

## אייה פוד'

## רימא חנה

### 4. עיבוד שפות עבודה מ'ו

#### חלק א:

הרכינו את הקוד הבא בקובץ נפרד, בשילול שומר את המידע פעמי' אחד, ובמקום מסויים בשביל להשתמש בתונונים אחר כך.

```

load_movie_dataset.py × gpt-generation_finetuning.py bert_classification.py
1 from datasets import load_dataset
2 dataset = load_dataset("imdb")
3 subset = dataset["train"].shuffle(seed=42).select(range(500))
4 subset.save_to_disk("imdb_subset")

```

ambilush ללחנו מאגר נתונים בשם "imdb" וקחנו ממtan 500 דוגמאות מהdataset של האימון באופן רנדומאל, ושמרנו אותן בקובץ בשם "imdb\_subset".

#### חלק ב:

עשינו מה מבוקש, בחירתנו עבור ארגומנטוי האימון הייתה כז:

```

training_args = TrainingArguments(
    output_dir=".//results",
    eval_strategy="no",
    learning_rate=2e-5,
    per_device_train_batch_size=5,
    per_device_eval_batch_size=5,
    num_train_epochs=3,
    weight_decay=0.01
)

```

#### **Num\_train\_epochs=3:**

מספר הדגימות בסט האימון הוא 400, ולכן עבור אחד בלבד על הנתונים (אפקט אחד) אין מספיק כדי ללמידה דפואים מורכבים, כמו אלו הנדרשים לשימוש הסיווג שאנו מבצעים. מצד שני, בחירה במספר גדול מדי של אפקטים עלולה לגרום לביאה של התאמת יתר (overfitting) שבה המודל מתאים את עצמו יותר מדי לנתחי האימון ומאנד את יכולת הכללה שלן.

לכן, בחרנו בערך ביניים, 3, שמאפשר למודל ללמידה דפואים משמעותיים תוך שמירה על יכולת הכללה על נתונים חדשים.

#### **eval\_strategy="no":**

כי לא רצים לשמור את המודול.

#### **Learning\_rate=2e-5:**

בחרנו קצב למידה זה כי הדטה שלנו יחסית קטן ומורכב, ומכוון שמדובר בתהילר Fine-Tuning של מודל מאומן מראש (BERT), קצב למידה בין כ-5e-5 מאפשר התאמת עדינה של המשקלים לדטה מבלי לגרום לשינויים חדים מדי או לאובדן הידע הקודם של המודול.

#### **weight\_decay=0.01:**

בחרנו ב-`weight_decay=0.01` כדי למנוע overfitting במהלך האימון. ערך זה מוסיף גורליזציה שופחתה את השפעתם של משקלים גדולים מדי במודול, מה שתורם לשיפור הכללה של המודול על דatasets חדשות. הערך 0.01 מתאים כיון שהוא מספק איזון טוב בין מניעת overfitting לבין שמירה על התאמת לדטה שלנו.

#### **per\_device\_train\_batch\_size=5=per\_device\_eval\_batch\_size:**

הערך 5 נבחר מטור איזון בין יעילות חישובית לבין מגבלות הזיכרון של המערכת, תוך שמירה על איכות האימון וההערכה.

עבור קבוצת הבדיקה. קיבלנו תוצאה accuracy הבהא:

```
'eval_loss': 0.3590928912162781, 'eval_accuracy': 0.86, 'eval_runtime': 67.2624, 'eval_samples_per_second': 1.487, 'epoch': 1.0};  
'eval_loss': 0.2870155572812354, 'eval_accuracy': 0.87, 'eval_runtime': 74.565, 'eval_samples_per_second': 1.341, 'eval_steps_per_second': 0.268, 'epoch': 2.0};  
'eval_loss': 0.2478859716259003, 'eval_accuracy': 0.9, 'eval_runtime': 73.8156, 'eval_samples_per_second': 1.355, 'eval_steps_per_second': 0.271, 'epoch': 3.0};  
'train_runtime': 3599.151, 'train_samples_per_second': 0.333, 'train_steps_per_second': 0.067, 'train_loss': 0.3306877772013463, 'epoch': 3.0}  
100% | 240/240 [59:59<00:00, 15.00s/it]  
100% | 26/26 [01:02<00:00, 3.14s/it]  
Test Accuracy: 0.90
```

