

מודלי AI להערכת מטלות ומבחנים עתירי טקסט: היתכנות ליישום בהשכלה הגבוהה

כנס מיט"ל יולי 2026

ד"ר אילן דניאלס רחימי

- **מגבלות ההערכה האנושית**
אתגרי סטנדרטיזציה, שונות בין בודקים והטיות מובנות בניתוח תוכן מורכב
- **נוכחות גוברת של כלי בינה מלאכותית**
שינוי הרגלי למידה עקב אימוץ נרחב של הטכנולוגיה בקרב סטודנטים
- **עומס מוסדי**
פער בין היקף המטלות לבין היכולת למתן בדיקה איכותית ומשוב בזמן סביר

רקע ומתודולוגיה

Project Based Research = Design Science Research

- פרדיגמת מחקר במערכות מידע לפיתוח והערכת פתרונות טכנולוגיים
- התוצר של הפיתוח הוא מוקד המחקר ובו בזמן הוא כלי המחקר.

שילוב בין

- Theory grounding - עיגון תיאורטי: להסתמך על תיאוריות, מודלים ומחקרים קיימים, ולא רק על התרשמות או ניסוי מעשי. Systematic evaluation
- הערכה שיטתית: לבדוק את התוצאות לפי שיטה מסודרת, קריטריונים ברורים, מדדים מוגדרים והשוואות עקביות

DSR רואה בפיתוח המערכת עצמה
תהליך מחקרי המייצר ידע חדש

מחזור המחקר



האם שילוב בינה מלאכותית אפשרי בקנה מידה רחב?



מוכנות סטודנטים לאימוץ פתרון בדיקה משולב AI - אמון

מוכנות
סטודנטים



אמון בכלי AI

מבנה
המערכת



הערכה אובייקטיבית



הערכה מדויקת



משוב מותאם אישית

מחקר
הערכה

הערכה
אנושית

הערכה
ע"י AI

מוכנות סטודנטים לאימוץ פתרון בדיקה משולב AI - טכנולוגיה

מוכנות
סטודנטים



שילוב AI בבדיקת
מטלות אקדמיות

מבנה
המערכת



אופן הבדיקה מועדף

מחקר
הערכה

מוכנות סטודנטים לאימוץ פתרון בדיקה משולב AI - טכנולוגיה

מוכנות
סטודנטים



תדירות שימוש אקדמי שבועי ב AI

מבנה
המערכת

הסתגלות 75% מהסטודנטים מוכנים למעבר למבחן ממוחשב

מחקר
הערכה

מוטיבציה 35% מוכנים לשלם עבור ציון מיידי ומשוב מפורט



40%



4.5-5%



5-7%



5-6%

מוכנות סטודנטים לאימוץ פתרון בדיקה משולב AI - פרופיל הלומד

מוכנות
סטודנטים

מתאם נמוך בין תדירות שימוש לתמיכה בשילוב Cor ↓

מתאם נמוך בין גיל לתמיכה בשילוב Cor ↓

מבנה
המערכת

אי-מובהקות בין דור Z לדור X ו Y t X

אי-מובהקות בין מכינה קדם אקדמית ללימודי תואר t X

מחקר
הערכה

תמיכה חוצת אוכלוסיות ואינה תלויה בתדירות שימוש 

מוכנות סטודנטים לאימוץ פתרון בדיקה משולב AI - סיכום

מוכנות
סטודנטים

הסתגלות גבוהה ☒

מוטיבציה גבוהה ☒

תמיכה בשילוב חוצת
אוכלוסיות ואינה תלויה
בתדירות שימוש ☒

שימוש תדיר של סטודנטים בAI ☒

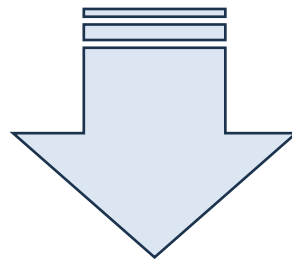
אמון בכלי AI ☒

תמיכה בשילוב AI ☒

Human-in-the-Loop ☒

מבנה
המערכת

מחקר
הערכה



מוכנות גבוהה לשילוב AI

כיצד נראה שילוב יישום של AI ככלי עזר להערכת מטלות ומבחנים?

