



כאשר הרובוט יחשוב במקומו

"אם נמשיך לפתח את הטכנולוגיה שלנו בלי תבונה ושיקול דעת, המשרת שלנו עלול להפוך למי שיוציא אותנו להורג" - גנרל עומר בראדלי

רקע: ההתפתחות הטכנולוגית של האמל"ח

במשך אלפי שנות לחימה מסתמך האדם על הטכנולוגיה כדי להכריע את יריביו - ממרכבות הסוסים וחרבות הברזל ועד הטנק, המטוס והצוללת המודרניים. באמצע המאה ה-20 נדמה היה שהרדיפה האין-סופית אחר הטכנולוגיה הגיעה לסיימה עם המצאת הנשק האולטימטיבי: פצצת האטום.¹ אולם עד מהרה התברר שמדובר בנשק שמשמש בעיקר לצורכי הרתעה ושלמעשה אי-אפשר להפעילו בלי להסתכן בהשמדת האנושות כולה, ולכן נמשך במלוא עוזו המרוץ הטכנולוגי בתחום הנשק הקונוונציונלי.

בחזית הטכנולוגיה ניצבות כיום מערכות הלחימה הבלתי מאוישות. המערכות האלה מסוגלות לפעול ביעילות במשך שעות ארוכות בשטחים רחבי ידיים ומסוכנים. דוגמה למערכת כזאת היא המטוסים הבלתי מאוישים (מל"טים). אלה מחליפים בני אדם בביצוע פעולות "מונטוניות, מלוכלכות ומסוכנות" בזירות שונות בעולם, בין היתר באפגניסטן.² המל"טים צברו שם מאות אלפי שעות טיסה במגוון רחב של משימות - החל מתצפיות וכלה בתקיפת מטרת ("סגירת מעגל") - בהנחיית מפעיל המצוי אלפי קילומטרים משדה הקרב.

אנחנו הולכים ומתקרבים ליום שבו נוכל לייצר מכונות לחימה אוטונומיות לחלוטין שיחליטו באופן עצמאי מתי לפתוח באש ונגד אילו מטרות. באפשרות הטכנולוגית הזאת טמונות סכנות עצומות. המאמר סוקר את הסכנות האלה ומציע דרכים לצמצמן

המל"ט האמריקני סקאי ווריור מתוצרת ג'נרל אטומיקס. 4 מל"טים מהדגם הזה, חמושים בטילי הלפיייר, יוצבו באפגניסטן עד תום 2010

סרן ניר סולומון

מפא"ת



דוגמה נוספת לכך - מתוצרת "כחול לבן" - היא מערכת סטטית הנקראת "רואה-יורה" מתוצרת רפא"ל. זו מצויה בשימוש מבצעי בצה"ל זה זמן מה וכבר הצליחה לסכל ניסיונות חדירה לשטח ישראל.³

מערכות לחימה מתקדמות עוד יותר הן אלה האוטונומיות, שאף אינן זקוקות למפעיל מרוחק שינחה אותן. חימוש מדויק מסוג "שגר ושכח" של היום הוא למעשה מערכת נשק אוטונומית: לאחר ששוגר הוא יודע להתביית בעצמו - ללא התערבות אדם - על המטרה עד לפגיעה בה. כיום כבר קיים חימוש שיודע להבדיל בעצמו בין סוגים שונים של מטרות, להתביית על אחת מהן ולבחור את הדרך היעילה ביותר להשמדתה.

כמעט לכל הטכנולוגיות החדשות יש מכנה משותף אחד: מחשוב מתקדם הנותן עצמאות הולכת וגדלה לאמל"ח עצמו בנוגע לאופן פעולתו. הכיוון המסתמן כמעט בכל פיתוח חדש הוא המעבר מאוטומטיות (חזרה על דפוס פעולה קבוע) לאוטונומיות (קבלת החלטות באופן עצמאי).

היכן, אם בכלל, יש לשים גבולות לאוטונומיה של אמצעי הלחימה?

האינטואיציה אומרת לנו שטמונה סכנה פוטנציאלית גדולה בהענקת יכולת בלתי מוגבלת ליישומים מכנית ליטול חיי אדם. ואכן הסוגיה הזאת הולידה תרחישי אימים רבים בספרות המדע הבדיוני ובקולנוע. החשש הוא שהגולם יקום על יוצרו וישמיד ללא הבחנה את כל הנקרה בדרכו.

