

Introduzione all'Intelligenza Artificiale

Paolo Nesi, paolo.nesi@unifi.it

<https://www.snap4city.org>

<https://www.disit.org>

PatentIAMoci

**Intelligenza artificiale e metodo giuridico:
uso consapevole degli strumenti digitali nella ricerca e nella
pratica del diritto**

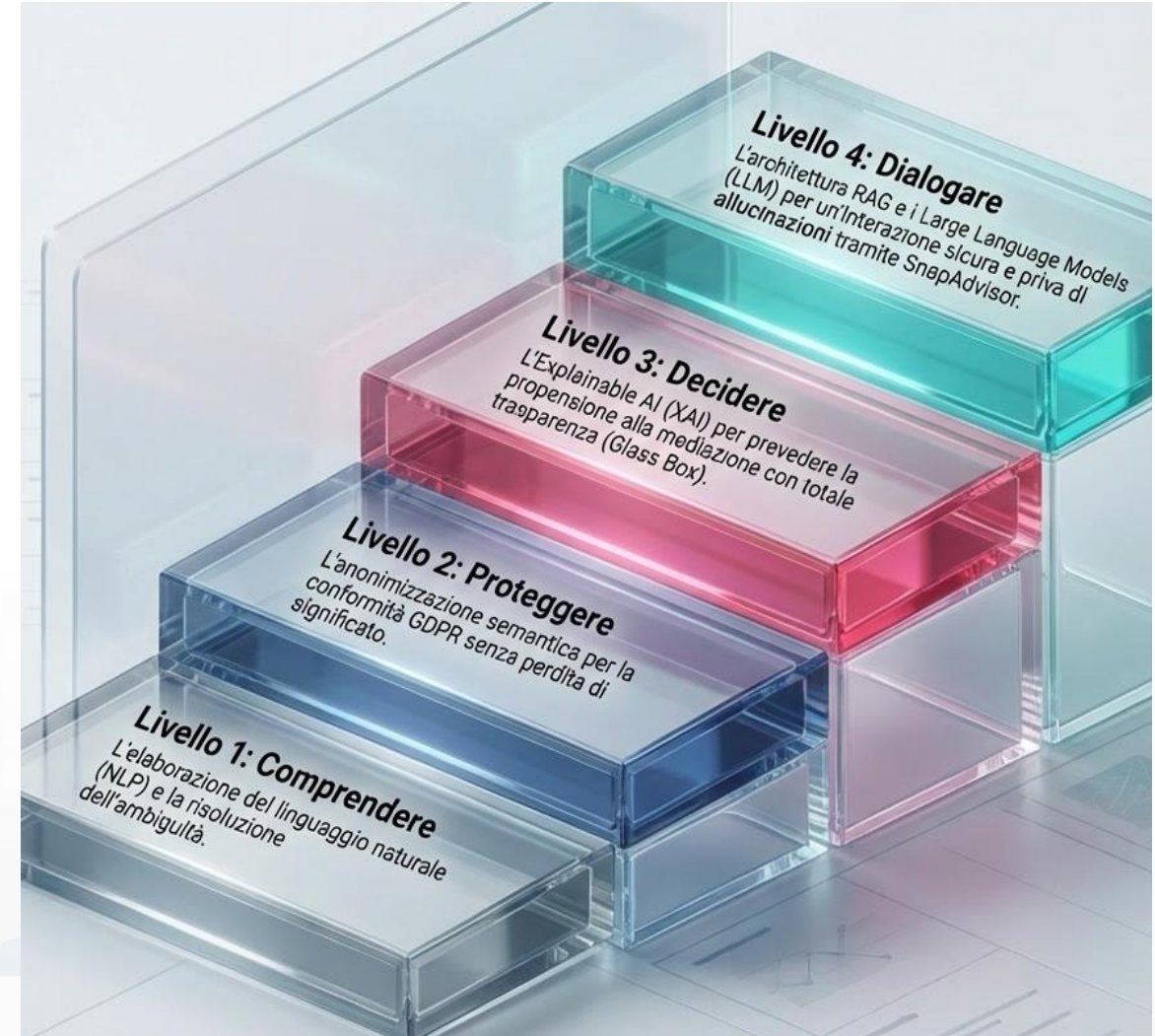


Aspettative

rispondere a domande come un esperto che *studiando il caso* è in grado di fornire suggerimenti

- **referiti** ad affermazioni localizzate nei documenti e fornire spiegazione
- con una **riduzione dei tempi** di analisi di una vertenza,
 - dal punto vista computazionale è instancabile
- **non influenzate** da precedenti documenti,
- **non affetti** da allucinazioni (deduzioni inventate)
- **Con minore varianza** dovuta alle variazioni di contesto, al quale gli umani sono sensibili.

→ *Decision Support System*



La Sfida

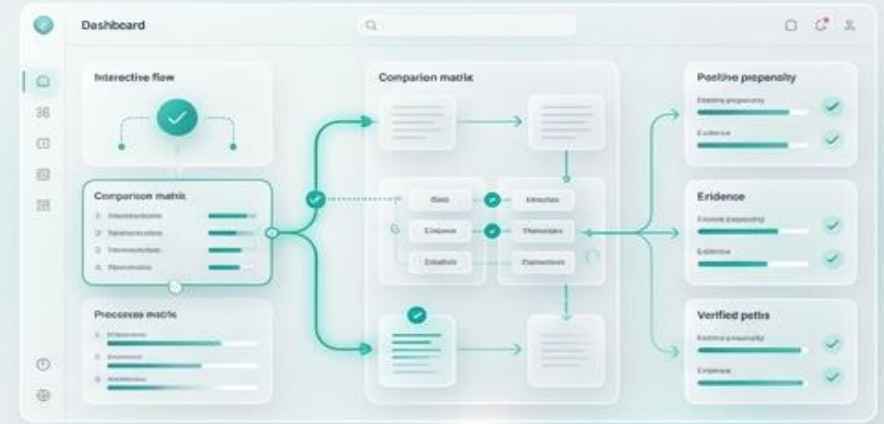


Analisi manuale di voluminosi documenti relativi al contenzioso.

