סיכוני סייבר למערכות בינה מלאכותית: אתגרים טכנולוגיים ודרכי התמודדות

לירן ענתבי וגיל ברעם

מערכות אוטונומיות מבוססות בינה מלאכותית הולכות ותופסות מקום משמעותי בחיי היומיום ובמגוון תחומי חיים: תעשייה, רפואה, משקי בית וכמובן התחום הביטחוני. מערכות אלו, בהיותן ממוחשבות, חשופות לטעויות בקוד, שמצידן עלולות להוביל לקבלת החלטות שגויה ולביצוע פעולות לא רצויות. בנוסף לכך, הן פגיעות למתקפות סייבר שעלולות להוביל לפגיעה בפעילותן או אף להשבתתן. מאמר זה בוחן את הסיכונים המאיימים על מערכות אוטונומיות כחלק ממרוץ חימוש בין המעצמות, ודן בצעדים בתחום המדיניות שיאפשרו להתמודד עם סיכונים אלה ברמה הלאומית.

מילות מפתח: בינה מלאכותית, לוחמת סייבר, ביטחון לאומי, מרוץ חימוש

מבוא

"בינה מלאכותית היא העתיד, לא רק עבור רוסיה אלא עבור האנושות כולה [...] יש בה הזדמנויות אדירות, אך גם איומים שקשה לחזותם. מי שיהפוך למוביל בתחום זה יהיה לשליט העולם". כך אמר נשיא רוסיה, ולדימיר פוטין, בהרצאה שנשא בספטמבר 2017.[1] ואכן, נראה כי מערכות אוטונומיות מבוססות בינה מלאכותית

לירן ענתבי היא עמיתת מחקר במכון למחקרי ביטחון לאומי, מנהלת את המחקר בתחום הטכנולוגיות המתקדמות והביטחון הלאומי ב־INSS, מרצה באוניברסיטת בן־גוריון ויועצת בתחום הטכנולוגיות המתקדמות.

גיל ברעם היא מנהלת המחקר בסדנת יובל נאמן למדע, טכנולוגיה וביטחון ועמיתת מחקר במרכז הסייבר הרב־תחומי ע"ש בלווטניק, אוניברסיטת תל אביב.

1 "Artificial intelligence is the future, not only for Russia, but for all humankind […] It comes with colossal opportunities, but also threats that are difficult to predict. Whoever becomes the leader in this sphere will become the ruler of the world": "'Whoever Leads in AI Will Rule the World': Putin to Russian Children on Knowledge Day", *RT World News,* 2017, https://www.rt.com/news/401731-ai-rule-world-putin.

הולכות ותופסות מקום משמעותי בחיי היומיום במגוון תחומי חיים: תעשייה, רפואה, משקי בית וכמובן בתחום הביטחוני. בהיותן מערכות ממוחשבות, הן חשופות לטעויות בקוד, שעלולות להוביל לקבלת החלטות שגויה ולביצוע פעולות לא רצויות. הן גם פגיעות למתקפות סייבר שעלולות להוביל לתוצאות לא רצויות, לפגיעה בפעילותן או אף להשבתתן. במקביל, מתרחב השימוש במערכות בעלות יכולות אוטונומיות, הדורשות אמנם רמה מסוימת של מעורבות אנושית בקבלת ההחלטה הסופית להפעלתן, אך תהליכי החישוב וההמלצה שלהן אוטונומיים, ולרוב אינם ניתנים להסבר.

מאמר זה בוחן את הסיכונים המאיימים על מערכות אוטונומיות ואת הדרכים להתמודד עימם ברמה הלאומית. חלקו הראשון נפתח בסקירת השימושים בבינה מלאכותית ובמערכות אוטונומיות בתחום הביטחוני, מתאר את מרוץ החימוש המתרחש בתחום זה ואת השפעתו על הזירה הבין־לאומית ודן בתמריצים לביצוע מתקפות סייבר על מערכות אלו. בחלק השני של המאמר מתוארות מתקפות סייבר אפשריות על מערכות מבוססות בינה מלאכותית – מתקפות ומניפולציות הייחודיות למערכות סייבר – וכן נסקרים שימושים שניתן לעשות בבינה מלאכותית בתחומי לוחמת הסייבר לצורכי הגנה והתקפה כאחת. לבסוף מציע המאמר צעדים בתחום המדיניות, שמטרתם היא לסייע בצמצום הסיכונים שעשויים להיות הרסניים ככל שהשימוש במערכות האוטונומיות מתרחב והתלות האנושית בהן גדֵלה.

בינה מלאכותית ומערכות אוטונומיות – התפתחות ושימושים מרכזיים

בינה מלאכותית היא ענף של חקר מדעי המחשב שקיים מאז שנות החמישים של המאה העשרים. אחת ההגדרות הפשוטות והמקובלות של בינה מלאכותית היא: "לגרום למכונה להתנהג בדרך שהייתה נחשבת לאינטליגנטית לו אדם התנהג כך".[2] בעשור האחרון, בין היתר עקב ההתקדמות בחקר מדעי המחשב, פיתוח חומרות ותוכנות מתקדמות בתחומי המחשוב והתקשורת ופריצת תחומים חדשים כמו מחשוב ענן ונתוני עתק, התאפשרה התקדמות משמעותית גם בתחום הבינה המלאכותית. במסגרת זו התפתחו תחומים כמו "למידת מכונה" ו"רשתות נוירונים עמוקות", שמאפשרים מגוון יישומים מתקדמים בתחומים שונים. בין השאר מדובר ביישומים בתחום פענוח התמונה, הפועלים בעולם הרפואה ומסייעים לנתח בדיקות מסוגים שונים; יישומים בתחום הקולי, המאפשרים פעולתם של "עוזרים חכמים", דוגמת "סירי" ו"אלקסה"; וכן אלגוריתמים רבים בתחום הניבוי, המציעים לקהל

2 Edward Geist and Andrew Lohn, "How Might Artificial Intelligence Affect the Risk of Nuclear War?", *RAND Corporation*, 2018, p. 9.

הרחב ברשת האינטרנט מוצרים או שירותים כמו אלה שהם רכשו קודם לכן או כאלה שהתעניינו בהם בעבר.

