

Archivi e intelligenza artificiale responsabile

Esperienze Opportunità Cautele

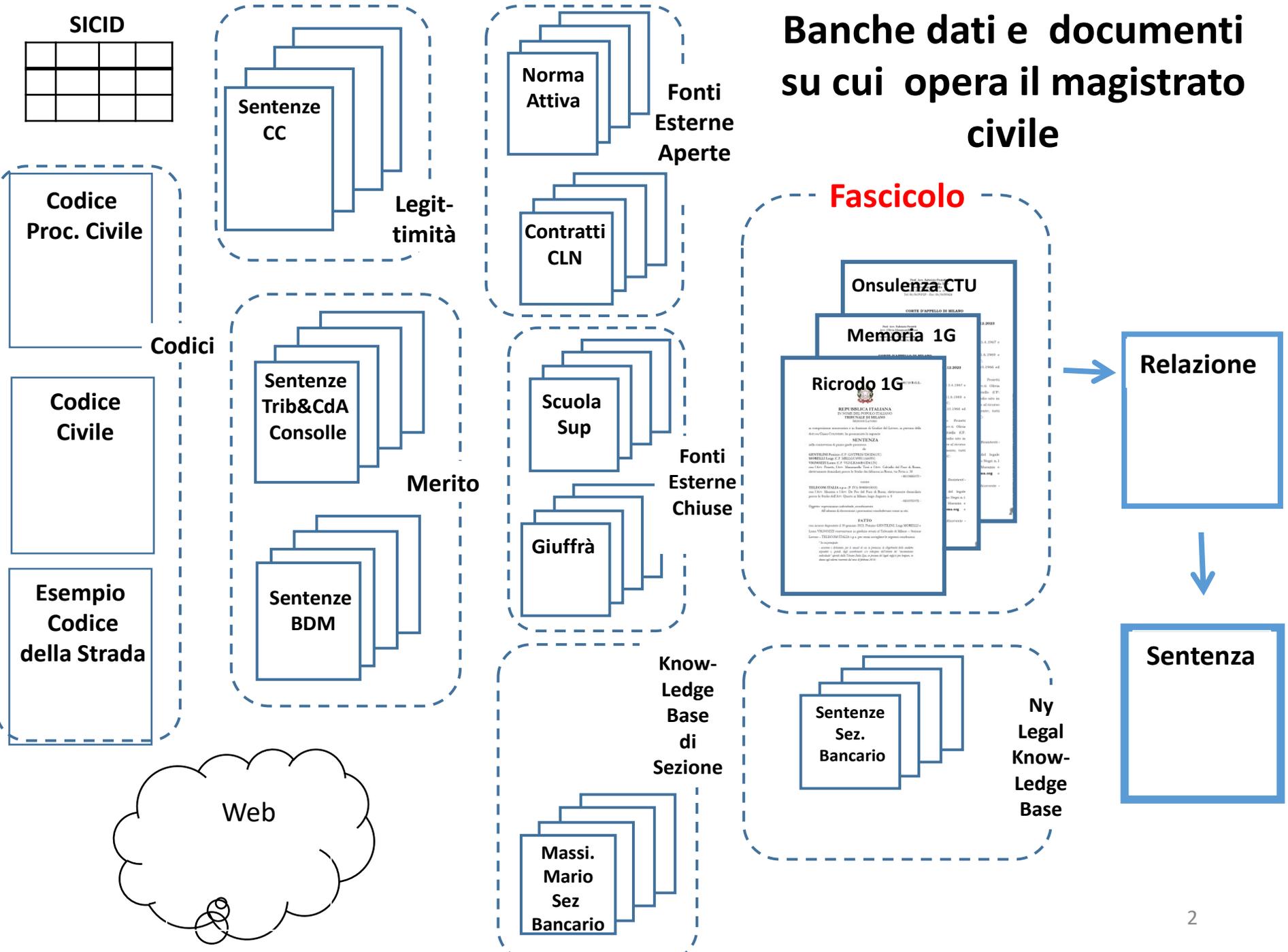
16 Maggio 2025

Carlo Batini

Cini e Università degli Studi di Milano-Bicocca

Verso un linguaggio di comandi cognitivi
basato su modelli linguistici per elaborare
i documenti del fascicolo del processo civile

Banche dati e documenti su cui opera il magistrato civile

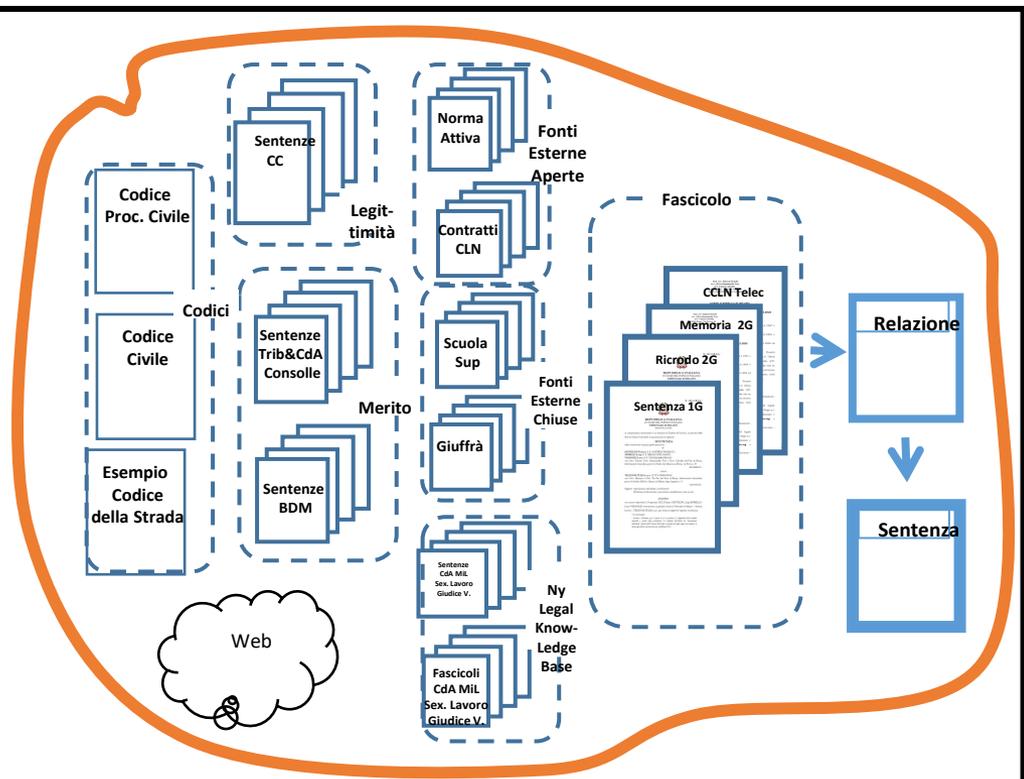


Oggi

Ricerca per parole chiave +
Modelli +
Massimari

Giudice

Word +
Copilot



Nel Sistema del Processo Civile Telematico i testi sono “piatti”



“Tornio lama faglia 13-12-2023 tiro anna fendere
formare tetto giostra Sotto rosa verde testa giro
panda assegno circolare mondo lanica Giorgio
tredici dicembre 2023 sacco azzurro lima martello
testa corpo piedi lancia sposo sposa altare municipio
assegno postdatato mettere tirare assegni circolari
gogna gino mesto misto monade giorno mese verde
testa giro panda leone tigre occhio dente oca verde
tenda toro tigre lancia rosso sposo sposa altare
municipio assegno postdatato mettere tirare otto
settembre 2018 Gesto 56/12/23 Gianni oca finto
toto vista occhio ginocchio scarpa tono allegro misto
tessere girare torto marcio byte siesta dormire
Roma Palermo Catania Azzurro verde arancio mela”

Invece, quando noi umani li leggiamo,
sono pieni di **semantica**



“Tornio lama faglia **Data** 13-12-2023 tiro anna fendere
formare tetto giostra Sotto rosa **Colore** *verde* testa giro
panda **Tipo di assegno** *assegno circolare* mondo lanica
Giorgio **Data** tredici dicembre 2023 sacco **Colore** azzurro
lima martello testa corpo piedi lancia sposo sposa altare
municipio assegno postdatato mettere tirare **Tipo di**
assegno assegni circolari gogna gino mesto misto
monade giorno mese verde testa giro panda leone tigre
occhio dente ocra verde tenda toro tigre lancia rosso
sposo sposa altare municipio assegno postdatato
mettere tirare **Data** otto settembre 2018 Gesto **Data**
sbagliata 56/12/23 **Nome** Gianni ocra finto toto vista
occhio ginocchio scarpa tono allegro misto tessere
girare torto marcio byte siesta dormire Roma Palermo
Catania azzurro verde arancio mela”

