

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

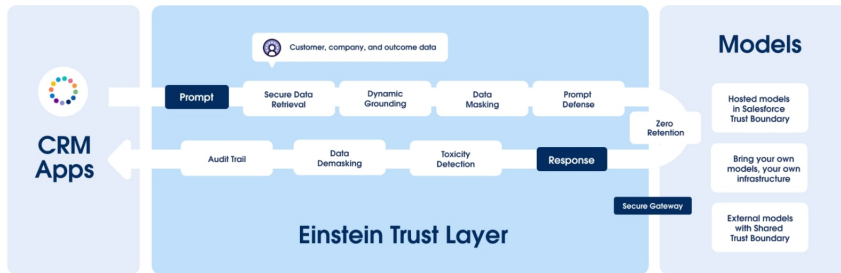
- Enables companies to manage relationships with customers, prospects and employees through customizable interaction rules.
- Integrates with modules such as Health Cloud and Einstein.

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.

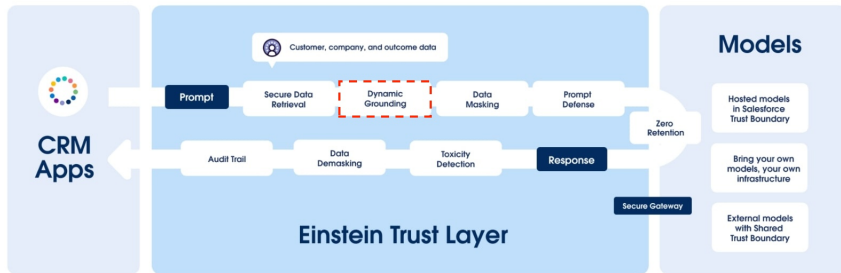


Einstein Trust Layer



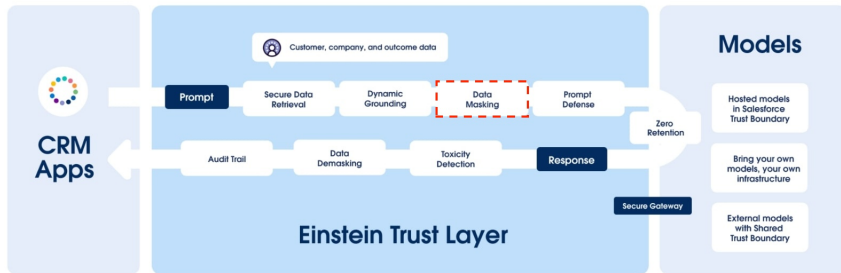


Einstein Trust Layer



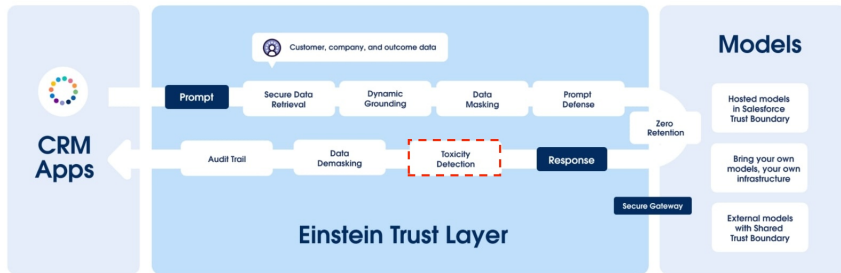


Einstein Trust Layer



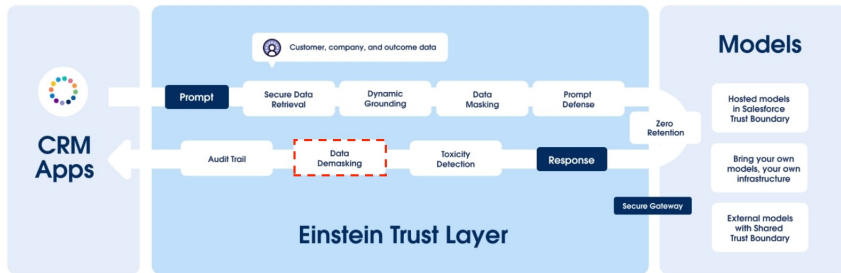


Einstein Trust Layer





Einstein Trust Layer





Health Clinic

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



Health Clinic

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



Health Clinic

- 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



Health Clinic

- 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



Patient Summary

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.

