

# בין "סיפורה" ל"סיפור"

## הדילמה האתית שמאחורי הבינה המלאכותית

קריאת מחשבות, אבחון מחלות, ראייה מבעד לערפל ואפילו מצלמה נטולת עדשה. ה-AI יהפוך להיות כלי הכרחי לפיתוחה של כל טכנולוגיה חדשה. אך לפוטנציאל הרב הזה יש גם סיכון מוסרי ואתי, עליו התריע נתן אלתרמן כבר בשנות ה-60. המחזה פרי עטו אמנם נחל כישלון מסחרי, אבל הלקה הנבואי של המשורר הלאומי רלוונטי מתמיד

פרופ' דוד מנדלוביץ

ויסייע לו להפיק תוצאות משופרות. ברמה המאתגרת יותר, AI יהיה כלי שיחליף את האדם המומחה בביסוס טכנולוגיה חדשה פורצת דרך ובשימוש בה.

לשם המחשה, בתחום התמחות, שקשור למערכות ראייה והבנת תמונה, תהליך חדירת ה-AI החל במישור תכנון מערכות ראייה. חלקי המצלמה השונים, כגון עדשות וחיישנים, מתוכננים כיום בעזרת כלי תכנון מבוססי AI. כנוסף לכך, ה-AI הפך להיות כלי מרכזי גם בשלבי עיבוד התמונה, שהכרחיים בכדי לעשותה מדויקת ואיכר תית יותר בשלב המתקדם עוד יותר, AI מבין עבורנו את התמונה שהמצלמה רכשה ומסייע

להעניק לנו תוכנות שיקשה עלינו להשיג בעצמנו מתוצר המצלמה. בכך הוא יכול להחליף את האדם המומחה. למשל, ויהיו מצביע סכנה בנהיגה, ויהיו פנים, איכון כוונות פליליות או אפילו הבנת רגשות של מצולם. הגביע הקדוש של תחום ה-AI הוא ליצור מערכת המתכללת את כל המשימות הללו בליבת חשיבה טכנולוגית אחת ובכך להשיג מערכות פשוטות, זולות וקטנות, המסוגלות להניב ביצועים פורצי דרך, שיפיקו תובנות שיהיו מעבר לאלו של בן אנוש, אפילו אם הוא מומחה. למשל, קריאת מחשבות, אבחון

מרווין לי מינסקי, מאבות הבינה המלאכור תית (AI) המודרנית, הגדיר את תחום ה-AI באופן הבא: "לגרום למכונה להתנהג בדרך שהייתה נחשבת לאינטליגנטית - לו אדם התנהג כך". היום אנו מבינים שההגדרה הזו מצומצמת מאוד ובמונחים רבים לעיתים התהפכו היוצרות והאדם

ישאף להתנהג בדרך שמכור נה מאוד אינטליגנטית תציע האחרונות הוא בוודאי מהמ"כ כוננים שכפריצות הדרך הטכנולוגיות וביחד עם מהפכת האינטרנט ומהפכת המידע הדיגיטלי הם מעצבים את המ"כ שך התפתחות האנושות. כמובן, שלעובדה זו יש משמעויות טכנולוגיות, חברתיות וכלכליות אדירות, אך למהפכת ה-AI יש גם משמעות אתית-מוסרית שתדון בהמשך.



פרופ' מנדלוביץ | צילום: איתמר זינור-קורפוטוניקס

להחליף את האדם המומחה כולנו מביטים קדימה ושואלים: לאן עולמות ה-AI מושכים אותנו? בהגדרה ההפוכה מזו של מינסקי, ברור ש-AI יחדור לכל תחום שהאדם נוגע בו ותנחה את האדם בכיוון פעולות הדורשות אינטליגנציה גבוהה. משימות הדבר שה-AI יהפוך להיות כלי הכרחי לפיתוחה של כל טכנולוגיה חדשה. ברמה הבסיסית, ה-AI יהיה כלי סיוע למתכנן

מחלות, ראייה מכעד לערפל ואפילו מצלמה נטולת עדשה.