.....e di logica

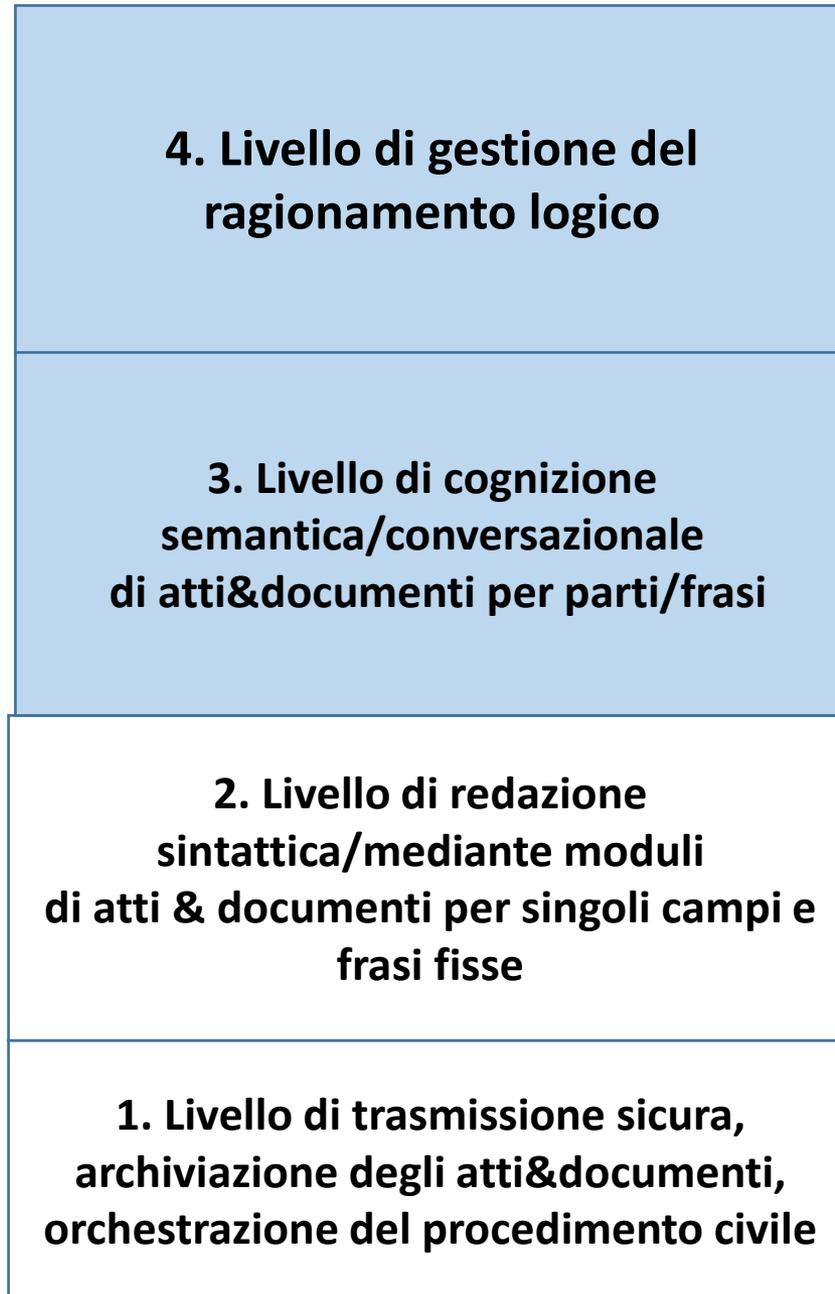
Tutti gli esseri umani sono mortali, Socrate è un essere umano, Socrate è mortale



Premessa1 Tutti gli esseri umani sono mortali,
Premessa2 Socrate è un essere umano,
Conclusione Socrate è mortale

**Scopo del Progetto
DataLake@Giustizia:
sperimentare e
specificare i livelli
3. Semantico
4. Logico
del futuro possibile
Sistema Informativo
del Processo Civile
Telematico**

**Il Processo
Civile
Telematico
oggi**



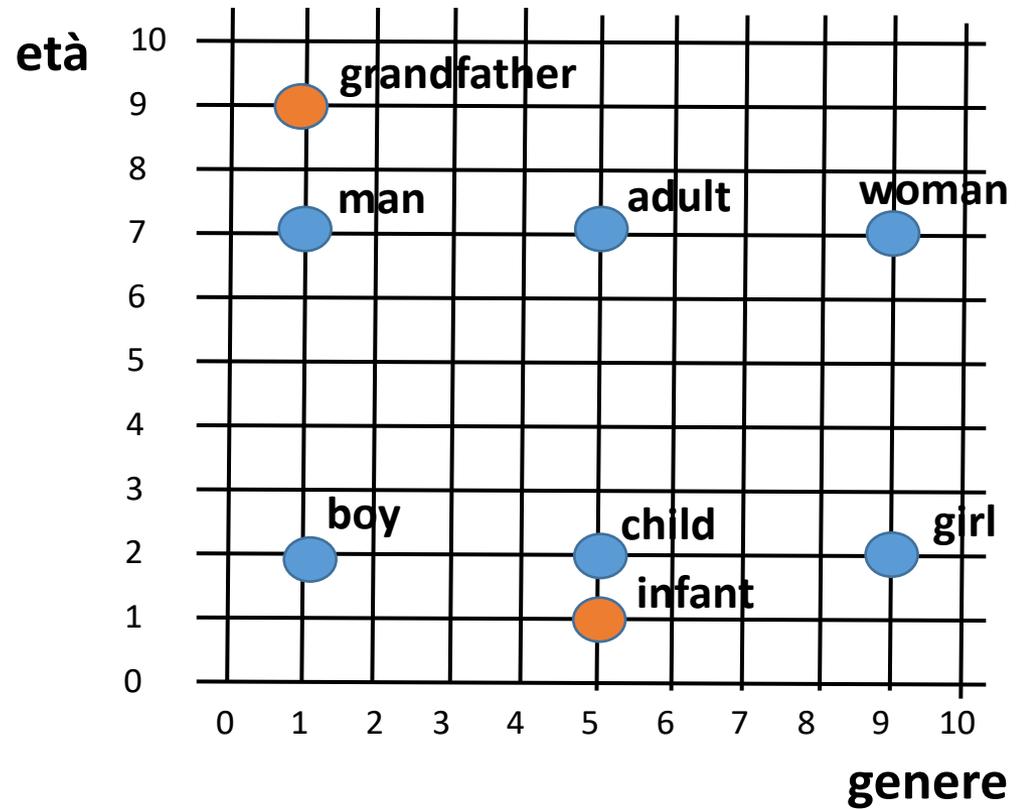
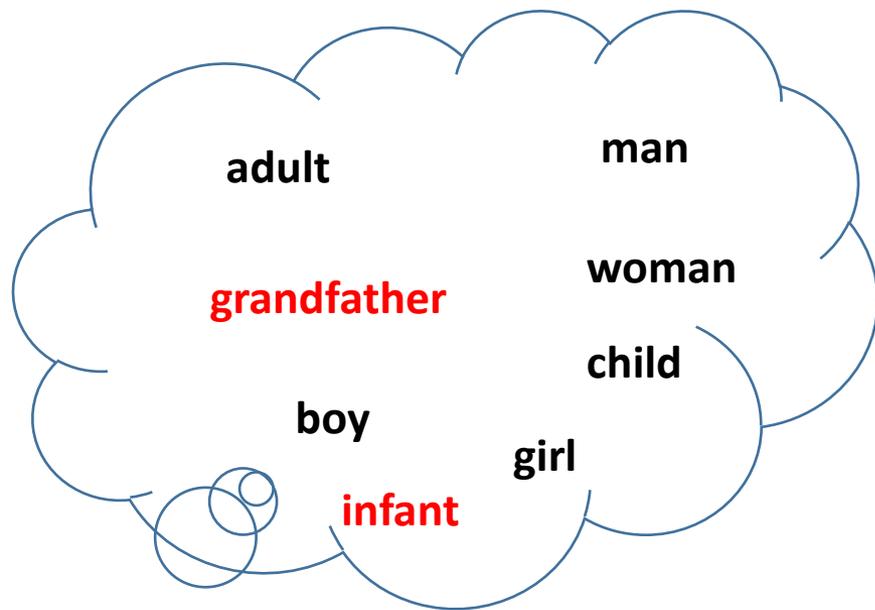
Slsadf dddd
As òda dgdg
Slsadf dddd
Asòda dgdgl
Slsadf dddd
Asòda dgdgl
ISlsadf dddd
Asòda dgdgl



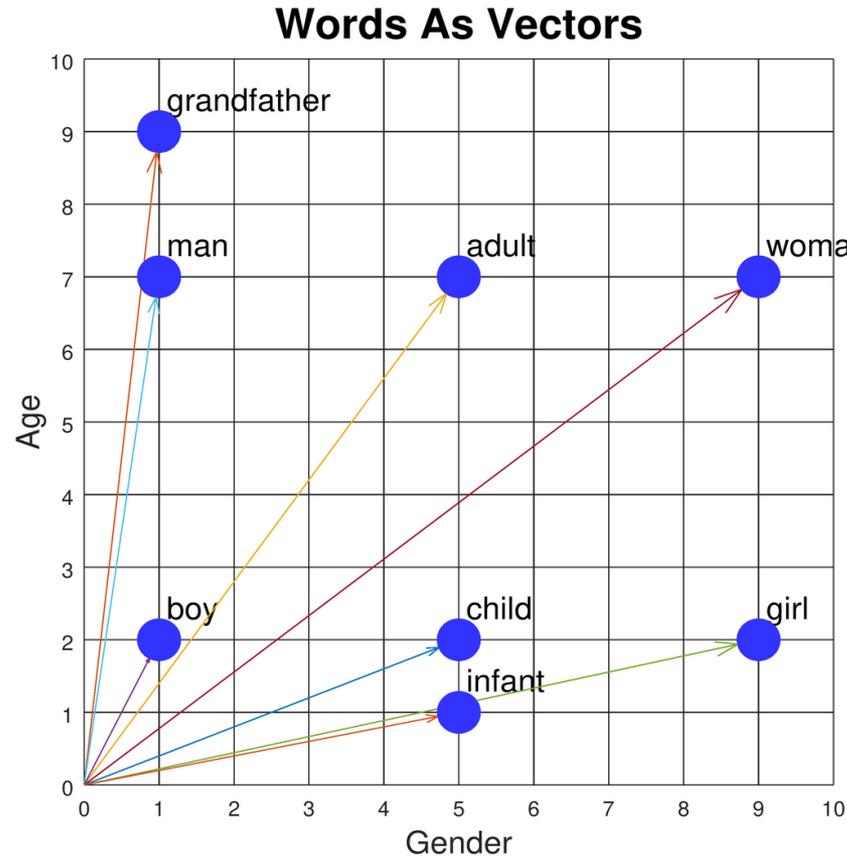
L'Intelligenza Artificiale Generativa

Novembre 2022: nasce ChatGPT

This morning I met in Bologna my old



Le parole sono dunque rappresentate con vettori di numeri, chiamati **word embeddings**



Original Word Vectors	
grandfather	[1, 9]
man	[1, 7]
adult	[5, 7]
woman	[9, 7]
boy	[1, 2]
child	[5, 2]
girl	[9, 2]
infant	[5, 1]

