

אפי | جاس | ISPA

אגודה ישראלית לפסיכומטריקה

الجمعية الإسرائيلية للبيكومتريكا

ISRAELI PSYCHOMETRIC ASSOCIATION

מבחנים ■ מדידה ■ הערכה

הכינוס השנתי ה-21 של האגודה הישראלית לפסיכומטריקה

11 בפברואר 2025

מוזיאון
ארצות
המקרא
בירושלים

בתמיכת:

מאל"ו

מרכז ארצי לבחינות ולהערכה (נ"ר)

المركز القطري للاختبارات والتقييم | NATIONAL INSTITUTE FOR TESTING & EVALUATION

מיסודן של האוניברסיטאות בישראל

האגודה הישראלית לפסיכומטריקה (אפי) היא גוף המאגד בתוכו אנשי מקצוע מתחום המדידה וההערכה החינוכית והפסיכולוגית.

מטרות האגודה:

- לעודד את הפעילות בתחום הפיתוח והיישום של מבחנים סטנדרטיים ושיטות להערכת ביצוע
- לקדם את המחקר והעשייה בתחומי המדידה החינוכית והפסיכולוגית
- לשכלל את שיטות ההערכה וכלי המדידה ולבסס את יישומם המושכל
- להתוות, להעריך ולהפיץ כלים מתודולוגיים עדכניים ושיטות הערכה חדשות
- להתריע על שימוש לקוי במבחנים ועל פרשנות מוטעית של תוצאותיהם
- לעודד קשר מקצועי, חילופי מידע ושיתוף פעולה בין העוסקים בתחום המדידה וההערכה

<p>התכנסות רישום וארוחת בוקר קלה דברי פתיחה וברכות – יו"ר אפי: ריטה סבר יו"ר הכינוס: חגית גליקמן, יוסי מחלוף</p>	<p>8:45–9:45 9:45–10:00</p>
<p>הרצאת אורחת <ul style="list-style-type: none"> Who and What is Conscious? Exploring the Possibility of Tests for Consciousness ליעד מודריק </p>	<p>10:00–11:00</p>
<p>בינה מלאכותית מול בינה אנושית – האם המבחנים מבחינים? <ul style="list-style-type: none"> How Well do LLMs Solve Chemistry MCQs in Hebrew? Benchmarking the Performance of Three Different Models אלעד יעקבסון, יעל שליפר, זיוה בר-דב, שלי רפ, רון בלונדר וג'ורא אלכסנדרון Applying Item Response Theory to Distinguish Between Human and Generative AI Responses to Multiple-Choice Assessments אלונה סטרוגצקי וג'ורא אלכסנדרון </p>	<p>מושב א' יו"ר: מירב אריאלי-אטאלי 11:00–11:40</p>
<p>הפסקת קפה</p>	<p>11:40–12:00</p>
<p>סימפוזיון – ממשקים בין בחינות ובינה מלאכותית: אתגרים והזדמנויות <ul style="list-style-type: none"> שימוש בבינה מלאכותית לפיתוח פריטים לבחינת ידע באנגלית יוני שריג, רועי שילו, אפי לוי, יעל מיכאלי וצור קרליץ העתיד במבחן: הערכה מחדש של בגרויות בעידן הבינה המלאכותית רונית נחמיה ונעמה מרז מהאנושי למכונה: אתגרים והזדמנויות בהערכת הבעה בכתב באמצעות בינה מלאכותית תמי סבג שושן, הדיל שחברי, הדס גלברט ובוועד רוזנבאום דיון </p>	<p>מושב ב' מנחה: צור קרליץ 12:00–13:30</p>
<p>מדידת ביצועים, מיומנויות, תפיסות והתאמת קריירה בקרב מבוגרים <ul style="list-style-type: none"> הבדלים מגדריים בביצוע מטלות כירורגיות בסימולטור לפרוסקופיה: השוואה בין מדדי ביצוע נועה גזית, גלעד בן גל ורון אלישר המחזור השני של סקר המיומנויות הבין-לאומי (PIAAC): טעימות מהממצאים, אתגרים והזדמנויות נעה שחורי איל וחיים פורטנובי תפיסות של איכות חיים, שלומות וסקרנות בגיל השלישי מנוחה בירנבוים, פאדיה נאסר אבו אלהיג'א, חני שלטון, הלנה קימרון ורובנה רוזנסקי שאלון התאמת קריירה בעידן ה-Big Data: חדשנות בייעוץ תעסוקתי רחל חביב וויטמן ופרדי אדיב </p>	<p>מושב ב' יו"ר: סול פיין 12:00–13:30</p>
<p>ארוחת צהריים</p>	<p>13:30–14:15</p>
<p>תוקף ניבוי וקשרים – מדידה בקרב תלמידי בית ספר ומורי דרך לעתיד <ul style="list-style-type: none"> הישגים באוריינות קריאה בכיתה ד': קשרים בין הישגים למשתנים אישיים ובית ספריים אליענה אביצור, ירון שלומי, פרנקלין זרומב, נורית ליפשוטט ויוסי מחלוף לקראת הערכה מעצבת אוטומטית של מורים במסגרת סימולציה: הערכה של קוהרנטיות הדיון בעזרת בינה מלאכותית יו"א אסאנו, ביאטה בייגמן קלבנוב וג'ימי מיקסקה בחינת התוקף של הליך האיתור של תלמידים מחוננים בישראל באמצעות מדדי הישג לימודיים דוד מעגן, יואל רפ וענבל שני הליך הייעוץ הפסיכומטרי למערך בחינות הרישוי למורי דרך – אתגרים, הישגים ולקחים דוד ציגלר, קרן גולדמן ורם נעמן </p>	<p>מושב ג' יו"ר: עינת נוטע קורן 14:15–15:45</p>
<p>שולחן עגול: פרס אפי – היבטים פסיכומטריים, חברתיים ותהליך בחירת הזוכה</p>	<p>מושב ג' מנחה: איתמר גתי 14:15–15:45</p>
<p>אסיפת חברי אפי</p>	<p>15:45–16:30</p>



Who and What is Conscious? Exploring the Possibility of Tests for Consciousness

ליעד מודריק
אוניברסיטת תל אביב

One of the greatest challenges for the science of consciousness is to provide the tools to determine which organisms and systems are conscious. Developing such tools is crucial not only for scientific reasons, but also for societal and practical ones. However, this is not an easy feat; in this talk, I will first describe the ongoing investigation of the neural basis of conscious experience, and then ask how it can inform the attempts to detect consciousness in non-trivial cases. I will present leading tests for consciousness, and explore different approaches to validating and developing such tests, based on neuroscientific theories of consciousness, on functional arguments, and on the natural kind approach. The talk will further highlight the entailed difficulties in the process, and ask how it can be applied to different populations, including artificial intelligence.



How Well do LLMs Solve Chemistry MCQs in Hebrew? Benchmarking the Performance of Three Different Models

אלעד יעקבסון, יעל שלייפר, זיוה בר-דב, שלי רפ, רון בלונדר וג'ורא אלכסנדרון
מכון ויצמן למדע

ההתפתחות המהירה של מודלי שפה גדולים (LLMs) מעוררת עניין מדעי וחשיבות מעשית בהערכת יכולותיהם במשימות אנושיות שונות. בהקשרים של חינוך מדעי, הבנת יכולות ה-LLMs לפתור מבחנים סטנדרטיים עשויה לשפוך אור על מאפייני השאלות שהופכים אותן למאתגרות במיוחד עבור בינה מלאכותית יוצרת. בנוסף, מכיוון שהביצועים של LLMs תלויים מאוד בשפה שבה הם פועלים, עם עדיפות ברורה לאנגלית על פני שפות אחרות, יש חשיבות רבה לביצוע הערכה של ביצועי LLMs במגוון רחב של שפות.

מטרתו המרכזית של מחקר זה הייתה לפתח בנצ'מרק בתחום החינוך למדעים בעברית, המתבסס על שאלות מתוך בחינות הבגרות בכימיה. הבנצ'מרק נבחן על שלושה מהמודלים המתקדמים ביותר, וביצועיהם הושוו לביצועי תלמידי תיכון. ספציפית, המחקר הונחה על ידי שאלת המחקר הבאה:

מהי רמת הביצועים של מודלים מתקדמים בפתרון שאלות מתוך בחינות בגרות סטנדרטיות בכימיה, וכיצד הם בהשוואה לביצועי תלמידי תיכון?

כדי להשיב על שאלה זו, פותח בנצ'מרק מקיף הכולל 120 שאלות רב-ברירה בעברית, שנלקחו מבחינות הבגרות בכימיה בישראל ב-15 השנים האחרונות (2009 – 2023), שמונה שאלות מכל בחינה. ביצועיהם של שלושה מודלים מתקדמים נבחנו על מדד זה: Gemini 1.5-Ultra, ChatGPT 4o, Claude 3.5 Sonnet. כל מודל נבחן בשני תנאים:

1. Zero-Shot: כל 120 השאלות הוצגו ל-LLM בו-זמנית, עם ההנחיה: "פתור את כל 120 השאלות. ספק רק את המספרים של התשובות הנכונות."

2. Few-Shot + Chain-of-Thoughts: המודל קיבל קובץ ובו 8 דוגמאות של שאלות עם פתרונות מפורטים. ההנחיה הייתה לפתור את השאלות בצורה דומה לדוגמאות, תוך מתן פתרונות מפורטים שלב-אחר-שלב.

