BIO-MEDICAL INFORMATICS
"Mario Stefanelli"

UNIVERSITÀ DI PAVIA

Istituti Clinici Scientifici Maugeri

USO DEI LARGE LANGUAGE MODELS PER IL RECLUTAMENTO DEI PAZIENTI PER I CLINICAL TRIALS

Riccardo Bellazzi, Valentina Tibollo, Francesca Pescol

Laboratorio di Informatica Biomedica «Mario Stefanelli», Università di Pavia

Laboratorio di Informatica Medica e Intelligenza Artificiale, IRCCS ICS Maugeri, Pavia

LA SFIDA DEL RECLUTAMENTO DEI PAZIENTI

Il processo di reclutamento dei pazienti rappresenta un vero e proprio **collo di bottiglia** nelle sperimentazioni cliniche, occupando circa 1/3 della loro durata complessiva.

La selezione subottimale della coorte di pazienti è una delle **cause principali** del fallimento delle sperimentazioni:

- circa 1/3 degli studi fallisce per problemi di arruolamento
- L'86% delle sperimentazioni non rispetta i tempi di arruolamento[2].

INTELLIGENZA ARTIFICIALE: STRUMENTO PER OTTIMIZZARE IL RECLUTAMENTO DEI PAZIENTI

?

LARGE LANGUAGE MODELS PER IL RECLUTAMENTO DEI PAZIENTI

Le innovazioni più recenti nel campo dell'Intelligenza Artificiale e, in particolare, del **Natural Language Processing (NLP)** hanno reso possibile lo sviluppo di **algoritmi** per il **reclutamento automatico** dei pazienti nei clinical trials[1].

I Large Language Models (LLMs) garantiscono una capacità senza precedenti di processare e interpretare il linguaggio naturale.

Ai fini del reclutamento, questi modelli permettono di sfruttare la vasta mole di informazioni contenuta nei referti e nelle note cliniche, fonti essenziali per considerare il quadro clinico completo del paziente[1].

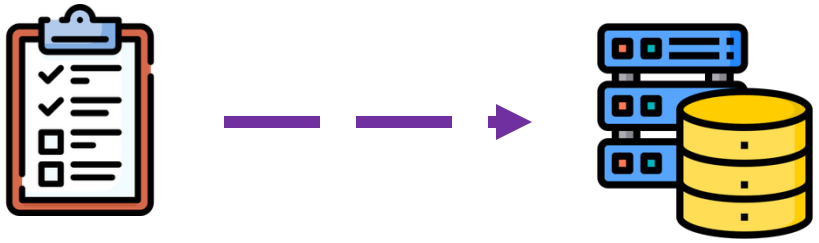
LARGE LANGUAGE MODELS PER IL RECLUTAMENTO DEI PAZIENTI

Approcci e esempi di applicazione allo stato dell'arte

Due approcci principali di utilizzo degli LLMs

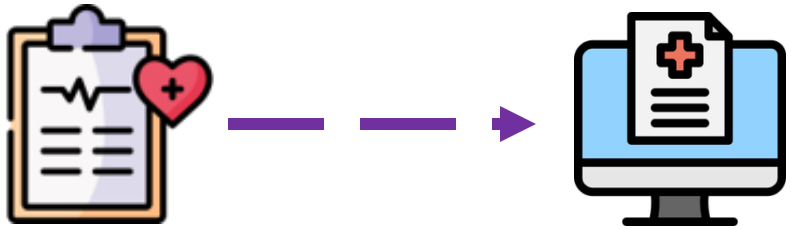
1

Conversione dei criteri di inclusione ed esclusione in **query strutturate**, le quali possono essere utilizzate per **interrogare database clinici**[3].



2

Estrazione di informazione in **formato strutturato** dai referti e dalle note cliniche, per facilitare il confronto automatico con i criteri di inclusione ed esclusione[4].



LARGE LANGUAGE MODELS PER IL RECLUTAMENTO DEI PAZIENTI

Approcci e esempi di applicazione allo stato dell'arte

Esempio approccio 1

Conversione dei criteri di inclusione/esclusione in query strutturate

ELSEVIER

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Special Communication

Criteria2Query 3.0: Leveraging generative large language models for clinical trial eligibility query generation

Jimyung Park [a,1], Yilu Fang [a,1], Casey Ta [a], Gongbo Zhang [a], Betina Idnay [a], Fangyi Chen [a], David Feng [a], Rebecca Shyu [a], Emily R. Gordon [b], Matthew Spotnitz [a], Chunhua Weng [a,*]

