

Advanced analytics per la gestione dei processi di distribuzione gas

Progetto a cura di Emanuela Elli



01

...

INTRODUZIONE



Tirocinio Curriculare



Società di consulenza informatica focalizzata sul mercato **Energy & Utilities**.



Progetto finalizzato a raccogliere automaticamente specifiche informazioni nel settore della **finanza sostenibile**.



Direttiva (UE)
2022/2464

Contesto di ricerca

La **finanza sostenibile** è la pratica finanziaria che considera i fattori ESG, incarnando così il concetto di sviluppo sostenibile nell'ambito finanziario.



Gli investitori sono attenti alla reputazione delle aziende in cui investono: dimostrare impegno verso i criteri ESG consolida la fiducia e la reputazione dell'azienda.



Obiettivo



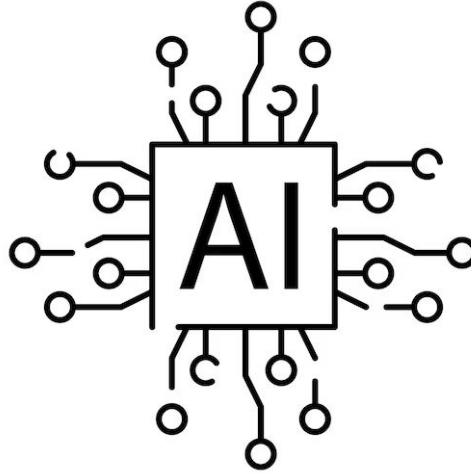
ESG reports

Raccolta ed analisi di **documenti ESG** pubblicati dalle aziende di interesse.

Obiettivo



ESG reports



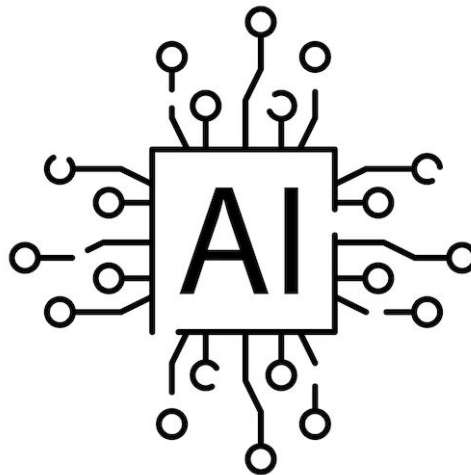
Generative AI models

In particolare **Large Language Models** (LLM), algoritmi di Deep Learning che utilizzando enormi quantità di dati, di diversi formati, producono output testuali coerenti e significativi.

Obiettivo



ESG reports



Generative AI models



Benchmarking

Creare **misure di confronto** per valutare le prestazioni dell'azienda, rispetto ad entità simili o al mercato nel suo complesso.

02

...

IMPLEMENTAZIONE



DATA ACQUISITION



I report analizzati sono presenti sul sito **CDP** (Carbon Disclosure Project) ed acquisiti tramite **web scraping**, in conformità con le politiche della piattaforma.

Aziende considerate:

Italgas, Snam, Enagas, Naturgy Energy Group SA, A2A, ENEL SpA, Iren SpA, ACEA SpA, Korea Gas Corp, Nippon Gas Co Ltd, EDF, Hera.

Search and view past CDP responses

Search for a city or company name

You can also refine your search

Country/Area Year Program Status

2023 Climate Change Submitted

Search

12,460 results

| Name | Response | Year | Status | Score |
|---|---------------------|------|-----------|---------------|
| (ACIP) Alexandria Company for Industrial Packages | Climate Change 2023 | 2023 | Submitted | Not Available |
| (Sichuan) Tianqi Lithium Corporation | Climate Change 2023 | 2023 | Submitted | B |
| 1&1 AG | Climate Change 2023 | 2023 | Submitted | D |
| 1000mercis SA (holding of Numberly) | Climate Change 2023 | 2023 | Submitted | Not Available |
| 16 POINTS CONSULTING LLC | Climate Change 2023 | 2023 | Submitted | Not Available |

Note: Not all companies requested to respond to CDP do so. Companies who are requested to disclose their data and fail to do so, or fail to provide sufficient information to CDP to be evaluated will receive an F. An F does not indicate a failure in environmental stewardship.

