



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Future
Artificial
Intelligence
Research

Verso un repository aperto di dati "nativi" italiani per modelli di AI generativa

Bernardo Magnini

Fondazione Bruno Kessler,
Trento

Napoli, 24 settembre 2024





Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Future
Artificial
Intelligence
Research

Dati per addestramento/valutazione di LLM in italiano

- Dati «nativi» italiani
- Traduzioni automatiche dall'inglese
- Dati sintetici, generati da LLM



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Future
Artificial
Intelligence
Research

Dati per addestramento/valutazione di LLM in italiano

- **Dati «nativi» italiani**
- Traduzioni automatiche dall'inglese
- Dati sintetici, generati da LLM



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Future
Artificial
Intelligence
Research

Contenuti di alta qualità

Dati nativi italiani per addestrare LLM/LMM

Attività Spoke 2



Massimizzare la qualità – minimizzare i rischi

Qualità dei contenuti

- **Accordi con editori:** testate giornalistiche, TV, libri
- Il caso dei piccoli editori trentini

Consapevolezza dei rischi dell'AI generativa

- Violazione del **diritto d'autore**
- Violazione della **privacy personale (GDPR)**
- Generazione di **contenuti socialmente inappropriati**



VENERDÌ 20 SETTEMBRE - 10:51

E' entrata nel vivo la terza edizione della kermesse che celebra il Trento Doc e che fino a domenica animerà i gioielli dell'arte di Trento e tutti i territori di produzione con eventi all'interno delle case spumantistiche, talk, degustazioni accompagnate dai migliori esperti al mondo e appuntamenti con i personaggi del mondo dello spettacolo. Un Festival spumeggiante" lo ha definito l'assessora provinciale all'agricoltura e promozione dei prodotti trentini Giulia Zanotelli, parlando di una "grande occasione per celebrare le eccellenze enologiche del Trentino



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Future
Artificial
Intelligence
Research

Verso un repository di dati per il dominio medico

Dati nativi italiani per addestrare LLM

Attività Spoke 2, collaborazione con IMN (progetto eCREAM)





LLM per il dominio medico

- Studio retrospettivo per sviluppare e validare un LLM nel dominio medico per PS
- 18 PS coinvolti
- Software di anonimizzazione installato nei PS
- Stimati 120M token per anno per PS
- **Primi dati Ospedale San Giovanni Bosco (Torino)!**

Dati pubblici

- PubMed, web, tesi in medicina, riviste scientifiche e divulgative
- Stimati 500M token

Motivo della visita QUESITO: crisi asmatica
Anamnesi patologica Verosimile asma bronchiale misconosciuta.

Dispnea da sforzo da almeno 2 settimane.

Accesso in PS per peggioramento dei sintomi.

Rx torace: non focolai. Eosinofilia: 940.

Già trattato con steroide ev, aerosolterapia.

Alla visita, riferito miglioramento clinico. Eupnoico. EOP: gemiti e sibili sparsi bilateralmente. SatO2 98% in aa. FC 90 R.

Barriera linguistica. Insegnata corretta modalità di assunzione della polvere inalatoria.

Fornito campione per inizio terapia.

Terapia consigliata:

- Deltacortene 25 mg 2 cpr al mattino per 5 giorni, poi 1 cpr al mattino per altri 5 giorni, poi 1/2 cpr al mattino per ulteriori 5 giorni e sospende

- Azitromicina 500 mg 1 cpr dopo pranzo per 6 giorni

- Pantoprazolo 20 mg 1 cpr alla sera

- Duoresp 160 mcg 2 inalazioni x 2 volte al giorno, mattino e sera (+ 2 al bisogno al pomeriggio), per almeno 1 mese

Rivalutazione presso il medico curante e programmare controllo pneumologico con

spirometria tra 1 mese sc.

Necessaria presenza di traduttore ai prossimi controlli, poichè capisce poco e male

l'italiano

Diagnosi: ASMA BRONCHIALE



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Future
Artificial
Intelligence
Research

EVALITA-LLM

Un nuovo benchmark per modelli del linguaggio in italiano

Collaborazione tra Spoke 2 e TP2



UNIVERSITÀ
DI TORINO



EVALITA-LLM

Benchmarking flessibile per LLM in Italiano

- 12 task “originali” italiani (da Evalita 2007-2023)
 - Comprensione del testo italiano
- **Multi-prompt** e single-prompt
- Task a **scelta multipla**: sentiment, test ammissione
- Task **generativi**: fai un riassunto, trova un sinonimo
- Zero-shot e few-shot
- Versione full e light

Prompt 1: domanda con due scelte

La parola '**minori**' nella frase '*Dopo innumerevoli provini sono stati scelti 15 attori per i ruoli principali, 200 per le parti minori e quasi 4000 comparse...*' ha lo stesso significato della parola '**minore**' nella frase '*Il direttore di Micromega ha detto che ci troviamo di fronte alla scelta fra un bene e un male minore.*'?

