נייר עמדה
שימושי בינה מלאכותית במלחמה

פרופ' יגיל לוי, פרופ' אורי שורץ

תקציר

נייר עמדה זה בוחן את השימוש במערכות בינה מלאכותית לייצור מטרות במלחמה, תוך התמקדות בניסיון שנצבר במלחמת עזה. הנייר מזהה ארבע קבוצות מרכזיות של סיכונים: שינוי אופני הייצור של הידע וההפללה באמצעות היגיון סטטיסטי ו"קופסה שחורה"; האצה והוזלה דרמטיות של ייצור מטרות, המאפשרות הרחבה חסרת תקדים של היקף התקיפות; הטיות וכשלים בַּדיוק הנובעים ממגבלות נתונים ומודלים; ועידוד "פתרונאות טכנולוגית" המחליפה התמודדות עם בעיות פוליטיות מורכבות בפתרונות חישוביים. נטען כי מערכות אלה כשלעצמן אינן גורמות להרג המוני, אך הן מספקות תשתית טכנולוגית ולגיטימציונית המאפשרת ומעודדת אותו. על בסיס ניתוח זה מוצע מתווה רגולטורי שנועד לצמצם פגיעה באזרחים ולחזק אחריותיות אנושית. עם ההמלצות המרכזיות נמנות: חובה לייצר "ידע משחזר" כתנאי לאישור מטרות, העלאת סף ההוכחות הנדרש להפללה, קביעת תקני מינימום

למעורבות אנושית, חיזוק הביקורת המשפטית והתאמתה למצב החדש, ופיתוח מנגנונים להתמודדות עם הטיות אוטומציה.

הקדמה

השימוש במערכות המבוססות על בינה מלאכותית לצרכים צבאיים בכלל ולצרכים מודיעיניים בפרט נתון במגמת עלייה בעולם המערבי ומעבר לו.[1] ישראל היא חלק ממגמה זו. במלחמת עזה (מאז 2023) עשתה ישראל שימוש נרחב במערכות AI לסיוע בייצור מטרות מודיעיניות – בפרט בתקיפות האוויריות בשלב הראשון של המלחמה – ובכלל זה, לצורך היתוך מידע על אנשים ועל מטרות פיזיות ממקורות שונים, אומדן ההסתברות שאדם או נקודת ציון יאושרו כמטרה, והמלצה על מטרות לאישור. ארצות הברית השתמשה במלחמה באיראן בשנת 2026 במערכות AI של פלנטיר ואנתרופיק למטרות דומות. שימוש זה כרוך באתגרים מוסריים ובסיכונים מגוונים, ועשויות להיות לו השלכות מגוונות ומרחיקות לכת.

על רקע זה קיים **מכון האוניברסיטה הפתוחה לחקר יחסי חברה־צבא** שתי סדנאות של חוקרים וחוקרות ב־18 בדצמבר, 2025 וב־16 בפברואר, 2026.[2] הסדנה התמקדה בשימושים בבינה המלאכותית לזיהוי "מטרות" אנושיות ופיזיות, ובמרכזן המערכות הצבאיות האלה: "הבשורה" (לזיהוי מטרות תשתיתיות), "לבנדר" (לזיהוי מטרות אנושיות) ו"איפה אבא" (לאיתור היעדים האנושיים בבתיהם). ההתמקדות במערכות אלה נועדה לשאול שאלות אתיות, פוליטיות וסוציולוגיות על ההשלכות הרחבות של השימוש בבינה מלאכותית למטרות מודיעין על מלחמה ועל חברות הנתונות במלחמה. הדיון המהותי בשאלות אלה אינו מתקיים בישראל כלל, וחלקים שלו אינם מעמיקים או נמצאים מתחת לרדאר הציבורי. מכאן הצורך לחולל דיון ציבורי. סדנאות אלה הניבו נייר עמדה מטעם המכון. נייר זה מבוסס על ניתוח של ממצאים מספרות המחקר ומפרסומים של מכוני מחקר, ארגוני זכויות אדם ותחקירנים בעיתונות, הנוגעים הן למלחמת עזה והן להקשרים אחרים.

מטרת המסמך היא לזהות את הסיכונים העיקריים של השימוש הנרחב בבינה מלאכותית לייצור מטרות, ולהציג מתווה של המלצות להתמודדות עם ההשלכות השליליות הללו. חלקו הראשון עוסק בהצגת הידע הקיים תוך מיפוי הסיכונים; חלקו השני הוא רשימה קצרה של הנחות יסוד המשמשות בסיס להמלצות, וחלקו השלישי מיועד להצגת המלצות מדיניות.

חלק א': סיכונים

להלן נמפה כמה תחומים שבהם משנות מערכות הבינה את אופן ההתנהלות של הלחימה. המיפוי מבוסס על דרך פעולתן של מערכות אלו בזירות הלחימה. הכנסתם של כלי AI לתהליך של ייצור המטרות מביאה לסיכונים מגוונים. זיהינו ארבע קבוצות מרכזיות של סיכונים: אלה הקשורים לשינויים באופני הייצור של הידע; להאצה ולהוזלה של ייצור המטרות; להטיות ולכשלים בַּדיוק; ולעידוד חשיבה פתרונאית באמצעות טכנולוגיה (technological solutionism).

1. הממד האפיסטמולוגי

מאחר שהאפיסטמולוגיה של הבינה המלאכותית שונה מהותית מזו שמשמשת להפללה בקרב חוקרי מודיעין אנושיים (או בדין הפלילי), היא משנה את משמעות ההפללה בשתי דרכים: התבססותה על היגיון סטטיסטי, וקופסה שחורה.

היגיון סטטיסטי: השימוש בבינה המלאכותית נועד, כלשונו של מפקד 8200 לשעבר יוסי שריאל, "לייצר (generate) את המידע שאין לך", כך שמידע לא־מפליל ידוע משמש לניבוי מידע מפליל שאינו ידוע.[3] זה ההיגיון של מה שמכונה "תקיפות חתימה". אלה החלו עוד בעשור הראשון של המאה ה־21, בסיוע כלים טכנולוגיים פשוטים יותר, כאשר ארצות הברית הכניסה אנשים לרשימות המועמדים להתנקשות על סמך דפוסי חיים הדומים לאלו הרווחים בקרב טרוריסטים, כפי שאלה השתקפו בחתימה הדיגיטלית שלהם. מהלך זה גבה מחיר גבוה בחיי

אזרחים לא־מעורבים ועורר ביקורת נרחבת בספרות.[4] גם במלחמת עזה שימשו מערכות בינה מלאכותית לייצור מטרות ולהפללתן שלא על סמך זיהוי ישיר של מעורבות בפעילות לחימה, אלא על סמך מאפיינים התנהגותיים עקיפים (כגון דפוסי תנועה ותקשורת) הקשורים במתאמים סטטיסטיים לפעילות טרור. בתחקיר של ה־*Washington Post* מתואר כיצד התבסס זיהוי מטרות אנושיות על דפוסי תקשורת, כגון החלפה תכופה של מספרי טלפון, קשרים עם מספרים חשודים או הופעה במסדי נתונים מסוימים.[5]

הבעייתיות בגישה זו היא שדפוסים אלה מאפיינים גם אזרחים רבים, לרבות עיתונאים, פעילי סיוע, קרובי משפחה של חמושים, או אנשים שנעקרו מבתיהם ונאלצו לשנות אמצעי תקשורת לעיתים קרובות (כפי שתועד רבות בספרות על מערכות של צבא ארצות הברית, כגון *Skynet*). התחקיר מוסיף כי גם השתייכות לקבוצות ווטסאפ רבות שימשה לעיתים כאינדיקציה מפלילה, חרף היעדר של קשר ישיר לפעילות לחימה. היגיון זה מבטא מהלך אמפיריציסטי־פרגמטי רדיקלי האופייני ל־AI וניתוח נתוני עתק באופן כללי: ויתור על הרצון להבין תופעות בשם הרצון לנבא אותן כדי להתמודד איתן.[6] יצוין כי טעויות קיימות בכל אופן של קבלת החלטות, אך אסור לאפשר קבלת החלטות שעלולות לעלות בחיי אדם על סמך היגיון שהוא סטטיסטי ונסיבתי בלבד, בלי הוכחות ישירות.

יש קושי מוסרי ומשפטי בהסתמכות בלעדית על היגיון סטטיסטי ונסיבתי לצורך הפללה או סימון מטרות אנושיות. קושי זה לא ייעלם גם אם נקבל את ההנחה שבנסיבות מסוימות, מערכות AI עשויות לשפר את הדיוק בזיהוי מטרות לגיטימיות וכך להפחית פגיעה בחפים מפשע. אם בהליך שיפוטי יש הטוענים שלעיתים אפשר להכיל היגיון הסתברותי במסגרת מנגנונים מוסדיים מורכבים של ביקורת, ערעור ובחינת ראיות,[7] הרי בהקשר מלחמתי, שבו החלטות מתקבלות בתנאי לחץ ובידי קצינים שאינם פועלים במסגרת שיפוטית מלאה, הסיכון שבהסתמכות על היסק אלגוריתמי מתעצם עוד יותר.

