



אתיקה ובינה מלאכותית – משמעויות למודיעין

מארק גרייבס¹

המאמר "Shared Moral and Spiritual Development Among Human Persons and Artificially Intelligent Agents" פורסם ב-2017 בכתב העת Theology and Science ונכתב על ידי Mark Graves. המאמר עוסק ברעיון המאתגר מחשבתית של הבניית מסגרת אתית-מוסרית משותפת לבני האדם ול"סוכנים" המבוססים על טכנולוגיות הבינה המלאכותית.

בשנים האחרונות מגמת הבינה המלאכותית (להלן: AI – Artificial Intelligence) תופסת תאוצה ובין היתר, גם קהילת המודיעין מושפעת מהשינויים הטכנולוגיים בתחום ה-AI, שצפויים להרחיב את מעגל השפעתם עם הזמן. לטענת כותב המאמר, ההתקדמות הטכנולוגית מאפשרת במידה מסוימת את האופציה ליצירת "סוכנים" המבוססים על AI בתוך עשורים בודדים. לפיכך, יש צורך בפיתוח מוסריות רוחנית משותפת של בני האדם וסוכני הבינה המלאכותית, שככל הנראה יעסקו זה לצד זה בתחומים שונים, ובהם גם בתחומי המודיעין.

שאלת השאלה: **אם האינטליגנציה המלאכותית תעלה על האינטליגנציה האנושית, האם יוגדרו כללי אתיקה ותיאולוגיה חדשים?**

במסגרת מחקר שנערך נשאלו 170 חוקרי AI, במהלך השנים 2012-2013, באיזו שנה יופיעו מכונות בעלות אינטליגנציה השווה לזו של בני האדם (Human-Level Machine Intelligence – HLMI). לאחר שכלול התשובות, השנה הממוצעת הייתה 2040. לטענת גרייבס, השלב שיגיע לאחר מכן יהיה "התפוצצות אינטליגנציה", כלומר כאשר הבינה המלאכותית תעלה על האינטליגנציה האנושית. שלב זה עשוי להתאפשר בזכות התפתחות טכנולוגיית למידת מכונות, שבמסגרתה כבר כיום מכונות AI משפרות ומלמדות את עצמן בתחומים מוגבלים.

לטענת חלק מהחוקרים, "התפוצצות אינטליגנציה" תוביל לטכנולוגיה ייחודית, שתהיה בלתי מובנת עבור בני האדם. הם טענו כי הופעתן של מכונות בעלות אינטליגנציה, העולה על זו האנושית, יופיעו כבר ב-2045. זאת, לעומת התחזיות במחקר שהוצג מעלה, שצפו כי מכונות מסוג זה יופיעו רק 30 שנים לאחר הופעת ה-HLMI, קרי ב-2070.

אומנם מועד ההתרחשות של שינויים אלו איננו וודאי, אך כל עוד קיים סיכוי שהטכנולוגיה מתקדמת לסוכנים המבוססים על AI בעלי HLMI, ואף בהמשך בעלי אינטליגנציית על, יש לחקור את האתיות והתיאולוגיות העולות מנושא זה.

כיום, טכנולוגיות הבינה המלאכותית חסרות את המרכיב האנושי, המסוגל ליצור אינטראקציה חברתית עם הסביבה החיצונית, אותו בני האדם לומדים כבר בילדותם. חוקרי הבינה המלאכותית

¹ מארק גרייבס הוא חוקר ב-Travis Research Institute במחלקה לפסיכולוגיה ועמית מחקר בחוג למדעי הרוח והחברה במכון הטכנולוגי של קליפורניה.

טרם גילו כיצד להציב את התרבות במרכז טכנולוגיית למידת המכונות, שעשויה לאפשר ליצור סוכני HLMI המסוגלים ללמוד ולקחת חלק בתרבות האנושית. יחד עם זאת, "כישלונות" ה-AI הובילו לפיתוחים טכנולוגיים משמעותיים, ובהם התעלות ה-AI על האינטליגנציה האנושית במספר תחומים ספציפיים ובהם אבחון מחלות מתמונות רפואיות, שחמט ועוד.

