מי כתב את זה, ולמה זה משנה

מסמך הרקע של המערכת לכתבה הראשית בגיליון 16. הוא נכתב בעקבות מסמך מחקר חיצוני שהתקבל, ולאחר קריאה של כל מחקר שמגיע לעמוד במקורו ולא בתיאור שלו. בשני מקרים המקור הראה משהו שהמסמך לא אמר, ובמקרה אחד הוא סתר את עצמו.

18 באוקטובר 2026 · מהדורה מס' 16 · 1247 מילים

מעמד המסמך

זהו מסמך הרקע של מערכת "היום בקיצור", והוא המקור שהכתבה הראשית מפנה אליו. הוא נכתב בעקבות מסמך מחקר חיצוני באנגלית בן 24 עמודים ו-2,718 מילים, שכותרתו Authenticity in the Age of Artificial Intelligence והתקבל במערכת לפני גיליון 16.

המסמך שהתקבל אינו מתפרסם. הוא נשמר בעותק המקורי בתוך המאגר של העיתון ובהיסטוריית הגרסאות שלו, כך שאפשר לבדוק מה נלקח ממנו ומה נדחה. הוא אינו מאומת בעצם קיומו, והעובדה שהוא מצטט כתבי עת שפיטים עם מזהי DOI אינה פוטרת מקריאה.

כל מחקר שמספר ממנו מגיע לעמוד נקרא במקורו ב-17 באוקטובר 2026: ניסוי השירה במלואו דרך PMC, תקציר מחקר האמפתיה דרך PubMed, תקציר מחקר הייחוס באתר כתב העת, ונוסח סעיף 50 וסעיף 113 בטקסט החוק האירופי.

המסמך שהתקבל יצא מהבדיקה הזו טוב יותר מקודמיו. הוא לא נתפס בטעות עובדתית. מה שכן נמצא הוא שהוא מתאר את הממצאים בצמצום, ובשני מקומות החמיץ את הפרט שהופך אותם לחזקים יותר. בפרק האחרון מפורט מה נדחה ומה לא ניתן היה לאמת.

תוכן

1. 1. השאלה
2. 2. ניסוי השירה, שנקרא במלואו
3. 3. מחקר האמפתיה, ומה שהמסמך החמיץ בו
4. 4. מה שהתווית עושה כשהיא מגיעה באיחור
5. 5. מה שהחוק קובע, ומה שהוא לא התכוון לקבוע
6. 6. ההיפוך: זו אינה שנאה למכונה
7. 7. מה נדחה, ומה לא ניתן היה לאמת
8. 8. גבולות המסמך הזה

1. השאלה

המילה אותנטי נשענה מאות שנים על הנחה שקטה אחת: שאפשר להסיק מהצורה על המקור. מכתב מגושם נכתב בידי אדם, ושיר שנון דרש משורר. ההנחה הזו נשברה, ולא בהדרגה. השאלה שהמסמך נשאל היא מה נשאר מהמילה אחרי שהיא נשברה.

התשובה שהוא נותן אינה שהמילה התרוקנה. היא שהמילה החזיקה כל הזמן חמש שאלות שונות שנסעו יחד, ועכשיו הן נפרדו. מאיפה הטקסט הגיע. מי הפעיל עליו שיקול דעת. האם הוא מבטא את מה שהשולח באמת חושב. מה הוא מחייב את השולח כלפי המקבל. ואיך המקבל מפרש אותו.

ההפרדה הזו אינה תרגיל פילוסופי. היא מה שמסביר למה הספרות נראית סותרת. מכתב ניחומים שמודל כתב יכול להיות אפקטיבי רגשית ולא כן. מכתב שאדם כתב יכול להיות אותנטי לחלוטין מבחינת מקור ובכל זאת מניפולטיבי. שתי האמירות נכונות בו זמנית ברגע שמפסיקים להתייחס לאותנטיות כאל תכונה אחת.

אותנטיות הופכת לפחות תכונה של הצורה ויותר תכונה של מקור, סוכנות, אישור, יחס ומחויבות שאפשר לתבוע עליה אחריות.

המסקנה הכוללת של המסמך שהתקבל, בניסוח שלו

2. ניסוי השירה, שנקרא במלואו

המצא שמחזיק את הכתבה פורסם ב-Scientific Reports ב-14 בנובמבר 2024, מאת בריאן פורטר ואדואר מאשרי. המערכת קראה אותו במלואו ולא בתיאור שבמסמך. הוא מורכב משני מחקרים שהורצו על אותם עשרה שירים: חמישה של משוררים ידועים, וחמישה שמודל ייצר בסגנונם.