קופסה שחורה: מערכות AI שונות ממערכות שקדמו להן ברמת המורכבות של המודל, שאינו כולל רק דפוסים וקורלציות שזוהו בידי חוקרים אנושיים אלא גם דפוסים מורכבים (כולל אינטראקציות בין משתנים מרובים) שזוהו אוטומטית בידי מערכות של בינה מלאכותית. במקרים אלה לא תמיד ברור מדוע סיווגה המערכת אדם או מקום כ"מטרה". מחקר של מכון RAND[8] ודוח המחקר של RUSI[9] מדגישים כי מערכות למידה עמוקה "אטומות מטבען להבנה האנושית" ואינן מאפשרות להבין כיצד התקבלה המלצה מסוימת. בהיעדר יכולת להסביר את אופן הפעולה של המערכת ואת השיקולים שהובילו לסימון מטרה, אי אפשר לדעת בדיוק על אילו נתונים או היגיון הסתמכה.[10] גם כאשר יש למקבלי ההחלטות האנושיים גישה למידע מודיעיני גולמי ששימש את מערכת הבינה המלאכותית (כפי שלטענת צה"ל נעשה במלחמת עזה), אין להם יכולת לשחזר במלואו את ההיגיון שהוביל את המערכת לחישוב האלגוריתמי של ההסתברות להפללה, בגלל עצם העובדה שהמלצות עשויות להתבסס לא על היגיון סיבתי, אלא על קורלציות מורכבות ורב־ממדיות בין מיליוני משתנים.[11] זו בעיה עקרונית: גם כלים אוטומטיים שנועדו להכניס הסבריות למערכות AI (למשל, להצביע על האלמנטים שהשפיעו על ההמלצה יותר מאחרים) אינם יכולים לשחזר במדויק את הסיבות להמלצה.[12] מצב זה עלול להציב אתגרים ממשיים ליכולת של חוקרים אנושיים לשחזר את ההיגיון שמאחורי המלצות המערכת, וליכולת לפקח עליה פיקוח משמעותי ולממש אחריות משפטית במקרה של טעויות קטלניות. במיוחד כשמדובר במערכות בעלות אופי התקפי, הדבר מעלה חשש לעמידה בחובות של הדין הבין־לאומי ההומניטרי ודיני זכויות האדם, ובפרט בחובה לערוך חקירות אפקטיביות של פעולות צבאיות.[13] הדבר מגדיל את הסיכון להישנות של טעויות, שכן אי אפשר לטפל בשורש הבעיה בלי להבינה לעומק.

הטיית אוטומציה: עקב האמור לעיל, מקבלי החלטות אנושיים נדרשים להתמודד עם המלצות שמבוססות על היגיון של מכונה, השונה מזה שמנחה אותם בדרך כלל. יש דרכים שונות להתמודדות עם פער

אפיסטמולוגי זה, אך בהיעדר תמריצים ברורים, יש חשש מרכזי שלא ייעשה בהן שימוש, לנוכח הנטייה האנושית הרווחת (והמתועדת היטב בספרות) לבטוח במערכות אוטומטיות, גם לנוכח מידע המערער על נכונות המלצותיהן.[14] הטיה זו תועדה גם בהקשרים צבאיים: המחקר של מכון RAND[15] ודו"ח המחקר של SIPRI[16] מתארים את הנטייה של מפעילים ומפקדים לתת אמון מופרז בהמלצת המערכות, במיוחד בתנאים של לחץ, עומס וקצב פעולה גבוה, ולהימנע מלהטיל ספק בתוצאות גם כאשר קיים מידע סותר המערער על נכונות ההמלצה. בתנאי לחץ ועומס אנשים נדרשים לבחור בין עמדתם לבין המלצת המכונה,[17] שכן הזמן והכלים אינם מאפשרים להם לשחזר את ההיגיון שמאחורי המלצות המכונה ואין להם די זמן לגבש עמדה משלהם על סמך חומר גולמי (למשל, כאשר מופעל עליהם לחץ לקבל החלטות בתוך שניות ספורות, כפי שקרה לפחות חלק מהזמן במלחמת עזה).[18] הסתמכות עיוורת על המלצות המכונה היא סכנה ממשית.

מהבחינה המשפטית פירוש הדבר שמערכות AI אינן עומדות בעקרונות יסוד של שלטון החוק: הקופסה השחורה, ההאצה והטיית האוטומציה פירושן שאין שיקול דעת אנושי החורג משקלול טכני, ואין אפשרות לביקורת ולאחריותיות על קבלת ההחלטות.[19] יתרה מזו, כתוצאה מעמימויות אלה נוצרת עמימות בייחוס אחריות משפטית ופיקודית בתהליכי קבלת החלטות המתבצעים בתיווך אלגוריתמי. הספרות מראה שכאשר מערכות בינה מלאכותית משולבות בקביעת מטרות, מתערערת היכולת לייחס אחריות ברורה לפעולות קטלניות. דוחות של SIPRI[20] ומסמכי הפרלמנט האירופי[21] מדגישים כי מסגרות הדין והפיקוח הקיימות מתקשות להתמודד עם תהליכי קבלת החלטות הנשענים על המלצות אלגוריתמיות, שבהן הסמכות, שיקול הדעת והאחריות נחלקים בין גורמים אנושיים לבין מערכות טכנולוגיות. בהקשר זה, הפרלמנט האירופי מצביע על סיכון מערכתי הנוצר כאשר אין שקיפות מספקת באופן הפעולה של המערכות, ואין יכולת אפקטיבית לבקרָה, לאכיפה ולייחוס אחריות לאחר מעשה. בפרט, מצב זה מפר את החובה המינימלית הקבועה בדין הבין־לאומי לחקור הפרות אפשריות של

המשפט ההומניטרי ואת שיקול הדעת של קבלת ההחלטות ביחס לזה של מפקד סביר.

באופן כללי יותר, השימוש במערכות בינה מלאכותית לייצור מטרות עלול לתרום ללגיטימציה של זיהוי המטרות, לנוכח האמונה הרווחת (אך המוטעית) בציבור ובקרב אנשי צבא כי מערכות אלו חפות מהטיה ומשיקולים זרים או ממצבים רגשיים כגון שאיפת נקם. הפיכת נתונים אנושיים לדאטה דיגיטלי גם מקדמת דה־הומניזציה דיגיטלית, שאף היא, בתורה, מסייעת ללגיטימציה ומחלישה דילמות אתיות.[22]

2. ממד ההוזלה וההאצה

הספרות מראה ששילוב בינה מלאכותית במערכות של קביעת מטרות שינה מהותית את קצב הפעולה הצבאי ואת היקף ייצור המטרות, ולאו דווקא את הדיוק. במקרה של ישראל, ההאצה של קצב הייצור וההגדלה של בנק המטרות היו סיבות מוצהרות להכנסת ה־AI. הצהרות רשמיות של צה"ל ותחקירים של מגזין *+972*[23] ו־*Washington Post*[24] מסכימים שתוצאה זו אכן הושגה: הכנסת כלי AI האיצה את קצב ייצור המטרות בשני סדרי גודל, ממצב שבו יוצרו עשרות מטרות בשנה (כמטרה בשבוע, לדברי הרמטכ"ל לשעבר אביב כוכבי) למצב שבו מערכות בינה מלאכותית מאפשרות יצירה של מאות מטרות בשבוע (בשנת 2021 כוכבי נקב במספר של 100 מטרות ביום).[25] בנובמבר 2023 הודיע צה"ל שייצר 1,200 מטרות בארבעת השבועות הראשונים של המלחמה. בעבר, קצב כזה לא היה אפשרי בשל המחיר היקר בשעות של עבודת אדם לייצור מטרה: כדברי שריאל בספרו, החוקרים האנושיים היו צוואר בקבוק.[26]

האצה והוזלה אלו כרוכות בכמה סיכונים:

ראשית, **<u>הסיכון לחיי אדם</u>**. גם אם נניח שמערכות ה־AI עולות על בני אדם בביצועיהן, וגם אם נניח שימוש זהיר ואחראי במערכות אלה, הן מועדות להביא להרג המוני של אזרחים לא־מעורבים, כטעויות

