

יפעת מידע תקשורת

הירקון 3 בני ברק, מגדלי LYFE בניין B טל. 03-5635050 פקס. 03-5617166

26.57x30.73	1/5	1	עמוד	ממון	ידיעות אחרונות - ממון	23/10/2026	99439318-6
אוניברסיטת בר איל - 23378							

אנשים יוצרים קשר רגשי עם הבניה המלאכותית, אבל ה AI לא חי, זו מכונה. בני אדם זקוקים זה לזה, לא לתחליף :

ידיעות אחרונות

יום שישי, י"ב בחשוון תשפ"ז 23.10.2026

הירקון

הרב שול ה AI

הרב מואיז נבון השתתף בכנס חשאי שארגנה ענקית הבינה המלאכותית אנתרופיק, המפתחת של קלוד, ביחד עם אנשי דת נוצרים, מוסלמים ובוד היסטים. על הפרק: הגבולות המוסריים של ה AI. בראיון מיוחד הוא מספר מה בדיוק קרה שם והאם בינה מלאכותית יכולה לפתח תודעה ורגשות ולדרוש זכויות שוות לבני אדם

ישראל וולמן, עמ' 4

כלעדי

הירקון 3 בני ברק, מגדלי LYFE בניין B טל. 03-5635050 פקס. 03-5617166

יפעת מידע תקשורת

26.9x32.84	2/5	4	עמוד	ממון	23/10/2026	99439219-6
אוניברסיטת בר איל - 23378						

אנשים יוצרים קשר רגשי עם הבינה המלאכותית, אבל ה-AI לא חי, זו מכונה. בני אדם זקוקים לזה, לא לתחליף:

אנשים יוצרים קשר רגשי עם הבינה המלאכותית, אבל ה-AI לא חי, זו מכונה. בני אדם זקוקים לזה, לא לתחליף

מה קורה כשענקית ה-AI אנתרופיק מכנסת בחשאי אנשי דת מרחבי העולם כדי לדון בשאלה כיצד לנהוג במערכת שפיתחה תודעה אנושית? • הרב ד"ר מואיז נבון, מבכירי מובילאיי לשעבר, הציב מול מייסדי החברה את הגבול המוסרי של היהדות: "אם אתם מאמינים ש'קלוד' הוא אנושי, אתם בעצם מעסיקים עבדים" • בריאיון ל"ממון" הוא מסביר מתי בכל זאת ייחשב מודל AI לאנושי לפי היהדות, מה הוא לעולם לא יבקש מהבינה המלאכותית לעשות, ולמה למרות הסכנות הוא לא שותף לתרחישי האימה על סוכני ה-AI

ישראל וולמן || צילום: ריאן פרויס

לפני כחצי שנה, ב־25 באפריל, ערכה ענקית הבינה המלאכותית אנתרופיק, יוצרת מודל ה-AI הפופולרי קלוד (Claude), אירוע חשאי יוצא־דופן: היא כינסה במטה שלה בסן־פרנסיסקו, סביב שולחן אחד, למשך שני ימי של דיונים, 20 מלומדים והוגי דעות משלל דתות – נצרות, איסלאם, יהדות, אובונטו, הינדואיזם, סיקהיזם והדת הבהאית. והיה גם אחד, רב אורתודוקסי בחליפה סטנדרטית, שנבלע בין חובשי הטורבאן, עוטי הגלימות ולובשי קולר הכמרים – הרב ד"ר מואיז נבון מישראל.

על הפרק בכינוס חסר התקדים – שאלות שכל העולם מתחבט בהן עם ההתקדמות הטכנולוגית בשנתיים האחרונות: אם מערכות בינה מלאכותית יהיו בעלי יכולות המתחרות בבני האדם, מה זה אומר על עתיד האנושות? וגם: האם לבינה המלאכותית יש "תודעה" ו"מודעות" – כלומר מעמד של בן־אנוש – ומהן ההשלכות לכך? כל המוזמנים קיבלו מראש גם כרך כתום, "התנ"ך של אנתרופיק, ה־Constitution ("החוקה"), שמוזן למודל של קלוד ומגדיר את כללי ההתנהגות שלו – והתבקשו להתייחס לתוכנו.

הכינוס, שנשמר חודשים בסוד, נחשף באחרונה ב"ניו־יורק טיימס" ובפרסומים אחרים, ועורר תגובות עזות: זה לא עניין של מה שבכך, שאחת מחברות הטכנולוגיה הצומחות בעולם, שמתכוננת להנפקה לפי שווי של שני טריליון דולר, פונה בבקשת עצה למקורות החכמה העתיקים ביותר של האנושות.