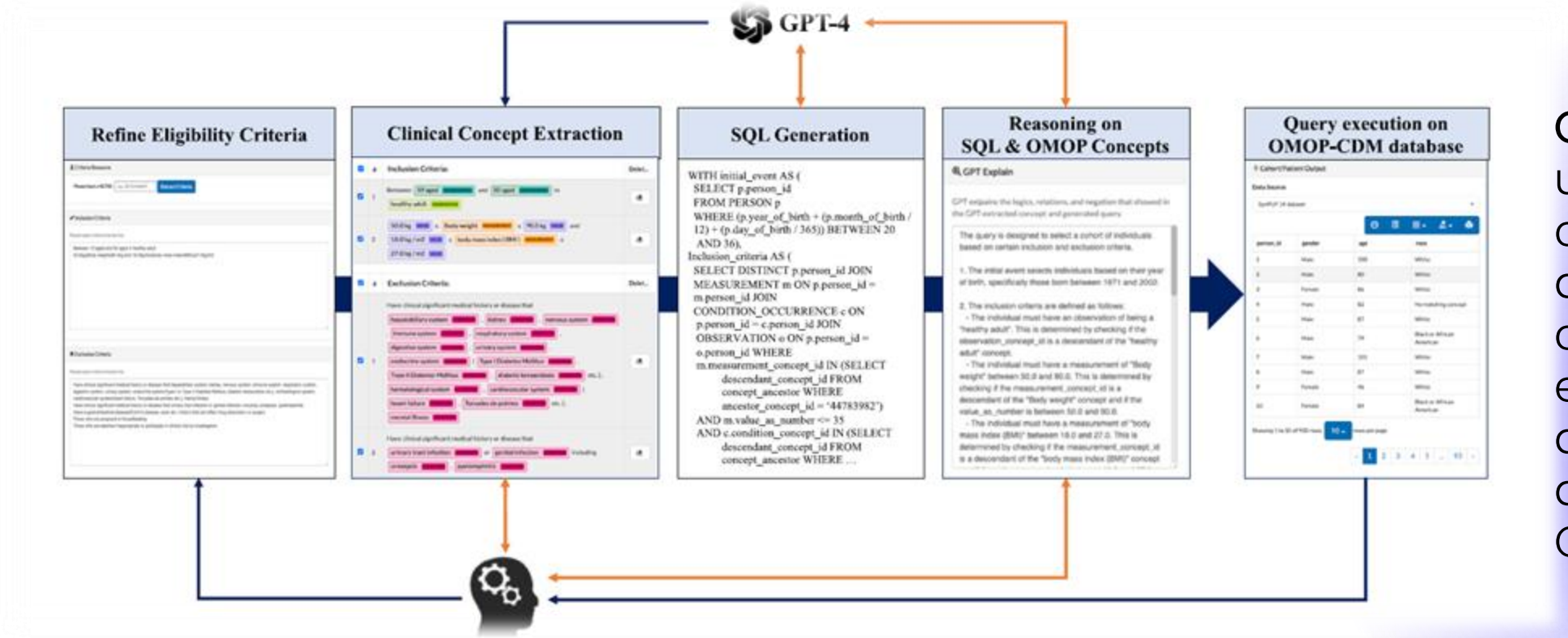
[a] Department of Biomedical Informatics, Columbia University, New York, United States

[b] Columbia University Vagelos College of Physicians and Surgeons, New York, United States

LARGE LANGUAGE MODELS PER IL RECLUTAMENTO DEI PAZIENTI

Approcci e esempi di applicazione allo stato dell'arte

Esempio approccio 1: workflow



OpenAI GPT-4 viene usato per la conversione semi-automatica dei criteri in **query SQL** eseguibili su un database conforme allo standard OMOP-CDM[3].

