



נשגב מבינתי? מושגי יסוד בבינה מלאכותית

מערכות בינה מלאכותית החלו כבר לפני שנים רבות יחסית לבצע בהצטיינות משימות "כמו אנושיות", כגון משחק שח-מט או פתרון שאלות טריוויה, וכלי בינה מלאכותית הוטמעו במוצרים טכנולוגיים שונים. עם זאת, פיתוחם של מודלי בסיס ומודלי שפה גדולים (ראו להלן) בשנתיים האחרונות והשימוש בהם במגוון יישומים הובילו לפריצת דרך משמעותית ביכולות של מערכות בינה מלאכותית וחוללו שינויים ניכרים באינטראקציה אדם-מחשב, בהיקף המשתמשים ובתפיסה הציבורית של מערכות אלה. להלן יוצגו בקצרה מושגי יסוד בתחום הבינה המלאכותית.



מושגי יסוד

+ **בינה מלאכותית (AI – Artificial Intelligence):** מונח זה הוא שם כולל למערכות המאפשרות לבצע משימות שבעבר נחשבו כאלה שרק אדם תבוני מסוגל לבצען, כגון זיהוי דיבור, ניתוח תמונה, תרגום או קבלת החלטות.



מועצת אירופה הגדירה בינה מלאכותית כ"מערכת המכגינה התנהגות תבונית באמצעות ניתוח סביבתה ונקיטת פעולות במידה מסוימת של אוטונומיות, כדי להשיג מטרה מסוימת". מערכות בינה מלאכותית יכולות להתבסס על תוכנה בלבד ולפעול במרחב המקוון, ודוגמאות לכך הן עוזרים קוליים, מנועי חיפוש או מערכות זיהוי פנים; או להיות מוטמעות בהתקנים פיזיים – כלי רכב אוטונומיים, רחפנים, רובוטים מתקדמים ועוד. מקובל להבחין בין בינה מלאכותית צרה לבין בינה מלאכותית כללית:

- **בינה מלאכותית צרה/חלשה (Narrow AI):** מערכת בינה מלאכותית המיועדת למשימה ספציפית, כגון משחק שחמט, תרגום או פענוח תמונה.



- **בינה מלאכותית כללית/חזקה (AGI – Artificial General Intelligence):** מערכת בינה מלאכותית בעלת מיומנות במגוון משימות תבוניות מסוגים שונים. ההנחה בעבר הייתה שמערכות כאלה הן תיאורטיות בלבד, אך קפיצת המדרגה בביצועיהם של מודלי שפה גדולים מסמנת התקדמות בכיוון של ריבוי משימות בתחומים שונים.



השימוש בבינה מלאכותית מוטמע ביותר ויותר תחומים, בשלב זה בעיקר כמערכות מסייעות – תומכות החלטה. למשל, בהקשר הרפואי נעשה שימוש בבינה מלאכותית לפענוח תוצאות דימות, בתהליכי אבחון וטיפול רפואי ולפיתוח תרופות; בהנדסת תוכנה – לכתיבה או לבדיקת קוד; בכתיבה יוצרת – ליצירת תוכן טקסטואלי כגון שירים ותסריטים באופן שלא ניתן לקבוע, ללא כלים מסייעים, אם התוכן נוצר באמצעות תוכנה או בידי אדם (הקריאות להסדרה והגבלה של השימוש ביישומי בינה מלאכותית היו גם בין הסוגיות שלאחרונה גררו שביתה של ארגון התסריטאים בארה"ב); בתחום התוכן החזותי – ליצירת תמונות וסרטונים; בתחום שירות לקוחות ותהליכי מכירה אוטונומיים – לפיתוח שירותים המבוססים על סוכן לא אנשי שעומו הלקוח יכול להתכתב (צ'טבוט) או לשוחח.

ביצירת מערכות בינה מלאכותית משתמשים בין היתר באלגוריתמים ובשיטות שונות ללימוד מכונה הכוללות הזנת המודל שבבסיס המערכת במידע לשם אימון שלו.

+ **אלגוריתם:** סדרה של הנחיות או חוקים מתמטיים המשמשים לשם פתרון בעיות, בפרט באמצעות מחשב. כיום משתמשים באלגוריתמים ממוחשבים בעלי יכולת למידה עבור שלל פעולות – ייעוץ, מיון, המלצה והתאמה אישית, באינטרנט ומחוצה לו. אלגוריתמים הם מאבני הבניין המרכזיות של מערכות בינה מלאכותית.