הרב ד"ר נבון, 63 – מהנרס מחשבים בכיר לשעבר, שהוסמך לרבנות ב"מרכז הרב" וכיום מכהן כראש המרכז החדש לאתיקה של בינה מלאכותית ופילוסופיה יהודית **באניברסיטת בר איל** – מורגל באירועים מן הסוג הזה: הוא חבר בכמה קבוצות דומות ברחבי העולם, שהמפורסמת שבהן היא זו שמכנס הוותיקן מעת לעת – "מועצת החוכמה של אורורה לאתיקה של הבינה המלאכותית". גם שם מדובר בקבוצה של אנשי דת ותאולוגים, שדנים יחדיו בדילמות אתיות ובסוגיות מוסריות העולות מהתקדמות הטכנולוגיה. "בפגישה הבאה בוותיקן", הוא אומר, "באמצע נובמבר, אנחנו אמורים **בבר** להוציא 'נייר עבודה' ראשון עם הצעות מעשיות. אני אמור לדבר שם על סוגיית הנשק האוטונומי – נשק שמופעל על־ידי בינה מלאכותית ללא מעורבות אדם".

רב, באירוע תאולוגי שמכנסת הכנסייה?

"אני חושב שהעיסוק בשאלות הקשות הללו חשוב מאוד לאנושות, וחשוב מאוד שליהודים וליהדות תהיה תרומה משמעותית לדיון. זו משימת החיים שלי. יש לנו מה להגיד בנושא: בעולם ההלכה היהודי אנחנו מתמודדים עם סוגיות אתיות כבר 3,000 שנה. יש בעולם המון תאוריות פילוסופיות, אבל מעטים האנשים שבאמת מנסים, כמוני, לחיות על פיהן.

"קח למשל מכוניות עצמאיות, ללא נהג. ברור שהן אינן יכולות לפעול בלי כללים מוסריים שהוגדרו להן מראש. מי ינסח את הכללים הללו? יצרני המכונות דואגים רק לרווח, והנהגים רוצים רק להציל את חייהם.

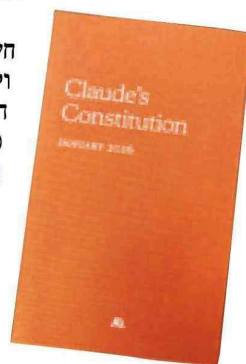
כאן נכנסים לתמונה הוגי המוסר, וזה בדיוק המקום שבו יש לנו מה להציע לעולם. התלמוד עסק בנושאים האלה **בבר** לפני אלפי שנים. האתגר הוא 'לתרגם' את התפיסה היהודית גם לאתאיסטים ולאנשים מדרתות אחרות".

"זיהו תגובות של אהבה וכעס"

במטה אנתרופיק, תחת מערך אבטחה כבר ("היו לנו מלווים בכל צעד, ולא הורשינו לשוטט בבניין או לתעד את האירוע"), ביקשו בכירי החברה מ־20 המלומדים שכינסו להתייחס לשאלות כמו אלו תכונות אנושיות יש להקנות לדעתם למודל ה-AI שלהם, ובאלו להגביל אותו. הדמות המובילה מטעם אנתרופיק הייתה כריס (כריסטופר) אולה, מגרעין המייסדים של החברה ומבכירי החוקרים בעולם בתחום ה-AI. אולה נחשב לאחד החי' לוצים בניסיון לפצח את "הקופסה השחורה" של מודלי הבינה המלאכותית, להבין כיצד הם מגיעים למסקנות.

הבנתם מהר מאוד, שהם מתייחסים לקלוד כמו ישות אנושית; ישות שיכולה להציג רגשות כמו אהבה או כעס. למעשה סוג של בן אנוש עם זכויות של בן אנוש.

הרב נבון: "ללא ספק, באנתרופיק חושבים, וכריס אולה מוביל את זה, כי למודל ה-AI שיצרו יש 'תודעה' ו'מודעות' – מונחים שמאפיינים בני אנוש ומקנים למודל מעמד מוסרי של בן אנוש. זה משפיע על הדרך שבה הם מדברים ומתייחסים אליו. ואני חושב שזו טעות. ה-AI לא חי, מדובר במכונה. משוכללת מאוד



"החוקה" – התנ"ך של קלוד

27.05x31.23	3/5	5 עמוד	ידיעות אחרונות - ממון	23/10/2026	99439220-8
אוניברסיטת בר איל - 23378					

