הנדסה להתנהגות פרו חברתית יחד עם סוכנים אוטונומיים בעזרת בינה מלאכותית (AI):

מאמר זה חוזה עתיד שבו סוכנים אוטונומיים(הנדסה של מכונות חכמים שתסייע לבני אדם) משמשים לטפח ולתמוך בהתנהגות פרו-חברתית בחברה היברידית של בני אדם ומכונות. התנהגות פרו-חברתית מתרחשת כאשר אנשים וסוכנים מבצעים פעולות כלשהם שעוזרות לאחר (לזולת). מעשים כגון סיוע לאחרים מרצון, תרומה לצדקה, מתן מידע או שיתוף משאבים, הם כולם צורות של התנהגות פרו-חברתית.

שאלות מחקר :

א. מה הם התנאים והמנגנונים שמובילים חברה של סוכנים אוטונומיים ושל בני האדם להיות יותר פרו-חברתיים?

ב. כיצד אנו יכולים להנדס ישויות אוטונומיות (סוכנים ורובוטים) שיובילו להתנהגויות אלטרואיסטיות (אנטי-אגואיסטיות) ושיתופיות יותר בחברה היברידית?

ההתעניינות ביישומי AI לטובת החברה אינה חדשה, והיו גל של התפתחויות ואירועים חדשים בשנים האחרונות. כמו כן ישנו את הארגון- "AI and Operations Research for Social Good", אשר מטרתו לחקור ולקדם את יישומי האינטליגנציה לטובת החברה ,למעשה, קרן XPRIZE ארגן את " AI לפסגה עולמית " שבו מציע גישות ליצירת ערים ברות קיימא, התמודדות ותגובות במקרי אסון, לטפל בהשפעה של אי שוויון, או לשפר את בריאות הציבור. העבודה המוצעת כאן הולכת בכיוון זה, שיש לה פוטנציאל לגרום להשפעה בחלק מתחומי היישום האלה.

כיצד ניתן להשתמש בסוכנים אוטונומיים כדי לטפח או לדחוף לשיתוף פעולה בחברה של בני אדם ומכונות? כיצד ניתן לעצב סוכנים אוטונומיים אשר נמצאים בסביבה של בני אדם, שיוכלו לקדם פעולה קולקטיבית במצבים שבהם זה לא יכול להתקיים באופן טבעי? כיצד נוכל לטפח שיתוף פעולה בארגונים, לעזור לאנשים להתמודד עם בריונות ברשת כאשר הם עדים לה, להיאבק בבעיית העוקבים, לגרום לאנשים לעסוק בטוב חברתי, לקדם הרגלים ברי קיימא, לשנות את האקלים, וכן הלאה? האם מערכות אוטונומיות יכולות למלא תפקיד מסוים בבעיות טכניות שונות? מספר מנגנונים שזוהו כתומכים בשיתוף פעולה, במצב שבין דילמות של שני אנשים לבין בעיות פעולה קולקטיביות רחבות היקף.

אנו מאמינים כי סוג חדש של מחשוב (סוכנים ורובוטים) יהיה מקושר עם היבטים של שקיפות , אחריות והשתתפות. ראשית, מחשוב להתנהגות פרו-חברתי תוגדר כ"מחשוב המכוון לתמיכה ולקידום פעולות המוטלות על החברה ועל אחרים". זהו מושג רחב העשוי לכלול דעות אלטרנטיביות שונות על איך להנדס פרו-מחשוב חברתי. כדי להפוך אותו למציאותי יותר, נתחיל בהצגת תרחישים פשוטים בהם ניתן להשתמש במחשוב פרו-חברתי.

הבה נמחיש מצבים פשוטים שבהם מחשוב פרו-חברתי, ובמיוחד סוכנים פרו-חברתיים, עשויים למלא תפקיד בשינוי הדינמיקה החברתית הלא-שיתופית השוררת בחברה היברידית של בני אדם ורובוטים.

"אפקט העומד מן הצד"-כאשר אנשים עדים למקרה שאדם זקוק לעזרה ואותם עדים לא מגישים עזרה כלל (מכל סיבה שהיא) יקראו כך או בכינוי "אפקט העוקב".

לפי מחקרים רבים עולה שככל שמספר האנשים העדים לאירוע מצער גדל, נכונותם לסייע קטנה. בתרחישים שמתווכת מחשב (למשל, מדיה חברתית) אנו עדים לעלייה בנכונות לסייע. כמות הזמן להתערבות עולה עם מספר האנשים העדים למצב (עוברי אורח וירטואליים).

מבחינה טכנולוגית ניתן לשאול האם ניתן לטפל באפקט זה, ובמיוחד אם:

האם מכונות אוטונומיות וסוכנים (במיוחד אם הם מגולמים בעולם הפיזי) ייחשבו כ"קהל" ב"אפקט העומד מן הצד "? כלומר, האם המכונות האוטונומיות האלה יגדילו את אפקט העוקב?

השפעה חברתית? האם מכונות / סוכנים יכולים להפגין התנהגויות (בין אם על ידי משחק או אי-משחק), שישפיעו על התנהגויות של אחרים (ושל בני-אדם)?

