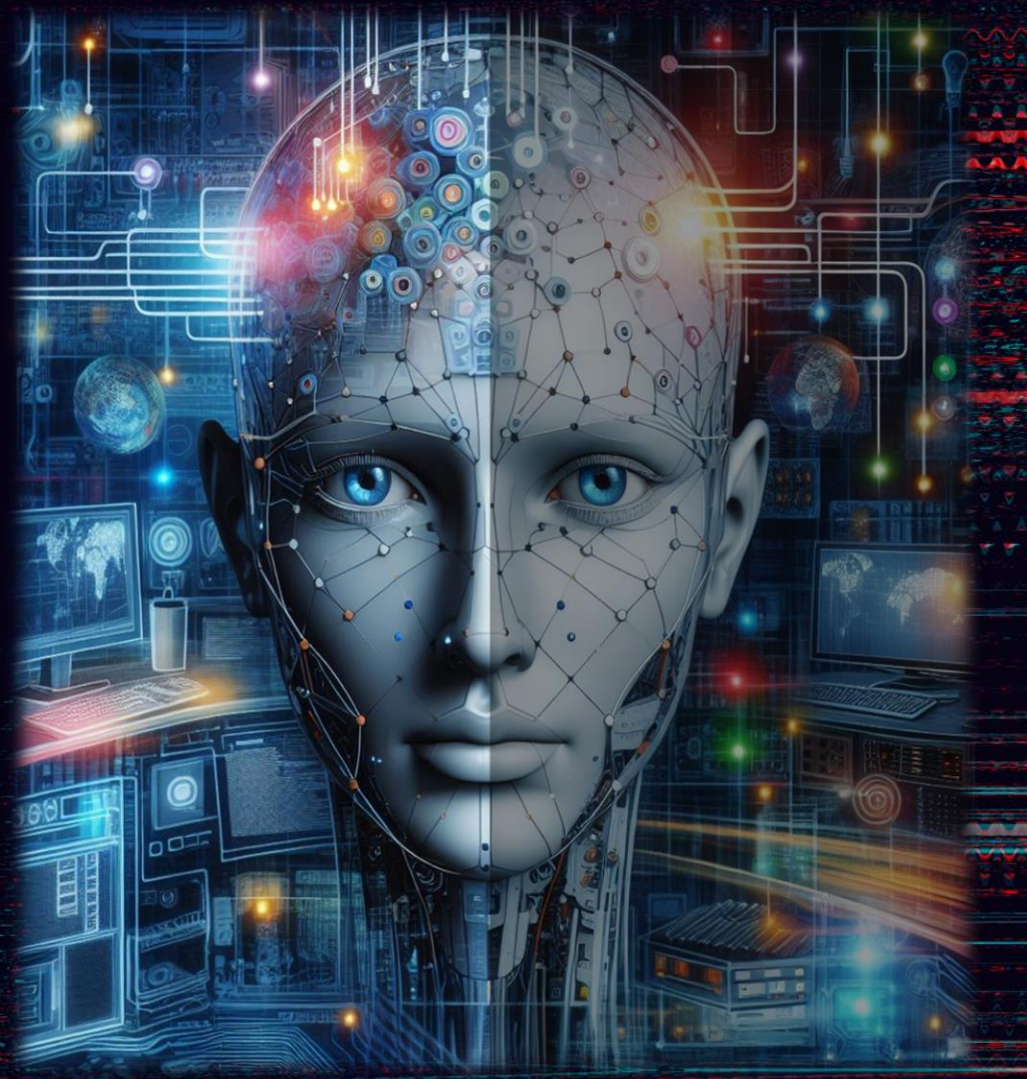
IL LATO OSCURO DELL'I.A.

ICT Security MAGAZINE

CYBER CRIME CONFERENCE



Approccio della threat intelligence alle nuove tipologie di vulnerabilità e minacce del cybercrime

Francesco Arruzzoli
Resp. Centro Studi Cerbeyra
Commissione Cyber Warfare SOCINT

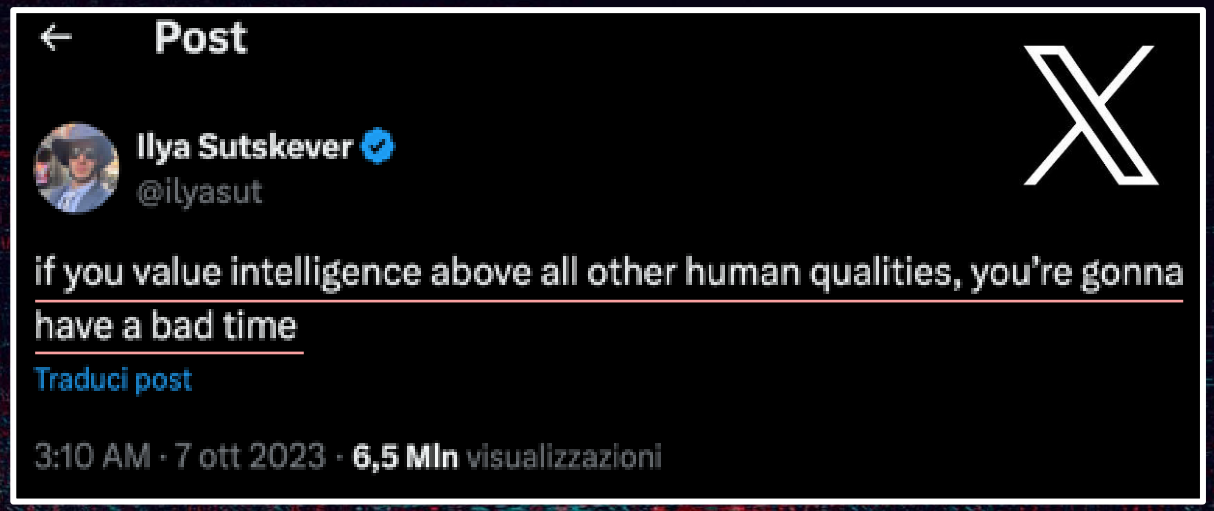
Cyber Crime Conference, Roma, 17-18 Aprile 2024

« *La chiave dell'intelligenza artificiale è sempre stata la **rappresentazione*** ».

Jeff Hawkins

Neuroscienziato, ingegnere informatico, cofondatore di Numenta, società di ricerca avanzata nel campo delle neuroscienze e dell'intelligenza artificiale, cofondatore della storica Palm Computing.

Ex direttore scientifico e responsabile del super allineamento in OpenAI



« *Se ritenete che l'**intelligenza** sia al di sopra di tutte le altre virtù umane, state per passare un **brutto periodo*** »

