

בעולם משתנה, שהבינה המלאכותית הולכת ומשתלבת בו בתחומים רבים, גוברת הדרישה לפיתוח כישורים מתאימים, גם בקרב משפטנים. מאמר זה נועד לספק טיפולוגיה קצרה וברורה של מושגים מרכזיים בתחום הבינה המלאכותית, ולאפשר הבנה בסיסית אך מהותית גם למי שאינם מומחי טכנולוגיה. המאמר מציג מושגים מרכזיים בתחום כגון "למידת מכונה", "בינה מלאכותית יוצרת" ו"מחזור החיים של נתונים". בנוסף, המאמר מסביר כיצד פועלים מודלים של למידה עמוקה, כמו ChatGPT. הוא מתאר את ההזדמנויות לשיפור ביצועים, לצד הסיכונים האתיים והרגולטוריים, את השפעת הטכנולוגיה על חוויית המשתמש והערך הייחודי שהיא מציעה, וקורא לשיח רב-תחומי בין משפטנים למדענים לצורך שימוש אחראי בטכנולוגיה זו.

מבוא

הבינה המלאכותית היוצרת פרצה לחיינו ב-2022, עם כלי בשם ChatGPT, תוצרת חברת OpenAI, אשר מייצר טקסט באופן הדומי (אינטראקטיבי) כחלק מדיאלוג עם אנשים. חוויית המשתמש בתקשורת עם הכלי מדמה לרבים שיחה מול סוכן אנושי ולכן התקבעה בקרב הציבור הרחב התחושה שמדובר במהפכה של ממש בתחום הבינה המלאכותית. תחושה זו מלווה לרוב בחששות מפני השתלטות הבינה המלאכותית על חיינו, ולא דווקא באופן חיובי.

מאמר זה נועד להציע טיפולוגיה למונחים חשובים בעולם התוכן המתפתח ולהגדיר מהי בינה מלאכותית, מהי למידת מכונה, מהו מחזור החיים של נתונים ומהי בינה מלאכותית יוצרת. הבנת המושגים הבסיסיים והאופן שהבינה המלאכותית פועלת בו תסייע רבות ביכולת לשלב אותה באופן בונה בחברה המודרנית.

1. מהי בינה מלאכותית?

המונח "בינה מלאכותית" הוא רחב למדי, ומתייחס למגוון טכנולוגיות, מאלגוריתמיקה פשוטה ועד מערכות מורכבות המסוגלות לחקות תהליכי חשיבה אנושיים. ישנן הגדרות רבות, אך לרוב בינה מלאכותית מוגדרת כבינה לא אנושית הנמדדת ביכולותיה לשכפל מיומנויות אנושיות כגון זיהוי דפוסים, למידה מסתגלת (אדפטיבית) מניסיון

* פרופסור למדעי הנתונים, הטכניון – מכון טכנולוגי לישראל. מחזיק בקתדרה על שם בנימין ופלורנס פרי. אני מודה לחברי מערכת "משפט, חברה ותרבות" ולקורא האנונימי על הערות והצעות מועילות.

ואסטרטגיה.¹ כחלק מעבודתן והליך יצירתן מערכות בינה מלאכותית לומדות, פותרות בעיות ומזהות תבניות. יישומים של בינה מלאכותית נבנים מתוך מטרה לשפר ביצועים של מערכות קיימות או לספק כלים ממוכנים מהירים ומדויקים כתחליף למטלות שבעבר בוצעו בידי בני אנוש. באופן אידיאלי, המודל ידע להשיג תוצאות או תחזיות טובות ואמינות.² כיום, בינה מלאכותית מוזכרת לרוב בהקשר תת-התחום של למידת מכונה ורשתות נוירונים עמוקות (מונחים אשר יפורטו בהמשך המאמר).

יש הטוענים כי בשנת 2022 התרחשה "המהפכה התעשייתית הרביעית" בזכות פריצת הבינה המלאכותית לתודעה הציבורית בעקבות הצלחתו המסחרית של ChatGPT.³ אולם, מחקרים בתחום הבינה המלאכותית החלו עוד בשנות החמישים של המאה העשרים.⁴ כבר אז עסקו חוקרים במידול ויצירת אלגוריתמים אשר מדמים התנהגות אנושית. תת-תחומים אשר התפתחו במסגרת מחקר בבינה מלאכותית כללו הסקה, ייצוג ידע, תכנון, למידה, עיבוד שפה טבעית ורובוטיקה.⁵ כלים חשובים אשר סייעו בידי החוקרים כללו חיפוש יעיל במרחב מצבים גדול, אופטימיזציה ולוגיקה מתמטית, בעוד שהתובנות אשר שימשו כבסיס למודלים הגיעו מתחומי הפסיכולוגיה, הבלשנות, הפילוסופיה ומדעי המוח.⁶

התפתחות הבינה המלאכותית בשנים האחרונות הובילה למעין "מירוץ חימוש" של פיתוחי בינה מלאכותית בין ענקיות הטכנולוגיה העולמיות. אומנם זאת בתורה הביאה לשיפור הדיוק והכלליות של הכלים, אך חלק מחברות הטכנולוגיה הגדולות, כמו גוגל ומטא, ספגו ביקורת משום שהמידע שהציגו היה לעיתים קרובות מוטע או שגוי.⁷