כדי להתמודד עם הבעיה הידועה של חוסר עקביות בתשובות של LLMs, יישמו גישת self-consistency: כל מודל הורץ חמש פעמים בכל אחד מהתנאים, והתשובה שהופיעה בתדירות הגבוהה ביותר נבחרה כתשובה הסופית.

בנוסף, כדי להתחשב בהבדלים אפשריים ברמת הקושי של השאלות בין השנים השונות של בחינות הבגרות, נערכה גם השוואה שנתית של ביצועי ה-LLMs מול ביצועי התלמידים.

הממצאים הראו שתלמידים השיגו ביצועים גבוהים משמעותית מכל המודלים על הבנצ'מרק שפותח, ששניים מהמודלים (Claude, ChatGPT) השיגו שיפור בביצועים שלהם בתנאי Few-Shot + Chain-of-Thoughts, וש-Claude עלה בביצועיו על שני המודלים האחרים.

ניתוח לפי שנים של ביצועי המודלים מול ביצועי התלמידים חשף פערים ניכרים. בעוד שביצועי התלמידים היו יחסית עקביים לאורך השנים, ביצועי ה-LLMs השתנו באופן משמעותי. בנוסף, לא נמצא קשר בין ביצועי התלמידים לביצועי ה-LLMs: בשנים מסוימות בהן התלמידים הצליחו, המודלים התקשו, ולהיפך. פער זה מרמז על כך שהגורמים המגדירים את רמת הקושי במבחן שונים בין תלמידים אנושיים לבין מודלי השפה.

תרומת מחקר זה היא גם בפיתוח מדד מהימן ראשון מסוגו להערכת ביצועי LLMs בחינוך מדעי בעברית, עם דגש על בחינות בגרות בכימיה, וגם בהדגשת המגבלות הנוכחיות של מודלי השפה בפתרון שאלות מדעיות ספציפיות בשפות שאינן אנגלית.



Applying Item Response Theory to Distinguish Between Human and Generative AI Responses to Multiple-Choice Assessments

אלונה סטרוגצקי וג'ורא אלכסנדרון
מכון ויצמן למדע

לבינה מלאכותית גנרטיבית פוטנציאל רב לקדם את החינוך, אולם ישנו גם חשש מבוסס שהיא תוביל לשימוש לא אתי והונאות בהיקף רחב שיובילו לפגיעה באיכות הלמידה ובמהימנות של מדידה והערכה בחינוך. אולם למרות השימוש הנרחב בשאלות רב-ברירה (MCQs) בכלי הערכה שונים, הנושא של זיהוי השימוש בבינה מלאכותית למענה על שאלונים מבוססי MCQ כמעט ולא נחקר, בניגוד לזיהוי השימוש בבינה מלאכותית גנרטיבית במשימות כתיבה עשירות בטקסט. אנו מציעים שיטה המבוססת על תיאוריית התגובה לפריט (IRT) כדי להתמודד עם הפער הזה. ההנחה שלנו היא שבינה מלאכותית ואנושית יציגו דפוסי תגובה שונים על סדרות של פריטי הערכה, כיוון שהבינה המלאכותית מושפעת בצורה שונה מלומדים אנושיים ממימדים שונים של הבעיות שמציגים הפריטים. על כן מענה הבינה מלאכותית למבחן יראה כסטיות מהדפוסים הצפויים של תגובות אנושיות. כדי למדל את הסטיות האלו אנו משתמשים בכלים מתחום ה-person-fit statistics, ובאופן ספציפי, אנו מודדים אותן באמצעות ארבעה מדדים: G , G^* , $U3$, $ZU3$. שהוכחו כיעילים לזיהוי הונאות.

כדי להדגים זאת, השתמשנו בנתונים של מענה אנושי לשני שאלונים: שאלון הכנה לבגרות בכימיה והפרק הכמותי של הבחינה הפסיכומטרית. על ידי שימוש במדדים שצוינו לעיל, הראינו שניתן להבחין באופן מובהק סטטיסטית בין התשובות האנושיות לבין המענה של הבינה המלאכותית. התשובות של הבינה המלאכותית נוצרו על ידי גרסאות הפרימיום של הצ'אטבוטים קלוד, ג'מיני, ו-ChatGPT-4o. כמו כן, הראינו כי ניתן להשתמש באותה שיטה בדיוק גם כדי להבדיל בין תגובות של הצ'אטבוטים השונים לאותו שאלון (ז"א, הצ'אטבוטים השונים מפגינים מודלים קוגניטיביים שונים).

התוצאות גם הראו כי קיימת רגישות של השיטה לרמת "הזיהום" של הנתונים על ידי התשובות של הבינה המלאכותית. רמת ה"זיהום" מוגדרת כיחס בין מספר התשובות שיוצרו על ידי הבינה המלאכותית למספר התשובות האנושיות שקיימות במאגר הנתונים. בחנו את השיטה על שני השאלונים תחת שלוש רמות שונות של "זיהום": 5%, 10% ו 25%. נמצא שככל שרמת הזיהום גדלה, קטנה האפקטיביות של השיטה להפרדה בין כל אחד מכלי הבינה המלאכותית לתגובות האנושיות. ההסבר הוא ברור: השיטה מבוססת על זיהוי תגובות הצ'אטבוטים כחריגות ביחס לנורמה, שמוגדרת על ידי כלל הלומדים. ככל שגדלה כמות התשובות שמקורן בצ'אטבוטים, כך תשובות אלו הופכות לפחות חריגות.

לסיכום, העבודה שלנו מספקת הן בסיס תיאורטי והן ראיות אמפיריות ליישום IRT לצורך זיהוי מענה של בינה מלאכותית בשאלונים מבוססי MCQ, ומצביעה על מגבלות המובנות בשיטה.



סימפוזיון

ממשקים בין בחינות ובינה מלאכותית: אתגרים והזדמנויות

צור קרליץ
מאל"ו

ההתפתחות המהירה של אלגוריתמים של בינה מלאכותית (AI) מביאה איתה הזדמנויות ואתגרים בכל תחום של חינוך. שימוש בבינה מלאכותית עשוי לייצל תהליכים, לאפשר תיסלום (Scaling-up) של משימות, לקדם הפקת ידע עשיר מנתונים מורכבים, לייצר תכנים ברמה גבוהה ועוד. עם זאת, היישום של בינה מלאכותית עשוי להיות מלווה בקשיים, צפוי לדרוש השקעה גדולה של משאבים, עלול להיות מושפע מהטיות ולהפיק תוצרים שאינם בהכרח מתאימים לצרכים. כמו בכל הטמעה של טכנולוגיה חדשה בעולם קיים, יש לבחון היטב את העלות מול התועלת.

בסימפוזיון שלהלן ננסה לתאר כיצד מהפכת ה-AI משפיעה על עולם הפסיכומטריקה, ובאופן ספציפי על יישומים במערכת החינוך בישראל. השפעה זו נבחנת משלושה היבטים המייצגים ממשקים שונים בהם הבינה המלאכותית עשויה להשתלב בתחום המדידה וההערכה החינוכית. ממשק אחד נוגע לשימוש בכלי AI לפיתוח פריטים, ממשק שני נוגע לשימוש בכלי AI בהעברת בחינות, וממשק שלישי נוגע לשימוש ב-AI בהערכת תוצרי הבחינה. הממשקים יוצגו על ידי אנשים מגופים שונים, המייצגים גם אתגרים שונים הנובעים מהצרכים היחודיים של כל ארגון וכל משימה. ההרצאות יציגו תפיסות שונות לשימוש ב-AI ומגוון של פתרונות מבוססי-AI לבעיות שונות מעולם המדידה וההערכה.

ההרצאה הראשונה עוסקת בשימוש בבינה מלאכותית לפיתוח פריטים לבחינת ידע באנגלית. ההרצאה תתמקד באופן בו נבנה תהליך סדור לייצור פריטים, תוך שימוש במגוון אלגוריתמים של עיבוד שפה טבעית ובינה מלאכותית יוצרת.

ההרצאה השנייה עוסקת בשילוב בינה מלאכותית במהלך בחינת בגרות במדע הנתונים. ההרצאה תתמקד בניסוי פורץ דרך, בו תלמידים בבחינת הבגרות ישתמשו בכלי בינה מלאכותית על מנת לפתור בעיות מורכבות בנייתוח נתונים, בדומה לאנליסטים בעולם האמיתי.

ההרצאה השלישית עוסקת באתגרים והזדמנויות בהערכת הבעה בכתב באמצעות בינה מלאכותית. ההרצאה תתמקד באתגרים בהערכת כתיבה חופשית, בתהליך המרת המחונן האנושי למחונן מבוסס AI ובהשלכות האפשרויות של בינה מלאכותית על תפקיד המורה, המוטיבציה של התלמידים והאתיקה בשימוש בהערכה מבוססת מכונה.

בסיום נערוך דיון המתייחס לשאלות המרכזיות בנוגע לשילוב של AI ובחינות, כפי שעולה משלושת ההרצאות.



סימפוזיון

שימוש בבינה מלאכותית לפיתוח פריטים לבחינת ידע באנגלית

יוני שריג, רועי שילון, אפי לוי, יעל מיכאלי וצור קרליץ

מאל"ו

רקע

ההתקדמות באלגוריתמים של עיבוד שפה טבעית ובינה מלאכותית (AI) יצרה הזדמנות לייעול תהליכי פיתוח פריטים. הרצאה זו עוסקת בשילוב האלגוריתמים הללו בשיטות מסורתיות על מנת ליצור יותר פריטים בפחות זמן, מבלי להתפשר על איכות. השאלה המנחה היא: "כיצד נכון להשתמש בכלי AI לפיתוח פריטים?"