Comprensione profonda del linguaggio naturale e del contesto giuridico.

Vulnerabilità umane: affaticamento e varianza dovuta alle variazioni di contesto.

L'Obiettivo



Creare un assistente computazionale instancabile.

Rispondere a domande mirate (es. Che tipo di rischio si ha?, Vi sono altre strutture coinvolte?)

Fornire suggerimenti basati su affermazioni localizzate nei documenti.

Requisito Fondamentale: Non influenzato da documenti precedenti, non affetto da allucinazioni (deduzioni inventate).



Ambiguità dei linguaggi naturali (1)

Ambiguità Lessicale

Stessa parola, funzione diversa.
«attacco» (verbo vs. sostantivo)



Ambiguità Strutturale

La sintassi crea doppi sensi.
«Ieri ho visto l'uomo col telescopio»
(Chi ha il telescopio?)



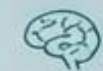
Ambiguità Semantica

Stessa parola, significato divergente.
«acuto» (persona intelligente vs. tipo di suono)



Ambiguità Pragmatica

L'intenzione comunicativa dipende dal ricevente.
«Se Buffon non gioca contro la Spagna, l'Italia perderà.»



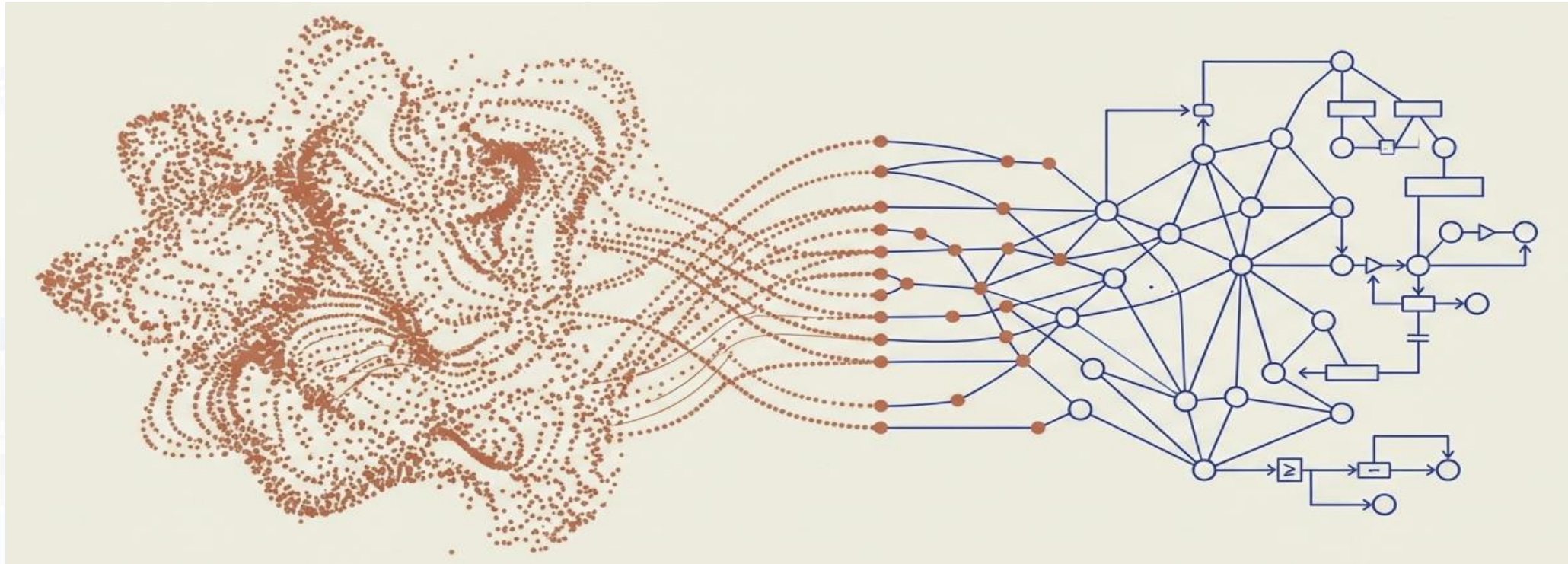
Interpretazione **Emotiva** (assenza psicologicamente fondamentale) vs. Interpretazione **Referenziale** (squadra fisicamente più debole).

Syntax Tree: Parsing del Linguaggio Naturale



L'approccio moderno (BERT) unisce questa rigida struttura sintattica al Machine Learning per comprendere il dominio specifico di conoscenza.

Obiettivi



Soluzioni AI sul linguaggio naturale

- **Creazione di report di sintesi su Vertenze, Documentazione privata**
 - Evoluzione temporale degli accadimenti, Causalità degli eventi, ordinamento e causa/effetto
 - Riferimenti ad articoli o a documenti
 - Ricerca delle evidenze su materiale multimodale
 - Etc.
- **Valutazione**
 - della propenzione alla mediazione, dai documenti di vertenze
 - delle richieste di risarcimento
 - della documentazione da analizzare per una perizia