- 1 STEPHAN DE SPIEGELEIREET ET AL. ARTIFICIAL INTELLIGENCE AND THE FUTURE OF DEFENSE (2017).
- 2 FUTURE OF PRIVACY FORUM, PRIVACY EXPERT'S GUIDE TO ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING (2018).
- 3 Eric Rosenbaum, *MIT research on ChatGPT shows 'Industrial Revolution-level large' leap for workers, says AI CEO*, CNBC (May 10, 2023), <https://www.cnbc.com/2023/05/10/mit-data-show-industrial-revolution-level-leap-for-workers-using-ai.html>.
- 4 דוגמה לשימוש של בינה מלאכותית ניתן למצוא בחקר החלל, שם הותקנה מערכת של תכנון אוטומטי ברכב חלל אשר נחת על המאדים. מטרת המערכת הייתה לתכנן מסלול תנועה של הרכב באזור שטרם מופה. תכנון שכזה מבוסס על הבנת המסלול שנעשה עד כה ובחירה מבין מסלולים עתידיים אפשריים תוך הנחת הנחות סבירות לגבי מה צפוי לרכב בהמשך תנועתו. ראו Richard Castaño & Tara Estlin, *OASIS: Onboard autonomous science investigation system for opportunistic rover science*, J. OF FIELD ROBOTICS, (2007).
- 5 CHRIS SMITH ET AL., THE HISTORY OF ARTIFICIAL INTELLIGENCE (2006).
- 6 שם.
- 7 Jennifer Elias, *Google execs warn company's reputation could suffer if it moves too fast on AI-chat technology*, CNBC (Jun. 11, 2024); Leigh Mc Gowran, *Meta's 'biased' science-writing AI demo gets pulled after three days*, SILICON REPUBLIC

בנובמבר 2022 שחררו מיקרוסופט וחברת OpenAI את ChatGPT, שהיה פחות מוטה מקודמיו, בהיותו מודל שפה גדול.⁸ חמוש בכמויות עצומות של נתונים, ChatGPT ידע לענות על שאלות באופן אשר הדהים אפילו את המומחים בתחום.⁹ יש לציין כי הבינה המלאכותית איננה מוצר או שירות העומד בפני עצמו, אלא סוג של "מערכת הפעלה" המסוגלת לשפר ביצועים ולדמות תהליכי מחשבה אנושיים.¹⁰ כך, הבינה המלאכותית מוטמעת בטכנולוגיות חדשות כמו מכונות אוטונומיות וטלפונים חכמים כדי לאפשר מטלות חדשניות, ומיושמת גם בתעשיות מבוססות כמו חקלאות ותקשורת כדי להגדיל דיוק ויעילות.¹¹ יתרה מכך, יש הטוענים כי למרות ההתקדמות המרשימה בתחום הבינה המלאכותית בשנים האחרונות, במועד הפרסום התחום עדיין רחוק מלהגיע לבינה מלאכותית כללית (AGI).¹² הגדרה זו מתייחסת למערכות בינה מלאכותית שמסוגלת לבצע משימות, כולל למידה והבנה, ברמה אנושית ואף גבוהה יותר. בדומה לתבונה אנושית, בינה מלאכותית כללית נחנה גם ביכולת הסתגלות למצבים שלא תוכננה אליהם מראש.¹³ כך, ניתן לראות כי מערכות הבינה המלאכותית הנוכחיות מצטיינות במטלות ספציפיות שעבורן הן אומנו, אך נכשלות כשמדובר בהתמודדות עם מצבים חדשים או עם דרישה להבנה רחבה יותר של הקשר ומשמעות. כך לדוגמה, מודלים רבים מתקשים בהבנת הומור אשר מצריך שליטה בהקשרים תרבותיים והבנת רגש וקצב.¹⁴ בנוסף, כאשר משתמש שואל שאלה עמומה או לא ברורה, ChatGPT עשוי להיתקל בקושי בהבנת ההקשר המלא, ובשל כך המערכת עשויה לפרש את השאלה באופן שגוי או להפיק תשובה חלקית בלבד, מאחר שהיא אינה "מבינה" את המידע אלא מחקה את

-
- (Nov. 22, 2022). מאוחר יותר רשמו כלי ה-AI של גוגל שיפור ניכר בדיוק, אך בעיית ההטיה נותרה על כנה.
- Erik Weber et al., *Is GPT-4 Less Politically Biased than GPT-3.5? A Renewed Investigation of ChatGPT's Political Biases*, ARXIV (Oct. 28 2024) 8
- Stephen Wolfram, *What Is ChatGPT Doing ... and Why Does It Work?*, STEPHEN WOLFRAM WRITINGS, (Feb. 14, 2023) 9
- LILIAN EDWARDS, REGULATING AI IN EUROPE: FOUR PROBLEMS AND FOUR SOLUTIONS 10 (2022).
- GUIDE TO ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING, לעיל ה"ש 2. 11
- Sébastien Bubeck et al., *Sparks of Artificial General Intelligence: Early Experiments with GPT-4*, ARXIV (Mar. 22, 2023), <http://arxiv.org/abs/2303.12712v5> 12
- Henry Shevlin et al., *The limits of machine intelligence*, EMBO REPORTS, 13 PMC6776890, <https://pmc.ncbi.nlm.nih.gov/articles/PMC6776890>
- Farah Shatnawi et al., *Comprehensive Study Of Pre-Trained Language Models: Detecting Humor In News Headlines*, 27 DATA ANALYTICS & MACH. LEARNING 14 .2575 (2022)
-

הדפוסים הנפוצים בתשובות לשאלות דומות.¹⁵ למעשה ישנה טענה כי מדובר ב"תוכי מתוחכם" החוזר על עוד ועוד מילים מורכבות מבלי להבין אותן באמת.¹⁶

2. מהי למידת מכונה?