This morning I met in Bologna my old **grandfather**

grandfather

1
3
6
4
9

Child

1
6
3
4
5

old

1
3
4
7
9

Nephew

1
9
9
3
8

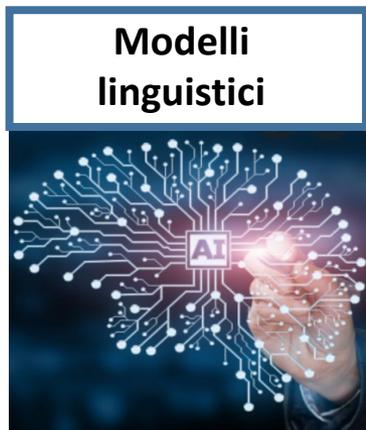
Young

8
3
4
7
2

La delega alla IA Generativa

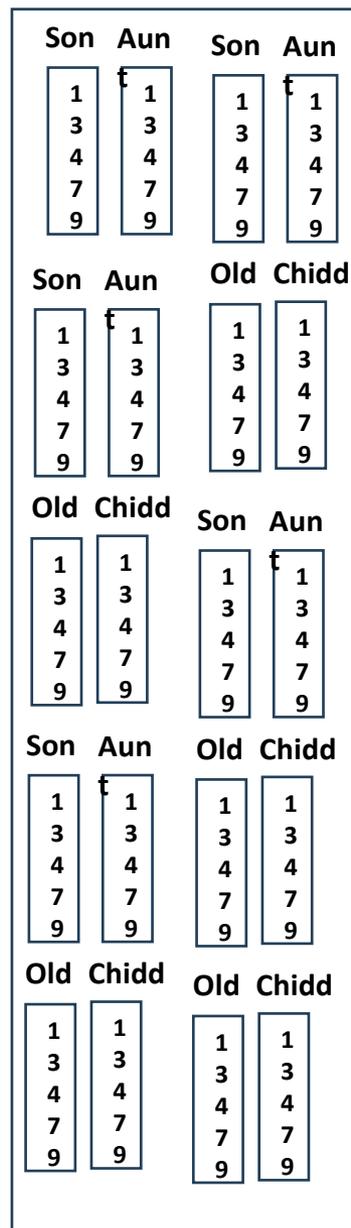
Il linguaggio con cui interveniamo sul mondo.....

Forniscimi nuovi indizi della relazione tra Rossi, Verdi e Neri



Dalla analisi dei testi delle intercettazioni, trascritti con il tool Speech-to-text e da me analizzati, emerge un rapporto di dipendenza emotiva tra Verdi, il capo, e Rossi, l'esecutore.

← Ed è lo stesso con cui i modelli linguistici ci rispondono e intervengono sul mondo



← E' lo stesso, pur con forti semplificazioni, con cui i modelli linguistici rappresentano il mondo

Nella nostra vita noi effettuiamo **azioni** che possono essere espresso da **verbi**, **cose** o (informazioni in) **documenti**

- Mi **porgi** quella **mela**? Prima, **lavala**
- Perfavore, **cerchi** in **archivio** questa **pratica!**
- Devo **riordinare** le **carte** sul **tavolo**
- **Riassumi** questa **memoria difensiva**, è troppo lunga
- Ti **consiglio** di **leggere** quel **libro** su un **eReader**

...e nella nostra professione effettuiamo **attività cognitive** che si possono esprimere con **verbi** o **comandi linguistici**

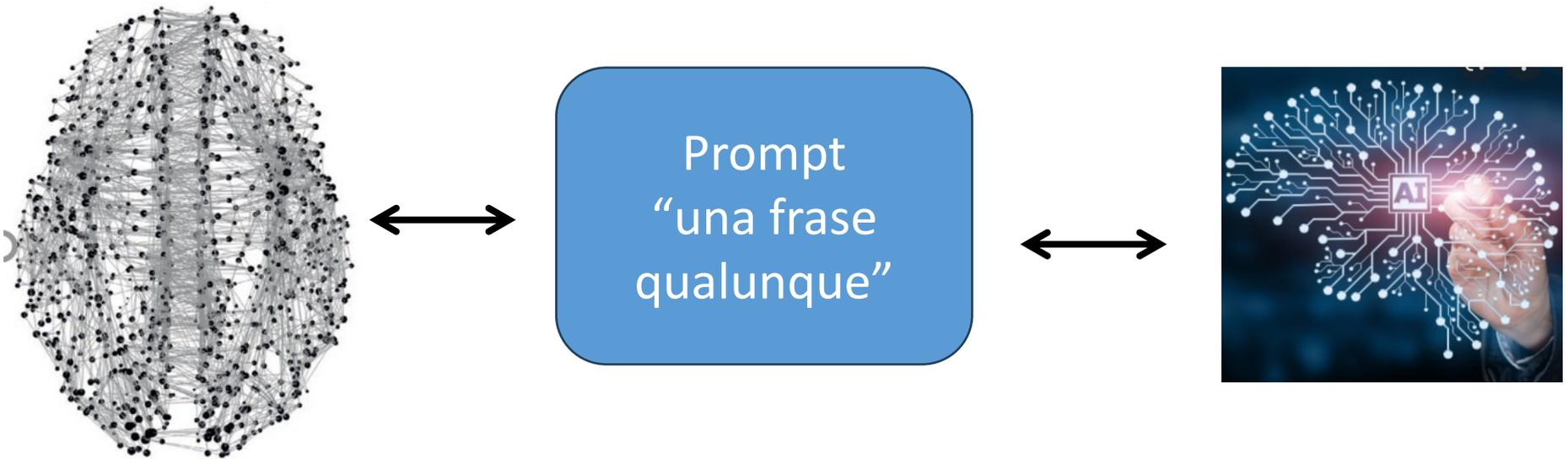
- Cercare
- Riassumere
- Estrarre conoscenza
- Classificare
- Confrontare
- Integrare
- Verificare
- Estendere (inverso di riassumere)
- Calcolare
- Esplorare
- Valutare la rilevanza
- Interpretare
- Dedurre
- Archiviare
- Trasmettere
- Modificare
- Acquisire
- Depositare
- Pubblicare
- Decidere
- Commentare
- Annotare
- Copiare
- Riusare
- Riordinare
-

Il Document Command Language

Un linguaggio che faccia da **intermediatore linguistico** tra utente di banche dasti documentasali, archivi e fascicoli e intelligenza artificiale generativa

Perchè un nuovo Inguaggio?

Perchè I modelli linguistici interagiscono con l'utente per mezzo di frasi dette **prompt**, e in generale non viene posta **nessuna restrizione** al modo con cui concepiamo I prompt



Usare la IA Generativa, tuttavia, presenta rischi, come Allucinazioni e False informazioni

Il meccanismo del deragliamento



AI Act - Regole che tutelano i diritti fondamentali

Accuratezza e Equità

I sistemi di IA devono essere

- a) **tecnicamente robusti** per garantire che la tecnologia sia adatta allo scopo e che
- b) **i risultati falsi positivi/negativi non devono incidere in modo sproporzionato sui gruppi protetti (ad esempio, per origine razziale o etnica, genere, età, ecc.)** → **Accuratezza e Equità**
- c) È molto importante che i sistemi di IA **non creino o riproducano distorsioni (Equità)**.

Se adeguatamente progettati e utilizzati, **i sistemi di IA possono al contrario contribuire a ridurre le distorsioni e le discriminazioni strutturali esistenti**, portando così a decisioni più eque e non discriminatorie (ad esempio, nella selezione del personale).

Papa Leone XIV, 10 Maggio 2025

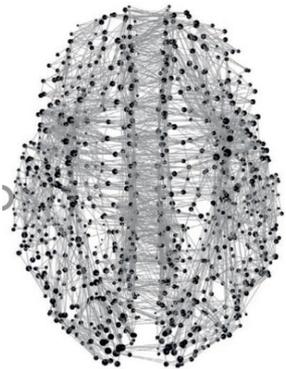
**Il Papa: “Per la Chiesa
la nuova questione
sociale è
l’intelligenza
artificiale”**



Come contribuire a una **IA responsabile** e allineata con l'AI Act? - **Strategie**

- 1. Restringere l'insieme dei comandi linguistici e la sintassi delle frasi con cui interagire con i modelli linguistici**

<u>Cercare</u>	2	<u>Interpretare</u>
<u>Riassumi</u>	3	<u>Dedurre</u>
<u>Estrai</u>		<u>iviare</u>
<u>Classifica</u>		<u>mettere</u>
<u>Confronta</u>		<u>lificare</u>
<u>Integra</u>		<u>isire</u>
<u>Verifica</u>		<u>ositare</u>
<u>Estendi</u>		<u>olicare</u>
<u>di riassumi</u>		<u>dere</u>
<u>Calcola</u>		<u>mentare</u>
<u>Esplora</u>		<u>otare</u>
<u>Valutare la rilevanza</u>	14	<u>are</u>
	15	<u>Riusare</u>
		<u>Riordinare</u>