האתגר המרכזי בפיתוח והמענה עליו

במעבר לבחינות ממוחשבות מתמשכות יש צורך ביצירת מאגר פריטים גדול מאוד, אך בו בזמן, לשמור על איכות פריטים גבוהה, מגוון של רמות קושי, וגיוון בנושאי הפריטים ומאפיינים אחרים. שימוש בבינה מלאכותית יוצרת מאפשר בניית מאגר פריטים גדול מאוד יחסית בקלות. עם זאת, התוצר שמתקבל לרוב אינו מגוון מספיק ואינו עומד בסטנדרטים של איכות המקובלים במאל"ו. בהרצאה נציג קווים מנחים לבניית תהליך (pipeline) שנותן מענה לאתגרים השונים בפיתוח פריטי ברירה למבחן באנגלית.

התהליך המוצע

התהליך כולל שלב מקדים ומספר שלבי פיתוח. בשלב המקדים מתבצעת חקירה תוכנית של דוגמאות פריטים, ונכתב רציונל לפיתוח הפריט. הרציונל כולל את מבנה הפריט, היכולות הנמדדות והנחיות "עשה ואל-תעשה" לכל מרכיב בפריט. תהליך הפיתוח של הפריטים נבנה כך שלכל אחד משלושת המרכיבים בפריט (גזע, מפתח, מסיחים) ישנו שלב "יצירה" ושלב "בחירה וסינון".

א. יצירת מועמדים לגזע הפריט

ליצירת הגזע, בנינו סוכן GPT-4, שפועל כמומחה ליצירת פריטים מהסוג המבוקש. ניסחנו הנחיה (prompt) שתיצור מגוון גזעים לפי מתווה מוגדר (לדוגמה, אמירות עובדתיות ברמה אקדמית).

ב. סינון מועמדים לגזע הפריט

הגזעים צריכים להיות מגוונים, נכונים מבחינה דקדוקית וללא חזרות. שני מודלים לתיקון שגיאות דקדוק (Gector¹ ו-Gramformer²) שימשו לסינון גזעים שאינם דקדוקיים. מודל הטמעת משפטים (AnglE³), יחד עם מדד דמיון סמנטי שימשו לשמירה על שונות הנושאים.

ג. בחירת מפתחות מתוך גזעים

כדי לבחור מפתח טוב, הסתמכנו על המבנה התחבירי של הגזע (בעזרת המנתח של ברקלי - benepar⁴) וייצרנו מפתחות לפי מספר כללים מבוססי-תחביר. הכללים נבנו לאחר ניתוח של פריטים קיימים ועל בסיס שיקולים הקשורים לאופי הפריט.

ד. בחירה וסינון של מפתחות

המפתחות המתקבלים עוברים סינון נוסף בהתבסס על הרכב חלקי הדיבר שלהם (באמצעות מתייג חלקי דיבר - spaCy⁵).

1 <https://github.com/grammarly/gector>

2 <https://github.com/PrithivirajDamodaran/Gramformer>

3 <https://github.com/SeanLee97/AnglE>

4 <https://github.com/nikitakit/self-attentive-parser>

5 <https://spacy.io/usage/linguistic-features#pos-tagging>



ה. יצירת מסיחים

השתמשנו ב-GPT Assistant ליצירת אפשרויות להשלמת הגזע. יישמנו מספר אסטרטגיות הנחיה שנבדלו בעיקר בהקשר שניתן ל-Assistant (כמה מילים מתוך הגזע). התקבלו מעל 150 מסיחים אפשריים לכל מפתח.

ו. בחירה וסינון של מסיחים

הסתמכנו על שני קריטריונים: (א) שתילת המסיח בגזע אמורה לייצר משפט לא דקדוקי; (ב) המסיח חייב להיות נכון דקדוקית. לבדיקת הקריטריון הראשון השתמשנו במודלים לתיקון דקדוק שהוזכרו לעיל. לקריטריון השני, הנחינו את ChatGPT לייצר מספר משפטים נכונים דקדוקית המכילים את המסיח. המשפטים עברו לבדיקה במודלים לתיקון דקדוק. אם ChatGPT לא הצליח לייצר אף משפט תקין הכולל את המסיח, אז המסיח נחשב לא תקין מבחינה לשונית ולכן לא קביל לשימוש.

התהליך אפשר לנו ליצור בנק של פריטים באיכות גבוהה, אשר עברו בהמשך למסלול הגהה אנושית.



סימפוזיון

העתיד במבחן: הערכה מחדש של בגריות בעידן הבינה המלאכותית

רונית נחמיה ונעמה מרוז

משרד החינוך

אנו מציגים את הניסוי הראשון מסוגו בארץ, ואולי בעולם, המשלב כלי בינה מלאכותית בבחינת הבגרות. הניסוי יאפשר לתלמידים לפתור בעיות מורכבות בניתוח נתונים, בדומה לאנליסטים בעולם האמיתי. זוהי הבגרות של העתיד, המפתחת מיומנויות הנדרשות לעולם העבודה המשתנה.

רקע

מגמת "מידע ונתונים" מכשירה תלמידי תיכון להתמחות בניתוח נתונים, אחד התחומים המבוקשים ביותר בעידן הדיגיטלי. מקצוע דאטה אנליסט כולל איסוף, עיבוד וניתוח כמויות עצומות של מידע להפקת תובנות משמעותיות. התלמידים במגמה לומדים להשתמש בכלים מתקדמים ולבצע ניתוחים מורכבים, כשהשלב הבא הוא שילוב הבינה המלאכותית.

למה חשוב שהתלמידים ילמדו לשלב כלים של בינה מלאכותית?

בעולם התעסוקה המודרני, היכולת לעבוד עם AI היא קריטית. מעסיקים מחפשים אנשי מקצוע שיודעים לא רק לנתח נתונים, אלא גם לנצל את עוצמת הבינה המלאכותית לפתרון בעיות מורכבות בזמן אמת.

מטרה

מטרתנו היא לפתח הערכה חינוכית חדשנית המתמקדת ביכולת התלמידים לשלב בינה מלאכותית בניתוח נתונים. זו אינה רק בחינת ידע, אלא הערכה הבודקת כיצד תלמידים משתמשים בטכנולוגיות מתקדמות לפתרון אתגרים מורכבים.

אתגרים ודילמות

האתגר המרכזי הוא לקבוע מה בדיוק יש להעריך וכיצד לוודא שהתלמיד מבין את פעולותיו ולא רק מסתמך על תשובות ה-AI. האתגרים הטכניים כוללים: הגבלות בבחירת מודל השפה, סוגיות תקציב וגישה למודלים, ושמירה על טוהר הבחינות במערכת סגורה.

הזדמנויות ומענים

הניסוי מבוסס על אינטראקציה עם AI בתפקידים שונים – כלקוח וכמנתח נתונים. התלמידים נדרשים להפעיל את הכלים במסגרת סימולציות, לשאול שאלות, ולהפיק תובנות והמלצות. הבינה המלאכותית משמשת כשותפה אנליטית, כשהתלמיד מוביל את התהליך וקבלת החלטות.

מה בעתיד?

השנה, תלמידים מ-18 בתי ספר ישתתפו בניסוי, ובשנה הבאה יורחב לכלל תלמידי המגמה. אנו צופים שבעתיד הקרוב, כל בחינות הבגרות ישלבו אלמנטים של AI, ויבחנו לא רק ידע אלא גם חשיבה ולמידה. מחקר מלווה יבחן את השפעת השימוש בכלים אלה על יכולות התלמידים ויספק תובנות על השפעת ה-AI על החינוך.



סימפוזיון

מהאנושי למכונה: אתגרים והזדמנויות בהערכת הבעה בכתב באמצעות בינה מלאכותית

תמי סבג שושן, הדיל שחברי, הדס גלברט ובוועז רוזנבאום
ראמ"ה

תקציר

הערכה של כתיבה חופשית היא אתגר הן ברמה האנושית והן בהערכה מבוססת מכונה. בהרצאה זו נסקור תחילה את המורכבות, האתגרים וההזדמנויות בשילוב של טכנולוגיות בינה מלאכותית (AI) להערכת הבעה בכתב של תלמידים. נתאר את המבדק הטכנולוגי, נפרט את תהליך המרת המחונן האנושי למחונן מבוסס AI ונציג ממצאים מהשימוש בכלי הממוחשב ומתוך מחקר תוקף המשווה בין הערכה אנושית להערכת מכונה. לבסוף, נדון בהשפעות על היבטים פדגוגיים מנקודת מבט של המורה תוך הצעת תובנות לשימוש מושכל ב-AI בהערכת כתיבה חופשית.

האתגרים בהערכת כתיבה

יכולת הכתיבה היא מיומנות מרכזית הנחשבת לפסגת האוריינות שמשפיעה על תחומים רבים. תלמידים רבים מתקשים בפיתוח מיומנות זו. כדי לשפר את הכתיבה נדרש לזמן לתלמידים התנסויות מרובות ומשוב מקצועי הממוקד בשיפור מתמיד. עם זאת, הערכה של כתיבה דורשת ידע מקצועי נרחב, מצריכה זמן רב ולעיתים מעוררת תחושה של חוסר ביטחון מקצועי בקרב מורים ביחס ליכולתם להעריך ולקדם כתיבה בצורה מיטבית. לנוכח האתגרים הללו, נדרש פיתוח פתרונות חדשניים שיסייעו למורים להתמודד עם המורכבות תוך שמירה על תפיסות פדגוגיות חינוכיות.