LARGE LANGUAGE MODELS PER IL RECLUTAMENTO DEI PAZIENTI

Approcci e esempi di applicazione allo stato dell'arte

Esempio approccio 2
Estrazione di informazione strutturata dai referti

npj | digital medicine

Article

Published in partnership with Seoul National University Bundang Hospital

https://doi.org/10.1038/s41746-024-01274-7

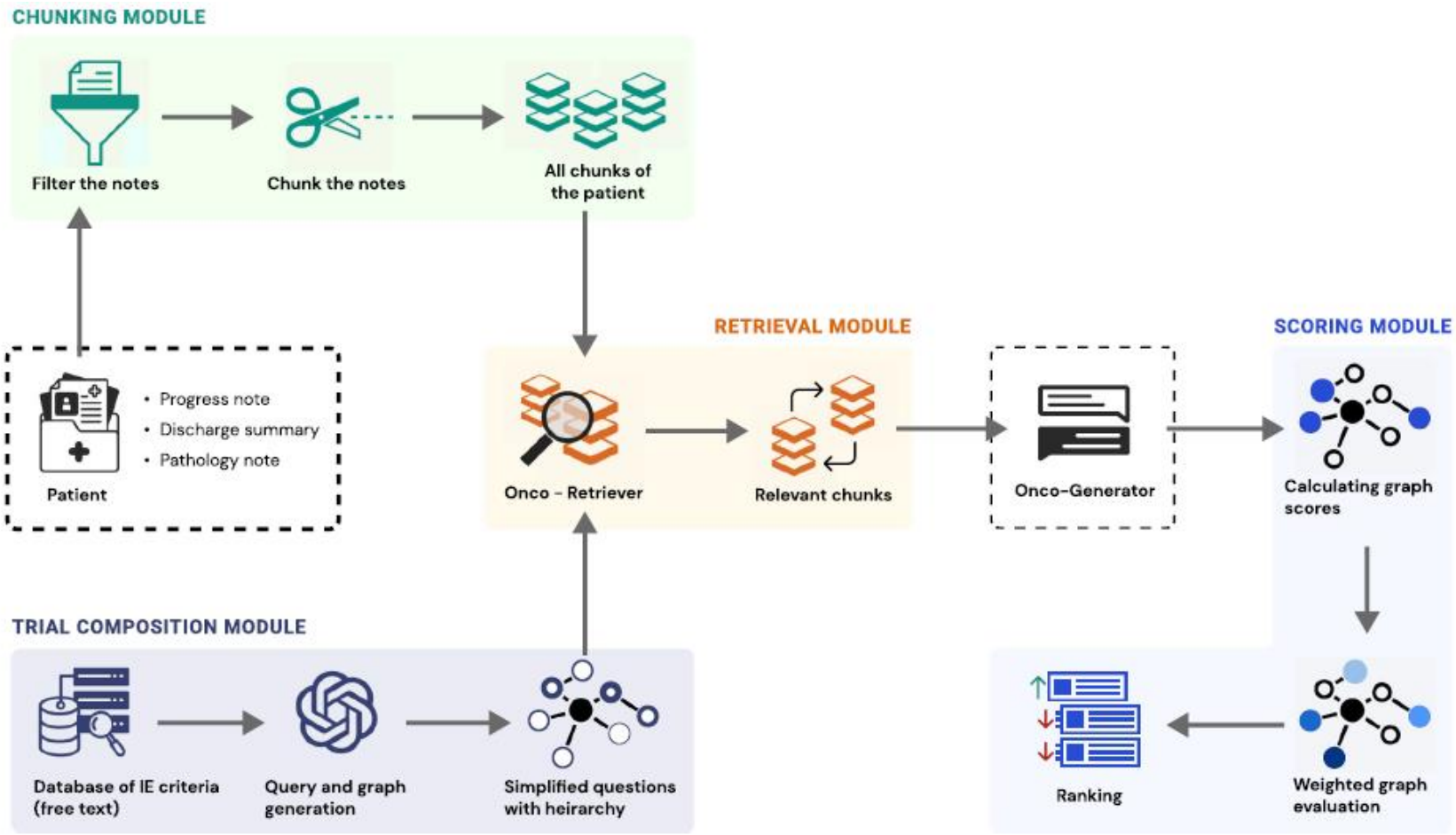
PRISM: Patient Records Interpretation for Semantic clinical trial Matching system using large language models

Check for updates

Shashi Gupta[1,6], Aditya Basu[1,6], Mauro Nievas[1], Jerrin Thomas[1], Nathan Wolfrath[2], Adhitya Ramamurthi[2], Bradley Taylor[2], Anai N. Kothari[2], Regina Schwind[1], Therica M. Miller[3], Sorena Nadaf-Rahrov[4], Yanshan Wang[5] & Hrituraj Singh[1]