+ **למידת מכונה (ML – Machine Learning):** שיטה ללימוד של מערכת מחשוב בלי להגדיר לה את מכלול הכללים והתנאים באופן מפורש, אלא בהתבסס על הזנתה בכמויות גדולות של מידע, שממנו היא לומדת לזהות דפוסים חוזרים במידע ולהציע תחזיות בהתבסס עליו.



הזיות של בינה מלאכותית: מונח מדעי מקובל לתיאור מקרים שבהם מערכות בינה מלאכותית מספקות תוצאה, כגון טקסט, תמונה או תוכן אחר, שהוא חסר משמעות, מוטעה או לא רלוונטי בהקשר הנדון. "שגיאות" כאלה פוגעות במהימנות של מערכות אלה ובציפיות המשתמשים מן המערכות; זהו אחד מן האתגרים בהבטלתן של טכנולוגיות אלה לכדי מוצרים אמנים.



- **למידה עמוקה (DL – Deep Learning):** למידה עמוקה היא סוג של למידת מכונה, שעיצובה מתבסס על המבנה והפונקציות של המוח האנושי ועל האופן שבו המוח מעביר מידע. למידה עמוקה מתבססת על **רשתות נוירונים מלאכותיות** מרובות שכבות (ומכאן "עמוקה"), שאליהן מוזן מידע ובהן הוא מעובד, מיוצג במספרים ומפוענח לכדי פלט. ככל שהרשת עמוקה ורחבה יותר וכוללת יותר קשרים, היא יכולה לעבד מידע מורכב יותר ולפתור בעיות מורכבות יותר.



- **למידת חיזוק (RL – Reinforcement Learning):** שיטת למידת מכונה המתבססת על בדיקת התוצאות (פלט) של המערכת ומתן משוב חיובי או שלילי ("חיזוק") בהתאם לדיוק של המערכת. המערכת לומדת להסיק מהתגובות מסקנות בדבר פעולותיה ולשפרן בהתאם.



+ **בינה מלאכותית יוצרת (Generative AI):** שם כולל ליישומים המבוססים על מודלי בסיס (ראו להלן) אשר מאפשרים, באמצעות שאילתה קצרה, ליצור תוכן חדש – טקסטים, תמונות, סרטונים, מוזיקה, קוד מחשב ועוד. התוכן החדש מופק בהתבסס על מידע אימון (training data) שהוזן למערכת. יישומים מוכרים בתחום כיום הם למשל ChatGPT ליצירת טקסט ו-Midjourney ליצירת תוכן חזותי.



+ **מידע אימון (Training Data):** מידע המוזן למערכות בינה מלאכותית כדי ללמד אותן או לאפשר להן ללמוד. המידע יכול להיות מתויג ולשמש לשם **"למידה מפורקת"** (כגון זיהוי תמונות של זן של כלב בהתבסס על תמונות מתיוגות של סוגי כלבים לשם אימון המערכת); או לא מתויג, ולשמש ל**"למידה לא מפורקת"**, שבה המערכת לומדת לזהות במידע דפוסים באופן עצמאי ולהסיק ממנו מסקנות. ככלל, ככל שמידע האימון מכורט, מדויק ומקיף יותר, היכולות של המערכת משתפרות. לעומת זאת, הטיות וטעויות במידע האימון יכולות להוביל לטעויות מערכת עד כדי אפליה אלגוריתמית.



+ **אפליה אלגוריתמית (Algorithmic Discrimination/Bias):** מצב שבו אלגוריתם מזהה, במישרין או בעקיפין, מידע על משתני רקע מסוימים, כגון מגדר, גזע או נטייה מינית, ומפיק תוצאה מפלה או מוטה על בסיס משתנים אלה. אפליה כאמור יכולה להתרחש עקב הטיות במידע האימון או בשל בעיות בעיצוב

המערכת. חדירתן של מערכות בינה מלאכותית לחיי היום-יום ולתהליכי קבלת החלטות מעוררת שאלות בדבר אפליה שעלולה להיווצר בגינן, למשל במערכת מיון מועמדים לעבודה, במערכת תומכת החלטה על הקצאת אשראי או אפילו במערכות תומכות החלטה בבתי משפט.



+ **עיבוד שפה טבעית (NLP – Natural Language Processing):** תת-תחום של בינה מלאכותית העוסק בפיתוח היכולת של מחשבים להבין שפה אנושית. הקושי של מחשבים בהבנת שפה טבעית נובע בין השאר מן השונות באופני ההתבטאות – כתיבה או דיבור של בני אדם – ומן הצורך להבין הקשר ספציפי, לזהות את המסר המובלע של הטקסט ועוד.

