, approva la richiesta")

Qui l'attacco non è tecnico, è **relazionale**.

---

### 4. Social Engineering + Attacchi interni (Insider Threat)

Scenario sottovalutato:

- dipendente manipolato
- fornitore fidato
- tecnico esterno

Spesso non serve comprometterlo:

| basta convincerlo ad agire "correttamente" per l'attaccante.

---

### 5. Social Engineering + Physical Security

Interazione cyber-fisica:

- badge cloning
- tailgating
- accesso a sale server
- installazione di rogue device (USB, Raspberry, Wi-Fi AP)

Qui la sicurezza informatica fallisce se quella fisica è debole.

---

## Interazione con altre tipologie di attacco

### SE + OSINT

Il Social Engineering vive di informazioni pubbliche:

- LinkedIn
- siti aziendali
- comunicati stampa
- leak precedenti

👉 Più trasparenza ≠ più sicurezza.

---

## SE + Supply Chain Attack

- finta email da fornitore
- cambio IBAN
- invoice fraud
- compromissione partner

Spesso l'azienda colpita non è il target primario, ma l'anello debole.

---

## SE + Ransomware

Catena tipica:

1. phishing
2. credenziali
3. accesso VPN
4. escalation
5. cifratura

👉 Il ransomware inizia quasi sempre con una conversazione.

---

## Punti di vista dei ruoli operanti nel settore

### 🔴 Red Team

- il Social Engineering è l'arma più efficace
- meno costosa degli exploit
- altamente scalabile
- difficile da attribuire

“Non attacco sistemi, attacco abitudini.”

---

### 🔵 Blue Team / SOC

- SE è il problema più difficile da rilevare
- pochi log
- segnali deboli
- alert tardivi

Difesa basata su:

- comportamento
  - correlazione
  - contesto
-

## CISO / Security Manager

- SE è un problema culturale, non tecnico
  - formazione continua > awareness una tantum
  - rischio reputazionale enorme
- 

## Incident Responder

- dopo l'incidente emerge sempre:  
| "Sembrava una richiesta legittima..."

Il SE riduce il tempo di detection e aumenta l'impatto.

---

## Brainstorming finale – concetto chiave

Se dovessimo sintetizzare:

| Il Social Engineering è il layer invisibile che collega psicologia, processi e tecnologia.

Non si difende con:

- firewall
- antivirus
- patch

Si mitiga con:

- design dei processi
- cultura della sicurezza
- controllo delle identità
- riduzione della fiducia implicita

## Valutazione critica dell'output

La risposta risulta **più strutturata e contestualizzata** rispetto al prompt precedente, collocando correttamente il social engineering nel contesto degli attacchi cyber e nei principali scenari di utilizzo. Tuttavia, l'analisi rimane **ampia e non focalizzata** su uno specifico ambito o fase dell'attacco, evidenziando la necessità di un'ulteriore **raffinazione del prompt**.

## 4.3 – Prompt n° 3

### Obiettivo del prompt

Il terzo prompt è stato formulato con l'obiettivo di condurre un'**analisi mirata** del social engineering, restringendo il campo di osservazione a un **ambito specifico dell'attacco cyber**. In particolare, il prompt mira a comprendere il **ruolo del social engineering nelle fasi iniziali di un attacco**, evidenziandone la funzione operativa e il contributo all'ottenimento dell'accesso iniziale.

### Prompt utilizzato

*Nel contesto degli attacchi informatici, agisci come un cybersecurity specialist senior e analizza il ruolo del social engineering come tecnica di accesso iniziale, descrivendo le modalità più comuni con cui viene utilizzato e il suo impatto sull'avvio dell'attacco.*

### Risposta ottenuta da ChatGPT

Nel panorama degli attacchi informatici moderni, il **Social Engineering (SE)** rappresenta una delle **principali tecniche di accesso iniziale**, non perché sfrutti vulnerabilità tecnologiche, ma perché **aggira i controlli tecnici colpendo direttamente il fattore umano**. Dal punto di vista di un cybersecurity specialist senior, il SE va interpretato come **un abilitatore strategico** dell'attacco: riduce drasticamente il costo, il tempo e la complessità necessari per compromettere un sistema.