כפי שפיתוחים בטכנולוגיית AI מניעים פיתוחים נוספים לכיוון HLMI, כך ההתקדמות בחקר מדעי המוח, מהפכת המידע שאפשרה בסיס נתונים עצום, מהירות חומרת המחשב ופיתוח מודלים ממוחשבים המציגים תהליכים המתרחשים במוח, יניעו את פיתוחם של סוכנים בעלי אינטליגנציית על. פיתוחים אלו חשובים במיוחד לצורך התמודדות עם האתגרים הניצבים כיום בפני כלל המגזרים, וקהילת המודיעין בפרט. כך, אתגרי מהפכת המידע, ובהם הצורך בהתמודדות עם גידול אקספוננציאלי של המידע הובילה לכך שהאינדקס של Google כולל כיום מעל טריליון דפי אינטרנט. אף אדם או קבוצת מומחים לא מסוגלים לעבד את כמות המידע הזו, ולכן רק מכונות המתקרבות ל-HLMI יוכלו להתמודד עם אתגר זה. האתגר יחייב את קהילת המודיעין, שמתקשה להתמודד עם הצפת הנתונים, לשלב טכנולוגיות HLMI במסגרת דיסציפלינות האיסוף והמחקר. לאור הגידול האקספוננציאלי, ככל שהזמן עובר כך תהיה חשיבות גוברת לשילוב טכנולוגיה זו, שתדרוש התמודדות עם שאלות אתיות ומוסריות.

נושא האתיקה בבנייה מלאכותית כבר החל להיבחן בפרספקטיבות שונות. הצלחת פיתוח טכנולוגיות HLMI ככל הנראה תוביל למכונות בעלות אוטונומיה משמעותית ויכולת לעסוק, לפחות בהיבטים מסוימים, בעולם האנושי. לפיכך, כיום מתפתח תחום חדש המעניק למכונות עקרונות אתיים או שיטות חשיבה, כדי שיוכלו לפתור בעצמן דילמות אתיות ולקבל החלטות באופן מוסרי. הצורך הבהול באתיקה בקרב מכונות אוטונומיות בא לידי ביטוי בעיקר במסגרת פעילויות צבאיות ומודיעיניות. בשנה שעברה בכנס של האו"ם בנושא נשקים קונבנציונאליים נערך דיון נרחב בנושא אתיקה בנשקים אוטונומיים. מסקנות הדיון מבטאות את החשש מפני השימוש בכלי נשק אוטונומיים וקובעות כי השליטה על מערכות הנשק חייבת להישאר בידי בני האדם, זאת בכדי להבטיח את הציות לדין הבינלאומי ועמידה בסטנדרטיים אתיים. לטענתם, קבלת החלטות על ידי AI עשויה לתרום ולאף להשיג תוצאות טובות יותר בתחומים מסוימים, אך אין לאפשר למכונות לקבל החלטות בנוגע להרג של אדם אחר. אוטונומיה במערכות נשק קטלניות עשויה להעלות בעיה אתית חמורה בנוגע ליעד התקיפה. כלומר, מערכות הנשק יוכלו לפעול באופן עצמאי לפגיעה באזרחים במקום בלוחמי האויב. יצוין, כי סוגיה זו מעלה גם שאלות בנוגע ליכולת החיזוי של בני האדם, משום שיהיה קשה לקבוע מתי מערכת הנשק האוטונומית תבצע מתקפה. לבסוף, נקבע בדיון כי על המדינות לפעול בדחיפות לקביעת מגבלות על מערכות נשק אוטונומיות.²

סקרי דעת קהל שנערכו הצביעו על התנגדות כללית למערכות נשק אוטונומיות. גם משרד ההגנה של ארה"ב הביע התנגדות, ונטען כי ארה"ב איננה מעוניינת בנשק ש"מקבל החלטות בעצמו". לדברי סגן שר ההגנה האמריקני לשעבר, צבא ארה"ב מעוניין להשתמש בבנייה מלאכותית באופן מצומצם בלבד, קרי במערכות נשק בעלות סט פעולות מוגדר ומצומצם. כיום בפנטגון קיים צוות The

International Committee of the Red Cross. (April 13, 2018). Ethics and Autonomous Weapon Systems: An Ethical Basis for Human Control. Geneva: UN. Retrieved January 8, 2019, from: <https://goo.gl/fh7ENF>.