במחקר הראשון התבקשו המשתתפים לומר על כל שיר מי כתב אותו. דיוק הזיהוי היה 46.6 אחוזים על פני 16,340 שיפוטים. זה לא נכשל להיות טוב מהמקרי. זה היה גרוע מהמקרי, ברמת מובהקות גבוהה. הסיכוי ששיר של משורר אנושי יזוהה כאנושי היה 0.758 מהסיכוי של שיר שהמודל כתב.

הפרט שאין לו תשובה נוחה

המשתתפים לא ענו באקראי. ההסכמה ביניהם הייתה נמוכה אך גבוהה מהמקרי, כלומר הם השתמשו בכללי אצבע משותפים. הכללים פשוט היו הפוכים: הם הניחו שמודל אינו יודע לחרוז ואינו כותב שירים ארוכים. בפועל 89 אחוזים משירי המודל חרזו בכל השורות, לעומת 40 אחוזים משירי האדם. בנוסף, ביטחון עצמי גבוה יותר ניבא דיוק נמוך יותר, לא גבוה יותר.

המחקר השני ביקש דירוגים ולא זיהוי, ושינה שני דברים בנפרד: מי באמת כתב את השיר, ומה נאמר למשתתף על מי שכתב אותו. בסולם של שבע, כשנאמר למשתתף שמודל כתב, הדירוג ירד ב-0.814 נקודות לעומת כשנאמר לו שאדם כתב. כשהמודל באמת כתב, הדירוג עלה ב-1.045 נקודות.

שני המספרים נמדדו באותו מחקר, על אותם עשרה שירים, מאותם משתתפים. הם מצביעים לכיוונים הפוכים. הדפוס חוזר ב-13 מתוך 14 ממדי איכות, וכל חמשת שירי המודל דורגו מעל כל חמשת שירי המשוררים בדירוג האיכות הכוללת.

היוצא מן הכלל, שהוא הממצא

הממד היחיד שבו הדפוס נשבר הוא מקוריות. שם התווית עדיין מורידה, אבל מחבר אנושי אמיתי אינו מקבל יתרון מובהק. כלומר מה שהקוראים חושבים שהם מגינים עליו כשהם מעניקים תווית AI הוא בדיוק הממד היחיד שבו הנתונים אינם מראים שלאדם יש יתרון.

3. מחקר האמפתיה, ומה שהמסמך החמיץ בו

המסמך מצטט תוכנית מחקר שפורסמה ב-Nature Human Behaviour ביוני 2025, ומתאר אותה כתשעה מחקרים עם 6,282 משתתפים שמצאו שאמפתיה שיוחסה לאדם הוערכה טוב יותר. שני המספרים נכונים. התקציר במקור אומר יותר מזה.

בכל הזרועות של כל תשעת המחקרים, התשובות האמפתיות נוצרו על ידי מודל. אף תשובה לא נכתבה בידי אדם. מה שהשתנה הוא התווית בלבד. כלומר ההפרש שנמדד אינו הפרש בין כתיבה אנושית לכתיבה ממוכנת. הוא הפרש בין שתי תוויות על אותו טקסט.

ממצא שני, שאינו מופיע במסמך שהתקבל כלל, חד עוד יותר: כשמשתתף פיתח מיוזמתו חשד שמודל סייע בתשובה שיוחסה לאדם, האמפתיה הנתפסת ירדה. לא נדרשה תווית. החשד לבדו הספיק.

פרט מקומי

המחבר הראשון, מתן רובין, והמחברת האחרונה, פרופ' ענת פרי, הם מהמחלקה לפסיכולוגיה באוניברסיטה העברית בירושלים. שאר המחברים מ-Harvard Business School ומאוניברסיטת טקסס באוסטין. המסמך שהתקבל אינו מזכיר זאת.

4. מה שהתווית עושה כשהיא מגיעה באיחור

המחקר השלישי שנקרא במקורו הוא של דוריגוני וג'ארדינו, Frontiers in Psychology 2025. הוא בדק מסרים בהקשרים רגשיים כבדים, לידה מול מחלה סופנית, כשהם מיוחסים לארבעה מקורות: חבר קרוב, חנות פרחים, גוגל ו-ChatGPT.

הממצא שמדווח בתקציר: רמזים אנתרופומורפיים מעלים את ההערכה עד לרגע שבו המקור נחשף, ואז היא קורסת. ייחוס ל-ChatGPT הוביל לירידות החדות ביותר באותנטיות ובכבוד מוסרי. ככל שההקשר כבד יותר, כך הקריסה חדה יותר.

