

האם מכונת השיוויון יכולה לעבוד? מעת ערן טוך

ביקורת על הספר The Equality Machine by Orly Lobel

הספר The Equality Machine של פרופ' אורלי לובל מתחיל במסר שמאוד קוסם לאנשים כמוני: We need to cut through the utopian/dystopian dualism. בעולם שהוא כמעט בי-פולרי ביחס שלו לטכנולוגיה, בין אנשים ואפילו כאינדיבידואלים, המסר הזה נקרא אפילו כמהפכני. קחו, למשל, את הטרנד של השבוע החולף, את ChatGPT. האם זה סוף שיעורי הבית בהיסטוריה האנושית או שזה כלי נפלא שיעזור לנו בתהליך הכואב של כתיבת מכתבי המלצה?

המסר המעודן מאוד קורץ למהנדסים כמוני, שמגיעים מהפקולטה להנדסה, ובונים מערכות כדבר הבסיסי שהם עושים. הספר עוסק בשאלה מאוד מהותית שאנו שואלים היום בעולם בו הטכנולוגיה משתנה בקצב הולך וגדל. בתוך המאניה-דיפרסייה הזו, השאלה הבסיסית היא איך פוליטיקה מתבטאת בטכנולוגיה. הספר נותן לנו דוגמא כזו.

הספר היה לי שימושי במיוחד כי הוא פותח דלת לעולם שלם של חוקרים, מתכנתים, ואקטיביסטים שמתכנתים התערבויות בעולם. למשל:

- המאמר של תמי כריכלי כץ
- המאמרים של אורלי לובל עצמה וניר גרינברג, למשל, על pro social behavior.

הספר מלא בדוגמאות כאלו, עם יצירתיות מתפרצת. הדוגמאות הם של טעינת פוליטיקה מסוימת לתוך טכנולוגיה, פוליטיקה שיכולה להתקיים בגלל מספר תכונות של AI/DS. אפשר לסווג את התכונות האלו למספר מימדים:

1. הן פורמליות. למשל, ההחלטות שלהן מבוססות על הגדרות ברורות, שלא תלויות בפסיכולוגיה הפנימית של מקבלי ההחלטות. לכן אנחנו יכולים לקבל מאמרים כמו של קליינברג שמנתחים את קבלת ההחלטות של מערכת "Age of Algorithms", ולא מאמרים מוזרים שמנסים להבין אם שופטים משחררים אנשים בערבות בשיעורים גבוהים יותר אחרי ארוחת הצהריים. (בסוגריים, אני חושב שצריך לחוקק חוק שאוסר על מי שעוסק באתיקה של AI לדבר על COMPAS. סבלנו מספיק)
2. הן מיידיות. כמו האפליקציה spot שמאפשרת לנהל ראיונות עם קורבנות של הטרדה מינית או כמו Safe-Up.
3. הן יודעות כל, ויכולות לדעת מה קורה בכל מקום. כמו OpenComp שיכולה לנתח משכורות בחברות ולזהות bias. כמו PhotoDNA שמוצאת חומר פורנוגרפי. הן אוגרות הכל. מייצרות Paper trail.
4. הן מאפשרות ביצוע ניסויים. למשל, המערכת שאשרת בנתה.

5. הן יכולות לבוא מה יש בעתיד. כמו Advanced Discovery, Nex, Botler, AwareHQ, and Emtrain. ניתן דוגמא נוספת מהמחקר שלי.
6. הן יכולות לכונן אנשים. מהוות פלטפורמה ואז מאפשרות nudging של prosocial behavior.

ההבנה של המסגרת הזו היא רלוונטית מאוד לדיון שאפשר לקיים, דיון שיכול לעזור לנו להפריד בין המסגרת שהטכנולוגיה נותנת לנו לבין היכולות הפוליטיות שלנו.

האם אפשר לחשוב על אפשרויות אופטימיות מבלי לחשוב על התהליך שממש מביא אותו לביטוי. כן, יש כאן בעיה בחשיבה אופטימית? בעיקר להיכנס לסחרור של Economies of Virtue. אנחנו לא רוצים להגיע למצב בו חברות מגיעות לכנסים כמו ACM FaCCT וכולם מדברים על AI Ethics, ומשקיעים במרכזי מחקר כאלו, אבל מה שמגיע בפועל לשטח אלו הם אלגוריתמים לזיהוי פנים שמופעלים על ידי המשטרה. יש כאן, במילים של Thao Phan, Jake Goldfein, Declan Kuch and Monique Mann סוג של "Economies of Virtue". מה זה אומר? זה אומר שחברות משקיעות כספים ומשאבים בתחושה שהנזק של AI הוא קטן, ואפשר לתקן אותו, מבלי שהפוליטיקה ממש מביאה את הקוד ל-production.

ניקה, כדוגמא, את נושא ה"מידע המושלם" ש-AI/DS נותנים לנו. יש הרבה דוגמאות שאורלי מעלה בה הטכנולוגיה נותנת לנו מידע מושלם, למשל המשכורות בחברה מסוימת וכל המידע הנלווה, כך שנוכל לדעת אם היתה הפליה. אבל כשאנחנו מסתכלים על המקום של מידע מושלם בהיסטוריה, ההיסטוריה די קודרת. אז מה זה מידע מושלם? זו הקלסיקה של surveillance capitalism. לגוגל ולפייסבוק יש מידע מושלם, האם זה מספיק? האם אנחנו, כחברה, נהנים מהמידע המושלם הזה? או שהוא משמש בעיקר למכור מודעות לטאבוניס לפיצות? מופנה לניצול של כל פיסה של תשומת לב שיש לנו.