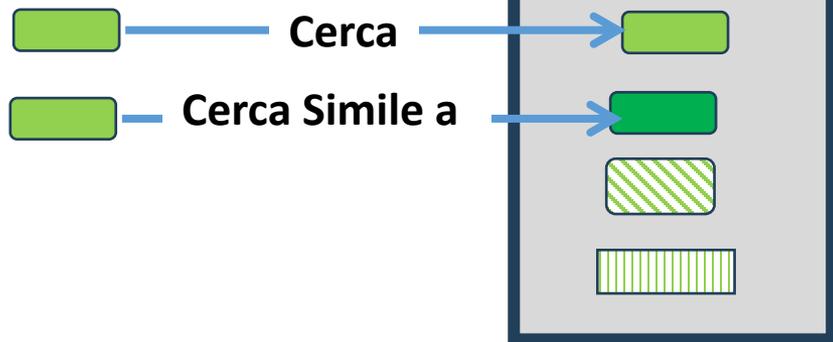


Rappresentazione iconica dei comandi

Cerca, Riassumi, Estrai, Confronta e Integra

 Parole, termini, frasi, periodi

 Il Fatto



Ricorso

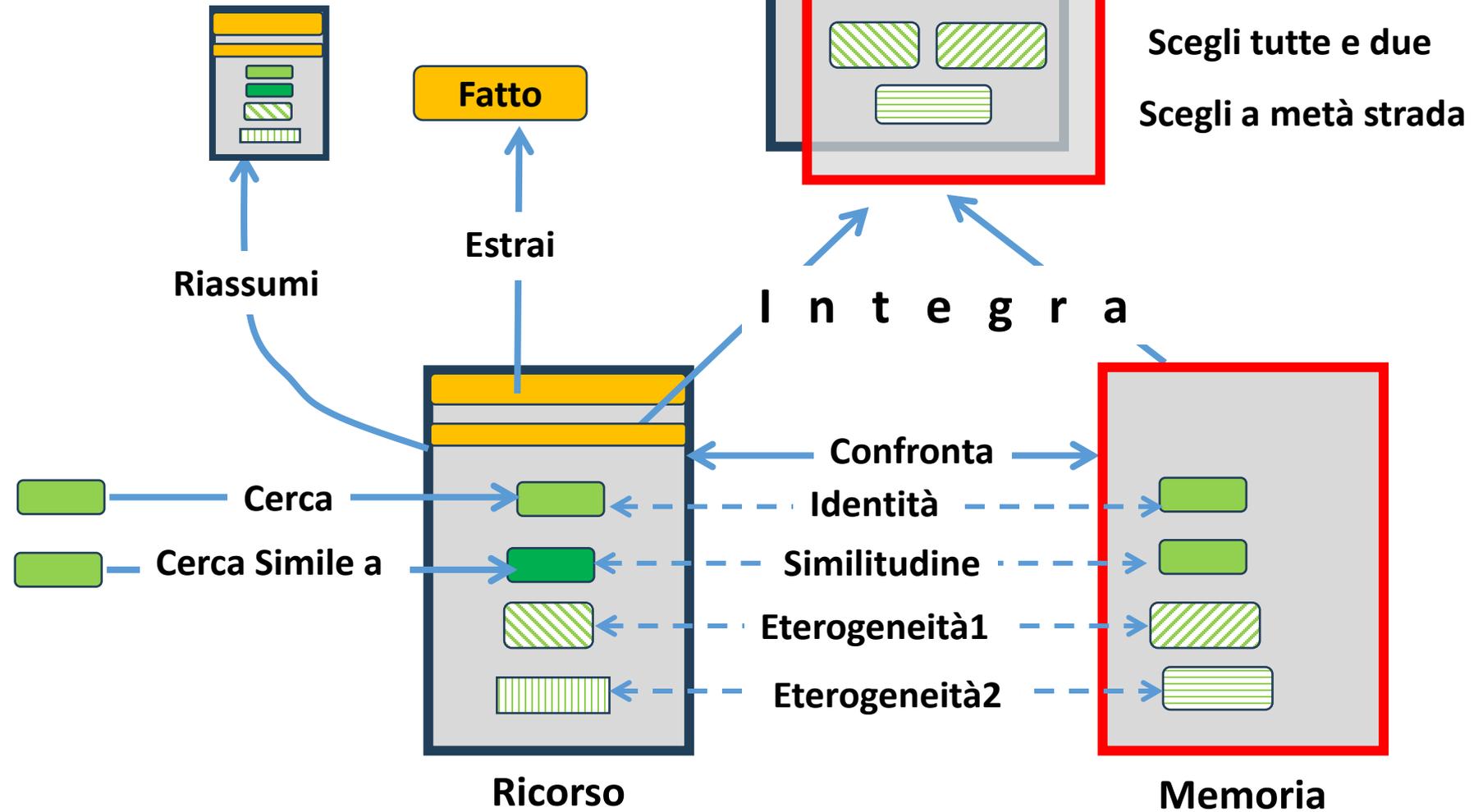
Esempio di Cerca frase simile a

Cerca **nel seguente testo italiano**, le frasi simili a **"assorbimento del superminimo"**

«.....sfsfg..... **assorbimento del superminimo**
.....fjsljalsdìqejfa..... **è stato assorbito il**
superminimo ..sdfjaldfj..... **mantenimento del**
superminimoslòfjqgfgqèerl **assorbibilità**
del superminimosldjfaljfa ...**non assorbimento**
del superimimo"

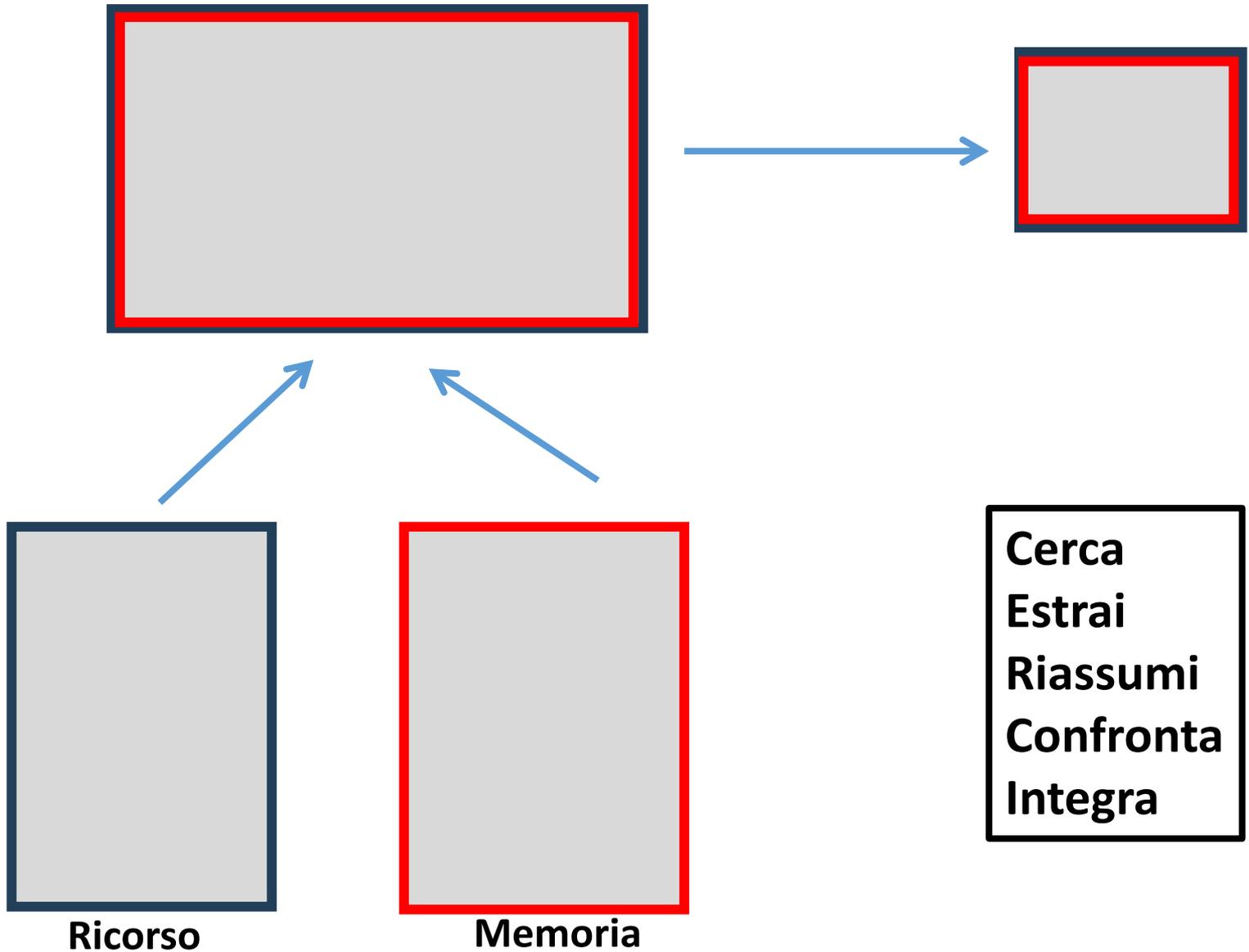
Rappresentazione iconica dei comandi Cerca, Riassumi, Estrai, Confronta e Integra