סכנה מהותית

המשמעות העסקית של עתיד כזה היא רבה ובוודאי בכוחה להשפיע על אומת חדשנות כמו ישראל באופן דרמטי. היכולת של ה-AI לעבור מרמת האלגוריתם המתוחכם לרמת המערכת הכוללת, משלב התכנון ועד לשיי

מוש בשטח, יכול להוביל להקמת סדרה רחבה של מיוזמים שידעו להציע פתרונות זולים ורבי ביצועים, תוך עקיפת השחקנים העולמיים הקיימים בשוק, שמצטיינים באספקט אחד של המערכת בלבד. אשר על כן, למדינת ישראל עניין מיוחד בתחום ה-AI כתחום התפתחות מרכזי. ברם, לפוטנציאל הרב הזה יש גם מימד מוסרי ואתי. המכוונות הללו, בהיותן בעלות

## דיפקטק ו-AI

הדה בושס כתבה ב"הארץ" שמדובר ב"כליל של קטעי-רעיונות, פסבדו-פילוסופיה חבוט, פשטנות מתמיהה ומביכה... הצגה כבידה, מיינעת, ארוכה, סטטית". אלתרמן חשב שהקהל והמבקרים החמירו עם המחזה ולכן בשנת 1966 פירסם מאמר אחרית-דבר שכוון תרנו "בין סיפרה לסיפור" ובו הסביר את תפיסת עולמו שהמכונה "פית" היא מכשיר דייקן, צייתן, מסור ורב עוצמה בעלת משמעות חיובית. המנהל לעומתו שילב שיקולי מוסר, שורים לסט שיקולי ההחלטה של פית, וכך נוצר הקונפליקט. במאמרו כותב אלתרמן: "אם יש גיבור חיובי במחזה הרי המחשב 'פית' הוא גיבור חיובי זה, ואין הוא אותה מפלצת טכנולוגית נדושה".

בהמשך מפציר אלתרמן באנשי הטכנולוגיה לא להירתע מלהמשיך ולפתח טכנולוגיות פורצות דרך, אשר במוסר נכון של מפתח חי הטכנולוגיות יביאו לעולם טוב יותר. המחשבה לעצור את פיתוחן של טכנולוגיות פורצות דרך אינו מועיל וכנראה גם אינו מעשי, ומכאן אנו מובלים למסקנה שהאתיקה מחייבת אותנו לפקח על יוצרי ומפעילי המכונות ולא על המכונות עצמן. בעולם הטכנולוגיה, המכונה, בעזרת רשת החשיבה שלה, מייצרת את מחשבתה כ"סיפרה" ואילו האדם בבסיס המוסר שלו מייצר את ה"סיפור". במחזה של אלתרמן התהפכו היוצרות וה"סיפרה", שהיא המכונה, מגלמת את הצד המוסרי האבסולוטי ואילו ה"סיפור", שהוא האדם, מנסה לכפות מוסר מעוות.

עתידה של האנושות תלוי בטכנולוגיה מומחים רבים קושרים את עתידו של תחום ה-AI בפתירת השאלות האתיות שהוא מציב ולעיתים אף דוחקים בפורצי הדרך הטכנולוגיים להאט את פעולותיהם. לדידי, זו גישה שאינה בת יישום לאורך זמן והתחבטות מעניינת שנעשתה לפני כמעט 60 שנה על ידי מי שמוכר יותר כמשורר לאומי יכולה לתמוך בגישה שכשעוסקים בטכנולוגיה כל המרכבה הרי זה משובח. כאפילוג אציין שהאיש שהקסט שלו רשמה יצירות מכונות כ"כוכבים בחוץ" ו"שמחת עניים", אשר ידע ללהטט בשפה העברית, תוך שילוב המקורות היהודיים והעברית המתחדשת, הכיר בעובדה שהאנושות הולכת בדרך הטכנולוגיה. הוא הניח שעתידיה של האנושות, לטוב ולרע, קשור בטכנולוגיה פורצת דרך. בכך הוא הסיק שהטכנולוגיה, שנתפסה אז ככלי שרת בידי האדם, תהפוך בעתיד להיות שותפתו הנאמנה של האדם לדרך. מהפיכת ה-AI מוכיחה עד כמה ראייתו הייתה מפוכחת ונכונה.

הכותב הוא פרופסור להנדסת חשמל באוניברסיטת תל אביב, מסציא ויוזם סדרתי המכהן כסגן נשיא סמסונג. שימש כסגן הרואי במשרד הפדע בין השנים 2008 ל-2010. מחקריו עוסקים באלקטרוניקה, צילום ממוחשב, עיבוד מידע אופטי, עיבוד וזיהוי אותות והולוגרפיה

Open AI בגין חששות עמיתיו שהוא מוביל את הפיתוח בקצב נמרץ מדי.

