REGIONE PIEMONTE

CSI your digital partner

WEBINAR
SICURI di essere **SICURI?**

Deepfake, frodi AI-driven e social engineering
di nuova generazione: come difenderci?"

Dall'Intelligenza Artificiale alla Fiducia Artificiale

Roberto Piazzolla - Clusit

Di cosa parleremo

Il bersaglio: la **fiducia**

L'evoluzione dell'**attaccante con l’IA**

Come nasce la **fiducia artificiale**

Difendersi dalla **fiducia artificiale**

Il bersaglio non sono i sistemi

Un attacco efficace contro un'organizzazione deve costruire un rapporto di fiducia.

- Fiducia nella mail
- Fiducia nella voce
- Fiducia in una procedura
- Fiducia nell’interlocutore

Come funziona una truffa

Conoscere

- Chi sei

Convincere

- Perché dovresti fidarti

Indurre ad agire

- Clicca. Rispondi. Autorizza. Paga.

Questo modello non è cambiato. È cambiata la velocità.

Cosa cambia con l'IA

Prima

- Ricerca manuale
- Scrittura manuale
- Telefonata umana

Oggi

- Ricerca automatica
- Scrittura automatica
- Voce sintetica
- Personalizzazione automatica

Cos'è un attacco agentico

In un attacco agentico l'IA, oltre a generare contenuti, esegue una sequenza coordinata di attività per raggiungere un obiettivo.

Un agente osserva

↓

Un agente decide

↓

Un agente Agisce

↓

Un agente verifica il risultato

↓

Un agente si adatta

Dipendenti virtuali al servizio dell'attaccante

Con l'IA è possibile simulare il lavoro di un intero gruppo.

- Ricercatore OSINT
- Analista
- Sviluppatore
- Copywriter
- Traduttore
- Operatore di contact center

Scenario: attacco ad una PA

La raccolta dati

Quasi tutte le informazioni necessarie per costruire una truffa credibile sono pubbliche. Ecco alcune possibili fonti:

- sito istituzionale
- amministrazione trasparente
- delibere
- determine
- gare
- LinkedIn
- comunicati stampa
- webinar registrati

Nei panni dell'attaccante

Cosa vede l'attaccante	Cosa deduce l'IA
Delibera firmata	Chi autorizza
Organigramma	Chi riporta a chi
Gara d'appalto	Fornitori coinvolti
Webinar registrato	Voce del dirigente
LinkedIn	Ruolo e linguaggio

Ho utilizzato solo informazioni pubbliche e ora ho una base di attacco.

La costruzione del bersaglio

L'agente individua:

- chi decide
- chi autorizza
- chi paga
- chi ha accesso

Direttore Generale

↓

Responsabile Acquisti

↓

Responsabile IT

↓

Fornitori

↓

Segreteria

La fiducia artificiale

L'IA produce:

- Mail credibili
- Siti credibili
- Documenti credibili
- Conversazioni credibili
- Voci credibili
- Video credibili

Il problema nasce dalla credibilità complessiva dell'ecosistema creato attorno alla vittima.

Il costo sta crollando

Prima	Oggi
100 persone → 100 mail generiche	100 persone → 100 mail diverse

La personalizzazione non è più costosa.
Questo cambia completamente il modello difensivo.

Come cambia la difesa

Prima: riconoscere il falso

- Errori grammaticali
- Mail sospette
- Telefonate strane

Oggi: verificare il vero

- Callback
- Doppio canale
- Verifica identità
- Procedure
- Segregazione approvazioni

Un cambio storico di paradigma

- Anni ‘90: Proteggiamo i sistemi.
- Anni 2000: Proteggiamo le reti.
- Anni 2010: Proteggiamo gli account.
- Primi anni 2020: Proteggiamo le identità.
- Dal 2026: Proteggiamo la fiducia.

Gli scenari futuri

Oggi

- Email credibili
- Voci sintetiche credibili
- Deepfake convincenti

Domani

- Agenti autonomi persistenti
- Conversazioni di lunga durata
- Identità digitali sintetiche
- Attaccanti virtuali sempre disponibili

Dopodomani

- Impossibile autenticare una persona dalla sola voce
- Impossibile autenticare una persona da una videochiamata
- Impossibile fidarsi di un'identità digitale senza verifiche aggiuntive

Q&A