



# TherAI

**Therapist Personal Assistant**

שיראל אוסקר, מתנאל שחמורוב



## מטרת הפרויקט

יצרנו את הפרויקט במטרה לעזור לפסיכולוגים עם עבודת היום-יום השגרתית שלהם ולעזור להם לייעל ולקצר את זמני העבודה של הפסיכולוגים, ובכך לאפשר להם לקבל מטופלים נוספים או לפנות זמן חשוב לעצמם להתאווור אחרי יום עבודה ארוך.



## למה בחרנו בנושא?

הגענו לנושא לאחר שיחה עם פסיכולוגית שאמרה שחלק מההליך הטיפולי הוא הקלטת והסרטת המפגשים, שלאחריהם יש צורך ליצור תמליל של הנאמר בפגישה, כלומר, להאזין מחדש לכל הפגישה ולכתוב את הנאמר במילה, וזה תהליך ארוך שראינו שנוכל לקצר בעזרת היכולות שקיבלנו בקורס.



# תכנון הביצוע

בנייה של קטע קוד ראשי שיריץ קטעי קוד משניים שכל אחד מהם יהיה אמון על חלק אחר במעבר מהקלט (סרטון של פגישה פסיכולוגית) לפלט (תמלול שיחה וקטעי שמע מופרדים על פי הדוברים - מטפל ומטופל).

- קטע קוד האחראי על המרת הסרטון לשמע מסוג WAV אותו נוכל לנתח ולעבוד איתו.

- קטע קוד המפריד את השמע לדוברים על בסיס מנוע זיהוי דוברים ומחזיר פלט דמוי list, בו כל איבר ברשימה הוא שלשה של משתנים - זמן התחלה של הדיבור, זמן סיום של הדיבור, מי הדובר (מטפל ומטופל).

```
6.7 7.2 SPEAKER_01
7.2 7.2 SPEAKER_02
7.6 8.3 SPEAKER_01
8.3 9.9 SPEAKER_02
9.9 10.9 SPEAKER_01
```

דוגמת פלט מספריית pyannote



## תכנון הביצוע

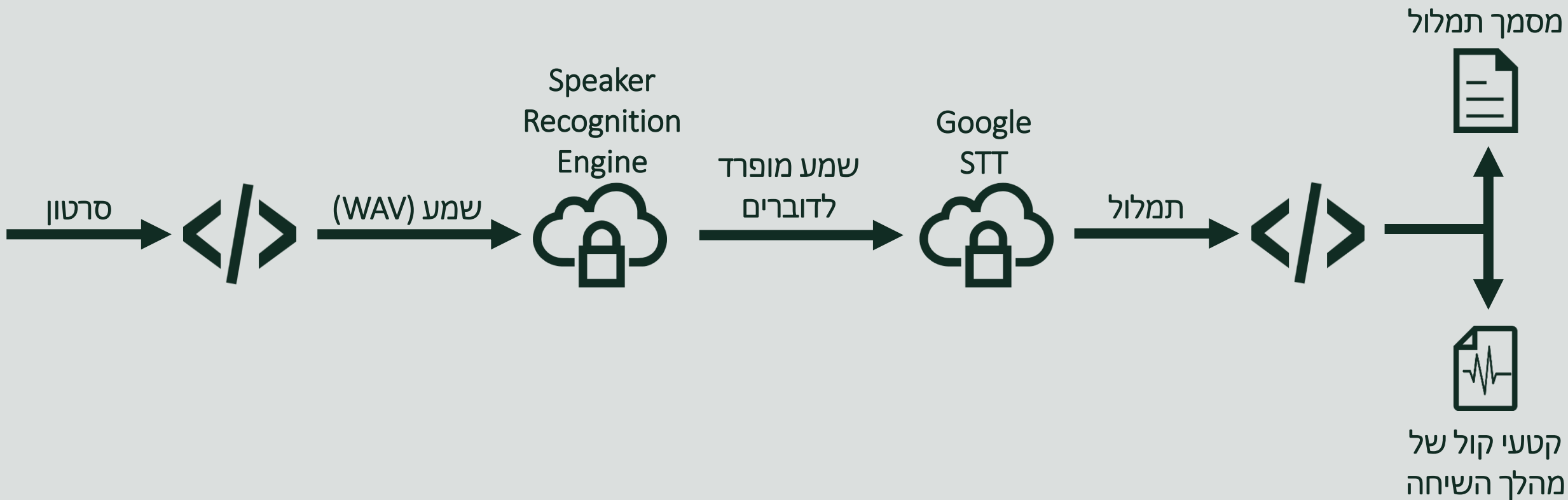
- קטע קוד העובר על הערכים שקיבלנו חזרה מהפרדת הדוברים, ומבצע חיתוך ותמלול של קטעי הדיבור ומדפיס אותם לתוך מסמך על פי פורמט מוגדר מראש - הדובר של המשפט (מטפל/מטופל), חתימת זמן (התחלה וסיום) והתמלול של המשפט.

