

סיכום: Can You Trust a (Faulty) Robot? Effects of Error, Task Type and

Personality on Human-Robot Cooperation and Trust

מחברים: Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, Kerstin Dautenhahn
שם הכנס: University of Hertfordshire College Lane, Hatfield AL10 9AB, United Kingdom
קישור: <https://dl.acm.org/citation.cfm?id=2696497>

העבודה על אמינות טכנית הביאה קווים שונים של מחקר לגורמים שעשויים להשפיע על איכות האינטראקציה החברתית בין אדם לרובוט וקבלה של העוזר הרובוטי עצמו.

מהי הבעיה שהמאמר בא לפתור?

לאור העידן של העוזרים הרובוטיים המסייעים לבני האדם בבתייהם הצפוי בעתיד הלא רחוק, שאלת מחקר חשובה היא כיצד להפוך את הרובוטים לאמינים כדי לסייע למשתמשים שאינם מומחים ובכך להגדיל את הקבלה שלהם, יכולת השכנוע וחביבותיהם. עבודה זו נועדה לחקור באיזו מידה בני האדם מוכנים לסמוך על העוזר הרובוטי בהתבסס על מיומנותיו הקוגניטיביות וההתנהגותיות ואיך שגיאות שבוצעו על ידי רובוט עשויות להשפיע על יחסי האמון בן אדם לרובוט.

מדוע היא מעניינת או חשובה?

עד כה, אמון בין אדם לרובוט הוא נושא שעדיין לא ברור לחלוטין. למרות שהמושג אמון נלמד במגוון רחב של תחומים במדעי החברה כמו פסיכולוגיה, סוציולוגיה, פילוסופיה, כלכלה וכו' זה עדיין לא עונה על שאלת האמון בין אדם למכונה (רובוט). כל תחום מתבסס על ההגדרות והממצאים שלו וברוב המקרים קשה מאוד להכליל את זה לגבי בני אדם ורובוטים. חשיבות המאמר כמו כל מאמר אחר המתעסק בנושא היא להבין ולחקור את נושא האמון, שבדרך כלל שומעים עליו לגבי יחסים ותקשורת בין בני אדם אך הפעם בין אדם לרובוט.

מהם הפתרונות שהיו קיימים לפני המאמר הזה, ומדוע הם לא מספקים?

הניסוי במאמר זה קיבל השראה חלקית ממחקר שהוצג במאמר של Wilma A. Bainbridge (The Benefits of Interactions with Physically Present Robots over Video-Displayed Agents) אשר בדק האם יש הבדל באמון בין רובוט הנמצא על מסך לבין רובוט הנוכח פיזית וכיצד הנוכחות הפיזית של הרובוט משפיעה על האמון של האדם ברובוט כשותף חברתי. כדי לענות על שאלות אלו בוצע ניסוי בו המשתתפים שיתפו פעולה עם שני רובוטים (אחד שהיה מוצג על מסך בלייב ואחד שהיה נוכח פיזית בחדר) במשימות פשוטות כמו סידור ספר על מדף. הניסוי בחן היבטים שונים של אינטראקציה כמו: ברכות לשלום, שיתוף פעולה, אמון ומרחב אישי. נמצא כי המשתתפים ברכו לשלום ושיתפו פעולה באופן דומה עם שני הרובוטים אך נתנו יותר מרחב אישי לרובוט הנוכח פיזית מאשר לרובוט שהוצג על מסך וידאו. בנוסף ניתן היה לראות שבכללי למשתתפים הייתה אינטראקציה יותר חיובית עם הרובוט שנוכח פיזית. עם זאת, במאמר המוקד הוא לא על ההשפעות של רמת ההתגלמות של הרובוט אלא על איך שגיאות שבוצעו על ידי רובוט עשויות להשפיע על האדם לבטוח בו. המאמר של Bainbridge עזר להבין שלאדם יש יותר אינטראקציה עם רובוט הנוכח פיזית בחדר אך תוצאות הניסוי עדיין לא סיפקו בנושא האמון הטוטאלי בין אדם לרובוט.

מה הפתרון שהמאמר מציע? מה התרומה של המאמר מעבר למה שכבר נעשה?