### להטיל הגבלות על מכונות

במאמר זה ארצה להציג זווית אחרת שטוענת שאין זו המכונה שמגלה רוע, אלא האדם שבנה והפעיל אותה. ארצה לגלול את שעון הזמן אחורה לשנת 1965, כלומר לפני כמעט 60 שנה. אז עלה על הבמה של תיאטרון הכימה המחזה "משפט פיתגורס", שנכתב על ידי נתן אלתרמן. המחזה, שנחל כישלון חרוץ, מתאר מכונת אינטגרציה מלאכותית שמנהלת על ידי יוצרה. המכונה ששמה פיתגורס (בקיזור "פית"), פענחה רצח שלא פוענח עד אז בזכות האינטליגנציה שהקנה לה המנהל שבנה והפעיל אותה. המנהל הופתע מהתגלית שהעמידה אותו בדילמה מוסרית והחליט למנוע מ"פית" לפרסם את התגלית אודות זהות הרוצח. לכן הוא ניסה לתכנת מחדש את "פית" שתשכח את התגלית. "פית", כמכונה צייתנית לחוקי הפעלתה המקוריים (הקוד המוסרי הטהור שהוטמע בה), לא הסכימה להתעלם מהתגלית המצביעה על הרוצח כיוון ש"רשת הניורונים" שמנחים אותה רואה את "המורס" באופן שונה אשר מחייב לגלות את הפענוח ולהעניש את הרוצח. הוויכוח עם המנהל נמשך ולבסוף המנהל החליט להרוס את המכונה במה שהעניק למחזה את שמו - "משפט פיתגורס". סיומו של המחזה הוא הריסת "פית" על ידי המנהל, כשמילותיה האחרונות של המכונה למנהל היו: "סכום ריבועי צלעותיו הקצרות של משולש ישר וווית..." (משפט פיתגורס הגיאומטרי), רוצה לומר, ש"פית" הצהירה שהיא מכונה שנאמנה לקוד שהוטמע בה - "הסיפרה" - ולא מוכנה לקבל את מוסר המנהל.

המחזה מאיר את הדילמה שבין ה"סיפרה", שזו תוצאת הקוד המתמטי של המכונה, לבין ה"סיפור", שמבטא את המוסר (המעוות?) של המנהל. זו לדעתי התייחסות מפתיעה, שפורסמה טרם זמנה, לגבי הדילמה של מורסריותו של האדם לעומת האתיקה של המכונה. הימים היו ימי אורול עם סיפרו "1984" ומכונות נחשבו לסכנה גדולה. התפתח חשש גדול ממכונות והיו כאמור קריאות להטיל הגבלות על מכונות. "משפט פיתגורס" העלה תזוה שהמכונה לא תהיה יותר רעה מהרוע של זה שבנה אותה ולכן את מאמציינו להגביל אתית אין אנו צריכים לנתב לניתוח המכונות, אלא לפיקוח על הבונים והמפעילים שלהן.

כאמור, המחזה נכשל והוצג 13 פעמים בלבד. המבקרים טענו שהמחזה אינו ממוקד ודן ביוטי דברים ולא בעיקרם. למשל, המבקר

"הגביע הקדוש של תחום ה-AI הוא ליצור מערכת המתכללת את כל המשימות בליבת חשיבה טכנולוגית אחת ובכך להשיג מערכות פשוטות, זולות וקטנות, המסוגלות להניב ביצועים פורצי דרך"

ציבורי ושיח תקשורתי. נשמעו קריאות שמוזכרות גם היום להאט ואף עצירת שכלול לים טכנולוגיים מהסוג של AI מפאת הסכנות הטמונות בהם. מהעת האחרונה זכורה לכולנו ההדחה (שבוטלה) של סם אלטמן מניהול

יכולת להיות יותר אינטליגנטיות מהאדם, נראות בעיני רבים כסכנה מהותית. הדיון על יכולתה של המכונה לרוע רוע דובר רבות, כאשר הדיונים הראשונים היו בעולמות המדע הבדיוני. כבר בשנות ה-50 של המאה הקודמת, מכונת "האח הגדול" ב-1984 של ג'ורג' אורול הייתה נושא פופולרי לדיון