בדרפ"א (DARPA – Defense Advanced Research Projects Agency) שבמשרד ההגנה האמריקאי מתייחסים אל בינה מלאכותית כאל "יכולת מתוכנתת לעבד מידע".[3] למרות ההגדרה, חשוב להבהיר כי לא כל מערכת מחשוב עושה שימוש בבינה מלאכותית. אלגוריתמים של בינה מלאכותית מעוצבים לצורך קבלת החלטות ועושים זאת לרוב באמצעות שימוש בנתונים שמוזנים להם בזמן אמת. לא מדובר במכונות פסיביות שמסוגלות לתגובות מכניות או קבועות מראש בלבד, כפי שהורגלנו בעידן האוטומציה (למשל, דלתות אוטומטיות), אלא במכונות שבעזרת חיישנים, נתונים דיגיטליים ואפילו קלטים מרוחקים, מסוגלות לשלב מידע ממקורות שונים, לנתח אותו באופן מיידי ולפעול בהתאם לתובנות הנגזרות מן הנתונים. הדבר מאפשר תחכום ומהירות בקבלת הנתונים ברמה שלא הייתה קיימת קודם לכן.[4]

מרבית היישומים הנפוצים בתחום הבינה המלאכותית בימינו שייכים לתחום הקרוי "למידת מכונה". תחום זה כולל שימוש באלגוריתמים סטטיסטיים המחקים משימות אנושיות קוגניטיביות על ידי גזירת חוקים בנוגע אליהן באמצעות ניתוח כמויות גדולות של נתונים בנושא. למעשה, האלגוריתם "מתאמן" על מידע קיים, ותוך כדי תהליך "האימון" יוצר מודל סטטיסטי משלו, שיוכל לבצע בעתיד את אותה משימה על נתונים חדשים שבהם לא נתקל קודם לכן.[5]

השימוש בטכנולוגיה של בינה מלאכותית הולך וגובר, ומדינות, חברות וגורמים ביטחוניים רבים מסתמכים כיום על מערכות כאלו לצרכיהם השונים. בין השימושים האזרחיים בבינה מלאכותית ניתן למנות שירותים שונים, כגון אפליקציות ניווט, אלגוריתמים המציעים סחורות או שירותים מותאמים, אלגוריתמים בתחומי הבנקאות והסחר הפיננסי, מערכות בתחומי אחזקה, לוגיסטיקה ועוד. כאמור, מערכות מבוססות בינה מלאכותית נפוצות גם בתחום הביטחון, וכבר כיום קיימים תחומים ביטחוניים רבים בהם נעשה שימוש במערכות כאלו. ביניהם ניתן למנות:

- **מודיעין:** שימושים רבים מאוד בבינה מלאכותית קיימים בתחום המודיעין. כיום נפוץ השימוש באלגוריתמים מתחום למידת מכונה ותחומים נוספים לצורך ניתוח

3 John Launchbury, "A DARPA Perspective on Artificial Intelligence", *TechnicaCuriosa*, 2017, https://machinelearning.technicacuriosa.com/2017/03/19/a-darpa-perspective-on-artificial-intelligence./

4 Darrell M. West and John R. Allen, "How Artificial Intelligence Is Transforming the World", *Brookings*, 2018, https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world./

5 Kelley M. Sayler and Daniel S. Hoadley, "Artificial Intelligence and National Security", *Congressional Research Service*, 2019, p. 2.

תמונות וטקסטים. כמו כן, נעשה שימוש באלגוריתמים לתרגום שפות, לניתוח וידיאו ואודיו ועוד. בין הידועים בפרויקטים מתחום זה נמצא "פרויקט מייבֶן" ("Maven") שבוצע בשיתוף פעולה בין "גוגל" ובין משרד ההגנה האמריקאי, שבמסגרתו נעשה שימוש בבינה מלאכותית לצורך ניתוח של תצלומי מל"טים.[6] בנוסף לארצות הברית, גם בסין ובמדינות נוספות מתקיימת פעילות לייצור מערכות לסינון מיטבי של תוכן מודיעיני והיתוך מידע ממערכות שונות לצורך הפקת מודיעין, תוך שימוש ביכולות בינה מלאכותית.[7]

- **לוגיסטיקה:** כמו בתחום האזרחי, יישומי בינה מלאכותית קיימים גם בתחומי הלוגיסטיקה הצבאית השונים. צבא ארצות הברית עושה שימוש במערכות כאלו כבר משנות התשעים של המאה הקודמת, ואחת מהן אף סייעה בתכנון ואופטימיזציה של העברת כוחות בזמן מלחמת המפרץ הראשונה. הדבר הוביל לחיסכון שהחזיר את ההשקעה שנעשתה במשך שלושים שנה במחקר הבינה המלאכותית.[8] בין המערכות החדשניות יותר ניתן למנות כאלו המסייעות בתחום האחזקה של מערכות באופן שלא היה מתאפשר בעבר, כמו למשל מערכות שמדווחות מראש על בלאי עתידי בחלפים ומאפשרות להחליף אותם על בסיס שימוש פרטני ולא על בסיס מידע סטטיסטי כללי, כפי שהיה מקובל עד כה. שיטה זו מובילה לחיסכון משמעותי, כמו גם להגברת הבטיחות.[9]
- **פיקוד ושליטה:** במערכות פיקוד ושליטה עתיד להיעשות שימוש הולך וגובר בבינה מלאכותית, בין היתר כמערכות מייעצות שיפעלו כמערכות מסייעות החלטה, בכפוף לגורם האנושי ובשיתוף פעולה איתו.
- **כלי רכב אוטונומיים:** נהיגה אוטונומית מזוהה בעיקר עם "מכוניות ללא נהג", שאבות הטיפוס שלהן מתחילים לצוץ על הכבישים במקומות שונים ברחבי העולם. למעשה, מדובר באחד האתגרים המרכזיים של דרפ"א עוד מהעשור הקודם, שאפשר את קפיצת המדרגה בתחום זה.[10] בעוד שבתחום האזרחי

6 Samuel Gibbs, "Google's AI Is Being Used by US Military Drone Programme", *The Guardian*, May 7, 2018, https://www.theguardian.com/technology/2018/mar/07/google-ai-us-department-of-defense-military-drone-project-maven-tensorflow. יש לציין כי הפרויקט הוביל להתנגדות בקרב עובדי "גוגל", לאור החשש שהידע שנוצר בו יסייע לא רק בניתוח חומרים מודיעיניים, אלא גם ביצירת מערכות נשק אוטונומיות שיוכלו לתקוף ללא מעורבות יד אדם.

7 להרחבה בנושא זה ראו: Stephen Chen, "Inside the AI Revolution that's Reshaping Chinese Society", *South China Morning Post*, June 29, 2017, https://www.scmp.com/news/china/society/article/2100427/chinas-ai-revolution-and-how-its-rivalling-us.

8 נורית כהן־אינגר וגל קמינקא, "והרי התחזית: צה"ל בדרך לצבא תבוני – מפת דרכים לאימוץ טכנולוגיות בינה מלאכותית בצה"ל", **בין הקטבים**, גיליון 18, 2018, עמ' 95.

9 Sayler and Hoadley, "Artificial Intelligence and National Security", p. 9.

10 DARPA, "The Grand Challenge for Autonomous Vehicles", 2019, https://www.darpa.mil/about-us/timeline/-grand-challenge-for-autonomous-vehicles.

השימוש הבולט הוא האוטונומיות הקרקעית, בתחום הביטחוני נעשה, כבר מספר עשורים, שימוש בכלים בלתי מאוישים באוויר, בים וביבשה, חלקם בעלי יכולות אוטונומיות שונות. לכלים אלה משמעות גדולה מאוד בשדה הקרב, והם יכולים להוות מכפיל כוח או להחליף גורם אנושי באזורי סכנה. יחד עם זאת, מרבית הכלים הללו, חרף יכולותיהם האוטונומיות, נסמכים במידה רבה על מעורבות והפעלה אנושיות.