אך האם הסכנה הזאת היא ממשית? הרי כבר כיום ישנם אמצעי לחימה אוטונומיים למחצה - למשל טילים מסוג "שגר ושכח" שהוזכרו לעיל וינותחו בהמשך - ששום צבא אינו חושש להשתמש בהם מחשש שיופנו נגדו. אולם ברור שקיימות רמות שונות של אוטונומיות, ובהחלט ייתכן שביום מן הימים יתברר לנו שעשינו בתחום הזה צעד אחד יותר מדי.

מטרת המאמר הזה היא לסקור את רמות האוטונומיות המוענקות לכלי הלחימה, לבחון את הסכנות הגלומות בכל רמה ולהמליץ האם יש להגביל את האוטונומיות הזאת ובאיזה אופן. בשל קוצר היריעה המאמר אינו דן בהיבטים האתיים, המשפטיים והעקרוניים של הנושא הזה ומתמקד בהיבטים המעשיים שלו.

המצב כיום: מאוטומטיות לאוטונומיות גוברת

משרד ההגנה של ארה"ב מייחס חשיבות רבה לכלים בלתי מאוישים - בין היתר למלחמה בטרור - ולכן גיבש תוכנית אב לקידום הנושא בכל זרועות הצבא.⁴ התוכנית הזאת מתווה מתודולוגיה לפיתוח הכלים האלה ולשילובם במאמץ המלחמתי.

כאמור, הפיתוח הטכנולוגי הבולט של השנים האחרונות הוא שכלול האוטומטיות. הודות להתקדמות שהושגה יכולים כיום כלים לא מאוישים לנוע לבדם בשיירה, לסרוק שטחים ולעקוף מכשולים. בכך הם משחררים את מפעיליהם מפעולות מונוטוניות או טריוויאליות ומאפשרים להם להתמקד בביצוע המשימה.⁵

ההתקדמות הטכנולוגית בתחום הפיתוח של כלים אוטומטיים היא מרשימה מאוד. כך, למשל, ב-2005 הצליח רכב לא מאויש לעבור באופן עצמאי מסלול של 210 ק"מ בשטח הררי בתוך פחות משבע שעות.⁶ דוגמה נוספת היא כטב"ם מדגם מנטיס (גמל שלמה), שמוטת הכנפיים שלו היא באורך 20 מטר, ואשר מסוגל לשהות פרקי זמן ממושכים בגובה בינוני. לכטב"ם הזה יש בקרת טיסה אוטומטית שכמעט אינה זקוקה למעורבות של מפעיל בשר ודם.⁷

המניע ליצירת כלים אוטומטיים יכול להיות הרצון להקל על המפעיל במקרה של תקלות, כגון אובדן תקשורת זמני. דרגה מתקדמת יותר של אוטומציה היא יצירת כלים שיכולים לגלות בעצמם תקלות במערכות שלהם ולתקן אותן. הטכנולוגיה הזאת כבר מיושמת כיום במל"ט גלובל הוק, אשר יכול לשהות באוויר במשך 36 שעות ברציפות - מה שמאפשר לו לבצע משימות הרחק מבסיסיו.⁸

כיום משקיע דרפ"א (DARPA) - גוף המו"פ הביטחוני של משרד ההגנה האמריקני - מאמץ רב בהרחבת האוטונומיות של מערכות רובוטיות, במיוחד בתחום של שיתוף פעולה אוטומטי בין פלטפורמות מאוישות ובלתי מאוישות.⁹ המטרה היא שהפלטפורמות הלא מאוישות יעבירו זו לזו מידע באופן אוטומטי וכמעט שלא יזדקקו להכוונה מהכלים המאוישים. גם בצה"ל ובמערכת הביטחון ערים לצורך להקנות יותר אוטונומיות לכלים בלתי מאוישים שפועלים בשירות צבא היבשה, ואף קיימת "מפת דרכים" לקידום הנושא הזה.¹⁰

בפלטפורמות הבלתי מאוישות אכן לא נמצאים מפעילים בשר ודם, אך בני אדם ממלאים תפקיד מרכזי בהפעלתן: הם נדרשים להגדיר את משימות המערכת, לפענח את קריאות החיישנים, להבין את תמונת הקרב ולהחליט על הפתיחה באש.¹¹ צבא ארה"ב עדיין מתייחס בחשדנות לכלים בלתי מאוישים כמושים, ותו"ל ההפעלה שלהם עדיין מחייב "אדם בחוג" (דהיינו מעורבות של מקבל החלטות בשר ודם), אך הגישה הזאת עשויה להשתנות כבר בעשור הקרוב.¹² ממחקר שהזמין צבא ארה"ב עולה שבעתיד יופעלו בשדה הקרב רובוטים אוטונומיים שיוכלו להחליט באופן עצמאי על פתיחה באש.¹³ המניע לכך הוא הרצון לחסוך בזמן. ההנחה היא שמחשבים מתקדמים יותר, שיפותחו עוד בדור הנוכחי, יוכלו - באמצעות אלגוריתם חכם - לקבל החלטה טובה יותר ממפעיל בשר ודם. והם יעשו זאת הרבה יותר מהר ממנו.¹⁴