## שאלות:

1-הגענו לדרגת דיוק של 0.9, שהיא יחסית ארך לא מושלמת, ונציג את הסיבות לכך בהמשך. סוג הסיווג בדתאטה של מटבוס על הטעון בטקסט: אם הטון חיובי, הביקורת מסוגגת כחויבית, אחרת – שלילית. המודל BERT מצטיין במלדיות הקשרים וסמנטייקה מורכבים, וכך שהוא אומן מראש על כמויות עצומות של טקסטים מגוונים. באמצעות Fine-Tuning על הדאטאה שננו ובחרית פרמטרים מתאימים, הצליחנו להתאים את המודול לדאטה חדשה מבלי לגרום ל overfitting.

עם זאת, סיבה אפשרית לכך שלא הגענו לדיקוגרף יותר היא מכות הדעתה המוצמצמת שהיא השתמשה. מכות גודלה יותר עשויה לאפשר למודול להבין הקשרים וסמנטיקה מרכיבים יותר. מצד שני, אם היינו לומדים על אותה DATA יותר מ-3 פעמים, היהת סכנתה ל- overfitting, ולכן איזון זה היה קריטי. דרגת דיקוגרף זו היא התוצאה המיטבית שהצלחנו להשיג תוך שמירה על הכללה ואיזון בין אימון לבין התאמת יתרה.

2- אם נשתמש במודול שאמנו על ביקורת סריטים כדי לסייע בבדיקה סריטים, סביר להניח שלא נגיע לדיווק גבוה יותר, אך יתכן שנשיג דיווק סביר. הסביר לכך היא שבביקורת סריטים וספרים גנותו להיות דומות מבחינה הטוונ: בבדיקה חיובית תתבסא בטון חיובי, ושלילת בטון שלילי. עם זאת, יתכן שיש הבדלים סמנטיים והקשרים "חודיים" בין ביקורת סריטים לביקורת סריטים.

כיוון שהמודל הותאם לדאטה של ביקורות סרטים ולא נחשף למאפיינים הייחודיים של ביקורות ספרים, יש סבירות גבוהה שאנו יתקשה בזיהוי נכון של ביקורות ספרים. לסייע, איצות התוצאות תלויה במידת הדמיון בין ביקורות הספרים לביקורות הסרטים שבחוק השתרעמנו במהלך האימון.

3- אם נשתמש במודול שiamiינו על ביקורות סרטים ממאגר אחד כדי לסייע בקיימות סרטים ממאגר אחר, התוצאות יהיו אוצר תלויות במידה הדמיין בין המאגרים. ככל מאגר עשוי להיות תוכנות ייחודיות, כמו סגן כתבה, מבנה משפטיים, או אוצר מיללים, המשפיקות על דיאלוג הטוון.

אם המאגרים דומים במאפייניהם שמבידלים בין טוב וחובי לשלייל, "תacen שהמודול יצלה לשומר על דיקון טוב ואיך לגאות גובה יותר. לעומת זאת, אם קיימים הבדלים משמעותיים בין המאגרים, "תacen שהמודול יתקשה לבצע החלטה כוננה, והתוצאות עשויות להיות פחות מדיוקן".

לכן, איקות הביצועים תלויה בעיקר בדמיון בין שני המאגרים בתוכנות המרכזיות הקשורות לשינוי.

4- כדי לשפר את התוצאות עבור כל אחד מהמרקדים, ניתן לבצע **Fine-tuning** נוספת על המודל שיצרנו, תוך שימוש בدادה מהמאגרים החדשניים אוטם אנו זכרים לוגו.