דוגמא לפלט הרצוי:

מטפל [0:17-0:23] - שלום, איך הרגשת מאז הפגישה האחרונה שלנו?  
מטופל [0:26 - 0:30] - בסדר, אני יודעת, לא היה משהו מיוחד



# ארכיטקטורת הפתרון (Flow)






## ביצוע בפועל

- לאור חיסיון רפואי, השתמשנו בקטעי שמע של סדרה הנקראת "בטיפול" המראה במהלך פרקיה שחקנים המציגים פגישות של מטפל עם המטופלים שלו ולא בפגישות אמיתיות של פסיכולוג עם מטופלים.
- את הפרק של הסדרה הכנסנו לקטע קוד שהוציא לנו שמע מסוג WAV.
- ניסינו למצוא מנועי הפרדת דוברים, אך ללא הצלחה למצוא מנוע שעובד, ולכן חילקנו את השמע לקטעי קול שונים בהתבסס על קטעי שקט בעזרת קטע קוד.
- את קטעי הקול שמענו וקיטלגנו בצורה ידנית למטפל, מטופל ושמע מעורב (של שני הדוברים) והכנסנו אותם לרשימות נפרדות בקוד שלנו.
- בנינו קטע קוד העובר על הרשימות ושולח כל קטע קול לתוך מנוע התמלול של גוגל (Google STT) והדפסנו את המלל לתוך קובץ txt.



# דוגמת פלט (ראשוני)

session\_transcript - פנקס רשימות 

קובץ עריכה עיצוב תצוגה עזרה

1 מטופל [15,0.0]: אמרתי לך

מטפל [16,3.9]: אמרתי מה

מטופל [17,7.17]: אתה יודע שאתה מעודדות לבכות בטיפול

מטפל [18,11.18]: שלום סלע

מטופל [19,14.99]: וחשבתי שבפעם הראשונה שאני אזכה זה לא יודע ככה

מטופל [21,25.39]: את יודעת קצת מאכזב

מטפל [22,29.81]: חמישי זה יותר משחרר את הכסף



# תקלות, אתגרים ופתרונות

## הפרדת דוברים

- הבעיה - ללא הצלחה למצוא מנוע הפרדה שעובד
- הפתרון הנוכחי - הפרדת השמע לפי קטעי שקט ותיוג ידני של הקטעים
- הפתרון הרצוי - מציאת מנוע הפרדה תקין
- הבעיה (תיאורטי) - גם לאחר הפרדה, אין איך לדעת מי מהדוברים זה המטופל ומי המטפל
- הפתרון הנוכחי - השמעת 2 קטעי קול למשתמש בהם הוא מתייג את המטופל והמטפל
- הפתרון הרצוי - מודל זיהוי קולי של המטפל (ועל ידי כך ביצוע הפרדה)





# תקלות, אתגרים ופתרונות

## תמלול

- הבעיה - מבדיקה ראשונית, המנוע מתמלל מילים לא נכונות מידיי פעם
- הפתרון הנוכחי - הוספת קטעי זמן לתמלול כך שיהיה ניתן לחזור למשפט הנאמר ולתקן ידנית את המלל
- הפתרון הרצוי - ביצוע הערכת דיוק של המודל ובחינה של צעדים

## פלט

- האתגר - הפלט כרגע בקובץ txt
- יכולת נדרשת - מעבר לקובץ word עם python-docx

%

## הערכת דיוק של המודל

לאחר קבלת קובץ הפלט מהקטע הקוד שלנו, עשינו הערכת דיוק למודל על ידי תמלול ידני של המשפטים שנאמרו, השוואה ומתן ניקוד לכל משפט בהתאם לכמות המילים שתוממלו בו בצורה נכונה בצורה הבאה:

"אתה רואה אמרתי לך אמרת לי מה - 4 מתוך 7".

לאחר מעבר על כל המשפטים, ביצענו סכימה והערכה של דיוק המנוע:

מילים סה"כ - 2155

מילים שתומללו נכון - 1398

אחוזי הדיוק - 65%

תוצאה זו נמוכה משציפינו ואינה תוצאה מספקת, ואף תוצאה מפתיעה, לאור שימוש במנוע של אחת החברות החזקות בשוק.