ובייחוד כנזק אגבי. כדי להדגים זאת, נניח כתרגיל מחשבתי מערכת AI היפותטית, כמעט מושלמת, שמפחיתה את מספר הטעויות (זיהוי אזרחים כחברים בארגונים חמושים) ב־90 אחוזים; אם השימוש בה יגדיל פי 300 את מספר המטרות שייוצרו ואת מספר התקיפות, למרות השיפור בדיוק ברמת המקרה היחיד, הרי מערכת זו עדיין תגדיל את מספר הזיהויים המוטעים פי 30. יתרה מזו, אם מערכת בצבא כזה או אחר זיהתה 50,000 מטרות, הרי היא תביא להרג המוני מחריד: גם אם אושר ערך של נזק אגבי זהיר ביותר, של לא־מעורב יחיד לכל מטרה, וגם אם בפועל שיעור ההרג של לא־מעורבים נמוך במחצית מסף זה, היא תביא להרג 25,000 לא־מעורבים. בפועל, כמצוין לעיל, המערכות שבשימוש לוקות במגבלות דיוק והטיות משמעותיות, ובמלחמת עזה נעשה בהן שימוש בלתי זהיר בעליל: אושרו תקיפות עם נזק אגבי גבוה בהרבה מזה שבתקופה שלפניה – פגיעה בעשרות אזרחים לצורך חיסול מחבלים בדרגות זוטרות ובינוניות, ובמאות אזרחים לצורך חיסול בכירים.[27] ואולם, הניסוי המחשבתי לעיל מעיד כי הרג אזרחים והרס תשתיות אזרחיות בהיקפים מזוויעים אפשריים גם בהיעדר כשלי דיוק ושימוש לא־זהיר, שכן סיבתם אינה מגבלות טכניות של המערכת (ולכן אינה ניתנת לפתרון באמצעות שיפור המערכת ודיוקה) אלא עצם ההאצה הדרמטית בקצב של ייצור המטרות. הדבר מתאפשר משום שההוזלה וההאצה הביאו ל**הורדת הרף**. המחיר הגבוה של ייצור מטרה בשעות אדם הבטיח תעדוף. כעת אפשר להגדיר עשרות אלפים או מאות אלפים של אנשים ומבנים כמטרות, דבר שלא היה אפשרי מעשית כאשר העלות של הפללת מטרה בשעות אדם הייתה גבוהה. במלחמת עזה, לפי תחקיר *+972*, ישראל הגדירה 37,000 אנשים כמטרות שזוהו כאנשי חמאס בהסתברות של מעל 90 אחוזים – לא רק מפקדים בכירים, אלא גם חיילים זוטרים, שוטרים, ואף עובדי שירות ציבורי במנגנונים אזרחיים של ממשל חמאס. בעבר, אנשים אלה לא היו מוגדרים כמטרות להתנקשות, שכן העלות לא הייתה מעשית.[28]

שנית, **לגיטימציה להרס ולהרג חסרי־אבחנה והסרת מגבלות על אסטרטגיה צבאית.** המתואר לעיל מאפשר להשתמש ב־AI כדי לעצב

מחדש אסטרטגיה צבאית ולהרחיב את תחום הלגיטימיות הרבה מעבר למה שהתכוונו דיני הלחימה. במקרה של ישראל, הרג כלל אנשי החמאס לא היה יכול להיות מוגדר כמטרה בלא היכולת הטכנית לנבא בעלות ריאלית מי הם אנשי החמאס הזוטרים כדי להתנקש בהם בזה אחר זה כיחידים; כך מעצבות יכולות המכונה את מטרות הלחימה. הן מאפשרות לפרק מהלכים אסטרטגיים כגון מחיקה של שכונה (אם כדי לפנות דרך לכוחות קרקע ואם כמטרה בפני עצמה) לאלפי פעולות נפרדות שלכאורה אפשר להצדיקן כשההפללה נעשית בדיעבד, אחרי סימון המטרה האסטרטגית (וזאת בלי להתעלם מהעובדה שהרס שיטתי בשדה הקרב, בעזה ובכלל, נעשה גם בלא שימוש בבינה מלאכותית). בלי לקבוע מסמרות בנוגע למניעים שעיצבו את המדיניות הישראלית במלחמת עזה, אפשר לומר כי באופן כללי, לשימוש במערכות AI אוטומטיות לייצור מטרות יש פוטנציאל לשמש כלי ללגיטימציה של הרג של המוני אזרחים או של דומיסייד (מחיקת ערים העלולה להגיע לכדי טיהור אתני), המאפשר למסגרן כניתנות להצדקה (ולו בדוחק) במסגרת החוק הבין־לאומי; פעולה מכוונת יחידה להשגה התוצאה המצטברת לא הייתה לגיטימית.

שלישית, הוזלה והאצה עלולות להוביל גם ל**<u>דה־הומניזציה דיגיטלית:</u>** ארגונים כמו Human Rights Watch[29] ו־Amnesty[30] מתארים כיצד ההצבה של כלל האוכלוסייה כמטרות פוטנציאליות מאפשרת לצמצם בני אדם לכדי פרופילים, ציונים ונקודות במרחב. בשל השילוב בין מערכות טרגוט, זיהוי פנים ומעקב ביומטרי, אוכלוסייה שלמה נעשית לחשודה מטבעה. חישוב מספר הנפגעים הפך גם הוא למשתנה כמותי שאפשר לנהל באמצעות אלגוריתמים (כפי שנעשה בצבא ארצות הברית, למשל). אפקטים אלה מצטרפים לאפקטים של עצם תרגום החלטות התקיפה לפלטים מספריים, ציוני הסתברות וחישובי נזק צפוי,[31] אשר מרחיקים את מקבלי ההחלטות ריחוק קוגניטיבי ורגשי מהפגיעה הצפויה באזרחים, שמקל על קבלת החלטות תקיפה, גם כאלה הצפויות לחולל נזק נרחב.[32]

לבסוף, **האצה כרוכה בהיעדר ביקורת אנושית**: מחקרים של IEMed[33] ו־RUSI[34] מצביעים על המחיר של האצה זו: ירידה באיכות המודיעינית והעדפה ברורה של כמות על פני איכות. אף שהמערכות נועדו במוצהר לפתור צוואר בקבוק אנושי, בפועל הן יצרו מפעל מטרות שבו אין זמן לבחינה מעמיקה של כל יעד, לבדיקה ביקורתית של אמינות המידע ולצמצום זמני האישור. עוד לפני מלחמת עזה דיווחו חיילים במנהלת המטרות הישראלית על לחץ שהופעל עליהם ועל תמריצים שהוענקו להם כדי להאיץ את הקצב של ייצור המטרות.[35] במלחמה דווח כי קצב האישור האנושי להוספת מטרה לבנק ירד לשניות ספורות לכל מטרה.[36] עקב כך, המעורבות האנושית נעשתה במקרים רבים פורמלית בלבד. דו"ח של SIPRI מזהיר כי במצבים אלה מתרחש היפוך ביחסי הכוח בין האדם למכונה: מערכות שנועדו לתמוך בקבלת החלטות מכתיבות בפועל את ההחלטה, ואילו תפקידו של האדם מצטמצם לאישור טכני.[37] מצב זה יוצר טשטוש מעשי בין מערכות תומכות־החלטה לבין מערכות נשק אוטונומיות, שכן אף שההחלטה מתקבלת פורמלית בידי אדם, בפועל שיקול הדעת האנושי מצטמצם לאישור מהיר של המלצת המערכת, בלי בחינה עצמאית של הנתונים או של ההקשר המבצעי.

3. ממד הדיוק

דיוקן של מערכות המלצה על מטרות, כמו של מערכות המלצה אלגוריתמיות בכלל, הוא מוגבל. אחת הסיבות לכך היא איכות הדאטה – הן הדאטה שעליו אומנו המערכות, והן הדאטה על המטרות הפוטנציאליות המשמש לקבלת החלטה (העיקרון המוכר: Garbage In, Garbage Out). כאשר המידע שעליו נסמכת המערכת חלקי, מיושן או פגום, גם הפלט המתקבל יהיה בהכרח בעייתי. בהקשר צבאי, משמעות הדבר היא קבלת החלטות על חיים ומוות המבוססות על מידע שגוי. בשנת 2021 קבע דוח JINSA כי מערכת ה־AI ששימשה לזיהוי מטרות חמאס במבצע צבאי באותה שנה אומנה אך ורק על פוזיטיבים; כלומר המערכת לא אומנה על פיסות מידע שקצינים אנושיים סיווגו לבסוף

כלא־מטרות, שכן נתונים אלה לא נשמרו – מה שהיה צפוי לפגוע באמינותה.[38] סיבה נוספת לטעויות היא עיבוד דאטה בכלי AI כגון השימוש בכלים לעיבוד שפה טבעית ותרגום. תחקיר של *AP News* תיעד טעויות שיטתיות במערכות עיבוד ותרגום של השפה הערבית, כולל פרשנות שגויה של מילים וביטויים והפקת משפטים שלא נאמרו בפועל. כשלים לשוניים אלה אינם רק תקלות טכניות נקודתיות, אלא משקפים מגבלות עמוקות של מודלים הסתברותיים הפועלים בהקשרים תרבותיים ולשוניים מורכבים ולכן מועדים לטעויות.[39] טעויות בתרגום ובפרשנות עושים גם חוקרים אנושיים, ואולם האצת התהליך פירושה ששיעור טעות נתון עלול לייצר היקף טעויות אבסולוטי גדול בהרבה. בעיות דיוק זוהו גם בכלים המשמשים לאיתור המטרה לאחר אישורה: בדוח של Human Rights Watch נמצא כי מערכות המבוססות על טריאנגולציה סלולרית אינן מספקות רמת דיוק שדי בה לקבוע נוכחות של חשוד בנקודה מסוימת, במיוחד בסביבה שבה תשתיות תקשורת נהרסו.[40] השימוש במערכת "איפה אבא", שנועדה לזהות מתי אדם נמצא בביתו, הוביל על פי העדויות לתקיפת בתים גם כאשר היעד לא נכח בהם בפועל, וכך להגברת הסיכון לפגיעה בבני משפחה ובשכנים.[41] מהמחקר של IEMed עולה כי בהיעדר הערכות נזק שיטתיות לאחר תקיפה, אין ביכולתה של המערכת לעדכן את הנחות היסוד שלה או לתקן דפוסי טעות חוזרים, והיא ממשיכה לשכפל כשלים לאורך זמן.[42]