נבחן מערכת המיועדת להתמודד עם אתגרי הניתוח והערכת התוכן
המאפיינים קורסים עתירי כתיבה

להלן נציג:

מנגנון הפעולה. כיצד המערכת מנתחת טקסט ומפיקה הערכה

מחקר הערכה. השוואה בין ביצועי ה-AI לבדיקת מתרגל בקורס אקדמי

מוכנות
סטודנטים

מבנה
המערכת

מחקר
הערכה

ארכיטקטורת המערכת

מובנות
סטודנטים

מבנה
המערכת

מחקר
הערכה

מערכת *Whizz.ai* להערכת עבודות אקדמיות

מאפייני המערכת

למידה מתמשכת: סינרגיה המאפשרת למערכת להשתפר לאורך זמן

Human-in-the-Loop: מערכת AI אינה מחליפה שיקול דעת מקצועי

שקיפות והגינות: הגנת פרטיות בסטנדרט רגולטורי GDPR

ארכיטקטורת המערכת

מובנות
סטודנטים

מבנה
המערכת

מחקר
הערכה

Knowledge Ingestion

עיבוד תשתית הידע ע"י הזנת ולמידת חומרי הקורס

Rubric Indicator

יצירת מחוון ע"י מקריטריונים כמותיים ואיכותניים

Few-shot Promptin

כיוול באמצעות דגימת עבודות

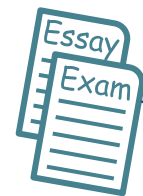
Human-in-the-Loop

ארכיטקטורת המערכת מרכיבי מנגנון ההערכה של Whizz.ai

מובנות
סטודנטים

מבנה
המערכת

מחקר
הערכה



סטודנט

דוגמאות
כיוול

דוגמאות עבודות
מדורגות בציונים
40, 50, 60, 70,
80, 90 ו-100.

מחווון
הערכה

מחווון מגדיר
קריטריונים
לחלוקת הנקודות

מאגר ידע
ייעודי

מאגר חומרי קורס:
סילבוס, מצגות, מאמרי
חובה, סיכומי שיעור,
חומרי עזר, תמלול
שיעורים

