

### מטלה 3

שם קוד הקבוצה- אריאלנפדןסקו& פרדמן

ת.ז חברה הקבוצה-

1. 313306078 אריאל נפדןסקו

2. 316420132 אבי פרדמן

קישור לפרויקט -

<https://github.com/aviferdman/LLMs-and-Multi-Agent-Orchestration---Assignment3.git>

ציון מוערך- 95 😊

### טופס הגשת הערכה עצמית

שם הסטודנט/ים: אריאל נפדןסקו& פרדמן

שם הפרויקט: Turing Machine

תאריך הגשה: 27.11.2025

הציון העצמי שלי: 95/100

הצדקה להערכת העצמית (חובה - 200-500 מילים)

בסעיף זה, הסבירו למה בחרתם בציון זה. כללו:

נקודות חזק: מה עשיתם במיוחד טוב? אילו מרכיבים הם בrama גבוהה?

נקודות חולשה: מה חסר או יכול היה להיות טוב יותר? (כנות מוערכתי!)

השערה: כמה זמן ומאזן השקעתם?

חדשנות: האם יש משהו ייחודי או מיוחד בעבודה?

למידה: מה למדתם מהפרויקט?

נקודות חזק:

הפרויקט שלנו מציג יכולות יוצאות דופן בכל ממד ההערכתה. הגיענו לניסוי אמפירי מكيف של 21 משפטים

עם 7 רמות שיבוש (20%-50%), כולל 63 פעולות תרגום ואימונות 100% (21/21) משפטיים עברו בדיקת

אימונות עם סטייה ממוצעת של 1.5% בלבד). התיעוד שלנו הוא בرمת Production מלאה: יצרנו מסמך

PRD מלא הכלול יעדים ומדד הצלחה; תיעוד ארכיטקטורה מפורט עם 3 מסמכי החלטות

ארכיטקטוניות (ADRs) המסבירים את בחירת המודל, עיצוב שרשרת התרגומים, וסטרטגיית

Methodology מהירות הcolałת אימונות שיטתי של שיעורי טעויות, וניתוח כמותי Multi-Agent

ואיכו תני מקייף. המערכת בנויה על ארכיטקטורת Multi-Agent מודולרית עם תקשורת מבוססת-קבצים הפרדת אחראיות ברורה (Claude לתרגום ויצירת משפטיים, Chosk Python לחישובים מתמטיים), ומנגנון הרחבה המאפשרים שינוי שפות ורשראות תרגום בקלות. תיעדנו את כל ה-Prompts בסוף הפרומטאים, ביצענו ניתוח עלויות מפורט, והכנו את הפרויקט כקוד פתוח עם רישיון ו-[Contribution Guidelines](#).

#### נקודות חולשה:

למרות שהפרויקט עומד בכל הדרישות, אנו מזהים מספר נקודות לשיפור. הניסוי התמקד בשרותת תרגום אחת (אנגלית→צרפתית→איטלקית→אנגלית) - הרחבת לשרשראות נוספות היהת מחזקת את הממצאים. מערכת ה-SSL certificate verification חוסמת CUT הרצה מחדש מחדש ברשומות ארגוניות - ניתן היה להוסיף fallback mechanism או הורדה מקומית של המודל. בנוסף, אף שתיעדנו 21 משפטיים, מוגן גדול יותר (100-50) היה מספק סטטיסטיקה חזקה יותר. הפרדת האחריות (Claude vs Python) אופטימלית מבחינה עיצובית, אך יוצרת תלות ב-API של Claude לתרגומים.

#### השקעה:

ה השקעה בפרויקט הייתה שימושית וככל היכן טכנולוגית Multi-Agent, מימוש שלושה סוכני תרגום אוטונומיים, יצירה ידנית של 21 משפטיים מגוונים עם הזרקת טעויות בשיעורים מדויקים, ביצוע 63 פעולות תרגום, אימות שיטתית של כל משפט, חישוב 21 מרחקים סמנטיים, ויצירת ויזואלייזציה 4-פאנל מקצועית. השקענו במתודולוגיה מחקרית קפדיית, תיעוד מקייף, וביקורות איכות מתחשכות.