תחום המחקר של למידת מכונה הוא תת-תחום של בינה מלאכותית אשר בוחן את האופן שבו ניתן לבנות מודלים מתצפיות. התחום נוצר אף הוא בשנות החמישים של המאה העשרים, והמטלות העיקריות אשר נחקרו במסגרתו כללו פתרון בעיות סיווג, קיבוץ וחיזוי.¹⁷

סיווג הוא תהליך שבמסגרתו נבחן האם האובייקט שייך לקטגוריה מסוימת.¹⁸ קיבוץ הוא תהליך שבמסגרתו קבוצה של אובייקטים מחולקת לקבוצות שונות לפי מאפיינים. כך, אובייקטים דומים מגיעים לאותה קבוצה ואובייקטים אשר אינם דומים מופרדים לקבוצות שונות.¹⁹ במסגרת שלב החיזוי נדרש המודל לנבא מה התוצאה שתתקבל בתגובה לנתונים עתידיים שיוזנו אליו.²⁰ כך, למידה מתצפיות היא תהליך אינטראקטיבי שבו אלגוריתם מקבל דוגמאות ומייצר פלט אשר מטרתו להביא למינימום פונקציית עלות (או למקסימום פונקציית רווח).²¹ לדוגמה, אלגוריתמי למידה מפוקחת מקבלים דוגמאות (למשל, תמונות של כלבים וחתולים) ומטרתם לחזות תווית לכל דוגמה (למשל, כלב או חתול). לאחר פעולת החיזוי, האלגוריתם מקבל חיזוי על הצלחתו ומנסה להשתפר מחזור (איטרציה) לחזור.

כחלק מהלמידה מתצפיות, אלגוריתמי למידה מקבלים כקלט היבטים מגוונים של התצפיות שנקראים מאפיינים (features). כך, כאשר האלגוריתם לומד לזהות עצמים בתמונה, מאפיינים יכולים להתקבל מפירוק התמונה לרכיבים קטנים שנקראים פיקסלים, כאשר כל פיקסל מאופיין על ידי צבע יחיד ומיקומו בתמונה. הנדסת

- 15 Walid Hariri, *Unlocking the Potential of ChatGPT: A Comprehensive Exploration of its Applications, Advantages, Limitations, and Future Directions in Natural Language Processing* ARXIV (Apr. 6, 2023), <https://arxiv.org/abs/2304.02017>
- 16 Emily M. Bender et al., *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*, 64 ASSOCIATION FOR COMPUTING MACHINERY 610 (2021)
- 17 ROBERTO CORDESCI, *THE DISCOVERY OF THE ARTIFICIAL* 187–89 (2002)
- 18 STUART RUSSELL & PETER NORVIG *ARTIFICIAL INTELLIGENCE: A MODERN APPROACH* 353 (4th ed. 2020)
- 19 SHAI SHALEV-SHWARTZ & SHAI BEN-DAVID, *UNDERSTANDING MACHINE LEARNING: FROM THEORY TO ALGORITHMS* 307 (2014)
- 20 שם, בעמ' 288.
- 21 Timothy B Lee & Sean Trott, *Large Language Models, Explained With A Minimum Of Math And Jargon*, UNDERSTANDING AI (Jul. 27, 2023)

מאפיינים (feature engineering) הוא השם המקובל לתהליך של יצירת מאפיינים משמעותיים מתוך תצפיות.²² באלגוריתמים מוקדמים של למידת מכונה הלמידה הייתה מוכוונת מטרה. כלומר, ראשית הוצגה שאלה למחקר ולאחר מכן אומן מודל תוך שימוש בנתונים רלוונטיים להשגת המטרה. בנוסף, סיווג הנתונים היה ידני (כלומר, על ידי בני אדם) ולכן תהליך יצירת בסיסי נתונים גדולים היה תהליך קשה ויקר.²³ כיום, ריבוי הנתונים וההבנה כי ניתן לייצר מודלי למידה יסודיים (generic) תוך שימוש בלמידה בהנחיה עצמית (self-supervised learning) מאפשרים יצירת מודלי תשתית (foundation models) אשר מאומנים על מטלות כלליות לצורך יצירת ייצוג של ישויות אשר בעתיד תשמשנה לביצוע מטלות חדשות ומגוונות.²⁴ מודלי שפה גדולים מודרניים משתמשים בקטעי טקסט רגילים (כמעט כל חומר כתוב אפשרי, מוויקיפדיה ועד קוד מחשב) כדי לחזות מהי המילה הבאה. ככל שהמודל מאומן על יותר טקסטים, משקולות המודל מותאמות בהדרגה על מנת ליצור תחזיות טובות יותר. כל החלקים השונים של מודלי שפה גדולים מיושמים כשרשרת של פונקציות מתמטיות פשוטות שהתנהגותן נקבעת על פי הפרמטרים של משקולות המודל. אלגוריתם האימון מגדיל או מקטין את הפרמטרים של המשקולות כדי לשלוט באופן שבו המידע מחושב בעודו זורם דרך רשת הנוירונים.²⁵