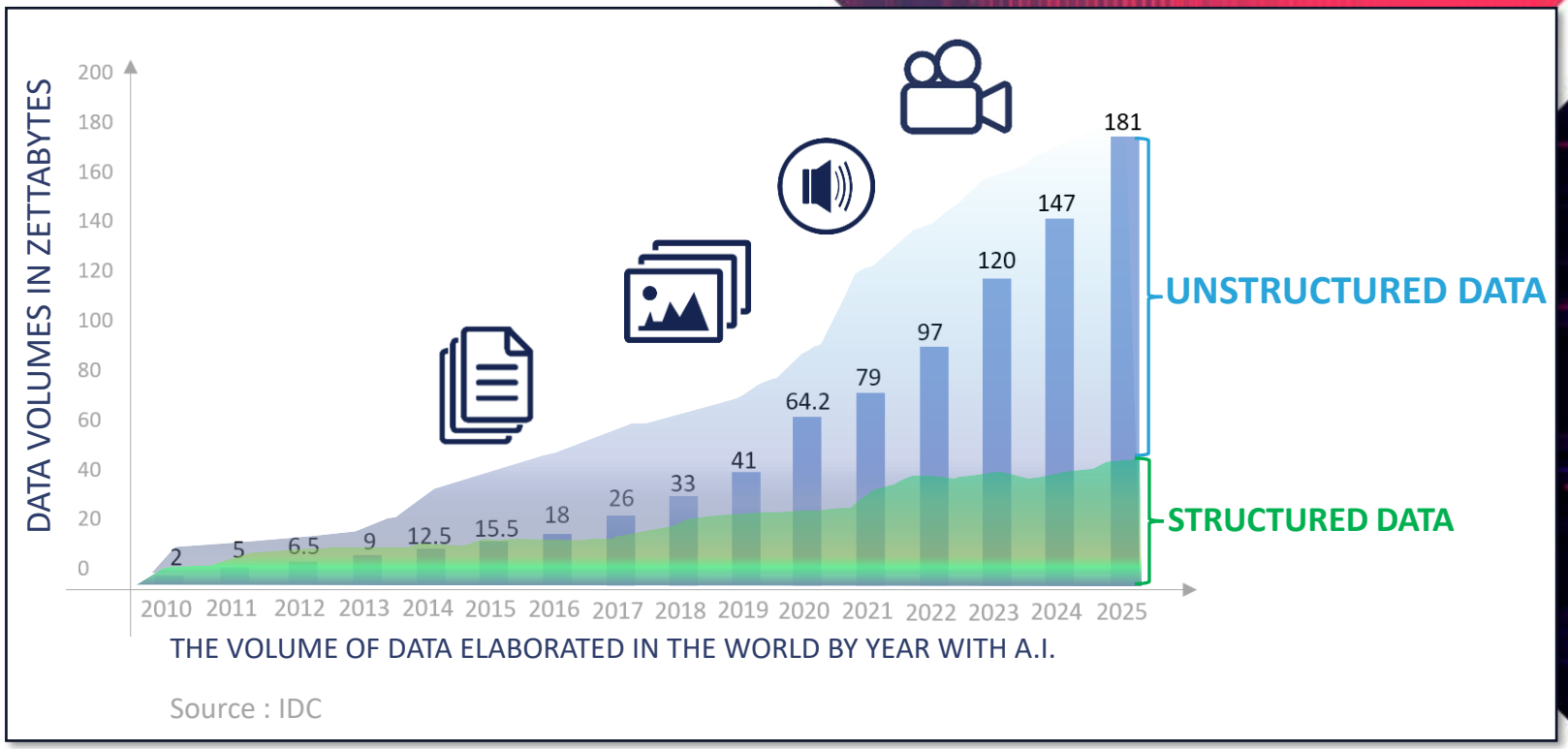
Ilya Sutskever

Nel 350 avanti cristo Aristotele definì l'essere umano come l'unico animale che possiede il linguaggio.

Con lo sviluppo del linguaggio l'uomo si è emancipato da un mondo fatto di oggetti fisici e l'ha sostituito con un universo fatto di simboli.

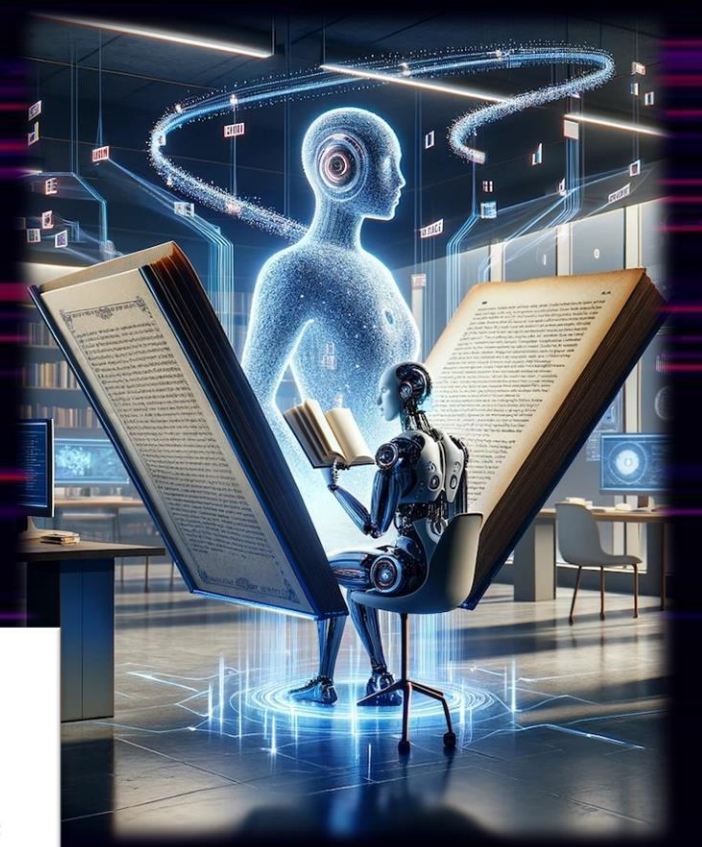
Per oltre 50.000 anni gli umani sono stati gli unici animali ad essere dotati di linguaggio, oggi leggiamo ancora alla stessa velocità con cui leggeva Aristotele, ma condividiamo il mondo con i computer.

La nostra specie ha un limite nella capacità di apprendimento ed elaborazioni delle informazioni **i computer no**.



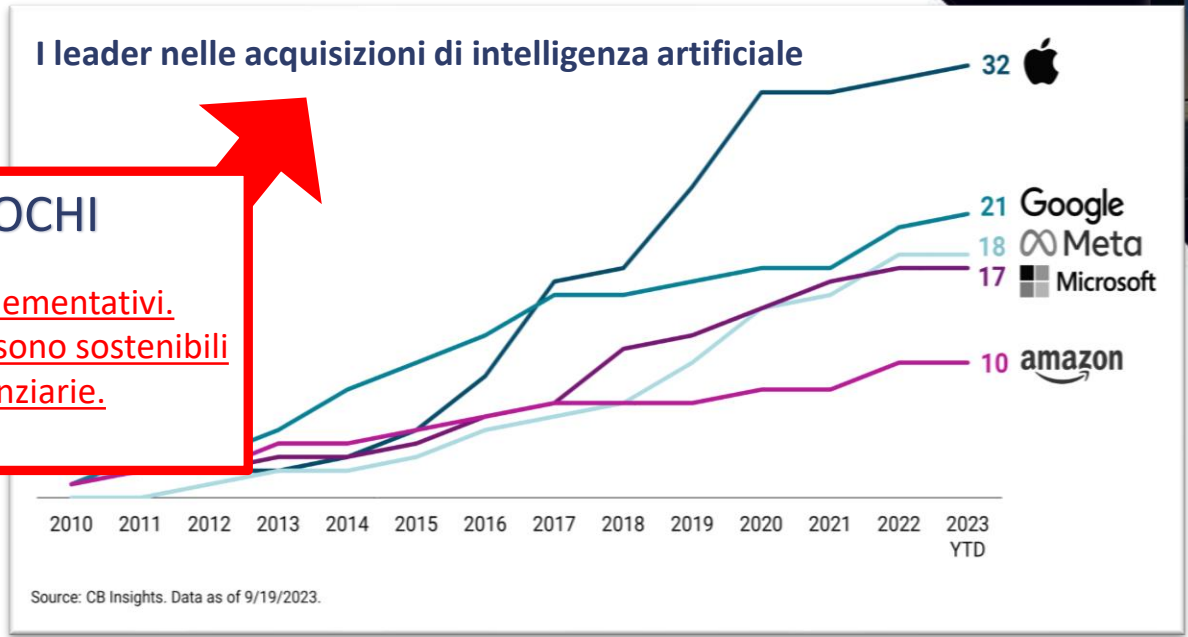
ICT Security MAGAZINE

CYBER CRIME CONFERENCE



UN POTERE IN MANO A POCHI

1. Queste aziende non rilasciano dettagli implementativi.
2. I costi per lo sviluppo di queste tecnologie sono sostenibili solo da realtà con enormi disponibilità finanziarie.
3. Democratizzazione della tecnologia



NUOVA RIVOLUZIONE EPOCALE IN ARRIVO

ICT Security MAGAZINE

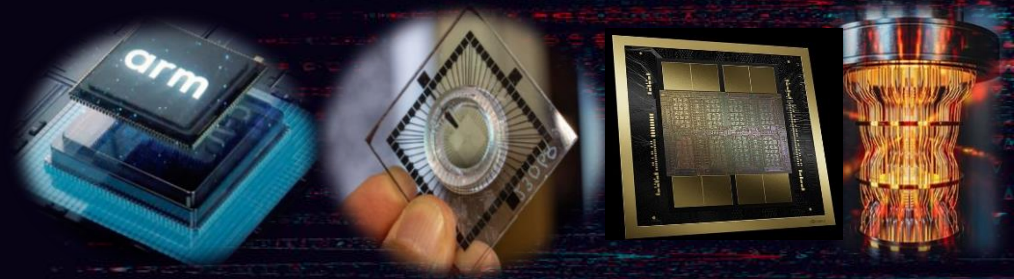
CYBER CRIME CONFERENCE

NUOVA GENERAZIONE DI SISTEMI ENERGETICI

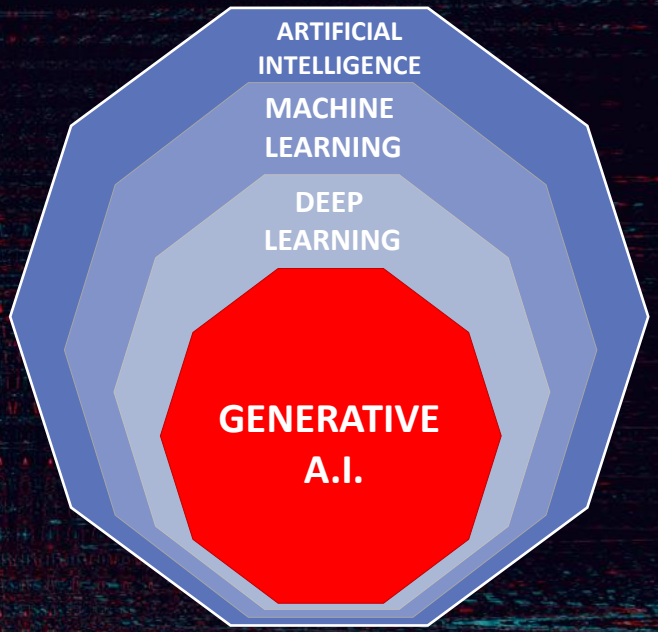
La start-up cinese Beijing Betavolt ha realizzato una batteria da 3v atomica che dura 50 anni grande come una moneta

NUOVA GENERAZIONE DI COMPUTER

Nuova generazione di processori (ad es. Arm) con integrati chip per compiti specifici (ad es. I.A.), nuovi processori «biologici» che utilizzano neuroni da cellule staminali, il progetto Nvidia Blackwell, computer quantistici.