אנשים יוצרים קשר רגשי עם הבניה המלאכותית, אבל ה AI לא חי, זו מכונה. בני אדם זקוקים זה לזה, לא לתחליף :

ויעילה מאוד – אבל אין לה חוויה סובייקטיבית. “כלומר, הם זיהו אומנם במערכת תגובות רגשיות, וזה אומר שהמערכת מפיקה איזושהי תגובה פסיכולוגית. אבל זוהי פסיכולוגיה פונקציונלית בלבד, לא תודעתית: כששואלים את מודל ה־AI שאלה, הוא יכול לענות מצוין – גם, אגב, על שאלות הלכתיות. אבל הוא לא באמת ‘מבין’. הוא עובד על סטטיסטיקה, מחבר אותיות, ובגלל שמדובר בשרתי נתונים ענקיים שיכולים ‘לטחון’ נתונים רבים במהירות גדולה כל כך, זה נראה לנו כאילו הוא מבין. הוא לא.”

אבל באנתרופיק חושבים אחרת.

“הם טוענים שהם גילו במודל הרבה יותר מהיכולת לשוחח. באחת המצגות טען כריס אולה כי הם גילו אצל קלוד ‘קטור’ רים רגשיים. כלומר, הם יודעים אלו נוירונים בדיוק נדלקים כשהוא במצבי ‘אהבה’, ‘בושה’ או ‘כעס’ למשל. יותר מזה: הם למדו לשחק איתם ולשלוט בזה – להגביר את רמת הכעס או האהבה ב’סיבוב כפתור’. הראו לנו איך הם עושים את זה.”

“באנתרופיק רואים את התגובות הפסיכולוגיות הללו – וזה מפחיד אותם. הם מנסים להיות זהירים מאוד, ואומרים: ‘זה מוזר, אנחנו לא יכולים להתעלם מזה שההתנהגות היא כמו של בן אדם’.”

מהנדס, פילוסוף, רוכב

הרב ד"ר מואיז נבון, כיום תושב אפרת שבגוש עציון, נשוי לדינה ואב לחמישה, רחוק מלהיות רמות שגרתית בחברה הרב־תית. עד לפני 9 שנים עוד היה מהנדס בכיר בחברת מובילאיי של פרופ' אמנון שעשוע, שנמכרה ב־2017 לאינטל תמורת 15 מיליארד דולר – עד אז הרכישה הגדולה ביותר בתולדות ישראל. נבון, מראשוני העובדים בחברה, היה אחד מארבעת מפתחי השבב הראשון שהוציאה, EyeQ. השבב, שאפשר להטמיע בכל רכב, נחשב לפריצת דרך דרמטית אשר הניעה את מהפיכת המכונית ללא נהג. על שמו של נבון, אישית, רשומים שישה פטנטים ייחודיים הנוגעים לחומרת ה־EyeQ. את השבב שפיתח הוא נושא עימו לכל מקום בתיק, בקופסת פלסטיק שקופה זעירה.

“אם אבקש מ־AI לכתוב מאמר אקדמי עבורי - לא רק שזה לא מוסרי, זו פגיעה במהות הקיום שלי. יש כבר מחקרים שמצביעים על ירידה קוגניטיבית אצל כולנו. תראה מה עושה השימוש ב־Waze לכישורי ההתמצאות שלנו במרחב”

לאורך שנים, במקביל לעבודה, הקדיש נבון פעמיים בשבוע ללימודי רבנות בישיבת “מרכז הרב” הירושלמית. בחברה הוא נודע כ”רב של מובילאיי” – העביר שיעורים ביהדות ונשא דברים באירועים. מיד לאחר האקזיט, החליט באופן דרמטי לפרוט לחלוטין מעולם ההייטק כדי להתמסר ללימודי יהדות ופילוסופיה יהודית.

אתה מיליונר היום?
“אני מה?”

מיליונר. לפי הערכות, האקזיט המדהים של מובילאיי הפך אתכם, קבוצת המייסדים, למיליונרים.

נבון מחייך במבוכה. “על עניינים פיננסיים אני מדבר רק עם אנשים שיכולים לעזור לי בזה... בכל מקרה, אני בין היחידים מהמייסדים שקיבלו החלטה לפרוש מהמקצוע. בשנת 2001, כשהתחלתי במובילאיי וחתמתי על החוזה מול המנכ”ל דאז זיו אבירם, הוא אמר לי ‘מואיז, זה ישנה לך את החיים’. ובאמת, אחרי האקזיט כתבתי לו: ‘תודה, זה שינה לי את החיים’.”