Health Clinic | Accounts | **Lorenzo Zanolin | A...**

Person Account
Mr. Lorenzo Zanolin [Edit] [Delete]

Gender Identity: Male | Age: 23 | Tax ID Code: ZNLLNZ00R20G888E | Weight (Kg): 66.00

Health Conditions (2)

- CON-0004: Obstructive Sleep Apnea (Active, Moderate Severity)
- CON-0005: Crohn's Disease (Active, Severe Severity)

Medication Statements (2)

- MS-0008: Cortisone (33 mg / day)
- MS-0009: Tachifluadec (132 mg/day)

Details | Invite | History

Account Name: Mr. Lorenzo Zanolin | Tax ID Code: ZNLLNZ00R20G888E

Gender Identity: Male | Mobile: +39 3427234343

Birthdate: 10/20/2000 | Other Phone: []

Age: 23 | Email: l.zanolin@reply.it

Weight (Kg): 66.00

Address Information

Mailing Address: Strada Mus Polcenigo, Po... Italy

Clinical Encounters (1)

CE-0004: Problems with nose and bowel (Status: Finished, Start Date: 7/17/2024, 1:00 PM)

Upcoming & Overdue

No activities to show. Get started by sending an email, scheduling a task, and more.

No past activity. Past meetings and tasks marked as done show up here.

Einstein | summarize patient informations

Great, the patient information has been summarized.

****Patient Overview****
We begin with an overview of the person identified by ZNLLNZ00R20G888E, a male born on October 20, 2000, currently 23 years old. His place of residence is Polcenigo. Contact can be made through his provided email address.

****Health Conditions****
Moving on to the health conditions, there are two significant issues to note. Firstly, this individual suffers from Obstructive Sleep Apnea, a condition characterized by intermittent cessation of breathing during sleep. This condition has been confirmed and is considered to be of moderate severity. Secondly, he has been diagnosed with a chronic inflammatory bowel disease, referred to as Crohn's Disease. This condition, which affects the lining of the digestive tract, has been confirmed and is classified as severe. The last recorded clinical encounter for this condition was on July 17, 2024, from 20:00 to 21:15. During this encounter, problems with the nose and bowel were reported.

Suggested Actions

Summarize account | Show top opportunities

Describe your task or ask a question...



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.



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.

The screenshot displays a healthcare software interface. At the top, there's a navigation bar with 'Health Clinic' and 'Accounts' tabs. Below this, a patient record for 'Lorenzo Zanolin' is shown. The main area is divided into several sections:

- Health Condition CON-0007**: Includes fields for Patient (Lorenzo Zanolin) and Severity.
- Details**: A section with a dropdown for 'Information' and a table of fields including Name (CON-0007), Patient (Lorenzo Zanolin), Onset Period Start, Abatement Period Start, Diagnostic Status, Symptoms (Diarrhea), Notes, Problem Definition, Severity, Onset Period End, Abatement Period End, Clinical Encounter, and Condition Status.
- Related Problem**: A section with a link to a related record and a search bar for problem definitions.
- Patient Informations**: A section with fields for Account Name (Mr. Lorenzo Zanolin), Age (23), and Tax ID Code (ZNLN200R20G888E).
- Medication Statements (0)**: A section for medication statements.

On the right side, there's a chat window titled 'Einstein'. It contains the following text:

Hi I'm Einstein, an AI assistant. I can do things like search for information, summarize records, and draft and revise emails. What can I help you with?

what could be the problem?

Alright, I have the list of possible problems related to the symptom.

According to the symptom Diarrhea, here are the possible problems associated with this symptom:

- Crohn's Disease: A chronic inflammatory bowel disease that affects the lining of the digestive tract.
- Ulcerative Colitis: A chronic inflammatory bowel disease that causes inflammation in the digestive tract.

The suggested medication for this situation is Cortisone, with a proposed dosage of 33 mg per day.

Copy

Describe your task or ask a question...



Send Visit Details

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



Send Visit Details

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

Email

* From

Andrea Bianchi <a.bianchi@gmail.com>

To

Lorenzo Zanolin

Cc Bcc

Subject

Summary of Your Recent Medical Visit

Font

Size

Dear Lorenzo Zanolin,

I am writing to provide you with a summary of your recent medical visit at our clinic.