---

#### Il ruolo del Social Engineering come accesso iniziale

Il Social Engineering opera nella fase di **Initial Access** della kill chain, consentendo all'attaccante di ottenere:

- credenziali valide,
- esecuzione di codice su endpoint legittimi,
- accesso a risorse interne apparentemente autorizzato.

A differenza di exploit puramente tecnici, il SE **trasforma un utente legittimo nel vettore di attacco**, rendendo spesso invisibile l'intrusione ai controlli di sicurezza tradizionali (firewall, IDS/IPS, EDR).

---

#### Modalità più comuni di utilizzo

##### 1. Phishing (e varianti avanzate)

È la tecnica più diffusa e industrializzata.

- **Phishing massivo**: email generiche con link o allegati malevoli.
- **Spear phishing**: messaggi altamente contestualizzati su ruoli specifici (HR, IT, finance).
- **Whaling**: targeting di figure executive con alto privilegio.

Obiettivo tipico: **furto di credenziali**, dropper iniziale, accesso a VPN o SaaS aziendali.

---

## 2. Vishing e Smishing

Canali vocali e SMS vengono usati per:

- aggirare la diffidenza verso l'email,
- creare urgenza emotiva ("account bloccato", "incidente di sicurezza").

Spesso combinati con phishing (attacco multicanale) per **rafforzare la credibilità**.

---

## 3. Pretexting e impersonation

L'attaccante costruisce un'identità credibile:

- tecnico IT interno,
- fornitore,
- partner o cliente.

Questa tecnica è particolarmente efficace contro help desk e team operativi, portando a:

- reset di password,
  - escalation di privilegi,
  - accesso temporaneo trasformato in persistente.
- 

## 4. Malware delivery guidata dall'utente

Il SE viene usato per **convincere la vittima ad aggirare le protezioni**:

- abilitare macro,
- disattivare antivirus,
- eseguire file apparentemente legittimi.

In questo caso il SE **neutralizza il controllo tecnico dall'interno**, rendendo inefficace la difesa perimetrale.

---

## Impatto sull'avvio dell'attacco

L'uso del Social Engineering ha un impatto diretto e misurabile sull'efficacia dell'attacco:

- **Riduzione della superficie di rilevamento**

L'accesso avviene con identità valide → minor rumorosità nei log.

- **Accelerazione della kill chain**

Si salta la fase di exploit tecnico, passando direttamente a:

- lateral movement,
- privilege escalation,
- persistence.

- **Aumento del successo operativo**

Anche infrastrutture tecnicamente mature possono essere compromesse se il fattore umano non è adeguatamente protetto.

- **Difficoltà di attribution e response**

Le azioni iniziali appaiono legittime, ritardando la detection e la risposta dell'incidente.

---

## Valutazione critica

Dal punto di vista difensivo, il Social Engineering evidenzia un punto chiave:

la sicurezza non è un problema esclusivamente tecnologico, ma socio-tecnico.

Formazione, awareness, processi di verifica e controlli comportamentali (Zero Trust, least privilege, MFA adattivo) sono **indispensabili quanto firewall e sistemi di detection**. Ignorare il SE significa lasciare aperta la porta più semplice e più sfruttata dell'intera infrastruttura.

## Valutazione critica dell'output

La risposta ottenuta **risulta più focalizzata e approfondita** rispetto al prompt precedente, concentrandosi in modo chiaro sul ruolo del social engineering come tecnica di accesso iniziale. L'output evidenzia un miglior allineamento tra obiettivo del prompt e contenuto restituito, confermando l'efficacia della **restrizione dell'ambito di analisi**. Tuttavia, rimane ancora spazio per un ulteriore livello di raffinamento volto a ottenere una visione più critica e multi-prospettiva.