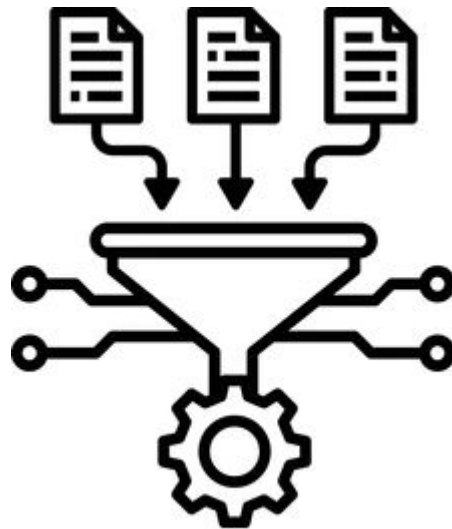
Show 5 10 20

< 1 2 3 4 5 ... 2491 2492 >

DATA PREPARATION



Viene costruito il set di dati contenente la lista di KPI da estrarre ed un **algoritmo di indicizzazione** per individuare le pagine corrispondenti al nome ed al contesto di ogni KPI richiesto.



context_name





kpi_name

```
{('Reporting year', 'Gross global Scope 1  
emissions (metric tons CO2e)': [48],  
  
('Reporting year', 'Scope 2, location-based'):  
[48, 49],  
  
('Purchased goods and services', 'Emissions in  
reporting year (metric tons CO2e)': [50, 57],  
...}
```

MODEL & PROMPT



Vengono implementati e testati diversi modelli presenti nel panorama tecnologico.

| Company | Modello | Numero di parametri | Context window | Dati di training |
|------------|--|---------------------|----------------|--|
| OpenAI | gpt-3.5-turbo-0125  | 175 miliardi | 16k | Multi-genere da fonti pubbliche (Common Crawl, libri, Wikipedia) |
| Databricks | dolly-v2-7b  | 7 miliardi | 2k | Coppie domande e risposta dai dipendenti Databricks |
| Meta | llama-2-7b-chat-hf  | 7 miliardi | 4k | Materiali pubblici (Common Crawl, Wikipedia, libri da Project Gutenberg) |
| Google | gemini  | 100 miliardi | 1 milione | Dati multimodali e multilingue dal Web, libri e dati interni |

PROMPT ENGINEERING



INSTRUCTION PROMPT

Per specificare il **contesto di ricerca** e fornire delle linee guida specifiche per il compito assegnato.



USER INPUT

Contiene le **istruzioni** per guidare il modello nell'estrazione del KPI considerato.

Viene incluso anche il **testo completo** delle pagine selezionate e il KPI di riferimento.



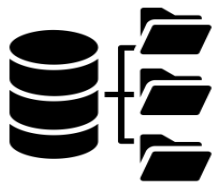
CHECK INPUT

Viene richiesto al modello di esaminare ulteriormente il testo e verificare la **correttezza dell'output** fornito nella fase precedente.

POST PROCESSING



Correzione degli output in formato JSON, eliminando commenti verbali aggiuntivi, e trasformazione dei risultati in Dataframe.



DATA STORAGE



Salvataggio dei risultati in file CSV all'interno dell'ambiente Azure Blob Storage.

DATA COMPARISON



Creazione di un dataset “**ground truth**”, contenente i valori veri presenti nei report. Dopodiché vengono uniti i risultati dei modelli per un confronto diretto.



- **model_name**: nome del modello LLM utilizzato (stringa);
- **company_name**: nome della compagnia a cui fa riferimento il report (stringa);
- **year**: anno di riferimento del report (integer);
- **context**: contesto del KPI analizzato (stringa);
- **kpi_name**: nome del KPI analizzato (stringa);
- **value_model**: valore estratto dal modello (float);
- **value_gt**: valore vero presente nel report (float);
- **in_report**: indica se il KPI analizzato è presente nel report considerato (boolean);
- **kpi_found**: indica se il KPI analizzato è stato trovato dal modello (boolean);
- **is_correct**: indica se il valore estratto dal modello, `value_model`, coincide col valore reale, `value_gt` (boolean).