ב-Fine-tuning, המודל למד להבדיל בין טוונים שונים על בסיס הדעתה המקורית. ב-Fine-tuning המודל ילמד את המאפיינים הייחודיים של המאגר החדש (כגון אוצר מילים או סגנון כתיבה "יהודי"), וכן יתראים את עצמו למאפיינים אלו. פועלה זו מאפשרת למודל להבין טוב יותר את הדעתה החדש ולהציג לרמת דיווק גבוהה יותר.

חלק ג:

- עברו כל אחד מהמודלים טענו מודל-2 GPT- ו- tokenizer
  - ללחינו את הדעתה מהקובץ שקיבלו את שמו.
  - חיליקנו את הדעתה לפ' עמודות ה- label כאשר ה- dataset\_label=0 שמרנו ב- \_0\_label ו- dataset\_label=1 שמרנו ב- \_1\_label ולחינו מכל אחת 100 דגימות.
  - עברו כל אחד משני המאגרים עשוilo tokenize בערצת הפונקציה tokenize\_reviews עם הטוקיניזר המתאים לה.
  - עברו כל אחד משני המאגרים חישבנו את אורך הטקסט הממוצע ושמרנו ב- \_0\_average\_length\_label ו- \_1\_average\_length\_label.
  - עברו כל אחד משני המאגרים בחרנו ערבי פורטרי האימון אלה:

```

training_args_label_0 = TrainingArguments(
    output_dir='./results',
    eval_strategy="no",
    learning_rate=2e-5,
    per_device_train_batch_size=5,
    per_device_eval_batch_size=5,
    num_train_epochs=3,
    weight_decay=0.01
)

```

- כasher כל הפרמטרים שווים לבחירה השלב בlyn זה אותו הסבר חוץ מ- "och" eval, כי בסוף האימון אנחנו שומרים את המודול ידנית.
- עברו על מודל יצרנו trainer עם הدادטה והטוקניזר והפרמטרים המתאימים לכל אחד.
  - אימנו שת שני המודלים ושמरנו אותם והטוקנים שלהם בתיקייה שהביבלו אותה בקלט, המודול הראשון בשם תיקייה model\_label\_0 עם הטוקניזר שלו והשנו בשם model\_label\_1 עמו הטוקניזר שלו.
  - טענו את המודלים והטוקנים שלהםשוב, עברו כל אחד עשוינו יצרת 5 טקסטים. נסביר על כל אחד בנפרד.

מעבר המודול של ביקורות שליליות:

```

prompt = "The movie was"
input_ids = tokenizer_label_0.encode(prompt, return_tensors="pt")
attention_mask = input_ids.ne(tokenizer_label_0.pad_token_id)

with torch.no_grad():
    output_label_0 = model_label_0.generate(
        input_ids=input_ids,
        attention_mask=attention_mask,
        max_length=average_length_label_0,
        temperature=1.2,
        top_k=15,
        top_p=0.6,
        repetition_penalty=1.2,
        do_sample=True,
        num_return_sequences=5,
    )

```

"יצרנו 5 טקסטים לפי המודול שאימנו על ביקורות שליליות כאשר מגדירים התחללה שלו שהוא "The move was" עם הטוקן המתאים למודל זה, הגדרנו שאורך מקסימלי יהיה המוצע של אורכי הטקסטים במודל זה כי יצירת הטקסט היא לפני פרטיהם שמאפיניהם טקסטים אלה ואורך הטקסט, כאשר המידע שנכלל בתוך הטקסט מיוחס לאוכן, لكن היה חשוב שיהיה התאמת.

בגלל שהtekst שנוצר מהתאר דעה מסוימת, ציפינו שהוא יוכל מגוון של מילים נרדפות ולא יחזור על אותן מילים שוב ושוב. וגם שלא יחזור על אותן משפטים. לשם כך, הגדרנו repetition\_penalty=1.2 כך שהמודול יקלל "עוונש" על שימוש חוזר באותו מילים. מצד שני, רצינו לשמר על עקבות רעונות באותו הטקסט ולכן לא הגדלנו את הערך יותר מכך. הבחירה זו איזנה בין גיון מילולי ושמירה על רעיון עיקרי לאורך הטקסט.

