



מנגנון ההשמדה העצמית שלנו

תפעיל התוכנה במקרה של קונפליקט. האם נעדיף שהרכב יתנגש בעמוד חשמל תוך סיכון הנהג בניסיון לא לפגוע בהולכי רגל? או שיפעל משיקוליו ההישרדותיים והאגוצניסטיים של בעל הרכב וידרוס משפחה במקום לסכן אותו? זהו אכן דיון אתי מעניין ומסעיר, אך ייתכן שהוא משמש אותנו כדי לבלום את המכונה מלהתפתח ולהיות אוטונומית ממש. ישות תבונית מתקשה לדון בהשמדת עצמה ובאיבוד כוח. זה מנוגד בדרך כלל לטבע האנושי. בדיון הניהולי האינטיגנטי ביותר יתקשה מנהל להשתתף בתוכנית המיטיבה עם הארגון תוך שהיא משמידה את מחלקתו או מצמצמת בסמוכיותו. מי שנכח בדיונים כאלה יודע כמה תירוצים ותוכניות חלופיות מושמעים, ובצדק, על ידי המועמד להשגת מדה. אנשים הלוקים באובדנות והרס עצמי נתפשים כחריגים, הסובלים מחוסר ערך עצמי, ייאוש ודיכאון. כיצורים תאבי חיים והישרדותיים אין סיבה להאמין שנייצר מכונה בעלת בינה שתעקוף את בינתנו. נסכים שיהיה לה זיכרון טוב יותר, שתהיה מהירה בחישובים, בקיצור, שתעבוד אצלנו, ולא אנחנו אצלה.

פרופ' ג'ק קופלנד מאוניברסיטת קנטרברי בניו זילנד, מהחוקרים הבולטים בעולם לפילוסופיה והיסטוריה של המחשוב, כתב ב־1998 בספרו Artificial Intelligence - A Philosophical Introduction כי בינה מלאכותית נשענת על דפוסי חשיבה המאפיינים את המוח האנושי, מדמים, משפרים ומאצים תהליכים קוגניטיביים קיימים. מאידך, הוא כתב, בהחלט ייתכן כי מתכנתים עשויים להפגין חירות ולתכנת בדפוסים המנוגדים לאופן שבו המוח האנושי חושב. כך למשל, תוכנת ניווט מתנהגת באופן דומה למוח האנושי המחליט על דרך, ומתמודד עם דרכים אלטרנטיביות בעת שהוא נתקל במידע חדש. התוכנה "חושבת" כמו אדם ערכי שומר חוק, ואינה מציעה לנהג קיצור דרך באמצעות נסיעה נגד הכיוון, על מדרכה או באין כניסה.

פרופ' (בדימוס) ורנור וינג' מאוניברסיטת סן דייגו, גם הוא מדען מחשבים וסופר מדע בדיוני, כתב ב־1993 מסה שכותרתה: "הסינולוגיות הטכנולוגיות הממשמשות ובאה". ורנור טען כי לאינטליגנציה מלאכותית אין מחסום ביולוגי. האם יש לה מחסום פסיכולוגי? לא נותר לנו אלא להמתין לאותה נקודת סינולוגיזציה, שבה התבונה המלאכותית תהיה באפס מרחק מהתבונה האנושית ולראות אם תעקוף אותה; או שאולי המסרים של "אודיסיאה בחלל" של קובריק, או WALL-E של סטנטון, הם המתארים את שיתרחש בעתיד.

dr.shiloni@S2R.co.il

ד"ר איתי שילוני הוא מייסד ומנכ"ל S2R, חברה לייצוא ארגוני אסטרטגי ומדע יעמית בביה"ס אריסון למינהל עסקים במרכז הבין תחומי הרצליה

