**Conclusion Report: A Comparison of Story Quality Between Plot Skeleton-Based Writing and Freeform AI Writing**

**Experiment Motivation**

במטרה להעריך את איכות התוצרים של האלגוריתם שפיתחנו ליצירת סיפורים, נערך ניסוי השוואתי בין סיפורים שנוצרו באמצעות האלגוריתם לבין סיפורים שהופקו באמצעות פרומפט זהה על ידי Claude 3.7. מטרת הניסוי הייתה לבחון האם הסיפורים האוטומטיים עומדים באותם קריטריוני איכות ואף עולים על הסיפורים הידניים, בהתבסס על הערכה זהה לפי פרמטרים מוגדרים מראש.

**אופן ביצוע הניסוי**

* יוצרו 50 סיפורים על ידי כל אלגוריתם (סה"כ 100 סיפורים), בהתבסס על Prompt אחיד.
* במקביל, נכתבו 100 סיפורים נוספים בצורה ידנית עבור אותם Prompts(ללא התבססות על שלד כל שהוא).
* כלל הסיפורים שנוצרו נבדקו ע"י המבקר האוטומטי.
* לצורך ההשוואה יצרנו 2 טבלאות, הראשונה הינה עבור השוואה בין 50 סיפורים הנוצרו באמצעות Plot Genie לבין 50 סיפורים ללא שלד, השנייה הינה עבור השוואה בין 50 סיפורים הנוצרו באמצעות Plotto לבין יתר 50 הסיפורים שנוצרו ללא שלד.
* כל קטגוריה הינה משוקללת לפי חישוב מחדש של היחס לפי המסמך שקיבלנו.
* כל שאלה בכל מדד משוקללת באופן שונה בהתאם להשפעתה על טיב הסיפור.
* 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 דירוג לכל סיפור על ידי המבקר האוטמטי, 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).

**תוצאות ההשוואה:**

**השוואה בין Plot Genie – No Skeleton**

**Plot Genie**

****

**No Skeleton**

****

|  |  |  |
| --- | --- | --- |
| **Statistical Measure** | **Plot Genie** | **No Plot Skeleton** |
| **Average** | 3.633 | 3.449 |
| **Median** | 3.6425 | 3.4625 |
| **STD** | 0.1149 | 0.181 |
| **Num Of Stories** | 50 | 50 |

להלן התוצאות המלאות של מבחני t לא מזווגים (Welch's t-test) להשוואה בין סיפורי Plot Genie לבין סיפורים ללא שלד (No Skeleton):

|  |  |  |  |
| --- | --- | --- | --- |
| **מסקנה סטטיסטית** | **p-value** | **t-statistic** | **מדד** |
| מובהקות גבוהה | 6.81E-12 | 7.882 | **Character** |
| מובהקות גבוהה | 2.57E-08 | 6.383 | **Conflict** |
| אין מובהקות | 0.118 | 1.576 | **Craft** |
| אין הבדל | 0.985 | -0.0187 | **Logic** |
| מובהקות גבוהה | 4.01E-08 | -0.0187 | **FINAL SCORE** |

**מסקנה מההשוואה:**

הממצאים מצביעים על כך ששימוש בשלד עלילה כמו Plot Genie תורם באופן מובהק לשיפור איכות הסיפור במדדים מהותיים כגון עיצוב דמות, בניית קונפליקט והציון הכולל של היצירה. לעומת זאת, לא נמצאה השפעה מובהקת על איכות הכתיבה הטכנית (Craft) או על ההיגיון הפנימי של הסיפור (Logic), מה שמעיד כי מרכיבים אלו תלויים פחות בשלד העלילתי ויותר בביצוע הספציפי של המודל הכותב.

**השוואה בין Plotto– No Skeleton**

**Plotto**

****

**No Skeleton**

****

|  |  |  |
| --- | --- | --- |
| **Statistical Measure** | **Plotto** | **No Plot Skeleton** |
| **Average** | 3.6429 | 3.415 |
| **Median** | 3.681 | 3.436 |
| **STD** | 0.141 | 0.186 |
| **Num Of Stories** | 50 | 50 |

להלן תוצאות מבחן T לא מזווג (Welch’s t-test) בין סיפורים שנכתבו בעזרת Ploto לבין סיפורים ללא שלד עלילתי (No Skelton):

|  |  |  |  |
| --- | --- | --- | --- |
| **מסקנה סטטיסטית** | **p-value** | **t-statistic** | **מדד** |
| מובהקות גבוהה | 2.92E-08 | 6.087 | **Character** |
| מובהקות גבוהה | 4.20E-10 | 7.267 | **Conflict** |
| מובהקות מובהקת | 0.0023 | 3.127 | **Craft** |
| מובהקות מתונה | 0.0381 | 2.102 | **Logic** |
| מובהקות גבוהה | 6.96E-10 | 6.89 | **FINAL SCORE** |

**מסקנה מההשוואה:**

השימוש בשלד העלילה Ploto מוביל לשיפור מובהק בכל מדדי איכות הסיפור – עיצוב הדמות, בניית הקונפליקט, איכות הכתיבה, ההיגיון הפנימי והציון הכולל. ממצאים אלו מעידים כי שלד עלילתי מסוג Ploto תורם תרומה רחבה ומהותית לכתיבה יצירתית, הן ברמת התוכן והן ברמת הביצוע הטכני של הסיפור.

**מסקנה סופית :**

הממצאים משני הניסויים מצביעים באופן עקבי על כך ששימוש באלגוריתם ליצירת שלד עלילתי תורם באופן מובהק לשיפור איכות הסיפור, בהשוואה לכתיבה ללא שלד כלל. הסיפורים שנכתבו באמצעות שלד אלגוריתמי הציגו ציונים גבוהים יותר במדדים מרכזיים כגון עיצוב דמויות, קונפליקט ועלילה, ולעיתים אף באיכות הכתיבה וההיגיון הפנימי – מה שמחזק את הטענה כי כתיבה מונחית-שלד מובילה לתוצרים ספרותיים איכותיים יותר.

**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 תיקיות הכוללת את כל הסיפורים שהופקו + דוחות ביקרות טיב הסיפור:**

1. Freeform Writing – כולל את כל הסיפורים ודוחות הביקורת עבורן עבור סיפור שנוצרו ללא שלד.
2. Plotto – כולל את כל הסיפורים ודוחות הביקורת עבורן עבור סיפור שנוצרו באמצעות שלד שנוצר ע"י אלגוריתם Plotto.
3. Plot Genie– כולל את כל הסיפורים ודוחות הביקורת עבורן עבור סיפור שנוצרו באמצעות שלד שנוצר ע"י אלגוריתם Plot Genie.