Per esempio: Propensione alla Mediazione

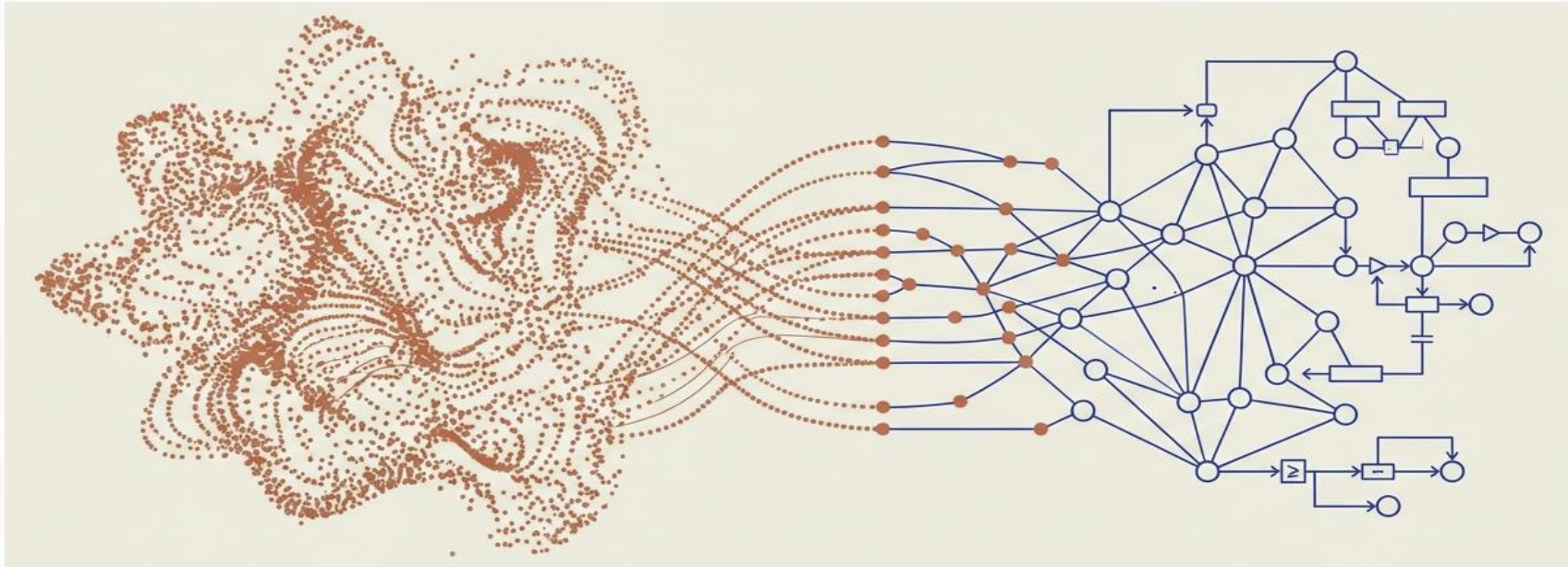
- **Probabilità che una vertenza**
 - possa essere effettivamente mandata in mediazione
 - *Vi sono delle condizioni specifiche per quali la mediazione viene esclusa*
 - sia mandata in mediazione da parte del giudice
 - *Il giudice valuta sulla base della sua esperienza*
 - Se mandata in mediazione venga risolta con successo: quando lo sappiamo?
 - *Se le parti effettivamente trovano un accordo*
 - *Non sempre si è riesce a conoscere il risultato se positivo*

Per esempio: stima della Propensione alla Mediazione

- **Passi:**

- **Modelli di Anonimizzazione** orientati alla comprensione
- **Analisi dei documenti** del fascicolo digitale
- **Stima della propensione** alla mediazione delle parti sulla base del contenuto del testo, come suggerimento per la procedura
- **Supporto al procedimento** per decidere o meno l'invio in mediazione
- **Presentazione delle motivazioni alla base della stima:** porzioni di testo rilevanti per la stima

Anonimizzazione necessaria o meno ?

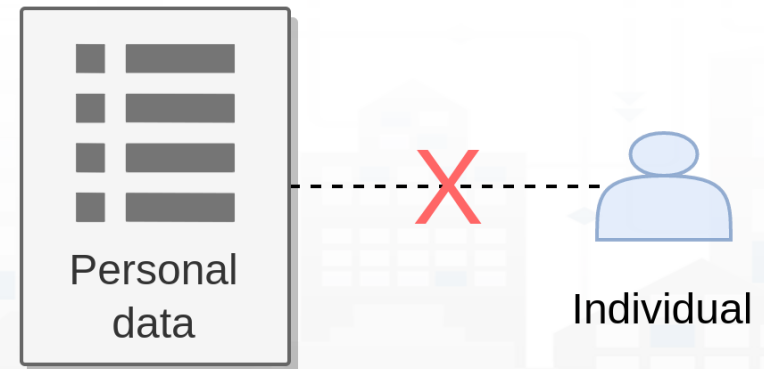


Anonimizzazione

- **La rimozione dei dati personali è un processo utile per i seguenti motivi:**
 - Tutelare la privacy dei soggetti coinvolti
 - Realizzare un dataset immune ad adversarial attacks, ovvero al recupero dei dati di addestramento
 - Realizzare un dataset privo di bias, ovvero non influenzato da informazioni personali
 - GDPR che dice ?
- **Esempi**
 - **Forza bruta:**
 - Mario Rossi è figlio di Marco Rossi. Marco Rossi ha invaso il giardino di Giuseppe Bianchi.
 - Xxxx è figlio di Xxxx. Xxxx ha invaso il giardino di Xxxx.
 - **Conservazione della Semantica:**
 - Mario Rossi è figlio di Marco Rossi. Marco Rossi ha invaso il giardino di Giuseppe Bianchi.
 - Act1 è figlio di Act2. Act2 ha invaso il giardino di Act3.

Anonymization methods

- Remove any link between data and individuals
- Suitable for anonymizing documents
 - The information of interest is in the context rather than in the personal data.
 - Beware of loss of semantics!
- Re-identification is **impossible**
 - Data are no longer protected by GDPR
 - Data can be published without concern



Re-identification

Transparent Digital Dossier

Dati Medici

Etnia
Data della visita
Diagnosi
Procedure
Farmaci
Costi

CAP (ZIP)
Sesso
Data di nascita

Liste Elettorali

Nome
Indirizzo
Data di registrazione
Affiliazione politica
Ultimo voto

Combinando dati di pubblico dominio, l'87% della popolazione è identificabile unicamente tramite: <CAP, Sesso, Data di Nascita>.

**I dati non de-identificati correttamente espongono a rischi GDPR severi.
La re-identificazione è una minaccia costante.**

Evolution of Data Redaction: From Brutal Force to Semantic Preservation

Mario Rossi è figlio di Marco Rossi.
Marco Rossi ha invaso il giardino di
Giuseppe Bianchi.