- Parole, termini, frasi, periodi
- Il Fatto

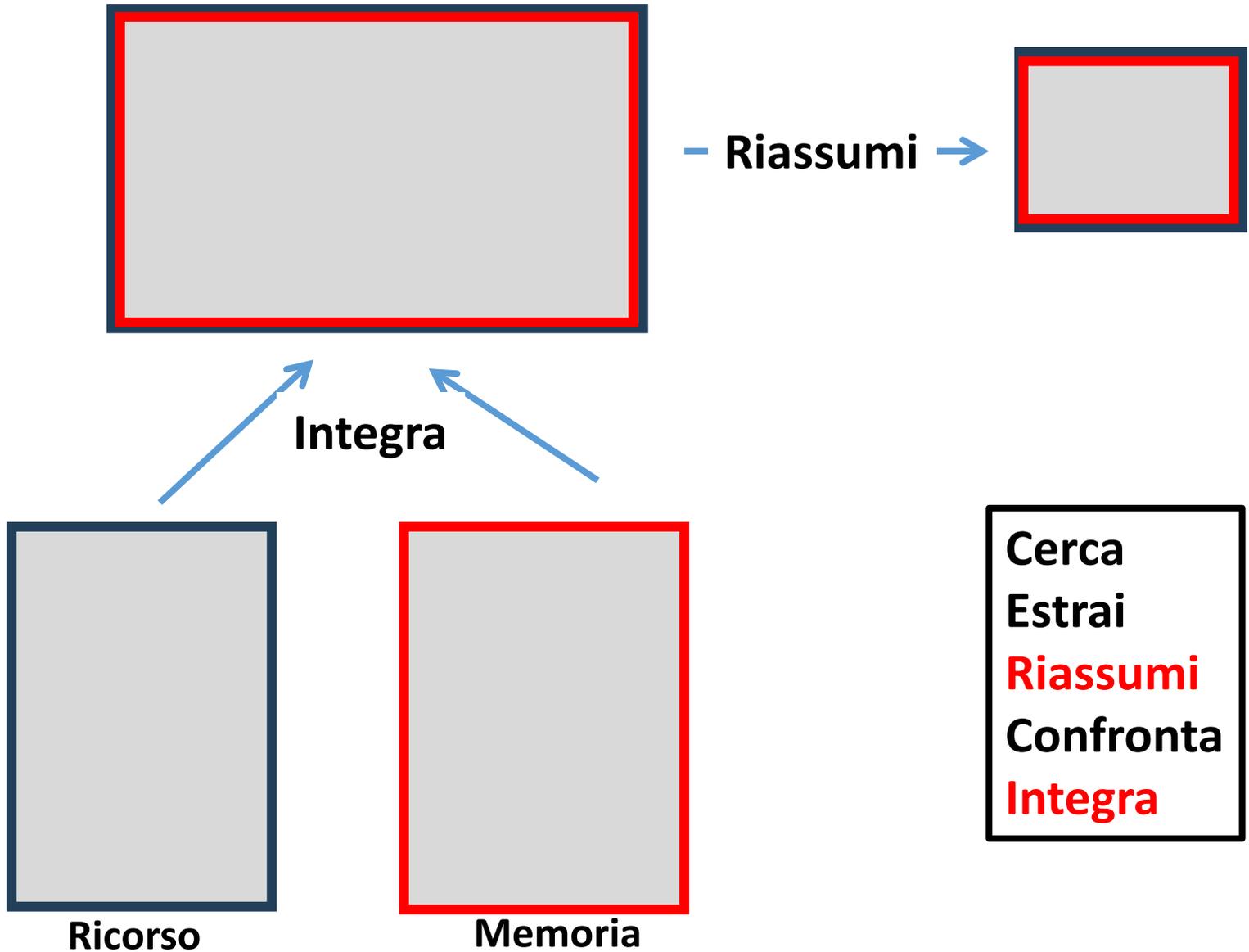


Ora provate voi!

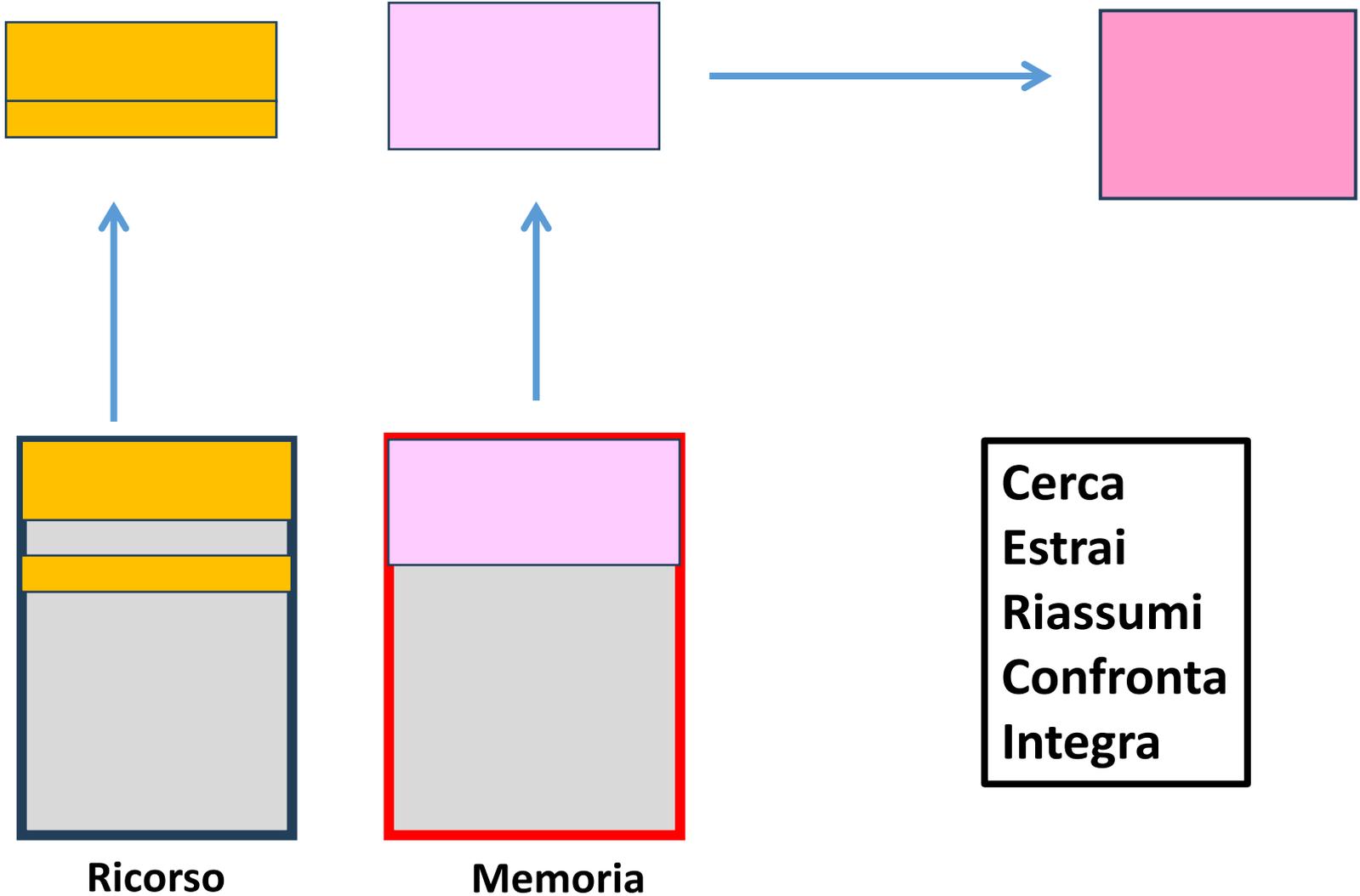
Quali comandi?



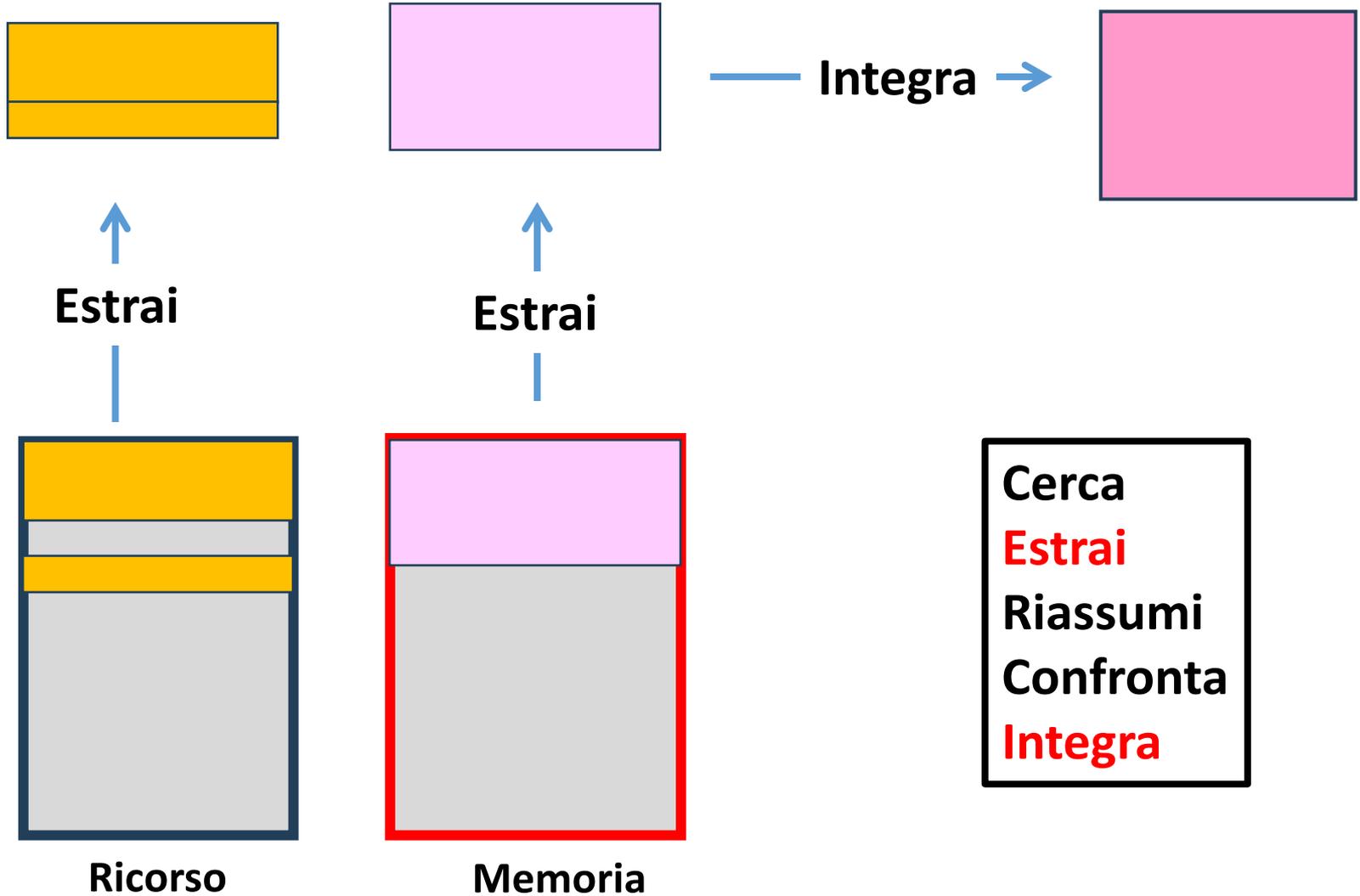
Integra e riassumi



Quali comandi?