באופן כללי יותר, אימון המערכת על פי קבלת החלטות אנושיות מבטיח שהמערכת תשכפל ואף תעצים הטיות אנושיות ולכן הן השתקפו בנתונים שעליהם אומנו (בכלל זה, דפוסים היסטוריים של הטיות מגדריות, אתנו־גזעיות ואחרות, שכבר מוטמעות במערכות האיסוף, הסיווג והתיוג), כפי שמודגש שוב ושוב בספרות,[43] במסמך קבוצת המומחים של האו"ם[44] ובדוחות נוספים של מכוני מחקר וארגוני זכויות אדם שנבחנו בסקירה. במובן זה, ההטיה אינה בהכרח תוצר של טעות מקרית או שימוש לקוי במערכת, אלא תוצאה ישירה של בחירות תכנוניות, הנחות יסוד וארכיטקטורת המודל עצמו. דוח Human Rights

Watch מדגיש כי כאשר מערכות טרגוט מאומנות על נתוני מעקב רחבים ומוטים, הן נוטות לסמן אוכלוסיות שלמות כבעלות סיכון גבוה, גם בהיעדר אינדיקציה קונקרטית לפעילות עוינת.[45] לפיכך, הטמעת הטיות במערכות אלה אינה תופעה שולית או מקרית, אלא כשל טכני מבני בעל השלכות מבצעיות, אתיות ומשפטיות ישירות. מסמך העבודה של קבוצת המומחים של האו"ם מדגיש כי הטיה במערכות בינה מלאכותית אינה תוצר של שלב יחיד, אלא מנגנון מצטבר היכול להיכנס לכל אחד משלבי מחזור החיים של המערכת: החל בשלב איסוף הנתונים, דרך אימון המודלים והערכתם, וכלה בשלב השימוש, התחזוקה והארכוב. לפיכך, גם תיקון נקודתי בשלב אחד אינו מבטיח נטרול של ההטיה כולה, אלא נדרש מענה מערכתי לכל אורך שרשרת הפיתוח וההפעלה.

טיפוח פתרונאות

השימוש ב־AI מעודד מסגור מסוים של בעיות פוליטיות, שבו בעיות מורכבות מצטמצמות לצורך בהרג סך כל המעורבים, או הרג כל מי שקיימת הסתברות למעורבותם או למעורבותם בעתיד. זה ביטוי לאידיאולוגיה, רחבה יותר, שייבגני מורוזוב כינה "פתרונאות טכנולוגית" (technological solutionism): האמונה שאפשר לתרגם כל בעיה חברתית, מורכבת ככול שתהיה, לבעיה טכנולוגית טכנית־חישוביות. בהקשר הצבאי, זו דרך להימנע מהתמודדות עם סכסוכים היסטוריים מורכבים, עם דינמיקות פוליטיות ותרבותיות פנימיות בחברות הנלחמות, ועם השפעת הלחימה עליהם, ובמקום זאת לעסוק בדרך לזהות "סיכונים" ו"לנטרלם". בישראל, השילוב בין שלטון ימין שאינו מעוניין בפתרון הסכסוך לבין הדומיננטיות של אליטת ההיי־טק בחברה בכלל ובצבא בפרט, עודד גישה זו (למשל, בבניית מכשול הגבול כמכשול טכנולוגי, גישה שקרסה בשבעה באוקטובר) ובהישענות על כלי AI בפרט. גישה כזאת מוכרת בחשיבה הצבאית־מדינית זה שנים, אבל התעצמה בישראל בשנות האלפיים. בשנים האחרונות, התרגום של השימוש במערכות הבינה המלאכותית להרג המוני לא

יכול להתקיים בלי ההנחה הפתרונאית בצבא ובהנהגה המדינית, כי יכולת טכנולוגית משוכללת יכולה לייתר את הצורך בפתרון הבעיות המדיניות שבשורש הסכסוך הישראלי־ערבי.[46] ואולם, מרגע שכלים אלה קיימים ומאפשרים לייצר אינסוף מטרות, הם מעגנים באופן חומרי את החשיבה הפתרונאית ומעמיקים את אחיזתה בצבא ומעבר לו, וגם לצבא וגם לפוליטיקאים יש אינטרס לגבות את פיתוח החשיבה הזו בתקציבים, וזאת על חשבון התמודדות מדינית עם הבעיות שיוצרת מציאות הסכסוך.

סיכום ביניים

השימוש בבינה מלאכותית מעצב מציאות חדשה. איננו מציעים להתייחס אליה כאל כשלים או תקלות אלא כאל תהליך מובנה המשנה את אופן ההתנהלות של הלחימה. יש לכך השלכות בעייתיות שהקרינו על היקפי ההרג וההרס במלחמה האחרונה (בלי להמעיט בחשיבותם של גורמים נוספים שתרמו להיקפי ההרג וההרס), והן היכולת לפעול בקצב מהיר מאוד של איתור מטרות, להקנות לגיטימציה להרג ולהרס בהיקפים שלא היה אפשר להשיג בלא האוטומציה, ולקדם דה־הומניזציה דיגיטלית של אוכלוסיות שלמות; אימוץ היגיון חדש כדי להגדיר מהי הפללה, שאינו חופף תמיד להיגיון של חוקרים אנושיים או לנורמות המשפטיות המחייבות; נטייה מובנית לאמץ את המלצות המערכת גם בלי להבינן; קידום תפיסות פתרונאיות המתרגמות בעיות מורכבות לצורך לזהות עוד ועוד מטרות ולפגוע בהן; וזאת בצד הטיות וכשלים המאפיינים מערכות אלגוריתמיות מסוג זה באשר הן. סכנות אלה הן גם מה שמכונן את קסמן של המערכות, המייעלות מאוד את ההרג אך גם מעניקות לו הילה לגיטימציונית ולכן מעודדות את השימוש הרחב בהן ואת המשך פיתוחן.

נדגיש כי מערכות AI כשלעצמן אינן גורמות להרג המוני. הן אינן אוטונומיות ואף לא מן ההכרח שזמינותה של היכולת תכפה את אופן השימוש שעושים בה בו מפעיליה ותשליך על ארגונו של הצבא (גישה

הנגזרת מדטרמיניזם טכנולוגי). נהפוך הוא, ההרג ההמוני בעזה מקורו בהחלטות אנושיות מובהקות. עם זאת, מערכות AI מעודדות קבלת החלטות כאלה, שכן הן מספקות להן תמיכה טכנולוגית ולגיטימציונית, ומאפשרות למסגר הרג המוני כאוסף של החלטות לגיטימיות על מטרות המופללות באמצעות מידע מודיעיני. כיוון שאנו מניחים את קיומו של שיקול דעת אנושי רחב, הוא שעומד במוקד עניינו בנייר זה.

חלקי המסמך שלהלן מציעים דרכים להתמודדות, אף אם חלקית מטבעה, עם הסכנות שמנינו למעלה.

חלק ב': נקודות מוצא למדיניות

הנקודות שלהלן הן הנחות מוצא לעיצוב המלצות המדיניות המפורטות בפרק הבא:

1. עצם הטמעתן של מערכות הבינה המלאכותית במלחמה היא מהלך בלתי הפיך, לפי הנחתנו. אם שיח אקדמי ביקורתי יערער על עצם ההטמעה הזאת, הרי שיח המכוון למדיניות, כמו נייר זה, חותר למזעור הסכנות הכרוכות בשימוש במערכות הבינה המלאכותית ובוודאי בהרחבתן.

2. התכלית של החתירה להגבלת השימוש במערכות בינה מלאכותית היא מוסרית: צמצום מהותי של הפגיעה בחפים מפשע. במלחמת עזה חלה פגיעה חסרת תקדים, גם בקנה מידה עולמי, באזרחים המוגנים בידי המשפט הבין־לאומי, לצד הרס מתחמים מאוכלסים שלמים. השימוש במערכות הבינה המלאכותית סייע לגרימת שיעורים אלה של הרג והרס.

3. צמצום מהותי של פגיעה בחפים מפשע משרתת גם את צורכי הלגיטימציה של המדינה בקרב הקהילה הבין־לאומית, לגיטימציה שנפגעה קשות במלחמת עזה.

4. הסיכונים המוצגים לעיל הם מהותיים, ואי אפשר להתמודד איתם באמצעות שיפור טכנולוגי בלבד (עמדה אחרת תהיה אימוץ מחדש

של הפרדיגמה הפתרונאית על כלל מגבלותיה); לשם כך דרוש שינוי רגולטורי המגדיר מחדש את תנאי הלגיטימציה להפעלת מערכות הבינה המלאכותית.

5. מועצת אירופה עיצבה אמנת מסגרת לבינה מלאכותית, לזכויות אדם, לדמוקרטיה ולשלטון החוק, המספקת תבנית נורמטיבית רלוונטית לשינוי מסוג זה,[47] אם כי המסמך האירופי אינו חל פורמלית על ענייני ביטחון במישרין. גם אם למסמך המועצה האירופית יש תרומה, חשוב לזכור כי הפעלת מערכות הבינה המלאכותית לצורכי לחימה נעשית בעולם המערבי בלי הסדרה משפטית ואתית המתמודדת באפקטיביות עם הסכנות שמנינו. פסגה בין־לאומית העוסקת ביישומים צבאיים של AI, שהתקיימה בהאג בשנת 2023, הסתיימה ב־Call to Action לא־מחייב בחתימת עשרות מדינות, שההסכמות שבו כלליות ומופשטות ביותר, כגון הצורך בפיקוח אנושי על מערכות AI ובשמירה על מגבלות הדין הבין־לאומי, והן רחוקות ממתן מענה מפורט לאתגרים שמציבים יישומים אלה. נייר זה צריך להתמודד עם לקונה זו וגם לתרום לצמצומה.[48]

6. כאמור בחלק הראשון, מערכות AI כשלעצמן אינן גורמות להרג המוני אבל מעודדות קבלת החלטות המחוללות אותו.