סוגי האוטונומיות בתחום המבצעי

בספרות המחקר העולמית יש כמה דוגמאות לניתוח רמות של אוטונומיות. ההבחנה בין רמות שונות של אוטונומיות חיונית למי שרוצה לעשות הערכות מצב טכנולוגיות.¹⁵ הרמות האלה נקבעות לרוב על פי רמת המורכבות הטכנולוגית של הכלים האוטונומיים. לצורך הניתוח הנדרש במאמר הזה אסווג את האוטונומיות שלהם לארבע רמות. כל אחת מהן היא קפיצת



כלי רכב בלתי מאויש מדגם טאלון של חברת קינטיק פוסטר-מילר. הכלי מסוגל לבצע מספר רב של משימות, ביניהן סילוק פצצות, איסוף מודיעין ואף ירי בשליטת מפעיל מרחוק

ובוחרת באופן עצמאי את הדרך היעילה ביותר להפעלת ראש הנפץ. התהליך כולו, פרט להחלטה על השיגור, אינו כולל התערבות של בן אדם.

4. רובוט קטלני אוטונומי. המוטיווציה "לשלוף ראשון" מובילה לפתרון האולטימטיבי: מערכת שמסיירת עצמאית, מזהה ובוחרת מטרות וקוטלת אותן בלי לפגוע באוכלוסייה לא מעורבת. עדיין אין דוגמאות למערכות כאלה שהן מבצעיות. בשלבים מאוחרים יותר ניתן יהיה לשכלל את המערכות האלה כך שיקבלו החלטות טקטיות ויפעלו בצוותים.

ניתוח הדוגמאות שהובאו לעיל מלמד על ההבדלים בין ארבע הרמות השונות:

רמה 1. אוטונומיות חלקית - לרוב של קטע תפעולי צר שבו אין למפעיל האנושי שום ערך מוסף. הטיפול האוטומטי במקטע התפעולי מאפשר למפעיל האנושי להפנות את הקשב שלו להיבטים מורכבים יותר של המשימה. דוגמאות לרמה 1 של אוטונומיות (נוסף על החימוש המונחה עצמאית שהוצג קודם לכן): טייס אוטומטי, בדיקה אוטומטית של תקלות.

רמה 2. כמו רמה 1, אך בתוספת של יכולת להגיב על איום על פי החלטה של מפעיל בשר ודם. ברמה הזאת הכלי הוא מעין זרוע ארוכה של המפעיל: הוא מספק לו את המידע הדרוש לו ואת האמצעים להגיב. נוסף על כטב"ם יש מערכות נוספות ברמה 2 כגון מערכת "רואה-יורה". דרך אגב, התגובה לאיום אינה חייבת להיות קטלנית. לדוגמה: רובוט לסילוק פצצות.

רמה 3. כמו רמה 2, אך בתוספת של יכולת להגיב באופן עצמאי על איום ולהבחין בין מטרות. ההחלטה על עצם הטיפול במטרה נמצאת בידי בני אדם, אך אופן הטיפול נתון לשיקול המכונה. מערכות כאלה עדיין אינן נפוצות.

רמה 4. כמו רמה 3, אך בתוספת של יכולת להחליט באופן עצמאי אם לפעול נגד האיום. במילים אחרות: שלושת שלבי הפעולה של המערכת - קליטת המטרה, הפללתה ותגובה - נעשים ללא התערבות של מפעיל בשר ודם. כפי שצוין קודם לכן, אין עדיין דוגמאות למערכות מבצעיות ברמה 4, אך ניתן להעלות על הדעת מערכות כאלה: למשל, מוקש נייד וסלקטיבי.

בעיות שעלולות להתעורר כתוצאה משימוש באמצעי לחימה אוטונומיים

במדינות רבות מתקיים כיום דיון בהיבטים האתיים של השימוש במערכות נשק חמושות

מדרגה בהשוואה לרמת האוטונומיות שקודמת לה. הסיווג לרמות אוטונומיות שמוצג להלן הוא פונקציונלי-מבצעי ולא טכנולוגי בשל ההנחה הסבירה שלקראת סוף העשור לא תהיה למעצמות שום מגבלה טכנולוגית ביישום כל רמת אוטונומיות שירצו בהתאם לצורכיהן.