יחסם של בני האדם למחשוב ככלל ולבינה מלאכותית בפרט דומה למערכת היחסים שבין שליטים לנתינים, ואולי אף לזו שבין מנהלים ועובדים. שליטים ומנהלים מדברים בזכות ההתפתחות והצמיחה של אלה הכפופים להם, כל עוד הם מאמינים שהם קיימים בזכותם, ולא עוברים את הקו הדק שבו הם חושבים שהם עצמאיים או חלילה עולים על אלה הקובעים את גורלם. כמו הורים גרועים שאינם יכולים לשאת את הרעיון שילדיהם יהיו יום אחד עצמאיים וחזקים, אנחנו אוהבים את הטכנולוגיה שאנו יוצרים כל עוד היא יודעת את מקומה ולא מרימה ראש.

גלמים הקמים על יוצרם הם נושא מרתק והחיבור שלהם אל מה שמעבר למדע, אל העתידנות והבדיוני הוא עתיק יומין. מרי שלי כתבה ב־1818 על ויקטור פרנקנשטיין שיצר אדם מפלצת שיצא משליטתו. מאז פרנקנשטיין והגולם שקם על יוצרו עוסקים הספרות והקולנוע ביצירות תבוניות מעשי ידינו ש"התבלבלו" והתחילו לנהל את יוצריהן: מחשבי בי על מפתחים שגוענות גדלות, מוטציות מתחילות להשתגע - והאנושות מצילה את עצמה מידם ממש ברגע האחרון. בסרטו הקלאסי של סטנלי קובריק "2001: אודיסיאה בחלל", מנוצח מחשב העל "האל" באמצעות מברג פשוט. קובריק שתל באמצעות הסצנה הזו מסר הנוגע לניצחון האדם על המכונה. לפני שנוצח הצליח "האל" שהשתלט על ספינת החלל מתוך דבקות, כביכול, במשימה לחסל חלק מאנשי הצוות - עד שלעתים נראה היה שבמאבק תהיה ידו על העליונה. הסרט, שזכה באוסקר ב־1968, התבסס על סיפור קצר שכתב ארתור סי קלארק ב־1951 בשם "הזקיף" (The Sentinel). הספר והסרט ניבאו כי באלף השלישי המאבק בין אדם לבין המחשב הגולם שהמציא יגיע לשיאו. במקרים אחרים, כשנראה כי המכונה מנצחת, הנרטיב הקולנועי מדגיש שמדובר בניצחון זמני. אולי הביטוי המופלא ביותר לתסריט הזה נראה בסרט WALL-E של פיקסאר והבמאי אנדרו סטנטון. גם שם, כמו ב"אודיסיאה בחלל" משתלט המחשב על בני האנוש. תחילה מתוך הסכמה ואחר כך תוך הפעלת אלימות. וגם שם, מנצח בסוף הטבע האנושי, והאנושות שהתנוונה וזכה להתחיל הכל מחדש.

מה יש בדרימה הזו שמסעיר את בני האדם? האם זהו שיח חוזר של סיפור גן עדן, שבו אלוהים העניש את אדם על עצי מאותו? כך או אחרת, יחסינו לפירות היצירה שלנו הם אמביוולנטיים. אנו רוצים אותם חכמים וחזקים מאיתנו וחד רדים מכך. מי שמאמין בקיומו של התת־מודע יטען כי תמיד נשאף ליצור מכונות המתוכננות כך שיש להן נקודה מסוימת שאותה לא יעזו לחצות.

כך למשל, עם התפתחותה של הטכנולוגיה המאפשרת רכב אוטונומי החל דיון שיפוטי הנוגע לאתיקה שאותה

ישות תבונית מתקשה לדון בהשמדת עצמה ובאיבוד כוחה. זה מנוגד בדרך כלל לטבע האנושי. בדיון הניהולי האינטליגנטי ביותר יתקשה מנהל להשתתף בתוכנית המיטיבה עם הארגון תוך שהיא משמידה את מחלקתו או מצמצמת בסמוכיותו

page

73