

הערכה בסיוע בינה מלאכותית: נקודות מפגש ופערים בין שיקול הדעת של המרצים לבין ChatGPT בעבודות גמר אקדמיות (פוסטר)

מרסלו דורפסמן

האוניברסיטה העברית בירושלים
marcelo.dorfsman@mail.huji.ac.il

AI-Assisted Evaluation: Convergences and Divergences between Teacher Judgment and ChatGPT in University Final Works (Poster)

Marcelo I. Dorfsman

The Hebrew University of Jerusalem
marcelo.dorfsman@mail.huji.ac.il

Abstract

Background: The emergence of large language models and generative artificial intelligence (GAI) tools has rapidly transformed educational practices. Their use remains an uncertain field, full of risks but also opportunities. **Objective:** this ongoing research provides a comparative analysis between assessments carried out by a university instructor and those generated by ChatGPT on twenty final projects within a graduate-level education program course. **Methodology:** We employ an interpretive multiple-case study design to analyze the convergence and divergence between human and ChatGPT-based evaluations of identical academic products. The comparison was conducted using the same evaluation rubric, maintaining the instrument unchanged in order to isolate interpretive differences effects. **Conclusions:** The results demonstrate a substantial overall correlation between human and automated grades, with an approximate margin of $\pm 5\%$. However, notable discrepancies are evident in their respective evaluations. The model perceives itself as emphasizing formal coherence and argumentative structure, whereas it perceives human evaluators as tending to value originality, contextual relevance, and personal reflexivity more highly. **Implications:** Extending this line of research can offer university instructors and evaluators valuable insights for optimizing the deployment of generative AI as a research and assessment aid within a framework that safeguards academic integrity and professional ethics. We examine the scope and limitations of AI as an assessment assistant, the epistemological tensions between professional judgment and automation, and propose guidelines for a more ethical and formative utilization of these technological tools.

Keywords: AI-assisted assessment, ChatGPT, Higher education, Assessment rubrics, Teacher judgment, AI epistemology.

תקציר

רקע: הופעתם של המודלים הגדולים ושל כלי בינה מלאכותית גנרטיבית שינתה במהירות את דפוסי ההוראה. השימוש בהם עדיין נתפס כתחום בלתי יציב ורווי סיכונים, אך גם כהזדמנות. **מטרה:** מחקר זה מציג ניתוח השוואתי בין הערכות שביצע מרצה באוניברסיטה לבין אלה שמבוצעות על ידי ChatGPT לגבי עשרים עבודות גמר בקורס לתואר מוסמך בחינוך. מטרתו היא לבחון, תוך שימוש במחווון משותף, את הדמיון והשוני בציונים ובהערות בין מעריך אנושי לבין הערכה מבוססת בינה מלאכותית. שאלות המחקר הן: 1. באיזו מידה הציונים המוענקים על ידי מעריך אנושי (מ"א) ועל ידי ChatGPT חופפים או מתפצלים כאשר משתמשים באותו מחווון בהערכת עבודות גמר בקורס אוניברסיטאי? 2. אילו הטיות או הדגשים נצפים בפרשנות של הקריטריונים להערכה – פורמליים, פדגוגיים ואתיים – בין המעריכים? **מתודולוגיה:** מדובר במחקר איכותני השוואתי, חקר מקרים מרובים, שנועד לנתח את נקודות המפגש והפערים בין הערכה אנושית לבין הערכה מבוססת ChatGPT. ההשוואה בוצעה באמצעות אותו מחווון הערכה, תוך שמירה על כלי מדידה קבועים כדי לבודד ככל האפשר את ההבדלים בפרשנות. **מסקנות:** הממצאים מצביעים על מתאם כולל גבוה בין הציונים האנושיים לציונים מבוססים בינה מלאכותית, בהפרש ממוצע של כ-5%. עם זאת, עולים גם פערים בולטים בכמה מקרים שנבחנו. הבינה המלאכותית רואה את עצמה כמעריכה קפדנית מן הבחינה הפורמלית, המדגישה היבטים צורניים הקשורים למבנה השפה, ובהם בהירות וצפיפות מושגית; בעוד שהמ"א מתואר על פי הב"מ כמי שמדגיש רכיבים יצירתיים, קונטקסטואליות של התשובות ואוטנטיות רפלקסיבית. בנוסף, הב"מ נוטה להדגיש עקביות פורמלית ומבנה טיעוני, בעוד שהמעריך האנושי (על פי הב"מ) נותן משקל רב יותר למקוריות ולרלוונטיות קונטקסטואלית. אם אכן האפיסטמולוגיה של הבינה המלאכותית מוגבלת להיבטים מסוימים של ההערכה, ייתכן שדווקא מגבלה זו היא המפתח לאינספור אפשרויות מעבר למודל. המחקר מראה כי הבינה המלאכותית יכולה לשמש כמראה של החשיבה ההערכתית האנושית, ולחשוף היכן מצויה למעשה המיומנות הפדגוגית של המעריך האנושי. **השלכות:** העמקת קו מחקר זה עשויה להעניק למרצים ולמעריכים באקדמיה תובנות משמעותיות לשם מיטוב השימוש בבינה מלאכותית כעוזרת מחקר והערכה, במסגרת שמירה על תקינות אקדמית ועל אתיקה מקצועית. בעבודה נדונים היקף השימוש והמגבלות של מערכות בינה מלאכותית כעוזרי הערכה, המתח האפיסטמולוגי שבין שיפוט מקצועי לבין אוטומציה, וכן מוצעים קווים מנחים לשימוש אתי בכלים אלה.

מילות מפתח: הערכה בסיוע בינה מלאכותית, ChatGPT, השכלה גבוהה, מחווון הערכה, שיפוט מקצועי של מרצים, אפיסטמולוגיה של הבינה המלאכותית.

מקורות

- Doğan, M., Celik, A., & Arslan, H. (2025). AI In Higher Education: Risks and Opportunities From the Academician Perspective. *European Journal of Education, 60*(1), e12863.
- Hall, K., & Severance, D. (2025). Development of Evaluation Criteria and Comparison of Performance Between Sophomore and Upper-Division Resident Assistants. *The Journal of College and University Student Housing, 51*(2), 3-22.
- Jauhainen, J. S., & Garagorry Guerra, A. (2025). Generative AI in education: ChatGPT-4 in evaluating students' written responses. *Innovations in Education and Teaching International, 62*(4), 1377-1394.
- Kalantzis, M., & Cope, B. (2025). Literacy in the time of Artificial Intelligence. *Reading Research Quarterly, 60*(1), e591.
- Khan, W. N. (2024). Ethical challenges of AI in education: Balancing innovation with data privacy. *AI EDIFY Journal, 1*(1), 1-13.
- Oh, J., Kim, E., Cha, I., & Oh, A. (2024). The generative AI paradox on evaluation: What it can solve, it may not evaluate. *arXiv preprint arXiv:2402.06204*.
- Román-Acosta, D. (2024). Exploración filosófica de la epistemología de la inteligencia artificial: Una revisión sistemática. *Revista Uniandes Episteme, 11*(1), 101-122.
- Spelda, P., & Stritecky, V. (2024). Why and how to construct an epistemic justification of machine learning? *Synthese, 204*(2), 74.