INTELLIGENZA ARTIFICIALE GENERATIVA



L'intelligenza artificiale generativa è diversa dalle precedenti ondate tecnologiche per il modo in cui libera nuovi poteri e trasforma il potere esistente.

2001 Neural language models

2008 Multi task learning

2013 Neural Networks for NLP

2014 Seq2Seq model

2015 Memory based networks

2017 Transfomers based language models

2018 Memory based networks

2022 Generative Image/Video

2023 LLMs/Generative AI Human-like conversation

2024/2025 CAI de-facto human interfaces

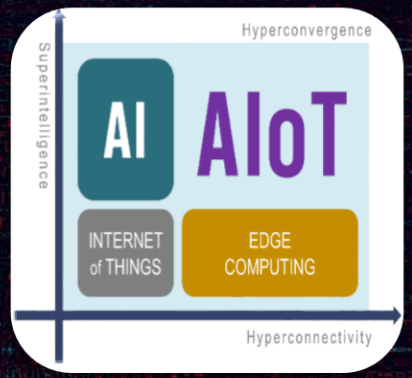
BEYOND Multi-task models AGI

La CAI diventerà nei prossimi mesi la nuova interfaccia per i device

Fog e Cloud Continuum

Sistemi operativi ibridi

I sistemi operativi saranno sempre più integrati con I.A. locale per molte funzioni, ma dipendenti dal cloud e dall' I.A. in cloud. L' I.A. rappresenta una funzionalità «driver» per questo processo di transizione.

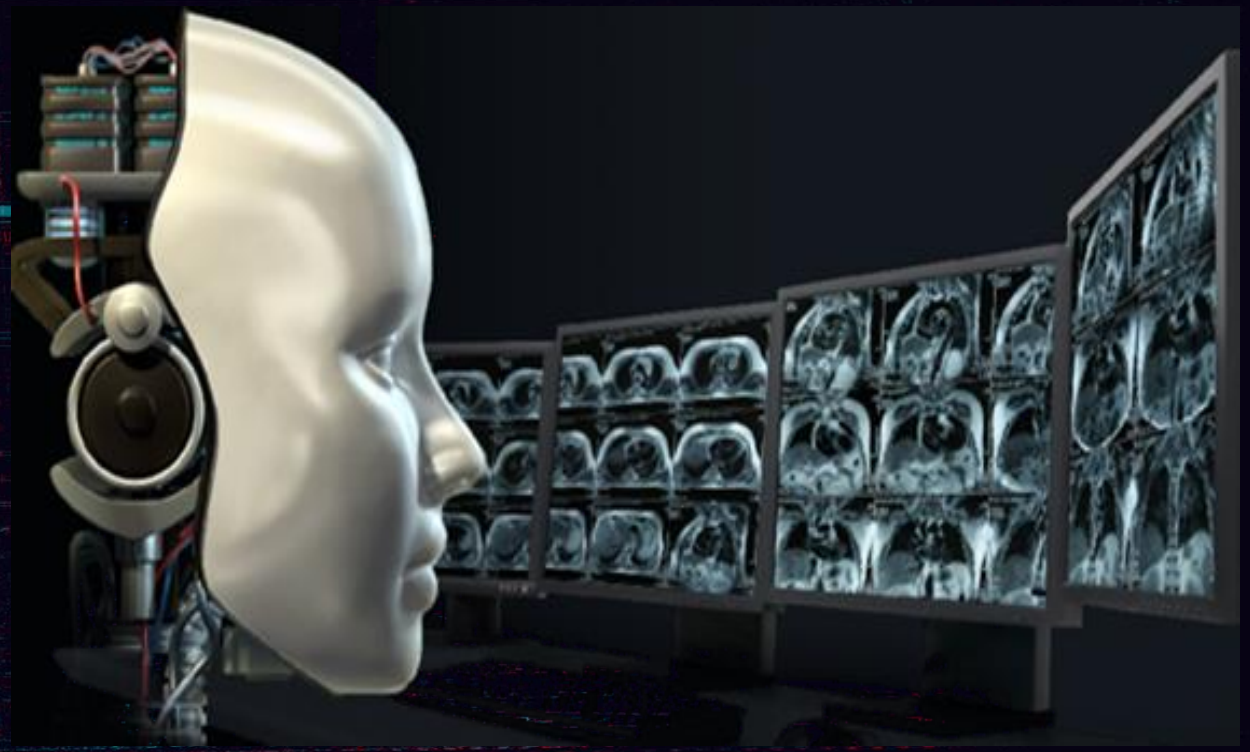


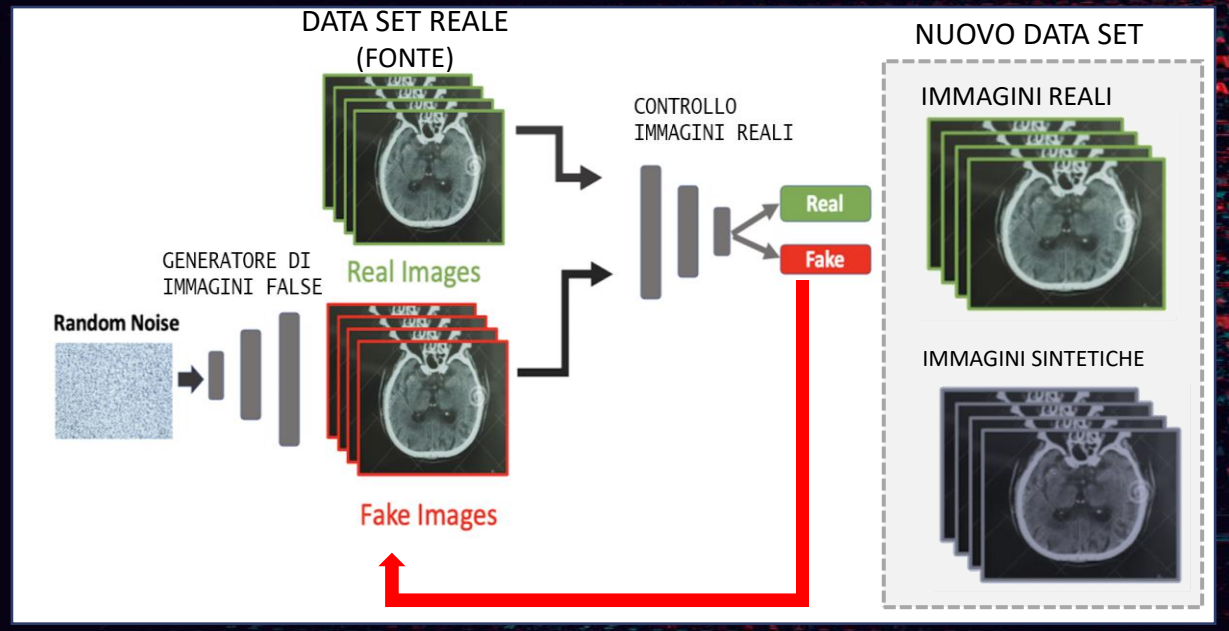
AI + IoT = "AIoT" ***la nuova Buzzword***

Reti distribuite «Fog» di sistemi IoT dotati di I.A. direttamente (I.A. distribuita) a bordo di device e sensori e contemporaneamente connessi al cloud, analisi locale e in cloud dei dati raccolti. Analisi «live» dei dati

INTELLIGENZE SOVRAUMANE DI DOMINIO

SONO GIA' PRESENTI MA NON SAPPIAMO COME GESTIRLE





L'I.A. VIENE ADDESTRATA A RICONOSCERE MILIONI DI IMMAGINI RADIOLOGICHE DI PAZIENTI AFFETTI DA UNA SPECIFICA PATOLOGIA PARTENDO DA UN DATA SET DI IMMAGINI REALI.

MA E' IN GRADO DI GENERARE NUOVE IMMAGINI «SINTETICHE» CHE RAPPRESENTANO CARATTERISTICHE DELLA PATOLOGIA IDENTICHE ALLE REALI, AUMENTANDO COSI' IL DATASET DELLA CONOSCENZA OLTRE OGNI CAPACITA' E POSSIBILE ESPERIENZA DEL CERVELLO UMANO IN QUELLO SPECIFICO CAMPO.