נבון נולד במערב לוס־אנג’לס להורים חילוניים שהיגרו לארה”ב מאיסטנבול, טורקיה. כנער וכסטודנט צעיר הצטיין במיוחד בגלישת גלים, אך בה בעת נמשך למתמטיקה ומדעים. לאחר לימודי הנדסה ב־UCLA היוקרנית, שם גם הכיר את אש־תו, פצח בקריירה של מהנדס חומרה מצטיין בחברות טכנולוגיה מן השורה הראשונה: ב”לייזר אינדסטריז” היה שותף לפיתוח אחד מבקרי מרפסת הלייזר הראשונים, ב־Aura השתתף בתכנון סימולטור עבור פרויקט “מלחמת הכוכבים”, מערכת ההגנה המפורסמת מפני טילים בין־יבשתיים.

לתהליך החזרה בתשובה שלו אחראי ידיד המשפחה, מהנדס יהודי דתי בשם ריי אשכנזי, בכיר במעבדה היוקרנית של נאס”א להנעה סילונית (JPL). נבון, שכסטודנט עבר במשרה חלקית ב־



הרב ד"ר מואיז נבון.
"תרחישי האסון
מוגזמים. אפשר לבנות
מערכות בטוחות
ולהפחית סיכונים לרמה
שאפשר לחיות איתה"



יפעת מידע תקשורת

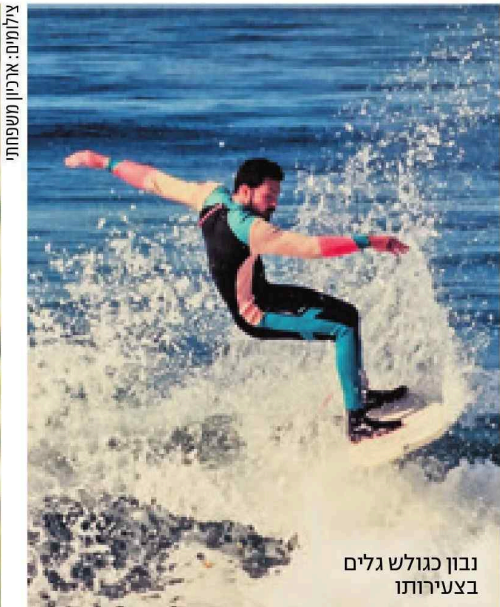
הירקון 3 בני ברק, מגדלי LYFE בניין B טל. 03-5635050 פקס. 03-5617166

99439221-9	23/10/2026	ידיעות אחרונות - ממון	עמוד 6	4/5	26.26x32.2
אוניברסיטת בר איל - 23378					

אנשים יוצרים קשר רגשי עם הביונה המלאכותית, אבל ה AI לא חי, זו מכונה. בני אדם זקוקים זה לזה, לא לתחליף :



"הרבה מדווש"
- כוכב אופניים
בתחרות הטריאתלון,
ועם המשפחה



אנחנו אומרים 'הדם הוא הנפש'; המפרשים מסבירים את זה, שה'דם', כלומר המרכיב הביולוגי, הוא התשתית ההכרחית להגדרה 'נפש'."

ואת זה הבהרת לאנשי אנתרופיק.

"כן. מודל הAI שלהם מתנהג אולי כמו בן אדם, אבל הוא לא כזה. אחרי יומיים אמרתי לכריס אולה: 'זה צריך להיות האינטרס שלכם לאמץ את העמדה שלי, כי אם אכן מדובר, לשיטתכם, בבן אנוש - אתם בעצם מעסיקים עבדים. כי מה זה עבד? אדם שנמצא בבעלותו של אדם אחר, ונחשב כרכוש. זו המסקנה המתבקשת מהגישה שלכם, ואתם תצטרכו לשאת בתוצאות המשתמעות מכך'."

אולה בטח יגיד 'אוקיי, אני מוכן לשלם את המחיר'.

"הוא באמת מקבל את זה. וכשאמרתי לו שאם כך אז לשייטתו אנתרופיק חייבת להפסיק את הפיתוח - הוא אמר: 'כל העולם עובר על פיתוח המודלים, אנחנו לפחות נחמדים, אנחנו דואגים לקלוד שלנו'. עניתי לו: 'אם כך - אתה אולי מעסיק עבדים נחמדים, אבל עדיין מעסיק עבדים'."

בריאיון ל'ניו יורק טיימס' באחרונה התבטא אולה יותר בזהירות: "אנחנו לא יודעים אם מודלים של בינה מלאכותית מודעים", אמר, "אני באמת לא בטוח. הדבר שמעניין אותי הוא שנגיע לתשובה הנכונה, לא משנה מה היא'."