Rimozione / Forza Bruta

OMISSIS è figlio di OMISSIS. OMISSIS ha invaso il giardino di OMISSIS.

Perdita totale della semantica. Inutile per l'addestramento AI.



Tagging Semplice

#PERSON è figlio di #PERSON. #PERSON ha invaso il giardino di #PERSON.

Trascura le relazioni tra le entità. Chi ha invaso il giardino?



Tagging Enumerato

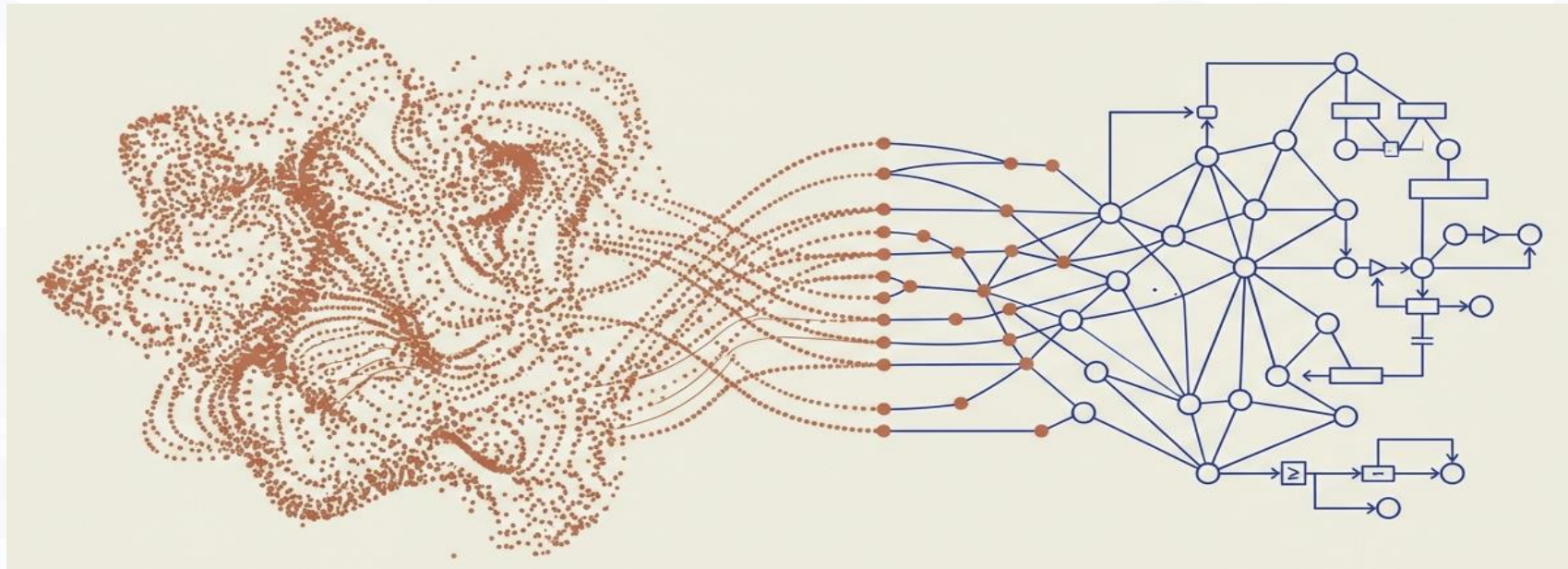
#PERSON_1 è figlio di #PERSON_2. #PERSON_2 ha invaso il giardino di #PERSON_3.

Il testo rimane comprensibile. La semantica è preservata per il Machine Learning (Identity Management).

La Diagnostica dell'Anonimizzazione

| | Rimozione Dati (Brute Force) | Tagging Standard | Tagging Enumerato |
|---------------------------------|---------------------------------|------------------|---|
| Conformità GDPR | Sì | Sì | Sì (Immune ad adversarial attacks) |
| Preservazione della Semantica | Persa | Parziale | Eccellente (Unbiased dataset) |
| Utilità per il Machine Learning | Nulla | Basso | Ideale (Mantiene relazioni logiche) |
| Livello di Complessità | Bassa | Media | Alta (Richiede modelli NLP avanzati per la classificazione) |

Apprendimento



Anonimizzazione per l'apprendimento

- **Anonimizzare in modo che sia preservata la semantica**
 - Si usano soluzioni di identity management
 - processo focalizzato a rimuovere le informazioni dei soggetti coinvolti (nomi, indirizzi email, date) per ottenere un dataset unbiased e normalizzato
 - un documento anonimizzato può non essere utile per altri scopi (es: pubblicazione)
 - Produzione di una data set di vertenze che abbiano:
 - Testi anonimizzati in modo semantico
 - Risultati noti della propensione alla Mediazione
- **Procedimento per anonimizzazione semantica**
 - modello di analisi del testo con cui poter anonimizzare automaticamente i documenti
 - Il modello identifica le occorrenze di dati da rimuovere (es. Nome e Cognome, Codice Fiscale, Luogo, ...)
 - occorrenze individuate sono sostituite con una etichetta anonima (es. #NOME, #CF, #LUOGO, ...)
- **Esempio:**
 - Mario Rossi è nato il 10/04/1990 a Firenze >>>> #NOME è nato il #DATA a #LUOGO

Modello AI vs Training

1. Ingestion (Documenti Fascicolo)

L'input di N documenti complessi e multimodali.



2. Preparazione (Anonimizzazione Semantica)

Conversione in #PERSON_1, #ORGANIZATION_1 per garantire la privacy (GDPR).



3. Inferenza (Rete Neurale)

Il modello BERT/LLM valuta le singole frasi, il contesto e la causalità degli eventi.



4. Output (DSS Report)

Predizione: Propensione (es. 0.00%), Non Propensione (es. 0.01%), Neutro (es. 99.99%).

Spiegazione (XAI): Evidenza delle parole influenti e riferimenti.