#### חדשנות:

הפרויקט מציג חידושים שימושיים. גילינו תבנית לא-LINIARITY מפתחית - שיא הסטיה הסמנטית ב-30% שיבוש (לא ב-50!), המuid על יכולת תיקון שגיאות מבניות ב-LLMs. הוכחנו שתוכן/דומין משפייע יותר משיעור השיבוש (תזמורת סימפונית ב-30% שיבוש: 0.824 סטיה; רותן היסטורי ב-40%: 0.295 סטיה). ישמןנו ארכיטקטורת Multi-Agent אמיתית עם תקשורת מבוססת קבצים, המדגימה Orchestration אוטונומי. הפרדנו ביצירה אוטומטית של טעויות ממודלים סטטיסטיים - Claude יוצר טעויות קוגניטיביות (substitution, deletion, transposition) ולא רנדומליות.

#### למידה:

הפרויקט העניק לנו הבנה עמוקה ב-Orchestration Multi-Agent, תקשורת בין-סוכנית מבוססת-קבצים, ומידה סמנטית עם Embeddings. למדנו כיצד LLMs מתפקידים כ"מתקני שגיאות" סטטיסטיים, מה ההבדל בין Semantic Similarity ל-Lexical Similarity, וכי צד עיצוב ניסויים מבוקר משפייע על תקיפות ממצאים. רכשנו ניסיון במתודולוגיות מחקר NLP, ניתוח סטטיסטי (Pearson correlation) על תקיפות ממצאים. הוכיחנו ניסיון במתודולוגיות מחקר: יכולות להפתח - הציפיות שלנו (!), וויזואלייזציה מדעית. הלקח המשמעותי ביותר: תוכאות מחקר יכולות להפתח - הציפיות שלנו (!).

(ליניאריות) נסתרו, והמצאים (U-shaped curve) חקרו התנהגות מורכבת של מערכות AI. למדנו חשיבות של אימות קפדי, תיעוד עמוקיק, והכנה לשכפול ניסויים.

### רמת הדקדקנות המבוקשת בבדיקה

על פי הציון העצמי שנתי, אני מבין/ה שרמת הבדיקה תהיה:  
100-90: דקדקנית ביותר - חיפוש "פילים בקונה", הקפדה על כל פרט

**מאפיינים:**

- קוד-ברמת ~~plugins~~ ~~production~~ ~~hooks~~ ~~extensibility~~, ~~hooks~~ ~~extensibility~~, ~~hooks~~ ~~arclitectural~~ ~~architecture~~
- תיעוד-מושלם ומפורט בכל היבט: PRD מקיים, תיעוד ארכיטקטורה מלא, README מקיים
- עמידה מלאה בתקן IEC 25010/ISO 25010
- בדיקות מקיפות עם כיסוי 85%+
- מחקר עמוקיק: ניתוח רגישות שיטתי, הוכחות מתמטיות, השוואת מבוססת-נתונים
- ניהול-אלאזיה בرمאה גבוהה-יותר עםashboard אינטראקטיבי
- ספר-פרומפטים מפורט ומתחודש
- ניתוח-עליזות מקיים עם המלצות אופטימיזציה
- חדשנות ויחידיות: רעיונות מקוריים, פתרון לבעה מורכבת
- תרומה להיליה: קוד פתוח, תיעוד לשימוש חוזר

### הצהרת יושר אקדמי (Academic Integrity Declaration)

אני מצהיר/ה בזאת ש:

- ההערכה העצמית שלי היא כנה ואמיתית
- בדקתי את העבודה מול כל הלקוחות לפני קביעת הציון
- אני מודע/ת שציון עצמי גבוה יוביל לבדיקה דקדקנית יותר
- אני מקבל/ת את העבודה שהציון הסופי עשוי להיות שונה מהציון העצמי
- העבודה היא פרי עובdotי/נו (של הקבוצה) ואני/ה אחראי/ה לכל תוצאה

חתימה: אריאל גולדנסקי ואבי פרידמן תאריך: 09.11.2025