A: Sì B: No

Prompt 2: aggiungo al prompt-1 breve descrizione del task

Devi determinare se una stessa parola usata in due frasi differenti ha lo stesso significato in entrambi i contesti. La parola '**minori**' nella frase '*Dopo innumerevoli provini sono stati scelti 15 attori per i ruoli principali, 200 per le parti minori e quasi 4000 comparse...*' ha lo stesso significato della parola '**minore**' nella frase '*Il direttore di Micromega ha detto che ci troviamo di fronte alla scelta fra un bene e un male minore.*'?

A: Sì B: No

Modello	Prompt 1	Prompt 2
zephyr-7b-beta	F1: 57.91	F1: 27.27

EVALITA-LLM

Benchmarking flessibile per LLM in Italiano

- 12 task “originali” italiani (da Evalita 2007-2023)
 - Comprensione del testo italiano
- **Multi-prompt** e single-prompt
- Task a **scelta multipla**: sentiment, test ammissione
- Task **generativi**: fai un riassunto, trova un sinonimo
- Zero-shot e few-shot
- Versione full e light

Prompt 1: domanda con due scelte

La parola '**minori**' nella frase '*Dopo innumerevoli provini sono stati scelti 15 attori per i ruoli principali, 200 per le parti minori e quasi 4000 comparse...*' ha lo stesso significato della parola '**minore**' nella frase '*Il direttore di Micromega ha detto che ci troviamo di fronte alla scelta fra un bene e un male minore.*'?

A: Sì B: No

Prompt 2: aggiungo al prompt-1 breve descrizione del task

Devi determinare se una stessa parola usata in due frasi differenti ha lo stesso significato in entrambi i contesti. La parola '**minori**' nella frase '*Dopo innumerevoli provini sono stati scelti 15 attori per i ruoli principali, 200 per le parti minori e quasi 4000 comparse...*' ha lo stesso significato della parola '**minore**' nella frase '*Il direttore di Micromega ha detto che ci troviamo di fronte alla scelta fra un bene e un male minore.*'?

A: Sì B: No

Single prompt	Prompt 1	Ranking	Single prompt	Prompt 2	Ranking	Multi prompt	Best	Average	Ranking
zephyr-7b-beta	57.91	1	zephyr-7b-beta	27.27	3	zephyr-7b-beta	57.91	42.59	3
gemma-2-9b-it	54.82	2	gemma-2-9b-it	66.13	1	gemma-2-9b-it	66.13	60.47	1
Meta-Llama-3.1-8B	38.28	3	Meta-Llama-3.1-8B	63.64	2	Meta-Llama-3.1-8B	63.64	50.96	2



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Future
Artificial
Intelligence
Research

MAIA - Multimodal AI Assessment

Un benchmark per modelli multimodali in italiano

Collaborazione tra Spoke 2, Spoke 1, Spoke 5 e TP2



UNIVERSITÀ
DI TRENTO



UNIVERSITÀ DI PISA



Istituto di Linguistica
Computazionale
"Antonio Zampolli"

Consiglio Nazionale delle Ricerche

MAIA (Multimodal AI Assessment)

Benchmarking Large Multimodal Models on Reasoning

- 100 video “nativi” Italiani (~30sec, CC-BY)
- 12 categorie di ragionamento temporale, spaziale, causale, controfattuale, implicito, pianificazione, etc.
- 2400 domande generate da persone
- 19200 risposte generate da persone (8 per domanda)
- Valutazione: dati di test “nascosti”
- Task 1: Multiple-choice (caption o foil)
 - *caption* generate via LLM da coppie Q+A
 - *foil* generate via LLM da caption
- Task 2: Visual Question Answering (generativo)
 - Valutazione basata su LLM



Controfattuale

Human Question *Che cosa succederebbe se il pizzaiolo facesse cadere la pizza a terra?*

Human Answer

Sarebbe infastidito e dovrebbe rifarla

Sarebbe costretto a raccogliere e gettare l'impasto

LLM Caption

Se facesse cadere la pizza a terra, il pizzaiolo sarebbe **infastidito** e dovrebbe rifarla

LLM Foil

Se facesse cadere la pizza a terra, il pizzaiolo sarebbe **contento** e dovrebbe rifarla

LLM Caption

Il pizzaiolo sarebbe costretto a raccogliarla e **gettare** l'impasto

LLM Foil

Il pizzaiolo sarebbe costretto a raccogliarla e **mangiare** l'impasto



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Future
Artificial
Intelligence
Research

Conclusioni

- Non rinunciare alla qualità dei contenuti
- Consapevolezza dei rischi dell'AI generativa
- Fondamentale la collaborazione tra ricerca, aziende e PA