

## מטלה 5 - אפשרות מספר 1

שם קוד הקבוצה - אריאל נפדנסקי ואבי פרדמן

ת.ז חברי הקבוצה -

1. 313306078 אריאל נפדנסקי

2. 316420132 אבי פרדמן

קישור לפרויקט -

<https://github.com/aviferdman/LLMs-and-Multi-Agent-Orchestration---Assignment5.git>

ציון מוערך - 89 😊

## טופס הגשת הערכה עצמית

שם הסטודנט/ים: אריאל נפדנסקי ואבי פרדמן

שם הפרויקט: מעבדה ראשונה - context window

תאריך הגשה: 10.12.2025

הציון העצמי שלי: 89/100

הצדקה להערכה העצמית (חובה - 200-500 מילים)

בסעיף זה, הסבירו למה בחרתם בציון זה. כללו:

נקודות חוזק: מה עשיתם במיוחד טוב? אילו מרכיבים הם ברמה גבוהה?

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

השקעה: כמה זמן ומאמץ השקעתם?

חדשנות: האם יש משהו "יחודי" או מיוחד בעבודה?

למידה: מה למדתם מהפרויקט?

## נקודות חוזק

הפרויקט מציג איכות גבוהה במספר ממדים מרכזיים. מבחינה טכנית, נבנתה ארכיטקטורה מודולרית המבוססת על Strategy Pattern ו-Abstract Base Classes, המאפשרת הרחבה קלה. הקוד כולל 181 בדיקות עם כיסוי של 93.69% — גבוה משמעותית מהיעד של 85%.

מבחינת תיעוד, נוצרו שבעה מסמכי תיעוד עיקריים עם למעלה מ-5,000 שורות, כולל PRD מקיף (2,268 שורות), תיעוד ארכיטקטורה מלא (1,077 שורות), וניתוח סטטיסטי מפורט (מעל 650 שורות)

הכולל נוסחאות LaTeX. המחקר מבוסס על 220 שאלות LLM אמיתיות שבוצעו בארבעה ניסויים מתוכננים היטב, עם ניתוח סטטיסטי שיטתי (ANOVA, רגרסיה, effect sizes) ומודלים מתמטיים.

הוויזואליזציות כוללות 13 גרפים באיכות פרסום (DPI 300) עם אנוטציות סטטיסטיות.

הפרויקט עומד בתקני ISO/IEC 25010 בכל הקטגוריות (Performance, Reliability, Usability, Maintainability, Security).

---

## נקודות חולשה

לצד האיכות הגבוהה, ישנם מספר תחומים הדורשים שיפור. ראשית, Dashboard אינטראקטיבי – הנדרש ב-PRD – לא יושם. שנית, ספר פרומפטים לא תועד במסמך נפרד, אף שהפרומפטים קיימים בקוד. שלישית, ניתוח העלויות מופיע ב-PRD אך לא בוצע בפועל עם נתונים אמיתיים.

בנוסף, הפרויקט מוגבל למודל אחד (Llama 2). אף שהארכיטקטורה תומכת במספר ספקים, לא בוצעו ניסויים עם מודלים נוספים. לבסוף, מספר יכולות מתקדמות לא יושמו, כגון Kaplan–Meier, ויזואליזציות תלת-ממדיות ואנימציות.

---

## השקעה

השקעתי בין 80 ל-100 שעות בפרויקט, כולל:

תכנון ותיעוד (25 שעות), פיתוח קוד (35 שעות), ביצוע 220 שאלות LLM (15 שעות), ניתוח סטטיסטי (10 שעות), ויזואליזציה (8 שעות), בדיקות (12 שעות) ו-5 polish (שעות).

ההשקעה הייתה משמעותית ורחבה בכל היבטי המערכת, עם תשומת לב גבוהה לפרטים ולאיכות המחקרית.

---

## חדשנות

הפרויקט כולל מספר תגליות מקוריות: זיהוי Performance Cliff א מפירי ב-2,500 tokens עם מודל מתמטי ( $R^2 = 0.42$ ), הפרכת "Lost in the Middle" בקנה מידה קטן, השוואה שיטתית של Chunking Strategies (יתרון של 7% ל-Semantic), והערכה כמותית של Context Management Strategies.

הפתרון מהווה מערכת מחקר אמפירית מלאה, reproducible, עם framework לשימוש חוזר.

---

## למידה

העמקתי בהבנת LLM context windows, מתודולוגיית מחקר הכוללת ניתוח השערות וסטטיסטיקה, יישום RAG pipeline, עקרונות הנדסת תוכנה, ויצירת ויזואליזציות באיכות פרסום. התובנה המרכזית: מחקר אמפירי איכותי דורש תכנון קפדני, ניתוח סטטיסטי מדויק ותיעוד מקיף לצורך reproducibility.

## סיכום

הפרויקט עומד בכל הדרישות העיקריות ואף מעבר לכך, עם תיעוד יוצא דופן, קוד איכותי ומחקר מעמיק. הציון 89 משקף איכות גבוהה מאוד לצד מספר תחומים לשיפור (Dashboard, ספר פרומפטים, יישום ניתוח עלויות, השוואה בין מודלים).

### הצהרת יושר אקדמי (Academic Integrity Declaration)

#### אני מצהיר/ה בזאת ש:

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

חתימה: אריאל נפדנסקי ואבי פרדמן, תאריך: 10.12.2025