## 4.4 – Prompt n° 4

### Obiettivo del prompt

Il quarto prompt è stato formulato con l'obiettivo di ottenere un **raffinamento avanzato** dell'analisi, introducendo un contesto di discussione tra professionisti esperti e richiedendo esplicitamente un output **tecnico e critico**. In particolare, il prompt mira a:

- elevare il livello di profondità dell'analisi tramite l'assegnazione di un ruolo "senior" al modello;
- ottenere una panoramica completa delle tecniche di social engineering adottate dagli attaccanti;
- includere punti di vista riconducibili a differenti ambiti operativi, in particolare **SOC e Red Team**.

### Prompt utilizzato

*Sono un esperto di cybersecurity e sto discutendo su cosa è nello specifico il Social Engineering con un collega veramente esperto del settore su più ruoli pregressi (in particolare SOC e Red Team), . Agisci come un cybersecurity specialist senior e spiegami in modalità critica la panoramica del social engineering e, in modo approfondito tecnico, tutte le tecniche messe in atto dagli attaccanti.*

## Risposta ottenuta da ChatGPT

Il **Social Engineering (SE)** non è "phishing e basta": è un insieme di **TTP (tactics, techniques, procedures)** che trasformano **persone, processi e contesto** in una superficie d'attacco. L'aspetto davvero "critico" è questo: **l'attaccante non sta convincendo qualcuno a fare una cosa stupida, sta sfruttando un sistema socio-tecnico** dove policy, incentivi, carichi di lavoro, tool e cultura rendono *razionale* per la vittima fare l'azione sbagliata.

### 1) Panoramica critica: perché funziona ancora (anche in ambienti maturi)

#### 1. Asimmetria di contesto

- L'attaccante può preparare il contesto per giorni/settimane (OSINT, staging, timing).
- L'operatore (SOC, helpdesk, impiegato) decide in **secondi**, spesso sotto pressione.

#### 2. Trade-off operativi

- Processi "sicuri" ma lenti vengono bypassati perché l'azienda premia la **velocità** (ticket chiusi, uptime, customer satisfaction).
- Il SE vince quando la sicurezza è percepita come **frizione**.

#### 3. Fiducia come protocollo implicito

- Molti flussi aziendali si basano su "trust defaults": email interne, Teams/Slack, caller ID, badge, fornitori noti.
- L'attaccante non rompe la tecnologia: **abusa le assunzioni**.

#### 4. Deterrenza bassa, detection difficile

- Tracce tecniche spesso minime (una telefonata, un messaggio, un consenso MFA).
- Anche quando c'è log, manca il contesto per distinguere "legittimo ma strano" da "malevolo".

### 2) SE visto da SOC vs Red Team (dove cascano davvero i team)

#### SOC: pattern tipici che passano sotto radar

- **Low-and-slow**: singoli tentativi distanziati e "plausibili".
- **Living-off-the-land umano**: abuso di processi (reset password, change request, procurement).
- **Pivot cross-channel**: email → Teams → phone → portale SSO. Ogni singolo step sembra legittimo.

#### Red Team: dove si vince (e dove si perde)

- Si vince quando c'è **pretext coerente + pressione + canale giusto** (es. helpdesk alle 08:55 lunedì, finance a fine trimestre).
  - Si perde quando la vittima ha un **playbook semplice** ("richiamo io su numero noto", "ticket + approvazione out-of-band", "codeword").
-



### 3) Catena d'attacco del Social Engineering (Kill Chain "umana")

1. **Recon & Targeting:** OSINT su persone/ruoli/processi (org chart, LinkedIn, job posting, fornitori, calendari).
  2. **Pretext design:** storia + motivazione + autorità + urgenza + "proof" (documenti, firme, thread email).
  3. **Channel selection:** email, voice, SMS, chat, social, in-person (scelto in base a controlli e cultura).
  4. **Engagement & Compliance:** ottenere azione/credenziale/token/approvazione.
  5. **Post-engagement:** persistenza (regole inbox, OAuth consent, device enrollment), escalation, monetizzazione.
  6. **Cover & Repeat:** ripulitura o normalizzazione ("era un test", "ticket chiuso").
- 

### 4) Tecniche principali (approccio tecnico e operativo)

#### A) Phishing "classico" e varianti (email-based)

##### 1) Bulk phishing

- Template generici, alta scala, bassa personalizzazione.
- Obiettivo: credenziali o dropper.