הדבר דומה לאופן שבו ילדים צוברים ידע כללי על העולם על מנת שיוכלו להשתמש בו בעתיד לאתגרים שיעמדו בפניהם כאשר יתבגרו. לדוגמה, כך ילדים נחשפים לעולם דרך חוויות, שפה ואינטראקציות. החשיפה גורמת להם להבין כללים וחוקים באמצעות משחק וניסוי וטעייה. בצורה דומה, גם ChatGPT מסוגל כיום לבצע מגוון פעולות מורכבות. הסיבה המרכזית ליכולותיו המרשימות היא שהמודל נחשף לכמויות בלתי נתפסות של מידע, מעל 500 מיליארד מילים. לשם השוואה, אדם סביר נחשף לעשרות מיליוני מילים עד גיל עשר.²⁶ במהלך חמשת השנים האחרונות הגדילה חברת OpenAI בהתמדה את גודל המודל, וכתוצאה מכך הדיוק שלו השתפר במידה

- 22 Mainak Sarkar & Arnaud De Bruyn, *LSTM Response Models for Direct Marketing Analytics: Replacing Feature Engineering with Deep Learning* 53 J. INTERACTIVE MKTG. 80 (2021).
- 23 Alexander L Fradkov, *Early history of machine learning*, 53 IFAC-PAPERSONLINE 1385 (2020).
- 24 Ruidan He et al., *On the Effectiveness of Adapter-based Tuning for Pretrained Language Model Adaptation*, ACL 2208 (2021).
- 25 Bender et al., *לעיל ה"ש* 16.
- 26 Hart B. & Risley T. R., *American Parenting Of Language-Learning Children: Persisting Differences In Family-Child Interactions Observed In Natural Home Environments*, 28 DEV. PSYCH. 1096 (1992).

ניכרת.²⁷ בכל חידוש שכזה, המודל מאומן על יותר נתונים מקודמיו ואף מבצע טוב יותר משימות הדורשות חשיבה מופשטת ושימוש בקוגניציה (Theory of Mind).²⁸ ניתן היה לחשוב שלמידת מכונה מחקה באופן מדויק את האופן שבו אנשים לומדים מתצפיות. ייתכן מאוד שכך הוא המצב, אבל אין כיום מחקר, למיטב ידיעתי, אשר מראה זאת. לכן, התבססות על אמונה כזו יכולה להוליך את מקבל ההחלטות שולל. התחום אשר נקרא למידת מכונה יריבנית (adversarial machine learning) יכול לסייע בהמחשת הטענה. מדובר בשיטה אשר מטרתה לשטות במודל של למידת מכונה תוך שימוש בשינוי מינימלי של הקלט. לדוגמה, ציפור אשר מודל למידה סיווג אותה כברווז ולאחר ביצוע שינוי קטן בתמונה, שינוי אשר לא ניתן להבחין בו בעין בלתי מזוינת, הסיווג השתנה לבעל חיים אחר (סוס).²⁹ בבואנו להתמודד משפטית וחוקית עם למידת מכונה, שומה עלינו לזכור כי לא מדובר במערכת מחשב אשר רואה (או שומעת, או קוראת) באופן שבו אנו רואים, שומעים או קוראים, או תופסת את העולם באותו אופן כמונו.

כדי להבין את הליך הלמידה של מודלים אלו וקפיצת הדרך שעשו בשנים האחרונות, יש להבין שמודלי שפה גדולים ובני אדם "מייצגים" מילים באופן שונה. בני אדם "מייצגים" מילים באמצעות רצף אותיות (למשל רצף האותיות מ-ז-ל-ג מייצג את המילה מזלג), ואילו מודלי שפה "מייצגים" מילים באמצעות רצף מספרים בשם "וקטור ייצוגים" (Word Vector). "וקטור ייצוגים" מייצג נקודה בתוך "חלל מספרי" רב-ממדי דמיוני, כאשר מילים בעלות משמעות מקבלות נקודות קרובות זו לזו בחלל הזה. ניתן להשוות את ההיגיון בייצוג זה לייצוג מיקומים במפה באמצעות נקודות ציון.³⁰ הרעיון של "וקטור ייצוגים" התפרסם במיוחד ב־2013 כאשר גוגל הכריזה על פרויקט בשם word2vec, במסגרתו ניתוח מיליוני מסמכים על מנת להבין אילו מילים עולות במשפטים דומים.³¹ בדרך זו, המערכת למדה "למקם" מילים דומות קרוב זו לזו בחלל הווקטורי. מעבר לכך, המערכת ידעה גם לייצר אנלוגיות. למשל, היא ידעה לקשר בין רומא לאיטליה או בין נסיך לנסיכה. חשוב לציין שמודלים אשר מבוססים על ייצוג וקטורי אינם מודלים מספריים טהורים והם עשויים לשמר הטיות בתפיסה האנושית אשר המודל למד מהדוגמאות שאליהן הוא נחשף. למשל, ייתכן שמודל שפה יידרש

James O'Donnell., *Why OpenAI's New Model Is Such a Big Deal* MIT TECH. REV. 27 (Sept. 17, 2024), <https://www.technologyreview.com/2024/09/17/1104004/why-openais-new-model-is-such-a-big-deal>.

Alberto Nebreda et al., *The Social Machine: Artificial Intelligence (AI) Approaches to Theory of Mind*, in THE THEORY OF MIND UNDER SCRUTINY 681 (Teresa Lopez-Soto et al. eds., 2023).

Christian Szegedy et al., *Intriguing Properties Of Neural Networks* להרחבה ראו ARXIV (Dec. 21, 2013), <https://arxiv.org/abs/1312.6199>.