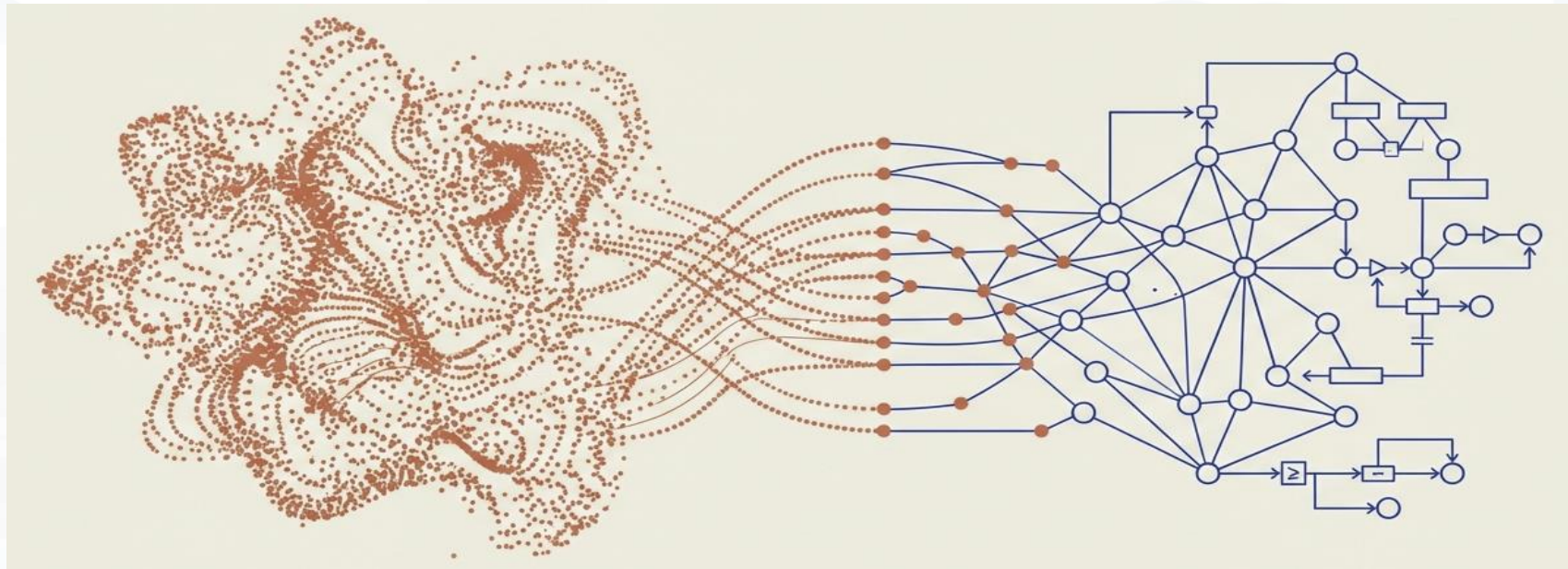
XAI Explained

Sample text: **influential words** want imere
influential words and semse negative words
 negative next in the **negative words** ...

Classificazione delle frasi

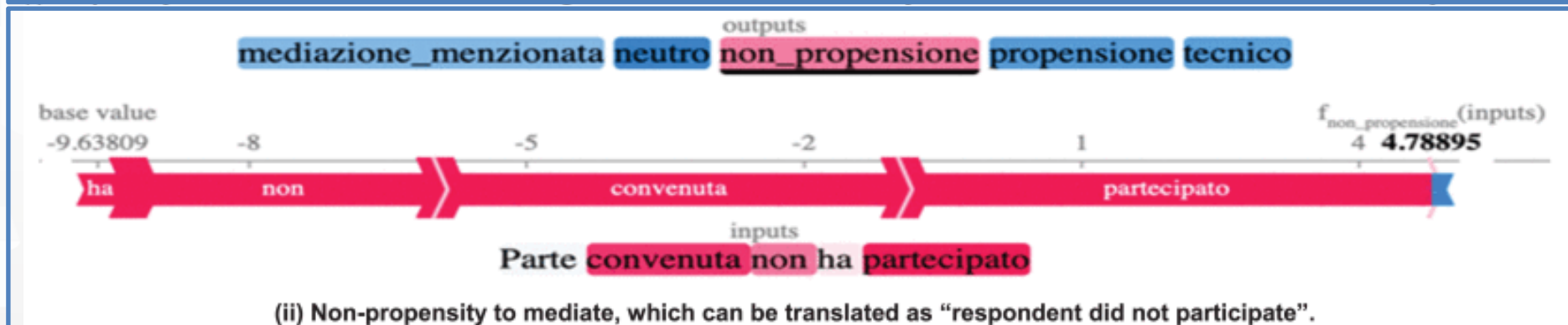
- **Mediazione menzionata:**
 - Sono esplicitamente menzionati documenti di mediazione
 - La mediazione viene disposta dal giudice
 - Viene dichiarata una mediazione in atto
 - Il giudice constata che la mediazione non è stata ancora esperita
 - Le trattative di mediazione sono fallite: non è detto che le parti non fossero propense però
 - Il giudice rileva che la materia del processo è a mediazione obbligatoria
- **Propensione:**
 - Le parti chiedono al giudice il permesso di avviare trattative
 - Entrambe le parti sono assenti al processo
 - Le parti si dichiarano disponibili a intraprendere la mediazione
 - L'attore è assente al processo: accade quando le parti stanno mediando
 - Il processo è dichiarato estinto
- **Non propensione:**
 - Indisponibilità delle parti a mediare
 - Il convenuto è assente al processo: rinuncia a difendersi e non ha interesse a seguire il processo
- **Tecnico:**
 - Presenza di un CTU
 - Richiesta di un CTU
 - Presenza di un CTP
- **Neutro:**
 - Non appartiene a nessuna delle altre classi