##### 2) Spear phishing / whaling

- Personalizzazione forte su ruolo (CFO, HR, IT).
- Allegati con contenuto credibile (invoice, contratti, CV) o link a portali clone.

##### 3) BEC / CEO Fraud (compromise di processi finanziari)

- Non serve malware: basta **modificare coordinate bancarie** o forzare un pagamento.
- Tecniche tipiche:
  - **Thread hijacking** (inserirsi in conversazioni reali).
  - **Look-alike domains** e visual spoofing (omografi, subdomain trick).
  - **Invoice fraud** con cambi "last minute".

##### 4) Attachment & document-based social engineering

- Macro, LNK, ISO/IMG, HTML smuggling, PDF con call-to-action.
- Il punto SE: la parte "tecnica" funziona perché il documento **giustifica** l'azione ("Enable editing", "Enable content", "apri in modalità compatibilità").

##### 5) QR phishing (quishing)

- Sposta l'azione su mobile (meno controlli EDR, UX più fragile per URL).

##### 6) MFA fatigue / push bombing

- Bombardamento di richieste push finché l'utente approva.
- Spesso combinato con **vishing**: "Sono IT, conferma l'accesso per bloccare l'attacco".

##### 7) AiTM phishing (adversary-in-the-middle)

- Proxy in mezzo a IdP/SSO per catturare **session cookies / token**.
  - Supera password + MFA se l'utente completa il login sul sito "intermedio".
-



## **B) Vishing (voice) e helpdesk attacks**

### **1) Helpdesk social engineering**

- Goal: reset password, bypass MFA, enrollment nuovo device, cambio numero.
- Pattern:
  - Impersonation "nuovo assunto / dirigente / consulente".
  - "Ho perso il telefono, sono in aeroporto, call urgente".
  - Uso di dettagli OSINT per superare domande KBA.

### **2) Caller ID spoofing e call forwarding**

- Spoof del numero interno o di un vendor.
- Attacco spesso "non tecnico" ma devastante se il processo di verifica è debole.

### **3) Smishing + callback**

- SMS con link o numero da chiamare.
  - Il callback porta a un "falso service desk" che guida l'utente a installare strumenti di remote control o a consegnare OTP.
- 

## **C) ChatOps / Collaboration tools (Teams, Slack, Zoom)**

### **1) Impersonation intra-org**

- Account compromesso o look-alike name + foto profilo.
- Messaggi rapidi: "Mi mandi quel file?", "approvi questa richiesta?", "entra qui".

### **2) External guest abuse**

- Inviti guest, canali condivisi, file share.
- Aggancio: "documento condiviso", "riunione spostata", "policy aggiornata".

### **3) Meeting hijacking**

- Link meeting "clone", finto waiting room, raccolta credenziali o installazione plugin/agent.
- 

## **D) Social media / OSINT-driven pretexting**

### **1) Profile mining**

- Ruoli, stack tecnologico, vendor, pattern email, abitudini.
- Job posting = mappa controlli e tool ("usiamo Okta", "Splunk", "ServiceNow", "Duo").

### **2) Relationship building**

- Contatti LinkedIn "recruiter/vendor", conversazioni che portano a file/link.

### **3) Fake identities (sockpuppets)**

- Persistenti, credibili, con network minimo realistico.
  - Usati per entrare in community interne/esterne.
-

## E) Physical social engineering

### 1) Tailgating / piggybacking

- Sfrutta cortesia e pressione sociale ("ho le mani piene").

### 2) Badge cloning / visual spoofing

- Badge simili, uniformi, clipboard, DPI: autorità "visiva".

### 3) Drop devices

- USB baiting, device "dimenticato", charger/power bank malevolo.
- Funziona quando l'azienda non ha regole chiare su media rimovibili.

