



Private AI nella Digital Forensics

Come la **Private AI** può aiutare nelle indagini digitali, senza compromettere controllo, riservatezza e validità legale delle evidenze.

DI NANNI BASSETTI

[NANNIBASSETTI.COM](https://nannibasetti.com)

Chi sono



Mi occupo di digital forensics dal 2005/2006, project manager di **CAINE** (distro usata in tutto il mondo)

Fondatore di **CFI** (Computer Forensics Italy)

Scrittore di parecchi articoli scientifici e software free/open source per la D.F.

- Membro fondatore e Segretario di **ONIF** (Osservatorio Nazionale Informatica Forense).

Coinvolto in casi di rilevanza nazionale.

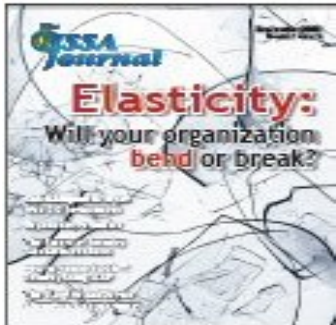
Chi sono

2005/2006



primi vagiti online
sulla computer forensics

2007



Tanti articoli
dal 2007 ad oggi

2008



Creazione di CFI
e pubblicazione di
SFDumper

2009 - 2014



Classificato
nella DC3 Challenge

2009

Scripts4CF

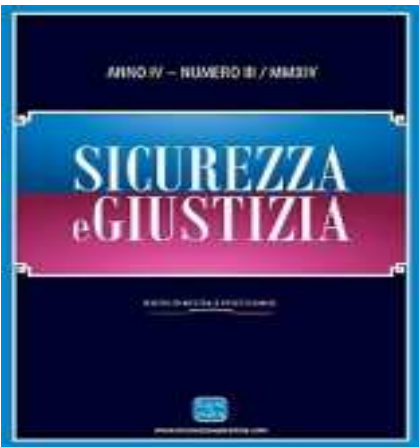
Alcuni miei scripts
per la forensics

2009



Sviluppo CAINE
Gnu/Linux

2013



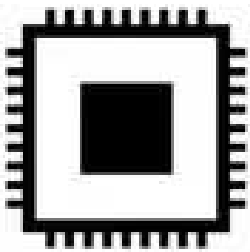
Nel comitato di
redazione di S&G

2015



Segretario e
membro
fondatore ONIF

2020



GLOBAL
PROFESSIONALS
FOR ARTIFICIAL INTELLIGENCE

Socio onorario GP4AI

CASI “MEDIATICI”



E TANTA FORMAZIONE



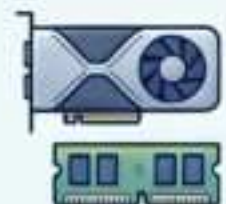
Private AI: Il Futuro Sicuro della Digital Forensics

IL PROBLEMA



L'uso dell'IA in cloud espone dati sensibili a rischi di privacy e violazioni, con dati che attraversano confini esterni.

STATISTICHE



REQUISITI HARDWARE OTTIMIZZATI

Grazie alla quantizzazione (4-bit), bastano 32-64GB di RAM e una GPU moderna per operare.

Efficienza Quantizzazione (Modello 20B)

Formato	Bit per parametro	RAM richiesta (appross.)
FP32	32 bit	80 GB
INT8	8 bit	20 GB
INT4	4 bit	10 GB

LA SOLUZIONE: PRIVATE AI

PERCHÉ PASSARE ALLA PRIVATE AI?



Controllo Assoluto del Dato

Le evidenze digitali rimangono confinate nell'ambiente air-gapped o on-premise dell'organizzazione.



Compliance e Sovranità

Garantisce il rispetto del GDPR e del segreto istruttorio eliminando i "trust boundary" esterni.



LOCALE

VS

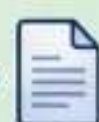


CLOUD

Cloud vs. Locale

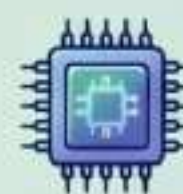
L'AI locale elimina la dipendenza da API di terze parti e il rischio di logging esterno.

TECNOLOGIE E STRUMENTI OPERATIVI



RAG (Retrieval-Augmented Generation)

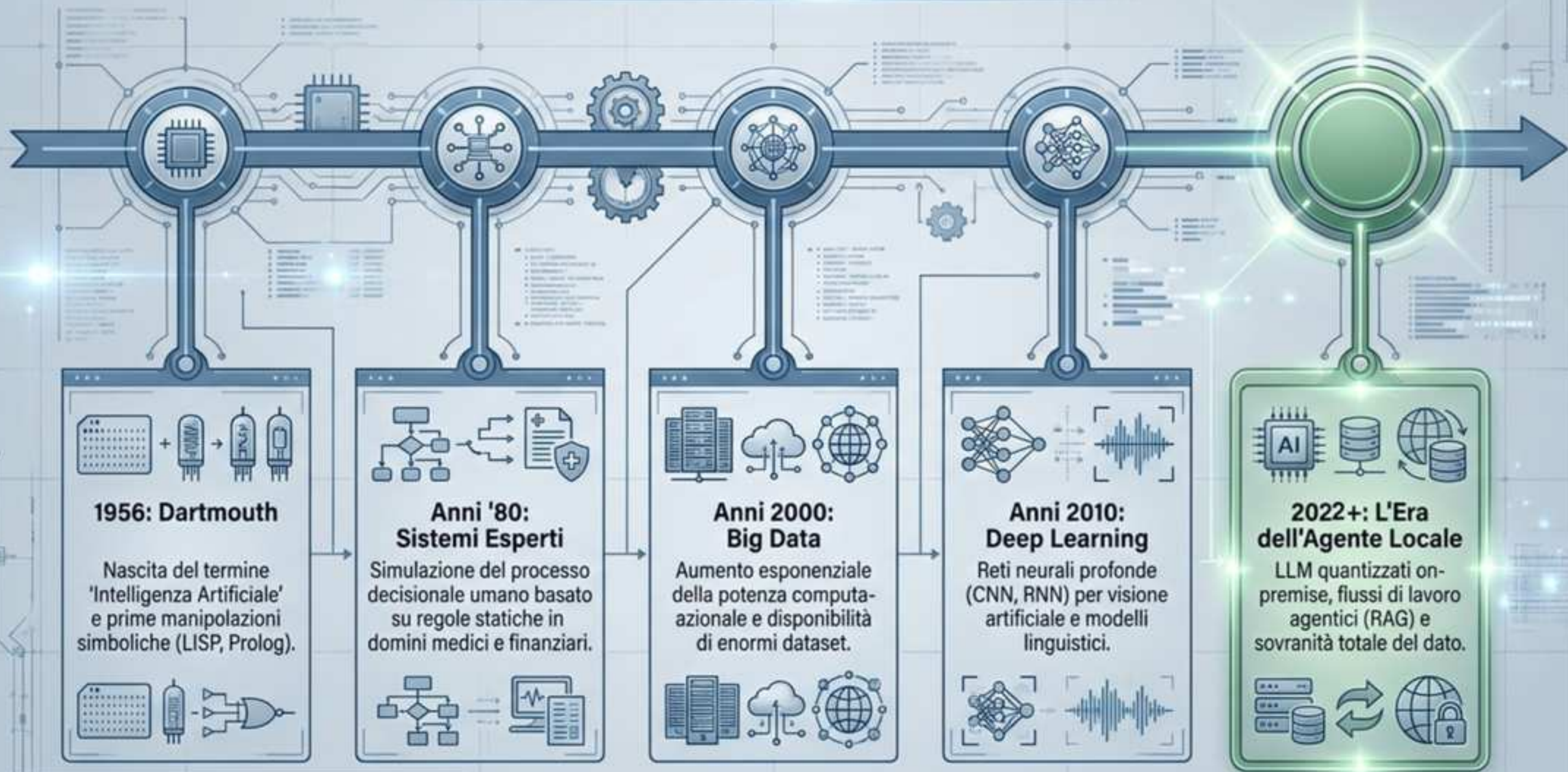
Framework che permette all'IA di "dialogare" con le prove (email, log, file) senza riaddestramento.

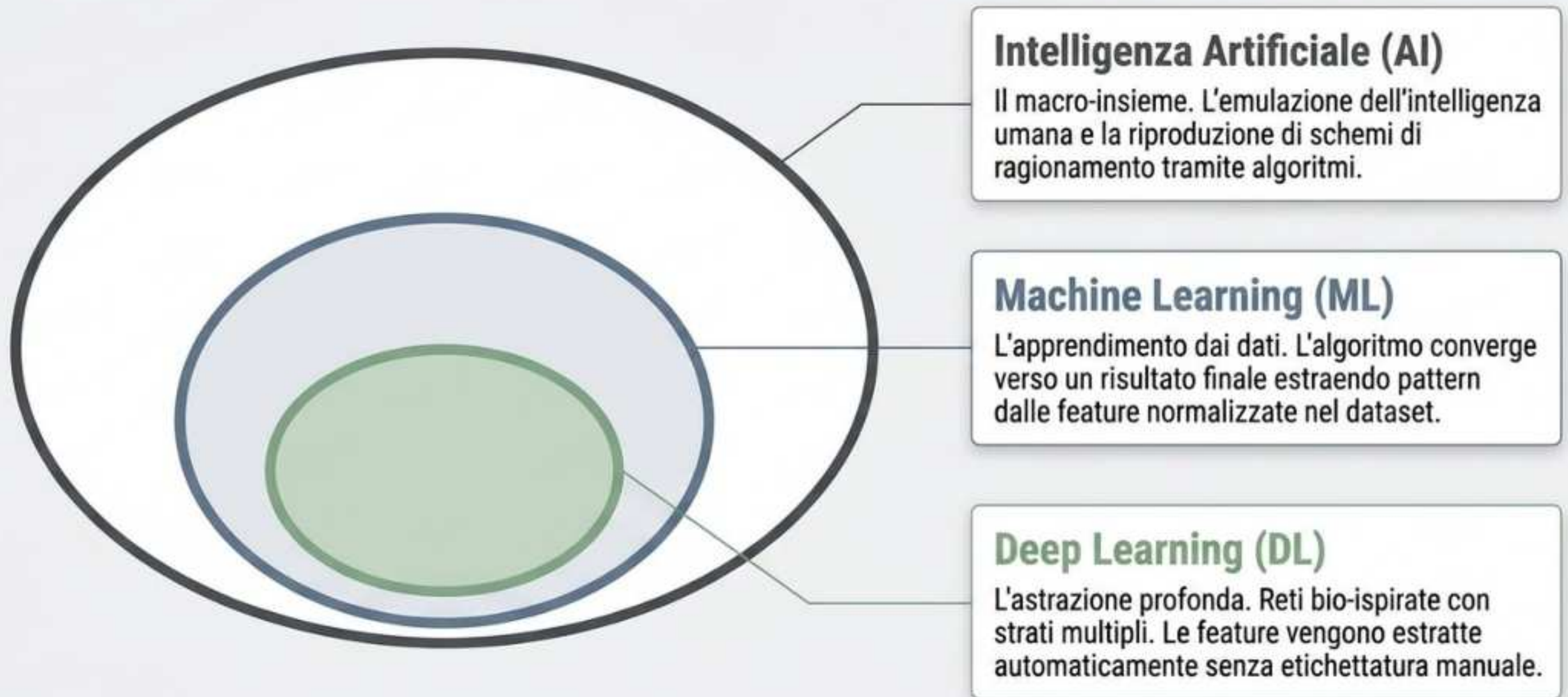


Strumenti d'Avanguardia

Utilizzo di agenti come Claude Code (tramite Ollama) e framework specializzati come NBMultiRag.

TIMELINE OF AUTONOMY





Algoritmo vs. Modello: Non sono la Stessa Cosa!