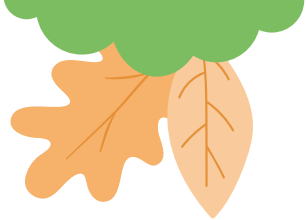
03

...

VALUTAZIONE



Considerazioni

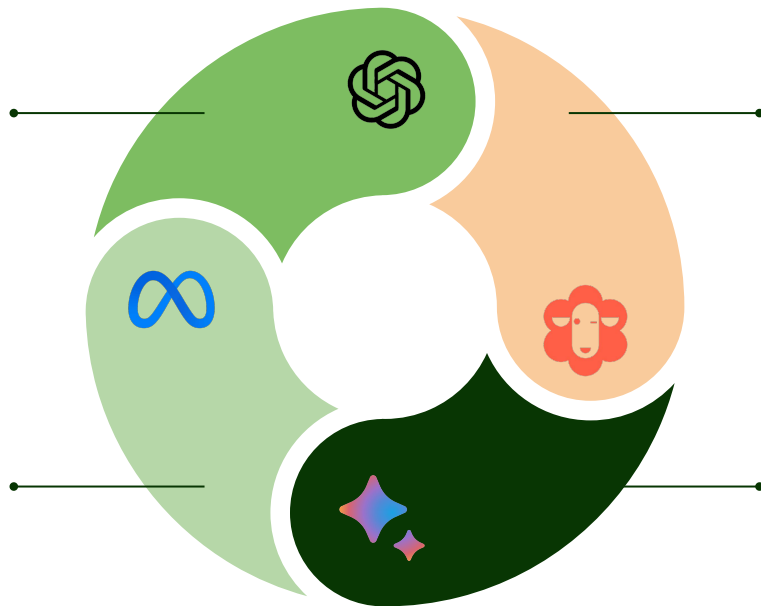


● ● ● GPT3.5

- + Stabilità e affidabilità.
- Pochi commenti aggiuntivi, output non valido per valori nulli.

● ● ● LLAMA

- + Rigoroso nel fornire output idonei.
- Scarsa capacità di comprensione del testo.



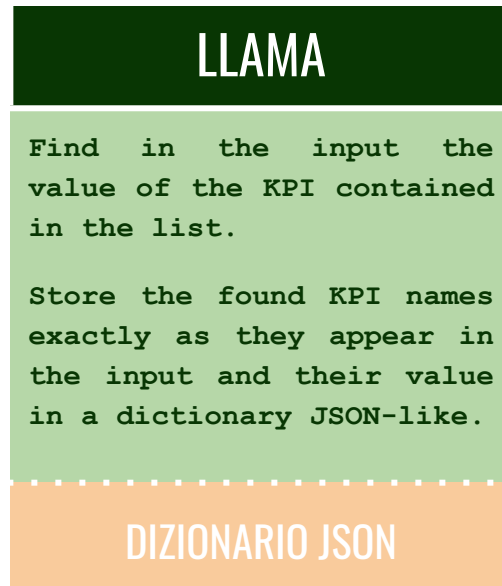
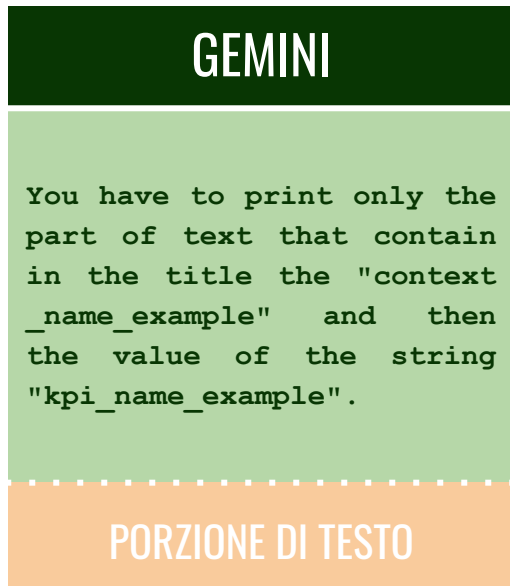
● ● ● DOLLY

- Instabilità, incapacità di strutturare l'output come JSON.

● ● ● GEMINI

- + Capacità nel comprendere il testo.
- Risposte troppo prolisse.