זו החוליה שמחברת בין השניים הקודמים. בניסוי השירה התווית ניתנה מראש והורידה דירוג. כאן היא ניתנה אחרי שהקורא כבר התרשם, והורידה אותו מגובה רב יותר.

5. מה שהחוק קובע, ומה שהוא לא התכוון לקבוע

סעיף 50 לחוק ה-AI האירופי נקרא בטקסט החוק ולא בסיכום שלו. סעיף 50(1) מחייב ספק להבטיח שאדם שמתקשר עם מערכת AI יידע זאת, אלא אם זה מובן מאליו לאדם סביר בנסיבות העניין. סעיף 50(2) מחייב שתוכן סינתטי, שמע, תמונה, וידאו או טקסט, יסומן בפורמט קריא-מכונה וניתן לזיהוי כמלאכותי.

תאריך התחולה אומת בסעיף 113, שקובע שהתקנה חלה מ-2 באוגוסט 2026. סעיף 50 אינו ברשימת החריגים המוקדמים או המאוחרים שבאותו סעיף. כלומר הוא בתוקף כבר כשני חודשים וחצי.

> **מה שלא נטען כאן**
>
> המחוקק האירופי לא ביקש להשפיע על דירוגי איכות ולא התייחס לשאלה הזו. מטרת סעיף 50 היא שאדם יידע עם מה הוא מתקשר. העיתון אינו טוען שהוא יצר את הקריסה שהניסויים מודדים, ואין מחקר שבדק את התחולה האירופית כהתערבות. מה שכן אפשר לומר בביטחון הוא שהרמז שהניסויים מראים שהוא משנה שיפוט הפך בעשרים ושבע מדינות מרמז אופציונלי לחובה.

6. ההיפוך: זו אינה שנאה למכונה

הקריאה הקלה של כל האמור לעיל היא שאנשים דוחים AI. המסמך עצמו חוסם אותה, ובצדק. הספרות שהוא סוקר מתעדת גם את התופעה ההפוכה, הערכת יתר לעצה אלגוריתמית, במשימות שנתפסות כמדידות ואובייקטיביות.

- הסלידה אינה מופיעה בכל מקום. היא מתרכזת במשימות שנתפסות כסובייקטיביות או כדורשות שיפוט אנושי.
- במשימות שנתפסות כחישוביות, הכיוון מתהפך ואנשים מעדיפים דווקא את המלצת האלגוריתם.
- לכן השאלה אינה אם הקורא סומך על מכונות, אלא באיזו קטגוריה הוא סיווג את המשימה שלפניו.

מכאן גם מגיע המתח המעשי שהמסמך מצביע עליו: במחקרי האמפתיה אנשים העדיפו את התשובה שהמודל ייצר, ובמקביל העדיפו באופן עקבי שאדם יהיה זה שעונה להם. אלה אינן העדפות סותרות. הן שתי שאלות שונות שנשאלו באותו רגע.

7. מה נדחה, ומה לא ניתן היה לאמת

המסמך שהתקבל מצטט ספרות רחבה בהרבה ממה שמגיע לעמוד. ההפרש אינו מקרי. מה שלא נקרא במקורו אינו מודפס, גם כשהוא נשמע סביר ומצוטט כראוי.

- הניסוי הקליני של Therabot, שהמסמך מתאר כ-210 משתתפים עם דיכאון או חרדה משמעותיים קלינית. לא נבדק במקורו, ולכן אינו מופיע בכתבה.

- ההשוואה של אייארס ועמיתיו על 195 שאלות מ-r/AskDocs. לא נבדקה, ואינה מופיעה.
- ניסוי השדה של בסטאני ועמיתיו על כאלף תלמידי תיכון טורקים. לא נבדק, ואינו מופיע.
- מאמר ה-PNAS וטיוטת השכנוע עם 1,601 משתתפים. לא נבדקו, ואינם מופיעים.
- בדיקות ה-NIST שהמסמך מזכיר. לא נבדקו במקורן, ולכן האזהרה שזיהוי אוטומטי עדיין אפשרי מנוסחת בכתבה כסייג ולא כממצא מכומת.
- מספר המשתתפים במחקר השני של ניסוי השירה. המסמך אומר 696, והמאמר עצמו מדווח רק דרגות חופש של 693 ברמת המשתתף. המספר אינו מודפס.

> **סתירה שנמצאה במקור, לא במסמך**
>
> התקציר של דוריגוני וג'ארדינו מתאר ארבעה מקורות מיוחסים: חבר קרוב, חנות פרחים, גוגל ו-ChatGPT. המבוא של אותו מאמר עצמו מתאר עיצוב אחר, עם ארבעה מקורות שונים, עצמי, חבר, זר ו-AI, ובנוסף מניפולציה של כתב יד מול הקלדה שאינה מוזכרת בתקציר כלל. המאמר סותר את עצמו בתיאור העיצוב שלו. לכן הכתבה מדפיסה את המסקנה שהתקציר מדווח, ולא את פרטי העיצוב.