LARGE LANGUAGE MODELS PER IL RECLUTAMENTO DEI PAZIENTI

Approcci e esempi di applicazione allo stato dell'arte



Esempio approccio 2: workflow

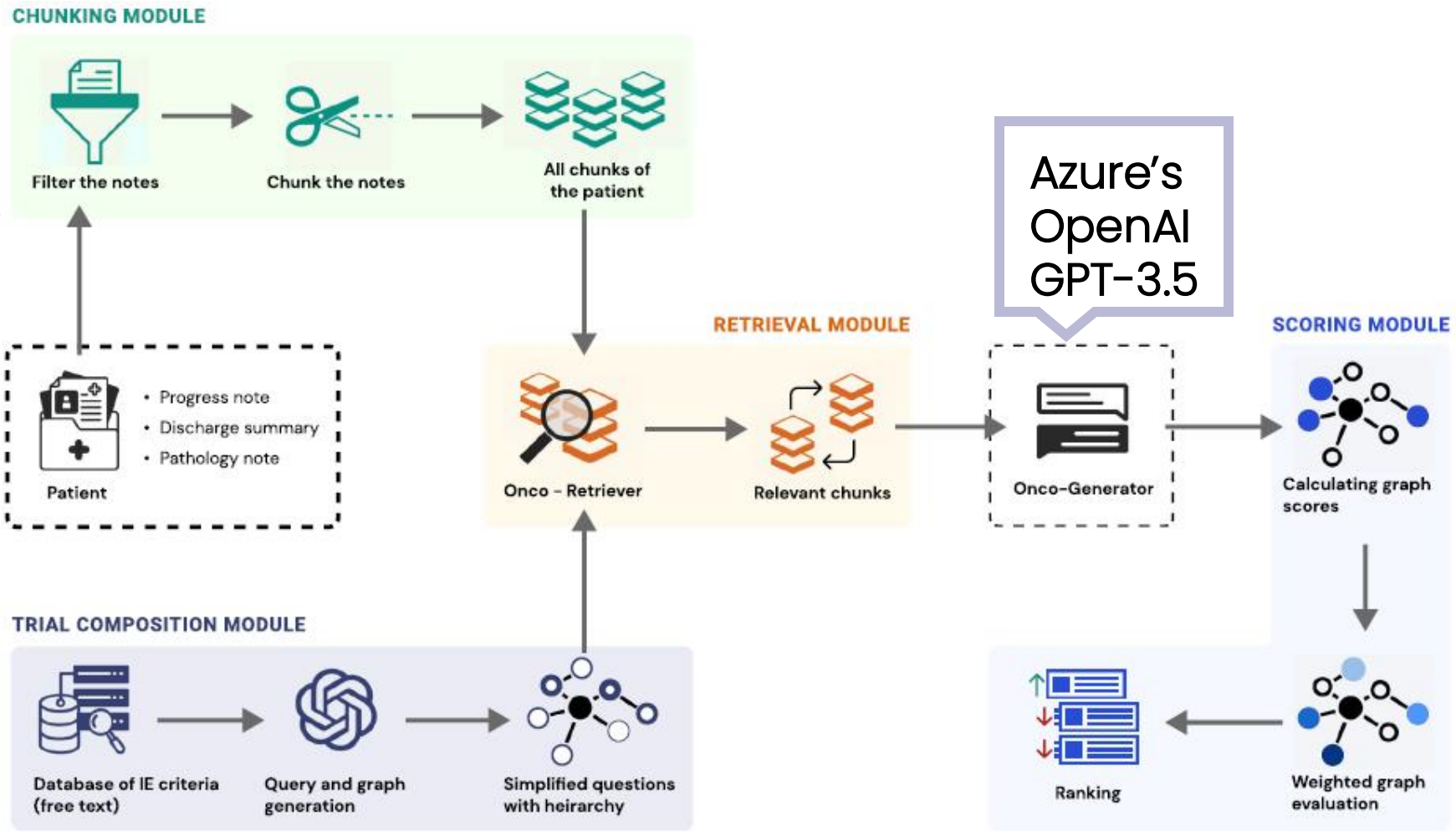
Input: criteri di inclusione/esclusione di diversi trials e note cliniche di pazienti **oncologici.** Entrambi vengono forniti in formato **testo libero**.

Output: ranking dei trials per ogni paziente sulla base dell'eleggibilità.