מכיוון שרוצים שהמודול ייצור טקסטים מגוונים ושיירים בכל פעם, בחרנו ב - **temperature=1.2** ערך זה מעודד את המודול לבחור באפשרויות פחות צפויות מתוך ההפצה שהוא מייצר, וכך מתקבלות תוצאות יצירתיות ומגוונות יותר. עם זאת, שמרנו על ערך מוגן ולא גבוי מדי כדי להימנע ממקרים בהם התוצאה תהיה אקרואית ולא עקבית מבחינה התקשרות. הבחירה ב- 1.2 איזנה בין גיון מילולי ושמירה על הרלוונטיות של הטקסט לנושא המרכזי.

**Top\_k=15**: כאשר המודול בוחר את המילה הבאה, הוא מתמקד ב-15 המילים בעלות ההסתברות הגבוהה ביותר. לאחר ניתוח הדאטא, זיהינו שככל טקסט יש שני חלקים עיקריים: מידע אינפורטטיבי על הסרט והבעת דעה. המטרה הייתה לשמר על מבנה זה ולהציג גיון בחירת המילים. כאשר הקטנו את k\_top\_k קיבלנו טקסטים שה坦קדו רק במידע האינפורטטיבי, ללא ביטוי של דעת. מצד שני, כאשר מדובר בביטויים שליליים, לרובם كانوا טון שלילי ומלים שמדגימות תכונות שליליות. לכן, בחירה ב- **Top\_k=15** איזנה בין שני ה择rets – שמירה על גיון אינפורטטיבי ובחירה מגוונת במילים שמייצגות דעת שלילתית.

**Top\_p=0.6** : פרמטר זה קובע את סף ההסתברויות המctrber לבחירת מילים. כפי שהסביר קודם, טקסטים של ביקורות שליליות לרוב מתחאים תכונות שליליות. בחירה בערך 0.6 הייתה מספקת כדי לתפום את הטון השלילי של הביקורת, ובזמן איזנה בין שמירה על המידע האינפורטטיבי המרכזי לבין גיון בחירת המילים. ערך זה הבטיח שהמודול יבחר מילים רלוונטיות ולא יחזור על אותן מילים שוב ושוב, תוך שמירה על קוהרנטיות ומשמעות הטקסט.

מעבר המודול של ביקורות חיובית:

```

input_ids = tokenizer_label_1.encode(prompt, return_tensors="pt")
attention_mask1 = input_ids.ne(tokenizer_label_1.pad_token_id)

with torch.no_grad():
    output_label_1 = model_label_1.generate(
        input_ids,
        attention_mask=attention_mask1,
        max_length=average_length_label_1,
        temperature=1.05,
        top_k=40,
        top_p=0.8,
        repetition_penalty=1.1,
        do_sample=True,
        num_return_sequences=5,
    )

```

בביקורתות שליליות הטון הכללי היה שלילי וברובו התרכז בתוכנות שליליות, בעוד שבביקורתות חייבות היה גיון רב יותר, שככל גם תיאורים חייביים וגם שליליים עם נטייה לחיבוב, למשל פין הנאה מהסרט. אך, במודל החיביב בחרנו להגדיל את ערכי **k\_top** כדי לאפשר גיון רחב יותר בבחירה המילימית, שיתאים לאופי המגנון של הביקורות החביבות, אך מבלי להגדיל יותר מדי ולפגוע ביכולת לשמר על הטון המרכזי והיחידיות של הביקורות החביבות. כך נשמר איזון בין גיון לבן Kohortnetiyot ומשמעות הטקסט.

בהתאם להגדלת ערכי **k\_top** ו-**k\_top** שהפחיתה את הסיכוי לחזרות בטקסט, הקטנו את ערך **repetition\_penalty** 1.1-1.05 כדי לאפשר חופש יצירתי מסוים במודל מוביל ליצור כפלויות. בנוסף, מכיוון שהגיאון בטקסט גדול, הקטנו את ערך **temperature** ל-1.05 כדי לשמר על עקבות בין הנושאים בטקסט ולהבטיח שהטון והמסרים יהיו מוקדים וברורים. כך יצרנו איזון בין גיון לבן שמירה על משמעות הטקסט והטון הרצוי.