- **מערכות אוטונומיות צבאיות:** מדובר באחד התחומים הבולטים המבוססים על בינה מלאכותית. מדינות רבות זיהו בעשורים האחרונים את הפוטנציאל הטמון בשימוש בכלים בלתי מאוישים לצורכי ביטחון, והן נוקטות צעדים שונים לרכש ולפיתוח מערכות כאלו באופן עצמאי. בין המדינות המובילות בתחום זה: ארצות הברית, ישראל, בריטניה וצרפת. בתוך המערכות האוטונומיות הצבאיות ניתן להתייחס לתת־קבוצה של מערכות נשק אוטונומיות. אלו מערכות המסוגלות לחפש, לזהות ולתקוף מטרות באופן עצמאי, ללא מעורבות אנושית.[11] מערכות אלו מהוות גורם "משַנה כללי־מִשחק", משום שהן ניתנות להפעלה כמעט ללא מעורבות אנושית, אך בה בעת מסוגלות לגרום לפגיעה קטלנית. תחום המערכות האוטונומיות הצבאיות מעורר ויכוח ציבורי נרחב, ובאו"ם אף מתקיימים דיונים בנוגע להגבלתו. למרות זאת, פיתוחו מואץ מאוד כיום, עד שיש מי שחושש שהעולם נמצא על סף מרוץ חימוש בתחום זה,[12] או אף בעיצומו. מדובר אמנם בתחום שנמצא בתחילת דרכו, אך מספר מדינות כבר רכשו ניסיון מבצעי בשימוש במערכות אוטונומיות מסוגים שונים. בין המערכות שצברו ניסיון ניתן למנות מערכות הגנה אווירית, דוגמת "פטריוט" האמריקאית או "כיפת ברזל" הישראלית. אלו הן מערכות בעלות אוטונומיות גבוהה מאוד, ואפילו יכולת לפעול באוטונומיות מלאה. יחד עם זאת, לאור החלטה עקרונית של המדינות המפעילות אותן, הן עדיין נסמכות על החלטת המפעיל האנושי הנמצא במעגל ההפעלה שלהן.[13] לצד מערכות ההגנה האווירית ניתן למנות את "החימוש המשוטט", דוגמת "הארופ" ("Harop"). מדובר במערכת אווירית בעלת יכולת טיסה, שהייה באוויר, איתור, מעקב ותקיפת מטרות ללא מעורבות

11 "Autonomous Weapon Systems – Q & A", *International Committee of the Red Cross*, November 12, 2014, https://www.icrc.org/en/document/autonomous-weapon-systems-challenge-human-control-over-use-force.

12 Billy Perrigo, "A Global Arms Race for Killer Robots Is Transforming the Battlefield", *TIME*, April 9, 2018, http://time.com/5230567/killer-robots.

13 "Losing Humanity: The Case Against Killer Robots", *Human Rights Watch*, November 2012, pp. 11-12, http://www.hrw.org/sites/default/files/reports/arms1112ForUpload_0_0.pdf

יד אדם, בין השאר באמצעות התבייתות על אותות מכ"ם.[14] מחקרים עדכניים מצביעים על מגמה, לפיה כלים אוטונומיים לחלוטין עתידים להפוך תוך כשני עשורים לאפשריים מבחינה טכנולוגית. במצב כזה, קיימת סבירות גבוהה לכך שכלים כאלה יהפכו לגורם מרכזי יותר בפעולתם של צבאות מודרניים.[15]

- **לוחמת סייבר:** זהו אחד התחומים המובילים בשימוש בבינה מלאכותית. בתחום זה נמשך השימוש בבינה מלאכותית מ"הדור הראשון",[16] במקביל להתפתחות יכולות גם מדורות מתקדמים יותר. אלגוריתמים מסייעים למנוע תקיפות במסגרת לוחמת סייבר, או לאתר ולהתריע מפני מתקפות על מערכות ממוחשבות שונות. יחד עם זאת, יכולות של בינה מלאכותית עשויות לשמש בדרכים מגוונות גם את הצד התוקף.

מרוץ החימוש בתחום הבינה המלאכותית[17]

מדינות רבות זיהו בשנים האחרונות את הפוטנציאל הטמון בתחום הבינה המלאכותית לכלכלותיהן ולביטחונן הלאומי. **ארצות הברית** נחשבת למובילה בתחום זה ופועלת לגיבוש אסטרטגיה מקיפה בנושא. אסטרטגיית ההגנה הלאומית של ארצות הברית כוללת מחויבות של משרד ההגנה האמריקאי להשקיע ביישומים צבאיים בתחום האוטונומיות, הבינה המלאכותית ולמידת מכונה, לצד שימוש

---

14 Dan Gettinger and Arthur Holland Michel, "Loitering Munitions", *The Center for the Study of the Drone at Bard College*, 2017, http://dronecenter.bard.edu/files/2017/02/CSD-Loitering-Munitions.pdf

15 יואב זקס ולירן ענתבי (עורכים), **השימוש בכלים צבאיים בלתי מאוישים עד שנת 2033: המלצות למדיניות לאומית על בסיס חיזוי טכנולוגי – הערכת מומחים**, מזכר 145, המכון למחקרי ביטחון לאומי, דצמבר 2014, http://www.inss.org.il/he/wp-content/uploads/sites/2/systemfiles/memo145_7%20%20%D7%A1%D7%95%D7%A4%D7%99%20%D7%9C%D7%90%D7%AA%D7%A8%20.pdf ; Paul Scharre, "Robotics on the Battlefield Part I: Range, Persistence and Daring", *Center for a New American Security*, May 2014, https://s3.amazonaws.com/files.cnas.org/documents/CNAS_RoboticsOnTheBattlefield_Scharre.pdf?mtime=20160906081925; Launchbury, "A DARPA Perspective on Artificial Intelligence".

16 John Launchbury, "A DARPA Perspective on Artificial Intelligence," *TechnicaCuriosa*, 2017, https://machinelea

17 נכון למועד כתיבת שורות אלו מתקיים מעין "מרוץ" בין המעצמות לפיתוח יכולות בינה מלאכותית מתקדמות. במקביל לכך יש הקוראים להפחית את השימוש במונח "מרוץ חימוש", שהוא בעל הקשר שלילי, בטענה כי יש לדבר על שילוב בין תחרות ובין שיתוף פעולה. הם גם קוראים לארצות הברית ולסין לפתוח בדיאלוג שיוביל לשיתוף פעולה בפיתוח בינה מלאכותית. להרחבה בנושא זה ראו: Tim Hwang and Alex Pascal, "Artificial Intelligence Isn't an Arms Race", *Foreign Policy*, December 11, 2019, https://foreignpolicy.com/2019/12/11/artificial-intelligence-ai-not-arms-race-china-united-states/.