LARGE LANGUAGE MODELS PER IL RECLUTAMENTO DEI PAZIENTI

Approcci e esempi di applicazione allo stato dell'arte

Azure OpenAI's Ada embedding model



Azure's OpenAI GPT-3.5

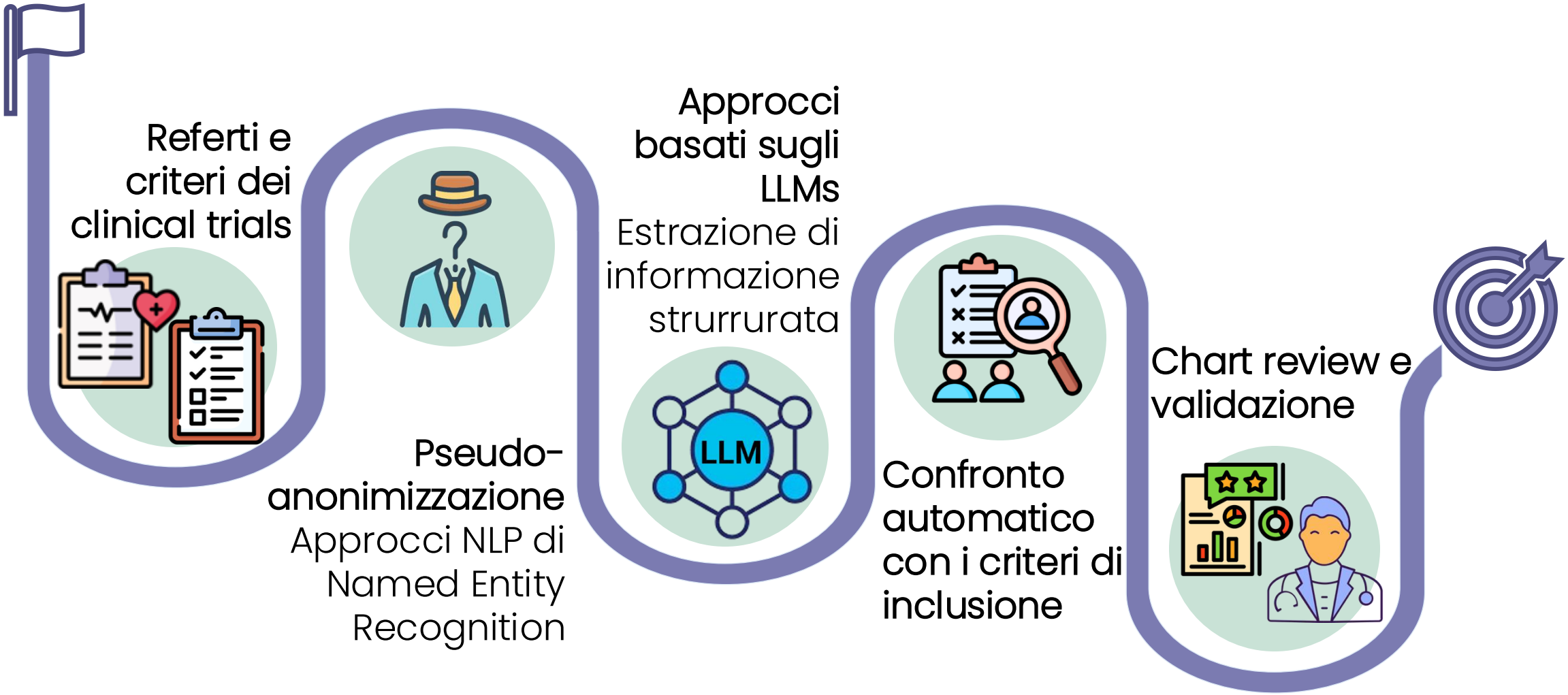
OpenAI GPT-3.5

Esempio approccio 2: workflow

APPLICAZIONE NELLA REALTÀ CLINICA DI ICS MAUGERI

APPLICAZIONE NELLA REALTÀ CLINICA DI ICS MAUGERI

Il processo



APPLICAZIONE NELLA REALTÀ CLINICA DI ICS MAUGERI

Primo caso clinico analizzato: ***Coronary Artery Disease (CAD)***

878 referti relativi a ricoveri presso ICS Maugeri Pavia di pazienti affetti da CAD, selezionati dagli esperti clinici

3 studi clinici relativi a CAD e focalizzati sulla valutazione dell'impiego di farmaci, alternativi alle statine, per diminuire livelli di lipidi, anch'essi selezionati dagli esperti clinici

Confronto delle performance di LLMs di diversa dimensione (complessità), sia proprietari sia open source

APPLICAZIONE NELLA REALTÀ CLINICA DI ICS MAUGERI

Primo caso clinico analizzato: ***Coronary Artery Disease (CAD)***

Istituti Clinici Scientifici Maugeri

Referto

NIH U.S. National Library of Medicine

ClinicalTrials.gov

Criteri

Prompt engineering per l'estrazione di informazioni strutturate (JSON)

JSON popolato con le informazioni richieste

Struttura tipica:

- RICHIESTA (con template JSON)
- INDICAZIONI (eventuali specifiche)
- REFERTO

Possibili integrazioni

- Definizioni e terminologie da ontologie mediche
- Chain of Thought

APPLICAZIONE NELLA REALTÀ CLINICA DI ICS MAUGERI

Studio pilota

- 10 referti sintetici relativi a pazienti ricoverati per **CAD** (ottenuti da una rielaborazione dei referti reali)