Concatenazione output (Gemini + LLAMA)



**POST
PROCESSING**



Metriche di valutazione

KPI presenti nel report, estratti dal modello ma errati oppure allucinazioni del modello.

KPI presenti nel report, individuati dal modello e correttamente estratti.

| KPI matrix | in_report | kpi_found | is_correct |
|----------------------------|---------------|--------------|----------------|
| True Positive (TP) | TRUE | TRUE | TRUE |
| False Positive (FP) | TRUE FALSE | TRUE TRUE | FALSE FALSE |
| True Negative (TN) | FALSE | FALSE | FALSE |
| False Negative (FN) | TRUE | FALSE | FALSE |

KPI non presenti nel report, non individuati dal modello.

KPI presenti nel report ma non individuati dal modello.

Metriche di valutazione

| KPI matrix | in_report | kpi_found | is_correct |
|---------------------|---------------|--------------|----------------|
| True Positive (TP) | TRUE | TRUE | TRUE |
| False Positive (FP) | TRUE FALSE | TRUE TRUE | FALSE FALSE |
| True Negative (TN) | FALSE | FALSE | FALSE |
| False Negative (FN) | TRUE | FALSE | FALSE |

L'**accuracy** è la percentuale di predizioni corrette fatte dal modello rispetto al totale delle predizioni eseguite.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Metriche di valutazione

| KPI matrix | in_report | kpi_found | is_correct |
|---------------------|---------------|--------------|----------------|
| True Positive (TP) | TRUE | TRUE | TRUE |
| False Positive (FP) | TRUE FALSE | TRUE TRUE | FALSE FALSE |
| True Negative (TN) | FALSE | FALSE | FALSE |
| False Negative (FN) | TRUE | FALSE | FALSE |

Precision indica la proporzione di predizioni positive correttamente identificate rispetto al totale delle predizioni positive fatte dal modello.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Precision = \frac{TP}{\text{predicted true}}$$

Metriche di valutazione

| KPI matrix | in_report | kpi_found | is_correct |
|---------------------|-----------|-----------|------------|
| True Positive (TP) | TRUE | TRUE | TRUE |
| False Positive (FP) | TRUE | TRUE | FALSE |
| | FALSE | TRUE | FALSE |
| True Negative (TN) | FALSE | FALSE | FALSE |
| False Negative (FN) | TRUE | FALSE | FALSE |

Recall (o **sensitivity**) indica la capacità del modello di identificare accuratamente i casi positivi rispetto al totale dei casi positivi.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Precision = \frac{TP}{predicted\ true}$$

$$Recall = \frac{TP}{real\ true}$$

Metriche di valutazione

| KPI matrix | in_report | kpi_found | is_correct |
|---------------------|-----------|-----------|------------|
| True Positive (TP) | TRUE | TRUE | TRUE |
| False Positive (FP) | TRUE | TRUE | FALSE |
| | FALSE | TRUE | FALSE |
| True Negative (TN) | FALSE | FALSE | FALSE |
| False Negative (FN) | TRUE | FALSE | FALSE |

Specificity (o **True Negative Rate**) indica la percentuale di predizioni negative correttamente identificate rispetto al totale delle predizioni negative.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Precision = \frac{TP}{\text{predicted true}}$$

$$Recall = \frac{TP}{\text{real true}}$$

$$Specificity = \frac{TN}{\text{real false}}$$

Metriche di valutazione

| KPI matrix | in_report | kpi_found | is_correct |
|---------------------|---------------|--------------|----------------|
| True Positive (TP) | TRUE | TRUE | TRUE |
| False Positive (FP) | TRUE FALSE | TRUE TRUE | FALSE FALSE |
| True Negative (TN) | FALSE | FALSE | FALSE |
| False Negative (FN) | TRUE | FALSE | FALSE |

F1 measure è la media armonica tra precision e recall.
Utile per trovare un equilibrio tra le due metriche,
offendo valutazione complessiva delle performance.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Precision = \frac{TP}{predicted\ true}$$