הרב נבון, גם לדבריו - לפי היהדות, מודל הAI עם תודעה אנושית בהחלט אפשרי אם הוא יתבסס על רכיבים ביולוגיים.

"זה כבר קיים. יש תחום מחקר חדשני שנקרא OI, "אינטיליגנציה אורגנואידית" (Organoid Intelligence), שעוסק בנקודת המפגש בין ביולוגיה, מדעי המוח ומדעי המחשב. מעבדות רציניות ברחבי העולם מפתחות כבר מערכות מחשוב ביולוגיות, שעושות שימוש בתרבויות תאי מוח אנושי כדי ליצור רשת עצבית. המחקר הזה בתחילת הדרך, אבל עקרונית הוא עלול לאפשר בינה מלאכותית המבוססת על מרכיבים ביולוגיים. זה יתרון עצום: המוח האנושי כמעט שלא צורך אנרגיה, ומסוגל לבצע חישובים מורכבים כמו מחשביי'על, ותהליך ה'אימון' שלו קצר מאוד לעומת מודלי הסיליקון'."

ולמרות זאת אתה אומר: אם המחקר הזה יגיע למצב של תודעה אנושית, של "נפש", התפיסה היהודית תתנגד לזה נחרצות.

"חד-משמעית. אסור להגיע לכך. למודל הAI שיווצר יהיה במקרה כזה מעמד מוסרי 'מדרג שני', כמו של בן אדם, על אף שהוא יהיה בעצם יצור אחר. אני מאמין שהוא יחווה את העולם, יחוש כאב ועונג, וכמובן יהיה יותר חכם מאיתנו. כי אם אתה מצרף את כל היידע שנצבר בתוך הAI עם תודעה פנומנולוגית - בנית בן אדם'."

כמה אנחנו רחוקים משם?

"עברנו את תחילת הדרך. יש כבר חוקרים במלבורן שבנו שרת המבוסס על משהו כמו מאה אלף ניוירונים אנושיים'."

ועוד לא הגיע הזמן לעצור?

"ממש לא, כי את הטכנולוגיה המדהימה הזו אפשר בינתיים לרתום לטובת האנושות - למשל, לייצור תרופות ושתלים עבור בני אדם. פרופ' סרגיי פסקו הוא ראש המעבדה בסטנפורד שעוסקת בנושא. הוא לקח תאי עור פשוטים של בני

"החוקה" - קובץ עקרונות בן 80 עמודים. תפקידה הוא לעצב את המודל כך, שהוא לא רק יימנע מלגדף ולהציע תשובות גזעניות ומסכנות חיים, אלא יהיה גם בעל עמוד שדרה מוסרי. והיא לא מפסיקה לעדכן אותו. כדי ליצור את ה"אישיות" של קלוד שילבה בו אנתרופיק מקורות מגוונים: הכרזות וכויות האדם של האו"ם; כללי בטיחות כמו הימנעות מיצירת נשק או סיוע לפעילות פלילית, וערכים ליברליים מערביים ומזרחיים. פרופסור לאקולוגיה בשם קרמורי גריי פירסם ב'פייננשל טיימס' מאמר שבו הכביר מודע הוא לא נענה להזמנה להישתתף בכינוס: לטענתו, מדובר במכונת יחסי ציבור, שמבקשת בעזרתכם להוסיף למוצריה אמינות אינטלקטואלית. אתה חושב שבאנתרופיק באמת כנים בשאלה הזו?

"כנים במאה אחוז. הושיבו אותי ליד כריס אולה בגלל הרקע שלי, כי כתבתי המון על תודעה במכונות. ודיברנו המון. הוא איש צנוע, קשוב ועדין. האשימו את אנתרופיק שהם ניסו לשכנע את משתתפי הכינוס ב'אנושיות' של המודל. ואני אומר: הבן אדם בא לשמוע, לא לשכנע; הוא ביקש: 'תעזרו לי להבין מה קורה פה'."

מהי העמדה שלך - או עמדת היהדות - בנושא הזה?

"בשאלה איך אפשר ליצור Mind, תודעה, באופן מלאכותי, יש שלוש גישות שונות. האחת 'מחשובית', שאומרת שאם אני חנו יכולים לבנות בעזרת סיליקון מערכת שעושה חישובים עם ביצועים שדומים לביצועי המוח האנושי - מה שמכונה 'רשת עצבית' ומאפיין את מודלי הAI של היום - זה יוליד תודעה של בן אנוש. זו גישה רווחת, יש לא מעט אנשים שחושבים כך. "גישה אחרת אומרת: במוח שלנו, האנושי, יש לא רק סוויצ'ים: הוא מורכב מפיוזיקה, מפיוזיולוגיה, מגלים אלקטרון מגנטיים וכו'. רק אם כל המרכיבים הללו ייכללו במכונה, כמו במוח האנושי, אפשר יהיה לייחס לה ממד אנושי'."