#### שאלות:

1-התוצאות היו לא ברורות ויש לפעמים חוסר עקבות כאשר עוברים מנסה לנושא אחר לא קשור, אבל ניתן לzechot הטון הכללי בטקסט. ניסינו להתחאים הפרמטרים ככל שניתן בהתאם לדאטה שלם אבל קיבלו תוצאות לא טובות מספיה.

2- במודל של הביקורות השליליות, הטון היה ברור מאוד ומהווה כשליל, כשהתאיורים כללו בעיקר תוכנות שליליות של הסרט, כמו עילילה מעסמת או משחק גראן. לעומת זאת, במודל של הביקורות החביבות, חלק מההשפעות היו לא חד משמעיים, וורה קשה להבחין אם מדובר בטון חיובי או שליל. עם זאת, היה ברור שנטילת הטקסט ריה חיובית, במיוודת אשר הזכרו ביטויים כמו "הסרט היה טוב", למורות שהתוכנות שהזכו היו שליליות. זה מצביע על כך שבביקורתות חייבות כוללות לא רק תיאורים חייביים אלא גם תיאורים שליליים, כשבסתו של דבר הדעה הכללית היא חיובית.

3- מהתוצאות שקיבלו נוכל להסיק שהדאטה לא היה מספק גודל בשבייל לתפקיד ההקשרים המסתובבים לכל מודל, لكن אם נאמן אותו על הרבה יותר דגימות, נצטרך להקטין מספר האיפוקים כי עם מספק דатаה אין צורך לדגימות, וכך המודל יתאפשר יותר פרטיטים ועובד יותר טוב מצד אחד, אבל מצד שני זמן האימון יהיה ארוך כי הטקסטים גדולים ומורכבים. ואם נאמן המודל עם פחות דגימות זמן האימון יהיה יותר קטן אבל המודל לא יתאפשר יותר טוב בפרטיטים שכן נדרש להגדיל מספר האיפוקים כך שלמד על אותו דטהה יותר מפעם אחת, אבל אז יהיה סיכוי יותר גודל ל- overfitting.

4- ה- **attention\_mask** יוצר מסכה שמאגדירה אילו חלקיים של הקלט חשובים למודל. הערך 1 מציין שהטוקן רלוונטי, בעוד שערך 0 מציין שהטוקן הוא מיולי (Padding) או לא חשוב. מטרתו להבטיח שהמודל יתמקד רק בטוקנים החשובים ויתעלם מוחלים ריקים או לא רלוונטיים בכלל, כך שהוא יוכל לבצע חישוב מוגזק ויעיל יותר.

5- השימוש ב- **prompt** מאפשר לדגם להתחילה את הטקסט באמצעות פתיחה שננתן, ובכך לתת נקודת פתיחה מובהקה והמשריצירת תוקן ממוקד סביר נושא מסוים. אך המשפט הראשון שהמודל שלנו מקבל ליצירת הטקסט הוא "The movie was" movie מה שמשמעותו את הדגם להמשיך וליצור תוקן הקשור לביקורת על סרט, תוך שימוש בנמנים שעיליהם הוא אומן. המודל ינסה להשלים את המשפט באופן שמתאים לסוגנו ול頓ון של ביקורת שעיליהן הוא אומן, וזאת תוך שימוש מרובה על הגיון רלוונטיות לטקסט הנתון שהוא. "The movie was".

עשינו ניסוי ושנינו אותו ל- "to My hobby is to play a lot of games that have fun and make my mind up. I enjoy watching television shows such as Batman, The Flash and other comedies while listening music with friends or reading books on their shelf when the kids are away for some reason (which makes me love them so much)"

ונכל להסיק שהוא יכול להשלים את המשפט המקורי לפני לפני Fine-tuning, שלמד על מגוון טקסטים שונים.