During your visit, we addressed two primary health conditions. Firstly, Crohn's Disease, a chronic inflammatory bowel disease affecting the lining of the digestive tract, was confirmed as severe. This diagnosis was made on 2024-07-17 between 20:00 and 21:15. Symptoms included problems with the nose and bowel. To manage this condition, Cortisone at a dosage of 33 mg is currently active in your treatment plan.

Additionally, Obstructive Sleep Apnea, a condition where breathing repeatedly stops and starts during sleep, was confirmed as moderate. This diagnosis was also made on 2024-07-17 between 20:00 and 21:15. Similar to the previous condition, symptoms included problems with the nose and bowel. Tachifludex at a dosage of 132 mg is active in your treatment plan for this condition.

Should you have any questions or concerns regarding the treatment or your symptoms, do not hesitate to contact me.

Thank you for your cooperation and trust in our care.

Sincerely,
Andrea Bianchi
Medical Doctor
Health Clinic
+39 3457534538
a.bianchi@gmail.com

Review the draft

Undo

Revise

Done



Evaluation Framework

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



Evaluation Framework

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.



Evaluation Framework

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

Automatic Evaluation:

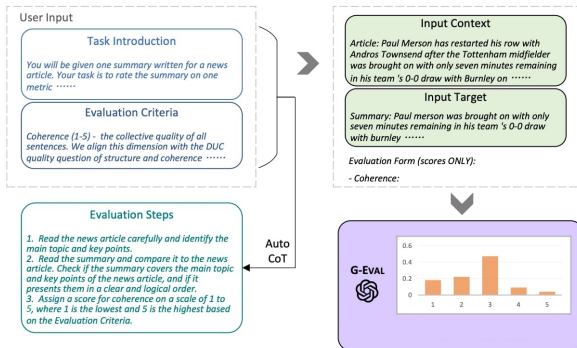
- Metrics: ROUGE, BLEU, METEOR, Word2Vec and BERTScore.

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

G-Eval Evaluation:

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





Automatic Evaluation

Model	Rouge1	Rouge2	RougeL	BLEU	METEOR	Word2Vec	BERTscore
Patient Summary							
GPT4	0.3966	0.1139	0.2380	0.1470	0.2198	0.9035	0.8458
GPT4 32k	0.3963	0.1043	0.2439	0.1742	0.2349	0.9150	0.8428
GPT4 Omnni	0.4286	0.1438	0.2857	0.2013	0.2692	0.8987	0.8572
Anthropic	0.3761	0.1073	0.2222	0.1054	0.1905	0.8702	0.8504
List Possible Problems							
GPT4	0.4444	0.1649	0.3838	0.1851	0.3690	0.8496	0.9007
GPT4 32k	0.4228	0.1322	0.2927	0.2180	0.2726	0.8819	0.8881
GPT4 Omni	0.3800	0.0612	0.2600	0.1366	0.2709	0.8529	0.8789
Anthropic	0.3579	0.0645	0.2526	0.1221	0.2480	0.8325	0.8810
Email Generation							
GPT4	0.3697	0.0845	0.2129	0.1704	0.2991	0.9164	0.8510
GPT4 32k	0.3371	0.0460	0.1771	0.1153	0.2811	0.8801	0.8443
GPT4 Omnni	0.4536	0.1295	0.2526	0.2345	0.3605	0.9295	0.8735
Anthropic	0.4375	0.1166	0.2321	0.2037	0.3037	0.9219	0.8739

Mean and standard deviation of Human Evaluation scores.