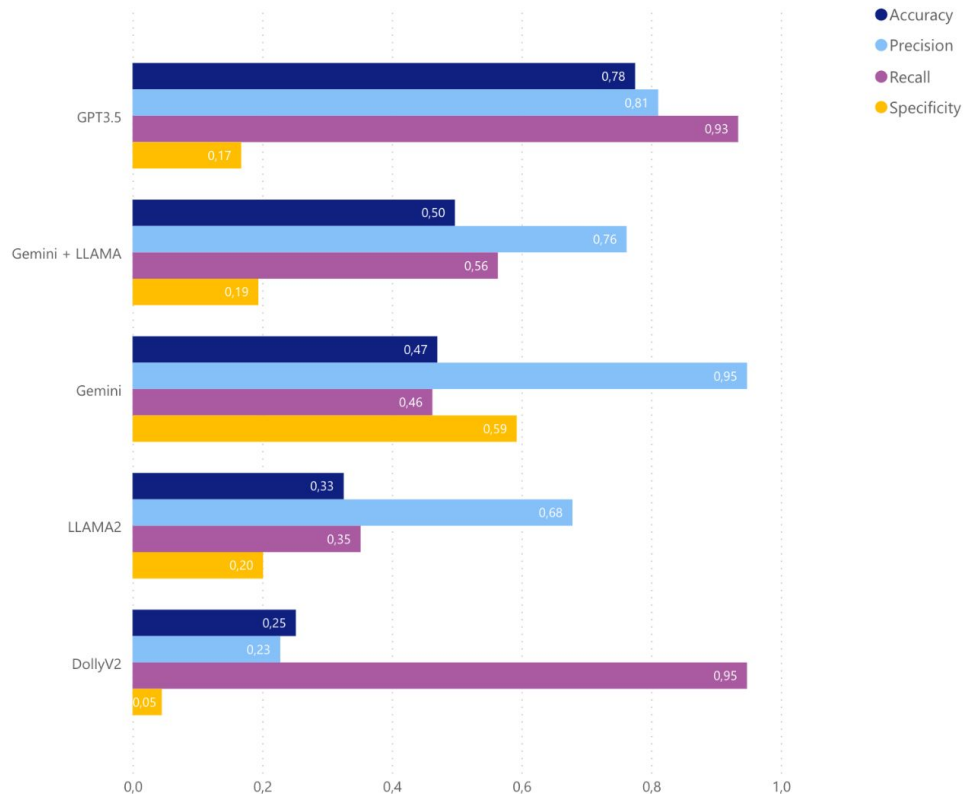
$$Recall = \frac{TP}{real\ true}$$

$$Specificity = \frac{TN}{real\ false}$$

$$F1\ measure = 2 \times \frac{precision \times recall}{precision + recall}$$

Risultati

Accuracy, precision e recall dei modelli utilizzati



- **GPT3.5** è il più affidabile, con elevati valori di *accuracy*, *precision* e *recall*.
- La combinazione di **Gemini** e **LLAMA2** offre prestazioni bilanciate, migliorando sia la *recall* di **Gemini** che la *precision* di **LLAMA2**.
- **Gemini** e **LLAMA2** offrono un equilibrio tra *precision* e *recall*, con **LLAMA2** adatto per situazioni in cui la specificità non è cruciale.
- **DollyV2**, eccelle nella recall identificando molti veri positivi a discapito di tanti falsi positivi.