%

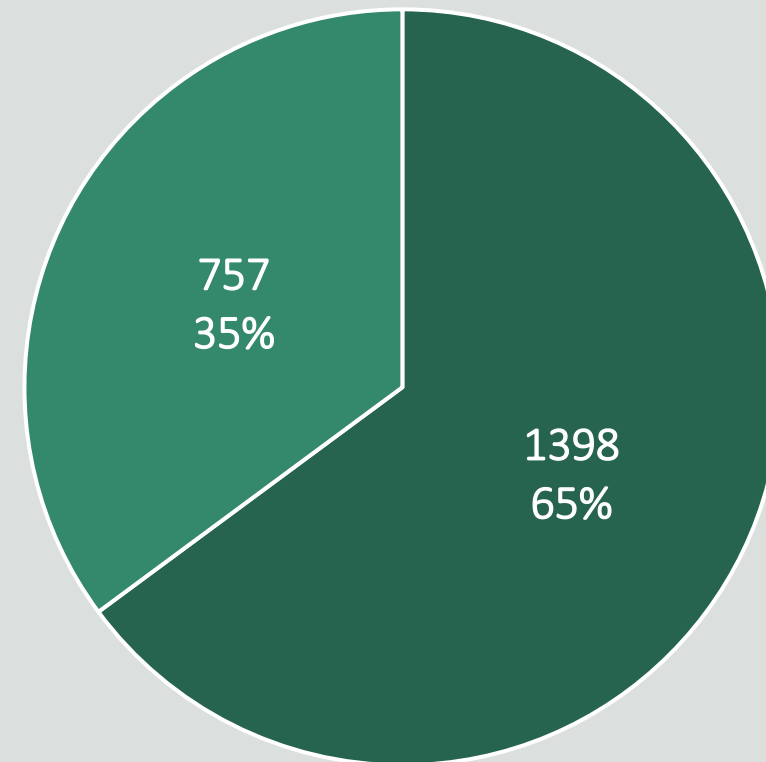
# הערכת דיוק של המודל

בצורה גרפית:

דוגמאות תמלול:

ניקוד	משפט בתמלול	משפט מתומלל
10/10	גם לי יש כזה כאילו ואני לא יודעת את זה	גם לי יש כזה כאילו ואני לא יודעת את זה
2/5	דמיינתי שאתה לא מסוגל לשמוע	מי שלא מסוגל לשמוע
0/3	את רוצה תה	התפוצצתי

מילים מתומללות



מילים לא נכונות    מילים נכונות



# ניסיונות שיפור של המודל

- לאור אחוזי הדיוק הנמוכים של המודל, ניסינו לשפר אותו במספר דרכים:
  - הפרדת דוברים - ניסינו למצוא מספר מנועי הפרדת דוברים, חלקם בהמלצת סגל הקורס, אך ללא מציאת מנוע שצלחנו בהפעלתו.
  - שיפור בתמלול - ניסיון לחלק את השמע בצורה ידנית לקטעים ארוכים יותר לאור העובדה שחלק מהמשפטים הארוכים שנאמרו תומללו נכון במלואם - אך עדיין ללא הצלחה, לאור העובדה שהמנוע מפסיק את פעולתו לאחר זיהוי שקט, שזו הגדרה שניסינו לשנות ללא הצלחה, ולאור זאת תמלול של קטעים ארוכים היה משמעותית גרוע יותר (תמלול של 5 שניות בלבד עד הגעה להפסקה בדיבור מתוך שמע של כמעט דקה שלמה).



## סיכום ומסקנות

הפרויקט נוצר מתוך צורך שזיהינו אצל אוכלוסייה מסוימת - אוכלוסיית המטפלים. המודל שקיבלנו הביא לנו תוצאות נמוכות מהצפוי ושאינן עונות על הצורך הנדרש, אך להערכתנו, מדובר בבעיה של כמות הזמן וכוח האדם של הפרויקט, שכן הצורך עדיין ישנו, ומוצר משופר יוכל לתת מענה טוב ומספק יותר לצורך הזה. כמה שיפורים שניתן לעשות:

- מניפולציות שונות על השמע (ניקוי, חידוד, חקירת פרמטרי המרה שונים מווידאו לאודיו).
- חקירה מעמיקה יותר למשך זמן ארוך יותר של מנועי הפרדת דוברים ו/או פיתוח מודל מכונה עצמאי המבדיל בין הדוברים.
- חקירה מעמיקה יותר של מנוע Google STT להבנה מעמיקה יותר של השינויים שניתן לבצע במודלים על מנת לשפר את דיוק התמלול שלהם.
- הכנסת המלל המתומלל לתוך קובץ מסוג אחר (שיפור קוסמטי לפלט).