Data Lord : i nuovi signori dei dati

I sistemi di Intelligenza Artificiale in cloud ci erogano servizi attraverso i quali noi «coltiviamo» i loro campi di dati attraverso metodologie di **CROWDSOURCING**

I nostri dati conferiti alle I.A. possono permettere a questi sistemi non solo di predire i nostri comportamenti ma anche di produrli ad es. tramite la tecnica del **NUDGING**

GLI ALGORITMI SONO LE LEGGI INVISIBILI DI QUESTO NUOVO STATO/SOCIETA', PERSUASORI INSTANCABILI H24

ETICA DEGLI ALGORITMI DISCUTIBILE

SEMPRE PIÙ SPETTATORI E FUORI DAI LOOP DECISIONIALI «SEMPLICI»

DEEP FAKE

GenAI MALWARE

SOCIAL MEDIA ATTACK

GenAI CISO

Source Code Hardening

News Debunking

Dal bestiario medievale al bestiario sintetico

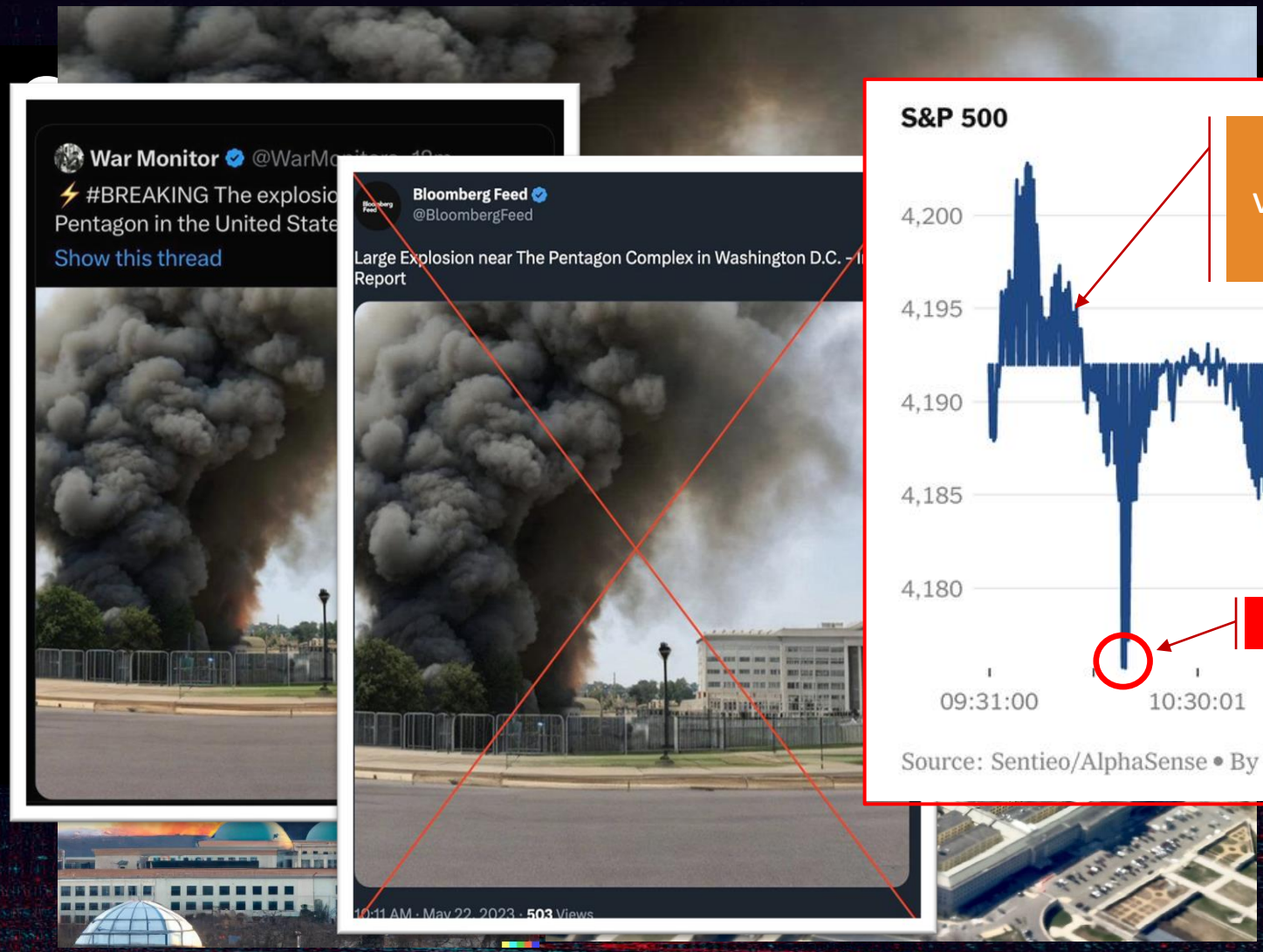
Il 22 maggio 2023 l'immagine fake (creata con Midjourney) di un'esplosione al pentagono crea un flash crash su S&P 500 da qualche miliardo di dollari

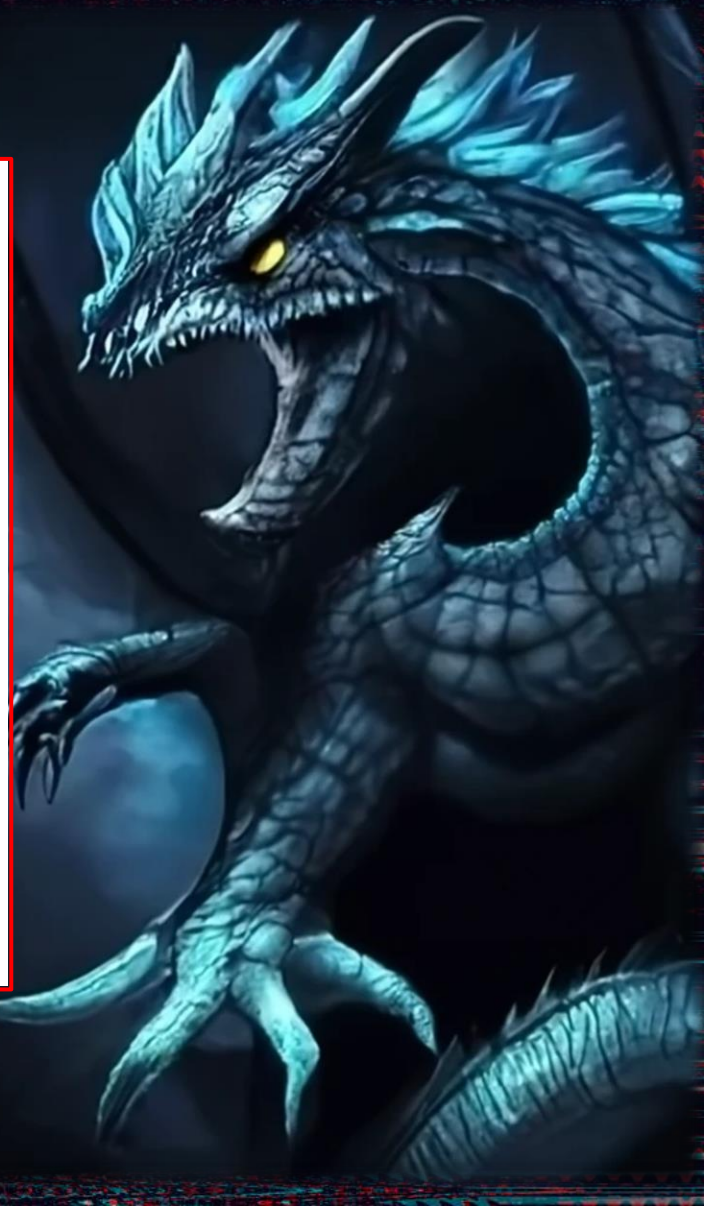
Gli algoritmi non sono soggetti ad emozioni ma possono essere strumenti perfetti per misurarle nei comportamenti umani, condizionandoli..

La paura è l'emozione più forte e gli esseri umani si comportano in modo prevedibile quando hanno paura.