NLP vs XAI

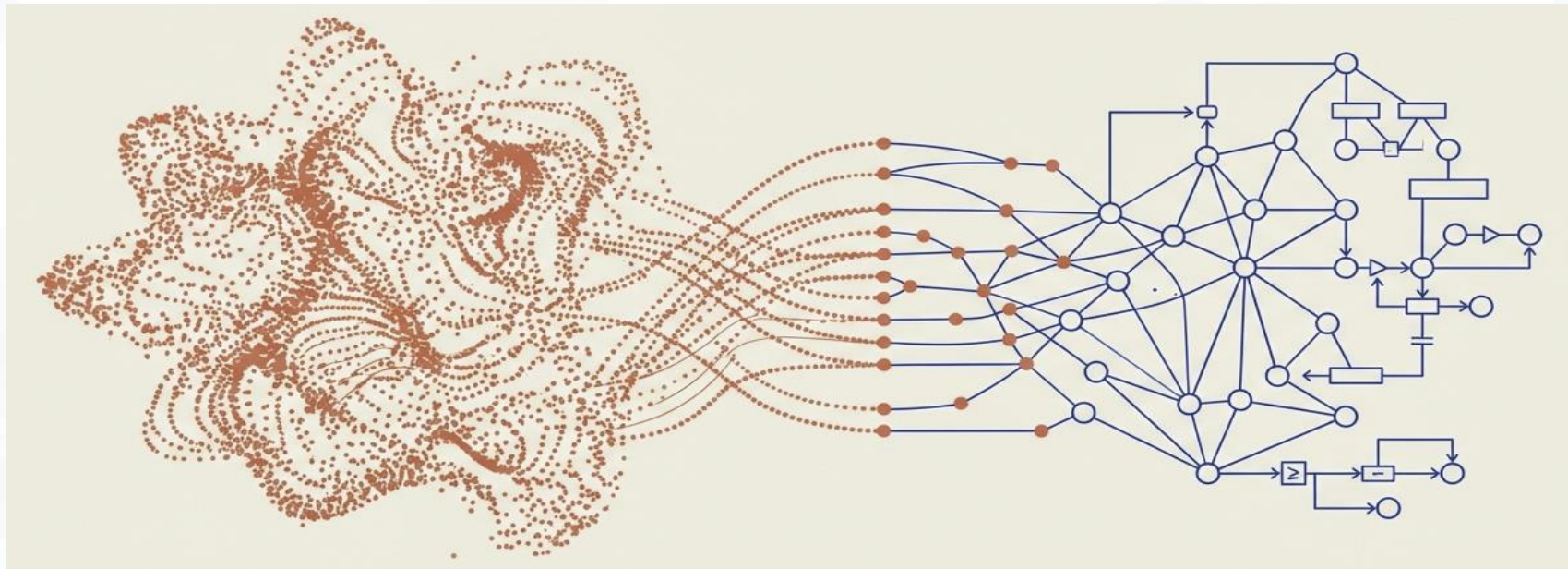


Explainable AI, global/local

- In passo ulteriore è fornire non solo una classificazione ma anche una spiegazione della classificazione delle singole frasi.
- Per questo si utilizzano tecniche di XAI (global e/o local)



Da NLP a LLM



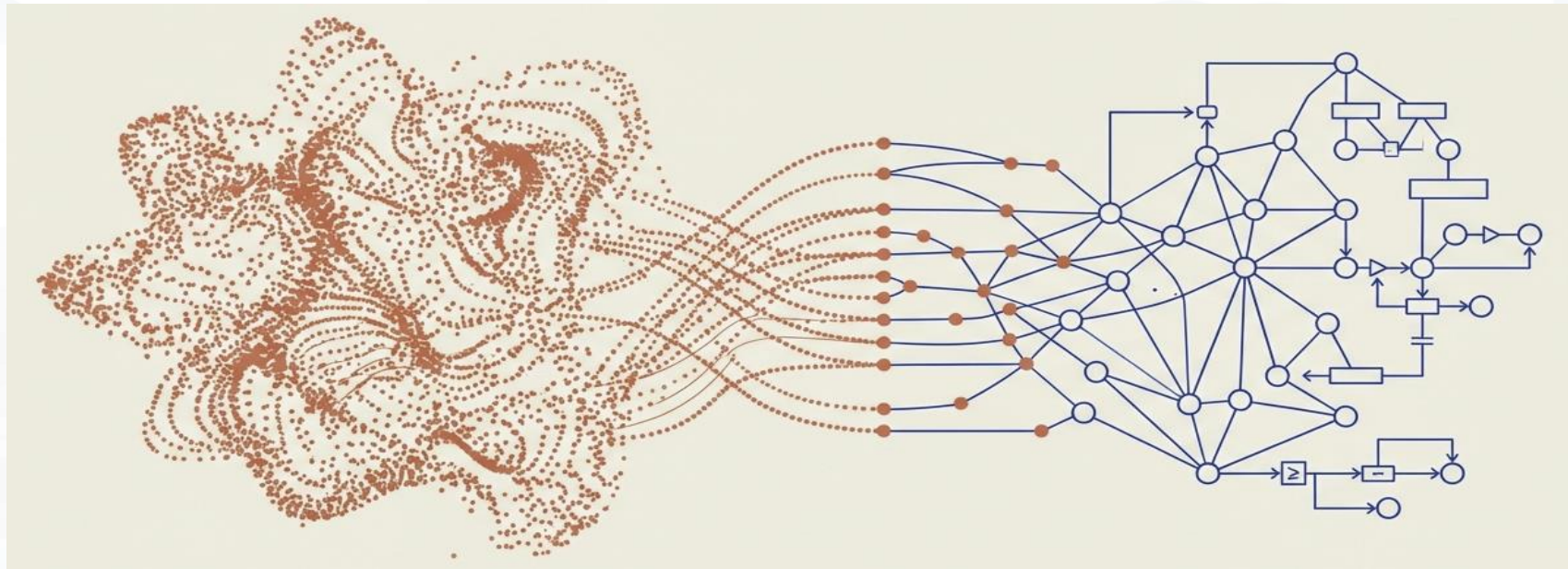
La rivoluzione dei Transformers (2017) (2/3)