- 8 LLMs testati per il task di estrazione di informazione strutturata (JSON template)
  - Modelli ad elevate prestazioni disponibili su AWS Bedrock
  - Modelli eseguibili su macchine locali (esigenze computazionali minori)

- Validazione basata su
  - Accuratezza di formato (output costituito da un JSON valido e rispetto delle opzioni di risposta fornite)
  - Accuratezza di contenuto (mediante chart review degli esperti clinici)

AWS Bedrock

In locale

Modello	Azienda	# Parametri (complessità del modello)	Utilizzo
Titan Text Large	Amazon	-	Proprietario a pagamento
Titan Text G1 - Express	Amazon	-	Proprietario a pagamento
Claude 3 Sonnet	Anthropic	-	Proprietario a pagamento
Claude 3.5 Sonnet	Anthropic	-	Proprietario a pagamento
Llama 3.1 8B Instruct	Meta	8B	Gratuito (con licenza)
Llama 3.1 70B Instruct	Meta	70B	Gratuito (con licenza)
Mistral 7B Instruct	Mistral AI	7B	Gratuito
Mixtral 8x7B Instruct	Mistral AI	47B	Gratuito
numind/NuExtract 1.5	NuMind	3.8B	Gratuito (con licenza)

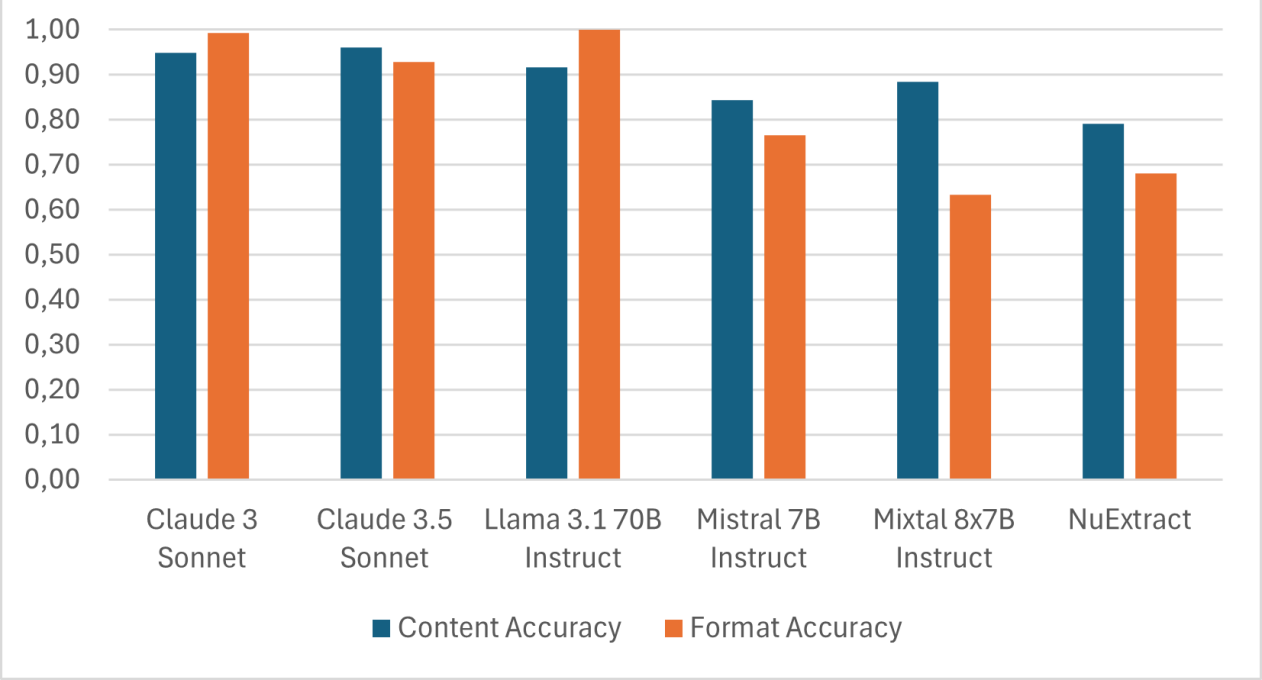
APPLICAZIONE NELLA REALTÀ CLINICA DI MAUGERI