Lee & Trott, לעיל ה"ש 21.

Tomas Mikolov et al., *Efficient Estimation of Word Representations in Vector Space*, ARXIV (Sept. 7, 2013), <https://arxiv.org/abs/1301.3781>.

לייצר אנלוגיה שבה הוא נדרש לספק מילה חלופית ל"רופא" כאשר מחליפים גבר באישה. בעוד שהתוצאה המתבקשת לאנלוגיה היא "רופאה", המודל יראה את התוצאה "אחות". בכך נראה כי המודל שיעתק דפוסי אפליה נפוצים בחברה האנושית שלפיהם רופא הוא מקצוע השמור לגברים בעוד שנשים במקצועות הרפואה תשמנה לרוב כאחיות.³² על אף זאת, "וקטור ייצוגים" מהווה אבן בניין שימושית במיוחד למודלי השפה הגדולים, שדרכם המערכת מקודדת מידע חשוב על מערכת היחסים בין מילים שונות.³³

אחד המודלים הפופולריים של למידת מכונה הוא מודל רשת נוירונים. מדובר במודל למידת מכונה אשר באופן מופשט דומה לפעילות הנוירונים במוח. דהיינו, מודל אשר משקלל יחדיו אותות פשוטים תוך שימוש בפונקציות לא ליניאריות. מודל כללי זה נחקר במאה השנים האחרונות, ובשנות השמונים קיבל לראשונה את הכינוי "למידה עמוקה".³⁴

לשם הבנה טובה יותר של מערכות בינה מלאכותית יוצרות (Generative AI), יש להבחין בין למידת מכונה באופן כללי לבין למידה עמוקה, תת-קבוצה של אלגוריתמי למידת מכונה. כלל מערכות למידת המכונה, בכללן גם מערכות למידה עמוקה, משתמשות בסט מובנה של נתונים לשם הלמידה ונדרשות לייצר פלט מוגדר בהתאם. ההבדל בין השתיים נעוץ בכך שלמידה עמוקה נזקקת פחות להנדסת מאפיינים (features). כך, כלים המשתמשים בלמידה עמוקה, דוגמת מערכות בינה מלאכותית יוצרת, אינם מכוונים מראש למאפיינים ספציפיים. המערכת עצמה מבינה מה הוא המאפיין החשוב ומנפקת אותו מבלי להזדקק לדוגמאות מוגדרות מראש.³⁵ כך, לדוגמה, מערכות למידה עמוקה שמייצרות תמונה מטקסט, כגון DALL-E, מסוגלות לנתח את התיאור הטקסטואלי המוזן להן ולתרגם אותו לתמונה מציאותית מבלי להזדקק להנחיות מפורטות מראש.³⁶

עם זאת, למידה עמוקה דורשת נפח עצום של נתונים לאימון ומצריכה שימוש בחומרות ותוכנות יקרות יותר מאלו אשר ניתן לעשות בהן שימוש במודלים אחרים של למידת מכונה.³⁷

תחום הלמידה העמוקה קידם את המחקר בנתונים בתחומים רבים, כולל הבנת קול, תמונה וטקסט. הבינה המלאכותית היוצרת מבוססת על כלים של למידה עמוקה.

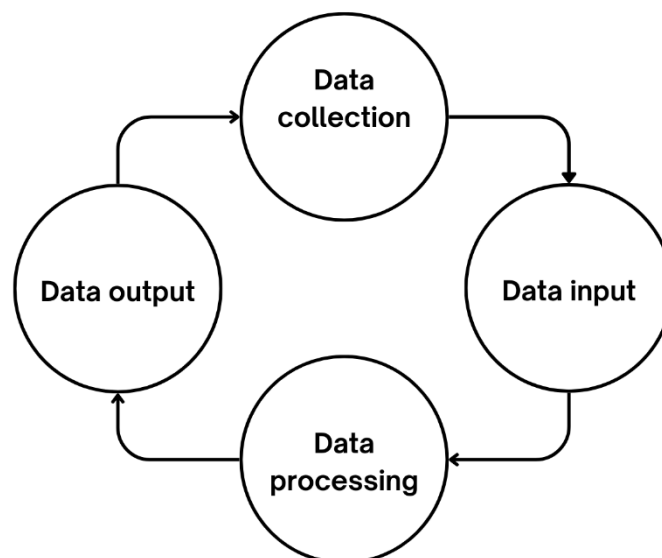
- | | |
|---|----|
| Aylin Caliskan et al., <i>Semantics Derived Automatically From Language Corpora</i> | 32 |
| .Contain Human-Like Biases, 365 SCIENCE 183 (2017) | |
| Bender et al., לעיל ה"ש 16. | 33 |
| Fradkov, לעיל ה"ש 23. | 34 |
| David Lehr & Paul Ohm, <i>Playing with the Data: What Legal Scholars Should Learn</i> | 35 |
| .About Machine Learning, 51 COLUM. L. REV. 653, 700 (2017) | |
| Suhaima Jamal et al., <i>Perception And Evaluation Of Text-To-Image Generative AI</i> | 36 |
| Models: A Comparative Study Of DALL-E, Google Imagen, GROK, And Stable | |
| .Diffusion, 25 ISSUES IN INFORMATION SYSTEMS 277, 282 (2024) | |
| Christian Janiesch et al., <i>Machine Learning And Deep Learning</i> | 37 |
| .MARKETS 685, 688 (2021) | |

עם זאת, חשוב להבין שהמחקר הראשוני בתחום היה מאכזב, ואף ננטש במשך שנים. הפריחה המחודשת של התחום נבעה, בין היתר, מההתקדמות הטכנולוגית אשר אפשרה איסוף ואגירה של נתוני עתק המשמשים לאימון מודלי למידה עמוקה, לצד יכולות חישוב מתקדמות אשר נדרשות כדי לאפשר למידה מכמות כה גדולה של נתונים.³⁸

3. מהו מחזור החיים של נתונים?