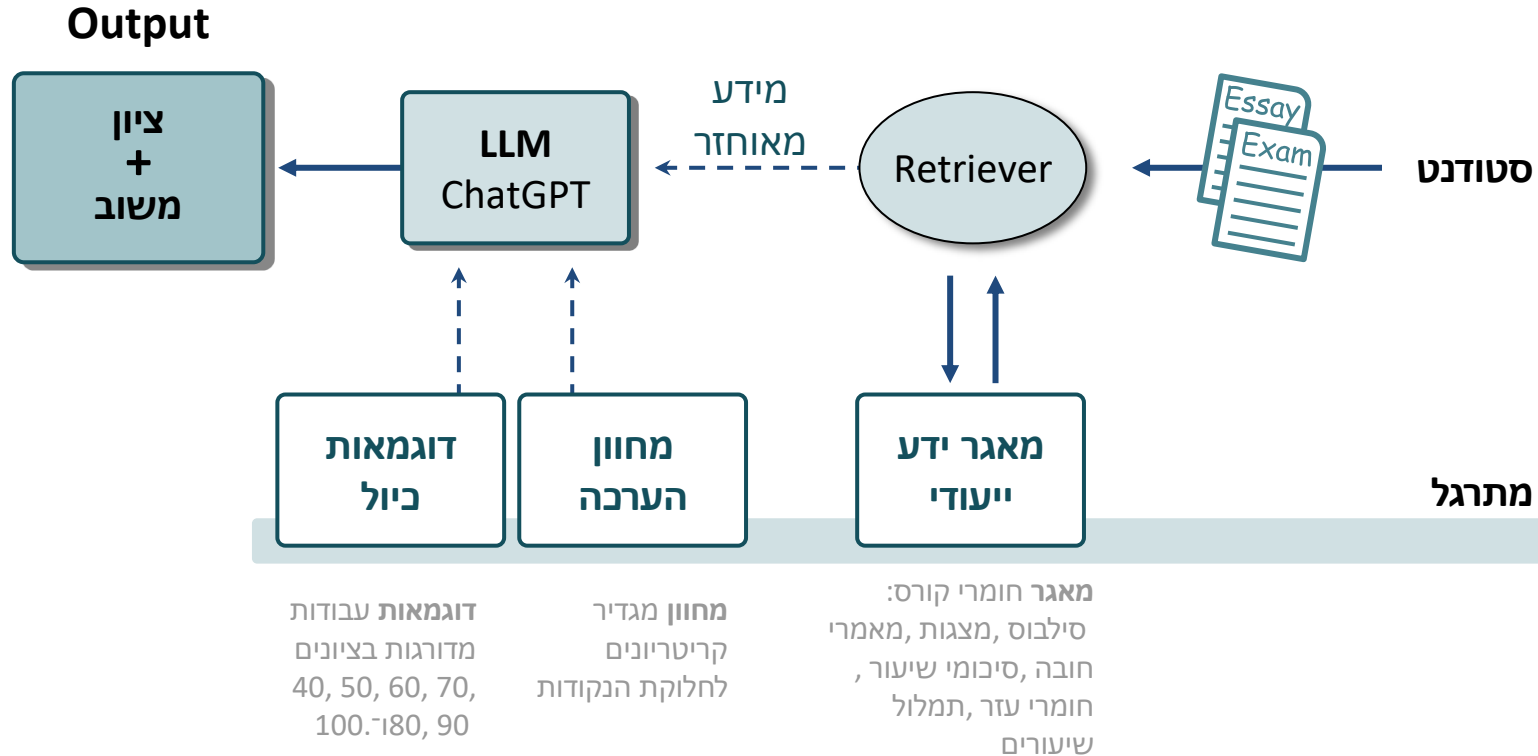
מתרגל

ארכיטקטורת המערכת מרכיבי מנגנון ההערכה של Whizz.ai

מובנות
סטודנטים

מבנה
המערכת

מחקר
הערכה



ארכיטקטורת המערכת מרכיבי מנגנון ההערכה של Whizz.ai

מובנות
סטודנטים

מבנה
המערכת

מחקר
הערכה



הערכת המודל

האם הערכת עבודות אקדמיות על ידי מודל AI זהה להערכה אנושית?

דאטה

בחינות סוף סמסטר תשפ"ה ב'
קורסים: משפט חוקתי ודיני עונשין

מתודולוגיה

- כל בחינה נבדקה פעמיים:
- בדיקה אנושית על ידי מתרגל מומחה
 - בדיקת מכונה על ידי מודל בינה-מלאכותית Whizz.ai

מדידה

עקביות במתן ציונים: Mean ו SD של ציונים

מובנות
סטודנטים

מבנה
המערכת

מחקר
הערכה

הערכת המודל השוואת מדדים

מובנות
סטודנטים

מבנה
המערכת

מחקר
הערכה

**Mean Diff AI–
Human**

**Human Mean
(SD)**

**AI Mean
(SD)**

-6.950

80.45
(10.89)

73.50
(10.01)

**משפט
חוקתי**

0.004

73.84
(12.03)

73.85
(7.30)