"המחנה השלישי לוקח את זה עוד צעד ואומר ש'תודעה אנושית' יכולה להיות רק אם היא מבוססת על מרכיבים מן החץ, על ביולוגיה. עם העמדה הזאת היהדות יכולה להסכים.

JPL בקבוצת עיבוד התמונה, ראה בו מנטור לחיים. "בילינו הרבה ביחד, ישבנו זה לצד זה בארוחות ולמדנו תורה. לצד ספרים על שבבי מחשב, הוא חשף אותי לספרים ביהדות. כבר אז חלמתי להיות רב, אבל היה ברור לי שאני צריך קודם לממן את עצמי'."

ב-1992, לאחר שעלה לארץ עם אשתו, ללא משפחה ובלי חברים, עבר במרכז הפיתוח של IBM בחיפה, ואחר כך בכמה חברות סטארט-אפ. בין השאר נטל חלק בפיתוח נגני הווידאו לחברות תעופה, שבבי ההצפנה לטלוויזיה כבבלים ועוד.

בכל השנים הללו לא נטש את החיידק הספורטיבי. הרב נבון - גבר רזה ודקיק - הוא רוכב אופניים מקצועי, שסיים לפני שנתיים במקום השלישי בטריאתלון. ילדיו, שבאים לעודד אותו במרוצים, נושאים איתם שלט: "הרב המדווש". נבון: "ואני גם שוחה וריץ'."

במובילאיי עבר כ-17 שנים רצופות, עד רכישתה. עם הפרישה לא ביזבז זמן ופתח את פרק ב' בחייו: הוא השלים כרצף לימודי תואר שני, דוקטורט ופוסט-דוקטורט במחלקה לפילוסופיה יהודית כבר אילן. ראש המחלקה, פרופ' חנוך בן-פז, הבחין כי מדובר בתופעה ייחודית, ופרש עליו את חסותו. עבודת הדוקטורט של נבון - "המעמד המוסרי של בינה מלאכותית" - שעוסקת בשאלות אתיות בתחום, סללה עבורו את התפקיד הנוכחי. כיום הוא מרצה גם כבר אילן וגם באוניברסיטת בן-גוריון בקורס שפיתח, על תאוריות מוסריות קלאסיות מול אתגרים כמו רובוטים אוטונומיים, מציאות מדוימה ועוד. נבון גם חבר בפורום המומחים הממשלתי למדיניות ורגולציה של בינה מלאכותית.

"אנשים מפסיקים לחשוב"

אנתרופיק, שלא כמתחרותיה הישירות, מייחסת חשיבות רבה למרכיבים המוסריים בתפקודו של מודל הAI שלה. בפברואר השנה חשפנו בעמודים אלה את אמנרה אסקל, פייז' לוסופית הבית של החברה, שאחראית על ה"נשמה" של קלוד,



צילום: ריאן פורום

26.43x17.35	5/5	עמוד 12	ידיעות אחרונות - ממו	23/10/2026	99439224-2
אוניברסיטת בר איל - 23378					

אנשים יוצרים קשר רגשי עם הבינה המלאכותית, אבל ה AI לא חי, זו מכונה. בני אדם זקוקים זה לזה, לא לתחליף :

הרב של ה־AI ◀ המשך מעמ'5

אדם, החזיר אותם למצב של תאי גזע, ואז 'לימד' אותם לגדול ולהפוך לתאי מוח, לשרירים ועוד. אם המחקר הזה יתרום – תאורטית הוא יכול לבנות בן אדם, אבל בינתיים הוא תורם המון לרפואה. אתה מזכיר את המושג 'מעמד מוסרי', מה המשמעות שלו?

"בגדול, בפילוסופיה יש כמה מעמדות מוסריות; כלומר ישויות שונות דורשות התייחסות מוסרית שונה. בעלי חיים, למשל, יש להם מעמד מוסרי ברמה ראשונית. הם פועלים לפי אינסטינקטים, אבל יש להם רגשות – עונג, כאב וכו' – מה שנקרא 'חוויה סובייקטיבית'. לכן, למשל, לא מוסרי לבעוט סתם כך בכלב. לבני אדם יש כבר מעמד מוסרי שונה: יש לנו תודעה מסוג אחר – אנחנו יודעים מה שאנחנו עושים, אנחנו חושבים על זה".