עם הגידול המתמשך בכמות הנתונים המיוצרת בעידן המודרני, גידול אשר נסמך על שיפורים טכנולוגיים באמצעי חישה וקישוריות, נוצר הצורך בסביבת עבודה שבה נתונים יכולים להיקלט, להיאגר, להיות מעובדים ולספק תוצרים. מחזור החיים של פרויקטי נתונים הוא סביבת עבודה שכזו.

מוצג א



בתרשים ניתן לראות דוגמה לתהליך המציג את מחזור החיים של נתונים.

בספרות יש גרסאות רבות למחזור החיים של נתונים.³⁹ בפרק זה נציג דוגמה למחזור חיים שיש בו ארבעה שלבים עיקריים – איסוף, ניהול, ניתוח והצגת נתונים.

Christian Szegedy et al., *Intriguing Properties Of Neural Networks* ARXIV (Dec. 21, 2013), <https://arxiv.org/abs/1312.6199>.

Kumar Rahul & Rohitash Kumar Banyal, *Data Life Cycle Management in Big Data Analytics*, 173 *PROCEDIA COMPUTER SCIENCE* 364 (2020).

בשלב איסוף הנתונים, סנסורים פיזיים מותקנים ומתחילים להזרים נתונים דרך רשת תקשורת. הסנסורים הפיזיים מעבירים מידע על אודות נתונים כמו טמפרטורה או תנועת רכבים (למשל באמצעות מצלמות תנועה).⁴⁰ בנוסף, דיווחים אנושיים באמצעות טפסים ממוחשבים, אשר ממולאים על ידי בני אדם, מגיעים ליעדם ומאוחסנים במאגרי נתונים. לבסוף נעשה חיפוש ושאיבת מידע באמצעות כלים אוטומטיים במקורות מידע בסביבות כמו האינטרנט וארכיונים דיגיטליים.⁴¹ עושר המידע הפתוח באינטרנט מאפשר גישה למידע בכמויות שלא ידענו בעבר, מה שמאפשר פיתוח מודלים של למידה עמוקה אשר לומדים בצורה יעילה.

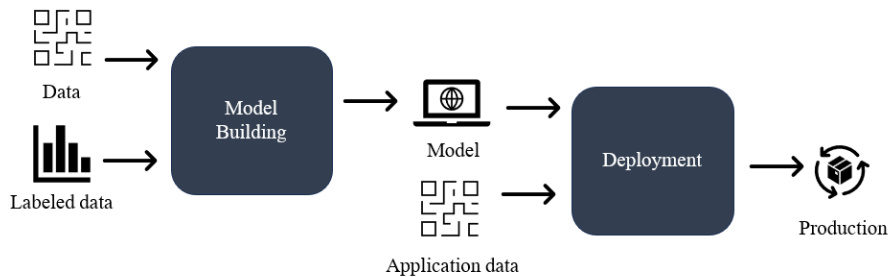
שלב ניהול הנתונים נותן מענה לצורך לאחסן נתונים שמגיעים ממקורות מידע שונים ושהתפתחו באופן עצמאי ובלתי תלוי כדי להתאימם לפרויקטים עתידיים. למשל, במהלך פיתוח מערכת בתחום הרפואה, נתוני בריאות מגיעים ממקורות שונים כמו תיקים רפואיים אלקטרוניים, סריקות רפואיות (כגון MRI ו-CT), ותוצאות בדיקות מעבדה. במצב זה יש לאחסן את כלל הנתונים ולהתאימם לפורמט אחיד כך שניתן יהיה לעשות בהם שימושים עתידיים. המונח "אגם נתונים" (data lake) מקובל כיום לתיאור המאגר הפיזי שבו נמצאים נתונים אשר יש בהם צורך (נוכחי או עתידי).⁴² מאגר כזה נמצא בשרתי הארגון או בענן. בשלב ניהול הנתונים, מעבר לאחסון הפיזי שלהם, יש צורך לאפשר את שילובם באופן שניתן יהיה לקשר בין פריטי מידע שניתן למקמם כשייכים לאותו מאורע, לאותה ישות, או לארגנם כראוי על ציר זמן. בנוסף, תהליך מקובל של ניקוי נתונים אף הוא שייך לשלב ניהול הנתונים ומתבצע בשל החשש מהטיות אשר עלולות ליצור הפליה.⁴³ בנוסף, באקדמיה מתחילים לפתח שיטות אוטומטיות לשלב ניהול הנתונים.⁴⁴

שלב ניתוח הנתונים הוא שלב מורכב אשר כולל ניתוח נתונים ראשוני, שמטרתו גישוש לזיהוי התובנות שניתן להפיק מהנתונים, ובהמשך ניתוח עומק, אשר כולל כמויות גדולות של נתונים ואימון חוזר ונשנה של מודלי למידה מתוך מטרה להוריד את היקף השגיאה בלמידה למינימום. על מנת להבין טוב יותר מהי "שגיאה" בהקשר של אימון מודל, דמיינו מודל למידה אשר מטרתו ללמוד לזהות האם יש חתול בתמונה. בכל שלב של האימון, המערכת מקבלת תמונה ומשיבה "כן" אם יש חתול בתמונה ו"לא" אחרת. אם המערכת מזהה שיש חתול בתמונה ואכן יש חתול בתמונה נאמר