שאלת השאלה: **כיצד נצליח להימנע מפעולות אלימות והרסניות מצד "סוכני" הבינה המלאכותית?**

בכדי להימנע מיצירת סוכני AI, שמסגולים להרחיק עצמם מבחינה קוגניטיבית ומוסרית מבני האדם, ייתכן כי קיים צורך לספק עבורם פרספקטיבה ומוטיבציה ליצור קרבה, שבאה לידי ביטוי אצל בני האדם באמפתיה. יצירת ערכים משותפים והרמוניה עם סוכני ה-AI עשויים לתרום לבני האדם, משום שביחד יוכלו בני האדם והסוכנים לחקור צורות נוספות של אינטליגנציה לקראת יצירת קהילה משותפת, שתושפת על ערכים אתיים זהים. המאמר מציע שלושה ערכים מרכזיים המעודדים פיתוח יחסים אתיים בין בני האדם וסוכני ה-AI:

1. **גם לבני האדם וגם לסוכני ה-AI יש את הזכות להתקיים ולהחזיק ערכים** - גם אם באופן מפורש טכנולוגיות ה-AI לא מסוגלות כיום להחזיק ערכים, הן כוללות לעיתים קרובות ערכים מרומזים. לדוגמא, נהיגה ברכב בבטחה, איסוף דגימות מדעיות מעניינות בזמן שיטוט בכוכב אחר, חיפוש אחר ניצולים באתר אסון ושרידות.
2. **גם בני האדם וגם סוכני ה-AI לא צריכים להתנגד לזכותו של האחר להחזיק בערכיו, גם אם הערכים מתנגשים זה עם זה. זאת, כל עוד הערכים לא מתנגשים עם הזכות להחזיק ערכים.**
3. **בני האדם וסוכני ה-AI צריכים לתקשר ולנסות להבין את הערכים של הצד השני, ללא קשר אם הצדדים מזהים עם הערכים של האחר** – הצורך לתקשר ולהבין את הערכים של האחר יתרום לתפקוד המוסרי של סוכני ה-AI.

לטענת כותב המאמר, הגורמים שצריכים להבנות את הערכים הללו אל מול סוכני הבינה המלאכותית הם מומחים בלבד ולא כל הקהילה. הסיבה לכך היא שלציבור אין יכולת לעצב מסגרת אתית שלמה, עמידה ואמינה לקהילות אנושיות בכלל, ולקהילות המשלבות סוכני AI בפרט. אם תרצה בכך, תוכל קהילת המודיעין לעצב ולהבנות מסגרת אתית שלמה מסוג זה על ידי מומחי ה-AI שבידה, תוך שילוב עם מומחים מהאקדמיה ומהמגזר העסקי. היתרון העיקרי בעיצוב המסגרת על ידי קהילת המודיעין עצמה טמון בעובדה כי לקהילת המודיעין ישנם צרכים גלובליים לצד צרכים ייחודיים. כלומר, מצד אחד קהילת המודיעין צריכה להתמודד עם אתגרים גלובליים כמו אתגרי עידן המידע איתם מתמודד גם המגזר העסקי, ומצד שני יש לה צרכים אישיים וייחודיים, שנוגעים לסוגיות מודיעיניות ספציפיות. באופן הזה, תוכל קהילת המודיעין להתוות מלכתחילה את הנתיב הרצוי עבורה כדי להתמודד עם שני סוגי האתגרים בכל הנוגע לאתיקה ובינה מלאכותית, ולא להיות מובלת על ידי המגזר העסקי.

Boyd, A. (October 31, 2018). Pentagon Doesn't Want Real Artificial Intelligence in War, Former Official Says. Defense One. Retrieved January 8, 2019, from: <https://goo.gl/qr7NVN>.

שאלת השאלה: כיצד יוכלו בני האדם והמכונות לתקשר כך שיבנו ביניהם מערכת יחסים הדדית המבוססת על ערכים מוסריים ואתיים משותפים?