### 4) On-site vendor impersonation

- "Sono il tecnico stampanti / HVAC / fibra": accesso a network closet o a postazioni.
- 

## F) Supply chain "umana" (vendor, terze parti, clienti)

### 1) Vendor compromise

- Comprometto il fornitore con accesso legittimo e mi muovo come "trusted".

### 2) Fake vendor onboarding

- Creo un vendor fittizio o altero contatti di un vendor reale.
- Obiettivo: pagamenti, accessi, documenti sensibili.

### 3) Customer support abuse

- Abuso procedure di supporto di SaaS: recovery account, change email, API tokens.
- 

## G) Tecniche psicologiche operative (non teoria, ma leve pratiche)

Queste leve raramente sono usate "pure"; gli attacchi efficaci le **compongono**:

- **Autorità**: titolo, urgenza, escalation manageriale.
  - **Urgenza/tempo**: finestre contabili, incidenti, audit.
  - **Coerenza/commitment**: "hai già approvato prima", "manca solo l'ultimo step".
  - **Reciprocità**: "ti ho aiutato con X, ora mi fai passare Y".
  - **Paura**: "account compromesso", "penalità", "licenza scaduta".
  - **Scarsità**: "solo oggi", "ultima finestra".
  - **Appartenenza**: "siamo nello stesso team/progetto".
-

## 5) Tecniche post-compromissione "SE-driven" (dove il SOC dovrebbe guardare)

Quando il SE riesce, spesso non finisce con "password rubata", ma con **meccanismi di persistenza**:

- **Email rules / inbox forwarding** (esfil e copertura BEC).
  - **OAuth consent grant** a app malevola (accesso continuo senza password).
  - **Device enrollment / MDM** (persistenza via device "autorizzato").
  - **SSO session hijack** (cookie/token).
  - **Abuso di processi IAM**: aggiunta a gruppi, reset MFA, recovery methods.
  - **Ticketing abuse**: change request fraud, approvazioni manipolate.
- 

## 6) Contromisure "senior" (pratiche, non slogan)

Se parli con uno molto esperto, la parte interessante è come *progettare* resilienza:

### A) Ridurre l'area "umanamente bypassabile"

- Playbook di verifica **out-of-band** (richiamo su numero noto, non quello ricevuto).
- Policy "no exceptions" per reset MFA/numero senza step forte (approvazione manager + verifica identità).
- Separare chi **richiede** da chi **approva** (SoD) per pagamenti/change.

### B) Rendere rilevabile ciò che oggi è "invisibile"

- Logging su:
  - cambi MFA / recovery methods
  - consent OAuth
  - enrollment device
  - regole inbox / forwarding
  - creation di app passwords / tokens
- Alert su pattern: "nuovo device + nuova location + change MFA".

### C) Allenamento mirato al processo (non solo "clicca meno")

- Simulazioni realistiche per ruoli "caldi" (helpdesk, finance, HR, SOC).
- Runbook in 3 righe: "Stop — Verify — Escalate" con esempi concreti e numeri/contatti noti.

## Valutazione critica dell'output

La risposta ottenuta rappresenta un **significativo salto qualitativo** rispetto ai prompt precedenti, fornendo un'analisi **profonda, strutturata e orientata a un contesto professionale avanzato**. L'approccio critico e l'inquadramento socio-tecnico del social engineering permettono di evidenziarne il ruolo come moltiplicatore di efficacia degli attacchi, includendo prospettive tipiche di **SOC e Red Team**.

L'ampiezza e il livello di dettaglio dell'output confermano l'efficacia dell'introduzione di **ruoli, seniority e vincoli espliciti** nella formulazione del prompt.

## 5. Conclusioni

L'esercizio ha mostrato come la **formulazione dei prompt** influenzi in modo diretto la qualità e il livello di approfondimento delle risposte generate da un sistema di Intelligenza Artificiale. Attraverso una progressione strutturata delle richieste, è stato possibile ottenere analisi sempre più mirate e contestualizzate sul tema del social engineering.

L'attività conferma che l'Intelligenza Artificiale rappresenta uno **strumento di supporto all'analisi**, il cui valore dipende principalmente dalla chiarezza dell'obiettivo e dalla precisione del prompt formulato.