המרת מחונן אנושי למחונן מבוסס בינה מלאכותית

עד כה, השימוש בבינה מלאכותית התמקד בעיקר בתהליכי הוראה ובהערכה של פריטים סגורים או משימות קצרות ומוגדרות. היכולת להעריך הבעה בכתב באמצעות בינה מלאכותית מסמלת פריצת דרך משמעותית. תהליך המרת המחונן האנושי למחונן מבוסס בינה מלאכותית מצריך הליך סדור ומוקפד הדורש התאמת המחונן לשפה טכנולוגית.

הכלי שפותח מבוסס על גישה פדגוגית-חינוכית המשלבת את יתרונות ההערכה האנושית עם אלו של ההערכה הממוחשבת. מצד אחד, הכלי מאפשר הערכה מהירה ומדויקת, ומצד שני, הוא מותיר למורה את שיקול הדעת לבחון ולאשר את ההערכה תוך התייחסות להקשרים תוכניים, להיבטים רגשיים ולמרכיבים דקדוקיים.

מנתוני השימוש עולה כי רוב המקרים המורים מאשרים את המשוב האוטומטי ללא שינויים מהותיים. ייתכן שהאישור נעשה מתוך הסכמה מלאה עם הבינה המלאכותית, ייתכן שלא נבדקו התוצרים באופן מעמיק, או שמא נובע הדבר מהיעדר ידע חלופי או רצון לערער על המשוב שהתקבל.

הממצאים של מחקרי תוקף בתוצרי כתיבה בעברית מצביעים על מתאם בינוני עד גבוה בין הערכות המערכת לאלו של מעריכים אנושיים, במיוחד בניתוח תוכן. עם זאת, קיימים עדיין אתגרים בולטים בהערכת היבטים תחביריים ודקדוקיים בשפה העברית המחייבים המשך פיתוח ושיפור.

השפעת הבינה המלאכותית בהערכה מנקודת מבט המורה והתלמידים

שילוב מערכות בינה מלאכותית בתהליך ההערכה מזמן למורים כלים חדשניים המפחיתים את עומס העבודה, משפרים את הדיוק ומפנים זמן לפעילויות משמעותיות כגון שיחות משוב אישיות ופיתוח תהליכי הוראה מותאמים. לצד ההזדמנויות הללו, יש לקחת בחשבון גם אתגרים פוטנציאליים. אף על פי שהדברים אינם מבוססים על נתונים מחקריים, אלא השערות, חשוב להדגיש חששות אפשריים כמו ירידה במעורבות האישית של המורה והיחלשות מיומנויות ההערכה העצמאיות. בנוסף, מערכות אלו עלולות להיות מוגבלות בזיהוי עומק רעיוני או ביטויים רגשיים – מרכיבים מרכזיים בתהליך הכתיבה.

סיכום

ההרצאה מדישה את השילוב הרגיש בין חדשנות טכנולוגית לערכים חינוכיים. כדי להבטיח שילוב מוצלח של AI בהערכה של הבעה בכתב, יש לאמץ גישה מאוזנת שמשכללת את תפקיד המורה ושמאפשרת לו להתמקד בפן האנושי והפדגוגי תוך ניצול היתרונות הטכנולוגיים של המערכת.



הבדלים מגדריים בביצוע מטלות כירורגיות בסימולטור לפרוסקופיה: השוואה בין מדדי ביצוע

גלעד בן גל ורון אלישר	נועה גזית
האוניברסיטה העברית	האוניברסיטה העברית
המרכז הרפואי האוניברסיטאי הדסה	מאל"ו

הבדלים בין גברים לנשים תועדו במגוון תחומים, כולל קוגניציה, אישיות והתנהגות, כאשר הן גורמים ביולוגיים והן גורמים סביבתיים נמצאו קשורים להבדלים אלו. אחד הממצאים הבולטים והעקביים ביותר הוא שגברים מבצעים טוב יותר מנשים מטלות המצריכות יכולות טכניות (כושר יד, קואורדינציה וראייה מרחבית). בהתחשב בכך שיכולות טכניות נחשבות חיוניות לביצוע פרוצדורות כירורגיות, לא מפתיע שהבדלים מגדריים דומים התקבלו גם במחקרים שבדקו ביצועיים כירורגיים בקרב מנתחים ומנתחות. עם זאת, "ביצוע כירורגי" (surgical performance) הוא מבנה תיאורטי מורכב, שיש לו הגדרות אופרציונליות אפשריות שונות (למשל מהירות ביצוע הפרוצדורה הכירורגית, מספר הטעויות בביצוע הפרוצדורה), שעשויות להוביל לתובנות שונות ביחס להבדלים מגדריים.

במחקר זה נערכה השוואה בין ההבדלים המגדריים המתקבלים במדדים שונים המשמשים להערכת ביצוע בסימולטורים כירורגיים. ממצאי המחקר מצביעים על כך שדפוס ההבדלים בין גברים לנשים תלוי בסוג המדד הנבחן, כאשר היתרון של גברים במטלות כירורגיות מתבטא רק במדדי ביצוע מסוימים. ממצאים אלו יכולים להוות בסיס לשיפור תהליכי ההערכה של מיומנויות כירורגיות ולפיתוח שיטות הדרכה מותאמות לאימון מיומנויות אלו תוך התחשבות בהבדלים מגדריים.



המחזור השני של סקר המיומנויות הבין-לאומי (PIAAC): טעימות מהמצאים, אתגרים והזדמנויות

נעה שחורי איל
ראמ"ה

חיים פורטנוי
למ"ס

בשנים 2022–2023 נערך ב-31 מדינות המחזור השני של סקר המיומנויות, סקר בין-לאומי השוואתי שמבצע ארגון OECD כחלק מתוכנית PIAAC (The Programme for the International Assessment of Adult Competencies). הסקר נועד לאמוד את המידה שבה אנשים בוגרים שולטים במיומנויות הנחשבות לרלוונטיות ונחוצות להשתלבות בעבודה ולתפקוד כאזרחים בחברה מודרנית. הסקר בודק את רמת השליטה של בוגרים (מגיל 16 עד 65) בשלוש מיומנויות יסוד: אוריינות קריאה (Literacy), אוריינות מתמטית (Numeracy) ופתרון בעיות אדפטיבי (Adaptive Problem Solving). המחקר כולל פרק מבחן ופרק שאלות רקע, מבנה המאפשר לבחון שאלות העוסקות במיומנויות עצמן (לדוגמה, כיצד הן מתפלגות בקרב בוגרים בהשוואה בין לאומית ובהשוואה למחזור הקודם) ושאלות הנוגעות בקשר בין מיומנויות לבין היבטים כלכליים-חברתיים שונים, להעריך את מערכות החינוך וההכשרה המקצועית ואת ההישגים הנובעים מהמדיניות החברתית ומהפרקטיקות הנהוגות במקומות עבודה בפיתוח המיומנויות הנדרשות ועוד.

בישראל השתתפו במחקר כ-6000 תושבי ישראל שמהווים מדגם מייצג של האוכלוסייה בגילי 16–65. שני פרקי סקר המיומנויות תורגמו לעברית ולערבית, והמחקר נערך באמצעות מחשב לוח (טבלט). כאמור, סקר המיומנויות הורכב משני חלקים. בחלק הראשון, שאלון הרקע, שאל הסוקר את המשיב על הרקע הדמוגרפי שלו (כגון גיל, מגדר, ארץ לידה, הרכב משק הבית); השכלה והשתתפות בפעילויות הכשרה מקצועית; מצב תעסוקתי והיסטוריה תעסוקתית; ושימוש במיומנויות שונות במקום העבודה. החלק השני היה מבחן מיומנויות אדפטיבי שבדק את רמת השליטה של המשיבים בשלושת התחומים שהוזכרו לעיל (כל משיב ענה על שני מקבצי יחידות בשניים מהתחומים). לכל משיב נזקפו ציונים במבחן המיומנויות (plausible values) בהתבסס על הפריטים שעליהם ענה, באמצעות המודל החד-פרמטרי של IRT.

במחקר נמצאה ישראל מתחת לממוצע OECD בכל שלושת התחומים שנבדקו. באוריינות קריאה חלה ירידה מובהקת של 10 נקודות ביחס למחזור הקודם, והפער מממוצע OECD התרחב. באוריינות מתמטית לא השתנה הציון של ישראל, אך הפער שלה מה-OECD מובהק. פתרון בעיות אדפטיבי הוא תחום חדש במחזור הנוכחי ולכן לא ניתן להשוותו למחזור הקודם, אך הישגי ישראל נמוכים ב-15 נקודות מממוצע OECD. ממצאים נוספים הם אחוז גבוה (למעלה משליש) של משיבים המאופיינים ביכולות נמוכות ופיזור ציונים גדול בכל תחומי המיומנויות. פערים בולטים נמצאו בין דוברי עברית ודוברי ערבית (ציון נמוך יותר ואחוז מתקשים גבוה יותר בקרב דוברי ערבית, בעיקר בקרב גברים) ובין חרדים ולא חרדים (הפער נובע ברובו מהישגים נמוכים יותר בקרב גברים חרדים בהשוואה ללא חרדים).

