

# AI lecture summaries

## מנחה פרופסור רועי פורן

### מגיש: יניב גלבר

מטרת הפרוייקט הייתה להראות יכולת של סיכום הרצאה בעזרת שימוש ב AI.

הפרוייקט חולק למספר חלקים:

1. הפקת תמלול
2. פיתחנו מנגנון המאפשר סנכרון מדויק בין התוכן המדובר בהרצאה לבין השקף שהוצג באותו רגע.
3. יצירת הסיכומים ויצירת ההרצאה הכוללת את הסיכום

1. ניסיתי מספר שיטות שונות להפקת התמלול להרצאות. תחילה היה ניסיון להשתמש במספר

מודלים אותם רועי הכיוון אותי אליהם ולאחר מכן למספר ניסיון להשתמש במודלים שקראתי עליהם באינטרנט. היו שתי בעיות מרכזיות כאשר ניסיתי להריץ מודלים על הרצאות. הבעיה המרכזית בשיטה זאת הייתה שהרצאה של מודל תמלול על הרצאות שזמנם הוא מעל לשעתיים לוקח כוח חישוב משמעותי אליו אין לי גישה. בעיה נוספת שהייתה שאפילו בניסיונות על זמן מוגבל של ההרצאות (היה ניסיון על זמן של 2 ו 5 דקות בנקודות שונות הרצאות) היה חוסר דיוק חריף במיוחד בתמלול ההרצאה, הסיבה לכך היא שרוב המודלים הטובים מאומנים באנגלית ועובדים רק על עליה ומודלים העובדים עבור שפות רבות אינן מדויקות כלל בעברית ונותנות ביצועים לא איכותיים.

בשלב זה חשבתי על כך שליוטיוב יש מודל מתקדם לתמלול ותרגום סרטונים, הפועל בשפות רבות וגם בעברית. את הפלט של תמלול זה אנשים לרוב פוגשים כאשר לוחצים בסרטונים על תרגום אוטומטי. היתרונות בשימוש במודל של יוטיוב הם שהמודל של יוטיוב מאומן באופן איכותי יחסית לשאר המודלים על קלט בעברית. בנוסף הרצאה של המודל על הסרטון קורת אצל השרתים של גוגל וניתן לקבל את הפלט שלהן בעזרת גישה לסטרים. יש לשים לב שגישה לסטרים של כתוביות שונה מגישה לסטרים של הוידאו. ביוטיוב הסרטון מחולק לכמה לסטרימים שונים וכל אחד מרכיב את הסרטון אותו אנו מקבלים. שם הקובץ עליו עבדתי הוא הפקת\_תמלולים\_לסרטונים. את הכתוביות שמרתי בקבצי iw.srt. לאחר הצגת התוצאות לרועי בשימוש במודל של יוטיוב לעומת המודלים האחרים קיבלתי אישור שהביצועים אכן איכותיים מספיק. יש לשים לב כי למרות שהמודל של יוטיוב משיג תוצאות טובות בהרבה מכל מודל אחר, התוצאות שלו בעברית עדיין אינן מדויקות באופן מוחלט ומכילות טעויות.

2. לחלק זה הייתי צריך להכין תמונות של העמודים במצגת, זאת מכיוון שלא ניתן להפעיל

השוות בין פריימים לבין העמודים כאשר הם אינם תמונות. לאחר מכן יצרתי מילון של אינטרוולים של זמני העמודים בהרצאה. על מנת לעשות זאת כל עשר שניות בדקתי האם התחלף עמוד, ואם כן לאיזה עמוד. את זה עשיתי בעזרת השוואה בין הפריימים שמופיע באותו רגע לבין עמודים שהיו קרובים לעמוד שזוהה בעשר השניות הקודמות, עשיתי זאת בדרך זו על מנת לחסוך בזמן עיבוד והרצה וכן ברור שהנחה זאת מתקיימת ברוב המוחלט של המקרים. את התוצאה שמרתי כח json בשם slide\_intervals.json.

לאחר מכן עשיתי פענוח של קובץ הכתוביות והפכתי אותו לרשימה של מילונים המכילים:

זמן התחלה, זמן סיום ומה נאמר באותו הזמן.

לבסוף עשיתי התאמה בין האינטרוולים של התמלולים לבין האינטרוולים של העמודים.

על מנת להיות שמרן ולאסוף כמה שיותר מידע על מה שנאמר בזמן כל מצגת, ההתאמות

בוצעו אם ישנה חפיפה בין אינטרוולים ולא רק הכלה מלאה. הרעיון של זה הוא שאם נאמר משהו קצת לפני הצגת עמוד אז יש קשר ישיר לאותו העמוד וכן לאסוף מידע זה יעזור

הסיכומים.

את התוצאות שמרתי בקובץ presentations\_words.json תחת התיקייה words\_to\_presentationsNew .

3. תחילה ניסיתי לייצר את הסיכומים עם BART, אך התוצאות לא היו טובות. הסיבה לכך היא שהמודל עובד פחות טוב בעברית ובנוסף מצריך טקסט נקי. לאחר מכן ניסיתי להשתמש במודלים דומים אך גם בהם לא הושגו סיכומים איכותיים, היו מקרים רבים בהם קיבלתי ג'יבריש. לאור העובדה כי המודלים הפועלים על עברית חייבים טקסט נקי לחלוטין, אותו לא ניתן להשיג, עקב מספר סיבות:

1 איכות התמלול איננה מדויקת לחלוטין

2 רעשי רקע שונים

3 שיח שאינו קשור ישירות לנדון באותו שלב, כגון שאלות הבהרה לגבי עמודים קודמים וכו

על מנת לפתור בעיה זו הסקתי כי ישנו מידע רב בעמודים הויזואלים עצמם ועל יד שימוש בהם ניתן לשפר רבות את הביצועים.

לכן לאחר חיפושים החלטתי להשתמש במודל gemini-1.5-flash של גוגל, המסוגל לקבל גם טקסט וגם תמונה כקלט.

ביקשתי ממנו לסכם לי את העמודים על פי מה שמשתקף מההרצאות וכן מהתמלול. קושי נוסף התגלה בחשיבות של פרומט מדויק, הדבר הקשה והאט על הליך מציאת הפורמט המדויק ביותר הוא מגבלת מספר השאלות בגרסא החינמית. על מנת להתגבר על מגבלה זו פתחתי מספר מיילים שונים.

את זה עשיתי בקובץ הכנת\_סיכומים .

לאחר מכן בקובץ יצירת\_הרצאות\_עם\_סיכומים יצרתי מצגת עם הסיכומים המתאימים לכל עמוד.