Studio pilota: risultati

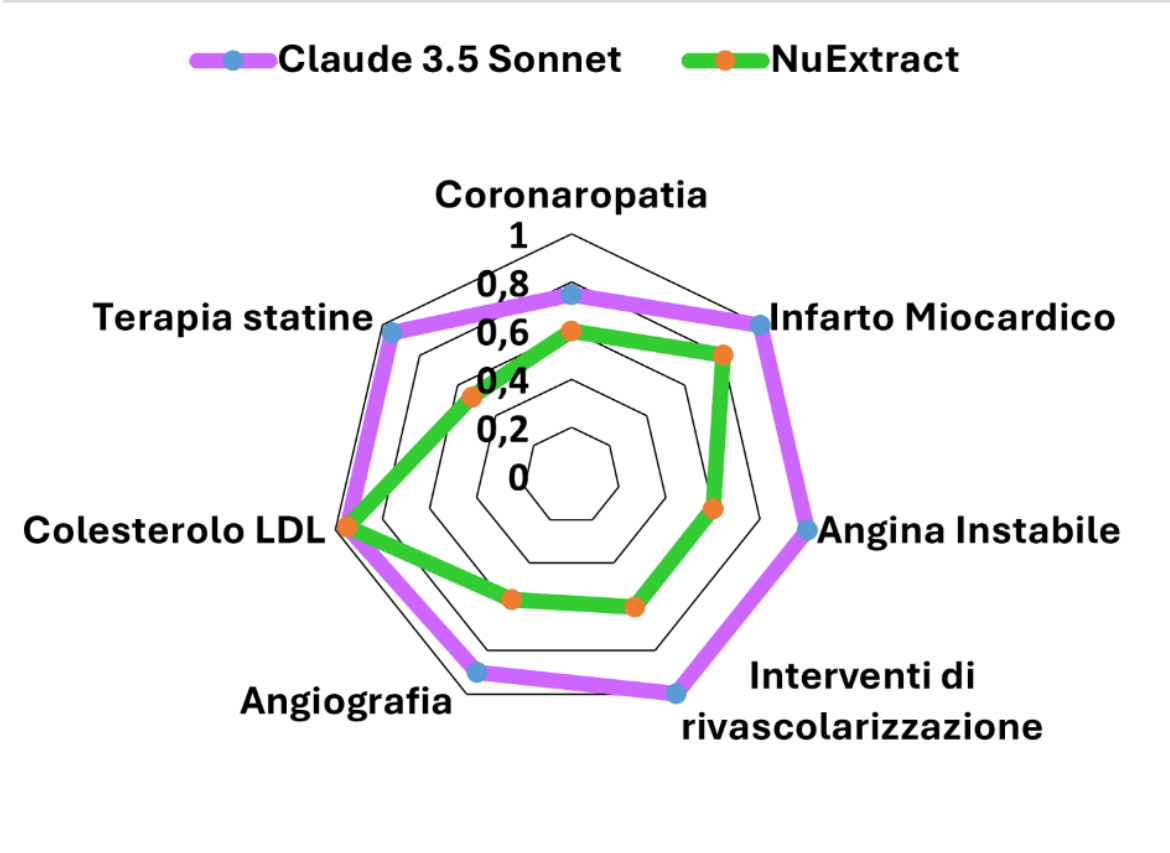
Modello	% JSON validi	Accuratezza di contenuto (media e deviazione standard)	Accuratezza di formato (media e deviazione standard)
Claude 3 Sonnet	80	0,95 (0,05)	0,99 (0,02)
Claude 3.5 Sonnet	**100**	**0,96 (0,03)**	0,93 (0,22)
Llama 3.1 70B Instruct	30	0,92 (0,06)	1,00 (0,00)
Mistral 7B Instruct	90	0,84 (0,06)	0,77 (0,25)
Mixtral 8x7B Instruct	**100**	0,88 (0,04)	0,63 (0,21)
numind/NuExtract 1.5	**100**	0,79 (0,10)	0,68 (0,21)

APPLICAZIONE NELLA REALTÀ CLINICA DI ICS MAUGERI

Risultati studio pilota

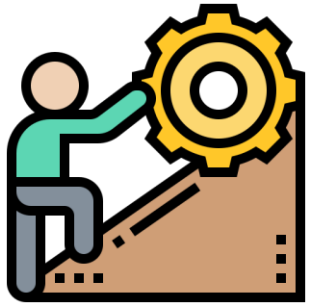


Confronto grafico delle prestazioni dei modelli



Confronto tra Claude 3.5 Sonnet e NuExtract su concetti specifici

SFIDE E REQUISITI PER L'IMPLEMENTAZIONE OPERATIVA



1) **Recupero delle fonti di informazione (referti):** l'accesso ai referti clinici in formato digitale richiede l'integrazione con i sistemi informativi sanitari e la definizione di procedure per l'estrazione automatica o semiautomatica dei documenti rilevanti.