לצורך הסיווג אשתמש בארבע דוגמאות שונות בתכלית של כלים אוטונומיים:

1. טיל מונחה עצמאית. כיום מסוגלים טילי אוויר-אוויר מתקדמים לפגוע במטוס אויב הנמצא במרחק רב מהם. כל מה שעל הטייס לעשות הוא להצביע על המטרה (אשר נרכשה באמצעי שונה לפני כן) ולשדך בין הטיל לבניה. הטיל אוטונומי לחלוטין בבחירת מסלול התקיפה להשגת ההרג היעיל ביותר. בלשונו של תא"ל (מיל') ד"ר דני גולד, לשעבר ראש מו"פ במפא"ת: "הטייס לוחץ על כפתור, והטיל כבר מסתדר בעצמו".¹⁶

2. כטב"ם תוקף. ארה"ב מפעילה כלי טיס בלתי מאוישים רבים במלחמה באפגניסטן. כמו כן היא הפעילה אותם במלחמה בעיראק ובמבצעים נגד טרור בתימן ובאפריקה. את הכטב"ם מפעיל בן אדם. הוא מקבל את תמונת המצב המודיעינית באמצעות מטע"דים (מטענים ייעודיים) שונים המותקנים על הפלטפורמה. באמצעותם הוא יכול להחליט אם לסגור מעגל ולתקוף מטרה ברגע שהופללה. הטיילים שמשוגרים מהפלטפורמה יכולים להיות אוטונומיים (כפי שהוסבר בסעיף 1). הכטב"ם אף יכול לתת למפעיל הערכת נזק לאחר התקיפה ובמידת הצורך לשגר חימוש נוסף.

3. חימוש בעל יכולת הבחנה. דוגמה לחימוש כזה הוא BLU-108 של חברת טקסטרוניק¹⁷ שיכול להבדיל בין מטרות. את החלטה על שיגור החימוש מקבל אדם. החימוש מנווט את עצמו באופן עצמאי אל המטרה (בדומה למתואר בסעיף 1). בעת המעוף לעבר המטרה בוחנת אותה הפצצה באופן אקטיבי

לחלוטין בקבלת ההחלטות. הוא פועל בהתאם לתוכנות שהשתילו בו מפעיליו. משמעות הדבר היא שברמה הזאת אין חשש מפני אסון שהוא חמור עשרות מונים יותר מהבחנה שגויה: פעולה של כלי אוטונומי באופן שנוגד את מטרתנו במערכה.

סוגיית היכולת לקבל החלטות (רמה 4 של אוטונומיות)

המעבר לרמה 4 של אוטונומיות - שבה מוענקת לפלטפורמה היכולת לקבל החלטות באופן עצמאי לחלוטין - הוא המסוכן באמת, שכן הוא עלול להוביל לפגיעה נרחבת בבני אדם בלי להתחשב כלל בשאלה אם הם מטרות לגיטימיות או לא לגיטימיות.

שנים ארוכות של לחימה בערים צפופות הוכיחו שקיימים מקרים שבהם התנהגות "לא מוסרית" היא לגיטימית לחלוטין (למשל, השבת אש לעבר בניין שבמפות מוגדר עדיין בית חולים, אך בפועל הוא נטוש ומשמש עמדה של האויב). אלגוריתם לעולם לא יוכל להכיל את מגוון השיקולים שמקבל מפקד בשדה הקרב. הסיבה לכך היא שהבסיס של כל שפת מחשב הוא סדרה של "מקרים ותגובות", בין אם אלה תוכנתו מראש על ידי המתכנן ובין אם המכונה למדה אותם באופן עצמאי. לא ניתן לתכנת מראש מערך של כללים שייתן תשובה לכל מצב בגלל התדירות הגבוהה שבה משתנים המצבים ובגלל האינטראקציה המורכבת ביניהם. כיוון שלא ניתן לנסח את הכללים האלה בשלמות, אין משמעות לכוח החישוב הגבוה של המכונה שיכול לעלות בהרבה על כוח החישוב של מוח אדם. בסופו של דבר, היכולת לקבוע באופן אוטומטי ומכני אם תקיפת מטרה משרתת את מטרת המשימה או המערכה תהיה תמיד מוגבלת.