בטכנולוגיות מסחריות פורצות דרך, וזאת במטרה לשמר את היתרון התחרותי הצבאי של ארצות הברית.[18]

בתחילת 2019 עדכן הבית הלבן את האסטרטגיה הלאומית למחקר ופיתוח של טכנולוגיות בינה מלאכותית, שפורסמה על ידי ממשל אובמה ב־2016. האסטרטגיה המעודכנת קוראת, בין היתר, לפתח שיטות אפקטיביות לשיתוף פעולה אדם–בינה מלאכותית ולהבטיח את השמירה על אבטחתן של מערכות בינה מלאכותית.[19] ארצות הברית משקיעה תקציבים גדולים בתחום זה ופועלת לשרטט אסטרטגיה רחבה לקידום ולהגנה על טכנולוגיית בינה מלאכותית ברמה הלאומית, תוך שיתוף פעולה בין הממשל, המגזר הפרטי, האקדמיה, הציבור ושותפויות בין־לאומיות.[20]

ביולי 2019 קרא "המרכז המשותף לבינה מלאכותית" של הפנטגון (Joint Artificial Intelligence Center – JAIC) לחברות להגיש רעיונות והצעות בטכנולוגיות של בינה מלאכותית לצורך הגנה בסייבר, שיכללו תיקון אוטומטי של חולשות אבטחה ברשתות צבאיות, איסוף מודיעין סייבר על פעילים ב"רשת האפלה" ועוד.[21] למרות מכלול מאמציה, ארצות הברית עדיין ניצבת בפני מספר אתגרים בתחום הבינה המלאכותית. בין הבולטים שבהם הַתחרות הגוברת עם סין, לאור השאיפה הסינית להפוך בתוך כעשור למובילת התחום.

סין מהווה, כאמור, מתחרה רצינית בתחום הבינה המלאכותית וכבר הוכיחה בעבר את יכולתה לקדם במהירות פרויקטים טכנולוגיים מתקדמים (לדוגמה, הפיכתה ליצרנית ויצואנית בולטת של מל"טים תוך עשור אחד בלבד). ההשקעה התקציבית הכוללת של סין במחקר ופיתוח של בינה מלאכותית אינה חשופה לציבור, אך לפי ההערכות היא מגיעה למיליארדי דולרים לכל הפחות. לפי חלק מההערכות, מדובר בהשקעות עתידיות בסך של 150 מיליארד דולר.[22] הדבר נובע, בין היתר, מאחד היתרונות הבולטים של סין בתחום זה, והוא היעדר כמעט מוחלט של גבולות או חסמים בין תחומים "אזרחיים" ל"ביטחוניים", בשל העובדה שבסין גם התחום האזרחי נתון לפיקוח הדוק של השלטון.

יתרון בולט נוסף של סין נובע מהעובדה שהיא אינה פועלת על פי ערכים מערביים, המעודדים דמוקרטיה, זכויות פרט ושמירה על הפרטיות, ולכן כבר שנים

18 James N. Mattis, "Summary of 2018 National Defense Strategy of The United States of America", *Department of Defense*, Washington, D.C., 2018, p. 5.

19 Aaron Boyd, "White House Updates National Artificial Intelligence Strategy", *Defense One*, June 22, 2019, http://bit.ly/2ZYY2U4.

20 "Artificial Intelligence for the American People", *The White House*, 2019, https://www.whitehouse.gov/ai/executive-order-ai.

21 "DoD's JAIC to Call for Private Sector Cyber Tech Pitches", *MeriTalk*, July 8, 2019, https://www.meritalk.com/articles/dods-jaic-to-call-for-private-sector-cyber-tech-pitches/.

22 Ibid.

רבות היא אוספת ומקודדת מידע על אזרחיה. תהליך זה הפך את סין למכרה עצום של נתוני עתק, ומוביל גורמים וחברות ברחבי העולם לעבוד יחד איתה כדי לקבל גישה אליהם. יש לציין כי הסינים אוספים מידע גם על אזרחים של מדינות אחרות באמצעות ביצוע תקיפות סייבר וגניבת מאגרי מידע עצומים, וזאת במקביל לניטור מידע ממערכות ויישומים מתוצרת סין המצויים בשימוש בקרב אזרחים במדינות רבות. כמו כן, החקיקה הסינית מאפשרת לגורמי הממשל הסיני להגיע לפסי הייצור של כלל היצרנים בסין ולהחדיר לתוכם "דלתות אחוריות" במידת הצורך. אותה חקיקה גם מחייבת את יצרני הטכנולוגיה הסיניים לתת לממשל הסיני את קוד המקור של הטכנולוגיה שלהם.[23]

יש הערכות, לפיהן סין תעקוף בעתיד את ארצות הברית בתחום הבינה המלאכותית ותהיה המדינה הדומיננטית בעולם בתחום זה. בנובמבר 2017 העריך אריק שמידט, אז יושב ראש חברת "גוגל", כי סין צפויה להשתוות כבר בשנת 2020 לארצות הברית ביכולות הבינה המלאכותית שלה, וכי בשנת 2025 היא אף צפויה לעקוף אותה בתחום זה.[24] הערכות עדכניות תומכות בהערכותיו של שמידט. נראה כי מספר המחקרים האקדמיים בנושא הבינה המלאכותית שמפורסמים על ידי חוקרים סינים ישתווה בשנת 2020 למספר המחקרים המפורסמים על ידי חוקרים אמריקאים.[25]

בנוסף לסין ולארצות הברית, גם **רוסיה** מנהלת תוכניות בתחום הבינה המלאכותית, ובשנת 2019 אף נוסחה בה אסטרטגיה לאומית לבינה מלאכותית.[26] יחד עם זאת, רוסיה מצויה בפיגור ביחס לארצות הברית ולסין: מלבד השקעה נמוכה בתחום זה יחסית למתחרותיה העיקריות, היא גם סובלת מבעיות שקשורות ל"אקוסיסטם" הטכנולוגי שלה.[27] על רקע זה יש הערכות לפיהן רוסיה תוכל להוביל רק בתחומים צרים וממוקדים של בינה מלאכותית, ולא את התחום כולו.[28]

---

23 המחברות מודות לד"ר הראל מנשרי, ראש תחום סייבר במכון הטכנולוגי חולון, על הערותיו המועילות בנושא זה.

24 Sam Shead, "Eric Schmidt on AI: 'Trust me, these Chinese People are Good'", *Business Insider*, November 1, 2017, https://www.businessinsider.com/eric-schmidt-on-artificial-intelligence-china-2017-11.

25 Tom Simontie, "China is Catching up to the US in AI Research – FAST", *Wired*, March 13, 2019, https://www.wired.com/story/china-catching-up-us-in-ai-research/.

26 Samuel Bendett, "Putin Orders Up a National AI Strategy", *Defense One*, 2019, https://www.defenseone.com/technology/2019/01/putin-orders-national-ai-strategy/154555/.

27 Ibid.

28 Andrew P. Hunter et al., "Artificial Intelligence and National Security: The Importance of an AI Ecosystem", *CSIS*, Washington, D.C., 2018, p. 48, https://www.csis.org/analysis/artificial-intelligence-and-national-security-importance-ai-ecosystem.