7. מאחר שההשפעה הבולטת ביותר של השימוש במערכות AI לייצור מטרות היא האצה והוזלה דרמטיים של התהליך, ומאחר שאלה סיכונים שאין להם פתרון טכנולוגי פשוט, יש צורך ברגולציה המביאה בחשבון את הסיכונים הכרוכים בהאצה ומציעה דרך לרסנה.

8. את המסמך הזה הנחו בעיקר שיקולים מוסריים, ואולם אין להניח משחק סכום־אפס שבו הפחתת סיכונים מוסריים מנוגדת בהכרח לשיקולים צבאיים פרגמטיים. מנקודת מבט צבאית, אין בהכרח תועלת צבאית בהרחבה דרמטית של הרג אזרחים והרס תשתיות

אזרחיות; יתר על כן, היא כרוכה בעלויות, בבזבוז תחמושת, בסיכונים להארכת הלחימה, ובסיכונים אסטרטגיים לטווח הרחוק.

חלק ג': פירוט ההמלצות

1. **<u>ידע משחזר כתנאי לאימוץ המלצות:</u>** ההמלצה המרכזית של המסמך היא לחייב כי לפני כל אישור של מטרה שהמליצה מערכת המלצה תידרש עבודה אנושית לפתיחת הקופסה השחורה באמצעות **ידע משחזר**, כלומר בחינה של חומרי גלם בכל מקרה, ניסיון לשחזר את ההיגיון שהנחה את המערכת בסיווג האדם או האתר כמטרה; וביסוס הפללה שאינה מבוססת על הקשות סטטיסטיות מעורפלות אלא על ראיות העומדות באמות המידה הנדרשות מהפללה אנושית, תוך השקעת שעות האדם הנדרשות לשם כך והטלת **חובת הנמקה** על מקבלי ההחלטות האנושיים. יש לכך חשיבות מרובה לנוכח קשיי ההסברתיות של מערכות בינה מלאכותית ("הקופסה השחורה"), כלומר הקושי להבין כיצד המערכת הגיעה להמלצות שאליהן הגיעה,[49] קושי שבהקשר של ייצור מטרות מעלה חשש לאי־עמידה בחובות של הדין הבין־לאומי ההומניטרי ודיני זכויות האדם. לטענת צה"ל, מערכת "הבשורה" בנויה כך שהיא מאפשרת ייצור ידע משחזר, שכן היא "מספקת לחוקר את המידע עליו היא מתבססת, באופן המאפשר לחוקר לאסוף ולהבין בקלות את המידע שעליו נשענה, ואז לערוך בחינה עצמאית אנושית של החומר המודיעיני".[50] אנו ממליצים שהידע הזה ישמש כבסיס לקבלת החלטות ולא רק לשחזור שלאחר מעשה.

2. לפי עיקרון זה, **המערכת תוכל להשתמש במתאמים סטטיסטיים רק כדי להחשיד – אבל בשום פנים ואופן לא כדי להפליל** (ובדומה לכך, יש לאסור גם על אנשים להשתמש במתאמים סטטיסטיים כבסיס לקבלת החלטות של חיים ומוות). שימוש בהמלצות AI בלא ייצור ידע משחזר ובלי לבסס הפללה על סטנדרטים אלה אינו לגיטימי מהבחינה המוסרית, ואף **אסור מהבחינה המשפטית,** שכן אין לראותו כעומד בעקרונות

המשפט הבין־לאומי (הדורשים זהירות, הבחנה בין מטרות צבאיות לאוכלוסייה אזרחית מוגנת, ואחריות אנושית לקבלת ההחלטות). המלצה זו תסייע להפחית את הסיכונים המתוארים לעיל:

- היא מבטיחה שהפללה לא תתבסס אך ורק על סטטיסטיקה (כלומר על ראיות לא־מפלילות קיימות המנבאות ראיות מפלילות שאינן קיימות). כך יימנעו תקיפות חתימה, שעוד לפני עידן ה־AI נכתב רבות על מחירן הקיצוני בחיי לא־מעורבים.
- היא מבטיחה שהאדם שבלופ לא יהיה חותמת גומי (בדומה לנהג העוקב אחרי הנחיות ווייז), אלא מי שמבצע את שיקול הדעת ומקבל את ההחלטה. כך יוסב מוקד הכוח לשליטה אנושית (from human-in-the-loop to human-in-power): המערכת מספקת נתונים והם נבדקים בידי הצוות האנושי כדי להצדיק את הפגיעה בהם.[51] כחלק מזה יש לצמצם את הביזוריות של המערכת המאפיינת אותה היום (לפחות בישראל) ותורמת לפיצול האחריות. יש צורך לקבוע מנהל.ת עם אחריות לכל התהליך ובפרט להחלטה הסופית. קבלת החלטות מבוזרת מחלישה את האחריות של כל דרג, ושילוב בינה מלאכותית בתהליך מוסיף ומעצים את עמעום האחריות. סכנת הביזור אף מתעצמת ככל שמורד דרג ההפעלה של כלי הבינה, למשל לדרג הטקטי של צבא היבשה, העתיד לשלב כלי בינה במערכת צבא היבשה הדיגיטלי (צי"ד).[52]
- בהמשך לכך, היא מאטה דרמטית את קצב ייצור המטרות, משום שאינה מאפשרת לוותר על שיקול הדעת האנושי, וכך מונעת את הורדת הרף וההפללה ההמונית של מטרות, שעשויה לשמש בנסיבות מסוימות כלי ללגיטימציה לאורביסייד (מחיקת שכונות),[53] להרג המוני, או לטיהור אתני.
- עם זאת, המלצה זו אינה אוסרת כליל על שימוש ב־AI כדי להחשיד מטרות ולתעדף את סדר בחינתן, וכך להתמודד עם

כמויות נתונים עצומות ועם שינויים מהירים בזמן מלחמה, בניסיון לזהות את המטרות בעלות הערך הצבאי הרב ביותר.

יודגש, כי ניסיון להשיג הסבריות אוטומטית, כחלק מאופן פעולת המערכת, לא ישיג מטרות אלה – הן מסיבות טכניות (ברמת מורכבות גבוה של מודלים אי אפשר לפתוח את הקופסה השחורה) והן מסיבות מהותיות – אוטומציה של הסבריות לא תחזיר את האדם ללופ, לא תסגור את הפער בין האפיסטמולוגיה של AI לאפיסטמולוגיה האנושית, ולא תמנע האצה אקספוננציאלית ואת האפשרויות לניצולה לרעה.

3. **<u>העלאת סף ההוכחות הנדרש להפללה מודיעינית:</u>** כפי שטוען דוד אנוך, האצת ייצור המטרות בידי AI עשויה לאפשר להשיג אותה התועלת הצבאית תוך התעקשות על סף הוכחה גבוה יותר.[54] למעשה, אם אכן מסוגלות מערכות AI להפחית את שיעור השגיאות ברמת ההפללה היחידה, הדבר עשוי לאפשר השגת תוצאות דומות תוך הפחתת מספר המטרות ולא הגדלתן. המלצה זו וקודמתה יסייעו למנוע את ההכפלה המעריכית של נזק אגבי שנבעה במלחמה האחרונה מהורדה משמעותית של דרג הבכירות של המטרות האנושיות.

4. **<u>תקני מינימום לזמן אדם לאישור מטרה:</u>** בהמשך לכך, וכדי להבטיח שההמלצה הקודמת תמומש, יש לקבוע תקן ברור ומושכל להיקף המינימלי של שעות האדם הדרוש לייצור הידע המשחזר לפני הכנסת מטרה לבנק המטרות. על תקן זה לשקף נאמנה את הזמן הדרוש לתהליכים אלה. המלצה זו תבטיח שאכן בן־אנוש יקבל את ההחלטה. לטענת צה"ל, על חוקרים אנושיים לקיים בדיקה עצמאית של המלצות המערכת.[55] ואולם, יש עדויות לכך שבמלחמת עזה, עקב הניסיון להאיץ את ייצור המטרות, חלק ניכר מהמטרות אושרו בתוך זמן קצר, שאינו מאפשר בדיקה משמעותית ושיקול דעת מהותי; הוא הופך את המעורבות האנושית בתהליך לנומינלית בלבד, המקנה לגיטימציה אך אינה משפיעה מהותית על ההחלטות המתקבלות.