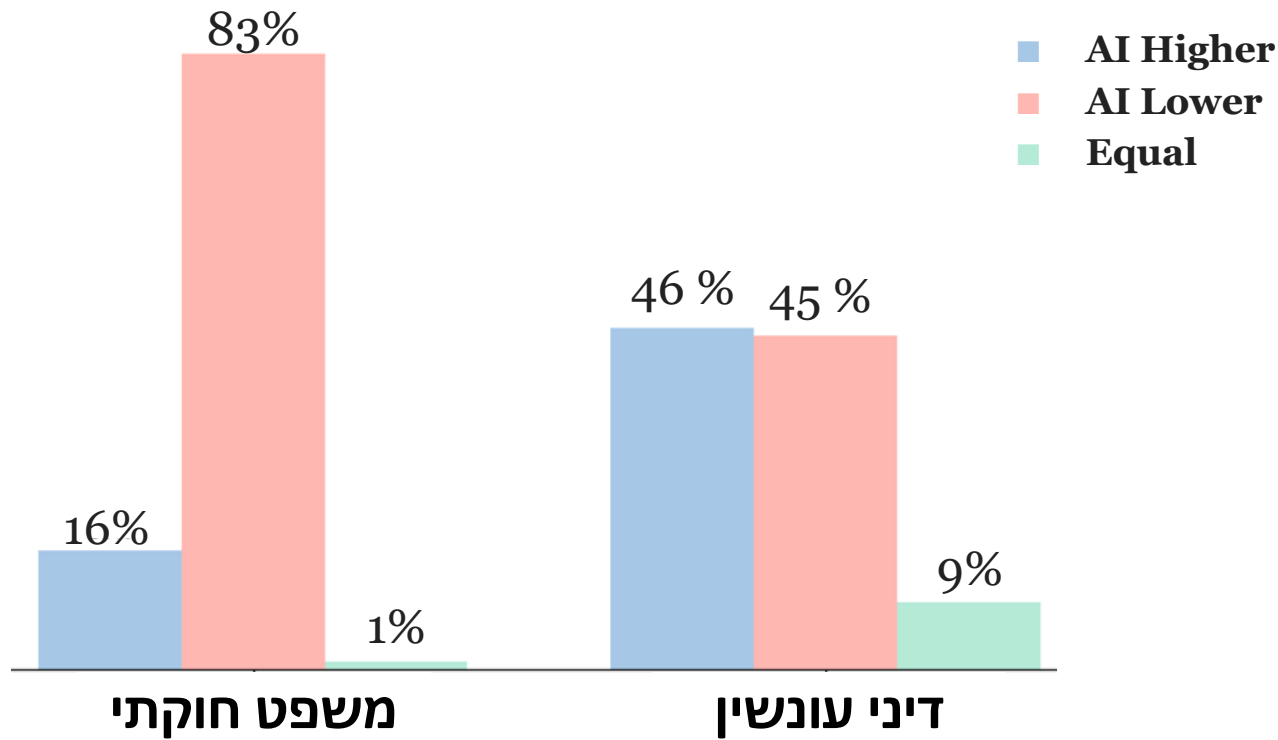
**דיני
עונשין**

הערכת המודל כיוון ההבדלים בין AI ל Human

מובנות
סטודנטים

מבנה
המערכת

מחקר
הערכה



הערכת המודל ניתוח ההבדל במדדים

מובנות
סטודנטים

התוכן שהוזן
למערכת

אופי
השאלות

מבנה
הרובריקה

**משפט
חוקתי**

פירוט כללים
ורכיבים

זיהוי והצגת
מרכיבים
ספציפיים

חומר רלוונטי
וממוקד

מבנה
המערכת

**מחקר
הערכה**

**דיני
עונשין**

פירוט פחות
פורמלי

דרשו דיון,
שקלול של
הטיעונים

חומר רלוונטי
רב אך לא
ממוקד

הערכת המודל בחינת ההבדל במדידה

מובנות
סטודנטים

מבנה
המערכת

מחקר
הערכה

הערכה ע"י המודל

- הערכה יותר אנליטית
- מאתר רכיבים חסרים

התוכן שהוזן למערכת

חומר רלוונטי
וממוקד

אופי השאלות

זיהוי והצגת
מרכיבים
ספציפיים

מבנה הרובריקה

פירוט כללים
ורכיבים

**משפט
חוקתי**

- אומדן כוללני
- הערכת איכות התשובה

חומר רלוונטי
רב אך לא
ממוקד

דרשו דיון ,
שקלול של
הטיעונים

פירוט פחות
פורמלי

**דיני
עונשין**

ההיבט מוסדי ליתרונות מערכת AI לבדיקת בחינות

יעילות תפעולית. צמצום בעומסי הערכה לטובת מחקר וההוראה

סטנדרטיזציה. משוב עקבי ואחיד המצמצם שונות בין בודקים

אמון ואובייקטיביות. חיזוק תפיסת ההוגנות והשקיפות במערך הציונים

ההיבט מוסדי ליתרונות מערכת AI לבדיקת בחינות

יעילות תפעולית.

סטנדרטיזציה.

אמון ואובייקטיביות.

המתאם בין הפתיחות הטכנולוגית של הסטודנטים לבין רמת הדיוק של המערכת בפועל מייצרת
הזדמנות לשינוי מודל ההערכה האקדמי

השילוב בין תשתית מוטיבציונית חזקה ליכולת ביצוע עקבית מוכיח כי ה AI-כבר אינו רק כלי עזר
תאורטי, אלא פתרון יישומי בשטח