בין המדינות הרבות המתחרות בתחום הבינה המלאכותית ניתן למנות גם את **ישראל**, הנחשבת למובילה טכנולוגית עולמית, במיוחד בתחומי סייבר וכלי טיס בלתי מאוישים. כיום אין לישראל אסטרטגיה סדורה לנושא הבינה המלאכותית, אם כי ועדה מטעם ראש הממשלה פועלת בימים אלה לביצוע מחקר מקיף בתחום, שעל בסיס מסקנותיו והמלצותיו ינוסחו אסטרטגיה ומדיניות ועשוי לקום מטה לבינה מלאכותית.[29] יתרונה המשמעותי של ישראל הוא ה"אקוסיסטם" הייחודי שלה, הכולל קשרים בין הממשלה, האקדמיה, התעשייה והצבא, וכן היכולת להגיב במהירות לשינויים בזירה. לכך מתווסף היתרון של מעבר ידע משמעותי בין הצבא ובין התעשיה וחברות אזרחיות, הנובע מהמודל הייחודי של גיוס חובה ושירות מילואים. מודל זה יוצר אצל חלק מהעובדים הזדמנות לרכוש ידע ולהעביר ידע בין מערכת הביטחון ובין התעשייה, באופן מתמשך ומַפרה.[30]

בישראל פועלות חברות רבות העוסקות בתחום הבינה המלאכותית. בחלקן מהווה התחום את ליבת החברה ובחלקן הוא מהווה מכפיל כוח או טכנולוגיה מאפשרת. לצד זה, הוקמו בישראל מרכזי פיתוח של חברות בין־לאומיות גדולות, כמו "אמזון", "אינטל", "מייקרוסופט" ו"אנבידיה", השמים דגש על בינה מלאכותית.[31] כמו כן, קיימות בישראל חברות מובילות בתחום הבינה המלאכותית המפתחות לא רק תוכנה אלא גם חומרה.[32] בתחום הביטחוני מפתחת ישראל בינה מלאכותית במסגרת משרד הביטחון, מפא"ת ויחידות טכנולוגיות שונות בצבא, וכן בתעשיות הביטחוניות.

מרוץ החימוש הבין־לאומי בתחום הבינה המלאכותית, הפיכתן של מערכות אלו לשכיחות, וההסתמכות עליהן בתחומי חיים רבים, מחייבים דיון באיומים עליהן ובדרכים האפשריות לאתר, לזהות, למנוע או לסכל אותם.

מתקפות סייבר ומניפולציות על מערכות בינה מלאכותית

איום הסייבר, ההולך וגובר בשנים האחרונות, מהווה איום גם על מערכות בינה מלאכותית. במקביל הוא מעלה חשש מניצול טכנולוגיות בינה מלאכותית גם לצורך הוצאה לפועל של מתקפות סייבר בקנה מידה הרסני הרבה יותר מזה שמוכר

29 "ועדת המדע: דיון ראשון על מוכנות הממשלה לתחום הבינה המלאכותית", *כנסת ישראל*, 2018, https://m.knesset.gov.il/News/PressReleases/pages/press04.06.18ec.aspx.

30 דפנה גץ ואחרים, **בינה מלאכותית, מדעי הנתונים ורובוטיקה חכמה**, דו"ח ראשון, מוסד שמואל נאמן למחקר מדיניות לאומית, 2018.

31 Amir Mizroch, "In Israel, A Stand Out Year for Artificial Intelligence Technologies", *Forbes*, March 11, 2019, https://www.forbes.com/sites/startupnationcentral/2019/03/11/in-israel-a-stand-out-year-for-artificial-intelligence-technologies/#13acbc7530a8.

32 כדוגמאות ניתן להביא את החברות הישראליות "מלאנוקס" ו"הבאנה לאבס", שנמכרו לחברות בין־לאומיות במיליארדי דולרים. לקריאה נוספת בעניין זה ראו: https://www.themarker.com/technation/.premium-1.8285726.

לנו עד כה. הסיכון גדל עוד יותר כשמדובר בפגיעה במערכות ביטחוניות שאינן מנותקות לגמרי מהרשת, וכן לאור השימוש ההולך וגובר של צבאות בטכנולוגיית בינה מלאכותית.

טכנולוגיית הבינה המלאכותית נחשבת כיום לחלק בלתי נפרד מהאפשרויות לביצוע תקיפות סייבר. יחד עם זאת, בינה מלאכותית יכולה להיות כלי להתמודדות יעילה יותר עם תקיפות סייבר, למשל באמצעות שימוש בטכניקות למידה עמוקה (Deep Learning) שמסוגלות לעקוב אחר פעילות חשודה ולסווג וירוסים שונים. מנגד, מערכות הבינה המלאכותית רגישות למתקפות סייבר עליהן ועלולות להיות נתונות למניפולציות שונות. מערכות אוטונומיות ומערכות בינה מלאכותית הן מערכות מחשוב, ולכן הן חשופות, כמו מערכות אחרות, למתקפות סייבר מן הסוגים המוכרים לנו ממערכות מחשוב רגילות. יחד עם זאת, בגלל אופיין המיוחד, המערכות האוטונומיות ומערכות הבינה המלאכותית חשופות גם למתקפות ייחודיות, בין היתר מהסיבות הבאות:

- הרצון לאפשר את פעולתן האוטונומית, ללא מעורבות אדם, משיקולים של יעילות, דיוק ומהירות, עשוי להשאיר אותן חשופות למתקפות. עם זאת, אפשר להעריך כי גם מערכות אלו יתנו חיווי למפעילים על אנומליות או זיהוי תקיפה.
- חלק מן המערכות המבוססות על בינה מלאכותית פועלות כיום באופן שאיננו יודעים להסביר או לנתח בדיעבד. הדבר מכונה אתגר ה"הסבריות" (explainability) או אתגר "הקופסה השחורה". הדבר משאיר פתח למתקפות שיהיה קשה לזהותן לפחות בחלק מהמקרים, מכיוון שמדובר במתקפה ולא בתהליך שהמערכת ביצעה באופן תקין.
- תהליכי ה"אימון" של מערכות בינה מלאכותית, המתבצעים על כמויות גדולות מאוד של מידע, משאירים פתח להכנסת מידע ש"יזהם" את התהליך באופן מכוון ויוביל לתוצאות מוטות או בלתי רצויות.

מספר דוגמאות מהעת האחרונה ממחישות את האיום: באפריל 2019 אמר ראש אגף המבצעים ב"מרכז המשותף לבינה מלאכותית" במשרד ההגנה האמריקאי, קולונל סטוני טרנט (Trent), כי הבעיות בהערכת איומי הסייבר נגד טכנולוגיות של בינה מלאכותית נובעות מחוסר מודעות בקרב מקבלי ההחלטות וממחסור בכלים ובשיטות לבדיקת חסינות הסייבר של מערכות בינה מלאכותית. לדבריו, אחד מתפקידיו של המרכז הוא לעודד פיתוח של כלים כאלה, שפיתוחם אינו נחשב לכדאי בענף המסחרי האזרחי.[33]

---

33 Theresa Hitchens, "Rush to Military AI Raises Cyber Threats", *Breaking Defense*, April 25, 2019, https://breakingdefense.com/2019/04/rush-to-military-ai-raises-cyber-threats/.