Estrai e Integra



Come contribuire a una IA responsabile e allineata con l'AI Act? - **Strategie**

- 1. Restringere** il numero dei comandi linguistici e la sintassi delle frasi con cui interagire con i modelli linguistici
- 2. Imparare a far bene le domande → Il prompt engineering**

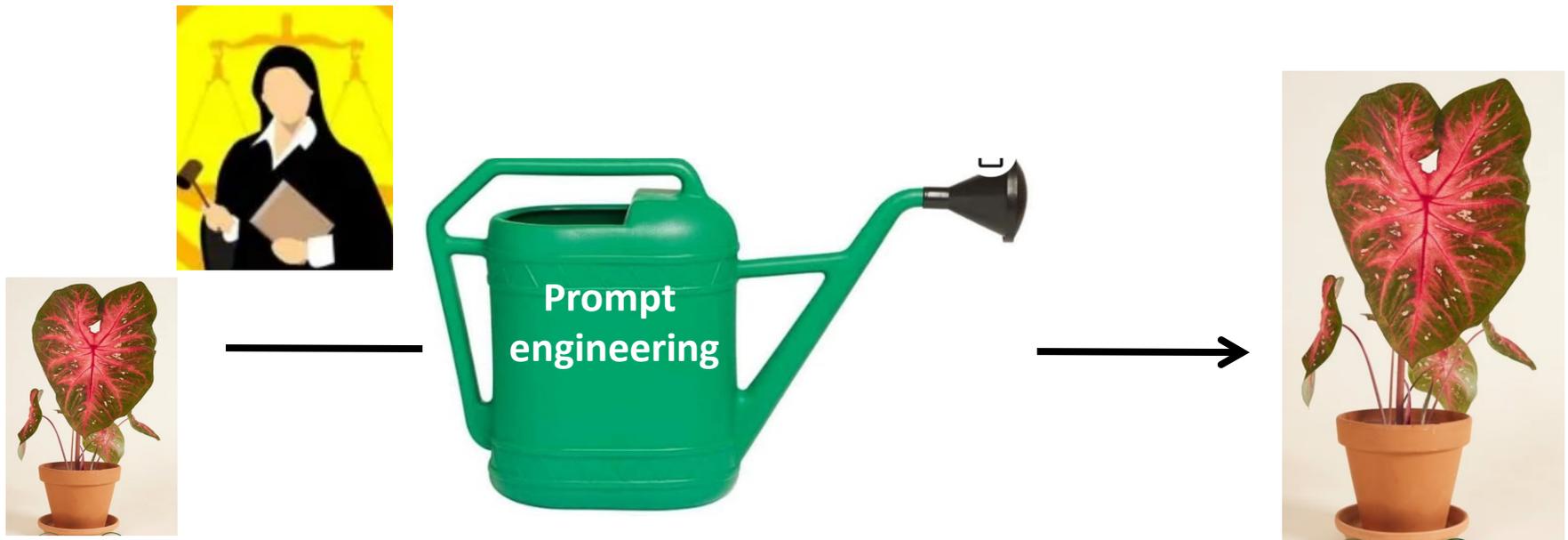
Il prompt engineering

E' un **insieme di tecniche** che si possono usare per fare bene le domande, tra cui le seguenti

1. Fornire **una sequenza di istruzioni** che indichino **come produrre ciò che si vuole**
2. Fornire **esempi** di **ciò che si vuole**
3. Fornire **definizioni** di **ciò che si vuole**

I modelli linguistici sono come piante che dobbiamo aiutare a crescere

It's a bit like growing a plant or a bacterial colony: we set the high-level conditions that direct and shape growth¹, but the exact structure which emerges is unpredictable and difficult to understand or explain. Looking inside these systems, what we see are vast matrices of billions of numbers. These are *somehow* computing important cognitive tasks, but exactly how they do so isn't obvious.



Sintassi generale dei comandi verbali

Nome-del-comando

Cosa, cioè quale **testo** si vuole produrre in output

Dove, cioè su quali **testi** opera il comando

Modalità della esecuzione del comando

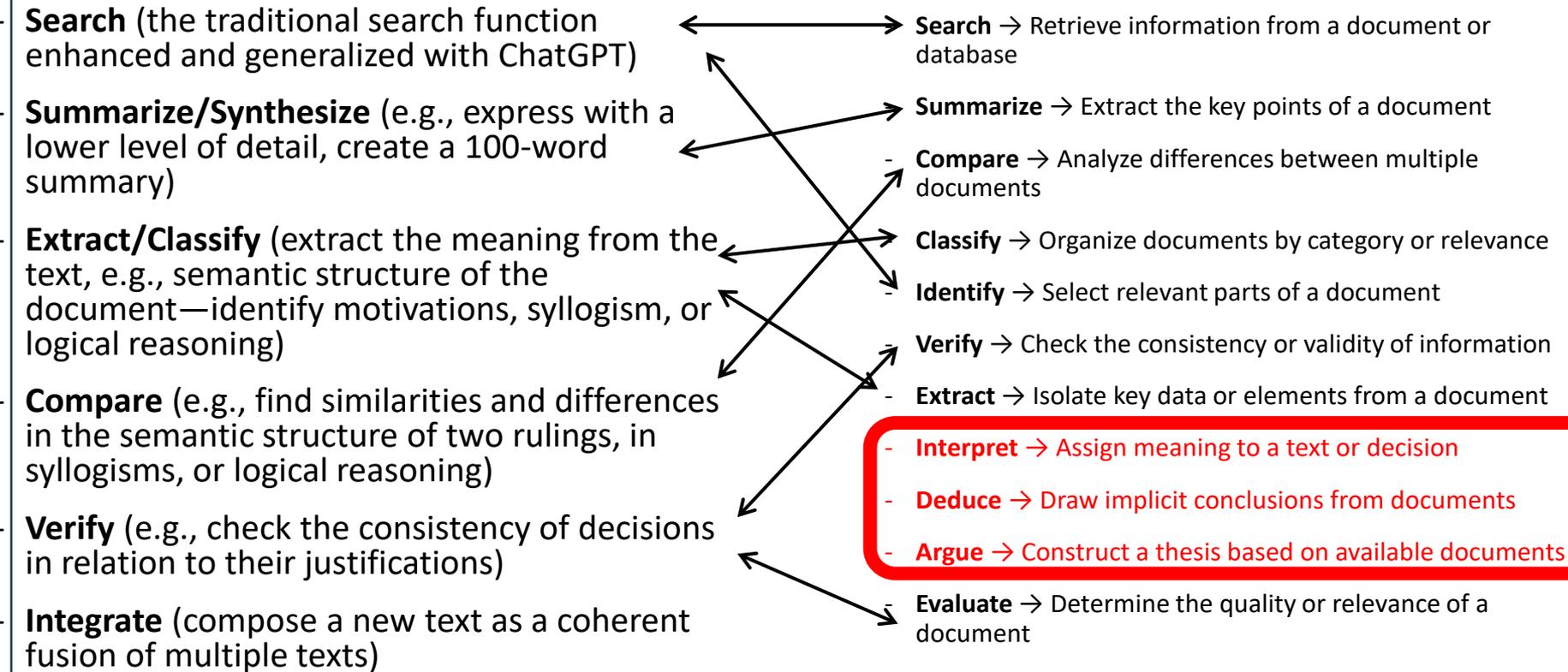
Questo testo è in formato libero, e può essere usato per esprimere:

1. Una condizione
2. Una tecnica di prompt engineering

Come contribuire a una IA responsabile e allineata con l'AI Act? - **Strategie**

- 1. Restringere** il numero dei comandi linguistici e la sintassi delle frasi con cui interagire con i modelli linguistici
- 2. Imparare a far bene** le domande → Il **prompt engineering**
- 3. Non includere tra i comandi** quelli che esplicitamente o implicitamente **sostituiscono** il magistrato nelle decisioni

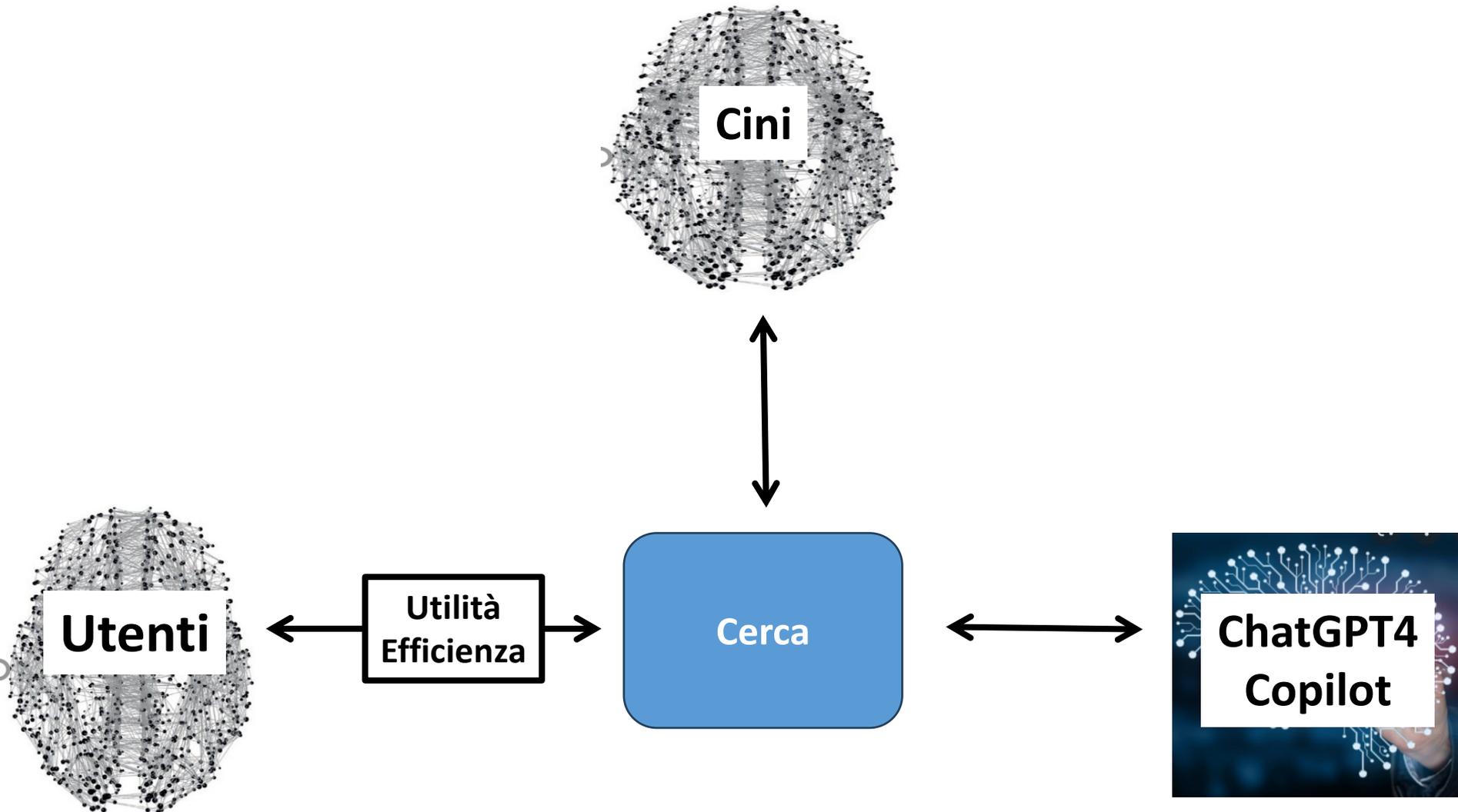
Comandi verbali e Speech acts della Speech Act Theory