יש לציין כי הרף שמציב הדין ההומניטרי לקבלת החלטה צבאית העלולה לפגוע בחפים מפשע הוא של "המפקד הצבאי הסביר", כלומר בעת בחינת החלטה מסוימת נשאל את עצמנו האם מפקד סביר היה מקבל אותה ההחלטה באותן הנסיבות.[56] לפיכך, אין הצדקה למצב שבו מערכת בינה מלאכותית בעלת אופי התקפי תורה על תקיפה של יחידים בלי כל מעורבות אנושית.[57] יתר על כן, הכללת גורם אנושי בתהליך קבלת ההחלטות חיונית מהבחינה האחריותיות. במצב שבו בוצע פשע מלחמה עקב החלטה שגויה של מערכת המבוססת על בינה מלאכותית, ואין גורם אנושי המעורב בקבלת ההחלטה, אי אפשר לקבוע מי הגורם שיישא באחריות על הפשע.[58] הגברת המעורבות האנושית גם תקהה את קסמן של מערכות הבינה המסייעות ללגיטימציה פנים־צבאית.[59]

5. **<u>הגדרה מחדש של יחידת הניתוח המשפטית:</u>** הדין הבין־לאומי תופס מידתיות במושגים של תקיפה יחידה, שהיא יחידת הניתוח. יחידות ניתוח רחבות מאומצות רק כשמדובר בפשעים נגד האנושות כמו ג'נוסייד, טיהור אתני או פשעים נגד האנושות. ואולם השימוש ב־AI מאפשר להאיץ את קצב ייצור המטרות, ולייצר עשרות אלפי או מאות אלפי מטרות, המשמשות ברובן לתקיפה יזומה. בתנאים אלה, גם אם כל תקיפה בפני עצמה היא מידתית על פי המשפט ההומניטרי (הנזק האגבי שלה קטן מהתועלת הצבאית שתביא), הרי כמכלול ובמצטבר הן מייצרות אפקט לא־מידתי. ההיגיון הפלילי שמאחורי פרסונליזציית הלחימה אינו מתאים להתמודדות עם מה שמתרחש כאשר מדיניות ההתנקשויות עוברת הרחבה משמעותית (Upscaling), מרמת המקרה היחיד למהלך מלחמתי נרחב. אלא שלעיתים, פעולות שאפשר להצדיקן כל אחת בנפרד אינן ניתנות להצדקה מוסרית ומשפטית כמכלול.[60] ההסתכלות הפרטנית (ראו למעלה) מגבירה את חופש הפעולה של מפעילי החימוש, בפרט בתנאים של ריבוי מטרות וקצב אש מואץ, בכך שהיא מכפיפה אותה לשיקול הדעת של המפקד הסביר ולסיטואציה מבצעית נתונה. לדוגמה, במסגרת הכללים הקיימים קטן הסיכוי שתישאל השאלה עד כמה נכון להרוג מספר חזוי של חפים מפשע כדי להשיג

את היעד לחסל את הנהגתו של ארגון מזוין כמכלול, ולא כל בכיר בפני עצמו. לפיכך, דרושות התאמות של המשפט הבין־לאומי למצב החדש, שיגדירו מחדש את יחידת הניתוח המשפטית ויאפשרו ביקורת משפטית החורגת מרמת התקיפה היחידה. כמובן, מהלך זה הוא קומה נוספת הנבנית על קומת היסוד שיש לשמרה בלי קשר לשימוש ב־AI: מחויבות של אמת לעקרונות המשפט הבין־לאומי, ובכלל זה מידתיות (להבדיל משימוש נומינלי במושגים אלה להצדקת פעולות שאינן ניתנות להצדקה במסגרתו). לכאורה זו המלצה מובנת מאליה, אך בעידן שבו עצם המושג חוק בין־לאומי נמצא תחת מתקפה בארצות הברית, בישראל ובמדינות נוספות, הן באמירות מפורשות והן בהתעלמות ממנו, אין מנוס מהכללתה.

6. **קריטריונים:** הקלות של ייצור מטרות, והצורך לקבל החלטות בתנאי לחץ עשויים לעודד תקיפה של אנשים ומבנים בשל עצם היכולת להגדירם כמטרות, ובלא תלות בתועלות ובמחירים הכרוכים בכך. כדי להפחית סיכון זה, ראוי כי צבאות יגדירו **מראש** קריטריונים אסטרטגיים ואתיים שאליהם יוכפפו תקיפות. עם שאלות אלה נמנות (ואין זו רשימה סגורה): (א) מטרות הלחימה שאותן אמורה כל תקיפה לשרת, והיקף המטרות שבהן יש לפגוע. ההגדרה של בכירות הלוחמים שבהם יש לפגוע תיגזר מזה ולא מיכולת המערכות. במסגרת זו נכון היה שתישאל בישראל השאלה האסטרטגית אם נכון לכלול 37 אלף איש כיעדי פגיעה במלחמה; (ב) מהי הגדרה של לוחם שאפשר לפגוע בו (כידוע ישראל מגדירה גם עובדי השירות הציבורי כלוחמים); (ג) מהי מידת הנזק האגבי המותר בתקיפת מטרות צבאיות ברמות שונות כנגזר מעקרונות המידתיות; (ד) מהו סוג החימוש שבו יש להשתמש במצבים שונים; (ה) אם יש לאפשר אזהרה מוקדמת כדי לצמצם פגיעה באזרחים; (ו) מהם המשתנים האסורים בשימוש המערכות לצורך הפללה, או שמראש יסומנו כבעייתיות (דוגמת טריאנגולציה סלולרית). קריטריונים אלה צריכים להיקבע מראש כדי למנוע כרסום בהם בעת מלחמה, כדי להתפרסם ולהיות נתונים לפיקוח משפטי ופוליטי (כולל פרלמנטרי) קפדניים.

7. **<u>התגברות על הטיות האוטומציה:</u>** מעורבות אנושית מוגברת המאטה את הקצב של ייצור המטרות ומייצרת ידע משחזר יכולה לסייע בהתגברות על תסמונת "הטיית האוטומציה", אך אין בה די להתגברות על הטיה זו. כפי שטען סמית' (Smith) בהשראת חנה ארנדט, כששיקול הדעת מועבר לאלגוריתמים, אובד הספק בדבר ייצוג המציאות וההנחות שלנו, וגוברת ההסתברות לאלימות.[61] משום כך נדרשת חשיבה ואימוץ של מתודולוגיות ספציפיות שיוטמעו בתהליכי העבודה ויצמצמו את ההטיה.[62]

אופן הצגת הנתונים: הצגת הסתברויות מספריות למקבלי ההחלטות האנושיים (למשל, מטרה חשודה ב־90 אחוזים) מעצימה את הטיית האוטומציה, לנוכח הנטייה האנושית לתת אמון במספרים הנתפסים כאובייקטיביים וכסמכותיים,[63] ולנוכח המסגור של "המלצה" שאימוצה כרוך בפחות לקיחת אחריות וסיכון מאשר אי־אימוצה (למשל, חברי מושבעים בארצות הברית נוטים להרשיע ברשלנות רפואית מומחים שדחו את המלצת ה־AI יותר מכאלה שאימצו אותה).[64] משום כך, אין די בהפיכת בן האנוש למאשר בלבד של נתוני המערכת. הפתרון לכך הוא שהמערכת תכונן סדר קדימויות על פי ההסתברות המשוערת (כלומר תציג תחילה את מי שלפי תחזיתה יאושרו כמטרות בהסתברות גבוהה), אך לא תחלוק עם משתמשי הקצה את ערך ההסתברות ולא תציג "המלצה" לאישור.

בהמשך לכך, וכדי למנוע את התרחיש שמתאר ברנסטיין, יש לטפח תרבות ארגונית המתגמלת על ספקנות ועל זיהוי טעויות ולא רק על תפוקה כמותית, ואף מינוי אנשים או צוותים שתפקידם לבקר ולהטיל ספק בזמן אמת (red teams), והם רב־תחומיים וכוללים גם דיסציפלינות שאינן חלק מצוותי העבודה כדי לאפשר גיוון חשיבה.

8. **<u>חיזוק מעמדם של משפטנים בקבלת ההחלטות:</u>** לכניסתה של הבינה המלאכותית יש פוטנציאל מובהק להחליש את השפעתם של משפטנים על קבלת החלטות תקיפה ולכן גם להחלשת הביקורת השיפוטית, וזאת משתי סיבות: הן בשל ההאצה, והפחתת שיקול הדעת האנושי בכללותו בעקבות האוטומציה החלקית של קבלת

החלטות, והן בשל עיגונן של נורמות משפטיות ברמת הקוד שעלולה להיתפס כמייתרת שיקול דעת משפטי – וזאת אף על פי ששיקול דעת אנושי הוא בגדר חובה משפטית. כדי לתקן זאת דרושה ביקורת משפטית אפקטיבית שאינה טקסית: זמן, כלים ומנדט אמיתי לדרג המשפטי להטיל וטו על החלטות, וכן ידע והבנה טכנית של הדרגים המשפטיים במערכות שמייצרות את המטרות.