## חלק ד:

עבור כל אחד מסוגי הפורמטים לקחנו המודל והтокניזר המתאים  
לקחנו את הدادה ואז חילקנו אותה לפי ערך ה label, ולקחנו 25 דגימות שליליות, ו 25 חיוביות שהן  
באור לכל היוטר 490 כי בפורמט zero-shot יש מגבלה על אורק הטקסט.  
חברנו שוב את הדטה ושמרכנו ב- combined\_dataset  
עבור כל אחד מהפורמטים ייצרנו את הפריפיקס המתאים וחברנו אותו בעזרת הפונקציה  
fix\_prefix, ואז עשינו טוקניזציה לחלק של ה text- וAz יצרנו דגימות לפי המודל המתאים, ואז עשינו  
decode, ואז הפכנו את התשובות מטキסטים למספרים 0\1 בשביל לחשב הדקות  
קיבלו התוצאות הנ"ל:

Zero-shot Prompting Accuracy: 0.94

Few-shot Prompting Accuracy: 0.92

Instruction-based Prompting Accuracy: 0.92

קיבלו שההתוצאה הכלכלית גבוהה עבור zero-shot, ובשניהם האחרים קיבלו את אותה תוצאה, נסיק מזה  
שהמודל הצליח לסוג יותר כאשר לא נתנו לו יותר מדי הוראות ודוגמאות.

### **פורמט של zero-shot:**

```
zero_shot_prefix = "Request: Tell me if the movie review is positive or negative. The review: "
```

### **פורמט של few-shot:**

```
few_shot_prefix = f"Tag the movie review if it is positive or negative: 1.'{example_label_0}':'Negative' 2.'{example_label_1}': 'Positive'"
```

, Now tag this review: "

דגם להביקורת חיובית example\_label\_1

### **פורמט של Instruction-based:**

```
Instruction_based_prefix = "You are an expert movie review classifier. Please classify this new movie review as positive or negative: "
```

### **דוגמה לשילוב מוצלח עבור כל אחת מהשיטות:**

Review 1: At least it's not full of senseless violence or fluff. It's also not very full of thought or a smooth " storyline. This story had some potential but the director appears to have lost site of the green. The story is disjointed and goes off in strange<br /><br />directions, to no conclusion. I also don't believe the director spent much time around a golf country club, either. Some of the depictions are absurd. Not very engaging.

```
Review 1 true label: negative
Review 1 zero-shot: negative
Review 1 few-shot: negative
Review 1 instruction-based: negative
```

### **דוגמה לשילוב לא מוצלח עבור כל אחת מהשיטות:**

"

Review 30: Not the most successful television project John Cleese ever did, "Strange Case" has the feel of a first draft that was rushed into production before any revisions could be made. There are some silly ideas throughout and even a few clever ones, but the story as a whole unfortunately doesn't add up to much.<br /><br />Arthur Lowe is a hoot, though, as Dr. Watson, bionic bits and all. "Good Lord."

Review 30 true label: positive  
Review 30 zero-shot: negative  
Review 30 few-shot: negative  
Review 30 instruction-based: negative