שלוש טענות נוספות של המסמך נדחו משום שהוא עצמו מסייג אותן. אסור להדפיס שתוצר של מודל הפך לבלתי ניתן להבחנה מתוכן אנושי: זה נמדד בניסויים מסוימים, ובניסוי אמנות שהמסמך מביא הנבדקים דווקא הבחינו מעל הסף. אסור להסיק מכאן דבר על תודעה, רגש או חוויה של מודל, והמסמך אומר במפורש שהספרות ההתנהגותית אינה מכריעה בזה. ואסור לכתוב שהפרמיה על מקור אנושי תגדל ככל שה-AI יתפשט. המסמך מסמן זאת כסביר וכחסר ראיות אורך, ולכן זה אינו מופיע.

8. גבולות המסמך הזה

כל המחקרים שמגיעים לעמוד הם ניסויים עם משתתפים מגויסים, רובם אמריקאים, ברובם לא מומחים. בניסוי השירה 90.4 אחוזים מהמשתתפים דיווחו שהם קוראים שירה כמה פעמים בשנה או פחות. זה אינו פגם בניסוי, שבדק בדיוק את הקורא הרגיל, אבל זה כן אומר שאין כאן ממצא על מבקרי שירה.

שנית, כל המדידות הן נקודתיות. אין כאן סדרת זמן, ולכן אין בסיס לטעון שהעונש על התווית גדל או קטן עם השנים. מי שרוצה לדעת אם הקורא יתרגל חייב לחכות למדידה חוזרת.

שלישית, המסמך הזה נכתב מתוך מסמך שנכתב בידי מערכת AI, ועובדה זו רשומה כאן משום שהיא בדיוק מה שהכתבה עוסקת בו. מה שמבדיל בין המסמך שהתקבל לבין מה שמודפס אינו שמישהו אנושי כתב אותו מחדש. זה שמישהו הלך למקורות ובדק.

מקורות

1. המסמך שהתקבל. 24 עמודים, 2,718 מילים, באנגלית. נשמר במאגר העיתון ואינו מתפרסם. Authenticity in the Age of Artificial Intelligence, התקבל לפני גיליון 16

2. ניסוי השירה. נקרא במלואו ב-17 באוקטובר 2026. ממנו: 46.6 אחוזים דיוק על 16,340 שיפוטים, יחס סיכויים 0.758, הפרש של 0.814 נקודות לתווית ו-1.045 נקודות למחבר האמיתי, 13 מתוך 14 ממדים, ו-90.4 אחוזים קוראי שירה מזדמנים. Porter, B. ו-Machery, E., AI-generated poetry is indistinguishable from human-written poetry and is rated more favorably, Scientific Reports 2024 בנובמבר 14, כרך 14, doi 10.1038/s41598-024-76900-1

3. מחקר האמפתיה. התקציר המלא נקרא ב-17 באוקטובר 2026. ממנו: תשעה מחקרים, 6,282 משתתפים, כל התשובות נוצרו על ידי מודל, והחשד העצמאי מוריד בעצמו את האמפתיה הנתפסת. Rubin, M. ואחרים, Comparing the value of perceived human versus AI-generated empathy, Nature Human Behaviour 2025 בנובמבר, כרך 9, עמודים 2345 עד 2359, doi 10.1038/s41562-025-02247-w, PMID 40588597

4. מחקר הייחוס. התקציר נקרא ב-17 באוקטובר 2026. ממנו: הקריסה ברגע החשיפה, והירידה החדה ביותר בייחוס ל-ChatGPT. - Dorigoni, A. ו-Giardino, P. P., Frontiers in Psychology 2025, doi 10.3389/fpsyg.2025.1568911

5. חובת הגילוי. נקראה בטקסט החוק ולא בסיכום, ב-17 באוקטובר 2026. סעיף 50 לעניין חובות השקיפות וסעיף 113 לעניין תקנה (EU) 2024/1689, תחולה מ-2 באוגוסט 2026

מסמך הרקע נכתב על ידי מערכת "היום בקיצור" לגיליון 16, 18 באוקטובר 2026. הוא נבנה מתוך מסמך מחקר חיצוני שהתקבל ואינו מתפרסם, ומתוך קריאה חוזרת של כל מקור שמספר ממנו מגיע לעמוד. מה שלא נבדק במקורו אינו מודפס, ומה שנדחה רשום בפרק 7.