תוצאות המחקר מצביעות על פערים משמעותיים בין קבוצות אוכלוסייה שונות בישראל, בדומה למחקרי ידע ומיומנויות בינ"ל אחרים שנערכים במערכת החינוך דוגמת פיזה, טימס ופירלס. מאפיין ייחודי של סקר המיומנויות הוא עושר נתוני הרקע שיכולים להסביר את ההישגים, הן בפרק השאלון של המחקר עצמו והן במאגרי מידע נוספים שמקושרים לממצאי הסקר. לצד הממצאים החשובים ומשמעותיים לקובעי מדיניות, נשאלת שאלת הרלוונטיות של מבחן ההישגים נוכח ההתפתחויות הטכנולוגיות המהירות בשנים האחרונות; נטען שמיומנויות יסוד אלו – אוריינות קריאה, אוריינות מתמטית ופתרון בעיות אדפטיבי – יוסיפו לעמוד בבסיס היכולת של בוגרים לתפקד ולשגשג כאזרחים וכמשתתפים פעילים בעולם העבודה ובחברה במאה ה-21.



תפיסות של איכות חיים, שלומות וסקרנות בגיל השלישי

מנוחה בירנבוים, פאדיה נאסר אבו אלהיג'א
אוניברסיטת תל אביב

חני שלטון

האוניברסיטה הפתוחה

הלנה קימרון
מכון אבני ראשה

רובנה רוזנסקי

מכללת לוינסקי-וינגייט

ההרצאה תעסוק בממצאי מחקר שבחן את הקשרים בין תפיסות איכות חיים (QoL), שלומות (WB) וסקרנות בקרב מבוגרים בגילאים 75 ומעלה. המחקר התבסס על שיטות מחקר כמותיות ואיכותיות, והראה כי לסקרנות יש תפקיד משמעותי בתמיכה בהזדקנות אדפטיבית.

שאלות המחקר

1. באיזה מידה קשורים הממדים השונים של סקרנות לתפיסת איכות החיים בגיל השלישי?
2. באיזה מידה קיים קשר בין גורמים דמוגרפיים לבין תפיסות הסקרנות ואיכות החיים בקרב מבוגרים?
3. כיצד מבוגרים סקרנים מבטאים במילותיהם את תפיסותיהם אודות הסקרנות, אהבת הלמידה ואיכות החיים שלהם?

שיטה

המחקר כלל שני רבדים:

הרובד הכמותי: מדגם של 199 מבוגרים השיבו על שני שאלונים: איכות החיים - CASP-12¹ וסקרנות DCR²-5. השאלונים כללו גם פרטים דמוגרפיים כגון מגדר, גיל, מצב משפחתי, מקצוע (והתעניינות מתמשכת בו), עיסוקים נוכחיים, תחומי עניין ושפת אם. בוצע ניתוח גורמים מאשש (CFA) לתיקוף המבנים הנבדקים וכן ניתוחי רגרסיה מרובה לבדיקת הקשרים בין ממדי הסקרנות, לבין ממדי איכות החיים והמשתנים הדמוגרפיים.

הרובד האיכותני: נערכו ראיונות עומק (חצי-מובנים) עם 10% מהמשיבים על השאלון - כולם מבוגרים סקרנים, בהם הם הביעו את תפיסותיהם לגבי סקרנות, אהבת הלמידה ותפיסת איכות החיים. מרבית המראיינים ידועים בזכות תרומתם לתחומים, כמו: אומנות, מדע, רפואה, ארכיטקטורה, קולינריה ועוד. הראיונות התבצעו והוקלטו באמצעות Zoom. בוצע ניתוח תוכן לקידוד המידע האיכותני.

ממצאים

- הממצאים מעידים על מאפיינים משביעי רצון של המדידה.
- בניתוח רגרסיה מרובה נמצאו קשרים מובהקים בין ממדי איכות החיים CASP-12 לבין ממדי הסקרנות. הממדים בעלי משקל המשמעותי ביותר היו סובלנות ללחץ או לאי וודאות (Stress tolerance) והנאה מגילוי (Joyous exploration).
- ניתוח נתיבים לבדיקת ההשפעות הישירות והעקיפות (דרך סקרנות) על איכות החיים הצביע על כך שסקרנות מהווה גורם מתווך בין התעניינות מתמשכת במקצוע לבין איכות החיים (מצ"ב תרשים המודל).
- הממצאים מהרובד האיכותני העלו התנהגויות המונעות על-ידי הסקרנות וכן את התחושות הנלוות לכך. נמצא שסקרנות וחדוות למידה בגיל השלישי מובילות לשביעות רצון גבוהה מהחיים, תחושת משמעות ותחושת הישג. המשתתפים הדגישו את החשיבות של התמודדות עם אתגרים והתייחסות חיובית לאי וודאות לצד תחושות שליטה עצמית ואוטונומיה לחקור ולהתפתח בתחומים חדשים.
- באופן זה הממצאים ברובד האיכותני המחישו את הממצאים מהרובד הכמותי, סיפקו מגוון הקשרים בהם מתבטא קשר בין סקרנות ואיכות החיים וכן הסברים לקשרים אלה.

1 CASP - Control, Autonomy, Self-realization, Pleasure (Hyde, Wiggins, Higgs & Blane, 2003)

2 5-DCR - Joyous exploration, Deprivation sensitivity, Stress tolerance, Social curiosity, Thrill seeking (Kashdan, et al., 2018).



מסקנות וחשיבות המחקר

המחקר הנוכחי מראה כיצד סקרנות וחדוות למידה משפיעות על הזדקנות אדפטיבית בקרב בני הגיל השלישי. סקרנות מניעה את הלמידה לאורך החיים. אלה שבאופן עקבי ממשיכים ללמוד וחווים הנאה מחקירה וגילויים, חווים תחושות חיוביות, תחושות של מימוש עצמי, משמעות והישג – תפיסות חיוביות של איכות החיים ושלמות.

ההשלכות של המחקר מכוונות לתוכניות התערבות בקרב אוכלוסייה של הגיל השלישי. למשל יש מקום להציע לבני הגיל השלישי להציג ולבטא את ההזדקנות האדפטיבית שלהם הנובעת מסקרנות שמובילה לחדוות למידה וגילוי המלווה בתחושות של מימוש עצמי, משמעות והישג. זה עשוי לעודד את המבוגרים אשר אינם חווים חדוות למידה וכן את המורים שאמורים לטפח אותה בקרב תלמידיהם.

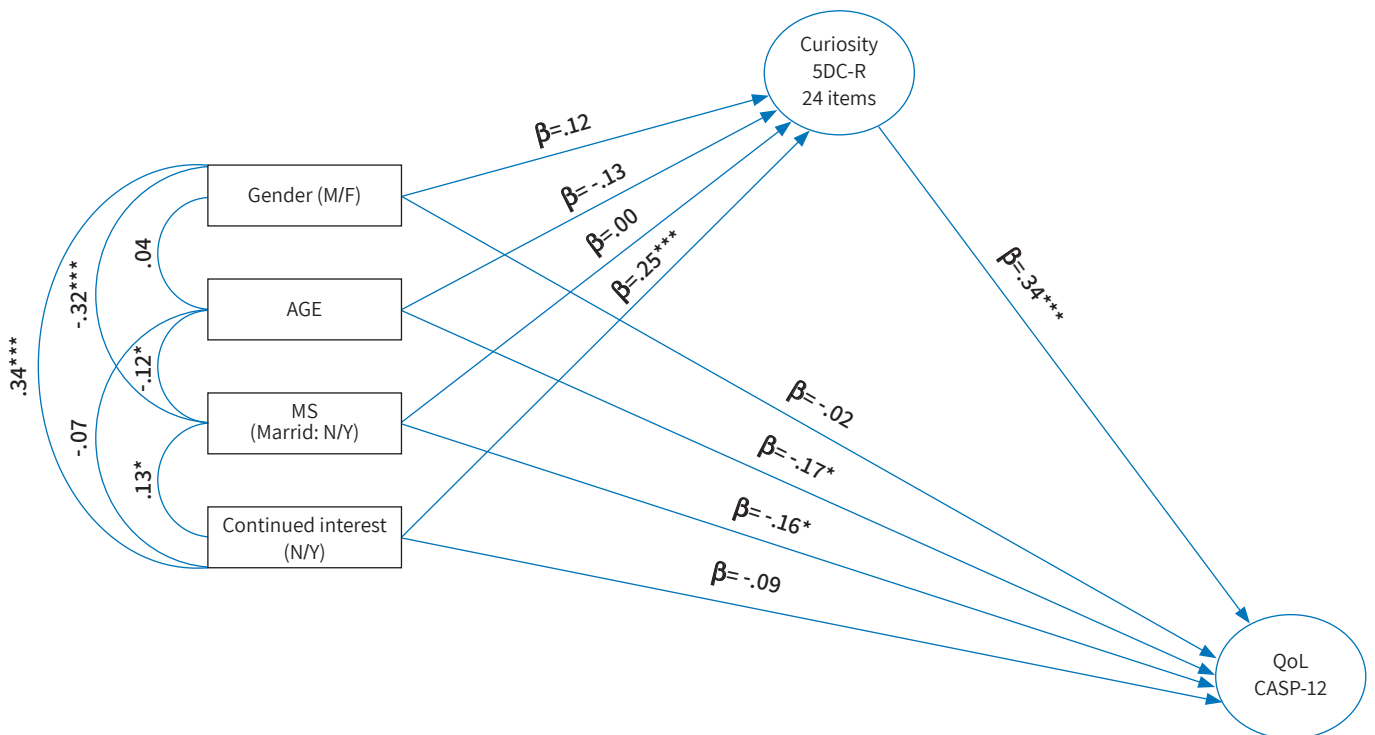


Figure: Path diagram – The mediating role of curiosity in predicting QoL



רעיון בפיתוח

שאלון התאמת קריירה בעידן ה-Big Data: חדשנות בייעוץ תעסוקתי

רחל חביב וויטמן
אוניברסיטת אריאל
פרדי אדיב
מכללה אקדמית הדסה

בעידן הנוכחי חווה עולם העבודה שינויים מואצים ביניהם: כניסת טכנולוגיות חדשות, העלמות מקצועות מסורתיים והופעה של מקצועות חדשים. תופעות אלו מציבות אתגר מיוחד בפני הפסיכולוגיה התעסוקתית במתן התאמה מדויקת, עדכנית ומותאמת אישית למתלבטים בבחירת קריירה.