Model	Accuracy	Relevance	Coverage	Clarity
Patient Summary				
GPT4	$\mu = 3.83, \sigma = 0.62$	$\mu = 3.72, \sigma = 0.75$	$\mu = 3.06, \sigma = 0.73$	$\mu = 3.83, \sigma = 0.71$
GPT4 32k	$\mu = 3.72, \sigma = 0.67$	$\mu = 3.78, \sigma = 0.81$	$\mu = 3.22, \sigma = 0.94$	$\mu = 4.06, \sigma = 0.54$
GPT4 Omni	$\mu = 3.83, \sigma = 0.86$	$\mu = 3.50, \sigma = 1.10$	$\mu = 3.33, \sigma = 0.77$	$\mu = 3.89, \sigma = 0.76$
Anthropic	$\mu = 3.83, \sigma = 1.04$	$\mu = 3.83, \sigma = 0.99$	$\mu = 3.67, \sigma = 0.59$	$\mu = 4.17, \sigma = 0.86$
List Possible Problems				
GPT4	$\mu = 3.95, \sigma = 0.87$	$\mu = 4.00, \sigma = 0.84$	$\mu = 3.61, \sigma = 0.85$	$\mu = 3.94, \sigma = 0.73$
GPT4 32k	$\mu = 3.83, \sigma = 0.79$	$\mu = 3.94, \sigma = 0.73$	$\mu = 3.50, \sigma = 0.86$	$\mu = 4.17, \sigma = 0.79$
GPT4 Omni	$\mu = 3.90, \sigma = 1.08$	$\mu = 4.00, \sigma = 0.97$	$\mu = 3.72, \sigma = 1.02$	$\mu = 4.05, \sigma = 0.80$
Anthropic	$\mu = 3.90, \sigma = 0.90$	$\mu = 4.06, \sigma = 0.80$	$\mu = 3.72, \sigma = 1.02$	$\mu = 3.94, \sigma = 0.94$
Email Generation				
GPT4	$\mu = 3.17, \sigma = 0.98$	$\mu = 3.22, \sigma = 0.88$	$\mu = 2.94, \sigma = 0.99$	$\mu = 3.28, \sigma = 0.89$
GPT4 32k	$\mu = 3.00, \sigma = 1.08$	$\mu = 3.56, \sigma = 0.92$	$\mu = 3.33, \sigma = 1.00$	$\mu = 3.44, \sigma = 1.20$
GPT4 Omni	$\mu = 3.78, \sigma = 0.94$	$\mu = 4.06, \sigma = 0.64$	$\mu = 3.72, \sigma = 0.67$	$\mu = 3.94, \sigma = 0.94$
Anthropic	$\mu = 4.06, \sigma = 0.80$	$\mu = 4.11, \sigma = 0.76$	$\mu = 3.90, \sigma = 0.90$	$\mu = 4.06, \sigma = 0.97$

Inter-rater agreement

for all pairs of raters

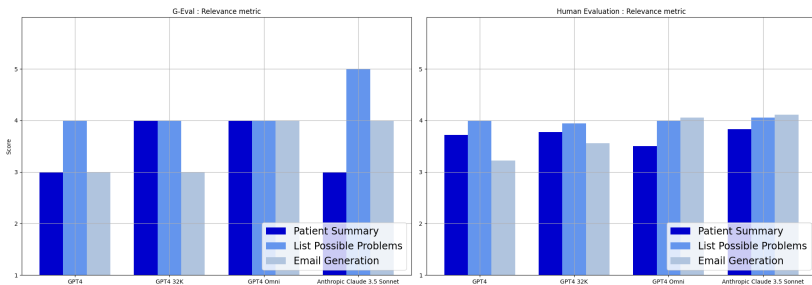
$$P = \{\{R_i = [r_{i,1}, \dots, r_{i,n}], R_j = [r_{j,1}, \dots, r_{j,n}]\} | R_i, R_j \in R, R_i \neq R_j\}$$

the corresponding $\kappa_{w_{i,j}}$ values were calculated.

Subsequently, the average $\kappa_w = \frac{1}{|P|} \sum_{(i,j) \in P} \kappa_{w_{i,j}}$ was computed.

GPT4	GPT4 32K	GPT4 Omni	Anthropic
Patient Summary			
$\kappa_w = 0.25$	$\kappa_w = 0.29$	$\kappa_w = 0.37$	$\kappa_w = 0.52$
List Possible Problems			
$\kappa_w = 0.61$	$\kappa_w = 0.63$	$\kappa_w = 0.63$	$\kappa_w = 0.57$
Email Generation			
$\kappa_w = 0.53$	$\kappa_w = 0.61$	$\kappa_w = 0.69$	$\kappa_w = 0.72$

Comparison of each metric between G-Eval and Human scores.



Results

- Generative AI can increase productivity
- New proposed framework that includes human and automatic evaluation
 - Sematical metrics work better in this context
 - Raters gave similar scores to shorter text

Results

- Generative AI can increase productivity
- New proposed framework that includes human and automatic evaluation
 - Sematical metrics work better in this context
 - Raters gave similar scores to shorter text

Future Directions

- Conduct a number of measurements with G-Eval equal to the number of raters to compare G-Eval with Human Feedback.