# הגדרה

המערכת היא אפליקציית תשאול והשוואת תשובות מודל LLM.

האפליקציה מקבלת מידע המוגדר כאמת. מתשאלת מודל עבור חלק מהמידע הנתון ומקבלת תחזית עבור החלק האחר. האפליקציה אומדת את יכולת המודל אל מול האמת הנתונה. האפליקציה מיישמת שיטה לשיפור המודל ע"י שינוי האמידה של המודל את המידע.

# תלויות

מידע גולמי נתון – dataset המכיל את המידע המוגדר כאמת.

בחירת מודל LLM – המודל אותו מתשאלים.

הגדרת מטריקת מרחק תחזית מהאמת.

סביבת ריצה – יכולות עיבוד CPU

בחירת אספקטים לניתוח – איזה אלמנטים בביקורת מחפשים

# הנחות

מודל ה-LLM אינו מצליח לענות על הדרישה לתחזית בצורה מושלמת

זמן התשאול לקבלת התחזית סביר מבחינת זמן ריצה

# מגבלות

מודל – עבור chatGPT 50$ תקציב

GPU – זמן ריצה לעיבוד מוגבל

אחסון – גודל המודל איתו עובדים במידה והוא מקומי

# דרישות פונקציונאליות

* קריאת ה-dataset ופרסור שלו
* תשאול המודל עם מידע
* אחזקת מבנה נתונים שיחזיק את התשובות של המודל
* שמירת התשובה של המודל בצורה פרסיסטנטית
* קריאת התשובות של המודל מהעבר
* חישוב מרחק התשובות של המודל מהאמת עבור כל רשומה
* תשאול המודל עם המידע עבור כל אספקט נבחן
* אחזקת כל התשובות של המודל עבור פריט מידע לכל האספקטים
* מיצוע משוקלל של התשובות לפי האספקטים
* השוואת כל זוגות המרחקים ויצירת סכימה כללית בכמה אחוזים היה שינוי לטובה / רעה
* ביצוע איטרציות על עבודת האפליקציה
* בדיקה ידנית של מספר דוגמאות לציונים לפי אספקטים

# דרישות לא פונקציונאליות

* יעילות
* עמידות בשגיאות
* ביצועים
* אמינות

# Use Cases

* קריאת dataset בצורת קובץ csv
* יצירת מבנה נתונים של ה-dataset באפליקציה
* פילטור המידע הדרוש לתשאול
* הוספת prompt למידע לתשאול
* תשאול המודל
* יצירת מבנה נתונים של התשובות
* אחסון התשובות בצורה פרסיסטנטית
* קריאת התשובות הפרסיסטנטיות
* חישוב המרחק של התחזית מהאמת
* יצירת מבנה נתונים של המרחקים
* אחסון המרחקים בצורה פרסיסטנטית
* קריאת המרחקים בצורה פרסיסטנטית
* יצירת מבנה נתונים עם האספקטים הנבחנים
* יצירת מבנה נתונים לכל המידע עבור כל אספקט
* יצירת מבנה נתונים של תשובות התשאול עבור האספקטים
* אחסון התשובות בצורה פרסיסטנטית
* קריאת התשובות הפרסיסטנטיות
* יצירת ממוצעים משוקללים של התשאולים לפי אספקטים לכל מידע
* יצירת מבנה נתונים לממוצעים המשוקללים כתחזית
* אחסון הממוצעים בצורה פרסיסטנטית
* קריאת הממוצעים הפרסיסטנטיות
* חישוב המרחק של התחזית מהאמת
* יצירת מבנה נתונים של המרחקים
* אחסון המרחקים בצורה פרסיסטנטית
* קריאת המרחקים בצורה פרסיסטנטית
* השוואת המרחקים של התשאול הרגיל והתשאול המפורק
* יצירת מבנה נתונים להצלחה / כישלון עבור כל מידע
* חישוב אחוז ההצלחה הכללי של הניסוי
* איטרציות ביצוע על התשאולים
* בדיקה ידנית של תשובות המודל לפי אספקטים

# דרישות משאבים

# ארכיטקטורה ודיזיין

שלבי עבודה