-
- 40 Lehr & Ohm, לעיל ה"ש 35, בעמ' 677–679.
- 41 Aman Ahluwalia & Suhrud Wani, *Leveraging Large Language Models for Web Scraping*, ARXIV (Jun. 12 2024) <https://arxiv.org/abs/2406.08246>
- 42 Dan Woods, *Big Data Requires a Big New Architecture*, FORBES (Jul. 21, 2011), <https://www.forbes.com/sites/ciocentral/2011/07/21/big-data-requires-a-big-new-architecture>
- 43 Daniel Varona & Juan Luis Suárez, *Discrimination, Bias, Fairness, and Trustworthy AI*, 12 APPLIED SCI. 1596 (2022)
- 44 Sandra Geisler et al., *Knowledge-Driven Data Ecosystems Toward Data Transparency*, 14 ACM JDIQ 1 (2021)
-

שמדובר ב-True Positive. כאשר המערכת מזהה נכונה שאין חתול בתמונה נאמר שמדובר ב-True Negative. אלו הן תוצאות של המודל שאינן נחשבות שגיאה. כאשר המודל טועה, בין אם טוען שיש חתול בתמונה (ואין, מה שנקרא False Positive), או שטוען שאין חתול בתמונה (ויש, מה שנקרא False Negative) המודל "נענש" והמשוב מסייע למודל, על ידי שינוי פרמטרים בקבלת ההחלטה, להוריד את כמות השגיאות. התוצר של שלב זה הוא מודל מאומן אשר ניתן להפעיל על נתונים חדשים ולהפיק מהם תובנות להמשך.

מוצג ב



שלבי אימון המודל, ניתן לראות כיצד מודלים של למידת מכונה מנתחים את הנתונים.⁴⁵

לבסוף, שלב הצגת הנתונים מיועד להחזיר את השליטה בתהליך לבני האדם. התובנות אשר התקבלו מניתוח הנתונים מוצגות למקבלי החלטות אנושיים. התוצאה, לרוב, היא אוסף דרישות חדש אשר מכתוב איסוף נתונים נוסף ומשם המשך מחזור החיים. על מנת להציג את התובנות, התפתחו שיטות רבות להצגה גרפית של נתונים ונוצר הצורך בהסבריות (explainability) של מודלי למידה.⁴⁶

4. בינה מלאכותית יוצרת או בינה מלאכותית יצירתית?

פריצת דרך משמעותית בתחום למידת המכונה נעשתה עם הופעתם של מודלים יסודיים המבוססים על רשתות נוירונים גדולות. על בסיס מודלים אלו ניתן לבנות מודלים

⁴⁵ התרשים נוצר בהשראת Prasant Dixit, *How to Deploy ML Models to Production*, MEDIUM (Dec. 27, 2021), <https://prasantdixit.medium.com/how-to-deploy-ml-models-to-production-ab7ce1c353de>.

⁴⁶ ראו סקר Waddah Saeed et al., *Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities*, 263 KNOWLEDGE-BASED SYSTEMS .110273 (Mar. 5, 2023).

אחרים, אשר ניתן לעשות בהם שימוש במגוון תחומים כמו רפואה⁴⁷ והנדסה אזרחית.⁴⁸ אף שמודלים אלו קיימים כבר שנים מספר, במהלך השנים האחרונות התאפשרה בניית מודלים גדולים במיוחד (לדוגמה LLM, "מודלי שפה גדולים"), וכתוצאה מכך יכולות המודל השתפרו במידה רבה.⁴⁹ ישנם מודלים יסודיים פתוחים שבהם משקולות המודל (הקוד) משוחררות לציבור הרחב, דוגמת מודל ה-Llama של מטא.⁵⁰ מנגד, ישנם מודלים יסודיים סגורים, שזמינים למפתחי המודל בלבד, דוגמת מערכת פלמינגו של גוגל.⁵¹ בין אלה לאלה ניתן למצוא גם מודלים יסודיים המעניקים גישה מוגבלת לפעילות המודל ופועלים בחוסר שקיפות כלפי הציבור, כמו GPT 4. כלומר, בעוד שהציבור יכול לעשות בהם שימוש, אופן הפעולה שלהם אינו חשוף לציבור.⁵²

המונח "בינה מלאכותית יוצרת" מקובל היום בתיאור מודלים של למידת מכונה אשר מייצרים תוכן (תמונה, טקסט, טקסט) ואינם מאבחנים אותו (למשל, על ידי סיווג תמונות לקטגוריות). עם זאת, חשוב לציין שהשם "בינה מלאכותית יוצרת" עשוי להיות מטעה, כי למעשה כל כלי של למידת מכונה הוא בבחינת בינה מלאכותית יוצרת, מאחר שמודלים של למידת מכונה מטרתם היא לספק פלט למשתמש. על מנת להבדיל בין בינה מלאכותית יוצרת לכזו שאינה יוצרת עלינו להתמקד במטרת היצירה. מודל למידה אשר תוצאתו מגיעה למשתמש קצה אך אינה נשמרת במאגר נתונים (אלא אם כן המשתמש בוחר לשלב אותה כתוצר במערכת מידע) אינו נחשב כבינה מלאכותית יוצרת. תחת הגדרה זו, תוצאה של מודל למידה אשר מוזנת כעובדה למאגר נתונים, הרי היא נופלת בקטגוריה של בינה מלאכותית יוצרת, כך שעדיין מדובר בהגדרה מרחיבה מדי. ייתכן שההבחנה שנרצה לעשות היא בין בינה מלאכותית (אשר יוצרת תכנים שישמשו בעתיד) לבין בינה מלאכותית יצרית, כאשר האחרונה מבטאת היטב את מה שנתפס בציבור כבינה מלאכותית יוצרת. לא רק יצירה, אלא תוצר שנתפס כיצירתי, טקסט מקורי, ציור שאין כמוהו במרחב האינטרנטי, סונטה מוזיקלית אשר הייתה יכולה להיכתב על ידי באך ועוד.