9. **<u>שיפור מערכות:</u>** אם נעשה שימוש במערכות בינה מלאכותית, יש להקפיד להשקיע בשיפורן המתמיד, ובכלל זה: לאמן מערכות על נגטיבים, כלומר על מקרים שהוחלט לסווגם כלא־מטרות; לתעד את שיטות הפיתוח של המערכות ואימון המודלים ולשתף מידע לשם זיהוי סיכונים אפשריים; לערוך סימולציות, פיילוטים וניסויים בתסריטים שונים; להקדיש משאבים מערכתיים לתחקירים ולזיהוי שגיאות, כדי לאפשר זיהוי שגיאות טיפוסיות של מערכות ה־AI ותיקונן (כגון דפוסי התנהגות חשודים שיכולים לאפיין גם את מי שאינם לוחמים); לתת משוב למערכות כדי לאפשר שיפור אוטומטי: הן משוב על מקרים שבהם חוקרים אנושיים לא אימצו את המלצות המערכת והן על מקרים שבהם המערכות אומצו והסתבר בדיעבד שאומצו בטעות; לזהות הטיות מובנות במערכות; ולזהות סיכונים הנובעים מהשילוב והאינטראקציה בין מערכות שונות. את הפיקוח והביקורת על המערכות יש להפקיד בידי צוותי פיקוח רב־תחומיים הכוללים גם דיסציפלינות שאינן חלק מצוותי העבודה כדי לאפשר גיוון חשיבה.[65] **עם זאת, יודגש כי שיפור המערכות כשלעצמו <u>אינו</u> מסוגל להתמודד עם הסיכונים העיקריים של השימוש במערכות בינה מלאכותית, ואין להתפתות לראות בו פתרון קסם.**

10. **<u>איסור על נשק אוטונומי:</u>** הסיכונים שנסקרו לעיל הביאו להמלצות לריסון השימוש ב־AI בייצור מטרות. מערכות נשק התקפיות אוטונומיות מאופיינות ברבים מסיכונים אלה (למשל הטיות המעוגנות בדאטה, פתרונאות טכנולוגית, היעדר שיקול דעת אנושי בהחלטות של חיים ומוות, אפיסטמולוגיה השונה מזו האנושית ולכן

מאתגרת נורמות מוסר אנושיות), לצד סיכונים נוספים שעלולים להביא להרג המוני בלתי־מכוון (כמו ה־flash crash או היכולת לשטות במערכות אלה),[66] ולכן קל וחומר שיש לשלול הטמעה של מערכות נשק אוטונומיות. מערכות אלו אינן מוקדו של מסמך זה, ואולם לנוכח הלכי רוח רווחים בצבא המצדדים במעבר לכלי נשק אוטונומיים, יש להדגיש את החשיבות שבדבקות בקונצנזוס הרווח במדינות מערביות כי יש למנוע מהלך זה, המעצים את הכשלים של מערכות הבינה המלאכותית ומחליש פיקוח אנושי.[67]

11. **הטמעה**: הדיון שהתנהל כאן אסור שיוגבל לשדה הביקורת ולאקדמאים המבקרים את המערכות הצבאיות מבחוץ: יש לקיים דיון בסוגיות אלה בתוך המערכות הצבאיות, ולעגן את הנורמות שהומלצו כאן לא רק בכללים משפטיים ובנהלי עבודה, אלא גם בתהליכים פנים־צבאיים של עיצוב תורת לחימה (תו"ל), קוד אתי וסוציאליזציה אתית. תהליכי עבודה לבדם אינם יכולים לתקן את המעוות. למשל, אין כל נוהל שיצמצם דה־הומניזציה של אזרחי אויב. התהליך המתבקש חייב לכלול הגבלות משפטיות, תו"ל מחייב וסוציאליזציה אתית, משולבים זה בזה ומפוקחים בידי הרמות הבכירות ביותר בצבא.

סיכום

פתחנו ואמרנו שישראל עשתה שימוש נרחב במערכות בינה מלאכותית לזיהוי מטרות במלחמת עזה. ההאצה וההוזלה הדרמטיות של ייצור המטרות שאפשרו מערכות אלה; הטיותיהן, טעויותיהן ומגבלותיהן; והלגיטימציה שהן מספקות – אחראיות לחלק ניכר מההרג וההרס חסרי התקדים שמלחמה זו יצרה. מטרת ההמלצות היא ליצור תנאים שיצמצמו בעתיד נזקים אלה באמצעות החזרה של ההיגיון האנושי והבקרה האנושית, והאטה של התהליכים כך שתִּימָּנע הצטברות של פעולות הנראות כל אחת כשלעצמה מידתית ולגיטימית ליצירת אפקט מצטבר של דומיסייד והרג המוני. אין זו נוסחת פלא, שהרי בני אנוש הם הנושאים באחריות לתוצאות ההרסניות של המלחמה. עם

זאת, חיזוק האחריות האנושית עשוי ליצור תנאים ממתנים שימנעו לגיטימציה חוקית למעשי זוועות בהיקפים חסרי תקדים וישפרו את הבקרה החיצונית על האופן שבו מפעיל הצבא את חימושו.

בזמן העבודה על מסמך זה היינו עדים לכניסה של טכנולוגיות AI לייצור מטרות בצבאות ברחבי העולם, ואין ספק שמגמה זו תלך ותגבר. הפקת לקחים מהמקרה הישראלי של מלחמת עזה היא חיונית אפוא, לא רק בהקשר המקומי, אלא בעולם כולו. אנו מקווים כי ההמלצות המופיעות במסמך זה יתרמו לתהליך של רפלקסיה על טכנולוגיות אלה ולריסון הסכנות הגלומות בהן.

הערות

1 Anthony King, "Digital Targeting: Artificial Intelligence, Data, and Military Intelligence", *Journal of Global Security Studies* 9, 2 (2024): 1–16

2 הסדנה כללה הצגת ניירות פתיחה מפי ד"ר סבסטיאן בן־דניאל, ד"ר ראובן גל, פרופ' יגיל לוי, מר אדם רז ופרופ' אורי שורץ. כן השתתפו בתהליך החשיבה ד"ר תמר ברקאי, מר אבנר גבריהו, ד"ר ליאור יוחנני, ד"ר טל מימרן וד"ר דמיטרי סקולסקי. את סקירת הספרות ערכה נועה המאירי. אנו מודים לפרופ' דוד אנוך על הערותיו המועילות. הנייר נכתב על דעת המחברים בלבד.

3 Y.S (Brigadier General), *The Human Machine Team: How to Create Synergy Between Human & Artificial Intelligence that Will Revolutionize Our World (*eBookPro Publishing, 2021), 9

4 לדיונים בשימוש בכלים אלה בצבא ארצות הברית ולהשלכותיהם הבעייתיות, ראו:

Hendrik Huelss, "Norms Are What Machines Make of Them: Autonomous Weapons Systems and the Normative Implications of Human-Machine Interactions", *International Political Sociology* 14, 2 (2019): 111–128; Anthony King, "Digital Targeting: Artificial Intelligence, Data, and Military Intelligence", *Journal of Global Security Studies* 9, 2 (2024): 1–16; Lucy Suchman, "Algorithmic Warfare and the Reinvention of Accuracy", *Critical Studies on Security* 8, 2 (2020): 175–187; Jutta Weber, "Keep Adding. On Kill Lists, Drone Warfare and the Politics of Databases", *Environment and Planning D Society and Space* 34, 1 (2015): 107–125; Lauren Wilcox, "Embodying Algorithmic War: Gender, Race, and the Posthuman in Drone Warfare", *Security Dialogue* 48, 1 (2016): 11–28; Christiane Wilke, "Seeing and Unmaking Civilians in Afghanistan", *Science Technology & Human Values* 42, 6 (2017): 1031–1060

5 Elizabeth Dwoskin,"Israel built an 'AI factory' for war. It unleashed it in Gaza", *The Washington Post*, December 29, 2024, Link

6 Rob Kitchin, "Big Data, New Epistemologies and Paradigm Shifts", *Big Data & Society* 1, 1 (2014); Yoav Mehozay and Eran Fisher, "The Epistemology of Algorithmic Risk Assessment and the Path Towards a Non-penology Penology", *Punishment & Society* 21, 5 (2018): 523–541

7 ראו: Vincent Chiao, "Algorithmic Decision-making, Statistical Evidence and the Rule of Law", *Episteme* 21, 4 (2023): 1241–1264; Lewis D. Ross, "Recent Work on the Proof Paradox," *Philosophy Compass* 15, 6 (2020)

8 Forrest Morgan et al., *Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World* (RAND, 2020)

9 Christina Balis and Paul O'Neill, *Trust in AI: Rethinking future command* (RUSI, 2022)

10 Yavar Bathaee, "The Artificial Intelligence Black Box and the Failure of Intent and Causation", *Harvard Journal of Law & Technology* 31 (2018): 889–938; Carlos Zednik, "Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence", *Philosophy & Technology* 34, 2 (2021): 265–288

11 למשל:

Jenna Burrell, "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms", *Big Data & Society* 3, 1 (2016)

12 כלים אלו מנסים להסיק זאת בדיעבד, למשל באמצעות ניסיון לזהות אילו שינויים בקלט היו מביאים לשינוי ההמלצה. לניתוח של (חלק) ממגבלותיהם המוסריות בהקשרים צבאיים ואחרים, ראו: Isaac Taylor, "Is Explainable AI Responsible AI?" *AI & Society* 40, 3 (2025): 1695–1704

13 UN Human Rights Committee (HRC), *General comment no. 31 [80], The nature of the general legal obligation imposed on States Parties to the Covenant*, CCPR/C/21/Rev.1/Add.13, May 26, 2004, para.8

14 Linda J. Skitka, Kathleen L. Mosier, and Mark Burdick, "Does Automation Bias Decision-making?", *International Journal of Human-Computer Studies* 51, 5 (1999): 991–1006

15 Morgan et al., Military Applications

16 Alexander Blanchard and Laura Bruun, *Autonomous Weapon Systems and AI-Enabled Decision Support Systems in Military Targeting: A Comparison and Recommended Policy Responses* (SIPRI, 2025)