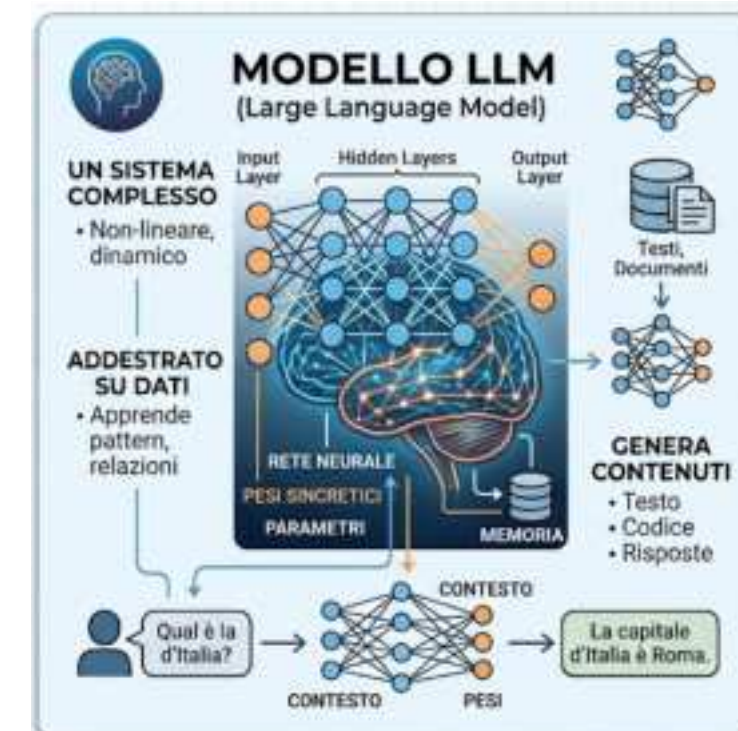
Algoritmo

Procedura che il computer segue
La "ricetta" per preparare un piatto



Modello

Prodotto finale dell'addestramento
La "pietanza finita" pronta per previsioni
Contiene pattern, relazioni e parametri



NUMERO DI PARAMETRI

Un modello di linguaggio con 100 miliardi di parametri significa che il modello stesso ha 100 miliardi di pesi o coefficienti che vengono utilizzati per effettuare previsioni linguistiche.

Questi parametri sono utilizzati per regolare il comportamento del modello e per apprendere le relazioni tra le parole e le frasi.

Requisiti Hardware e Quantizzazione

Una Private AI non richiede infrastrutture proibitive. La **quantizzazione** (4-bit o 8-bit) rende praticabile l'uso di LLM locali anche su workstation forensi.

Formato	Bit/parametro	Modello 20B	RAM richiesta
FP32	32 bit	20B × 4 byte	80 GB
FP16/BF16	16 bit	20B × 2 byte	40 GB
INT8	8 bit	20B × 1 byte	20 GB
INT4	4 bit	20B × 0,5 byte	10 GB

Configurazione minima: CPU 8–16 core, RAM 32–64 GB, SSD NVMe 1–2 TB, GPU opzionale (8–12 GB VRAM per modelli 7B). Per ridurre allucinazioni: temperatura bassa (0.1–0.2) e Top-K contenuto.

Il mito del supercomputer milionario



1 La magia della Quantizzazione

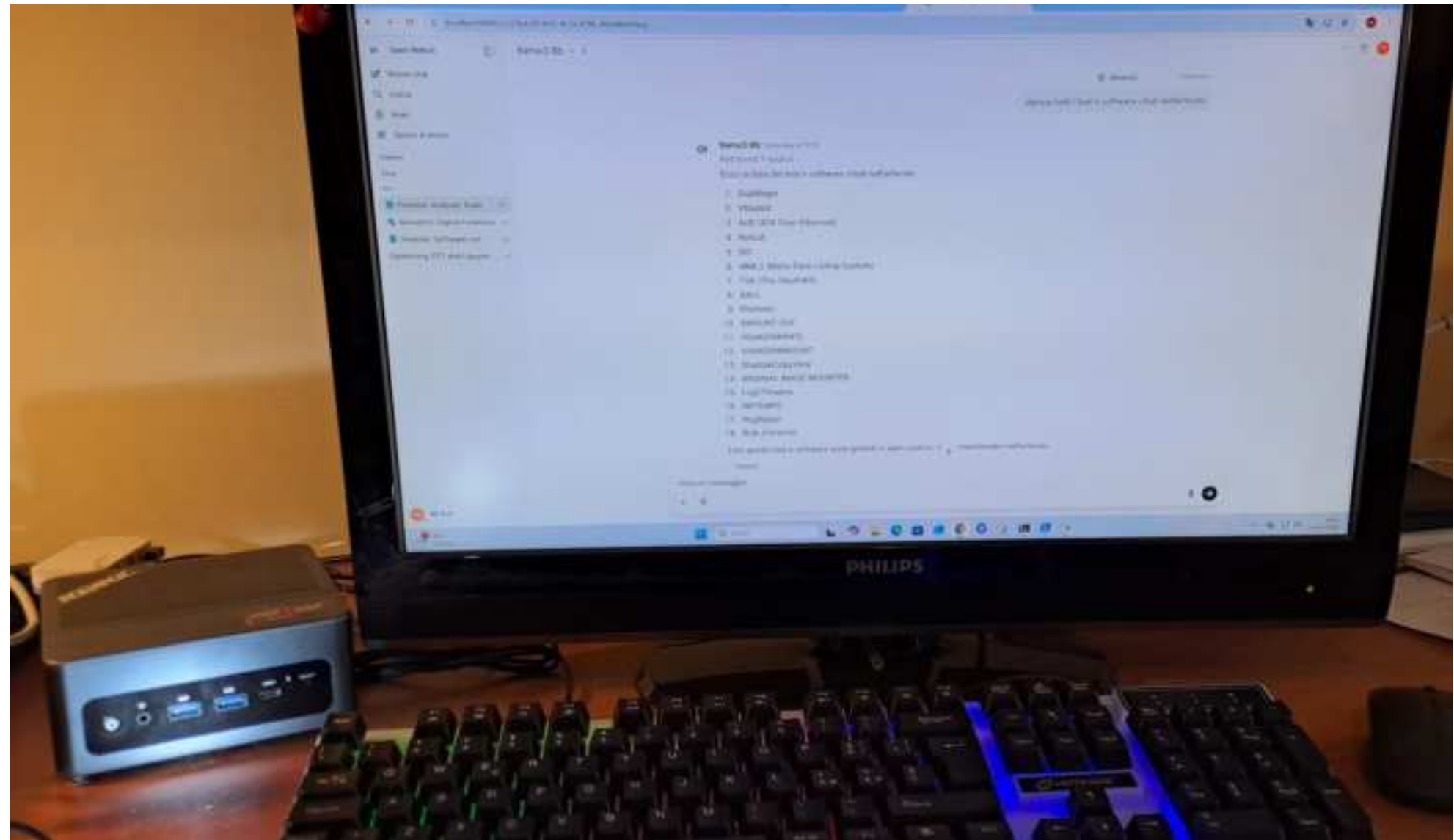
I cervelli digitali possono essere compressi senza perdere la loro intelligenza.

2 Numeri alla mano

Un modello da 20 miliardi di parametri passa da 80 GB a soli 10 GB di memoria richiesta.

3 Hardware accessibile

Bastano un buon computer da ufficio (32-64 GB di RAM) e una scheda grafica standard. Niente server farm.



POSTAZIONE PER AI PRIVATA – UN MINIPC CON 64Gb di RAM, 2Tb NVME disco fisso, AMD Ryzen 9 HX 370 AI + AMD Radeon 890m come GPU.

IA APPLICATA ALLA DIGITAL FORENSICS

ACQUISIZIONE DATI INTELLIGENTE

Ingestione dati assistita dall'IA



ANALISI AUTOMATIZZATA DI LOG E ARTEFATTI

Processa massicci log file, rileva rapidamente nuovi formati



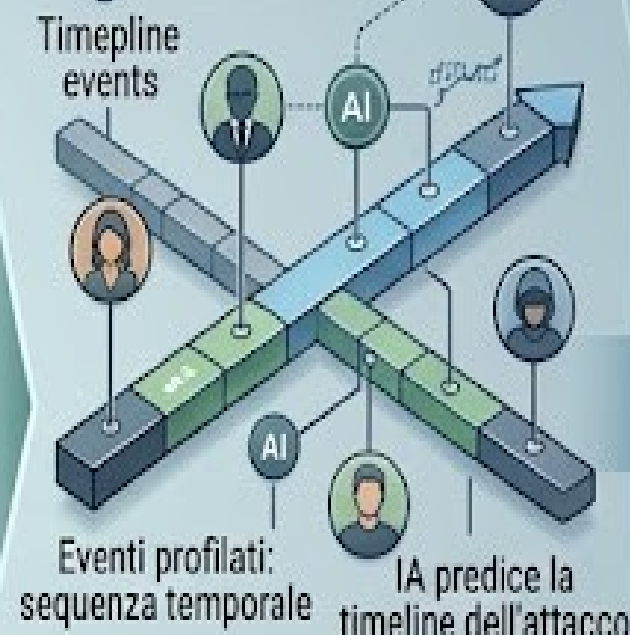
TRIAGE E SCOPERTA DI PROVE

Potente funzione di ricerca: elaborazione del linguaggio naturale



FORENSICS PREDITTIVA E ATTRIBUZIONE

IA predice un attacco: scenario di attacco



Hacker Profile



Digital Footprints

REPORTISTICA E PRESENTAZIONE

CONCLUSIONI GENERATE DALL'IA



Summary sintetizzato automaticamente al termine



L'IA cambia il mondo, ma a quale prezzo?



La Luce

- Maggiore **efficienza** e **risparmio di tempo** (es. Sanità, PA, Industria).
- Decisioni **oggettive** basate sui **dati**.



L'Ombra

- Il rischio della **Scatola Nera**: non sappiamo come la macchina decide.
- Problemi di **conformità GDPR**, **privacy** e **bias cognitivi**.

L'innovazione esige la Explainable AI per illuminare la Black Box

I Pilastri Etici

- **Beneficenza:** Generare valore reale per l'uomo.
- **Non maleficenza:** Evitare i danni sistemici.
- **Autonomia:** Garantire il rispetto delle scelte individuali.
- **Giustizia:** Assicurare equità e inclusione sociale.



I Rischi della Black Box

- **Bias:** Dati distorti generano inevitabilmente decisioni distorte.
- **Discriminazione:** Es. riconoscimento facciale e selezione del personale.
- **Privacy:** Criticità nella gestione dei dati secondo le direttive GDPR.
- **Disoccupazione:** Rischio tecnologico e necessità di reskilling.

La Spiegabilità e Controllabilità (XAI) è il ponte essenziale per generare fiducia nei sistemi intelligenti.

IL CLOUD LUCI ED OMBRE



Un abbonamento infinito: Usare sistemi cloud significa pagare un pedaggio continuo per ogni singola domanda o operazione, perdendo la sovranità dei propri dati.

La soluzione trasparente: leggere la ricetta



Gli Ingredienti

- ✓ I Dati.
- ✓ Gli ingredienti grezzi su cui il modello ha studiato. Spesso tenuti segreti.



La Cucina

Strumenti di Training.
Come i dati vengono filtrati, preparati e cucinati.



La Torta

Il Modello Finale. I Pesì.
Il cervello pronto all'uso.

Un'IA è davvero **trasparente (Open Source)** solo se possiamo ispezionare tutti e tre i passaggi.



- Where
- All 250
 - Local 237**
 - Cloud 17

- Capabilities
- Tools 96**
 - Thinking 46**
 - Vision 46**
 - Embedding 13
 - Decision 8

- Download size
- Any**
 - Up to 8 GB
 - Up to 16 GB
 - Up to 32 GB
 - Up to 64 GB

Models

Search 250 models

9 of 250 models

qwen3.8

Qwen3.8 delivers substantial gains across coding, professional work, research, and long-horizon agentic tas...

vision tools thinking 27b

muse-glimmer

Meta's latest open model built for always-on local agents. 30B parameters, licensed under Apache 2.0 and r...

vision tools thinking 30b

qwen3.8-flash-next

This experimental preview of the architecture that will unde...

vision tools thinking

qwen3.8

3.9M Downloads Updated 2 weeks ago

Qwen3.8 delivers substantial gains across coding, professional work, research, and long-horizon agentic tasks.