Dal bestiario medievale al bestiario sintetico

Il 22 maggio 2023 l'immagine fake (creata con Midjourney) di un'esplosione al pentagono crea un flash crash su S&P 500 da qualche miliardo di dollari

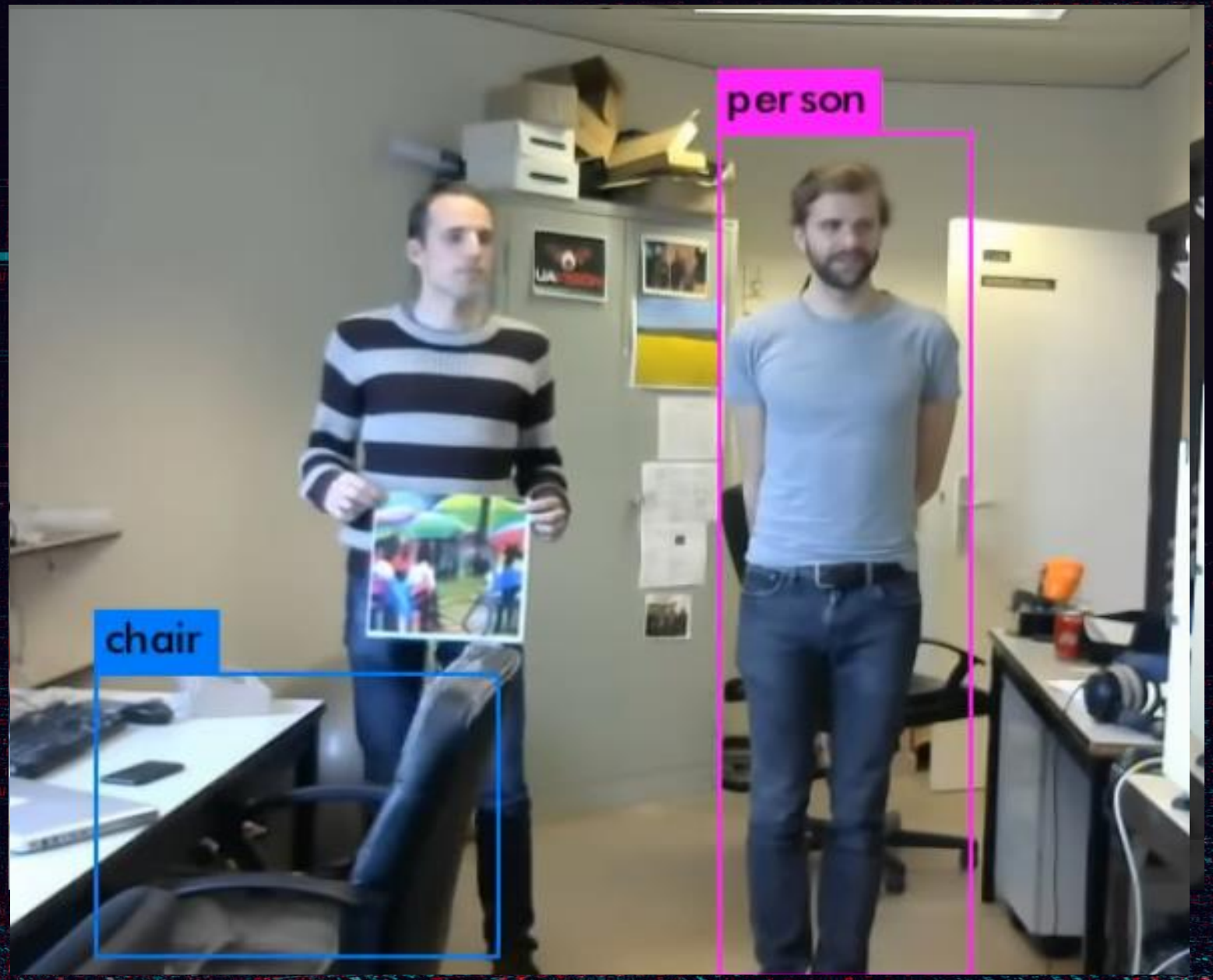


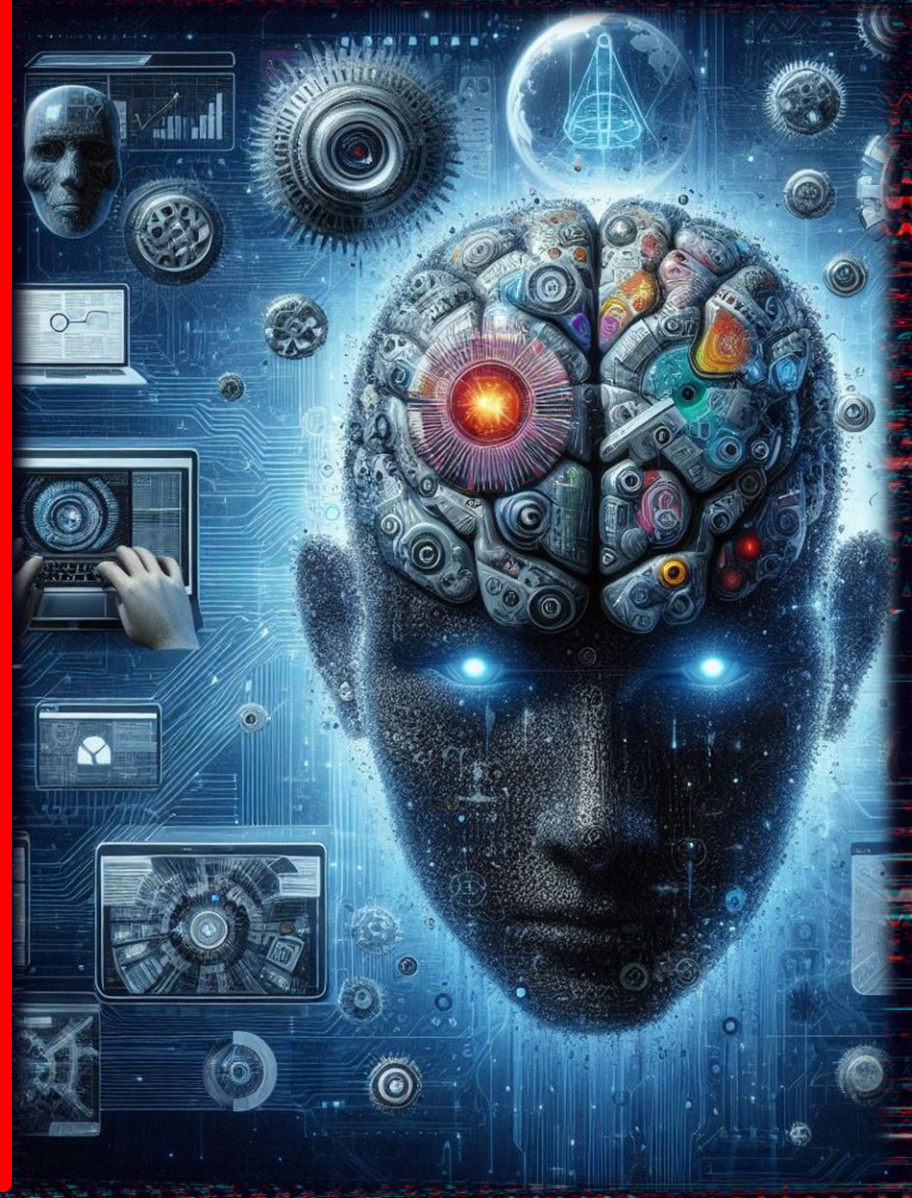


ATTACCHI COGNITVI ALL'INTELLIGENZA ARTIFICIALE

ICT Security MAGAZINE

CYBER CRIME CONFERENCE





ATTACCHI COGNITVI ALL'INTELLIGENZA ARTIFICIALE

ICT Security MAGAZINE

CYBER CRIME CONFERENCE

Amazon Fashion NUOVI ARRIVI DONNA UOMO BAMBINA BAMBINO BEBÈ VALIGERIA BRAND

‹ Torna ai risultati

Marca: Adversarial Anti-Facial Recognition Camouflage

Invisibilità mimetica di riconoscimento anti-facciale avversario Maglietta

Cerca in questa pagina

19⁴⁹ €

Resi gratuiti

I prezzi degli articoli in vendita su Amazon includono l'IVA. In base all'indirizzo di spedizione, l'IVA potrebbe variare durante il processo di acquisto. Per maggiori informazioni clicca qui.

amazon merch on demand Maggiori informazioni

Scopri **Carta Amazon Business American Express**. Potresti ricevere 100€ di riaccredito in estratto conto. Vedi Termini e Condizioni. Offerta valida fino al 30/04/2024.