17 Sarah Lebovitz, Hila Lifshitz-Assaf, and Natalia Levina, "To Engage or Not to Engage with AI for Critical Judgments: How Professionals Deal with Opacity When Using AI for Medical Diagnosis", *Organization Science* 33, 1 (2022): 126–148

18 Yuval Abraham, "'Lavender': The AI Machine Directing Israel's Bombing Spree in Gaza", *+972 Magazine*, April 3, 2024, Link

19 Frank Pasquale, "Legal Automation, Rule by Law and the Rule of Law" (paper presented at ICON-S Annual Conference, June 27, 2018, Hong Kong) https://bit.ly/2DlB3wa

20 Laura Bruun, Dr Marta Bo and Netta Goussac, *Compliance with International Humanitarian Law in the Development and Use of Autonomous Weapon Systems: What does IHL Permit, Prohibit and Require?* (SIPRI, 2023)

21 Sebastian Clapp, *Defence and artificial intelligence*, Briefing No. 769580 (European Parliamentary Research Service, 2025)

22 יגיל לוי, "בינה מלאכותית כמערכת לגיטימציה", בתוך: **המלחמה האלגוריתמית: בינה מלאכותית, לגיטימציה והפוליטיקה של האלימות במלחמת עזה**, עריכה: יגיל לוי (מכון האוניברסיטה הפתוחה לחקר יחסי חברה־צבא, 2026), 61–70

23 Yuval Abraham, "'A mass assassination factory': Inside Israel's calculated bombing of Gaza", *+972 Magazine*, November 30, 2023, Link

24 Dwoskin, Israel built an 'AI factory' for war

25 רון לשם, "אביב כוכבי: 'צה"ל לא דומה למה שהיה לפני עשר שנים. יש מודיעין בזמן אמת כמו ב'מטריקס', היקף הפצצות והטילים גדל", **ידיעות אחרונות**, 23 ביוני 2023

26 Y.S, The Human Machine Team

27 Patrick Kingsley et. Al., "Israel loosened its rules to bomb Hamas fighters, killing many more civilians", *New York Times*, December 26, 2024

28 Abraham, Lavender: The AI Machine

29 Human Rights Watch, "Questions and Answers: Israeli Military's Use of Digital Tools in Gaza", September 10, 2024, Link

30 Amnesty International, "Israel's apartheid against Palestinians: a cruel system of domination and a crime against humanity", February 1, 2022, Link

31 Noah Sylvia, *The Israel Defense Forces' use of AI in Gaza: A case of misplaced Purpose* (RUSI, 2024)

32 Joanna L D Wilson, *AI, war and (in)humanity: the role of human emotions in military decision-making*, ICRC Humanitarian Law & Policy Blog, February 20, 2025

33 Noah Sylvia, *The Israeli Military's Use of AI in Gaza: Operational Efficiency at the Cost of Humanity* (IEMed, 2025)

34 Sylvia, The Israel Defense Forces' use of AI

35 יניב קובוביץ, '"ההצלחה נמדדת בכמה מטרות חדשות מייצרים': כך נבנה הבנק של צה"ל בעזה", **הארץ**, 15 בדצמבר 2019

36 Abraham, Lavender: The AI Machine

37 Blanchard and Bruun, Autonomous Weapon Systems

38 JINSA's Gemunder Center Gaza Assessment Policy Project, *Gaza Conflict 2021: Assessment: Observations and Lessons* (The Jewish Institute for National Security of America – JINSA, October 2021)

39 Sam Mednick, Garance Burke and Michael Biesecker, "How US Tech Giants Supplied Israel with AI Models, Raising Questions about Tech's Role in Warfare", *AP News*, February 18, 2025

40 Human Rights Watch, Questions and Answers

41 Abraham, Lavender: The AI Machine

42 . Sylvia, Operational Efficiency at the Cost of Humanity

43 על ליקויים דומים בשימוש ב־AI בשיטור אזרחי ראו:

Sarah Brayne, "Big Data Surveillance: The Case of Policing", *American Sociological Review* 82, 5 (2017): 977–1008

44 United Nations, Group of Governmental Experts on Lethal Autonomous Weapons Systems, *Addressing Bias in Autonomous Weapons*, UN Doc. CCW/GGE.1/2024/WP.5 (March 4, 2024), working paper submitted by Austria, Belgium, Canada, et al

45 Human Rights Watch, "Gaza: Israeli Military's Digital Tools Risk Civilian Harm", September 10, 2024, Link

46 יגיל לוי, **יורים ולא בוכים – המיליטריזציה החדשה של ישראל בשנות האלפיים** (למדא עיון, 2023), 102–105

47 Council of Europe, "The Framework Convention on Artificial Intelligence", September 5, 2024, Link

48 Government of the Netherlands, "REAIM 2023 Call to Action" Link

49 Bathaee, The Artificial Intelligence; Zednik, Solving the Black Box Problem

50 דובר צה"ל, בקשתך למידע בנושא מידע בעניין מערכת 'הבשורה', 24 בפברואר 2024, קישור

51 Elise Li Zheng et al., "From Human-In-The-Loop to Human-In-Power", *The American Journal of Bioethics* 24, 9 (2024): 84–86

52 יורם גביזון, "אלביט זכתה בחוזה של 100 מיליון דולר לפיתוח הדור הבא של צבא היבשה הדיגיטלי", **דה מרקר**, 9 בפברואר 2026

53 Martin Coward, *Urbicide: The Politics of Urban Destruction* (Routledge, 2009)

54 המלצה זו נשענת על הצעותיו של דוד אנוך. ראו: David Enoch, "Smaller Chance of Mistake, More Innocents Dead: A Different Problem with Using AI Systems in War (and elsewhere)", (unpublished manuscript/ working paper)
הניתוח של אנוך מוביל למסקנה כי אפשר להצדיק שימוש במערכות AI לשם שיפור הדיוק של תקיפות, אך לא בהכרח לשם הרחבה דרמטית של היקף המטרות והתקיפות. הרחבה כזאת משנה את 'רמות האכיפה' של הפעלת הכוח ועלולה להגדיל את מספר האזרחים הנפגעים, גם כאשר רמת הדיוק המקומית משתפרת. בהקשר של עזה, דינמיקה זו נקשרה גם להרחבה של מאגרי המטרות ולהורדת דרג הבכירות של מטרות אנושיות.

55 דובר צה"ל, בקשתך למידע

56 בג"ץ 769/02 **הוועד הציבורי נגד העינויים בישראל נ' ממשלת ישראל**, פ"ד נב(2) 507, 512 (2006); Ian Henderson and Kate Reece, "Proportionality Under International Humanitarian Law (IHL): The 'Reasonable Military Commander' Standard and Reverberating Effects", *Vanderbilt Journal of Transnational Law*, 51, 3 (2018): 835–855

57 Robert D. Sloane, "Puzzles of Proportion and the Reasonable Military Commander: Reflections on the Law, Ethics, and Geopolitics of Proportionality", *Harvard National Security Journal* (2015): 299–343

58 Kevin Jon Heller, "The Concept of 'The Human' in the Critique of Autonomous Weapons", *Harvard National Security Journal* (2023): 1–76

59 ראו: לוי, בינה מלאכותית כמערכת לגיטימציה

60 אנוך (ר' הערת שוליים 54). טוען כי לפער זה בין המוסריות של כל תקיפה בנפרד למוסריותן המצטברת יש השלכות על הדין הרצוי. ברוח דומה, הוא מצביע על הקושי שבבחינת המידתיות של כל תקיפה בנפרד, שכן היקף התקיפות הכולל והקלות של ייצור מטרות עשויים להשפיע על מאזן המידתיות במקרה היחיד.

61 Brian Smith, "Civilian Casualty Mitigation and the Rationalization of Killing", *Journal of Military Ethics,* 20, 1 (2021): 47–66.

62 The Decision Lab, "Why Do AI Systems Make Responsibility Feel like No One's Job? The Accountability Diffusion in AI, explained", accessed February 20, 2026, Link

63 Theodore M. Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton University Press, 1996)

64 Michael H. Bernstein et al., "Randomized Study of the Impact of AI on Perceived Legal Liability for Radiologists", *The New England Journal of Medicine AI* 2, 6 (2025)

65 Association of American Medical Colleges – AAMC, "Protect Against Algorithmic Bias", accessed February 21, 2026, Link

66 Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* (New York: W.W. Norton & Company, 2019)

67 ראו: ראובן גל ויעקב שנרב, "דילמות אתיות הקשורות בהפעלת מערכות נשק אוטונומיות חמושות בישראל: בחינה אל מול תר"ש 'תנופה'" בתוך **תנופה ותורפה: קריאות חברתיות בדוקטרינה הצבאית של ישראל**, עריכה: עפרה בן־ישי, יגיל לוי ורינת משה (פרדס, 2024), 127–156

על הכותבים

פרופ' יגיל לוי הוא פרופסור למדיניות ציבורית וסוציולוגיה פוליטית באוניברסיטה הפתוחה וראש מכון האוניברסיטה הפתוחה לחקר יחסי חברה-צבא
yagille@openu.ac.il

פרופ' אורי שורץ הוא סוציולוג המתמחה בחקר החברה הדיגיטלית, תיאוריה סוציולוגית וסוציולוגיה תרבותית באוניברסיטת בר אילן, ומחבר הספר Sociological Theory for Digital Society (Polity, 2021)
ori.schwarz@biu.ac.il