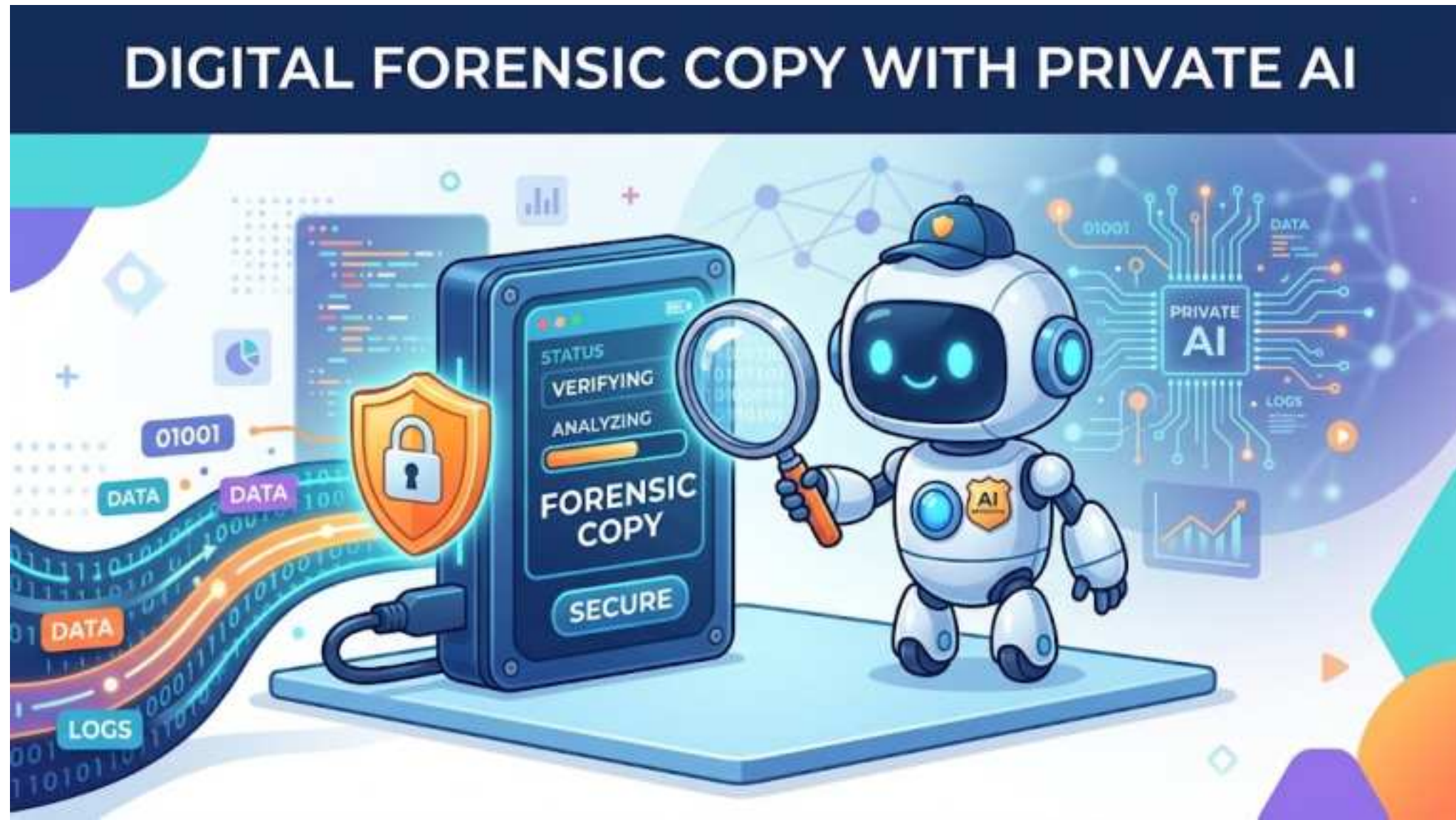
vision tools thinking 27b

CLI cURL Python JavaScript

```
ollama run qwen3.8
```

Copy icon

COPIA FORENSE CON AI PRIVATA



MACHINE LEARNING PER L'IDENTIFICAZIONE DEI FILE SOSPETTI

Mi sono ispirato ad un paper del 2019 ad opera di Xiaoyu Du e Mark Scanlon

“Methodology for the Automated Metadata-Based Classification of Incriminating Digital Forensic Artefacts”

[https://www.researchgate.net/publication/](https://www.researchgate.net/publication/334192020_Methodology_for_the_Automated_Metadata-Based_Classification_of_Incriminating_Digital_Forensic_Artefacts)

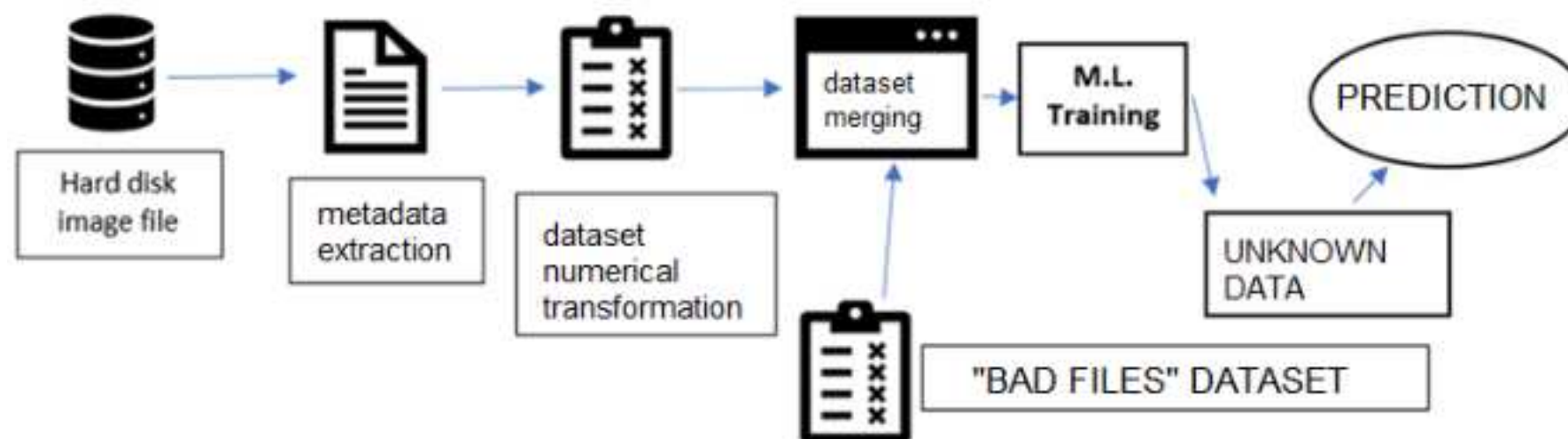
[334192020_Methodology_for_the_Automated_Metadata-Based_Classification_of_Incriminating_Digital_Forensic_Artefacts](https://www.researchgate.net/publication/334192020_Methodology_for_the_Automated_Metadata-Based_Classification_of_Incriminating_Digital_Forensic_Artefacts)

Cercherò di rendere più “*tangibile*” e visuale il concetto espresso nel paper e nella premessa dell’articolo, ossia quello di classificare/identificare un file dannoso tramite lo studio dei metadati ed artefatti, in modo che anche se un file ha cambiato nome, ha un contenuto leggermente differente da quello già noto e presente nei database, lo possiamo trovare addestrando un machine learning a capire dai metadati se quel file è vicino ai metadati di qualcuno già noto e sospetto.

MACHINE LEARNING AND DIGITAL FORENSICS

Insomma, una specie di “profiling” del file, si osservano molti profili, chiaramente non psicologici, ma tecnici e quando ci si trova di fronte a qualcosa di nuovo, ma che ha un “comportamento” e delle caratteristiche simili ai “bad files”, la macchina dovrebbe segnalarlo.

Chiaramente l'esempio ed il codice qui riportato è rudimentale ed a scopo didattico, ma può servire a comprendere meglio lo schema:



FACE RECOGNITION



```
# -*- coding: utf-8 -*-  
"""
```

Created on Fri Mar 26 09:18:08 2021

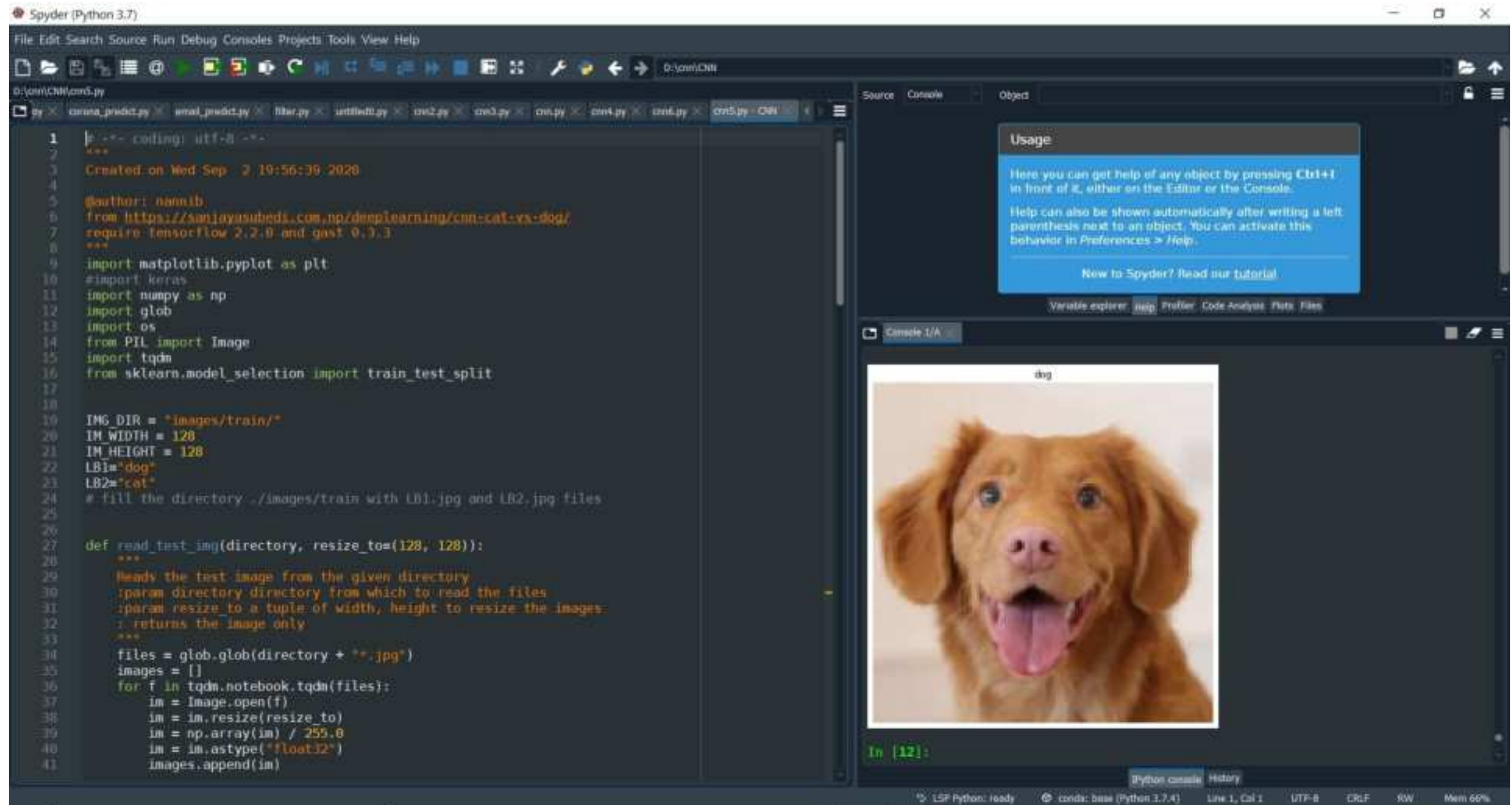
```
@author: nannib  
"""
```

```
import face_recognition  
import cv2
```

```
image = face_recognition.load_image_file("folla.jpg")  
img = cv2.imread("folla.jpg")  
for (top,right,bottom,left) in  
face_recognition.face_locations(image):  
    cv2.rectangle(img, (left,top), (right,bottom), (0,0,255))
```

```
cv2.imshow("folla", img)  
while True:  
    if cv2.waitKey(1) & 0xFF == ord("q"):  
        break  
cv2.destroyAllWindows()
```