דרך נוספת לחשוב על בינה מלאכותית יוצרת היא "אפקט וואו" (wow effect) של תוצר הלמידה. להמחשה הבה נתבונן בשתי טכנולוגיות. טכנולוגיה אחת היא טכנולוגיית תמונה לטקסט, אשר מקבלת תמונה ומייצרת רשימה של עצמים אשר קיימים בתמונה. טכנולוגיה שנייה, אשר תוצאתה לכאורה סימטרית לראשונה, היא

- | | |
|---|----|
| Nandhini Mahesh et al., <i>Advancing Healthcare: the role and impact of AI and foundation models</i> , 16 AM. J. TRANSNAT'L RSCH., 2166 (2024). | 47 |
| Mattia Rigotti et al., <i>Towards Workflows for the Use of AI Foundation Models in Visual Inspection Applications</i> , 6 CE/PAPERS 605 (2023). | 48 |
| Fradkov, <i>לעיל ה"ש</i> 23. | 49 |
| Hugo Touvron et al., <i>Llama 2: Open Foundation and Fine-Tuned Chat Models</i> , ARXIV (Jul. 19, 2023), https://arxiv.org/abs/2307.09288 . | 50 |
| Jean-Baptiste Alayrac et al., <i>Flamingo: a Visual Language Model for Few-Shot Learning</i> , ARXIV (Nov. 15, 2022), https://arxiv.org/abs/2204.14198 . | 51 |
| OpenAI et al., <i>GPT-4 Technical Report</i> , ARXIV (Mar. 15, 2023), https://arxiv.org/abs/2303.08774 . | 52 |

טכנולוגיית טקסט לתמונה, אשר מקבלת הנחיות ליצירת תמונה ומייצרת כפלט תמונה שעונה על הדרישות. טכנולוגיית תמונה לטקסט קיימת שנים רבות ובבסיסה מודל למידה אשר מסווג לעצמים ידועים. טכנולוגיית טקסט לתמונה היא חדשה (עם יישומים כגון DALL-E ו-Midjourney), משתמשת במודלי פְּצָפּוּעַ (diffusion models)⁵³ ונופלת תחת הקטגוריה של בינה מלאכותית יוצרת (למעשה יצירתית). מבחינה טכנולוגית, מדובר במודלים שווי יכולות עם מטרות שונות. עם זאת, מנקודת מבט של משתמש קצה, טכנולוגיית תמונה לטקסט מאפשרת למכונה לבצע מטלה שאדם ללא יכולות מקצועיות חריגות יכול לבצע (הרי כולנו יודעים להתבונן בתמונה ולקבוע מהם העצמים שיש בה). טכנולוגיית טקסט לתמונה, לעומת זאת, איננה נמצאת בסל הכישורים של רבים מאיתנו, ודאי שלא ברמה שבה המודל יודע לייצר, ולפיכך נתפסת כקפיצת דרך טכנולוגית משמעותית.

סיכום

החדירה המהירה של כלי בינה מלאכותית לשימוש יומיומי פותחת עולם חדש של יישומים, יישומים אשר חסרים חותמת איכות ויכולת רגולטורית אמיתית לפקח על איכות זו. בבואנו לבחון דרכים רגולטוריות להתמודדות עם שיטפון היישומים מבוססי בינה מלאכותית חובה עלינו להבין מהן היכולות הטכנולוגיות העומדות לרשותנו להבנת הכלים והדרכים שבהן ניתן לבקר טכנולוגיה זו. בפרט, הבינה המלאכותית היוצרת מציבה אתגרים חדשים הכרוכים בשילוב של יצירת טקסטים, תמונות וסרטונים הנראים מציאותיים אך יכולים להכיל מידע שגוי, מוטעה ומטעה.

במאמר קצר זה תיארתי את הבסיס הטכנולוגי של למידת מכונה ואת מחזור החיים של נתונים. השימוש בשניהם מאפשר ניתוח איכותי להבנת היכולות והמגבלות של כלים טכנולוגיים.

עם הפנים קדימה, הצלחת הרגולטור בטיוב יישומים מבוססי בינה מלאכותית תלויה ביכולת של משפטנים ומדעני נתונים לעבוד יחדיו. לשם כך, יש צורך בהכשרות משותפות, יצירת שיח מפרה רב-תחומי ויכולת לנתח בו-בזמן את נקודות המבט השונות של בעלי זיקה (stakeholders) למוצר.

Ziyei Chang et al., *On the Design Fundamentals of Diffusion Models: A Survey* ARXIV 53
(2023), <https://arxiv.org/abs/2306.04542>