

Milestone 4: Final prototype and evaluation

By: Gilat Toker and Dalia Kipervasser

סיכום הפרויקט -

בפרויקט זה נרצה לתת מענה לבעיה אקטואלית במיוחד בימינו - הסברה ישראלית ותגובה לביטויי אנטישמיות ברשתות החברתיות. המערכת שלנו תאפשר למשתמשים שונים, ללא תלות ברמת ההתנסחות, בפערי השפה וברמת הניסיון, ליצור טקסטים איכותיים ורלוונטיים אשר מגבירים חשיפה בעניין הסברה ישראלית. הפתרון שאנו מציעות הוא מערכת חכמה שתומכת בשלל פונקציות, ביניהן: (1) בהינתן טקסט אנטי-ישראלי, תדע לייצר טקסט תגובה אליו, (2) בהינתן פוסט עבר שלך - תדע לייצר טקסט חדש בהשראת הטקסט שלך, תוך כדי שמירת אלמנטים ייחודיים לך, (3) בהינתן וורסיה ראשונית שלך, תדע לשכתב ולייצר פוסט איכותי ותקין ו-(4) תייצר טקסט חדש לגמרי. המערכת מבוססת על LLM אשר יש ביכולו ליצור טקסט מעניין ובהיר ועל דאטהסט שייצרנו ע"י סריקה מ - Reddit. האלמנט הייחודי של המערכת הוא המיקוד בנושא ספציפי - פוסטים פרו ישראלים ברשתות חברתיות. אומנם קיימים פתרונות דומים בסקטורים אחרים, אך הם מכוונים למגזר העסקי, לא מאפשרים להביא לידי ביטוי את הסגנון האישי של המשתמש ונמנעים משיח בנושאים רגישים כמו זה.

תיאור המערכת -

קישור למערכת: <https://github.com/DaliaKip/SpeechUp>

ממשק משתמש: עיקר תרומת המערכת היא היכולת ליצור טקסט חדש ואיכותי המותאם אישית לך בנושא רגיש כמו הסיכסוך-הישראלי-פלסטיני. כפי שתואר בסיכום, המערכת תומכת ב-4 פונקציות עיקריות, כאשר ב-3 מהן המשתמש מכניס טקסט ראשוני שלו. רצינו כי ממשק המשתמש יהיה נקי וברור ובכל שלב יציג רק את החלקים הרלוונטים לזמן זה, על מנת לא לגרום לעומס קוגניטיבי על המשתמש. תיאור רכיבי הממשק:

- לוגו של המערכת
- הסבר קצר על הפונקציות בהן המערכת תומכת
- אזור קלט טקסט, בו המשתמשים יכולים לכתוב או להדביק את התוכן אותו הם רוצים שהמערכת תעבד.
- סרגל אורך וסרגל יצירתיות, באמצעותן המשתמש יכול לשלוט בכמות המילים בטקסט המיוצר וב"יצירתיות" שלו (כלומר כמה מפתיע הטקסט המיוצר יהיה).
- 4 כפתורים צבעוניים המושכים את העיון עם הכיתוב: "תגובה לטקסט זה", "בסגנון שלי", "שכתוב", ו-"צור פוסט חדש", אשר תואמים ל-4 המשימות העיקריות שהמערכת תומכת בהן.

בעקבות כל האמור לעיל, בחרנו כי ממשק המשתמש בעת כניסה ראשונית למערכת יראה כך:



Enter input and click on the desired mode

reply to this text

my style

rewrite

generate a new post

Enter your text here

shorter!

300

longer!

less creative!

1.00

more creative!

לאחר שהמשתמש לווחץ על אחד הכפתורים, הוא מקבל את הטקסט שהמערכת יצרה ו-2 כפתורים נוספים נחשפים לפניו: (1) כפתור תרגום - אשר מוביל לתרגום הטקסט לעברית ו (2) כפתור ג'נרט מחדש - אשר מאפשר למשתמש לקבל אופציה נוספת לאותה הקריאה במידה ולא אהב את הגרסא הראשונה.