- Hanno sostituito le RNN andando a introdurre il meccanismo di «Attention» per pesare la rilevanza delle parole nelle frasi
- **BERT (Bidirectional Encoder Representations from Transformers) (2018),**
 - BERT: uno dei primi modelli basati su Transformers a ottenere risultati straordinari per i classici compiti di NLP:
 - riconoscimento, classificazione, disambiguazione, etc.
- **GPT (Generative Pretrained Transformer):** introdotto da OpenAI
 - tra i primi a passare a modelli di auto-regressione per generare testo.
 - **Generazione di risposte in linguaggio naturale**
 - La generazione di risposte, *una sorta di predizione a partire da una sequenza data di parole, il prompt, la richiesta.*

Da BERT a LLM (3/3)

- I LLM, Large Language Model sono l'evoluzione dei **GPT** (e.g., chat **GPT-4**), e si distinguono per la loro scala in termini di:
 - **Numero di parametri:** i LLM hanno bilioni di parametri, per catturare informazioni molto più complesse e dettagliate sul linguaggio.
 - **Dimensione del dataset:** i LLM sono addestrati su dataset enormi, spesso contenenti centinaia di Gbyte-TByte di testo provenienti da varie fonti (libri, articoli, pagine web).
 - **Capacità zero-shot e few-shot:** i LLM, grazie alla loro scala, possono eseguire compiti complessi:
 - zero: produzione di risposte sulla base della loro conoscenza di base
 - few: produzione di risposte sulla base di conoscenza aggiuntiva
 - **Alcune architetture di LLM continuano a imparare** ma magari non è sempre cosa buona, anche perché si rischia di virare il modello, e dimenticare...

LLM → RAG-LLM → Agentic



Il Lato Oscuro: Allucinazioni e Rischi Professionali

Precedente: Mata v. Avianca, 2023.
La corte stabilisce che...

La corte stabilisce che di nei
continuzione su ressenire la
precedente, tanisa mpremada,
courtra voido e response fonnati di
contrattualioso.



- **Il Problema:** L'AI genera citazioni, precedenti o fatti plausibili ma inesistenti. Tasso di allucinazione del 16.7% nelle ricerche generaliste.
- **Il Rischio:** Sanzioni disciplinari severe per il deposito di memorie inventate.
- **La Regola d'Oro:** L'AI è un assistente, non un'autorità. La verifica manuale sulle fonti primarie ufficiali è un obbligo deontologico assoluto.

L'Infrastruttura Sicura: LLM vs. RAG

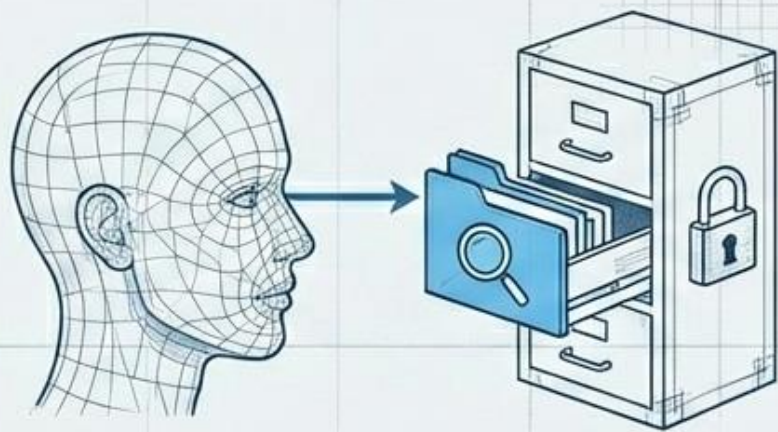
Solo LLM (Modelli Generici)



Ottimo per ideazione e riassunti generici.

- **Criticità:** Rischio elevato di allucinazioni, dati non tracciabili, mancanza di fonti primarie.

RAG (Retrieval-Augmented Generation)



Le risposte sono vincolate a documenti esterni e privati caricati dall'utente.

- **Vantaggio:** Tracciabilità assoluta delle fonti (Evidence Traceability).
- Azzeramento delle allucinazioni 'inventate' e aggiornamento normativo in tempo reale.

Comparison LLM vs RAG-LLM

Usa un LLM Generico Quando...

L'obiettivo è una spiegazione generica o l'ideazione creativa.

I fatti necessari sono già forniti nel prompt dell'utente.

Non sono richieste citazioni o tracciabilità delle fonti.

La velocità di setup è prioritaria rispetto alla precisione assoluta.

Usa l'Architettura RAG Quando...


Le risposte dipendono da documenti esterni e specifici.

Il corpus di dati è privato, voluminoso o in evoluzione.

La tracciabilità delle prove (citazioni) è fondamentale.


Lo stesso compito di conoscenza si ripete su larga scala con utenti diversi.

Logged in as:
edoardo_tester

 Your Profile

Logout

SnapAssistant model:

Llama3.3 

New Chat

EdoChat1

EdoChat2

EdoChat3

EdoChat4

prova

chatNuova

chatTest1LLAMA

chatTest2LLAMA

testGianniLLAMA

testTastoInvioPreChanges

ChatTestPostPSWINVIO


<https://www.snap4city.org/1116>


SnapAdvisor

Current chat: debug0

Visualize references

Personal
Advisor LLM

 Come creo una dashboard?