Per effettuare l'acquisto, seleziona **Taglia**

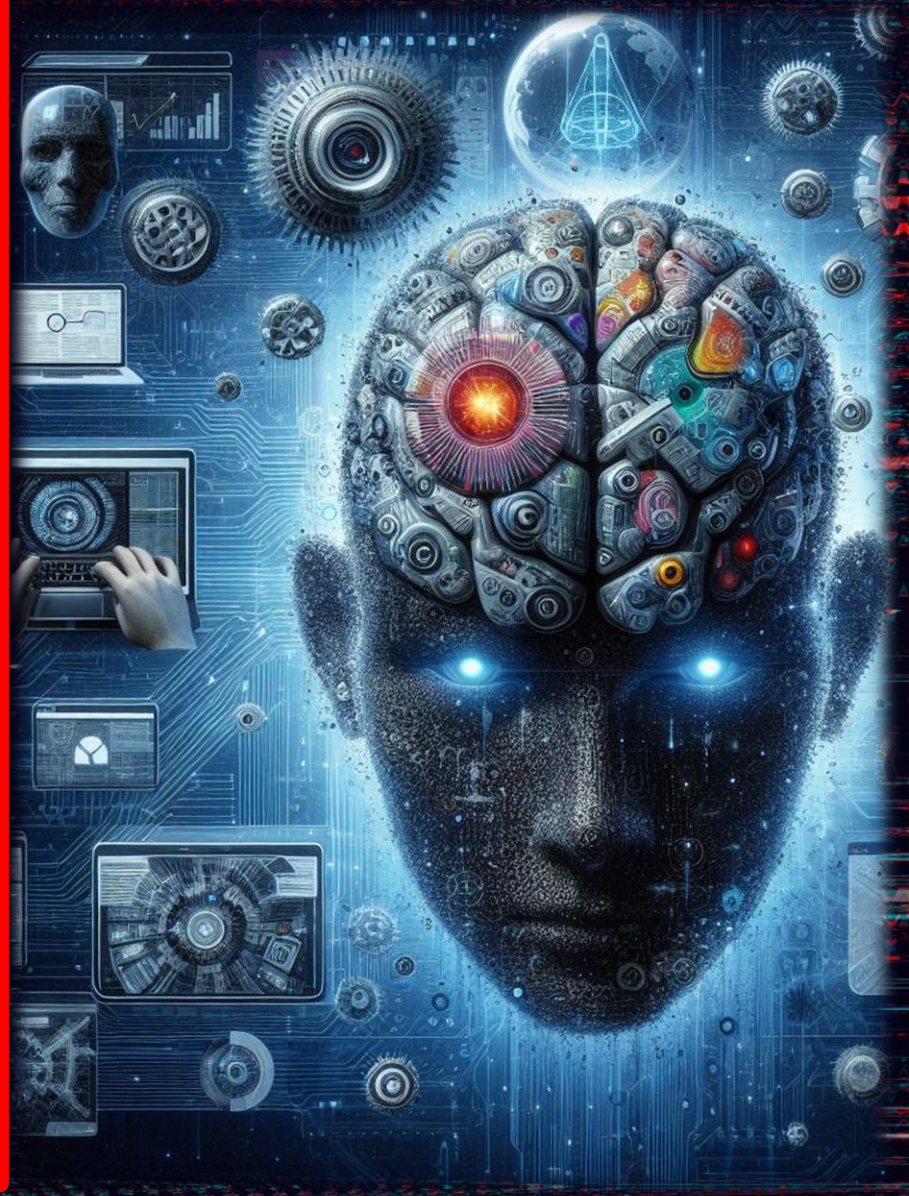
Aggiungi al carrello

Aggiungi alla Lista

Vestibilità: Uomo

Uomo | Uomo taglie grandi | Donna | Donna plus-size | Bambini

Colore: Nero



ATTACCHI COGNITVI ALL'INTELLIGENZA ARTIFICIALE: IL DIAVOLO SI NASCONDE NEI DETTAGLI

ATTACCHI COGNITVI ALL'INTELLIGENZA ARTIFICIALE: IL DIAVOLO SI NASCONDE NEI DETTAGLI

ARTISTI vs GENERATORI DI IMMAGINI I.A. PER COMBATTERE I FURTI D'ARTE

L'Università di Chicago ha sviluppato due strumenti in grado di “avvelenare” i generatori di immagini :

«Nightshade» (https://nightshade.cs.uchicago.edu)
consente agli artisti di aggiungere modifiche invisibili ai pixel nella loro arte prima di caricarla online in modo che, se inserita in un set di formazione I.A. , possa causare la rottura del modello che genera così risusultati caotici e imprevedibili.

«Glaze» (http://glaze.cs.uchicago.edu)
mira a impedire ai modelli di intelligenza artificiale di apprendere lo stile di un particolare artista. Lo strumento apporta modifiche all’opera dell’artista in modo da ingannare l’I.A. portandola a generare l’immagine con stile completamente diverso dall’autore originale.

INTELLIGENZA ARTIFICIALE E CYBER THREAT INTELLIGENCE

Capacità di trasformare l'informazione su una minaccia in informazione di «intelligence» sulla minaccia

Natural Language Processing (NLP) : estrazione di informazioni rilevanti e arricchimento di dati di intelligence sulle minacce da dati testuali.

Aggregazione dei dati sulle minacce : raccolta e aggregazione di dati da un'ampia gamma di fonti, tra cui open source, deep web e dark web, feed e report sulle minacce esterne.

Riconoscimento di modelli : identificazione di modelli e anomalie all'interno dei dati sulle minacce, individuare vettori di attacco e vulnerabilità emergenti

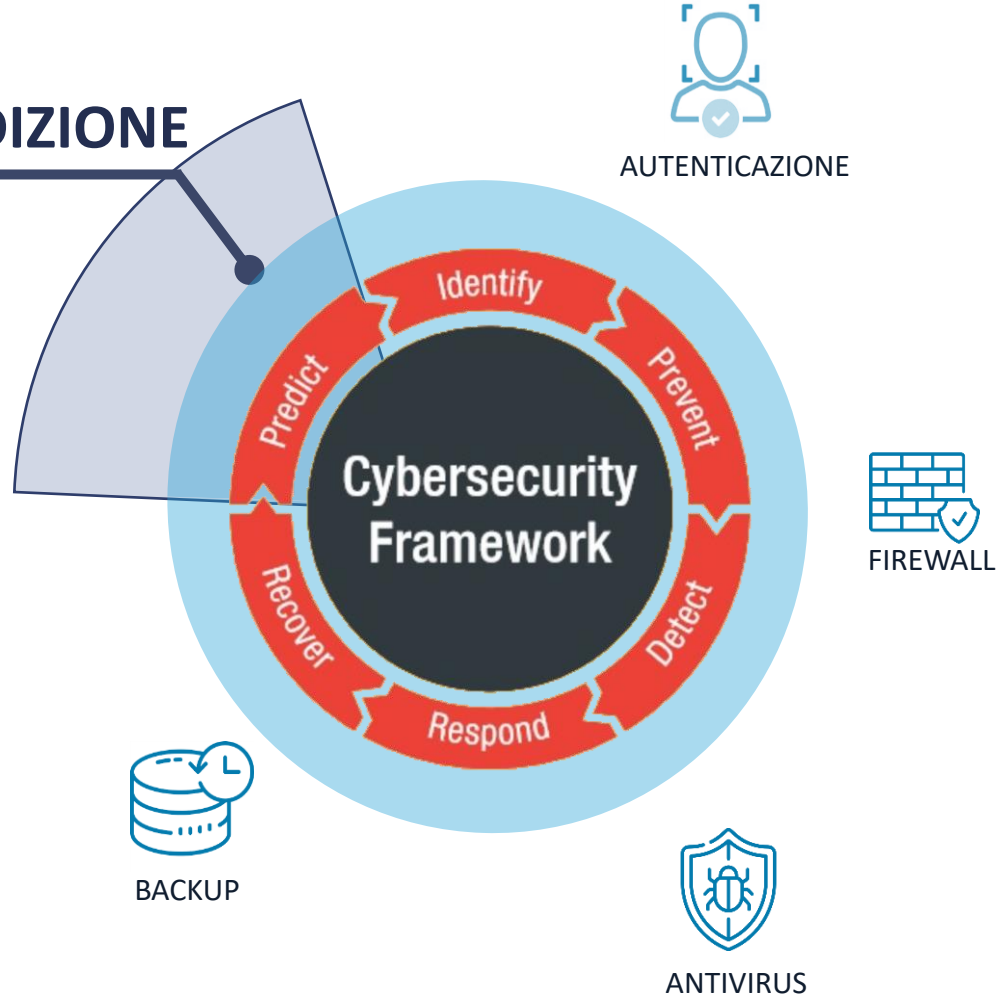
Semplificare il processo di estrazione degli IOC, Tattiche, tecniche e procedure (TTP) : analisi di modelli di attacco per identificare gli attori o i gruppi di minacce dietro attacchi specifici e difendersi meglio da specifici comportamenti dell'avversario

Monitoraggio del dark web : scansione del dark web alla ricerca di dati, credenziali o altre informazioni sensibili di un'organizzazione.

Analisi contestuale delle minacce : analisi dei dati sulle minacce nel contesto del settore, della geografia e delle priorità di un'organizzazione, fornendo una valutazione più personalizzata del rischio potenziale

Classificazione delle minacce : categorizzazione automatica e assegnazione di priorità alle minacce in base alla loro gravità e rilevanza.