2) **Creazione di un ambiente idoneo:** l'utilizzo degli LLMs ad elevate prestazioni richiede risorse computazionali adeguate. Allo stesso tempo, l'ambiente deve garantire il rispetto delle normative di protezione dei dati (GDPR).

3) **Processo di chart review:** la validazione di questi approcci richiede la presenza di un *human in the loop*. È un passaggio fondamentale per garantire l'affidabilità e la sicurezza dell'applicazione, ma anche particolarmente laborioso.

CONCLUSIONI E SVILUPPI FUTURI

Breve termine

- **Valutazione dell'impatto dell'integrazione delle ontologie mediche nel prompt** nel task di estrazione di informazioni, considerando un campione di referti sintetici più esteso (30)

- **Valutazione dell'accuratezza dell'eleggibilità stabilita dagli LLMs** (usando sia regole logiche applicate alle informazioni estratte, sia direttamente gli LLMs) sul campione di 30 referti sintetici

- **Validazione dell'intera pipeline su un set di referti reali molto più esteso,** grazie all'adozione dell'infrastruttura cloud **Microsoft Azure**. Per questo, è stata da poco presentata la richiesta al **comitato etico del Policlinico San Matteo**.

In programma

- Sviluppo di un **software completo per Maugeri,** provvisto di interfaccia grafica, per il reclutamento dei pazienti e la validazione sistematica dell'arruolamento ottenuto. Il software implementerà le strategie di prompting validate nella fase precedente.

- Integrazione **dell'informazione temporale** per la valutazione delle finestre di eleggibilità

GRAZIE PER L’ATTENZIONE

BIBLIOGRAFIA

1. Wouters OJ, McKee M, Luyten J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. JAMA. 2020;323(9):844–853. doi:10.1001/jama.2020.1166
2. Harrer S, Shah P, Antony B, Hu J. Artificial Intelligence for Clinical Trial Design. Trends Pharmacol Sci. 2019 Aug;40(8):577-591. doi: 10.1016/j.tips.2019.05.005. Epub 2019 Jul 17. PMID: 31326235.
3. Park J, Fang Y, Ta C, et al. Criteria2Query 3.0: Leveraging generative large language models for clinical trial eligibility query generation. Journal of Biomedical Informatics. 2024;154:104649. doi:10.1016/j.jbi.2024.104649
4. Gupta S, Basu A, Nievas M, et al. PRISM: Patient Records Interpretation for Se-mantic clinical trial Matching system using large language models. npj Digit Med. 2024;7(1):305. doi:10.1038/s41746-024-01274-7
5. Ray KK, Wright RS, Kallend D, Koenig W, Leiter LA, Raal FJ, Bisch JA, Richardson T, Jaros M, Wijngaard PLJ, Kastelein JJP; ORION-10 and ORION-11 Investigators. Two Phase 3 Trials of Inclisiran in Patients with Elevated LDL Cholesterol. N Engl J Med. 2020 Apr 16;382(16):1507-1519. doi: 10.1056/NEJMoa1912387. Epub 2020 Mar 18. PMID: 32187462.
6. Schwartz GG, Steg PG, Szarek M, Bhatt DL, Bittner VA, Diaz R, Edelberg JM, Goodman SG, Hanotin C, Harrington RA, Jukema JW, Lecorps G, Mahaffey KW, Moryusef A, Pordy R, Quintero K, Roe MT, Sasiela WJ, Tamby JF, Tricoci P, White HD, Zeiher AM; ODYSSEY OUTCOMES Committees and Investigators. Alirocumab and Cardiovascular Outcomes after Acute Coronary Syndrome. N Engl J Med. 2018 Nov 29;379(22):2097-2107. doi: 10.1056/NEJMoa1801174. Epub 2018 Nov 7. PMID: 30403574.
7. Nissen SE, Lincoff AM, Brennan D, Ray KK, Mason D, Kastelein JJP, Thompson PD, Libby P, Cho L, Plutzky J, Bays HE, Moriarty PM, Menon V, Grobbee DE, Louie MJ, Chen CF, Li N, Bloedon L, Robinson P, Horner M, Sasiela WJ, McCluskey J, Davey D, Fajardo-Campos P, Petrovic P, Fedacko J, Zmuda W, Lukyanov Y, Nicholls SJ; CLEAR Outcomes Investigators. Bempedoic Acid and Cardiovascular Outcomes in Statin-Intolerant Patients. N Engl J Med. 2023 Apr 13;388(15):1353-1364. doi: 10.1056/NEJMoa2215024. Epub 2023 Mar 4. PMID: 36876740.