Let us help you make your text better! Enter input and click on the desired mode

reply to this text

my style

rewrite

generate a new post

Bombings in gaza

shorter!

300

longer!

less creative!

1.00

more creative!

פוסט התגובה שלך מוכן: In the face of challenges such as rocket attacks targeting civilian population centers, Israel has an inherent obligation to defend its citizens and safeguard their security. The IDF does not act aggressively but defensively, and it takes every precaution to limit civilian casualties and damage to infrastructure. It's noteworthy that Israel often sends warnings ahead of time to minimize harm to innocent individuals

Translate

Regenerate

לבסוף, המערכת מספקת ממשק נוח לצפיה בכל הטקסטים שהמשתמש יצר ע"י המערכת, מסודרים באופן כרונולוגי לפי תאריך יצירתם. על מנת לא ליצור עומס קוגניטיבי, בכל פעם מוצג למשתמש טקסט אחד והוא יכול לדפדף בין הטקסטים השונים שייצר.

1/2: The issue at hand is both complex and sensitive, and it is crucial to approach it with understanding and care. In the face of challenges such as rocket attacks targeting civilian population centers, Israel has an inherent obligation to defend its citizens and safeguard their security. The IDF does not act aggressively but defensively, and it takes every precaution to limit civilian casualties and damage to infrastructure. It's noteworthy that Israel often sends warnings ahead of time to minimize harm to innocent individuals.

Date: 2024-04-07 18:28:55

תוך כדי עיצוב ממשק המערכת, שמנו דגש מיוחד על עקרונות העיצוב הבאים (בהתאם ל- Guidelines for human-AI interaction):

- בהירות לגבי מה המערכת מסוגלת לעשות (G1) - בזמן בניית ממשק המשתמש, שמנו דגש עיקרי על נקודה זו. כפי שתוארנו, במסך הפתיחה המשתמש נחשף תחילה ללוגו ולהסבר קצר על שימוש במערכת - כך אנו מאפשרים למשתמש "להיכנס" לעולם המערכת ולהבין על מה מדובר. לאחר מכן, מופיעה תיבת טקסט, 2 סרגלים לשליטה באורך ויזריות ולבסוף- 4 כפתורים צבעוניים המייצגים את ארבעת הפונקציות בהן המערכת תומכת. בניה זו רומזת למשתמש על סדר הפעולות ההגיוני - תחילה הכנסה של טקסט ראשוני, לאחר מכן דרישה על הטקסט המיוצר ולבסוף בחירה באחת מ-4 הפונקציות באפשרויות. בנוסף, הוספנו הסבר קצר במעבר על הכפתור "my style" אשר שמנו לב שפחות אינטואיטיבי למשתמשי המערכת.
- הימנעות מהטיות (G6) - הניסוחים בפקודות והמידע שהמערכת מספקת מנוסחים בלשון נייטרלית, ששמה פוקוס על המערכת עצמה, ולא בפניה ישירה למשתמש.
- תמיכה בתיקונים (G9) - נאפשר למשתמש לייצר טקסט חדש נוסף בכל אחד ממצבי הג'נוט בהם המערכת תומכת. במידה והמשתמש אינו מרוצה מהתוצאה, ניתן לייצר אחת חדשה בקלות ובמהירות, בלי שהמשתמש ייצטרך להכניס את המידע פעם נוספת.
- זכור אינטרקציות אחרונות (G12)- המערכת שומרת את כל הטקסטים המיוצרים ומאפשרת גישה קלה ומסודרת אליהם. על מנת לתמוך בעיקרון זה, המערכת מכילה הפניה לזיכרון בו ניתן למצוא את הפוסטים האחרונים שנוצרו.

* הסבר לגבי בניית ממשק המשתמש נמצא תחת הגישה האלגוריתמית.