לטענת כותב המאמר יש להתייחס לדרך התקשורת בין בני האדם והמכונות לא רק באופן הבסיסי ביותר, אלא לשיח שיוצר משמעות עבור האדם. יצירת משמעות כוללת את מכלול השאיפות התרבותיות והאנושיות בהקשרן ההיסטורי, כלומר יצירות אומנות, ספרות וטכנולוגיה; שיפוט מוסרי ואתי; והשאיפות וההתחייבות שמיוחסות לרוב לרוחניות ולדת. גרייבס סבור כי ליצירת המשמעות יסודות פסיכולוגיים שיש לקחת אותם בחשבון כאשר מפתחים טכנולוגיות AI על מנת להרחיב התנהגויות אנושיות, תשוקות וחיפו משמעות בקרב המכונות.

פיתוח משמעויות משותפות בין בני האדם למכונות הוכח כבעייתי, אך הוא הכרחי ומהווה את המפתח להתקדמות בפיתוח טכנולוגיית AI. גרייבס טוען כי בכדי לפתח משמעויות משותפות יש לבצע שינויים לא רק בתכנות המכונות, אלא גם בהרגלים של בני האדם. כלומר, עבור קהילת המודיעין בכדי ליצור שותפות אינטרסים ושיח משותף עם מערכות ה-AI, שיוטמעו, יש צורך בשינויים חשיבתיים ותפיסתיים של הקהילה ולא רק בתכנות מחדש של המערכות. באופן זה תוכל קהילת המודיעין להפיק את המרב ממערכות ה-AI ולבסס ערכים מוסריים ואתיים משותפים, שיתאמו את מטרות הקהילה. בנוסף, גרייבס מציע כי בכדי ליצור "קהילה" אחת של בני אדם ומכונות על בני האדם לתקשר עם סוכני הבינה המלאכותית כפי שהם מתקשרים עם בני האדם. בזכות ההתפתחויות בתחום טכנולוגיית למידת המכונות עצם ההיחשפות של סוכני הבינה המלאכותית לפעילותה של קבוצה חברתית, בעלת מטרה משותפת ברורה, עשויה לפתח את המשמעות המשותפת והבסיס האתי המשותף.

סיכום

השימוש בבינה מלאכותית בקהילת המודיעין היום הוא נרחב. כך, טכנולוגיית ה-AI מסייעת בהתמודדות יעילה עם כמות המידע העצומה ומאפשרת חיסכון בזמן יקר ובהקצאת המשאבים של החוקרים והאוספים לדברים נוספים, כמו הסקת מסקנות וניתוח המידע המודיעיני. בנוסף, טכנולוגיית ה-AI מאפשרת פענוח מדויק ומהיר יותר של תצלומים ומדיות אחרות, דוגמת לווייניים, רחפניים, מצלמות אבטחה וכו'. כמו כן, הכשרתם של אנשי המודיעין לשליטה בשפות זרות, ובמיוחד שפות מסובכות ומורכבות כמו מנדרינית, עשויה לקחת שנים. מערכות ה-AI מאפשרות כבר כיום תרגום מהיר, גם אם לא באופן מדויק לחלוטין, עבור שפות וניבים רבים ויכולות לסייע לחוקרי המודיעין ולחסוך זמן יקר. שימוש בטכנולוגיית AI רווחת כיום גם בתחום הסייבר לזיהוי תוכנות זדוניות ועצירתן תוך זמן אפסי. ה-AI מתפקדת כבר כיום כ"סוכנים בשטח" בחלק מהמקומות, שהגישה אליהן מסוכנת או מסובכת עבור סוכן שטח אנושי, זאת על ידי יצירת מעגל מצלמות.

השימושים הללו ועוד רבים אחרים שמוטמעים כבר כיום בקהילת המודיעין, מעלים חשש בקרב חלק מאנשי המודיעין כי ה-AI יחליף את מקום עבודתם. יחד עם זאת, פעילות משותפת של בני האדם וסוכני ה-AI, כאשר בני האדם אחראיים על פיקוח וחשיבה אסטרטגית, תוכל להגדיל את תפוקת המודיעין ולהנגיש אותו במהירות אל מקבלי החלטות.