



רובוט חילוץ משדה הקרב. צילום: חברת TATRC

בשנת 2009 הציעו חוקרים מהמכון הפוליטכני של קליפורניה להטמיע קוד אתי, שיהיה מבוסס על משהו דומה לשלושת חוקי הרובוטיקה של אסימוב ברובוטים לשימושים צבאיים. לדברי החוקרים אז, רובוטים צבאיים אוטונומיים יהיו חייבים לציית לקוד לוחמים או שהם יהפכו לאדוני האנושות.

בשנים שחלפו מאז, כפי שהדבר בא לידי ביטוי גם במסגרת מבצע צוק איתן, נעשה שימוש רב בטכנולוגיה, ובין היתר, הראו כיצד שימוש בכטב"מים המצויידים באמצעי ראיית לילה ואף (חלקם) באמצעי התקפה. היום, עיקר התבונה של המערכות הללו היא בסיוע לניווטם וניהוגם, אך כל הקשור לתפעולם נעשה בידי מפעילים אנושיים.

בעתיד הקרוב, ישויות בעלות בינה מלאכותית יוכלו להיות "יצורים מוסריים יותר מבני-אדם" – או לפחות כאלו שמקבלים החלטות טובות יותר בבואם לטפל בדילמות מסוימות.

התכנון של רובוטים בעלי יכולת מוסר עצמאית עשויה לכלול הטמעת יכולת מצפונית שפותחה באופן מושלם על מנת להבדיל בין טוב לרע, ולפעול על פי כך. רובוטים עתידיים אלו עשויים להיות מוסריים יותר מאשר בני-האדם.

בשנת 2002, הרובוטיקאי Veruggio Gianmarco טבע את המונח רובו-אתיקה (roboethics) – הטמעה של מוסר אנושי במסגרת התכנון, הייצור והתפקוד של רובוטים – תוך הדגשת יעד המחקר שאליו יש לשאוף. באותו זמן, המוסר של בינה מלאכותית נחלק לשני תת-תחומים:

מוסר של מכונות: הענף העוסק בהתנהגות של גורמי מוסר מלאכותיים;
מוסר של רובוטים: הענף האחראי לשאלות באשר להתנהגותם של בני-אדם – כיצד הם מפתחים, בונים, משתמשים ומתנהגים כלפי רובוטים כלפי יצורים אחרים בעלי בינה מלאכותית. תחום זה כולל גם את התהייה באשר לאפשרות לתכנת רובוטים כך שיהיו בעלי קוד אתי שיוכל לכוון את התנהגותם בהתאם לנורמות חברתיות המבדילות בין טוב לרע.

באופן הגיוני, על מנת שניתן יהיה ליצור רובוטים עצמאיים מבחינה מוסרית כעין אלו, החוקרים חייבים להסכים על מספר אמות מידה בסיסיות: מהי יכולת מוסרית ומה על בני-האדם לצפות מהרובוטים העובדים לצידם בעודם חולקים עימם את תהליך קבלת ההחלטות בתחומים כגון רפואה ולחימה. בה בעת, צצה ועולה שאלה נוספת: מהי אחריותם של בני האדם ביצירת בינה מלאכותית בעלת עצמאות מוסרית? והשאלה המחקרית העיקרית: מה עלינו לצפות מרובוטים בעלי יכולת מוסרית?

אחת הסוגיות החשובות ביותר, הגם ששאלות רבות אודותיה נותרו ללא מענה, היא באשר לרעיון של 'יכולת מוסרית'. הפרופסורים Malle .F Bertram מאוניברסיטת בראון ו-Scheutz Matthias מאוניברסיטת טוענים הם זה במאמרם. "חברתיים רובוטים של מוסרית יכולת": שכותרתו מחקרי מאמר השנה פרסמו Tufts כי יכולת מוסרית מורכבת מארבעה מושגים רחבים:

ליבה מוסרית: "מערכת של הנורמות ושל השפה והמונחים שנועדו להגדיר את הנורמות הללו, לרבות מושגים ושפה של מוסר ורשת של נורמות מוסר";
פעולה מוסרית: "מודעות למוסר ולתוצאות" – התאמת פעולותיו של הגורם לנורמות;
הכרה ורגש מוסריים: "קביעת החלטות ונקיטת פעולות מוסריות", התגובה התחושתית להפרת נורמות ושיפוט מוסרי;
תקשורת מוסרית: הנמקה, הצדקה, התמקחות והשלמה עם הפרות של מוסר.

תכנון של רובוטים בעלי יכולת מוסר עצמאית עשוי להיות מעורר השראה ומרתק, אולם ללא ספק הוא יהיה מאתגר. המהנדסים חייבים לעבור דרך שלבים הכרחיים. ראשית, הם חייבים לבנות ייצוג ממוחשב של מערכות מוסר ולהטמיע מונחים ואוצר מילים בתחום של מוסריות אל תוך הארכיטקטורה של הרובוט. בשלב הבא, עליהם לפתח אלגוריתמים שיוכלו להכיל באופן ממוחשב הכרה וקבלת החלטות מוסריות.

בתרחיש מעשי, לדוגמה, מערכת בעלת יכולת מוסרית עצמאית שנועדה למטרות תעבורה רפואית תצטרך להחליט אם שינוי נתיב התנועה שלה מנקודת ביקורת א' לנקודת ביקורת ב' היא הדרך הטובה ביותר להשגת התכלית שלה, שהיא העברת אספקה לאזור מוכה אסון או לשדה הקרב.

{loadposition content-related}