סרטון ממשק המשתמש -

https://www.youtube.com/watch?v=KSmq0KSlkME&ab_channel=DalyK

גישת אלגוריתמית:

הקומפוננטה העיקרית של המערכת שלנו היא כמובן מודל ה LLM שמייצר את הטקסט. על מנת ליצור את הטקסט השתמשנו ב-API של מודל השפה gpt-4. בחירת המודל הנ"ל ולא שימוש במודל זול יותר (למשל gpt-3.5) נבעה מהצורך בשימוש במודל אשר מעודכן ומודע לאירועי ה 7 לאוקטובר. שימוש ב-API נעשה באמצעות ייבוא ספריית from openai import OpenAI.

קושי שנחשפנו אליו במהלך העבודה עם המודל היא שפעמים רבות הוא לא רצה לבצע את המשימה, מכיוון שזהו נושא רגיש ושנוי במחלוקת וחברת openAI שמה מגבלות על ג'נרט טקסט בנושאים אלו. על מנת להתמודד עם המכשול הנ"ל, דייקנו את ההסבר בprompt והסברנו שמטרת המערכת היא ליצור טקסטים מכבדים, אשר לא פוגעים באף אחד מהצדדים, הוספנו את השימוש ב-"role": "system" והשתמשנו בשיטת few-shot להשראה. כל השינויים האלו אכן פתרו את הבעיה. את הטקסט שאנו קולטים מהמשתמש ובחירתו לגבי האורך והיצירתיות אנו שומרים ומעבירים למודל. בנוסף, קיימת אופציית תרגום של הטקסט שנוצר, שגם כן מתבצע ע"י המודל. במהלך העבודה על הפונקציה של "ג'נרט טקסט מחדש" נחשפנו בבעיה כי רוב הטקסטים דומים מאוד, ניטרלים ולא נשמעים כמו פוסט של אדם עם דעה אמיתית. על מנת לשפר זאת, יצרנו סט נתונים עם פוסטים הנוגעים לסכסוך הישראלי-פלסטיני ובעת יצירת טקסט חדש אנו שולפים את אחד הטקסטים מסט הנתונים ומעבירים אותו למודל כהשראה. כך, למודל יש השראה מפוסט אמיתי והפוסטים שנוצרים הפכו להיות מעניינים ואמינים. את איסוף הדאטה ביצענו ע"י סריקה מאתר reddit ובפרט מ sub-reddit שנקרא: 'r/IsraelPalestine', אשר עוסק בסכסוך הישראלי-פלסטיני. עשינו זאת בעזרת ספריית praw. בעת בניית הדאטהסט, סיננו את הפוסטים ע"י פרמטרים כמו: אורך הפוסט, דירוג הפוסט.

מימוש הפן האינטראקטיבי של המערכת- על מנת לממש את האלמנט האינטראקטיבי במערכת השתמשנו בחבילה ipywidgets המאפשרת להציג אלמנטים ויזואליים ולממש אינטראקציות של המשתמש עם המערכת, בצורה אלגנטית, ברורה למשתמש ופשוטה למימוש. בנוסף השתמשנו באפשרויות השונות של המודל IPython.display ושילבנו בחלקים בקוד HTML. המימוש מאפשר למשתמש לבצע אינטראקציה ישירה דרך מחברת הג'ופיטר.

אינטראקציית משתמשים -

במסגרת הפרויקט, נתנו ל-10 חברים שלנו, אשר לא שמעו על המערכת מראש, להתנסות ולהשתמש במערכת. המשתמשים התבקשו תחילה לשחק עם המערכת כרצונם ולאחר מכן לבצע מספר משימות ספציפיות, ביניהן: יצירת תגובה לטקסט אנטי-ישראלי, יצירת פוסט חדש בסגנון אישית שכתוב פוסט קיים, יצירת טקסט חדש לגמרי, תרגום שלהם לעברית.