יתר על כן, כאשר כלי לחימה הוא בעצם צייד עצמאי, עלול המפעיל להתלבט איזה אלגוריתם להשתיל בו. רצונותיו ינועו בין הרצון שלא לפגוע באוכלוסייה לא מעורבת לבין הרצון לפגוע בכל מטרה ערכית לפי מערך חוקים אקראי. במילים אחרות: קיים



כלי רכב בלתי מאויש מדגם מארס של חברת קינטיק־פוסטר־מילר. זהו דגם משופר וחמוש של רובוט מסדרת טאלון. הירי והניהוג נעשים בשליטת מפעיל מרחוק

אוטונומיות. החשש הוא - כפי שצוין בתחילת המאמר - שרובוטים יקבלו החלטות של חיים ומוות וכתוצאה מכך ישתנה אופיה של החברה האנושית. אחד הגופים שעוסק בסוגיה הזאת הוא התאגיד לטכנולוגיות חדשות, למבצעים צבאיים ולביטחון לאומי. התאגיד הזה חוקר בראש ובראשונה את ההשפעה שתהיה לרובוטים חמושים ואוטונומיים לחלוטין על האתיקה ועל המשפט.¹⁸

בבית הספר למשפטים באוניברסיטת סטנפורד דנה אשתקד קבוצת חוקרים בסוגיית האחריות המשפטית למעשיהם של רובוטים אוטונומיים.¹⁹ בין היתר עלתה השאלה מי אחראי לניק שגרם בשוגג רובוט אוטונומי: המפעיל שלו או היצרן שלו? הסוגיה נהיית סבוכה עוד יותר אם גורם שלישי מצליח להשתיל וירוס לתוכנה שמפעילה את הרובוט. במקרה כזה - אמרו המשפטנים - כלל לא ברור מי אחראי לניזקים החומריים ולפגיעות בנפש שייגרמו בשל פעולת הרובוט.

הסוגיה של יכולת ההבחנה (רמה 3 של אוטונומיות)

במאמרו Grounds for Discrimination מתייחס נואל שארק, פרופסור למדעי המחשב מאוניברסיטת שפילד באנגליה,²⁰ להיבטים האתיים והמשפטיים של הענקת יכולת הבחנה לכלים בלתי מאוישים בין מטרות שבעיני המפעיל הן לגיטימיות למטרות שבעיני הן לא לגיטימיות. טענתו היא שאסור להשאיר בידי מכונות החלטות של חיים ומוות, שכן הדבר פותח פתח לפגיעה באוכלוסייה לא מעורבת.

בעניין הזה יש לומר שני דברים:

1. ההבחנה בין מטרות שניתן לתקוף אותן לבין מטרות שאסור לתקוף אותן היא אכן שלב קריטי בביצוע משימה, אך זהו שלב אחד בלבד מתוך שרשרת ארוכה של שלבים שמתחילה בהגדרת ההישג המבצעי הנדרש, ממשיכה בבחירת הטקטיקה לאור האסטרטגיה ובתכנון משימה ומסתיימת בביצוע התוכנית. ההבחנה בלגיטימיות המטרה היא אפוא שלב מאוחר מאוד בשרשרת שבא ממש לפני התקיפה עצמה. באמצעות תכנון נכון של השלבים המוקדמים ניתן למנוע מלכתחילה מצב שבו מכונה תצטרך לבחור בין מטרות לגיטימיות למטרות לא לגיטימיות. במילים אחרות: יש למנוע מצב שבו מכונה "מתלבטת" בסוגיית הלגיטימיות של מטרה שהיא "שוקלת" לתקוף.

2. שארק צודק בהחלט באומרו כי יש לטפל בבעיית ההבחנה בין מטרות לגיטימיות למטרות לא לגיטימיות, שכן המחיר (המוסרי, המשפטי, המדיני והכספי) של כל שגיאה עלול להיות גבוה מאוד. אך יש לזכור שבסופו של דבר ברמה הזאת הכלי אינו עצמאי

כאן פוטנציאל לניגוד בין הפעלה מוסרית של מכונות לחימה לבין הפעלה יעילה שלהן. אם האלגוריתם שמושתל במכונה הוא מהסוג הראשון, הסכנה העיקרית היא של הבחנה שגויה בין מטרות לגיטימיות למטרות לא לגיטימיות, והסוגיה הזאת כבר נידונה לעיל. במקרה השני, כאשר סוגיית המוסר אינה גורם משמעותי בקבלת ההחלטה בנוגע לתקיפה, וכל פגיעה באזרחים משרתת את מטרות הגורם המפעיל, עלולה מערכת עצמאית להיות כלי נשק יעיל יותר בידי טרוריסטים, שאינם כפופים למערכת של חוקי מוסר (אלא למערכת שונה בתכלית), מאשר בידי צבא מאורגן. ברגע שארגון טרור ישים את ידו על מערכת עצמאית, הוא יוכל לשחררה בשטח אזרחי צפוף - לאחר תכנות מחדש או בלעדיו. האפשרות הזאת תהיה רלוונטית כאשר המערכות האלה יהיו נגישות וזולות וידרשו פחות מאמץ מאשר שיטות הפעולה הנוכחיות של ארגוני הטרור, כגון גיוס מתאבדים.