אם הסוכנים יכולים להיות בעלי השפעה חברתית על בני אדם, האם הם יוכלו לפעול נגד פעולת "אפקט העוקב"? אם כן, איך נוכל לבנות טכנולוגיה בשביל זה?

ניסוי בשני סוכנים פועלים זה עם זה: המציע ניחן במשאב כלשהו ויש לו להציע חלוקה עם המגיב. אם המשיב דוחה את ההצעה, אף אחד מהשחקנים לא מרוויח דבר. אם ההצעה מקובלת, הם מתחלקים. בהקשר של UG (משחקים

אולטימטומים), רק החלוקה השוויונית, שבה גם המציע וגם המשיב מקבלים שכר דומה, נחשבת לתוצאה הוגנת. מחקרים רבים מעידים כי אנשים הוגנים במשחקים אולטימטומים.

"מטא-אנליזה" ביצעו 100 ניסויים שכללו למעלה מ -5,000 נבדקים גילתה כי באופן כללי ההזדמנויות של תקשורת אנושית הגדילו באופן משמעותי את שיעורי שיתוף הפעולה . הייחודיות של תהליך ההתלבטות האנושית משפיעה גם על רמות שיתוף הפעולה שנצפו. כאשר אנשים מקבלים החלטות מהירות ואינטואיטיביות בתרחיש משותף, יש יותר שיתוף פעולה מאשר כשאנשים מקבלים את החלטותיהם לאחר זמן מה לדיון ולהשבה מחדש.

אנו מאמינים כי מחשוב פרו-חברתי יכול להביא את ההזדמנות להנדס מערכות הוגנות, תוך שימוש במקצועות פסיכולוגיה וביולוגיה אבולוציונית. לדוגמה ,שרוב הרובוטים מסייעים בתיאום בין אוכלוסיות של בני אדם וסוכנים. בכדי להנדס את הסוכנים האוטונומיים נצטרך לחקור את הנושאים הבאים:

• ביצוע מחקרים ניסיוניים עם בני אדם וסוכנים המשתמשים בדילמות חברתיות כדי להבין את התנאים והמצבים בהם מופיעים התנהגויות פרו-חברתיות;

• התנהגויות ספציפיות בהנדסה (ואולי אף פתולוגיות) בתרחישים הראשונים לסימולציה חברתית על מנת לבחון את ההשפעות על אוכלוסיות;

• ביצוע מחקרים עם בני אדם וסוכנים וירטואליים בעולמות וירטואליים. סוכנים יכולים להיבנות כפרו-חברתיים (בהתחשב בתוצאות הקודמות), דבר הגורם להתנהגות פרו-חברתית.

• הנדוס רובוטים חברתיים כסוכנים פרו-חברתיים כדי לבדוק אותם במרחבים פיזיים טבעיים, שבו בני אדם וסוכנים מתקיימים יחדיו.

למעשה, אנו רואים באמפתיה כחיונית לטפח פעולות פרו-חברתית ולכן האמפתיה תצטרך להיות מסונתזת בסוכנים וכמו כן למדל אותם כהיוריסטיים(כלל חשיבה פשוט המבוסס על הגיון פשוט ואינטואיציה) כדי להבין אחרים - בעיקר באינטראקציה קבוצתית.

על הסוכנים להיות מסוגלים להבחין בין התנהגויות טובות ורעות בהקשרים ספציפיים, אשר יש להתייחס אליהם הן במהלך תהליך קבלת ההחלטות שלהם והן בשפיטת פעולות הסוכנים הסובבים.

בעזרת תיאוריה של המוח ניתן ליצור מודלים של המצב הפנימי של סוכנים ובני אדם ולדון עליהם.

כל היכולות הללו יהוו את אבני הבניין של הסוכנים שיאפשרו להם לקבוע את מידת הכדאיות של מאורע לסביבה וכמו כן להוסיף הערכות אישיות למאורע הקיים.

השאלות שנשארו לעתיד:

האם מכונות אוטונומיות וסוכנים יכולים לערער (או לחזק) את הקשרים החברתיים והתרבותיים הקיימים בחברה ולרוקן (או להגביר) את רמות ההגינות ?

האם "מכונות אוטונומיות" חסרות ייחוס סיבבתי אנושי, מה שיוביל אותן לפטור מהתנהגויות לא הוגנות?

האם יוכלו המכונות לעסוק בסנקציות ו / או בהסדרים הדדיים, המכוונים לעתים קרובות כהוגנות בחברות אנושיות?

נושא המאמר ברובו תאורטי וכמו כן התרחישים שהוצגו בו עם זאת ,המאמר צופה אל העתיד ובעזרת בינה מלאכותית ותחומים נוספים על מנת להפוך את החברה לחברה יותר שיתופית הוגנת וחברתית יותר.

לעניות דעתי, כותב מאמר זה חוזה את המובן מאליו שבו לטכנולוגיה ומערכות אוטונומיות יהיה חלק אינטגרלי בחיי הפרט ובחברה כלל.

לכן אני סובר שאם התנאים שצוינו לעיל יתקיימו – קיום הטכנולוגיה להנדוס הסוכנים והשמה בחברה עם כל התנאים והמנגנונים- ניתן להפוך את החברה לחברה פרו-חברתית אך לא צריך להתנער מהאחריות שלנו כבני אדם לחנך לעזרה הדדית וחמלה אנושית.