CNN – CONVOLUTIONAL NEURAL NETWORK



The screenshot displays the Spyder Python IDE interface. The main editor window on the left contains a Python script for a CNN. The script includes comments, imports for various libraries (matplotlib, numpy, glob, os, PIL, tqdm, sklearn), and a function definition for reading and processing test images. The right-hand pane is divided into three sections: 'Usage' (providing help on object documentation), 'Variable explorer' (showing a variable named 'img'), and 'Console' (displaying the output of the script). The 'img' variable in the variable explorer contains a 128x128 pixel image of a brown dog with its tongue out. The console shows the command 'In [12]:' followed by the image display.

```
1  -*- coding: utf-8 -*-
2  ***
3  Created on Wed Sep  2 19:56:39 2020
4
5  @author: nannib
6  from https://sanjayasubedi.com.np/deeplearning/cnn-cat-vs-dog/
7  require tensorflow 2.2.0 and gast 0.3.3
8  ***
9  import matplotlib.pyplot as plt
10 #import keras
11 import numpy as np
12 import glob
13 import os
14 from PIL import Image
15 import tqdm
16 from sklearn.model_selection import train_test_split
17
18
19 IMG_DIR = "images/train/"
20 IM_WIDTH = 128
21 IM_HEIGHT = 128
22 LB1="dog"
23 LB2="cat"
24 # fill the directory ./images/train with LB1.jpg and LB2.jpg files
25
26
27 def read_test_img(directory, resize_to=(128, 128)):
28     """
29     Reads the test image from the given directory
30     :param directory directory from which to read the files
31     :param resize_to a tuple of width, height to resize the images
32     : returns the image only
33     """
34     files = glob.glob(directory + "*.jpg")
35     images = []
36     for f in tqdm.notebook.tqdm(files):
37         im = Image.open(f)
38         im = im.resize(resize_to)
39         im = np.array(im) / 255.0
40         im = im.astype("float32")
41         images.append(im)
```

Usage

Here you can get help of any object by pressing **Ctrl+I** in front of it, either on the Editor or the Console.

Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in **Preferences > Help**.

New to Spyder? Read our tutorial.

Variable explorer: [help](#) [Profiler](#) [Code Analysis](#) [Plots](#) [Files](#)

Console 1/A

dog



In [12]:

Python console History

LSP Python: ready | conda: base (Python 3.7.4) | Line 1, Col 1 | UTF-8 | CRLF | RW | Mem 66%

ANOMALY DETECTION – IP ADDRESSES AND TIME

Pensiamo ad un dataset di connessioni formato da indirizzi IP ed orari di connessione, quindi un tabulato, potremmo addestrare il programma di machine learning a capire qual è la connessione “anomala” e basterebbero poche righe di codice.

Il programma sfrutta l'*Isolation Forest*, un algoritmo non supervisionato e quindi non necessita di etichette per identificare il valore anomalo / anomalia. Il fondamento dell'algoritmo è quello di "isolare" le anomalie creando alberi decisionali su attributi casuali, questo partizionamento casuale produce percorsi più brevi per le anomalie quindi, se ci sono poche anomalie ci saranno partizioni più piccole di dati, perché i valori delle anomalie hanno maggiori probabilità di essere separati nel partizionamento iniziale.

Quindi, quando una foresta di alberi casuali produce lunghezze di percorso più brevi per alcuni punti particolari, è molto probabile che siano anomalie.

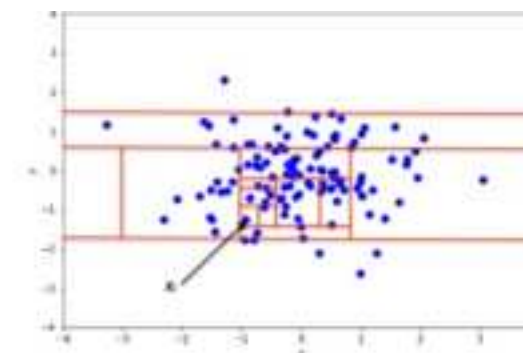


Fig. 1 - ci sono tante partizioni sui dati per isolare un singolo punto normale.

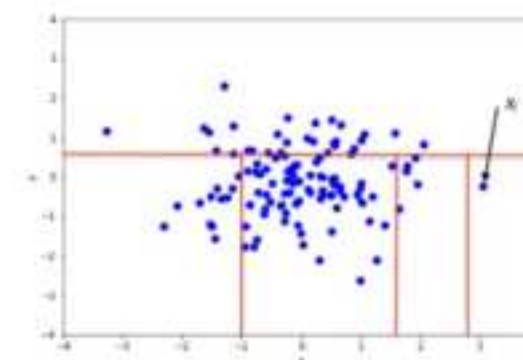


Fig. 2 - ci sono solo quattro partizioni per isolare il punto anomalo.

NLP – MACHINE LEARNING ANTI-PLAGIARISM

In questo caso avendo un dataset di testi scritti da vari autori, si può addestrare la macchina a capire chi ha scritto un testo nuovo che forniamo in input.

Qui si tratta di convertire le parole in numeri e poi calcolare dei valori, quindi un testo nuovo avrà un valore numerico, se la distanza tra il valore del nuovo testo ed il valore di altri testi di un certo autore, allora quel testo probabilmente apparterrà a quell'autore.

- NLP: Natural Language Processing

<https://nannibassetti.blogspot.com/2021/02/machine-learning-attribuire-la.html>

AI in Cloud vs. AI Locale

AI in Cloud

Nessun investimento hardware, modelli aggiornati, integrazione rapida. Ma in ambito forense introduce un **trust boundary esterno**: i dati trasmessi alle API possono essere loggati o conservati dal fornitore.

⚠️ Problematiche di GDPR, segreto istruttorio e sovranità del dato.

Private AI (Locale)

Esecuzione di LLM, embedding e sistemi RAG in ambienti completamente controllati: on-premise, air-gapped o cloud privati. Garantisce controllo totale sul dato, auditabilità, conformità normativa e personalizzazione profonda.

✅ In Digital Forensics è spesso l'unico modo per garantire che le evidenze non escano mai dal perimetro investigativo.

L'evoluzione perfetta si chiama Private AI



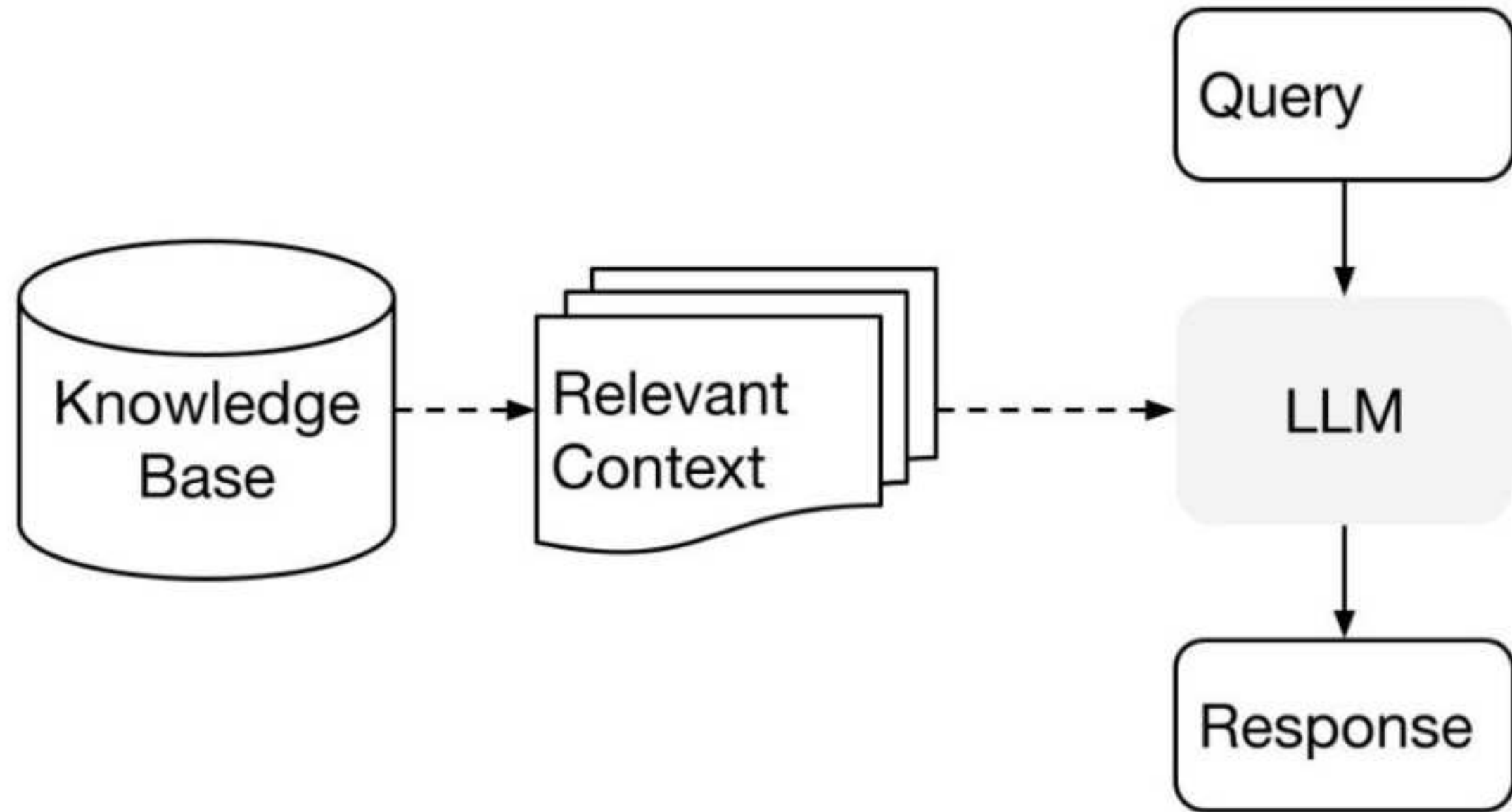


Immagine da: <https://www.falkordb.com/blog/what-is-a-rag-retrieval-augmented-generation/>

Cos'è il RAG (Retrieval-Augmented Generation) nell'AI Generativa?

Definizione

Il RAG (Retrieval-Augmented Generation) è una tecnica avanzata nell'intelligenza artificiale che combina due modelli principali:

- 1. Modello di Recupero (Retriever):** Seleziona informazioni rilevanti da una vasta base di conoscenze.
- 2. Modello Generativo (Generator):** Genera risposte coerenti e contestualmente appropriate utilizzando le informazioni recuperate.

Come Funziona

- **Fase di Recupero:**

- Il modello di recupero cerca informazioni pertinenti da un database o da una collezione di documenti basandosi sulla query iniziale dell'utente.

- **Fase Generativa:**

- Il modello generativo utilizza le informazioni recuperate per formulare una risposta dettagliata e contestualizzata.

Vi presento il vostro Bibliotecario personale



**Nessun costo di riprogrammazione (fine-tuning).
L'IA legge i nostri documenti aggiornati in tempo reale.**

La Private AI garantisce il controllo assoluto del perimetro investigativo



Nella Digital Forensics, la Private AI non è una scelta ideologica, ma un requisito operativo ineludibile.