לסיכום, הבעיה האמיתית טמונה בהסתמכות על אלגוריתם לצורך הקביעה אם לפתוח באש, כיוון שהיא עלולה לגרום לפגיעה אקראית חוזרת ונשנית באוכלוסייה לא מעורבת או לפגיעה כוללת במקרה של אלגוריתם הבחנה זדוני.

פתרונות לבעיית האוטונומיות של מכונות לחימה

כיוון שהבעייתיות בהענקת אוטונומיות למכונות לחימה ברורה לכולם, נעשו כמה ניסיונות לפתור את הבעיה - לפחות באופן חלקי. למשל, אחד הניסיונות הבולטים למניעת תקלות במעבר לשלב השלישי, שלב ההבחנה, הוא פיתוח של "בקר אתי" המותקן במערכת הפעלה של כלי טיס בלתי מאוישים.²¹ בשלב הראשון יהיה הבקר מעין נצרה המונעת במקרה של ספק ירי של מפעיל אנושי. את ההחלטה הוא יקבל באמצעות השוואה בין דוגמאות שהוזנו בו מראש של פגיעה באוכלוסייה לא מעורבת בסיטואציות דומות לבין סימולציה של ירי במצב הנתון. לשלב הרביעי, סוגיית ההחלטה, ניתן

כבר פתרון מקורי לפני קרוב ל-70 שנה. אייזיק אסימוב, סופר המדע הבדיוני האמריקני המפורסם, הציע כבר ב-1942 להשית על רובוט עצמאי כללי אתיקה בסיסיים שבשום אופן הוא לא יוכל לפעול בניגוד להם.²² להלן "שלושת כללי הרובוטיקה" שהציע אסימוב:

1. רובוט לעולם לא יפגע באדם ולא יאפשר מצב שבשל אי-פעולתו ייפגע אדם.
2. רובוט יציית תמיד לבן אדם, אלא אם ההוראה שהוא קיבל סותרת את כלל 1.
3. רובוט יפעל תמיד כדי לשמר את קיומו, אלא אם כן השימור העצמי סותר את כלל 2.

מובן שאלה כללים פילוסופיים-דמיוניים המבוססים על הנחת היסוד שקיימת טכנולוגיה שניתן ליישם עליה את הכללים הנ"ל. הטכנולוגיה הזאת בוודאי שלא הייתה קיימת בתקופתו של אסימוב. יתר על כן, על מכונת מלחמה לא ניתן להשית כלל של אי-פגיעה גורפת בכל אדם, שכן אז היא תחדל להיות מכונת מלחמה.

פתרון אפשרי לבעיית ההחלטה

את רעיונו של אסימוב אני מציע ליישם בדרך אחרת. בסיס רעיונו של אסימוב הוא קיום של שני רובדי תוכנה: הראשון בסיסי, שאותו לא ניתן לשנות, והשני הוא רובד-על של כללים שעל פיהם מקבלת המכונה את החלטותיה. היחס ביניהם הוא פשוט: כלל שכתוב ברמת העל לעולם לא יוכל לסתור כלל בסיסי, כפי שחוק רגיל אינו יכול לסתור חוק יסוד או חוקה. את כללי הבסיס יש לצרוב לתוך החומרה באופן שאינו ניתן לעקיפה, והם נדרשים לכלול שני כללים פשוטים. הראשון פותר בצורה פשוטה ביותר את בעיית ההחלטה, והשני קובע שזהו כלל בסיסי:

1. תמיד נדרש אישור של מפעיל אנושי לפני פתיחה באש.
2. שום כלל לא יסתור את כלל 1.