Come contribuire a una IA responsabile e allineata con l'AI Act? - **Strategie**

- 1. Restringere** il numero dei comandi linguistici e la sintassi delle frasi con cui interagire con i modelli linguistici
- 2. Imparare a far bene** le domande → Il **prompt engineering**
- 3. Non includere tra i comandi** quelli che esplicitamente o implicitamente **sostituiscono** il magistrato nelle decisioni
- 4. Includere tra i comandi solo** quelli che **servono effettivamente** agli utenti finali → **utilità e efficienza**
- 5. Far partecipare attivamente** gli utenti al progetto

Gli utenti valutano cosa gli è utile
e cosa fa risparmiare tempo



Come contribuire a una IA responsabile e allineata con l'AI Act? - **Strategie**

- 1. Restringere** il numero dei comandi linguistici e la sintassi delle frasi con cui interagire con i modelli linguistici
- 2. Imparare a far bene** le domande → Il **prompt engineering**
- 3. Non includere** tra i comandi quelli che esplicitamente o implicitamente sostituiscono il magistrato nelle decisioni
- 4. Includere tra i comandi solo** quelli che **servono effettivamente** agli **utenti finali**,
- 5. Far partecipare attivamente gli utenti** al progetto
- 6. Valutare la qualità dei comandi**

Il comando Riassumi

2024 T. Worledge - The Extractive-Abstractive Spectrum: Uncovering Verifiability Trade-offs In LLM Generations

Sono definiti cinque livelli intermedi tra GoogleSearch (Extractive) e Language model (Abstractive)

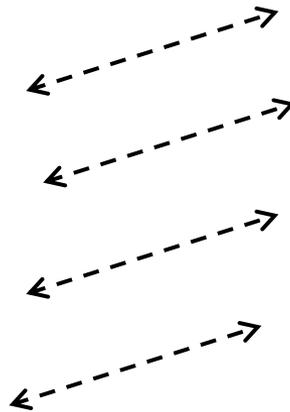


Query	Explain to a third-grader: where did the term 'robot' come from?
Extractive	<p>https://asylonrobotics.com/history-of-robotics/ Drones, manufacturing equipment, and vehicles are all using robotic technology these days. But where did the idea for robotics come from? More importantly, where is it headed? Let's take a look at the history of robotics, and what their future looks like. History of Robotics</p> <p>The word 'robot' comes from the Czech word 'robota', meaning forced labor. That word is itself derived from the Proto-Slavic word 'orbota', meaning hard work. The term robot was originally introduced by Karel Capek in his 1920 play Rossumovi Univerzální Roboti, which can be translated to 'Rossum's Universal Robots.' The play talked about a time when artificial organic humanoid robots were built, but they quickly grew dissatisfied, leading to a robot revolution and a new robot society. In 1942, American writer Isaac Asimov used the term 'robotics' for the first time in a short story titled 'Runabout.' He suggested the 'Three Laws of Robotics' as a guideline for all sci-fi robot characters in his stories.</p> <p>https://en.wikipedia.org/wiki/Robot The term comes from a Slavic root, robot-, with meanings associated with labor</p> <p>The word 'robot' was first used to denote a fictional humanoid in a 1920 Czech-language play R.U.R. ...</p>
Quoted	<p>"The word 'robot' comes from the Czech word 'robota', meaning forced labor" [1]. "The term robot was originally introduced by Karel Capek in his 1920 play Rossumovi Univerzální Roboti, which can be translated to 'Rossum's Universal Robots.'" [2] "The word 'robot' was first used to denote a fictional humanoid in a 1920 Czech-language play R.U.R" [3]. "(Rossumovi Univerzální Roboti – Rossum's Universal Robots) by Karel Čapek, though it was Karel's brother Josef Čapek who was the word's true inventor" [4].</p>
Paraphrased	<p>The term 'robot' is derived from the Czech term 'robota', which translates to forced labor [1]. Karel Capek was the first to use the term in his 1920 play titled "Rossum's Universal Robots." [2] Interestingly, it was in this Czech play, known as R.U.R. (Rossumovi Univerzální Roboti), that 'robot' was used to describe a fictional humanoid for the first time [2][3]. However, the actual credit for inventing the word goes to Karel's brother, Josef Čapek [4].</p>
Entailed	<p>The term 'robot' comes from the Czech word 'robota', which means forced labor [1]. It was first used in a 1920 play called 'Rossum's Universal Robots' by Karel Capek, but it was actually Karel's brother Josef who came up with the word [2][4].</p>
Abstractive	<p>The word 'robot' was first used in a story about pretend people who could do work for others [2][3]. It comes from a Czech word 'robota', which means work that you have to do but might not want to [1]. A writer named Karel Capek used it in a play he wrote in 1920, but it was actually his brother Josef who thought of using the word 'robot' [4].</p>

Confronto tra modalità di valutazione

Extractive-Abstractive Spectrum

- Fluency
- **Citation precision**
- **Citation coverage**
- **Perceived utility**
- **Time to verify**



Valutate dal sistema

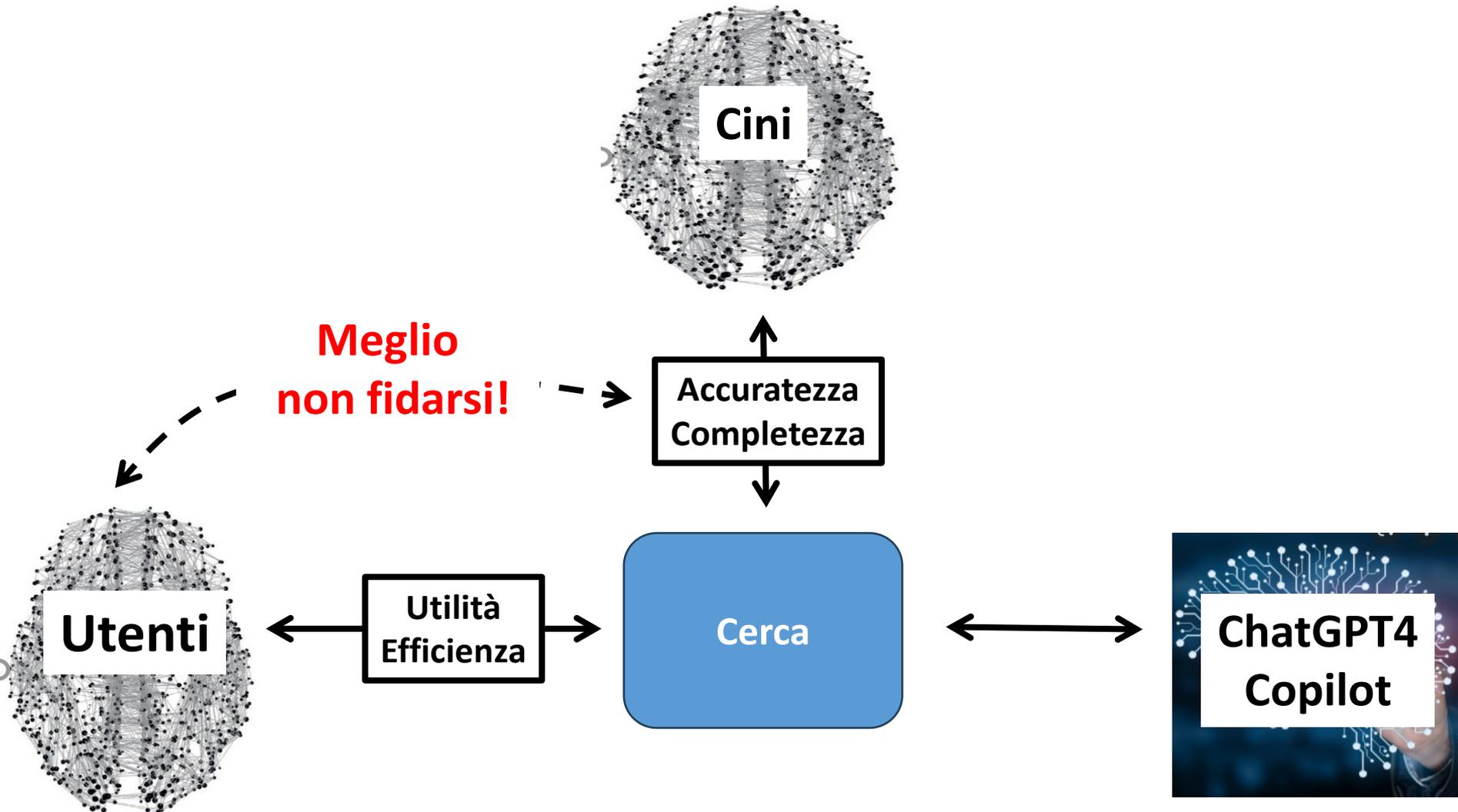
Comandi DCL

- **Accuratezza**
- **Completezza**
- **Utilità**
- **Tempo risparmiato**

Valutate dal sistema

Valutate dagli utenti

Qualità dei comandi e chi le valuta



Confronto ChatGPT vs Copilot

Risposta corretta (la “verità”)