לדברי מי שהיה ראש אגף המחקר ב"סוכנות לביטחון לאומי" (NSA), פרדריק צ'אנג (Chang), המרוץ לפיתוח טכנולוגיות צבאיות המבוססות על בינה מלאכותית יגדיל משמעותית את מרחב התקיפה, אולם ממשלות עדיין אינן מודעות למרבית חולשותיהן של המערכות הללו. צ'אנג הזהיר שתוקפים עלולים להטעות את מנגנון הזיהוי של המערכת באמצעות "תשומות נגד" (Adversarial Inputs), "להרעיל" את המידע המשמש ללימוד המערכת (Data Poisoning) ולהסתנן לתוכה במטרה להבין את מערכת ההפעלה שלה ולפגוע בפעילותה.[34]

בנוסף, השילוב בין אמצעי לוחמת סייבר ובין טכנולוגיית בינה מלאכותית עלול להוביל לפיתוח סוגים חדשים של נוזקות. לדוגמה, חוקרים בחברת IBM פיתחו נוזקה מבוססת בינה מלאכותית בשם DeepLocker, במטרה להבין כיצד ניתן לשלב מספר מודלים קיימים של בינה מלאכותית כדי ליצור סוג חדש ויעיל יותר של נוזקה שטרם נראה. הנוזקה החדשה מסווה את מטרתה עד שהיא מגיעה אל היעד, אותו היא מזהה באמצעים כגון זיהוי קול או זיהוי פנים. נוזקות מסוג זה נחשבות ליעילות במיוחד, שכן הן מסוגלות להדביק מיליוני מערכות מבלי להתגלות, וזאת בניגוד לתקיפות סייבר, שלעיתים עשויות להיות רחבות היקף ו"רועשות" ("spray and pray" approach).

מכיוון שמערכות אוטונומיות אמורות לפעול ללא מעורבות יד אדם, או במעורבות נמוכה מאוד שלו, מניפולציה או מתקפה יעילה עליהן עלולה לא להתגלות במשך זמן רב. הייחודיות של נוזקה משולבת בינה מלאכותית כמו זו שהוזכרה, לעומת נוזקות קיימות, נעוצה בכך שבשל תחכומן יידרשו מומחיות עמוקה וכלי פורנזיקה מתקדמים כדי לזהותן.

נוזקת DeepLocker שינתה את כללי המשחק בכך שהיא "מחביאה" את פעילותה מאחורי אפליקציות נפוצות, כמו אפליקציות לשיחות וידאו. זאת ועוד, השימוש שהיא עושה בבינה מלאכותית כמעט ואינו מאפשר לזהות אותה ולבצע הנדסה לאחור של הקוד שלה. DeepLocker תתחיל לפעול רק אם תזהה את המטרה שנבחרה, והיא תעשה זאת באמצעות שימוש במודל של "רשת נוירונים עמוקה" (Deep Neural Network). מודל זה תוכנן לפעול רק כאשר הוא יזהה קלט ספציפי, כלומר רק כאשר יזהה את הקורבן המיועד.[35]

דוגמה נוספת לשימוש עוין בטכנולוגיית בינה מלאכותית לצורך תקיפה סייבר הוצגה במחקר שפורסם בשנת 2017, במסגרתו הפעילו חוקרים כלי בינה מלאכותית

---

34 Ibid.

35 Marc P. Stoecklin, "DeepLocker: How AI Can Power a Stealthy New Breed of Malware", *SecurityIntelligence*, August 8, 2018, https://securityintelligence.com/deeplocker-how-ai-can-power-a-stealthy-new-breed-of-malware/.

כדי לזהות סיסמאות של משתמשי הרשת החברתית "לינקדאין". סקירה של 43 מיליון פרופילים של משתמשים אפשרה לחוקרים לזהות 27 אחוזים מהסיסמאות.[36]

חוקרים מעריכים כי מערכות הבינה המלאכותית יאפשרו לבני האדם לבצע מתקפות שלא ניתן היה לבצע לפני השימוש במערכות מסוג זה. בנוסף, תוקפים יוכלו לאתר מקורות חדשים למתקפות על מערכות בינה מלאכותית על ידי מציאת נקודות תורפה חדשות בהן. תרחיש נוסף הוא תקיפה של מערכות שאחראיות על מספר כלים אוטונומיים. תקיפה כזאת עלולה לגרום לשיבוש נרחב, שישפיע על יותר מאשר כלי אחד.[37]

אפשר להצביע על שלושה סוגים עיקריים של תקיפות על מערכות בינה מלאכותית: (1) הכנסת מידע שגוי למערכת, כך שתסיק מסקנות שגויות; (2) שיבושים זעירים בתמונות (או קלטים אחרים) שהמערכת מעבדת, באמצעות הכנסת פריטים גלויים או שיבוש בלתי נראה בפיקסלים של התמונה, כך שהיא תסווג את התוצרים באופן שגוי; (3) שיבוש בשיפוט מידע באמצעות חבלה פנימית באמצעי המיון של המערכת, וזאת במקום התמקדות במידע בודד שמוזן למערכת.[38]

להלן פירוט של כמה מסוגי המתקפות הייחודיות למערכות בינה מלאכותית:

- **מניפולציה במידע המוצג למערכות הבינה המלאכותית (Adversarial Attacks)**: זוהי טכניקה להטעיית מערכות בינה מלאכותית (machine learning classifier) באמצעות ניצול רגישותן למניפולציה במידע המוזן להן ומשַמש לאימון המערכת. בדרך זו יכולים התוקפים לייצר קֶלֶט הנראה כבעל סיווג מטעה, וכן "הפרעות" למידע המוזן למערכת, כדי לגרום לסיווג שגוי. השינויים כמעט ואינם נראים בעין אנושית. אחד המחקרים מצא כי ניתן "לשַטוֹת" בקלות רבה ב"רשתות נוירונים עמוקות" באמצעות הכנסת קֶלֶט שגוי.[39]
- **"הרעלת מידע" (Data Poisoning)**: מדובר בטכניקה שבה התוקף מכניס נתונים שגויים ומשבש באופן שיטתי את המידע המוזן לצורך אימון המערכת. לשם כך,

36 Matthew Hutson, "Artificial Intelligence Just Made Guessing Your Password a Whole Lot Easier", *Science*, September 15, 2017, https://www.sciencemag.org/news/2017/09/artificial-intelligence-just-made-guessing-your-password-whole-lot-easier.

37 Allan Dafoe, "AI Governance: A Research Agenda", *Future of Humanity Institute*, University of Oxford, 2017, p. 5, https://doi.org/10.1176/ajp.134.8.aj1348938 ; Miles Brundage et al., "The Malicious Use of Artificial Intelligence: Forecasting, Prevention and Mitigation", 2018, p. 20, https://img1.wsimg.com/blobby/go/3d82daa4-97fe-4096-9c6b-376b92c619de/downloads/MaliciousUseofAI.pdf?ver=1553030594217.