בהקשר זה, פותח השאלון הנוכחי –שאלון התאמת קריירה. השאלון מבוסס על ניתוח נתוני עתק (Big Data), המשלב דיווח עצמי עם אלגוריתמים מתקדמים לניתוח והתאמה ומטרתו מתן הכוונה מדויקת למתלבטים בבחירת קריירה על מקצועות בעלי רמת התאמה גבוהה.

השאלון מתבסס על נתוני O*NET, אחד ממאגרי המידע המקיפים בעולם בתחום התעסוקה, המכיל מידע עדכני על מגוון רחב של מקצועות, דרישותיהם, הכישורים הנדרשים והמאפיינים האישיותיים המתאימים לכל תפקיד. החישוב של מענה המשתמש נעשה באמצעות ניתוח אלגוריתמי מורכב המשווה את המענה עם נתוני מאגר המידע. באמצעות שילוב זה, השאלון מאפשר למשתמשים לקבל המלצות מותאמות אישית על מסלולי קריירה פוטנציאליים, בהתבסס על נטיותיהם, כישוריהם והעדפותיהם האישיות.

השאלון כולל מספר שלבים:

1. דיווח עצמי: המשתמשים מתבקשים לספק מידע על נטיותיהם התעסוקתיות, תחומי העניין, הכישורים והעדפות העבודה שלהם באמצעות סולם דרוג פשוט.
 2. ניתוח אלגוריתמי: התשובות מעובדות באמצעות אלגוריתמים מתקדמים, המשווים את הנתונים האישיים למידע הקיים ב-O*NET, ובוחנים את מידת ההתאמה למגוון מקצועות.
 3. הצגת תוצאות: המשתמשים מקבלים דוח מפורט בשפה ברורה, המציג את המקצועות המתאימים ביותר עבורם, כולל מידת התאמה כמו גם מידע על דרישות התפקיד, הכישורים הנדרשים והזדמנויות פיתוח קריירה.
- השאלון מתעדכן באופן שוטף, תוך התחשבות בשינויים בשוק העבודה. בנוסף, המשתמשים יכולים לבחור את היקף המענה, בהתאם לצרכיהם האישיים, ולקבל תוצאות מהירות וכלליות או מעמיקות ומדויקות יותר. כפיתוח עתידי מתוכנן שימוש בנתוני משוב מיידים של המשתמשים על התוצאות שקיבלו בלמידת מכונה להעלאת דיוק והתאמת תוצאות השאלון.
- בכנס נציג את הרציונל שמאחורי פיתוח השאלון, את המתודולוגיה הייחודית לניתוח הנתונים, דוגמאות מעשיות לשימוש בו, והאופן שבו השאלון תורם לשיפור הייעוץ התעסוקתי בעידן המודרני. הכלי מהווה פריצת דרך בתחום הפסיכומטריקה, ומציע אמצעי מתקדם ומדויק לסיוע למתלבטים בבחירת מסלול הקריירה המתאים ביותר עבורם.



הישגים באוריינות קריאה בכיתה ד': קשרים בין הישגים למשתנים אישיים ובית ספריים

אליענה אביצור, ירון שלומי, פרנקלין זרומב, נורית ליפשוטט ויוסי מחלוף
ראמ"ה

מחקר פירלס (Progress in International Reading Literacy Study – PIRLS) הוא מחקר בין-לאומי מרכזי, שמטרתו להעריך את אוריינות הקריאה של תלמידי כיתות ד' ברחבי העולם. המחקר, הנערך על ידי הארגון הבין-לאומי להערכת הישגים בחינוך (IEA), מבוצע במחזוריות של אחת לחמש שנים מאז 2001. מחזור פירלס 2021 בישראל כלל כ-5,000 תלמידים מכ-200 בתי ספר, המהווים מדגם ארצי מייצג. המחקר כלל שני סוגי טקסטים עיקריים, סיפורי ומידעי, בעברית או בערבית, בהתאם לשפת ההוראה בבית הספר. המחקר כלל גם שאלוני רקע ועמדות שמולאו על ידי התלמידים, הוריהם, מוריהם לשפה ומנהלי בתי הספר.

במחקר הנוכחי נבדקו קשרים אפשריים בין מגוון משתנים אישיים ובית ספריים לבין הישגים באוריינות קריאה, וזאת באמצעות מודלים של גרסיה ליניארית היררכית.

המשתנה המוסבר היה הישגי התלמיד באוריינות קריאה, שנמדדו באמצעות חמישה ערכים סבירים (Plausible Values). כל מודל נותח חמש פעמים, כל פעם בהתבסס על אחד מן הערכים הסבירים, ותוצאות הניתוחים שוקללו לכדי תוצאה אחת. המשתנים המסבירים הוכנסו למודל בשני שלבים: תחילה נבנה מודל בסיס שכלל משתני רקע (מגזר שפה, רקע חברתי-כלכלי ומגדר). לאחר מכן נוספו למודל הבסיס שש קבוצות משתנים: (1) עמדות התלמידים כלפי קריאה, (2) הכשרת המורים, (3) עמדות המורים, (4) דיווחי מנהלים על האקלים הבית-ספרי, (5) הסביבה הביתית וגיל הכניסה למסגרות לגיל הרך, ו-(6) עמדות ההורים כלפי קריאה. בדקנו אם המשתנים בכל קבוצה קשורים להישגים בקריאה, ובאיזו מידה הם תורמים להסבר השונות בהישגים מעבר לתרומה של משתני הרקע שנכללים במודל הבסיס (שונות מוסברת תוספתית). ככל שהשונות המוסברת התוספתית גדולה יותר, כך המשתנים שהתווספו למודל מרכזיים יותר וקשורים יותר להישגים בקריאה.

ממצאי המחקר הראו כי למשתני הסביבה הביתית וגיל הכניסה למסגרות לגיל הרך, עמדות ההורים כלפי קריאה, והכניסה למסגרות חינוך בגיל הרך קשר חזק ומשמעותי להישגי הקריאה, תוך בקרה על משתני הרקע. מודל הסביבה הביתית והכניסה למסגרות לגיל הרך הסביר את אחוז השונות התוספתית הגבוה ביותר (37%), בעיקר בזכות משתנים כמו יכולת לבצע פעילויות אורייניות עם כניסת התלמיד לכיתה א' וגיל הכניסה לגני ילדים. גם תדירות העיסוק בפעילויות אורייניות בבית נמצאה כמנבא מובהק לאוריינות הקריאה בכיתה ד'.

המודלים שהכילו משתנים הנוגעים לעמדות ההורים והתלמידים הסבירו גם הם אחוז ניכר של שונות תוספתית (36% ו-31%, בהתאמה). במודלים אלו המנבאים החזקים ביותר היו ציפיות ההורים בנוגע להשכלת ילדם, אהבת הקריאה בקרב ההורים וביטחון ביכולת לקרוא בקרב התלמידים.

המודלים של עמדות מורים ושל השכלת המורים הסבירו אחוז ניכר מהשונות בהישגים (24% ו-18% בהתאמה) אך המשתנה היחיד שהשפעתו היתה מובהקת היה תפיסת המורים את מידת המוכנות של התלמידים ללמידה כחסם להוראה איכותית של השפה. האפקט היה חזק יותר בקרב דוברי הערבית.

דיווחי המנהלים על האקלים הבית-ספרי הסבירו את אחוז השונות התוספתית הנמוכה ביותר (16%), והמשתנה היחיד שהיה מובהק במודל זה הוא תחושת מוגנות ומשמעת בבית-הספר.

מחקר זה פותח צוהר למשתנים המנבאים הישגים באוריינות קריאה בקרב תלמידי כיתות ד', ומצביע על כיוונים לתוכניות התערבות לשם שיפור ומיצוי יכולות התלמידים בתחום מרכזי זה.