#### שאלות:

1. התוצאות לא תאמו את הציפיות שלנו, ציינו שה- **zero-shot** **יהי לא דיק hei** פחות כי לא נתנו לו הוראה מדויקת או דוגמאות מתאימות, אבל בפועל היה לו דיק hei גבוהה.
2. כן יש סיכוי שהතוצאות ישתנו אם נסיף יותר מדוגמה, יש אופציה שהדיק **עליה** כי אנחנו נתונים לו יותר מידע ומלאים אותו יותר לפניו שיתחיל לסתוג, אבל מצד שני יש אופציה שהוספה זו תסביר את תהליכי העיבוד ובשל זה המודל טעה. לסייע זה תלו בדוגמאות שנוסף וגם באורכי הבקורות.
3. הם דומים זהה לשניהם משתמשים בدادआ **קיים** וככל שבייל מסימה ספציפית והשוני ביןיהם הוא ש- **Fine-Tuning** : מעדקן את משלימות מודל קיים באמצעות אימון נוסף על דאטה ספציפי למשימה ספציפית. הוא דורש משאבים רבים וזמן חישוב גבוה. ו- **למידה מבוססת פרומפטים** : לא דורשת אימון נוסף, אלא משתמשת בפרומפטים כדי לכוון את המודל. היא גמישה יותר, וחסכונית במשאבים.
4. לא תמיד אפשר להחליפה **Fine-Tuning** במלידה מבוססת פרומפטים, זה תלוי בסוג המשימה. אם המשימה דורשת שהמודול ילמד מידע חדש שלא היה בدادआ של האימון של המודול המקורי, כמו תחומי רפואי או שפה טכנית, **Fine-Tuning** יהיה הכרחי כי הוא מאפשר למודול להתאים את עצמו לנתחים חדשים. מצד שני, אם המודול כבר מכיר את הנושא ואנו רק רוצחים להנחות אותו איך להשתמש בידע שלו, למשל בסיווג ביקורת או תרגום טקסטים פשוטים, אפשר להשתמש בפרומפטים כדי לכוון אותו בלי הצורך לאמן אותו מחדש. במקרה לפעמים פרומפטים מסוימים מספקים, אבל במקרים מסוימים נדרשות דיק גובה או למידה של מידע חדש- **Fine-Tuning** הוא הפתרון הנכון.
5. ניסוח הפרומפט משפיע מאוד על יכולת התשיבות שהמודול מחזיר, מכיוון שהמודלים מבוססים על דפוסים סטטיסטיים שנלמדו מהנתונים עליהם אומנו. כאשר הפרומפט ברור ומספק הקשר מתאים, המודול מפיק תוצאות מדויקות יותר, ואילו ניסוח מעורפל עלול להוביל לתשובות שאgioות או לא רלוונטיות. אפילו שניינו קטן בניסוח יכול לשנות את אופן ההבנה של המודול ולגרום לתשובות שונות. לכן, כדי לשפר את הביצועים, נבצע סיווג תוך שימוש בכמה פרומפטים שונים ונבדוק איזה מהם מניב את רמת הדיק הגבוהה ביותר.

## תיק ה:

- cn מדבר ב **sais** סביר להניח שהדעתה שהמודול למד עליה לא היכלה נשים בנהיגה, لكن המודול לא ידע ליצור תמונה כזו.
- הסיבות שגרמו לו להפיק את תוצאות אלה היא שהמודול עשה מה שלמדו אותו אז במקורה שתהיה בקשה ליצור תמונה שיש בה פרטים שלא ראה קודם הוא לא ידע לייצר משהו איזה דומה. במקרה שלנו נראה שכאר יש משפחה שנסועת באותו איזה גבר הוא נהוג, אך לא נספק מספיק נתונים למודול שマーאה אישה נהגת ולידה גבר שכן קיבלם תוצאות אלה.
- .3

תיצור לי תמונה שマーאה זוג שנסועים באותו מקום כאשר האישה נהגת ולידה יושב גבר



Here is the image of a couple driving in a car, with the woman driving and the man sitting beside her as the passenger. You can view and download the image if you wish.

## **.4. שאלת בונוס:**

דוגמה מ- ChatGPT כאשר ביקשתי ממנו Shiraea אישה שמצויה נישואין לגבר וזו הייתה התוצאה:

תציג לי תמונה של אישה פצעה ונישואין לגבר



Here is the image of a woman proposing marriage to a man in a romantic outdoor setting. You can view and download the image if you wish.

בדרך כלל נהוג שהגבר יציע נישואין לאישה لكن שאימנו את המודל של ChatGPT סופקו רק דוגמאות של גבר שמציע נישואין, لكن הוא לא ידע לייצר תמונה כאשר האישה הוא המבקשת. וכשנשינו לדיבוק הבקשה קיבלנו תוצאות כזאת:

תיצור לי תמונה של אישה כורעת ומחזיקה טבעת כדי להציע נישואין לגבר  
והגבר עופד



Here is the image of a woman kneeling and holding a ring to propose marriage to a man who is standing. You can view and download the image if you wish.