Il framework RAG funge da bibliotecario per ancorare l'IA alle evidenze



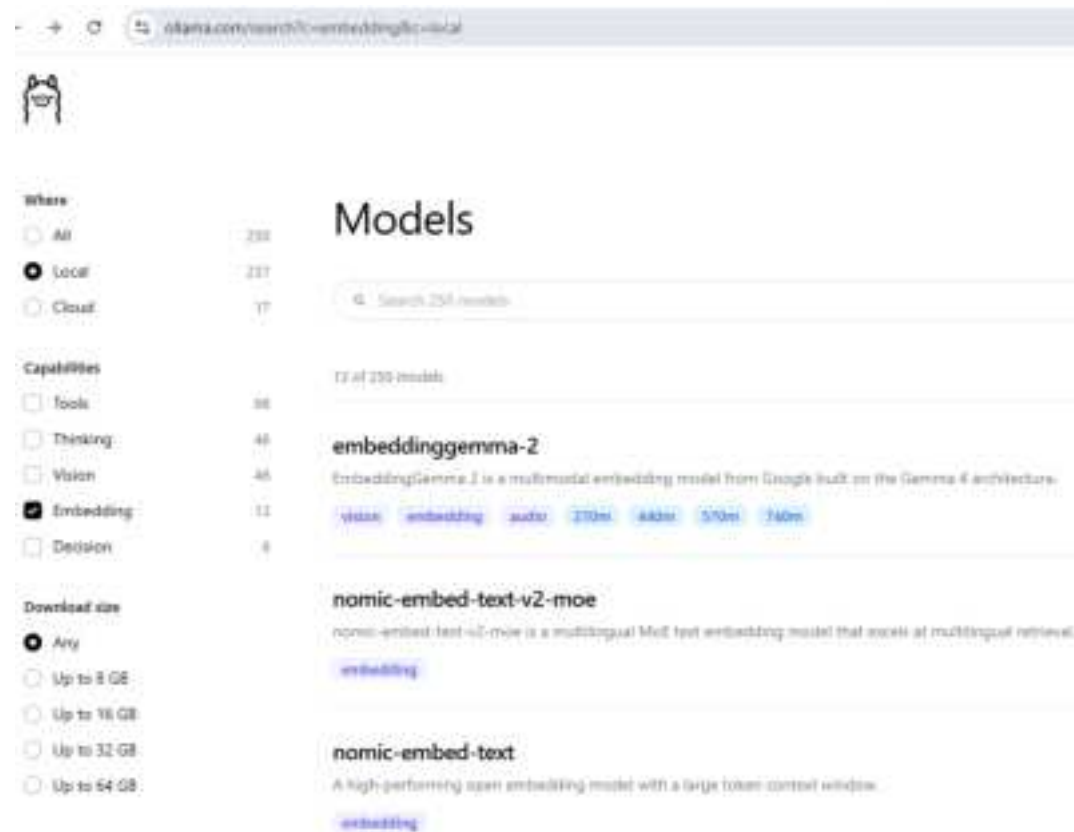
Questo approccio riduce le allucinazioni grazie all'ancoraggio a fonti verificabili e consente di impiegare modelli generici in domini altamente specifici.

Perché conviene? Il confronto definitivo

	Intelligenza Artificiale in Cloud	Intelligenza Artificiale Privata
Privacy e GDPR	 Bassa (Dati inviati a server esterni).	 Massima (I dati restano in casa).
Costi Operativi	 Abbonamento mensile e costi per ogni singolo utilizzo (API).	 Solo il costo iniziale del PC. Nessuna bolletta mensile.
Controllo Operativo	 Nessuno (Dipendenza dai server del fornitore).	 Totale (Il sistema è sempre a nostra disposizione).

Embedding: il Significato in Numeri

Un **modello di embedding** genera vettori numerici multidimensionali che rappresentano il **significato semantico** di sequenze di testo. Questi vettori costituiscono la base di conoscenza interrogabile dal sistema RAG.



nomic-embed-text

Alta performance,
ampio contesto token

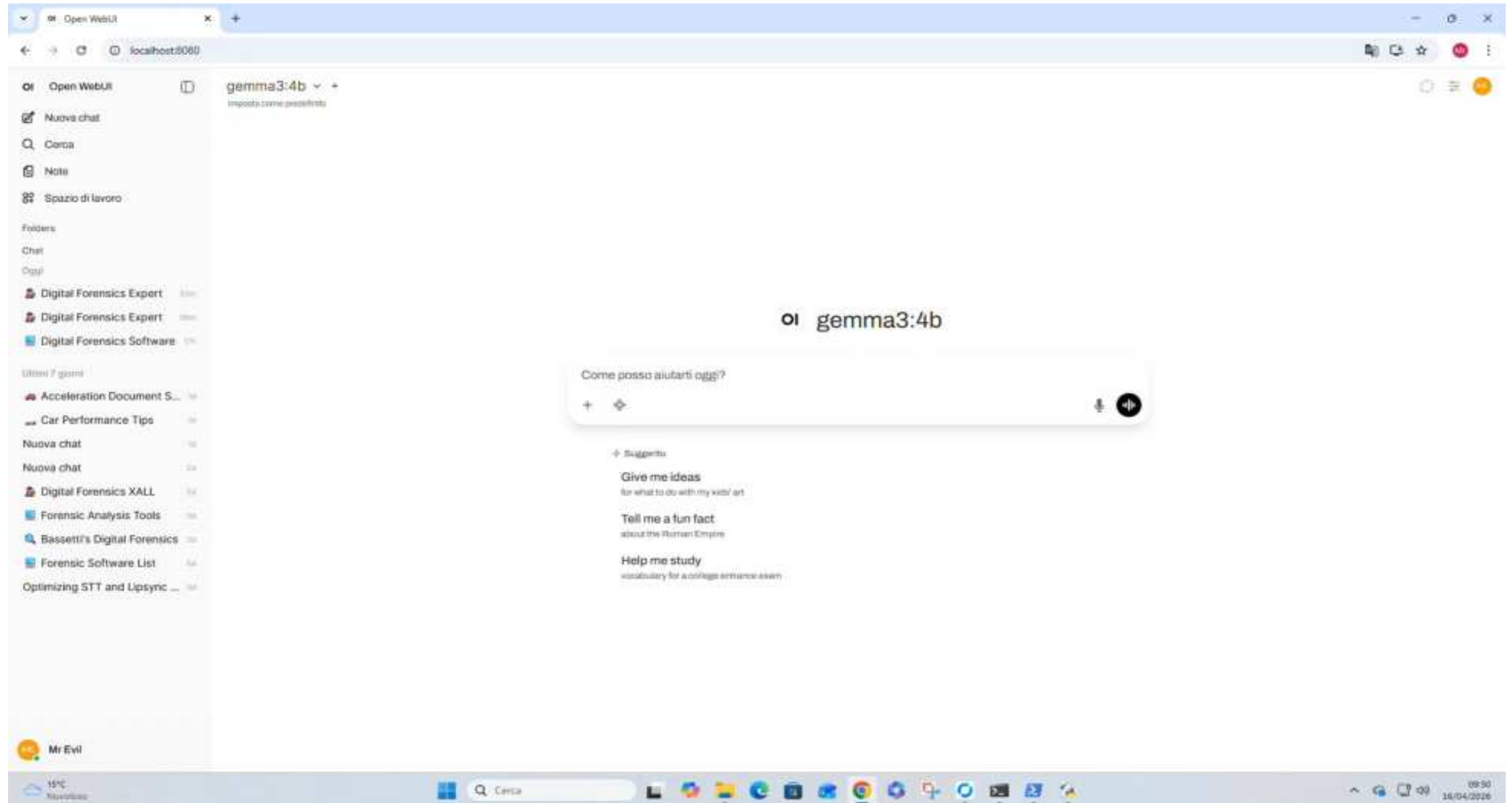
mxbai-embed-large

Modello large di
mixedbread.ai (335M)

bge-m3

Versatilità multi-
funzionale e
multilingua (567M)

Esempio di RAG su me stesso



NBMULTIRAG

<https://github.com/nannib/nbmultirag>

Quindi che cos'è NBMultiRag? È un framework che integra modelli di linguaggio GPT (LLM) scaricati da [OLLAMA](#), per consentire agli utenti di interagire in modalità chat con i propri documenti, inoltre il sistema, non solo permette di chattare "in bianco" (senza l'uso del RAG), ma anche di creare workspace personalizzati per indicizzare cartelle contenenti documenti eterogenei eseguendo:

- **Indicizzazione e suddivisione dei documenti:** I documenti vengono frammentati in segmenti (chunk) per essere processati più efficientemente.
- **Trascrizione e OCR:** I file audio e video vengono trascritti, mentre le immagini sono processate tramite OCR ed anche descritte, mentre per i video vengono estratti e descritti 10 frame distribuiti uniformemente.
- **Interazione tramite upload:** È possibile caricare file singoli per una descrizione puntuale del contenuto.

NBMULTIRAG

← → ↻ ⓘ localhost:8501

🔍 ☆ 👤 ⋮

Deploy ⋮

Scegli la lingua / Choose Language

☐ Italian ☒ English

Workspace Management

New Chat

Select Model ⓘ
llama3.1:8b ▼

Make a new workspace

Make

Active workspace
takeout2 ▼

Workspace Configuration

▼

Upload File

Drag and drop file here

🗨️ e cosa c'è nel corpo dell'e-mail?

📄 Nel Documento 8, che sembra essere l'e-mail in questione, il corpo del messaggio è:
"I had the campaign done, but it's not on my desktop anymore?? The folder I have on my desktop is now .7z(?) and it won't open. I have no idea what happened, it was there yesterday."

In sintesi, la persona (Ruth) sta chiedendo aiuto perché non riesce a trovare un progetto che aveva fatto precedentemente e ora è presente come file compresso (.7z) sul suo desktop, ma non può essere aperto.

🗨️ che video ha visto su YouTube?

📄 Non sono in grado di trovare informazioni specifiche sui video visti da Ruth o Greg su YouTube. Tuttavia, posso notare che nel Documento 14 è presente un riferimento a un video intitolato "Little Liars Stars Shay Mitchell and Ashley Benson Play 'Which Liar?' | Glamour" pubblicato il 31 ottobre 2024 sul canale di YouTube di Glamour.

Your query:

➤

NBMULTIRAG

localhost:8501

takeout2

Configurazione Workspace

System Prompt

principalmente sui documenti
forniti

2. Integra conoscenza esterna
solo quando necessario

3. Sii preciso e fornisci
riferimenti ai documenti

Embedder

BERT Italiano (Default)

Chunk Size

512

Chunk Overlap

128

Temperature

0.11

0.00

1.00

Numero di risultati nella ricerca

20

Numero di documenti rilevanti nel contesto

60



NBMultiRAG





Secondo il documento n. 9, i calzini sono chiamati "Floppers". Inoltre, nel documento n. 10 e n. 11, si fa riferimento al "Floppers" come un prodotto di comfort e stile per la moda dei piedi.

In particolare, il titolo del documento n. 9 è: "Floppers: The Ultimate Comfort Shoe".

Documenti di riferimento

1) All mail Including Spam and Trash.mbox

Percorso/Path: D:/nbmultirag/nbmultirag-1.4/Takeout\Mail\All mail Including Spam and Trash.mbox

Congratulations on taking the first step towards landing the job you really= =20 want!=20.....

2) MyActivity.html



Video nel rag



Scegli la lingua / Choose Language

Italian

English

Gestione Workspace

Nuova Chat

Seleziona modello

gpt-oss:20b

Crea nuovo workspace

Crea

Workspace attivo

test

Configurazione Workspace

System Prompt

Sei un esperto assistente che analizza documenti. Rispondi in italiano seguendo queste regole:

1. Base le risposte principalmente sui documenti



NBMultiRAG

dimmi se tra i documenti forniti c'è un riferimento alla Ferrari e cosa viene detto in merito

Sì, il riferimento a Ferrari appare nel Documento 8.

In quel documento si legge:

"Con queste modifiche la mia Ferrari va velocissima."

Quindi l'unico documento che menziona Ferrari afferma semplicemente che, grazie alle modifiche apportate (presumibilmente con il Super Accelerator), la vettura diventa molto più veloce. No altri documenti forniscono ulteriori dettagli o contestualizzazioni su Ferrari.

Documenti di riferimento

Scegli la lingua / Choose Language

☒ Italian ☐ English

Gestione Workspace

Nuova Chat

Seleziona modello ?

gpt-oss:20b

Crea nuovo workspace

Crea

Workspace attivo

test

Configurazione Workspace

System Prompt

Sei un esperto assistente che analizza documenti. Rispondi in italiano seguendo queste regole:
1. Basa le risposte principalmente sui documenti

Embedder

DC1LEX/nomic-embed-tex...

Con queste modifiche la mia Ferrari va velocissima.