לקראת הערכה מעצבת אוטומטית של מורים במסגרת סימולציה: הערכה של קוהרנטיות הדין בעזרת בינה מלאכותית

ג'ימי מיקסקה
ETS Research Institute

ביאטה בייגמן קלבנוב
ETS Research Institute
Edusoft

יויא אסאנו
אוניברסיטת פיטסבורג

השלב הראשון של העבודה שעליה אנו בונים התרחש לפני כ-10 שנים במסגרת של פיתוח מבחן תצפיתי למועמדים להוראה (National Observational Teacher Examination (NOTE) ב-ETS. המבחן נועד למדוד לא רק ידע שביסוד הפרקטיקה, אלא גם למדוד את איכות הפרקטיקה עצמה, תוך התמקדות ביכולת של המועמד להוביל דיון אפקטיבי בכיתה. כדי להשיג מטרה זו, המבחן השתמש בכיתה מדומה עם חמישה אוטורים של תלמידים שהופעלו על ידי מומחה סימולציה מיומן ומוסמך; האוטורים שוחחו בזמן אמת עם הנבחן (האנושי) בעזרת מודולצית קול דיגיטלית שאפשרה למומחה הסימולציה לדבר בקול שונה עבור כל תלמיד. לכל אוטור היה פרופיל ידע ואישיות סטנדרטי והוא (או היא) השתתפו בדיון בהתאם לפרופיל. בנוסף, פותח מחוון רב-ממדי כדי להעריך את ביצועי הנבחנים. המבחן החדשני לא זכה להתעניינות מספקת של הלקוחות בשל העלות הגבוהה שנבעה מהזמן שהושקע בכל נבחן על ידי מומחה הסימולציה וגם על ידי הבודק. מאז נמשכה העבודה במסגרת של מחקר על הערכה מעצבת למורים ולתלמידי הוראה.

ההתקדמויות הטכנולוגיות הגדולות של העשור האחרון עשויות להוריד עלויות של מבחן סימולציה באופן משמעותי. מחקרים חדשים שמשמשים בבינה מלאכותית יוצרת כדי לדמות בני אדם בתפקידים שונים, כולל תלמידים, מראים תוצאות מבטיחות. טכנולוגיה מבוססת בינה מלאכותית מהווה גם בסיס מבטיח לייעול הערכה של ביצועי המורה בסימולציה. בעבודה הקודמת שלנו, הראינו שלמידת מכונה המבוססת על כ-150 דיונים עם הערכה אנושית מפורטת יכולה להניב הערכה אוטומטית מדויקת למדי עבור שנים מחמשת הממדים של המחוון שנחקרו עד כה.

בעבודה הנוכחית, אנחנו מתמקדים בממד שלישי – זה של קוהרנטיות הדין, כלומר, מיומנות המורה בניית דיון כדי, מצד אחד, לתת לתלמידים הזדמנות לבטא ולשתף את החשיבה שלהם ולתת את דעתם על רעיונות של תלמידים אחרים, ומצד שני, למקד את הדיון לעבר מטרת הלמידה הספציפית שנקבעה לשיעור הזה. בנוסף לממד החדש, אנו גם בוחנים את האפשרות של הערכה אוטומטית ללא אימון על דיונים קיימים עם ציונים אנושיים תוך שימוש בלמידת zero shot.

התוצאות שלנו מראות שהערכה כזו אפשרית על ידי בניית הוראות (prompt engineering) בסגנון שרשרת מחשבה (chain-of-thought) המתמקדת בהיבטים ספציפיים של קוהרנטיות, מבקשת ציטוטים מהדיון כדי לתמוך בהערכה של כל היבט ומיישמת אסטרטגיית השוואה לדיון אחר (במקום מתן ציון באופן ישיר). עם זאת, ישנן מגבלות. בפרט, כאשר משתמשים בתבנית ההוראות (prompt blueprint) שפותחה במסגרת משימת דיון אחת למשימת דיון אחרת, אנו רואים ביצועים נמוכים משמעותית. לפיכך, פיתוח תבנית הוראות כללית שתכסה דיונים בנושאים שונים בדומה למחווון האנושי דורש מחקר נוסף.



בחינת התוקף של הליך האיתור של תלמידים מחוננים בישראל באמצעות מדדי הישג לימודיים

דוד מעגן יואל רפ ענבל שני
למ"ס מאל"ו אוניברסיטת חיפה

איתור מחוננים ברמה ארצית הוא הליך מורכב המתבצע מדי שנה במערכת החינוך הישראלית לתלמידים צעירים, ומטרתו לאתר תלמידים בעלי יכולות אינטלקטואליות גבוהות, כישרונות ייחודיים או פוטנציאל יוצא דופן.

האיתור נעשה באמצעות תהליך דו-שלבי; בשלב א' מועברים מבחנים של הישגים לימודיים לכלל תלמידי השכבה, ובשלב ב' מועברים מבחני יכולת קוגניטיביים לבעלי ההישגים הגבוהים של שלב האיתור הראשון. הנבחנים בעלי ההישגים הגבוהים ביותר מומלצים לתוכניות מחוננים ובעלי ההישגים המעט פחות גבוהים מומלצים לתוכניות מצטיינים. מטרתו העיקרית של הליך האיתור התלמידים המחוננים היא לספק לתלמידים בעלי יכולות אינטלקטואליות גבוהות הזדמנויות חינוכיות מותאמות שיאפשרו להם למצות את מלוא הפוטנציאל שלהם באמצעות מתן מענה לצורכיהם האקדמיים, הרגשיים והחברתיים הייחודיים. ישנה גם הסכמה רחבה כי איתור מחוננים הוא בעל ערך רב לחברה ולמדינה בכך שהוא מאפשר לטפח את ההון האנושי העתידי של מדינת ישראל, ובכך לתרום לחוסן המדעי, התרבותי והכלכלי שלה.

המחקר הנוכחי מבקש לבחון את תוקפו של הליך האיתור התלמידים המחוננים בישראל באמצעות ניתוח הקשר בין ההישגים במבחני שלב ב' של הליך האיתור בגיל הצעיר, לבין ההישגים הלימודיים העתידיים במבחנים מערכתיים סטנדרטיים הנהוגים במערכת החינוך. שאלות המחקר מתמקדות בשתי נקודות מבט משלימות: מהו הקשר הפרוספקטיבי שבין איתור מוקדם כמחונן ובין ההישגים הלימודיים בהמשך התהליך החינוכי בשנות בית הספר, ומהו הקשר הרטרופקטיבי. כלומר, האם התלמידים בעלי ההישגים הגבוהים ביותר במדדים המערכתיים הסטנדרטיים – אותרו כמחוננים בילדותם.

באמצעות מעקב אורך רב-שנתי, הוצמדו לנתוני הנבחנים בשלב ב', הישגיהם במבחני המיצ"ב בכיתות ה' ו-ח', הישגיהם במבחן פיזה הבינלאומי בגיל 15 (כיתות ט'-י'), והישגיהם במבחני הבגרות ובבחינה הפסיכומטרית. בכל אחת ממערכות הבחינה הללו נקבע לכל נבחן האם היה באחוזון 95 ומעלה (חמשת האחוזים העליונים בהתפלגות, להלן העשרימון העליון) בהתפלגות ההישגים. מהכיוון הרטרופקטיבי נבדק האם התלמידים הנמצאים בעשרימון העליון במבחני המיצ"ב, במבחן הפיזה, במבחני הבגרות ובבחינה הפסיכומטרית – נבחנו בשלב ב' של הליך האיתור המחוננים בצעירותם, ואם כן – האם הומלצו לתוכניות המחוננים (קרי, אותרו כמחוננים) או המצטיינים, או שכלל לא הומלצו.

נמצא כי בקרב נבחני שלב ב' שהומלצו לתוכניות המחוננים בילדותם, קיים שיעור נמוך, יחסית למצופה, בחמשת האחוזים העליונים במבחנים הללו. בקרב מומלצים לתוכניות למצטיינים, השיעור, כצפוי, נמוך יותר, ובקרב נבחני שלב ב' שלא הומלצו כלל – השיעור הוא הנמוך ביותר.

מנקודת המבט הרטרופקטיבית, נמצא כי בקרב התלמידים בעלי ההישגים הגבוהים ביותר במבחנים הסטנדרטיים המערכתיים (הנבחנים הנמצאים בחמשת האחוזים העליונים) – קיים שיעור גבוה יחסית שכלל לא נבחן בשלב ב' לאיתור מחוננים, או נבחן ולא הומלץ לתוכניות המחוננים או המצטיינים, כאשר רק שיעור נמוך מקרב בעלי ההישגים הגבוהים ביותר, הומלץ לתוכניות המחוננים או המצטיינים.

ממצאים אלו מסייעים בהבנת הקשר בין איתור מוקדם של מחוננים להצלחה לימודית לאורך זמן – שאלה תיאורטית חשובה בעולם המחקר של איתור, טיפוח ומימוש הפוטנציאל של מחוננים. כמו כן, הממצאים שופכים לראשונה אור על יעילות ותוקפו של הליך האיתור המחוננים במתכונת המתקיימת בישראל מזה מספר עשורים, ומעלים שאלות והסברים אפשריים שונים.



רעיון בפיתוח הליך הייעוץ הפסיכומטרי למערך בחינות הרישוי למורי דרך - אתגרים, הישגים ולקחים

דוד ציגלר
מאל"ו

קרן גולדמן ורם נעמן
משרד התיירות

הוראת דרך היא מקצוע הכפוף למערך רישוי ממשלתי, והעוסקים בו נדרשים לרישיון והסמכה מטעם משרד התיירות. בכל שנה לומדים כמה מאות תלמידים בקורסי הכשרה למורי דרך, המועברים כיום בשבעה מוסדות. ההכשרה נמשכת שנה וכוללת 300 שעות לימוד פרונטלי וכ-60 ימי סיור ברחבי הארץ. נושאי ההכשרה בתכנית מגוונים, וכוללים היסטוריה, ארכיאולוגיה, גיאוגרפיה, אדריכלות, דתות, אקולוגיה, תרבות ואמנות, דידיקטיקה ומתודולוגיה.