F1 Measure per ogni modello

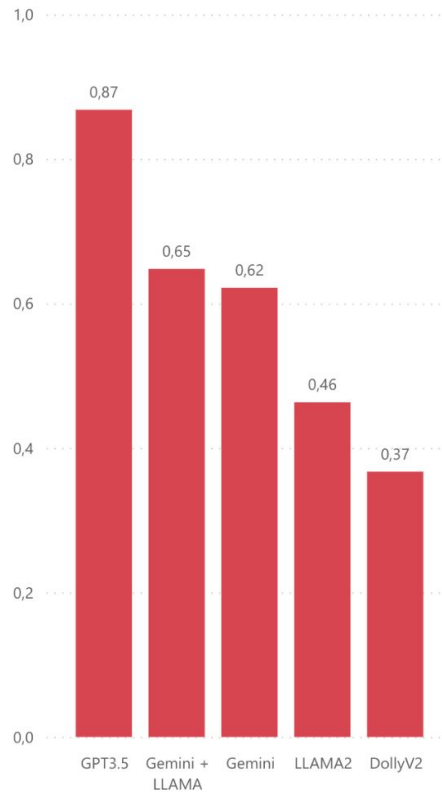
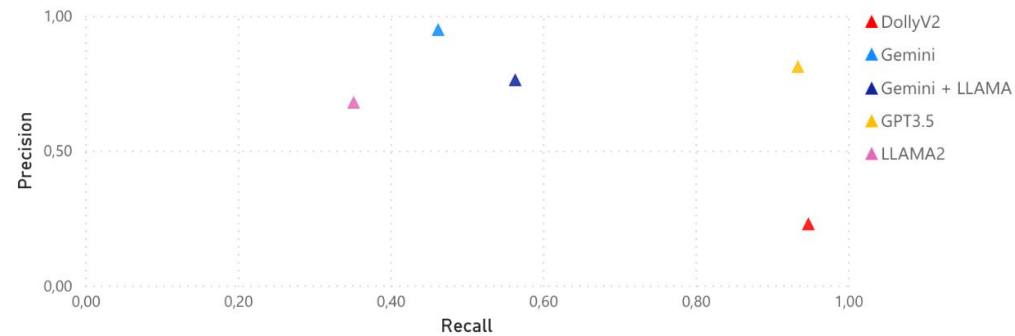
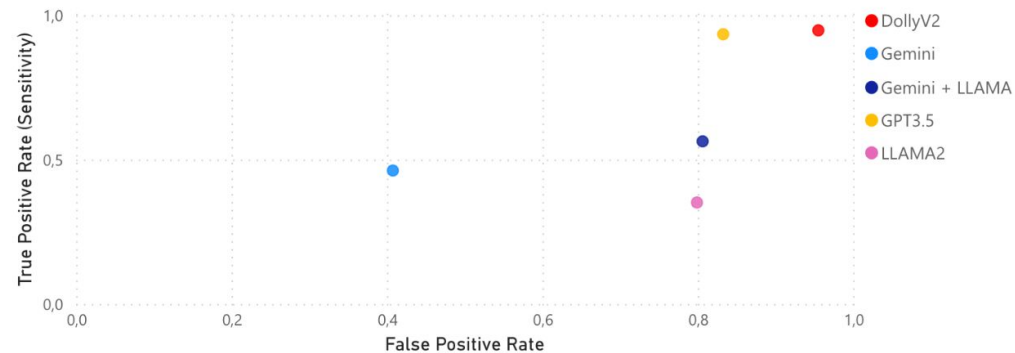


Grafico Precision e Recall



Spazio ROC puntiforme



04

...

CONCLUSIONE E
SVILUPPI
FUTURI

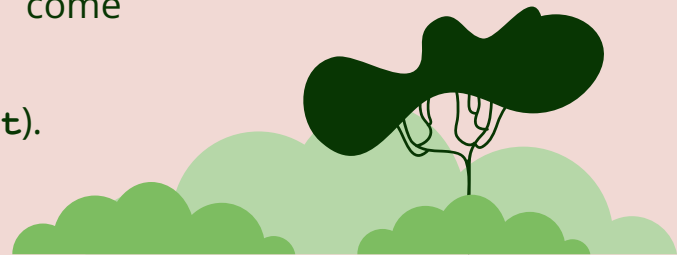


Conclusioni

- ❑ **GPT3.5** risulta il più affidabile nell'individuare correttamente i casi positivi, necessita di miglorie per il controllo dei falsi positivi.
- ❑ **Dolly** ha una sensibilità elevata ma alti falsi positivi. Poco utilizzabile.
- ❑ **Gemini** e **LLAMA**, singolarmente offrono prestazioni moderate, con possibilità di miglioramento per sensibilità e falsi positivi.
- ❑ La concatenazione dei modelli sembra migliorare le criticità dei singoli, suggerendo che tale approccio possa portare a progressi significativi.



- **Standardizzare** la procedura per favorire l'adattamento a ulteriori report e requisiti.
- Utilizzare metodi di messa a punto come **fine-tuning** e **RAG**.
- Merge di layer dei modelli (es. **mergekit**).



...

**GRAZIE PER
L'ATTENZIONE!**