Quindi l'unico documento che menziona Ferrari afferma semplicemente che, grazie alle modifiche apportate (presumibilmente con il Super Accelerator), la vettura diventa molto più veloce. No altri documenti forniscono ulteriori dettagli o contestualizzazioni su Ferrari.

e che si vede nel filmato fornito?

Nel materiale video riportato nel Documento 5 si vede un toy car (auto giocattolo) posizionato sul pavimento.
Il file è strutturato come una serie di frame numerati:

Frame 0: a toy car is on the floor

Frame 17: a toy car is sitting on the floor

Frame 34: a toy car is sitting on the floor

Frame 52: a toy car is sitting on the floor

Frame 69: a toy car is sitting on the floor

Frame 87: a toy car is sitting on the floor

Frame 104: a toy car is sitting on the floor

Frame 122: a toy car is sitting on the floor

Frame 139: a toy car is sitting on the floor

Frame 157: a toy car is sitting on the floor

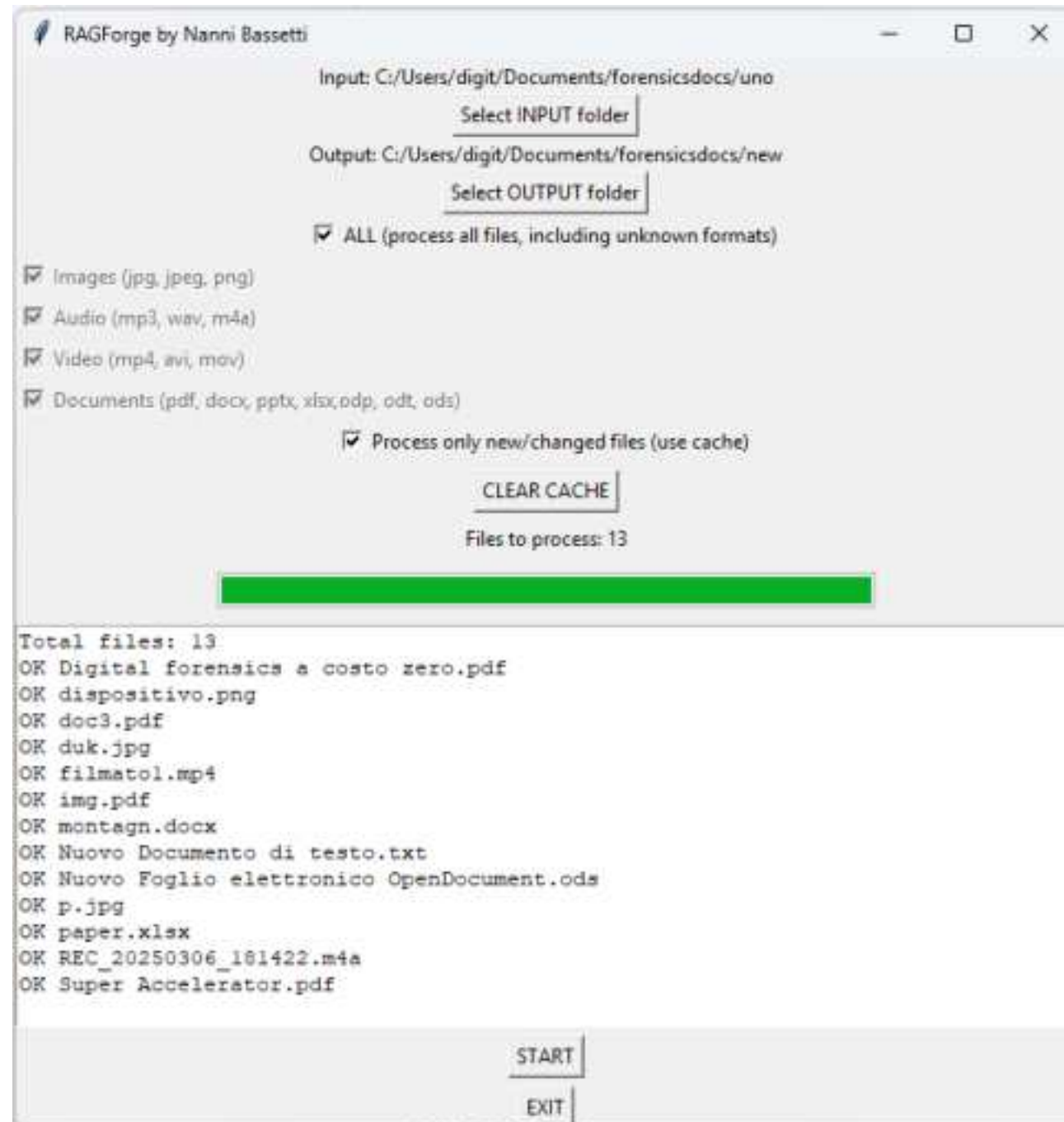
Non viene mostrata alcuna Ferrari né alcun veicolo reale; l'unico veicolo visibile è l'auto giocattolo.

Nessun altro documento fornisce indicazioni sul contenuto di un filmato.

Documenti di riferimento

Scrivi la tua domanda:

RAGForge - <https://github.com/nannib/RAGForge>



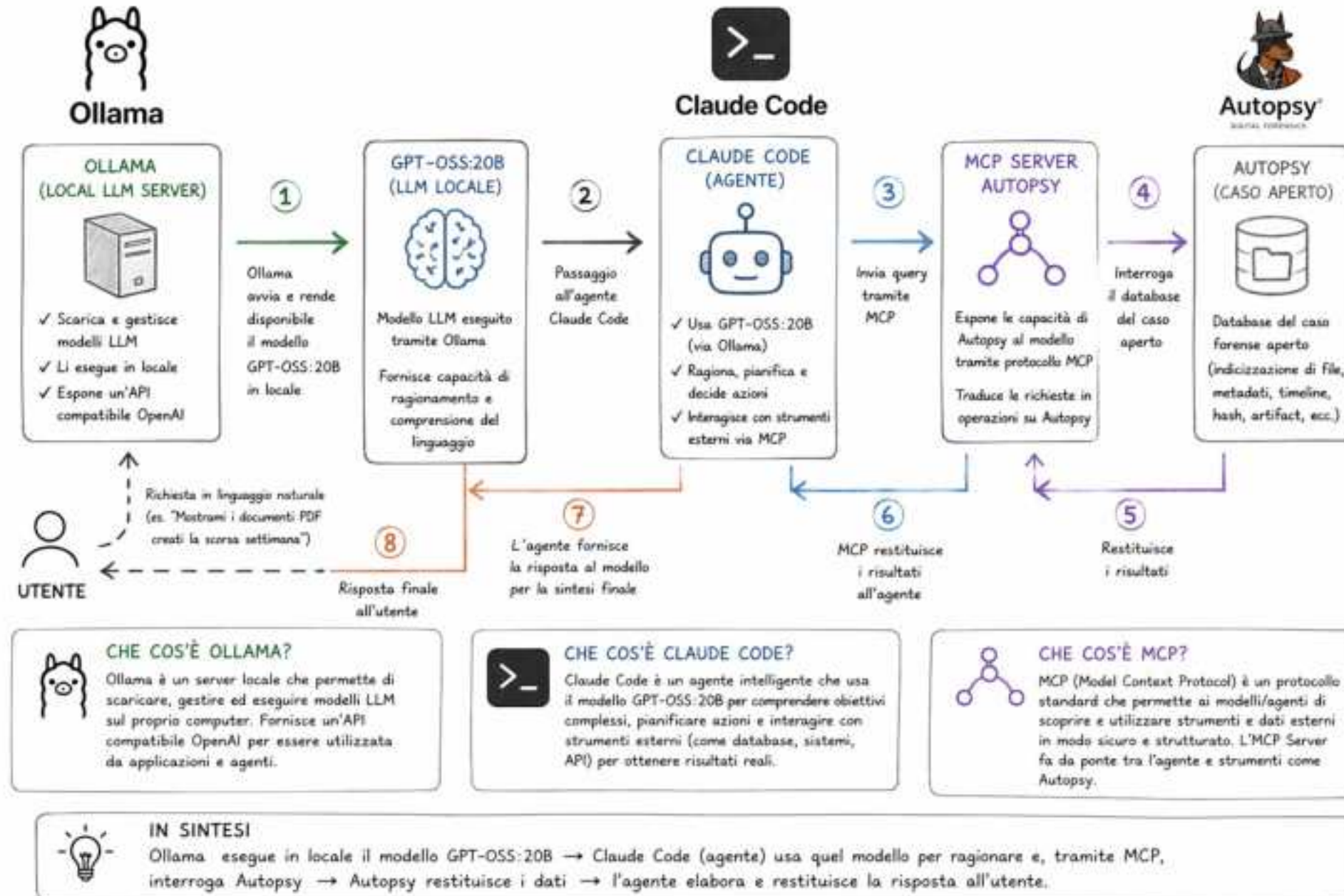
Generatore di una pipeline multimodale di estrazione e normalizzazione contenuti (OCR + ASR + captioning + document parsing + export testuale per ingestion). Il programma permette di estrarre testi da una directory di input contenente vari tipi di file: docx, pdf, mp4, wav, m4a, jpg, png, ecc.. In una cartella di output sono scritti tutti le parti testuali in file nominati come i file originali ma con estensione TXT.

Il programma inserisce anche la descrizione dei frame di un filmato, la descrizione di un'immagine, la trascrizione audio, ecc.. Una volta che la cartella di output si è popolata, allora si può usarla come directory di RAG, per programmi come open-webui, ecc.ecc..

OCR (Optical Character Recognition)
ASR (Automatic Speech Recognition)

AGGIUNGIAMO L'MCP SERVER DI AUTOPSY A CLAUDE CODE

```
claude mcp add --transport stdio autopsy -- "C:\Program Files\Autopsy-4.23.0\bin\autopsy-mcp-stdio.exe"
```



Possiamo anche creare modello personalizzato con la finestra di contesto allargata:
creiamo un modelfile.txt:

From gpt-oss:20b
PARAMETER num_ctx 65535

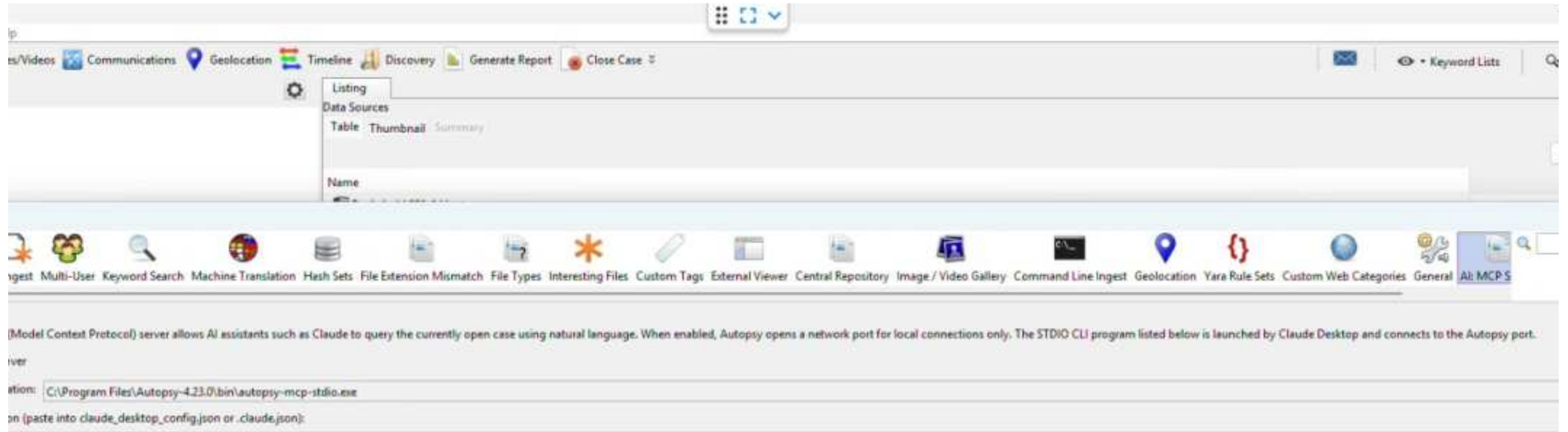
poi lanciamo:

ollama create gptoss65k -f modelfile.txt

L'LLM è il motore, ma l'Agente AI è il veicolo completo, capace di navigare, interagire e raggiungere destinazioni. L'MCP è il linguaggio dei cartelli e dei semafori che permette all'auto di interagire con il mondo fuori dal cofano

- <https://www.autopsy.com/autopsy-4-23-0-release-claude-ai-assistant-mcp-cyber-triage-integration/>
- https://sleuthkit.org/autopsy/docs/user-docs/4.23.0/claude_mcp_page.html

Avviamo Autopsy e nelle opzioni attiviamo MCP Server, poi riavviamo Autopsy ed apriamo un caso



```
{
  "mcp": {
    "server": "C:\\Program Files\\Autopsy-4.23.0\\bin\\autopsy-mcp-studio.exe"
  }
}
```

Copy to

OK

Apply

ollama launch claude

Claude Code v2.1.114

Welcome back!



qwen3.5:cloud with low effort · API Usage Billing
-\\Desktop\\esercizi\\LexarIecce

Tips for getting started
Run /init to create a CLAUDE.md file with instructions for Cla...