38 Jian hua Li, "Cyber Security Meets Artificial Intelligence: A Survey", *Frontiers of Information Technology and Electronic Engineering* 19, no. 12 (2018): 1462-1474, https://doi.org/10.1631/FITEE.1800573.

39 Mesut Ozdag, "Adversarial Attacks and Defenses against Deep Neural Networks: A Survey", *Procedia Computer Science*, 140 (2018): 152-161, https://doi.org/10.1016/j.procs.2018.10.315.

צריך שתהיה לתוקף גישה למידע המשמש לאימון המודל. שיבוש המידע עלול להיטיב עם התוקפים או להפלות קבוצות אחרות – למשל מודלים המחשבים פרמיות ביטוח או הענקת הלוואות.[40]

- **מתקפות להתחמקות מגילוי (Evasion attacks)**: אלו מתקפות שבהן מבצע התוקף מניפולציה על יכולת הסיווג של מודלים במטרה להתחמק מזיהוי. מתקפות כאלו נועדות לעקוף מודלים המשמשים כמסַנני ספאם, כמזַהֵי קודים זדוניים או כמנטרֵי תעבורת רשת ומגלֵי אנומליות.[41]
- **גניבה והעתקה של מודל ה־AI (Model Extraction)**: אלו הן מתקפות שבהן התוקף שולח דוגמאות של נתונים למודל המערכת ומנתח את הפֶלֶט במטרה לבנות מודל משל עצמו.[42]
- **מתקפות על תיוגי Watermarks**: Watermarks הוא מושג המתייחס למתקפות שבהן התוקף מוסיף פיקסלים מיוחדים לתמונה במטרה לגרום למודל להגיב באופן מסוים.[43] מתקפה אפקטיבית מסוג זה עשויה להיות בעייתית מאוד כאשר מדובר במערכות מודיעיניות או במערכות נשק.

מערכות המבוססות על למידת מכונה מחזיקות לעיתים במידע רגיש שיכול להיות יעד לתקיפות סייבר, כמו מערכות לזיהוי פנים, שהתוקפים יכולים להוציא מהן מידע על האנשים המזוהים במערכת. תקיפות כאלו יכולות להתבצע בשני שלבים שונים של בניית המערכת: בשלב אימון המערכת ובשלב הבחינה והסקת המסקנות מפעילותה. יחד עם זאת, קיימים אמצעי הגנה שונים בתחום זה, שאפשר להכניסם בשני השלבים הללו.[44]

---

40 Jacob Steinhardt, Pang Wei Koh and Percy Liang, "Certified Defenses for Data Poisoning Attacks", *Advances in Neural Information Processing Systems*, no. i (December 2017): 3518-3530 ; Patrick Hall, "Proposals for Model Vulnerability and Security", *O'Reilly Media*, 2019, https://www.oreilly.com/ideas/proposals-for-model-vulnerability-and-security.

41 Battista Biggio et al., "Evasion Attacks against Machine Learning at Test Time", *Lecture Notes in Computer Science*, Part 3 (2013): 387-402, https://doi.org/10.1007/978-3-642-40994-3_25 ; Erwin Quiring and Konrad Rieck, "Adversarial Machine Learning against Digital Watermarking", *European Signal Processing Conference* (September 2018): 519-523, https://doi.org/10.23919/EUSIPCO.2018.8553343.

42 Florian Tramèr et al., "Stealing Machine Learning Models via Prediction APIs", *Cornell University*, October 2016, http://arxiv.org/abs/1609.02943.

43 Romain Artru, Alexandre Gouaillard and Touraj Ebrahimi, "Digital Watermarking of video streams: Review of the State-Of-The-Art", *arXiv preprint arXiv:1908.02039*, August 2019.

44 Qiang Liu et al., "A Survey on Security Threats and Defensive Techniques of Machine Learning: A Data Driven View", *IEEE Access* Vol. 6 (2018): 12103-12117, https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8290925.

אפשר לבצע התקפות של "הרעלת מידע" בשלב האימון, שיכללו הכנסת מידע שגוי למערכת ושינוי תוויות המידע. הדרך להגן על המידע בשלב זה היא באמצעות פיתוח אמצעים שיסננו מידע שגוי הנכנס למערכת (data sanitization), שימוש בתרחישים כדי ללמוד על פעולותיו הצפויות של היריב (adversarial training), זיקוק אמצעי הגנה (defense distillation), שילוב שיטות (ensemble methods) וחיזוק הפרטיות באמצעים שונים (differential privacy).

בשלב הבחינה אפשר לבצע סוגים שונים של התקפות, ביניהם התקפות להתחמקות מגילוי (evasion) על ידי שימוש באמצעֵי תקיפה שמסוגל לחמוק מאמצעי הזיהוי של המערכת. שיטה נוספת לתקיפה היא התחזות (Impersonation), המאפשרת לגורמים עוינים להתחזות לגורמים לגיטימיים כדי להיכנס למערכת ולשבש את המידע. שיטה שלישית לתקיפה היא היפוך (inversion). שיטה זו מאפשרת לגנוב מידע רגיש מתוך המערכת. בתקיפות כאלו ניתן להגן על המערכת באמצעות שימוש ברעש כ"מסך עשן" להגנה על מידע, או שימוש רנדומלי באמצעי הגנה במהלך האימון (differential privacy).[45]

כפי שניתן לראות, קיים מגוון רחב של מתקפות אפשריות על מערכות אוטונומיות ומערכות בינה מלאכותית, שחלקן ייחודיות לסוג זה של מערכות ושונות ממתקפות על מערכות מחשוב "רגילות". יש צורך להבחין בין המתקפות הרלוונטיות למערכות ממוחשבות באופן כללי ובין מתקפות בעלות ייחודיות לתחום הבינה המלאכותית והמערכות האוטונומיות, כפי שביקשנו להבהיר במאמר זה.

הסיכון במתקפות על מערכות בינה מלאכותית עולה מאוד, משום שלעיתים קרובות אין אדם שיוכל לזהות את הבעייתיות שנוצרה תוך פרק זמן קצר או סביר, וזאת עקב מאפייני המערכת. לבעיה זו מצטרפת בעיה נוספת שהוזכרה לעיל – אתגר "הקופסה השחורה" או ה"הֶסבריוּת". אתגר זה מתייחס לעובדה שחרף התוצאות המוצלחות של פעולות שונות של מערכות בתחום למידת מכונה או למידה עמוקה, לא ניתן להסביר את הדרך שעשתה המערכת כדי להגיע לתוצאה.[46] היעדר השקיפות מקשה לבצע ביקורת על פעולת מערכות באופן כללי. כאשר מוסיפים לכך את כמויות המידע ואת הקצב שבו המערכות הללו מעבדות אותו, ניתן להבין מדוע גם פיקוח אנושי שיוצב עליהן עשוי להיות למראית עין בלבד. מומחים ואנשי צבא טוענים כי על בני אדם להיחשב אחראים (accountable) לפעולותיהן של מערכות הבינה המלאכותית, אולם זוהי טענה שמחייבת דיונים

---

45 Ibid.

46 Richard Gall, "Machine Learning Explainability vs Interpretability: Two Concepts That Could Help Restore Trust in AI", *KDnuggets*, 2018, https://www.kdnuggets.com/2018/12/machine-learning-explainability-interpretability-ai.html.