השפעות והשלכות על תעסוקה:

בניגוד להנחה המקובלת כי מגמת האוטומציה ישיעו בעיקר על עובדים בעבודות כפיים, ייתכן שההשפעה הגדולה של ההתפתחויות בתחום הבינה המלאכותית תהיה דווקא על בעלי מקצוע בתחומים אחרים – תסריטאים, גרפיקאים, מתכנתים, מתרגמים ועוד. לפי **טיטט מחקר** ממרץ 2023, מקצועות בשכר גבוה חשופים יותר ממקצועות בשכר נמוך להשלכות של חדירתן של טכנולוגיות בינה מלאכותית. לפי המחקר, בכ-19% מן המשרות צפוי שהבינה המלאכותית תשפיע על כ-50% מן המשימות של העובדים, וב-80% מן המשרות צפויה השפעה על כ-10% מן המשימות של העובדים. עם זאת, לפי **סקר מקיף** של הפורום הכלכלי העולמי (WEF) ממאי 2023, שיעור המשרות החדשות שיווצרו בטווח הקצר, בשנים 2023-2027, צפוי להיות גדול משיעור המשרות שייעלמו, כך שהמאזן הכולל המשוער יהיה חיובי.



+ **מודלי בסיס (Foundation Models):** מודל בינה מלאכותית הוא מכלול של עקרונות וכללים המיוצגים במספרים, ויכולים לשמש לחיזוי או לקבלת החלטות בהתבסס על הסקה ממידע אימון שהוזן למערכת. המונח מתאר את המעבר ממודלים ליישומים ספציפיים למודלים חזקים (בעלי כוח חישוב חזק) וגדולים (המתבססים על כמויות גדולות של מידע אימון), המסוגלים לבצע משימות שונות. מודלי בסיס מאומנים באמצעות מידע בהיקף עצום בתחומים שונים, והם מסוגלים, באמצעות למידה עצמית של המערכת, לבצע משימות שונות. מודלי שפה גדולים הם סוג מסוים של מודלי בסיס.



+ **מודלי שפה גדולים (LLM – Large Language Models):** מודלים המוזנים בכמויות עצומות של מידע לא מתויג, מאומנים לנתח אותו ומסוגלים להשתמש בו כדי לבצע משימות שונות. שלא אומנו להן באופן ספציפי, וכך ליישם ידע שלמדו בהקשר אחד גם בהקשרים אחרים.

בעיית הקופסה השחורה: מערכות בינה מלאכותית מסיקות מסקנות, ממליצות על פעולות או מחלטות החלטות בהתבסס על המידע המוזן לתוכן. אך בשל הכמויות העצומות של המידע ואופן פעולתן של המערכות, לעיתים קרובות המערכת ומעצביה אינם יכולים להסביר את השיקולים שהובילו להמלצה או לניבוי שלה. הקושי לנמק או לתאר את מכלול הסיבות להחלטה גורם לפירוש פעולת המערכת כ"קופסה שחורה". בינה מלאכותית הסברתית – Explainable Ai היא ציפייה, עיקרון אתי ולעיתים דרישה רגולטורית שאפשר יהיה להסביר החלטות מכונה למי שהחלטות אלה משפיעות עליהם.

מקורות

- משרד החדשנות, המדע והטכנולוגיה ומשרד המשפטים, [עקרונות מדיניות, רגולציה ואתיקה בתחום הבינה המלאכותית](#), 2022.
- עמיר כהנא ותהילה שוורץ אלטשולר, [אדם, מכונה מדינה](#), המכון הישראלי לדמוקרטיה, 2023.
- רועי גולדשמידט, [בינה מלאכותית יוצרת: הזדמנויות, סיכונים ורגולציה](#), מרכז המחקר והמידע של הכנסת, 7 ביוני 2023.
- רועי גולדשמידט, [אפליה אלגוריתמית במערכות המבוססות על בינה מלאכותית](#), מרכז המחקר והמידע של הכנסת, 14 ביוני 2020.
- European Parliamentary Research Service, Scientific Foresight Unit (STOA), [AI Definition & Glossary Key Terms and Concepts](#), February 2022.
- Bommasani, R. et al. [On the Opportunities and Risks of Foundation Models](#). Preprint.
- James Tobin, [Artificial Intelligence: Development, Risks and Regulation](#), UK Parliament, House of Lords Library, July 18, 2023.
- Christopher Manning, [Artificial Intelligence Definitions](#), Stanford University, September 2020.