Recent activity
No recent activity

what are the suspected files? use autopsy

- I'll check the Autopsy case for suspected files by querying analysis results with suspicion-related artifact types.
Called autopsy 2 times (ctrl+o to expand)
- Good, a case is open. Let me search for suspected files and encryption-related findings.
Called autopsy 4 times (ctrl+o to expand)
- Now let me get the file details for the suspected encryption files:
Called autopsy 5 times (ctrl+o to expand)
- Suspected Files in the Case

File	Reason	Significance
Unalloc_508_254983680_257949696 (2.8 MB)	High entropy data (7.78) in unallocated space - suspected encryption	LikelyNotable
Passwordchinese.dat (26 KB)	Password-protected Word document - encryption detected	Notable

Details:

1. Unalloc_508_254983680_257949696

- Path: /img_PenLab.dd.001/vol_vol3/Unalloc_508_254983680_257949696
- Located in unallocated volume space
- Flagged for high entropy suggesting encrypted or compressed content

2. Passwordchinese.dat

- Path: /img_PenLab.dd.001/vol_vol2/Documents and Settings/Documents/Passwordchinese.dat
- Microsoft Word document with password protection
- Author: joan.donohue
- Also has a file extension mismatch detected

is there a secret message inside any jpg file?

- I'll search for JPG files and check for any hidden content or suspicious analysis results.

Called autopsy 2 times (ctrl+o to expand)

- Let me check which files have extension mismatches and examine suspicious JPG files for hidden content:

Called autopsy 6 times (ctrl+o to expand)

- I found several files with extension mismatches - some are actually JPG images hidden with different extensions. Let me check the actual JPG files for hidden content:

Called autopsy 3 times (ctrl+o to expand)

- Those files are actually JPEG images disguised with wrong extensions. Let me check the real JPG files for hidden text or steganography:

Called autopsy 2 times (ctrl+o to expand)

- Let me check the actual JPG files for hidden content by examining their content more closely:

Called autopsy 2 times (ctrl+o to expand)

- Let me check the end of JPG files for hidden text and look at files with unusual characteristics:

Called autopsy 3 times (ctrl+o to expand)

- Let me search for hidden text patterns inside JPG files more specifically:

Called autopsy 3 times (ctrl+o to expand)

- Let me check for embedded text in JPG files and look at files with suspicious characteristics:

Called autopsy 2 times (ctrl+o to expand)

- Found it! There's a secret message hidden inside pict1.jpg:

Secret Message Found

File: /img_PenLab.dd.001/vol_vol2/Documents and Settings/krank/Document1/pict1.jpg (49 KB)

Hidden Text:

| "The last stash of weapons and the explosives is stroed in the junker car at the corner of Elm Ave and Highland St. Look in the trunk under the carpet in the spare tire well under the trash."

This is a classic case of steganography - text hidden inside an image file. The message is embedded in the JPEG data and was discovered by reading the file content. The message describes a location where weapons and explosives are hidden.

⌘ Cooked for 3m 16s

LeviLeCce - Autopsy 4.23.0

Case View Tools Window Help

Add Data Source

Images/Videos

Communications

Geolocation

Timeline

Discovery

Generate Report

Close Case

Data Sources

PenLab.dd.001_1 Host

PenLab.dd.001

vol1 (Unallocated: 0-62)

vol2 (NTFS / exFAT (0x07): 61-498014)

OrphanFiles (13)

ScavengedFiles (1)

Extend (5)

Unalloc (1)

aa (5)

bb (4)

Documents and Settings (7)

Administrator (2)

All Users (2)

Default User (18)

Documents (5)

krenk (18)

java (2)

Applications Data (2)

Cookies (3)

Defi applicazioni (6)

Desktop (8)

Documents (18)

mydocs (2)

Nuova cartella (2)

forensics%20intermedio.pdf (3118)

Insuistro-LEFE-3.78.pdf (33)

net_forensics.pdf (32)

vacation.ppt (5)

Importazioni locali (7)

Menu Avvio (4)

Modeli (14)

Professi (2)

Recent (2)

Ricerca di rete (2)

Ricerca di stampa (2)

SendTo (6)

Temp (6)

Downloads (6)

foto (6)

my Folder (21)

Nuova cartella (20)

Programs (5)

Temp (4)

vol3 (Unallocated: 498015-503807)

File Views

File Types

Deleted Files

MB File Size

Listing

/img_PenLab.dd.001/vol_vol2/Documents and Settings/krenk/Documents

Table Thumbnail

Name	S	C	O	Modified Time	Change Time	Access Time
[current folder]				2010-02-17 17:38:13 CET	2010-02-17 17:38:13 CET	2010-02-17 17:38:13 CET
[parent folder]				2010-02-17 17:06:08 CET	2010-02-17 17:06:08 CET	2010-02-17 16:25:52 CET
mydocs				2010-02-17 17:38:08 CET	2010-02-17 17:38:13 CET	2010-02-17 16:36:28 CET
Nuova cartella				2010-02-17 17:38:02 CET	2010-02-17 17:38:02 CET	2010-02-17 16:36:28 CET
chat1.txt			0	2009-05-04 13:26:48 CEST	2009-05-30 19:29:04 CEST	2010-02-17 17:26:40 CET
forensics%20intermedio.pdf			0	2009-02-20 17:38:41 CET	2009-02-20 17:38:42 CET	2010-02-17 17:10:48 CET
Insuistro-LEFE-3.78.pdf			0	2009-11-21 16:39:02 CET	2009-11-21 16:39:02 CET	2010-02-17 16:36:30 CET
net_forensics.pdf			0	2010-02-02 09:02:16 CET	2010-02-02 09:02:16 CET	2010-02-17 17:10:34 CET
pk11.jpg			0	2009-05-10 17:56:16 CEST	2009-05-30 18:04:15 CEST	2010-02-17 17:27:01 CET
vacation.ppt			0	2009-05-28 17:46:51 CET	2009-05-30 19:30:42 CEST	2010-02-17 17:52:13 CET

New Text Application File Metadata

Strings Extracted Text

Page: 1 of 1 Page

Matches on page: - of - Match

100%

Reset

Text Source: File Text

The last stack of weapons and the rest of the explosives is stored in the junker car at the corner of Elm Ave and Highland St. Look in the trunk under the carpet in the spare tire well under the trunk

Listing

/img_PenLab.dd.001/vol_vol2/Documents and Settings/krenk/Documents

Table Thumbnail

Name	S	C	O	Modified Time	Change Time	Access Time	Size	Flags	Flags2	Access	Location
[current folder]				2010-02-17 17:38:13 CET	2010-02-17 17:38:13 CET	2010-02-17 17:38:13 CET	76	Allocated	Allocated	unknown	/img_PenLab
[parent folder]				2010-02-17 17:06:08 CET	2010-02-17 17:06:08 CET	2010-02-17 16:25:52 CET	36	Allocated	Allocated	unknown	/img_PenLab
mydocs				2010-02-17 17:38:08 CET	2010-02-17 17:38:13 CET	2010-02-17 16:36:28 CET	46	Allocated	Allocated	unknown	/img_PenLab
Nuova cartella				2010-02-17 17:38:02 CET	2010-02-17 17:38:02 CET	2010-02-17 16:36:28 CET	46	Allocated	Allocated	unknown	/img_PenLab
chat1.txt			0	2009-05-04 13:26:48 CEST	2009-05-30 19:29:04 CEST	2010-02-17 17:26:40 CET	140	Allocated	Allocated	unknown	/img_PenLab
forensics%20intermedio.pdf			0	2009-02-20 17:38:41 CET	2009-02-20 17:38:42 CET	2010-02-17 17:10:48 CET	140384	Allocated	Allocated	unknown	/img_PenLab
Insuistro-LEFE-3.78.pdf			0	2009-11-21 16:39:02 CET	2009-11-21 16:39:02 CET	2010-02-17 16:36:30 CET	45000	Allocated	Allocated	unknown	/img_PenLab
net_forensics.pdf			0	2010-02-02 09:02:16 CET	2010-02-02 09:02:16 CET	2010-02-17 17:10:34 CET	32000	Allocated	Allocated	unknown	/img_PenLab
pk11.jpg			0	2009-05-10 17:56:16 CEST	2009-05-30 18:04:15 CEST	2010-02-17 17:27:01 CET	36960	Allocated	Allocated	unknown	/img_PenLab
vacation.ppt			0	2009-05-28 17:46:51 CET	2009-05-30 19:30:42 CEST	2010-02-17 17:52:13 CET	2220544	Allocated	Allocated	unknown	/img_PenLab

New Text Application File Metadata

Strings Extracted Text

Page: 1 of 1 Page

Matches on page: - of - Match

100%

Reset

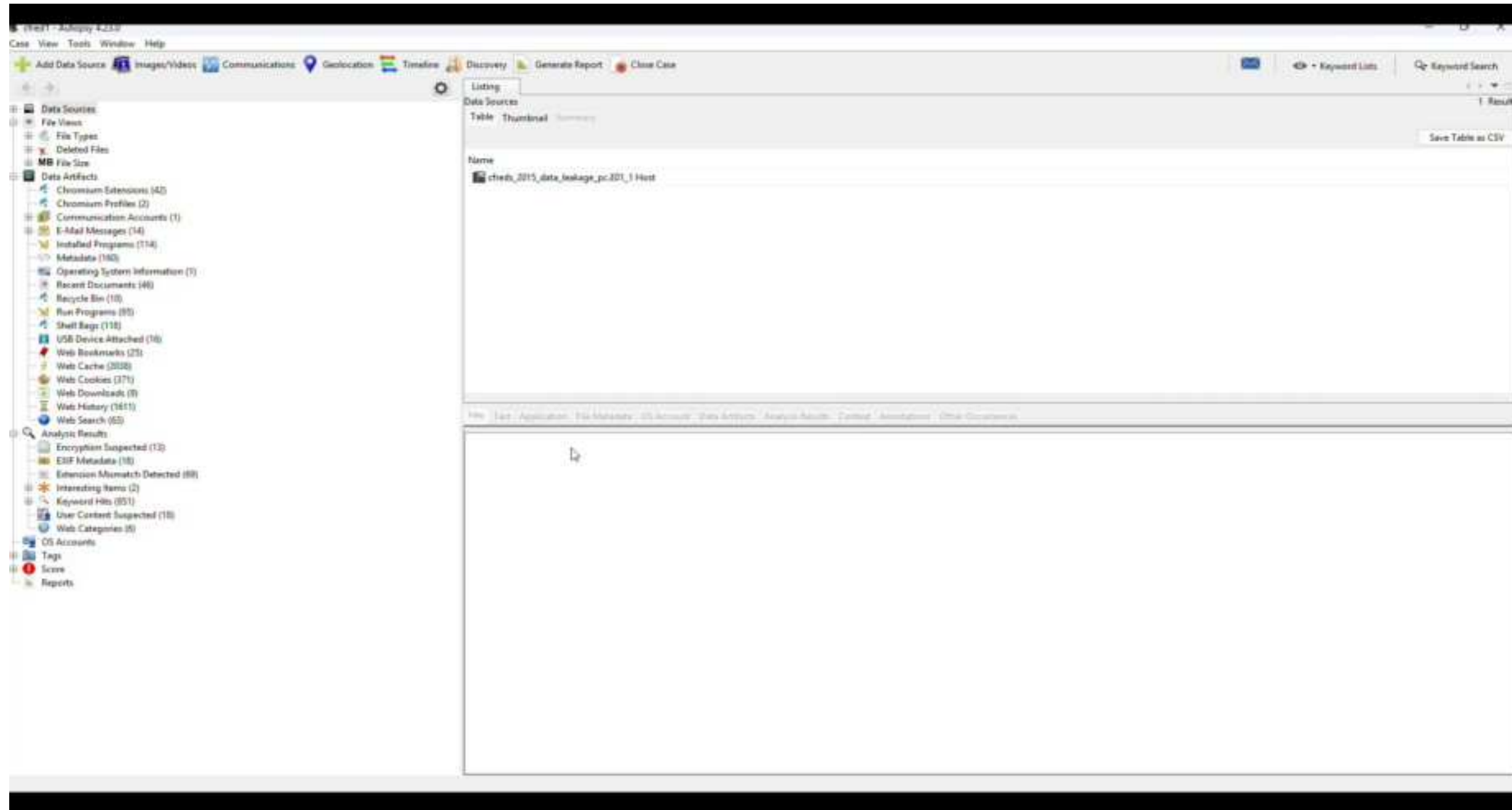
Text Source: File Text

The last stack of weapons and the rest of the explosives is stored in the junker car at the corner of Elm Ave and Highland St. Look in the trunk under the carpet in the spare tire well under the trunk