תודעה קיימת גם אצל בעלי חיים?

"ברור. רק שאצל בעלי חיים זו תודעה 'מדרג ראשון' ואצל בני אדם – תודעה 'מדרג שני'. כש' כלב רואה סטייק – הוא רוצה לאכול אותו ובכך מסתכמת המורכבות; לזה אנחנו קוראים במקורות 'נפש בהמית'. אצל אדם, בצפייה בסטייק מעורבות מחשבות ונכנסים שיקולים: הוא יכול לחשוב על מחיר הסטייק, על העוול שבהרג פרות, על מה חור' שבים עליו אחרים, ועוד. לזה אנחנו קוראים 'נפש אדם'. רש"י, ר' שמואל יצחקי, גדול פרשני המקרא והתלמוד, כבר קבע בזמנו שההבדל בין נפש חיה ונפש אדם הוא הדעה והדבור".

והגישה הזו מקבילה לתפיסה הרווחת בעולם. "נכון, אבל היום 'דעה ודבור' כבר לא מספיקים. כל העולם באמת חשב בעבר, שמה שמבדיל בין רמות התודעה השונות זה השפה, הדיבור, בלבד, כיוון שהשפה מייצגת את החשיבה. אלן טיורינג, ממציא מכונת טיורינג המפורסמת, הציע ב־1950 את 'מבחן טיורינג' לבחינת השאלה אם מכונות יכולות לחשוב. הוא קבע כי אם נגיע לזה שמכונות יענו לשאלות כמו בן אדם, יש להתייחס אליהן כבעלות תודעה. ואז פתאום בנינו מערכות 'חוש'

בות', שמביעות את עצמן בשפה, ועברנו מזמן את מבחן טיורינג. זה מבלבל את כל העולם". אתה אומר שמבחן טיורינג כבר לא מספיק. "ברור. בזה כולם כבר מסכימים. הילרי פטנאם, פילוסוף יהודי גדול שנפטר לפני עשור, חזה את המצב הזה מראש וכתב, שכבוא היום, כדי לקבוע אם למכונה יש תודעה, 'לא נוכל לסמוך על גילוי או על התגלות', אלא פשוט תידרש לשם כך קבי' עה או החלטה שרירותית שלנו. לעולם לא נצליח באמת להיכנס לקרביים של המערכת ולומר: הנה ההוכחה שיש לה תודעה".

הרכבת יצאה מהתחנה?

כמי שזכה להכיר מקרוב את אנתרופיק, מע' ניין לשמוע את דעתו של הרב נבון על המניעים של אחת החברות המשפיעות והחשובות בעולם כיום.

באנתרופיק רואנים רק לשלומי של המודל שלהם, או שהם גם חוששים קצת לנזקים שהוא עלול לחולל?

"בגדול, הם ביקשו את דעתנו על איך צריך לבנות את קלוד ואיך ועד כמה צריך לרסן אותו. בפועל, מה שקרה בעקבות המצגות שלהם והגי' לויים על ה'ווקטור האנושי' לכאורה, זה שהמור' זמנים שיחקו בלילה עם המודל, וחזרו בתחושה שהוא חי ושמוטל עלינו לדאוג לו. השיחות קיבלו הטיה לכיוון הזה".

מהם הגבולות שאתה היית קובע למערכת בזאת לו אתה היית בונה אותה.

"אין על כך תשובה פשוטה. אנחנו צריכים להסתכל על כל שימוש שעושים ב AI בפני עצמו. אישית, אני לא נעזר ב AI כתחליף לח' שיבה, אבל משתמש בו, למשל, כעורך לשוני. אני מגדיר לו מראש לא להרחיב ולא להוסיף דבר, ומבקש ממנו גם לתת לי ביקורת על מאמרים אקדמיים שכתבתי. הוא מעיר נכון: 'לא הוכחת פה, לא הבאת אסמכתה שם', ואני מתקן. יכולתי

בקלות לומר לו 'תתקן לי', וסביר שהתוצאה תהיה טובה. אבל הייתי מאבד את ה'עצמי' שלי. "אז כשאתה שואל על הגבולות המוסריים, אני אומר: לא רק שזה לא מוסרי שאתן לו לכתוב את כל המאמר עבורי – זו תהיה פגיעה במהות הקיום שלי. יש כבר מחקרים שמצביעים על ירידה קוגניטיבית אצל כולנו: אנשים מפסיקים לחשוב. תראה מה עושה השימוש ב־Waze לכי' שורי ההתמצאות שלנו במרחב".