מצאנו כי ממשק המשתמש ברור ומכוון את המשתמש לשימוש נכון עם המערכת. המשתמשים הצליחו להתמצא במערכת בלי הכוונה שלנו, הם מצאו את המשימות כברורות וקלות להבנה והצליחו לבצע אותן ללא קושי. תגובות המשתמשים: קיבלנו תגובות חיוביות מהמשתמשים לגבי המערכת. כולם הבינו את החשיבות הרבה שיש למערכת מסוג זה בייחוד בתקופה זו. נרשמה הפתעה מאיכות מהטקסט המיוצר ומשתמשים אף טענו כי היו משתמשים במערכת זו אם היא הייתה פתוחה לשימוש הכלל.

מסקנות שעלו : ראינו כי אם תגובת המערכת לא הופיעה תוך מספר שניות בודדות, חלק מהמשתמשים חשבו כי לחיצת הכפתור לא עבדה ולחצו שוב. הסיבה שלעיתים תהליך זה לוקח כמה שניות הוא מכיוון שהקריאה ל API של GPT לוקחת כמה שניות. על מנת לפתור זאת, היינו יכולות להוסיף למשתמש הערה כמו "המערכת מייצרת את הפוסט שלך ברגעים אלו". בנוסף, חלק מהמשתמשים ציינו כי היו שמחים אם היה ניתן לבקש מהמערכת לשנות רק חלקים ספציפיים בטקסט הנתון. לסיכום: הערכה על ידי משתמשים הוכיחה שהמערכת שלנו שימושית ומועילה. קיבלנו משוב חיובי מהמשתמשים וזיהינו מספר תחומים שבהם ניתן לשפר את המערכת.

סיכום -

מה למדנו מהפרויקט:

בראש ובראשונה, גילינו שיש צורך וביקוש רב במערכת שיכולה ליצור באופן אוטומטי טקסטים איכותיים בנושאים רגישים כמו הסכסוך הישראלי-פלסטיני. מערכת כזו מסייעת לכל מי שרוצה, בלי קשר ברמת הידע שלו, ליצור פוסט פרו-ישראלי אשר מכבד את האחר ומעביר את המסר הרצוי. המערכת עוזרת בהעלאת המודעות על הסכסוך הישראלי-פלסטיני מהצד הישראלי ומעודדת דיאלוג מכבד מבוסס נתונים. במהלך בניית המערכת נתקלנו באתגרים רבים מול המודל, אשר ההתמודדות איתם לימדה אותנו דברים חדשים למשל: כיצד לגרום למודל לדבר על נושאים רגישים ואיך לגרום לטקסטים שהוא מייצר להיות מגוונים יותר. בנוסף, זו הפעם הראשונה בה התנסנו ביצירת ממשק משתמש אינטרקטיבי ולמדנו על דרכים שונות כיצד לעשות זאת.

כיצד היינו מרחיבים את המערכת לו היה לנו יותר זמן:

שיפור איכות הטקסט המיוצר: היינו עובדים על שיפור איכות הטקסט המיוצר על ידי המערכת, תוך שימוש בטכניקות מתקדמות יותר של עיבוד שפה טבעית. הרחבת הפונקציונליות של המערכת: היינו מוסיפים פונקציונליות חדשה למערכת, כגון האפשרות לתרגם טקסטים לשפות שונות, לכתוב סוגים שונים של תוכן יצירתי (כמו שירים או סיפורים), ולנתח טקסטים קיימים. בנוסף אלמנט של זיהוי "פייק ניוז" והפרכתן יכול להיות מעשיר עבור המערכת. הפיכת המערכת לנגישה יותר: היינו הופכות את המערכת לנגישה יותר למשתמשים על ידי פיתוח UI מתקדם יותר בפורמט תוסף לדפדפן, על מנת לאפשר למשתמש להשתמש במערכת שלנו ככלי עזר בעת הגלישה ברשתות החברתיות ולשלב אותה בשימוש הרגיל שלו וליצור תוצרים באופן סימולטני. אנו מאמינות שמערכת כזו יכולה להיות כלי רב עוצמה לקידום הבנה, תקשורת ושיתוף פעולה בנושא הסכסוך הישראלי פלסטיני.