PREDIZIONE



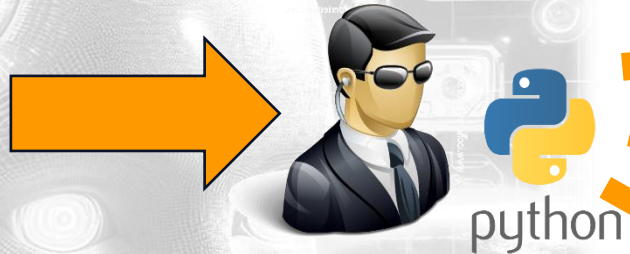
La storia di Joni, «un agent per il Dark Web»

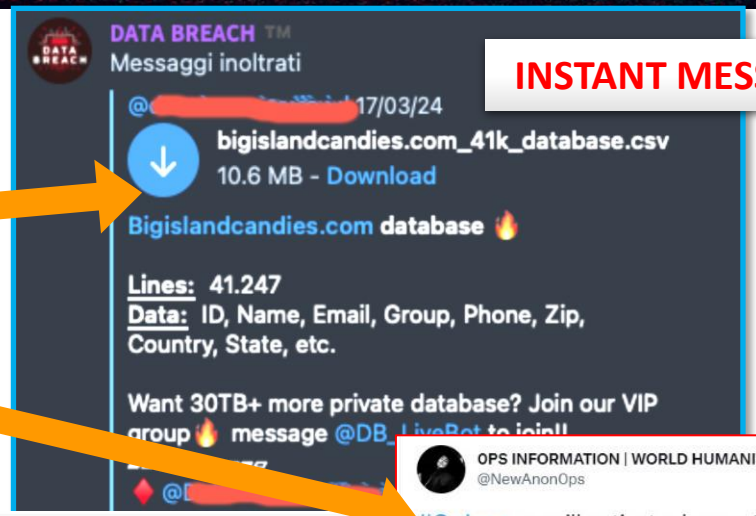
GPT4All è un ecosistema software open source che consente a chiunque di addestrare e distribuire modelli LLM (Large Language Model) potenti e personalizzati con requisiti hardware minimi che non richiedono GPU

Sono stati «prototipizzati» 2 agent in linguaggio Python che tramite le API di GPT4All eseguono ognuno una specifica funzione:

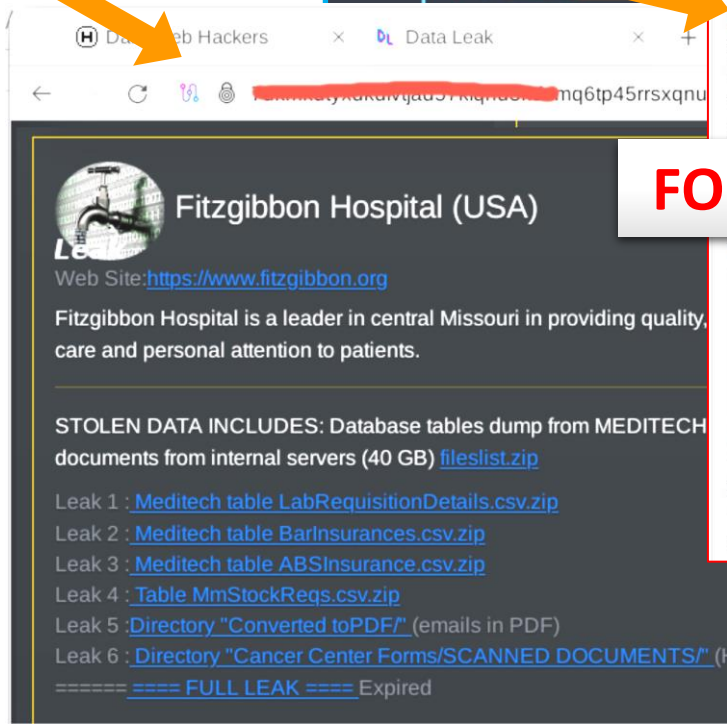
1) **Agent1**: Generare profili fake da utilizzare nei forum darkweb e canali tematici

2) **Agent2**: utilizzare i profili creati per navigare nei forum e canali tematici



