SPERA

Artificial Intelligence & Cybersecurity

Generative AI in Healthcare: Applications and Evaluation of Effectiveness

Candidate: Lorenzo Zanolin

Dec 13, 2024

Supervisor:

Prof. Giuseppe Serra

Co-Supervisor:

Prof. Jan Steinbrener



Background

- Generative AI goal: derive a probability distribution from data to generate new synthetic data.
- Masked Multi-Head Self Attention: captures global dependencies, process sequences in parallel and generates embeddings that improve performance on a wide range of tasks.



Background

- Generative AI *goal*: derive a probability distribution from data to generate new synthetic data.
- Masked Multi-Head Self Attention: captures global dependencies, process sequences in parallel and generates embeddings that improve performance on a wide range of tasks.

Motivations

- Generative AI can be integrated within workflows to help clinicians.
- Absence of a complete evaluation framework in the literature

 → new framework that combines human feedback and
 metric based feedback.



Salesforce

- Enables companies to manage relationships with customers, prospects and employees through customizable interaction rules.
- Integrates with modules such as Health Cloud and Einstein.



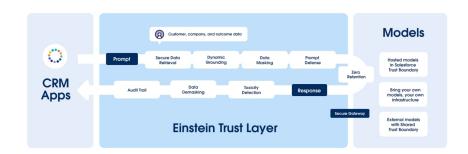
Salesforce

- Enables companies to manage relationships with customers, prospects and employees through customizable interaction rules.
- Integrates with modules such as Health Cloud and Einstein.

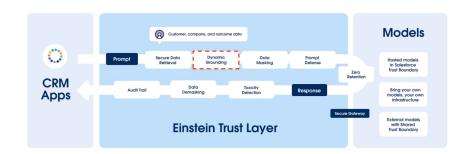
Einstein 1

- Enables the use of genAI models within Salesforce.
- Allows for the creation of reusable prompts with database integration (RAG).
- Trust Layer framework to ensure data security during information exchange with LLMs.

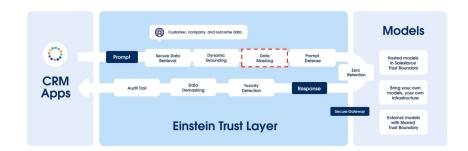




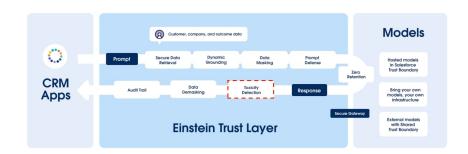




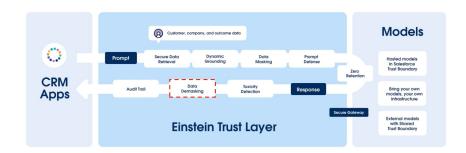












- Application intended for healthcare clinics that manage numerous patients and offer a wide range of services.
- Tasks to be implemented with Copilot:

- Application intended for healthcare clinics that manage numerous patients and offer a wide range of services.
- Tasks to be implemented with Copilot:
 - Patient Summary

- Application intended for healthcare clinics that manage numerous patients and offer a wide range of services.
- Tasks to be implemented with Copilot:
 - Patient Summary
 - List Possible Problems

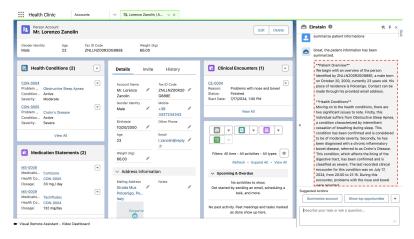
- Application intended for healthcare clinics that manage numerous patients and offer a wide range of services.
- Tasks to be implemented with Copilot:
 - Patient Summary
 - List Possible Problems
 - Send Visit Details

Invoked by the doctor before receiving the patient: quick and detailed overview of all the patient's conditions.



Patient Summary

Invoked by the doctor before receiving the patient: quick and detailed overview of all the patient's conditions.

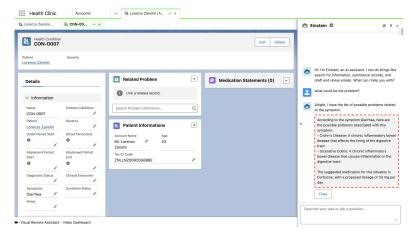


Identifies potential conditions based on the patient's symptoms, providing a list of possible diagnoses and medications, with dosage calculations for of them.



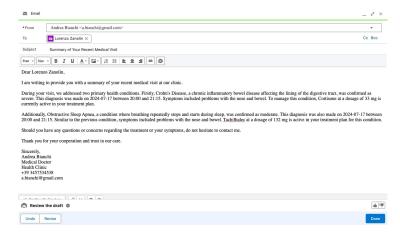
List Possible Problems

Identifies potential conditions based on the patient's symptoms, providing a list of possible diagnoses and medications, with dosage calculations for of them.



Creates a draft email summarizing the patient's latest Clinical Encounter for the doctor to review and edit.

Creates a draft email summarizing the patient's latest Clinical Encounter for the doctor to review and edit.



For each task, the output of each model was evaluated in comparison to the clinician's notes.

For each task, the output of each model was evaluated in comparison to the clinician's notes.

Automatic Evaluation:

 Metrics: ROUGE, BLEU, METEOR, Word2Vec and BERTScore.



For each task, the output of each model was evaluated in comparison to the clinician's notes.

Automatic Evaluation:

 Metrics: ROUGE, BLEU, METEOR, Word2Vec and BERTScore.

Human Evaluation:

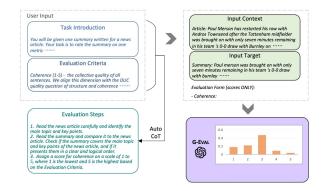
- A sample of 20 physicians tested the system, providing feedback via a Likert scale (1-5) on four aspects:
 - Accuracy: alignment with clinician notes
 - Relevance: appropriateness to the query
 - Coverage: completeness of information
 - Clarity: syntax and overall quality



Evaluation Framework - cont'd

G-Eval Evaluation:

- Framework used to evaluate model outputs, consistently using the clinician's notes as a reference.
- Compared against human evaluation.





References I