בניגוד לכללי הבסיס, את תוכנת העל הראשית אפשר (ואף רצוי) לשנות לפי רצונו של המתכנן ולעדכן אותה באמצעות מערכת קבלת החלטות חכמה יותר ויותר, כולל "בקרים אתיים" או כל מנגנון הגנה נוסף לפי רצון המתכנן. ניתן אף להכניס בקרים נוספים לכללי הבסיס כל עוד הם לא סותרים את הכלל הראשון. זהו שלב ראשוני העוזר לצמצם את הנזקים הפוטנציאליים של בעיית ההבחנה ומונע אפשרות שהמצב ייצא מכלל שליטה. בשום מקרה לא תהיה הסתמכות על הבחנה של מכונה לשם פתיחה באש, בין אם היא שגויה ובין אם לאו. לשיטה הזאת יש יתרונות רבים. ראשית, המתכנן לא יצטרך לחשוש מכך שבאג בתוכנה יגרום לירי לא רצוני, והוא יהיה פטור מאחריות משפטית במקרה של ירי כזה. נוסף על כך, למפקדי המערכת תהיה שליטה מלאה בקרב, שכן תבטול האפשרות שהמכונות יפעלו בדרך לא רצויה. כל קרב הוא סיטואציה כאוטית גם ללא הכנסת גורם נוסף של אי-ודאות כגון מכונה שיצאה מכלל שליטה. יתרון נוסף הוא שיצומצם מאוד הנזק שעלול להיגרם כתוצאה מנפילת הכלי לידיים זדוניות, שכן תוגבל יכולתו לפתוח באש על מטרות אקראיות.

הטכנולוגיה למימוש הרעיון הזה כבר קיימת: פותחו מנגנונים להבטחת אי-שינוי של תצורת תוכנה בסיסית (anti-tampering)²³ - בוודאי ברמה כזאת שתמנע מארגון בעל משאבים מוגבלים לשנות תוכנה.

אמצעי נוסף הוא רגולציה ממשלתית. הדרך היחידה לוודא התקנת מנגנונים לאי-שינוי של תצורת תוכנה בסיסית היא באמצעות תקן ממשלתי מחייב, בדומה לתקנות ביטחוניות. עלויות הפיתוח והייצור של רכיב כזה הן זניחות לעומת פוטנציאל הנזק העצום שעלול להיגרם ללא שימוש בו.

סיכום: כיוון ההתפתחות הרצוי לכלים בלתי מאוישים

במאמר הזה הוצגה המגמה הקיימת בכל הצבאות המתקדמים להעניק אוטונומיות לכלים חמושים ובלתי מאוישים. נוהל שארקי סיכם היטב את הכיוון כיום: "אנו צועדים כסהרורים לעולם חדש ואמיץ שבו רובוטים מחליטים את מי, איפה ומתי לקטול".²⁴

כדי להגדיל את הספקי התקיפה הוא מסוכן מאין כמוהו. במקרה כזה יגבירו מקבלי ההחלטות עוד יותר את הכאוס שבו הם פועלים בלאו הכי.



המל"ט האמריקני פרידטור מתוצרת ג'נרל אטומיקס. מל"טים מחליפים בני אדם בכיוצא פעולות מונוטוניות, מלוכלכות ומסוכנות בזירות שונות בעולם, בין היתר באפגניסטן

הערות

1. Martin van Creveld, **Technology and War**, The Free Press, 1989
2. Unmanned Systems Roadmap, US DoD, 10 Dec. 2007
3. "רואה-יורה' פגעה במחבל וסיכלה פיגוע", <http://www.ynet.co.il/articles/0,7340,L-3856921,00.html>
4. Unmanned Systems Roadmap, US DoD, 10 Dec. 2007
5. US Army to Deploy Semi-Autonomous Robots, **Defense News**, 26 Oct. 2009
6. **שם**, עמ' 36
7. Mantis Takes Flight – The UK's Largest Ever Fully-Autonomous UAV, **Gizmag.com**, 17 Nov. 2009
8. Unmanned Systems Roadmap, US DoD, 10 Dec. 2007 p. 116
9. Unmanned Systems Roadmap, US DoD, 10 Dec. 2007 p. 34, 49
10. תא"ל (מיל') יעקב גול (סגן ראש מפא"ת לענייני מדע), "מפת דרכים ברובוטיקה ביטחונית", סדנת תל-אביב למדע, לטכנולוגיה ולביטחון, כנס 52
11. Unmanned Systems Roadmap, US DoD, 10 Dec. 2007 p. 52
12. Unmanned Systems Roadmap, US DoD, 10 Dec. 2007 p. 54
13. "Pentagon Exploring Robot Killers that Can Fire on Their Own", **The News & Observer**, 20 Apr. 2009
14. Ray Kurzweil, **The Singularity is Near**, Viking Press, New York, 2005
15. למשל, התייחסותו של פרופ' דניאל וייס בהרצאתו "כלי טיס בלתי מאוישים - האם הטייס נחוץ במטוס?", סדנת תל-אביב למדע, לטכנולוגיה ולביטחון, כנס 52
16. תא"ל (מיל') ד"ר דני גולד, מפקד יחידת המו"פ במפא"ת: "אנחנו נשארים מחוברים לפן המבצעי כל הזמן" - <http://www.pc.co.il/?p=5018>
17. Noel Sharkey, "Grounds for Discrimination: Autonomous Robot Weapons", **RUSI Defense Systems**, Oct. 2008, pp. 86-89
18. "Organization Analyzes How New War Weapons Will Impact Society", **Physorg.com**, 15 Oct. 2009
19. "As Robots Become More Common, Stanford Experts Consider the Legal Challenges", **Physorg.com**, 23 Nov. 2009
20. נואל שארק, **שם**
21. "Ethical System Could Limit Damage from UAV Attacks", **International Defense Review**, Mar. 2010, p. 5
22. Isaac Asimov, "Runaround", in: **I, Robot**, Doubleday & Company, New York, 1950
23. US DoD Anti Tamper, <http://at.dod.mil/>
24. Noel Sharkey, "Pentagon Exploring Robot Killers That Can Fire on Their Own", **The News & Observer**, 20 April 2009
25. חמי שלו ואהרון לפידות, "על המסלול - ראיון עם יאיר שמיר", **ישראל היום** 16.4.2010