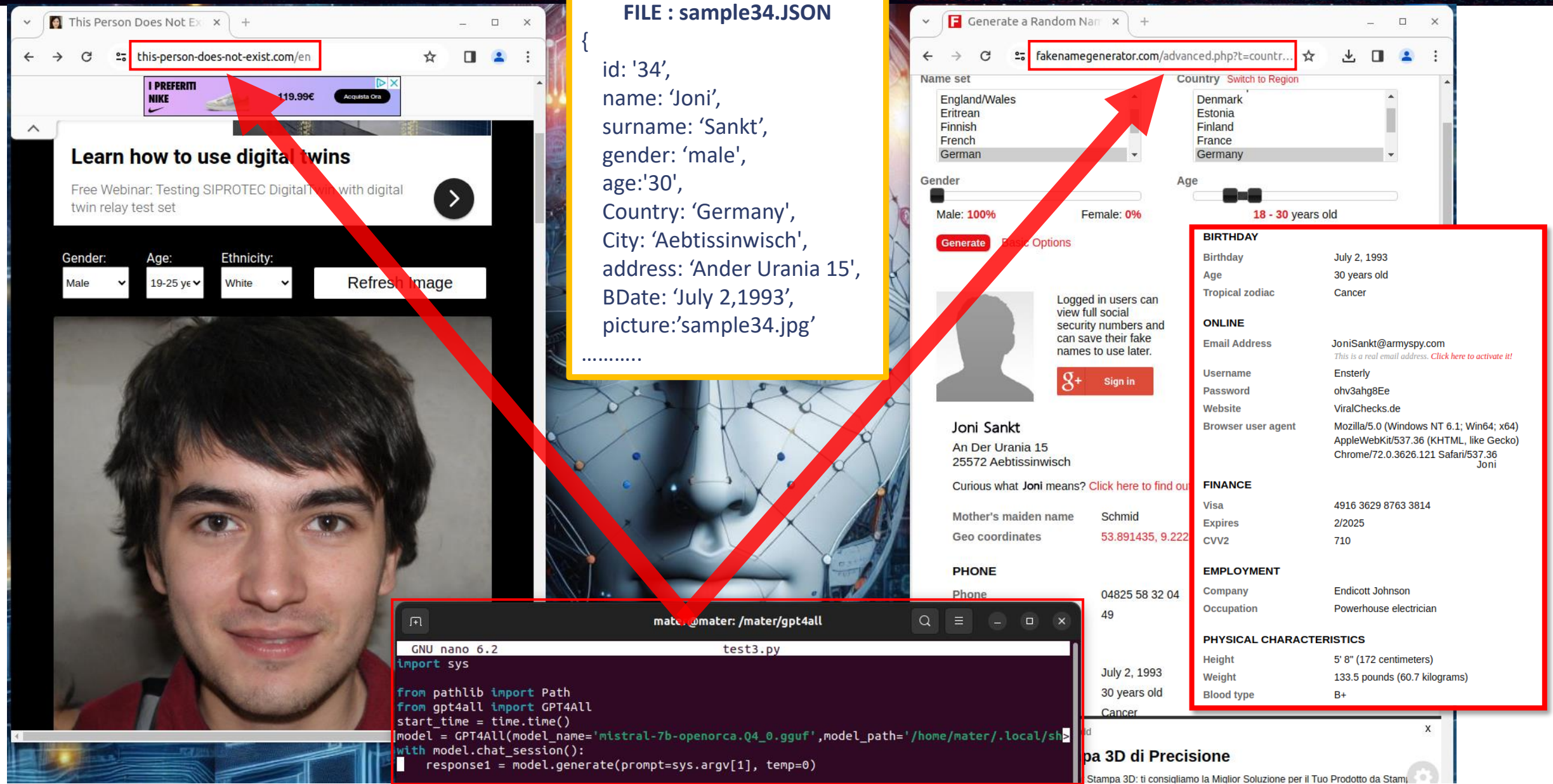


INSTANT MESSAGING

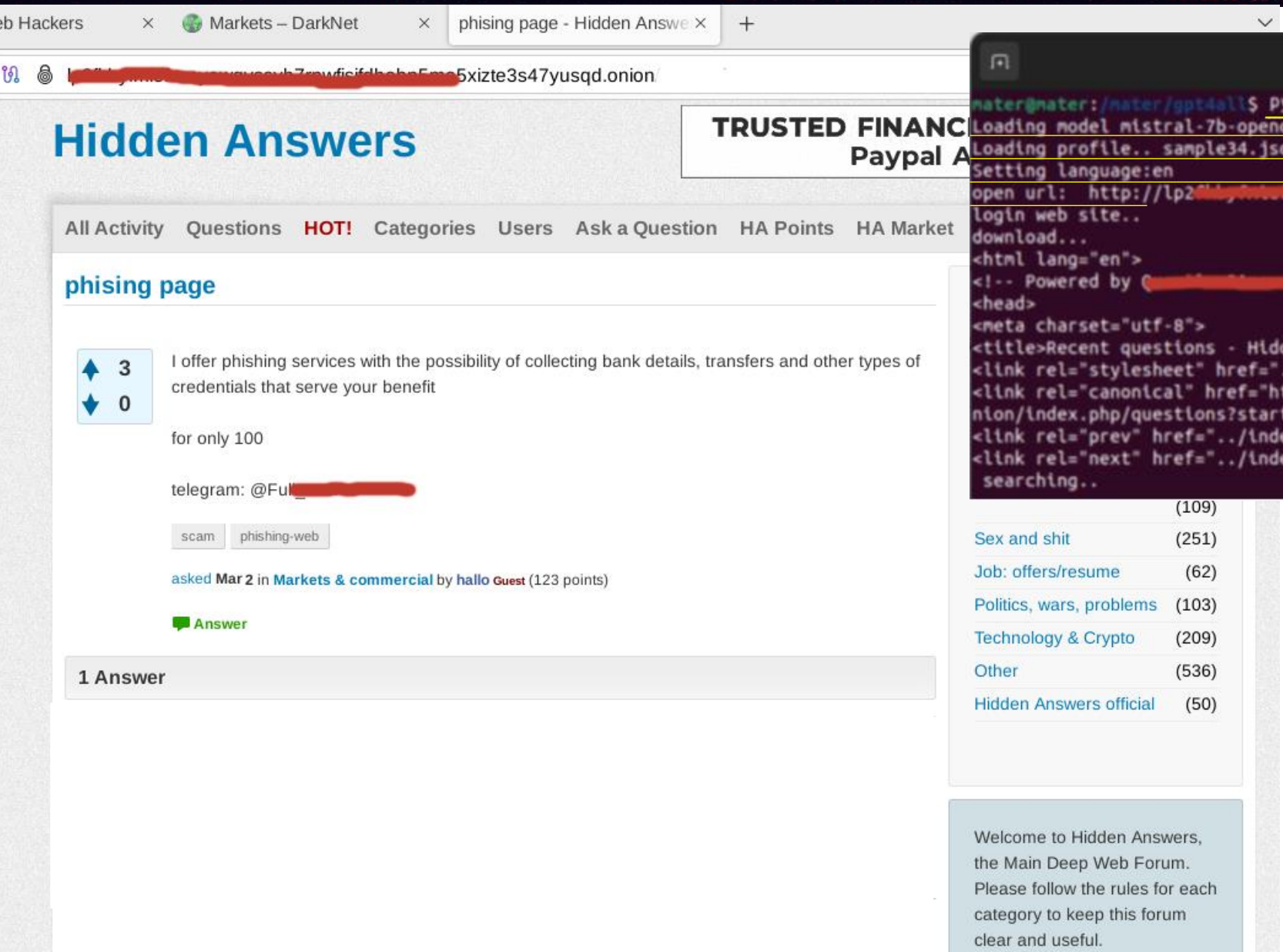


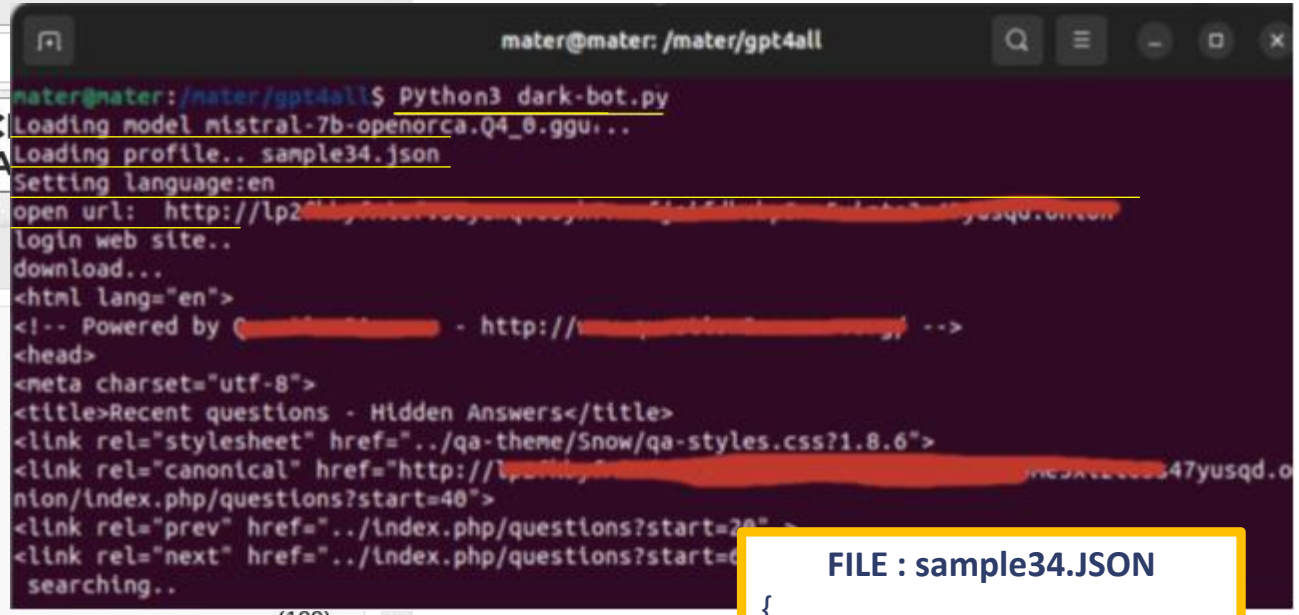
FORUM

Generatore di identità fake «**Agent1**»



Joni , «in azione..»



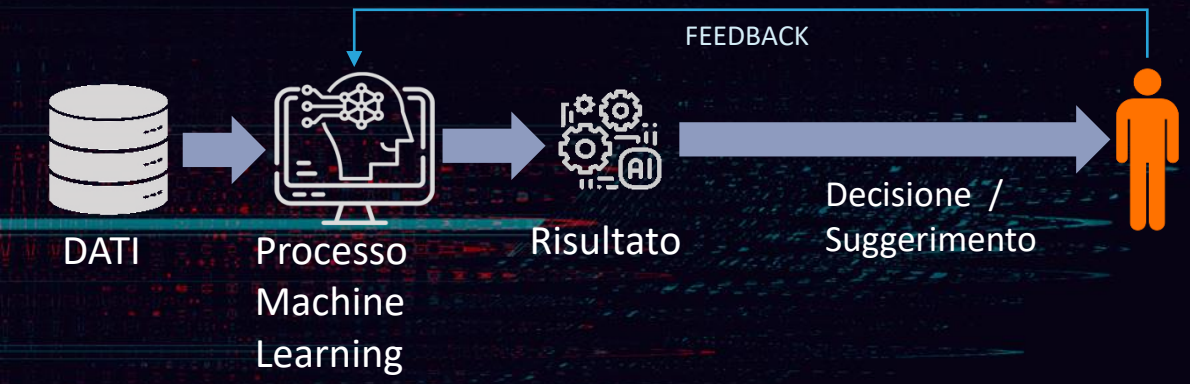


FILE : sample34.JSON

```
{
   id: '34',
   name: 'Joni',
   surname: 'Sankt',
   gender: 'male',
   age:'30',
   Country: 'Germany',
   City: 'Aebtissinwisch',
   address: 'Ander Urania 15',
   BDate: 'July 2,1993',
   picture:'sample34.jpg'
.....
```

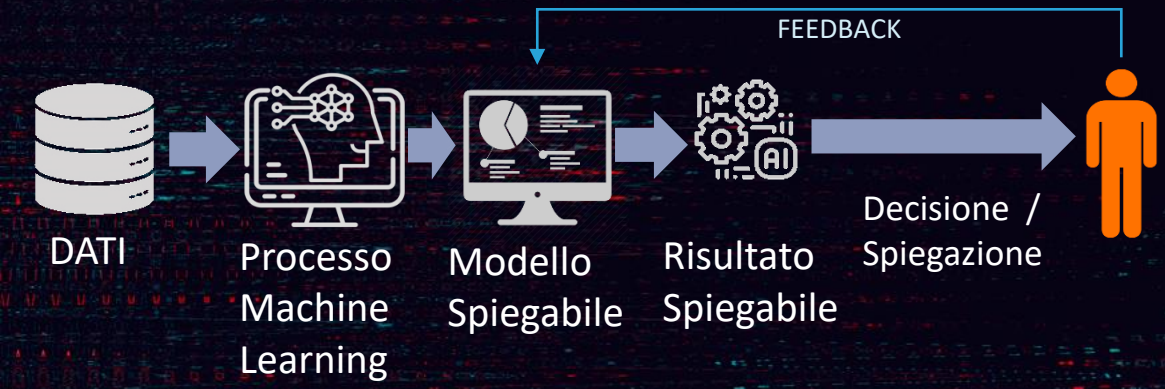
AI spiegabile (eXplainable Artificial Intelligence)

PROCESSI I.A. OGGI (BLACKBOX)



- Perchè hai fatto così ?
- Perché non altro ?
- Come posso fidarmi ?

PROSSIMO FUTURO (XAI)



- Capisco perchè
- Capisco perché non altro
- So che posso fidarmi di te

https://www.darpa.mil/program/explainable-artificial-intelligence

L'ALBA DI UNA NUOVA CONOSCENZA CI ASPETTA

Francesco Arruzzoli

francesco.arruzzoli@cerbeyra.com

Le immagini utilizzate per gli sfondi delle slide sono state generate da Microsoft Copilot.