**דוח מסקנות: בחירת כלי הבינה המלאכותית המתאים ביותר להעשרת הסיפור**

**מוטיבציה לניסוי**

כחלק ממטרות הפרויקט, נבצע הרחבה לשלד הסיפור בעזרת בינה מלאכותית. לצורך בדיקת הכלי המתאים ביותר שיוכל להעשיר עברונו את הסיפור בצורה הטובה והמעניינת ביותר, ביצענו ניסוי שבו בחנו את כלי בינה מלאכותית הבאים :Grok 3, Claude 3.7, ChatGPT 4.5.

**אופן ביצוע הניסוי**

* יצרנו 3 שלדי סיפור תקינים (מבחינת כל ההיבטים שתוארו בדוח המתאים)
* את אותם השלדים הוספנו ל-prompt שיצרנו.
* מכל כלי הבינה המלאכותית שהצגנו למעלה ייצרנו סיפורים.
* עבור כל אחד מהסיפורים שייצרנו הכנו טבלה המכילה את הקטגוריות הבאות: דמויות, קונפליקט, מלאכת הכתיבה והיגיון.
* כל קטגוריה הינה משוקללת לפי חישוב מחדש של היחס לפי המסמך שקיבלנו.
* כל שאלה בכל מדד משוקללת באופן שונה כך שסכום כל המשקלים בכל מדד הינו 1.
* טווח ציוני השאלות נאים בין 0 – 4, כאשר 0 – לא רלוונטי, 1 – חוסר הסכמה,

2 – הסכמה חלקית, 3 – הסכמה מרובה ו- 4 – הסכמה מוחלטת.

**מהלך ההשוואה:**

בוצע ניתוח של שלושה סיפורים מכל בינה מלאכותית בצורה ידנית, כאשר המדדים בהם התרכזנו בהשוואה הם: דמויות, קונפליקט, מלאכת הכתיבה והיגיון.

עבור כל אחד מהמדדים השתמשנו בשאלות שהופיעו במאמר להערכת תסריטים תוך התאמת השאלות לסיפורים (ניתן לראות את השאלות בנספח א'), [קישור למאמר](https://www.specscout.com/score).  
הביקורת התבצעה ע"י 2 מבקרים שונים עבור כלל הסיפורים (3 סיפורים של ה-Grok, 3 סיפורים ה-Claude ו-3 סיפורים של ה-ChatGPT), עבור כל בינה מלאכותית חושב ממוצע של שלושת הציונים, ולאחר מכן בוצע ממוצע על שני הממוצעים שיצאו לכל מבקר.

**תוצאות ההשוואה:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Grok 3 | Claude 3.7 | ChatGPT 4.5 |
| מבקר 1 | 2.6 | 3.5 | 3.13 |
| מבקר 2 | 2.89 | 3.15 | 2.84 |
| ממוצע משוקלל | 2.745 | 3.325 | 2.985 |

**מסקנות מההשוואה:**

לאחר השוואת כלל הציונים כלי הבינה אשר הביא את ממוצע הציונים הטובים ביותר הינו: **Claude 3.7.**

**כעת נציג את הסבר קצר עבור כל מדד:**

1. **דמויות:**

במדד זה סוג השאלות שנעזרנו בהן ע"מ לתת דירוג, הן שאלות הבודקות את רקע הדמויות והתפתחותן במהלך העלילה.

1. **קונפליקט:**

במדד זה סוג השאלות שנעזרנו בהן ע"מ לתת דירוג, הן שאלות הבודקות את העניין בקונפליקט המרכזי והתפתחותו לאורך העלילה ובתתי הקונפליקטים המסתעפים ממנו במהלך העלילה.

1. **מלאכת הכתיבה:**

במדד זה סוג השאלות שנעזרנו בהן ע"מ לתת דירוג, הן שאלות הבודקות האם הסיפור מנוסח בשפה תקנית ומודרנית.

1. **היגיון:**

במדד זה סוג השאלות שנעזרנו בהן ע"מ לתת דירוג, הן שאלות הבודקות האם יש היגיון בעלילת הסיפור, למשל האם קיימים חורים או נקודות לא ברורות בעלילה.

**בנוסף לדו"ח זה, מצ"ב 2 קבצים אקסל (נספחים א' + ב') הכוללים את הציון של כל מבקר לכל אחד מהסיפורים, וקובץ Word נוסף (נספח ג') המכיל את ה Prompt שנשלח עבור כל סיפור ואת הפלט שהתקבל עבור כל כלי בינה מלאכותית.**