בשנת 2019 הוחלט לעדכן את מבנה ההכשרה ותכניה, ובין השאר למקדה במיומנויות ההדרכה של מורי הדרך לעתיד כדי שהתכנים והידע יועברו למודרכים באופן מעניין, יצירתי וחוויתי. השינוי בהכשרה לווה גם בשינויים בהליך הרישוי, ובמרכזם הפיכת שלב בחינת הרישוי המעשי להערכת ביצוע – "Performance Assessment" – במהלכה הנבחן מדריך סיור בשטח אל מול בוחנים ממשרד התיירות, ולצד הערכת הידע מוערך השימוש בפועל במיומנויות ההדרכה של מורה הדרך שנלמדו בתכנית ההכשרה. הרציונל לשינוי היה שצפייה בהדרכה בפועל תאפשר למעריכים לאמוד באופן ישיר ידע, כישורים ומיומנות מעשיות בתחום ההדרכה, ושאומדן זה יהיה מנבא מהימן ותקף יותר ליכולתו של המועמד להצליח בעבודתו כמורה דרך.

הפיתוח של בחינות הרישוי, העברתן, גיוס המעריכים והערכת תוצרי הבחינה והנבחנים מתבצעים על ידי צוות עובדים מצומצם ממשרד התיירות, שאין לו רקע והכשרה בתחום המדידה וההערכה.

בשנת 2021, פנה משרד התיירות למאל"ו, המרכז הארצי לבחינות ולהערכה, לנוכח הניסיון שצבר מאל"ו בהערכת ביצוע (למשל במבחני המיון לבתי הספר לרפואה) בבקשה לקבל ליווי מקצועי בביסוס תהליך הרישוי החדש. הליווי המקצועי כלל מיסוד הליכי פיתוח כלי המדידה ותפעולם, סיוע בגיבוש מדיניות כללית ופתרונות פרטניים למגוון בעיות פסיכומטריות מעשיות בהליך, ומענה לביקורת מצד נבחנים, מוסדות הכשרה ובעלי עניין נוספים, בקרה על איכות הליך הרישוי, שיפור וקיבועו כהליך הוגן, מהימן ותקף.

בהרצאה יוצגו הליך הרישוי וכלי ההערכה שבהם נעשה שימוש. אחר כך יתואר תהליך הייעוץ שנעשה על ידי מאל"ו ועיקרי הצעדים המעשיים שננקטו בעקבות המלצותיו, וכן נתונים סטטיסטיים על מהימנות כלי ההערכה שבהם נעשה שימוש בבחינות הרישוי. לסיכום יוצגו תובנות, קשיים והתלבטויות שעלו במהלך הייעוץ והשינוי במערך בחינות רישוי פעיל.



שולחן עגול

פרס אפי – היבטים פסיכומטריים, חברתיים ותהליך בחירת הזוכה

איתמר גתי

האוניברסיטה העברית בירושלים

פרס אפי נועד לעודד מחקר בתחום המדידה וההערכה ובפרט הפסיכומטריקה בקרב סטודנטים לתארים מתקדמים. לאחרונה, נעשו שינויים מבורכים בנוהל הגשת המועמדות וההחלטה על הענקת הפרס. עם זאת, עדיין עולות שאלות ביחס לכללי השיפוט והענקת הפרס, ולכן נראה שיש מקום לדון בהם במסגרת "שולחן עגול" בכנס אפי.

הדיון יתייחס להיבטים הפסיכומטריים (מה הם הקריטריונים של מחקר פסיכומטרי מצטיין), להיבטים האקדמיים והחברתיים (גובה הפרס, אפשרות להגשה חוזרת) ולייעול תהליך השיפוט וקבלת ההחלטה (מי הם השופטים, הבחנה בין הערכת איכות העבודה לבין תהליך בחירת הזוכה). משתתפי השולחן העגול מוזמנים להציג נושאים נוספים לדיון.

היבטים פסיכומטריים

- מהם הנתונים עליהם מבססים את ההחלטה: הגשה אנונימית או גלויה?
- כיצד מעריכים מהו מחקר ראוי לפרס בתחום הפסיכומטרי (מהם הקריטריונים), ואיך מעריכים עד כמה המחקר המועמד לפרס עונה על קריטריונים אלה? מהם המאפיינים של מחקר מצטיין?
- האם הקריטריונים הם זהים או שונים בין עבודת גמר ודוקטורט?
- האם יש להתחשב במידע נוסף, מעבר לעבודת המחקר שהוגשה לפרס? (למשל, הציון של העבודה, חוות דעת של השופטים על העבודה במסגרת מוסד הלימודים בה הוגש, הצגה קודמת בכנס אפי או בכנס בחו"ל, וכמובן פרסום בכתבי עת).

היבטים אקדמיים וחברתיים

- האם גובה הפרס על עבודת גמר ועבודת דוקטורט צריך להיות זהה?
- האם להתמקד רק בסוג מסוים של עבודה בכל שנה (גמר או דוקטורט), או לפתוח בכל שנה ערוצי הגשה מקבילים אך נפרדים הן לעבודות דוקטורט והן לעבודות גמר?
- האם יש מקום להעניק פרס גם אם הוגשה רק עבודה אחת?
- האם ניתן להגיש לפרס בשנית, מחקר שהוגש לפרס בעבר אך לא זכה?
- מהו משך הזמן שיש להקציב בין אישור עבודת הגמר או הדוקטורט לבין המועד בו ניתן להגיש מועמדות?
- האם כדאי להוסיף משוב על העבודה למועמד שלא זכה בפרס? למשל, "עבודתך הוערכה כטובה מאוד".
- האם יש מקום להעדפה מתקנת?

היבטים של תהליך השיפוט

- האם יש להפריד את תהליך הערכת איכות העבודות מתהליך קבלת ההחלטות לגבי הזוכה/הזוכים?
 - האם יש לחלק את תהליך הערכת האיכות לשלב של סינון ראשוני ושלב של הערכת איכות על פי הקריטריונים והאינדיקטורים של איכות?
 - האם לפרק את תהליך קבלת ההחלטה לשלבים: ראשית, דירוג העבודות באופן בלתי תלוי על ידי חברי וועדת הפרס על בסיס הערכות האיכות, ושנית לערוך דיון ולקבל החלטה על בסיס סיכום הדירוגים?
 - מי הם השופטים בכל שלב?
- הנושא יעניין חלק מחברי אפי, חברי ועדת הפרס בהווה ובעבר, מועמדים שקיבלו ושלא קיבלו את הפרס, סטודנטים בשלבי ההכנה של מחקרם וכן מנחים של הסטודנטים העורכים מחקר בתחום.

אופי הדיון: פתיחה כללית לנושא ולתהליך הדיון, לאחריו יוצגו הנושאים. לגבי כל נושא יתואר המצב הקיים, חלופות אפשריות, דיון ביתרונות ובחסרונות של כל חלופה, סיכום לגבי החלופות הראויות ליישום פוטנציאלי, ומעבר לנושא הבא.



אנשי קשר ומציגים בכינוס אפי 2025

זוא"ל	שם	
ritasever9@gmail.com	ריטה סבר	יו"ר אפי
hagitglickman@gmail.com	חגית גליקמן	יו"ר הכינוס
yossy.machluf@gmail.com	יוסי מחלוף	
liadmu@gmail.com	ליעד מודריק	מרצה אורחת
meirav.attali@gmail.com	יו"ר - מירב אריאלי-אטאלי	מושב א' בינה מלאכותית מול בינה אנושית - האם המבחנים מבחינים?
elad.yacobson@weizmann.ac.il	אלעד יעקבסון	
alona.faktor@weizmann.ac.il	אלונה סטרוגצקי	
tzur@nite.org.il	מנחה - צור קרליץ	מושב ב' סימפוזיון: ממשקים בין בחינות ובינה מלאכותית: אתגרים והזדמנויות
joni@nite.org.il	יוני שריג	
ronitne@education.gov.il	רונית נחמיה	
tamisa@education.gov.il	תמי סבג שושן	
finesaul@gmail.com	יו"ר - סול פיין	מושב ב' מדידת ביצועים, מיומנויות, תפיסות והתאמת קריירה בקרב מבוגרים
gazit.noa@mail.huji.ac.il	נועה גזית	
noash@education.gov.il	נעה שחורי איל	
helenak@avneyrosha.org.il	הלנה קימרון	
havivrach@gmail.com	רחל חביב וויטמן	
einat.notea@gmail.com	יו"ר - עינת נוטע קורן	מושב ג' תוקף ניבוי וקשרים - מדידה בקרב תלמידי בית ספר ומורי דרך לעתיד
elianaav@education.gov.il	אליענה אביצור	
bbeigmanklebanov@ets.org	ביאטה בייגמן קלבנוב	
maagan@cbs.gov.il	דוד מעגן	
yoel@nite.org.il	יואל רפ	
sz.inbal@gmail.com	ענבל שני	
david@nite.org.il	דוד ציגלר	
itamar.gati@mail.huji.ac.il	מנחה - איתמר גתי	מושב ג' שולחן עגול: פרס אפי - היבטים פסיכומטריים, חברתיים ותהליך בחירת הזוכה
ritasever9@gmail.com	מארגנת הכינוס - ריטה סבר	
meital@nite.org.il	פרוייקטורית הכינוס - מיטל אילון	







תגובות והצעות
בנושא האגודה הישראלית לפסיכומטריקה

נא להפנות אל:

ד"ר ריטה סבר, יו"ר אפי | ispaisrael@gmail.com

אתר אפי | www.ispa.org.il

עיצוב ועריכה גרפית | יעל פרץ | מאל"ו