דוגמא נוספת היא ה-Credit Score. למרות שיש מידע מושלם על הלווים, המחקרים מראים שאין כאן תרומה ממש למוביליות חברתית או אפילו לניהול סיכונים של חברות, כפי שראינו ב-sub prime mortgage crisis. [אני לא חושב של אורלי היתה תשובה רלוונטית לנושא הזה]. יש גם פרדוקס די בסיסי בעידן שבו יש יותר ויותר ידע, אבל יותר ויותר misinformation. נראה שיש לנו, כחברה, נוגדנים כלפי ידע.

כדי שנוכל לעשות את הקישור הזה בין המסגרת ה"מושלמת" שהטכנולוגיה נותנת לנו לבין מה שעושים אנחנו צריכים לשאול משהו פשוט יותר: מה צריך לעשות כדי שבאמת הערכים שלנו יתבטאו בטכנולוגיה הזו. השאלה היא לא לגבי הפוטנציאל של הטכנולוגיה, נושא שאני חושב שבו הטיעון של אורלי הוא משכנע ומעניין. אני לא רוצה לטעון כאן כי טכנולוגיה היא אגנוסטית. ברור שגם המסגרות בה טכנולוגיה מפותחת היום כוללת בסיס פוליטי (למשל, העובדה שרוב הטכנולוגיה מפותחת כיום על ידי חברות פרטיות ולא על ידי ממשלות, כפי שהממשלה של איינדה רצתה לעשות עם פרוייקט ה-Cybersyn ב-1971). האם אפשר לתת לחברות לעשות את מה שהן עושות כי הן מורכבות מאנשים נהדרים?

השאלה הראשונה היא האם חברות הטכנולוגיה יכולות להביא את הקוד ה"טוב" יעבור מה-test ל-production? האם חברות יעשו את זה בעצמן? למי שקורא את הספר באופן נאיבי, יכולה להיווצר

התחושה שאם נחכה מספיק זמן, חברות הטכנולוגיה יעשו את "הדבר הנכון" כי הן כוללות את "האנשים הנכונים". שאורלי מסייגת את התחושה הזאת, אבל אני חושב שמדובר רק בהסתייגות, לא בטיעון המרכזי של הספר. אורלי מזכירה את Timnit Gebaru ואת Frances Haugen, אבל אנחנו צריכים לזכור שהן פוטרו ממקום העבודה שלהם. אנחנו צריכים להכיר בעובדה ש-OpenAI מגייסים מיליארדים, אבל יישומי אתיקה מוצגים רק בכנסים אקדמיים.

העובדה שהטכנולוגיה קיימת ואפשר ליצור אותה, לא אומרת שכך יצרו אותה. יש משהו בדינמיקה של חברות פיתוח הטכנולוגיה שמונע מהן לעשות את זה בעצמן, בצורה עצמאית. במאמרים שפרסמה אשורת איילון, סטודנטית שלי, יחד עם מיכאל בירנהאק, ועירית הדר מאוניברסיטת חיפה, ראינו את הדינמיקה בה מהנדסים בונים מסגרות, כמו client server או blockchain, והדין הפוליטי מתבצע באותן מסגרות. אנחנו גם רואים איך היבטים של תרבות הנדסית מגיעים באופן טבעי למדי לאופן שבו הטכנולוגיה מיושמת. למשל, כיוון שהתוכנה מפותחת בחוסר ודאות, יש נטייה לאסוף עוד מידע. איך אפשר להגיע לכאן?

אם אי אפשר לסמוך על חברות הטכנולוגיה, אז על מה אפשר לסמוך? האם יש מקום לתפקיד של גולציה וחקיקה. האם זה המקום בו רוב הפתרון קיים? יש נקודה בספר, שהיא מאוד משנית אבל אולי מצביעה על משהו יותר עמוק בעיניי. למשל, כאשר מדובר על הכלי שהממשלה השוויצרית פיתחה לחברות לבדיקה של שכר שווה, אולי לא מזכירה את העובדה שהכלי הזה עובד במסגרת של חקיקה מ-2020, לפיה חברות מעט 100 עובדים חייבים לבצע בחינה של המשכורות על פי מדדים ולפרסם את התוצאות לעובדים ולבעלי מניות (בישראל יש חקיקה דומה, אבל עם מדדים מאוד לא ברורים). לא ברור איך בהיעדר של חקיקה וקונפליקט ממשי בין האקוסיסטם הטכנולוגי לבין הממשלה אפשר לעשות משהו. קונפליקט כזה הוא לא מובן מעליו, והוא בוודאי לא "טבעי". ראו, למשל, את הקונפליקט בין אפל וגוגל מצד אחד וממשלות (צרפת, גרמניה, אוסטרליה) מצד שני על פרוטוקול איתור המגעים בתקופת הקורונה. הוא נגמר בניצחון מוחלט של חברות הטכנולוגיה, ושום חקיקה לא היתה פותרת את הבעיה.