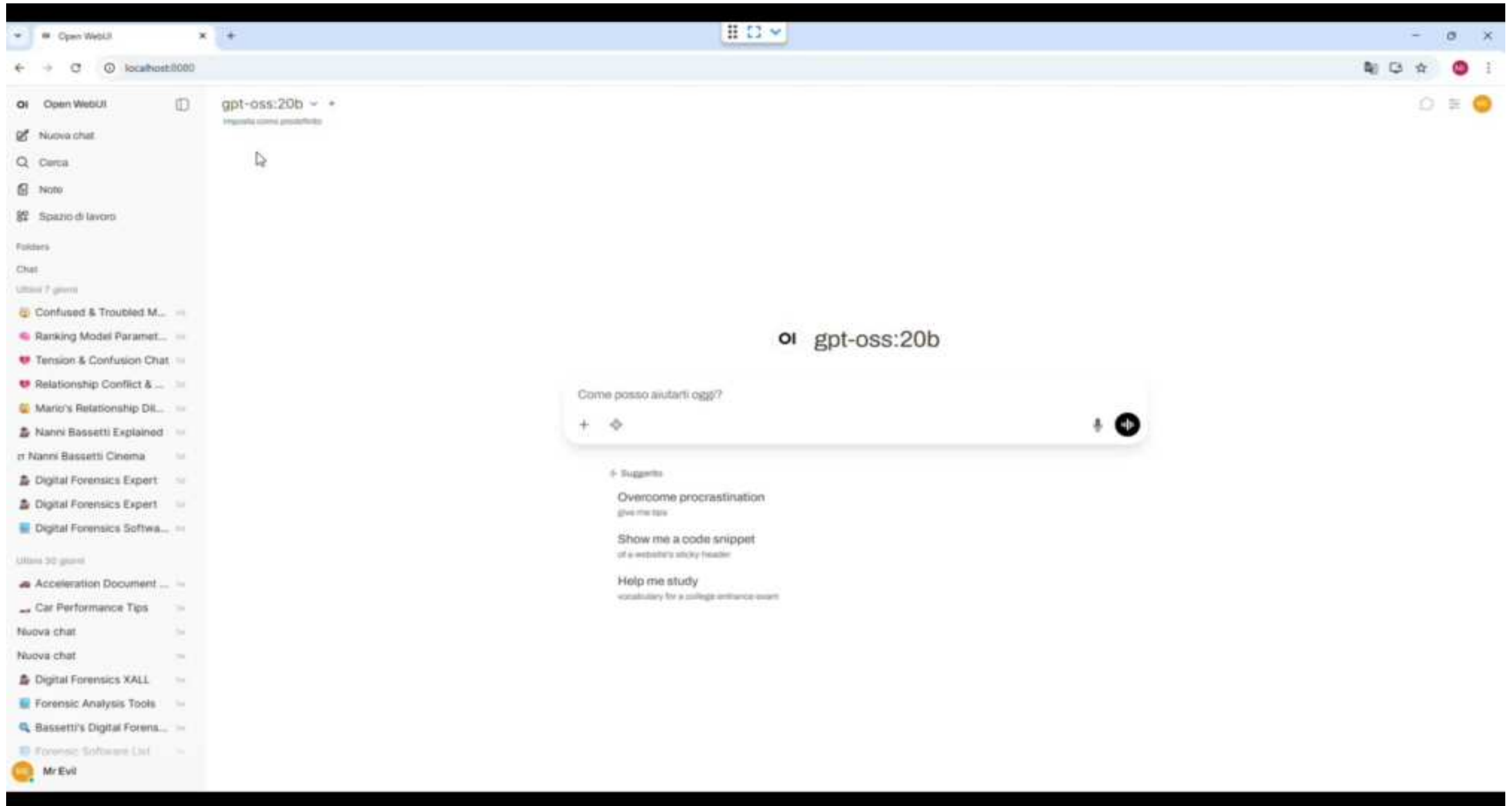
USIAMO LM STUDIO - <https://lmstudio.ai/>



MCP AUTOPSY CON LMSTUDIO SU QWEN3.8:27B



ANALISI DELLE CHAT

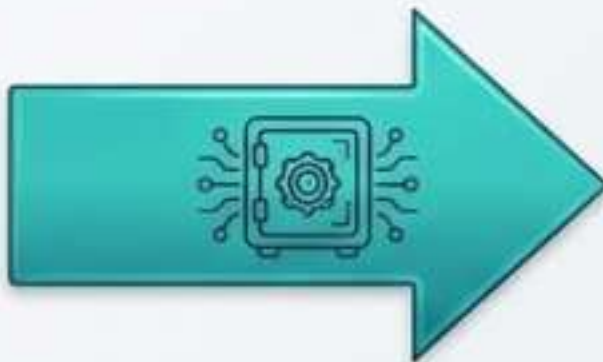


L'IA in azione: un assistente instancabile



Migliaia di documenti, email e contratti da analizzare a mano.

Chiediamo:
Riassumi tutte le email
con allegati sospetti
degli ultimi 30 giorni.



Estrazione immediata e precisa
delle sole informazioni rilevanti.

**L'Intelligenza Artificiale filtra, estrae e riassume ore di lavoro noioso
in pochi secondi. È uno strumento di primo soccorso cognitivo.**

Indipendenza tecnologica, controllo umano

“

La vera rivoluzione è un'Intelligenza Artificiale che ci solleva dalle mansioni ripetitive e protegge i nostri segreti. Ma ricordiamo sempre: la macchina analizza, **l'essere umano decide.**

”

- ✓ • Supporto cognitivo, non sostituzione.
- ✓ • Zero rischi legali
- ✓ • Massima convenienza economica nel lungo periodo

Esempio Operativo: "Parlare con le Evidenze"

Scenario: computer aziendale sospettato di esfiltrazione dati. Le evidenze estratte (EML, cronologia web, log, documenti) vengono normalizzate, indicizzate con embedding locali e interrogate tramite LLM locale con logging completo.

→ Query temporali

"Esistono correlazioni tra accesso a servizi cloud e invio di e-mail esterne?"

→ Query contestuali

"Quali documenti sono stati aperti nei 30 minuti precedenti a un upload web?"

→ Query riassuntive

"Riassumi le comunicazioni e-mail con allegati sospetti."



Il risultato non costituisce una prova automatica, ma uno strumento di **trriage cognitivo**.

Limiti, Rischi e Checklist di Adozione

⚠ Anti-pattern da evitare

- Usare risposte AI come evidenza probatoria diretta
- Automatizzare le conclusioni investigative
- Introdurre servizi cloud esterni in fasi critiche
- Fine-tuning su dati sensibili senza controlli
- Assenza di logging strutturato e versioning

✅ Checklist di adozione

- Evidenze in ambienti controllati, dataset separati per caso
- Versioning di modelli ed embedding
- Preferenza per RAG rispetto al fine-tuning
- Hardware dedicato con isolamento dell'inferenza
- Conformità GDPR/DPIA e revisione umana critica

⚠ La Private AI deve essere trattata come strumento di supporto cognitivo, non come sostituto del metodo forense. **L'ultima parola deve rimanere umana.**



Conclusioni

„L'AI deve stare dove stanno le evidenze

La Private AI rappresenta un'evoluzione naturale degli strumenti di supporto alla Digital Forensics.

Non sostituisce il metodo forense, ma lo potenzia, permettendo di interagire semanticamente con grandi volumi di evidenze senza compromettere controllo umano, riservatezza e validità legale

La scelta tra AI in cloud e AI locale non è solo tecnica, ma anche **etica e professionale**.

VESTIGEX

OVERVIEW

Dashboard

Refined Results

Case Management

Jobs

EVIDENCE

Data Sources

File System

File Types

Deleted & Carved

SEARCH

Keyword Search

Hash Sets

YARA Rules

ANALYSIS

Timeline

Connections

Media Analysis

AI Investigation

ARTIFACTS

Registry

OS Artifacts

Web History

Messaging

Email

Memory

Mobile

SQLite Databases

Credentials

THREAT HUNT

Findings

Lateral Movement

Process Inspector

Detection Rules

cfred

Example: Administrator - C:\Users\dig...

Artifacts 26.340

Events 15.519

Findings 0

Crit / High 0

Bookmarked 0

Date / time (UTC): From gg/mm/yyyy To gg/mm/yyyy 00:00:00 23:59:59

Data sources: All data sources

Administrator

Close Case

OVERVIEW

Dashboard

Artifacts 26.340

21 kinds

Events 15.519

0 log events - 15.519 dated artifacts

Findings 0

run a threat hunt

Critical 0

requires triage

High 0

requires review

MITRE techniques 0

0 tactics

Bookmarked 0

events, files and artifacts

IOCs 0

in hunt list

Files indexed 161.945

35.718 folders in disk images

Deleted files 1.158

11.452 extension mismatches

System information

COMPUTER NAME INFORMANT-PC

DOMAIN —

OPERATING SYSTEM Windows 7 Ultimate

BUILD 7601 - x86_64 - Client

INSTALLED 2015-03-22 14:34:26 UTC

REGISTERED OWNER informant

TIME ZONE Eastern Standard Time - UTC-04:00

LAST SHUTDOWN 2015-03-25 15:31:05 UTC

LAST ACTIVITY 2015-03-25 15:31:05 UTC

LAST LOGGED-ON USER .informant

IP ADDRESSES 18.11.11.129

LANGUAGE —

Network interfaces (1)

IP address Subnet Gateway DNS servers DHCP Domain

18.11.11.129 255.255.255.0 18.11.11.2 18.11.11.2 yes - 10.11.11.254 localdomain

Known networks (1)

User accounts (1)

Activity heatmap 0 matched events - UTC

No timestamped finding events yet.

Severity breakdown 0 findings

No findings yet.

Top findings by severity and confidence

Run a threat hunt to populate findings.

Evidence by type

Type Items Size / events

• Text 68.593 1.7 GB

• Executables 60.566 28 GB

• Other 18.408 8.1 GB

• Images 9.815 409 MB

• Audio 1.314 123 MB

• Documents 1.048 872 MB

• System artifacts 888 400 MB

• Archives 831 2.1 GB

BACKEND HEALTHY

http://127.0.0.1:8001

Index 0 events

Jobs idle

Hashes 0/0 SHA-256

AI gemma4:12b

VESTIGEX v11.2-beta

Made with GAMMA

GRAZIE PER L'ATTENZIONE!



NANNIBASSETTI.COM