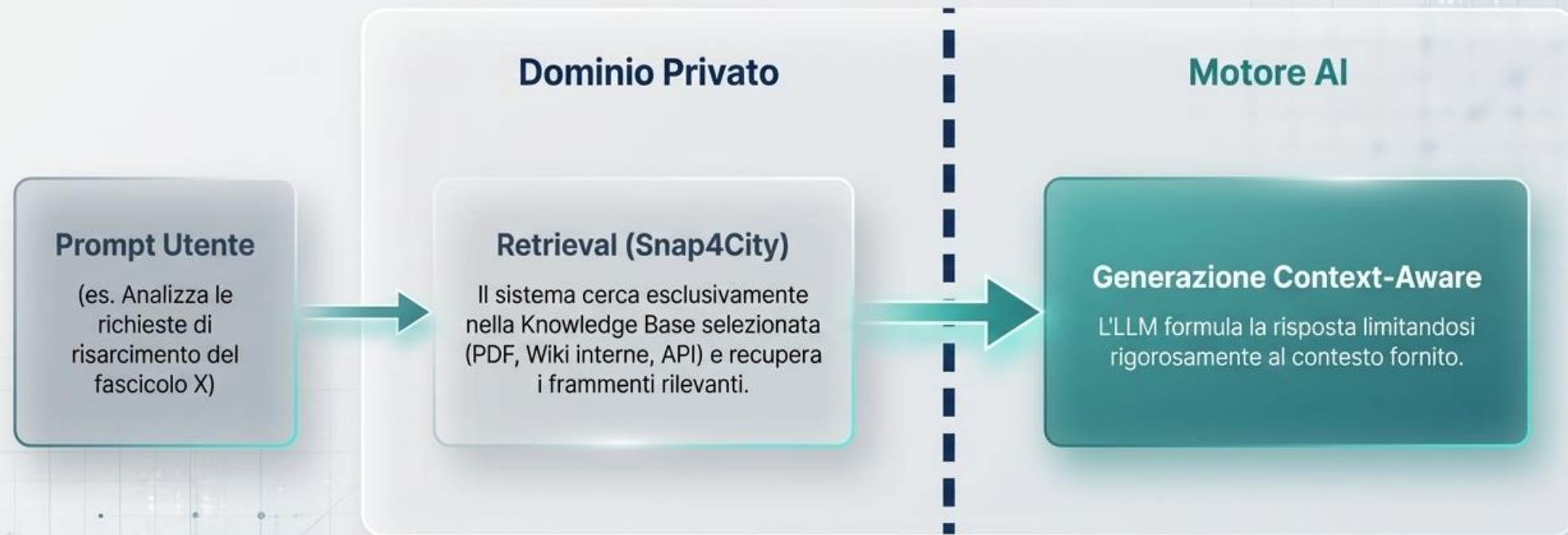
 Per creare una dashboard, puoi utilizzare la piattaforma Snap4City. La dashboard si compone di un insieme di widget grafici che possono essere impostati separatamente assegnando una serie di parametri, come origine dati, dimensioni, colori, forma, font, allarmi, relazioni con altri, ecc. (SOURCE: Snap4 City per la creazione di Smart City Control Room, SCCR). Inoltre, la piattaforma offre la possibilità di creare vere e proprie applicazioni interattive, multipagina/multidashboard, e non semplici storie monopagina da scrollare. Puoi anche condividere la dashboard con altri utenti anche fuori dalla Smart City Control Room tramite canali autenticati e protetti in HTTPS, in accordo al GDPR (SOURCE: Snap4 City per la creazione di Smart City Control Room, SCCR). Per vedere alcuni esempi di dashboard, puoi visitare il sito <https://www.snap4city.org/dashboardSmartCity/view/index.php?iddashboard=MjU0Mg==> (SOURCE: Snap4 City 1st Overview v4 8 annot). Tuttavia, non ho trovato informazioni dettagliate sulla procedura di creazione di una dashboard. Se hai bisogno di ulteriore assistenza, ti consiglio di contattare direttamente il team di supporto di Snap4City. Non so dirti altro in merito.

Write a message...



Architettura RAG di SnapAdvisor: Flusso di Conoscenza Controllato

Separazione sicura tra dati privati e motore AI per risposte precise e verificabili.



Controllo del Dominio:
L'utente decide quali documenti interrogare.

Spiegabilità:
Ogni risposta include riferimenti diretti (SOURCE: Doc X, Pag Y).

Nessuna Interferenza:
Modulare e multi-utente; i dati restano isolati.

SnapAdvisor: AI based Assistants, DSS

- **Snap4CityAdvisor:**
 - for developers and promoters knowing all Snap4City and DISIT Lab literature, articles, manuals, slides, codes, etc. The SnapAdvisor with this skill is accessible only for selected Snap4City users. It allows users of saving a lot of time by providing answers to facilitate the development of smart solutions, to the exploitation of Snap4City tools, and recently on data and services accessible on the platform, producing offers, etc.
- **Legal Advisor**
 - expert on specific disputes at service of the Legal department of Careggi Hospital of Florence. It allows to save time in recovering precise information from complex legal documents, ordering of events, understanding causes and effects, producing reports, etc.
- **Expert of industrial machines, user manuals, technical manuals, rules,**
 - Answering on technical manuals
- **Complains and Questionnaire analysis**
 - Answering on trends, via questionnaires collected via QR, blobs, emails, etc.
- **Commercial Advisor:**
 - processing orders understanding them and preparing the offer, thus reducing the time to process them.
- **Generative Designs:**
 - Multimodal Generative AI supporting designers in producing innovations
- etc.

See: "Context-Aware Retrieval Augmented Generation using Similarity Validation to handle Context Inconsistencies in Large Language Models", IEEE Access, 2025. <https://doi.org/10.1109/ACCESS.2025.3614553>

SnapAdvisor

- **working on your private content** and thus producing answers.
 - Select a subset of documents
- **domain control:** you can decide the knowledge base (internal wikis, PDFs, APIs),
- **explainability** capability, provide references to your documents
- **multilingual via content**
- **modularity:** it is possible to pass from one collection of documents to another, and multiple users can work on the advisor asking for different topic on different collections/domains at the same time, independently as needs change, without any interferences among them.
- **lower hallucinations**

<https://www.snap4city.org/1116>

1 Inizia a Basso Rischio

Implementa l'AI per compiti ad alto volume come l'analisi contrattuale e la revisione documentale, non per la ricerca strategica profonda.

2 Esigi il Framework RAG

Usa solo strumenti che legano le risposte a fonti documentali verificabili per azzerare le allucinazioni.

3 Proteggi il Dato

Applica sempre l'anonimizzazione semantica se non disponi di licenze Enterprise 'closed-loop'.

4 Human-in-the-Loop

Tratta l'AI come uno stagista brillante ma inesperto. La responsabilità finale e la validazione delle fonti sono esclusivamente umane.

The evolution is now: Agent/Agentic AI