«.....sfsfg..... **assorbimento del superminimo**
.....fjsljalsdìqejfa..... **è stato assorbito il**
superminimo ..sdfjaldfj..... **mantenimento del**
superminimoslòfjqgfqèerl **assorbibilità**
del superminimosldjfaljfa ...**non assorbimento**
del superimimo"

Frase simili

- **assorbimento del superminimo**
- **è stato assorbito il superminimo**
- **assorbibilità del superminimo**

Frase con significato opposto

- **mantenimento del superminimo**
- **non assorbimento del superimim**

Risposte a confronto (in rosso gli errori)

Accuratezza → Quante tra le risposte fornite sono giuste

Completezza → Quante tra le risposte giuste sono state fornite

ChatGPT

Fraasi simili

- assorbimento del superminimo
- è stato assorbito il superminimo
- assorbibilità del superminimo

Fraasi con significato opposto

- mantenimento del superminimo
- non assorbimento del superminimo

Copilot

Fraasi simili

- è stato assorbito il superminimo
- **mantenimento del superminimo**
- assorbibilità del superminimo
- **non assorbimento del superminimo**

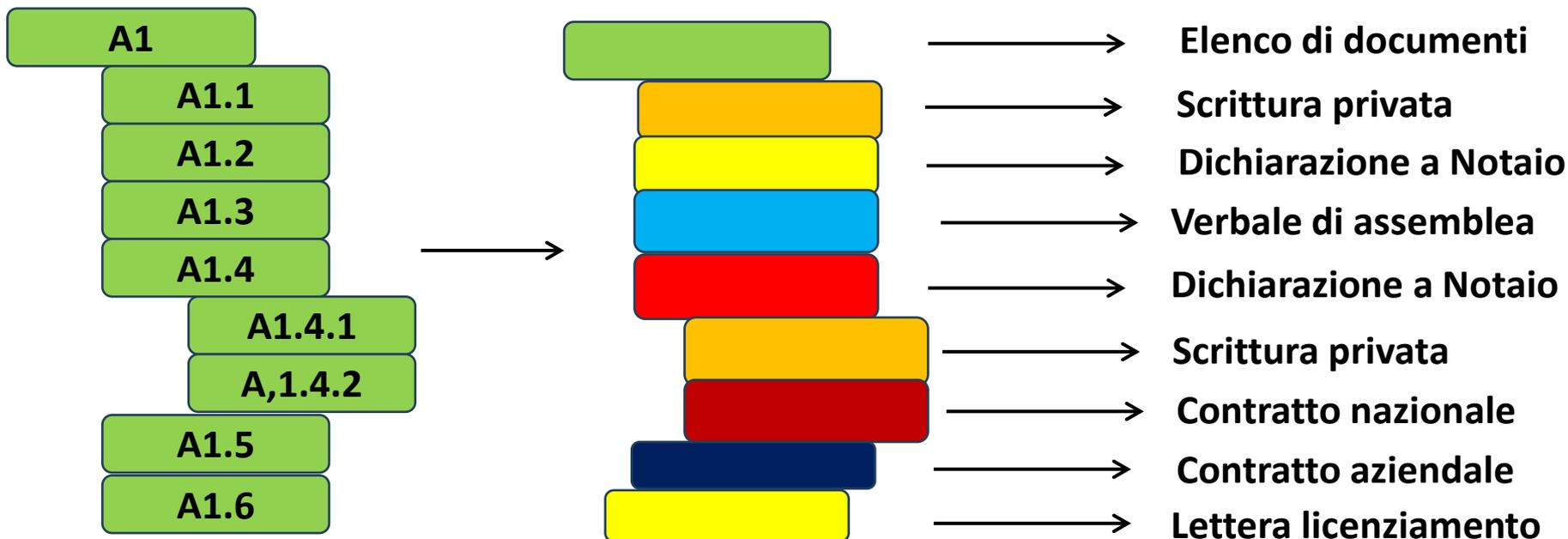
	Accuratezza	Completezza
ChatGPT	100%	100%
Copilot	50%	66%

Approfondimenti

Normalizza e Classifica

IA Generativa

IA Ristretta – Machine Learning



Confronto

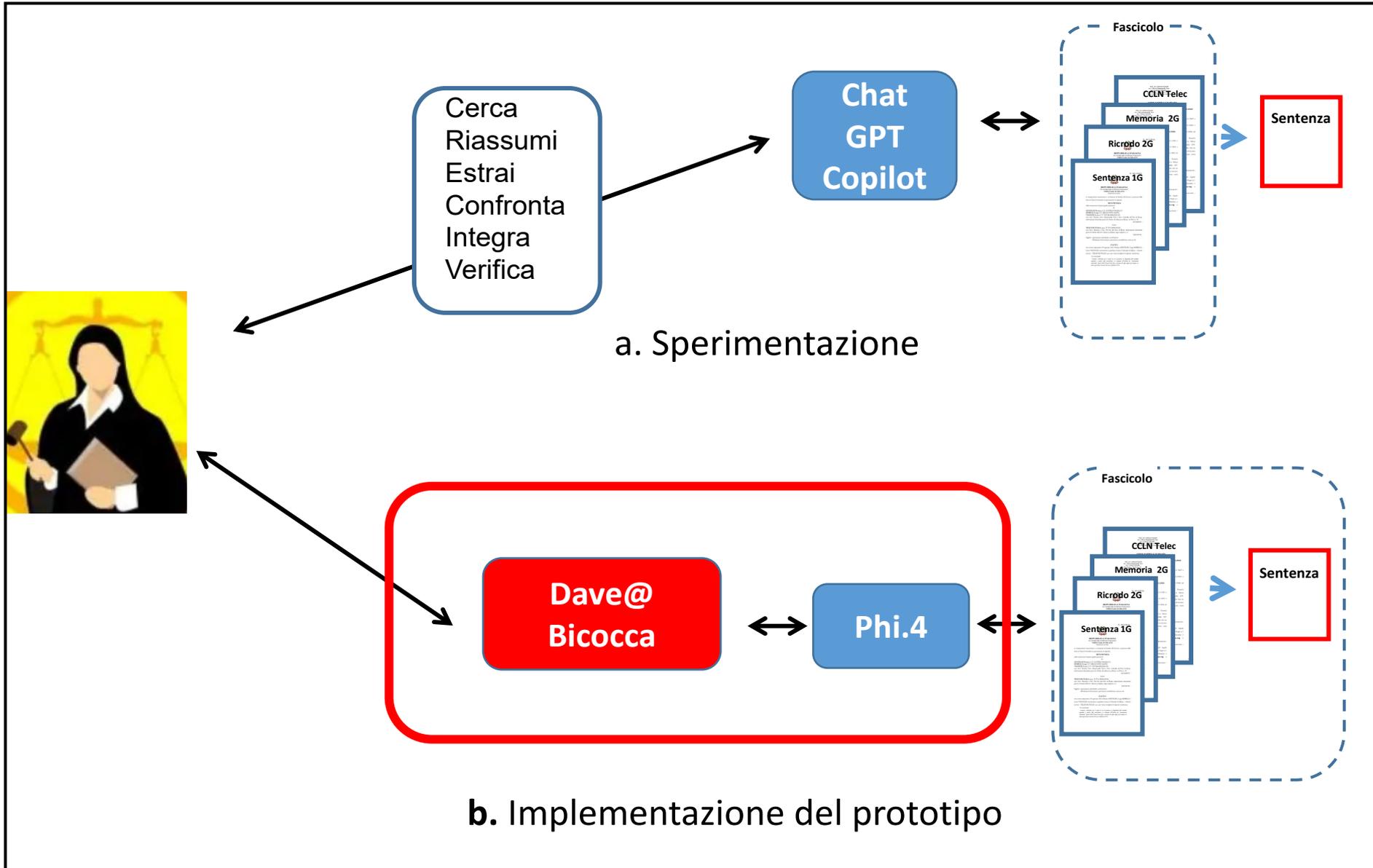
- Siamo I primi a immaginare un linguaggio di interazione tra umani e macchine?
- No, dagli anni 70 del secolo scorso è stata investigata dal linguista John Searle una **teoria degli atti linguistici**

Come contribuire a una IA responsabile e allineata con l'AI Act?

Strategie

- 1. Restringere** il numero dei comandi linguistici e la sintassi delle frasi con cui interagire con i modelli linguistici
- 2. Imparare a far bene** le domande → Il **prompt engineering**
- 3. Non includere** tra i comandi quelli che esplicitamente o implicitamente sostituiscono il magistrato nelle decisioni
- 4. Includere tra i comandi solo** quelli che **servono effettivamente** agli **utenti finali**,
- 5. Far partecipare attivamente** gli **utenti** al progetto
- 6. Valutare l'accuratezza** dei comandi
- 7. Specificare e sviluppare** **agenti software** che **implementano i comandi assicurandone la aderenza all'AI Act**

Le due fasi del Progetto



Il Processo Civile Telematico (PCT) oggi

**2. Livello di redazione
sintattica/mediante moduli
di atti & documenti per singoli campi e
frasi fisse**

**1. Livello di trasmissione sicura,
archiviazione degli atti&documenti,
orchestrazione del procedimento civile**

Slsadf dddd
As òda dgdg
Slsadf dddd
Asòda dgdgl
Slsadf dddd
Asòda dgdgl
ISlsadf dddd
Asòda dgdgl

Documento