נוספים, מכיוון שבני אדם אינם יכולים לאתר ולזהות את כל הסיכונים והפגיעויות הקיימים במערכות אלו.[47]

כדי ליצור פיקוח אפקטיבי ויכולת התמודדות עם הבעיות, יש צורך בהשקעת מאמצים ותשומת לב לנושא זה כבר בשלב המחקר והפיתוח של המערכות, וזאת כדי לנסות ולגבש מנגנונים אפקטיביים לבקרה אנושית עליהן. מאמצים לפתור את אתגר ה"הֶסבריוּת" נעשים מזה זמן,[48] אולם עד שיושג פתרון טכנולוגי מסַפק בתחום זה, יש צורך במנגנונים אחרים להתמודדות עם היעדר השקיפות, כולל מתחומי הרגולציה או המשפט. הדבר חשוב בעיקר כאשר מדובר במערכות אסטרטגיות, או בכאלו שתוצאות פעולותיהן עשויות לפגוע בחיי אדם.

סיכום והמלצות

השימוש במערכות מתחום הבינה המלאכותית, ובכלל זה במערכות מבוססות למידת מכונה ולמידה עמוקה, הולך והופך לנפוץ יותר ויותר במגוון תחומי חיים, ובתוכם התחום הביטחוני. מערכות אלו פועלות בקצב מהיר ובאופנים שלעיתים קרובות מקשים על בקרה אנושית עליהן. קושי זה מצטרף לשאיפה לאפשר למערכות אלו לפעול באוטונומיות מלאה כדי לחסוך במעורבות אנושית, וזאת ממגוון סיבות. במצב זה חשוב להבין מהן המתקפות האפשריות על מערכות אלו ולפתח אמצעי שיבוש שעשויים לפעול עליהן. כמו כן, חשוב להגיע להבנה של המתקפות הייחודיות למערכות מסוג זה, כדי שניתן יהיה לפתח מנגנוני פיקוח והגנה אפקטיביים שיאפשרו את פעולתן התקינה של אותן מערכות ואת האמון בהן. כדי להשיג את כל אלה, ניתן לפעול במספר אפיקי מדיניות.

ישראל, שהיא מעצמה בתחום הסייבר ואחת המדינות המובילות בעולם בתחום הבינה המלאכותית,[49] היא בעלת פוטנציאל גדול להפוך לגורם מוביל בתחום מחקר הסייבר והגנת הסייבר גם בכל מה שקשור לבינה מלאכותית ולמערכות אוטונומיות. לאור זאת, להלן מספר המלצות למדיניות עבור ישראל (שעשויות להיות רלוונטיות גם עבור מדינות נוספות):

---

47 Connor McLemore and Charles Clark, "The Devil You Know: Trust in Military Applications of Artificial Intelligence", *War on the Rocks*, September 23, 2019, https://warontherocks.com/2019/09/the-devil-you-know-trust-in-military-applications-of-artificial-intelligence/.

48 Zelros AI, "A Brief History of Machine Learning Models Explainability", *Medium*, 2018, https://medium.com/@Zelros/a-brief-history-of-machine-learning-models-explainability-f1c3301be9dc.

49 אורי ברקוביץ, "השקעה של 2 מיליארד ש' בשנה בעיר חכמה, בחקלאות ובאקדמיה: כך מתכננת ישראל להפוך למעצמת בינה מלאכותית", **גלובס**, 18 בנובמבר 2019, https://www.globes.co.il/news/article.aspx?did=1001307714.

- יש להגדיר תקנים בתחום המערכות המבוססות על בינה מלאכותית. במסגרת זו יש להטמיע במערכות אלו אמצעים שיאפשרו לפקח עליהן, או לוודא שלא בוצעו בהן תקיפה או מניפולציה. התקנים שיוגדרו צריכים לחול לא רק על מערכות ביטחוניות, אלא גם על מערכות אזרחיות קריטיות (ורצוי שגם על מערכות שאינן קריטיות, אך מסייעות בשמירה על שגרה). כמו כן, יש צורך בתקצוב לאומי של הגנת מערכות בתחומים שבהם קיים כשל שוק ואין תמריץ מסחרי שפותר את העניין.
- על הגורמים הרלוונטיים במערכת הביטחון להשקיע במחקר שיאפשר מיפוי ואיתור מתמשך של התחומים שבהם ייתכנו פגיעויות ומתקפות ייחודיות במערכות אוטונומיות. כמו כן, יש להשקיע בפיתוח של פתרונות ייעודיים בתחום זה.
- על הגופים הרלוונטיים במדינה להגדיר נהלים להשגחה על מִבצעי סייבר מבוססי בינה מלאכותית, כדי להימנע מהשלכות בלתי רצויות של מבצעים כאלה. הנהלים שיוגדרו צריכים להיות מגובים באמצעי אכיפה אפקטיביים.[50]
- רצוי לקיים תרגילים משותפים של בעלות ברית, בהם יִבָּחנו את יכולות ההגנה של הבינה המלאכותית. התרגילים יחשפו חולשות שיש לתקן ויעניקו למערכות את המידע שהן זקוקות לו כדי ללמוד ממנו.
- יש להרחיב את הדיון הבין־לאומי בנושא במטרה ליצור שיתופי פעולה בין מדינות בעלות עוצמה בתחום הבינה המלאכותית ובעלות אינטרסים משותפים (Like Minded Nations). זאת כדי לנסות ולהשפיע על הזירה הבין־לאומית כולה, לאור העובדה שקשה להשיג הישגים מהירים במוסדות בין־לאומיים כמו האו"ם. כן יש לגבש אמנות בין־לאומיות בתחום זה. חשיבותו של הנושא מתחדדת עוד יותר כאשר לוקחים בחשבון את השונוּת התרבותית הקיימת בין מדינות ואת השפעתה האפשרית הן על עיצוב הבינה המלאכותית במדינות שונות והן על הגדרת המושג "מוסרי" בקבלת החלטות בתחום זה.

יישום המלצות מדיניות אלו, ונוספות שניתן לגבש לאור הכרת הבעיה, עשוי לסייע במניעה של מתקפות בעלות פוטנציאל פגיעה. פוטנציאל זה הולך וגדל ככל שמערכות מבוססות בינה מלאכותית הופכות לנפוצות יותר ולאחראיות לתפקודים קריטיים שונים. טיפול בנושא מבעוד מועד עשוי לסייע במניעת תוצאות הרסניות ברמות הלאומית והבין־לאומית.

---

50 Mariarosaria Taddeo and Luciano Floridi, "Regulate Artificial Intelligence to Avert Cyber Arms Race", *Nature*, 556, no. 7701 (April 16, 2018): 296-298, https://doi.org/10.1038/d41586-018-04602-6.