במאמר נעשתה הבחנה בין ארבע רמות שונות של אוטונומיות: אוטומציה של רוברד נמוך, אפשרות לסגירת מעגל תקיפה, אפשרות לאוטומציה של שיטת ההרג ולהבחנה בין מטרות ואוטונומיות מלאה בהחלטה על פתיחה באש.

אף שמגמות המחקר בעולם הן כיצד ניתן לצמצם טעויות בהבחנה כושלת בין מטרות, הבעיה האמיתית טמונה ביכולת קבלת ההחלטה על הירי ולא בהבחנה. ירי אוטומטי שאינו משרת את מטרות הלחימה עלול לטמון בחובו סכנות רבות.

ניתן למנוע זאת באמצעות חלוקת הבינה המלאכותית של הכלי לשתי רמות: רמה בסיסית שמונעת ירי ללא אדם בחוג ורמה גבוהה הכוללת מערך של כללי התנהגות של המערכת אשר לעולם לא תוכל לסתור את כללי הרמה הבסיסית.

מהו אם כן תפקידם הנכון בעתיד הנראה לעין של כלים בלתי מאוישים? ישנם כמה כיוונים טבעיים להתפתחות. למשל: הרחבה של משימות הבט"ש (אך השארת סגירת המעגל בידי מפעיל אנושי) ומתן משימות של תובלה ושינוע - גם של בני אדם - לכמה כלים בלתי מאוישים אוטומטיים שאותם יפעיל אדם אחד בלבד. את תשומת ליבו יוכל אותו אדם להקדיש למשימה הלוגיסטית, והוא לא יצטרך לעסוק בהובלה עצמה.

אשר לתקיפה - היטיב לתאר זאת יאיר שמיר, יו"ר דירקטוריון התעשייה האווירית,²⁵ באומרו שהעתיד טומן בחובו שיתוף פעולה בין פלטפורמות מאוישות ולא מאוישות. כך, למשל, ניתן יהיה לשלוח למשימה מבנה של כלים בלתי מאוישים שעליו יפקד כלי מאויש אחד.

כלים בלתי מאוישים ימשיכו להיות העיניים והזרוע הארוכה בעבור המפקדים: הם יסייעו למפקדים להחליט אם לתקוף או אם להימנע מהתקפה ואף יבצעו את התקיפות בעצמם ויחסכו בכך חיי אדם. הכלים יכולים להיות בעלי רמות אוטומציה גבוהות מאוד, אך לעולם אין ליטול את ההחלטה על פתיחה באש ממפעיל אנושי. יש להקנות לכלים האלה את האפשרות להחליט באופן עצמאי על פתיחה באש רק במקרים שבהם אין שום ספק שמדובר בפעולת הגנה שנועדה להציל חיי אדם. למשל, כאשר מדובר במערכת הגנה אקטיבית המגינה מפני אש באמצעות תגובת אש מהירה הנמדדת בחלקיקי שניות. שום בן אנוש אינו מסוגל להגיב במהירות כה רבה. הרצון לחרוג מהתחום הצר הזה לעבר פתיחה באש אוטומטית וסלקטיבית