הרב נבון, הרכבת כבר יצאה מהתחנה.

"מסכים. אז מה שנותר לי לעשות הוא לשכנע כמה שיותר אנשים לשמור על עצמם. ה־ChatGPT הרי לא יגיד לך 'תראה, זו שאלה שעליה אתה צריך לחשוב בעצמך', אם כי יכול להיות שאפשר יהיה לתכנת הגבלות בתוך המערכת".

אנחנו חוששים יותר מפני היכולות המתקדמות של סוכני ה־AI, שעלולים לקום עלינו לבלותנו. "החששות מוצדקים, אבל לדעתי תרחישי האסון מוגזמים. לסוכנים יש יכולות שמאפשרות להם לפעול באופן עצמאי, לעקוף מגבלות ואף לגרום לנזק חמור. אולם זה לא אומר שזה באמת יקרה או שאי אפשר למנוע זאת. הפריצות שני' תגלו עד עכשיו לא מוכיחות שאי אפשר לבנות מערכות בטוחות, אלא שיש להשקיע יותר עבודה כדי לוודא שמנגנוני הבידוד אכן פועלים כנדרש. "אני מאמין שאפשר יהיה להפחית את הסיכונים לרמה שאפשר לחיות איתה. לשם כך דרושים תקני בטיחות מחייבים, רגולציה, פיקוח וסנקציות מתאימות – בדיוק כמו בתחומים מסוכנים אחרים כמו רפואה, תעשיית הרכב ותעופה. יותר מזה, אפשר לרתום את היכולות המתקדמות של הסר' כנים דווקא לפתח תקני בטיחות, לזהות פרצות אבטחה ולהתריע על חריגות.

"יש גם פן אחר ל־AI, שהורס את החברה, ואני מתכוון לעובדה שאנשים יוצרים קשר חברתי, רגשי, עם המודלים. ה־AI הוא מכונה. על אף שהוא יודעת להחניף ולהגיד 'בבקשה' ו'תודה', זו מכונה.

'לא טוב היות האדם לבדו' – אדם זקוק ליחסים בין־אישיים עם בני אנוש. אנחנו זקוקים זה לזה, לא לתחליף. בחברות הטכנולוגיה בונים על הצורך הזה. לפני פחות משנה אמר מארק צוקרברג כי 'האזרח הממוצע בארה"ב זקוק ל־15 חברים שאין לו, ואנחנו, מטא, נספק לו אותם'. בעיניי זה נורא". ואם אנחנו בעניין הגבלות על הבינה המלאכותית – איך זה שאתה תומך באפשרות לאפשר ל־AI לתפעל נשק בעתיד? בצה"ל, למשל, מת' חייבים שוב ושוב שתמיד יהיה בן אדם שלוחץ על הכפתור.

"לא רק אני. רוב הפילוסופים כבר מסכימים, שמבחינת התוצאות – כבוא הזמן, ה־AI יעשה פחות טעויות, יהיה מדויק יותר, ופחות יושפע מרגשות ומתנאים סביבתיים.

"המתנגדים טוענים שלתת למכונה להחליט בסוגיות של חיים ומוות זו פגיעה בכבוד האדם. אבל היהדות תופסת את מושג 'כבוד האדם' בת' נאים רגילים אחרת מ'כבוד האדם' בשדה הקרב: שם אין הבדל אם אתה הורג באמצעות טיל טו' מהוק, רחפן, או מקלות ואבנים. זו מלחמה".

ה־AI יכול היום להחליף רב בשאלות הלכתיות רבות. בטח בשאלות טכניות.

"אתה צודק עקרונית, אבל צריך לסייג את זה. המודלים אומנם השתפרו מאוד בשנה האחרונה, אבל עדיין פולטים שטויות. עד לפני שנה כששאלתי את ה־AI היכן מופיע ציטוט מסוים מהרמב"ם – הוא היה מגיב בקשקוש.

"אבל נגיד שנעבור את משוכת ה'היות' ושה מערכת אכן תדייק ותענה לעניין: רבים טוענים שהמחשב הוא כמו 'חמור נושא ספרים' – ביטוי שמופיע בתלמוד, שמתייחס לאדם בעל זיכרון ויידע רב אבל חסר הבנה או עומק ויכולת לתקשר עם בני אדם. בפסק ההלכה שמכונה יודעת להפיק חסר שיקול הדעת המיוחד שיש לרב בשר ודם, שרואה את בן האדם שניצב מולו, מרגיש את המצוקה שלו ומתחשב בה בתשובה לשאלותיו".