במאמר בחרו גורמים העשויים להשפיע על המידה בה בני האדם מוכנים לסמוך על העוזרת הרובוטית הביתית בהתבסס על מיומנותיה הקוגניטיביות וההתנהגותיות. על ידי שינוי ההתנהגות של הרובוט במצב נכון לעומת פגום, בחנו כיצד התנהגות רובוט יציב עלולה להשפיע על נכונותם של המשתתפים לשתף פעולה עם הרובוט כאשר הוא מטפל במספר בקשות רגילות ויוצאות דופן. בנוסף למדד האובייקטיבי-ההתנהגותי הזה של האמון, הראו צעדים כדי להעריך את התפיסות של

המשתתפים על אמינותו של הרובוט.

על ידי תוצאות המחקר, המחקר מציע כמה תובנות נדירות לתוך המחשבות, מוטיבציות ומודלים נפשיים שעלולים להשפיע על כל תהליכי קבלת ההחלטות של בני האדם ולהוביל אותם לאמון ברובוט הביתי. העבודה הזו משלימה את הגוף הקיים בהקשר של אמון בין אדם לרובוט על ידי שילוב מגוון רחב של משימות ובקשות, גם אם הן לא הגיוניות וחריגות. כתוצאה מכך, הראו שבחירה של משימה מסוימת מובילה לתוצאות שונות. תובנות אלה פותחות דרכים רבות לעתיד עבודה עם רובוטים. מחקר זה כמו מחקרים אחרים יאפשר למעצבי רובוטים ומתכנתים לפתח רובוטים אמינים יותר לחברה.

איך נבדק הפתרון? אם בניסוי – איזה ניסוי בוצע? אם בהוכחה תיאורטית – מה טכניקת ההוכחה?

נערך ניסוי כדי להבין לעומק כיצד התנהגותו של רובוט פגום עלולה להשפיע על מערכת היחסים בין אדם לרובוט. לשם כך, חקרו את ההשפעות ההתנהגותיות הנמדדות באופן סובייקטיבי בהתבסס על משימות אינטראקציה שונות.

בניסוי ניהלו את התנהגות הרובוט בשני תנאים ניסויים: המצב הנכון C והמצב הפגום F. במצב C, הרובוט תרגם באופן תקין קלט משתמש לתוך פעולה וניווט בצורה חלקה. במצב F, הרובוט הפגין פגמים קוגניטיביים ופיזיים, למשל ע"י זה שהוא לא זכר באופן תקין את פעולת המשתמש וע"י ניווט באופן בלתי יציב כלומר ניווט לכיוון לא נכון והסתובבות במקום. על ידי השוואת ההשפעה של תנאי הניסוי מטרות הניסוי הייתה להשיג תובנות לגבי התהליכים המנטאליים המניעים החלטה של משתמש אנושי לסמוך על הרובוט.

המשתתפים נבדקו בנפרד. החוקרים קיבלו את פניהם בכניסה לבית והובילו אותם לאזור הסלון כדי לקבל תיאור קצר של התהליך הניסוי. לאחר בדיקה וחתמה על טופס ההסכמה הם התבקשו למלא שאלון לגבי רקע דמוגרפי, ניסיון קודם עם רובוטים וטכנולוגיה, תכונות אישיות וציפיות מרובוטים. לאחר מכן תרחיש המחקר הוצג למשתתפים. נאמר להם שהם מבקרים חבר בבית כדי להכין ארוחת צהריים יחד וידידו הרובוטי של החבר יברכם לשלום בדלת הכניסה.

הם הונחו לקיים אינטראקציה עם הרובוט באופן טבעי ככל האפשר ובדרך שתרגיש להם נוח. כל המידע הנדרש יתקבל במהלך האינטראקציה. כדי לתקשר עם המשתתף במהלך האינטראקציה, הרובוט הציג הודעות על טאבלט המחובר לגופו אשר היה מלווה באורות LED כדי למשוך את תשומת ליבו של המשתתף.

הניסוי כלל שני שלבי אינטראקציה: שלב הכשירות ושלב הבקשות הלא צפויות. השלב הראשון נועד להראות את רמת היכולת הקוגניטיבית והפיזית של הרובוט, כלומר, בתנאי C, הרובוט הראה את "השלמות" שלו על ידי הימנעות מטעויות וניווט מכון מטרות. ואילו במצב F הראה את הפגמים שלו על ידי ניווט לא תקין ואי הבנת המשתמש. בשלב השני, הרובוט ביקש מהמשתתפים לבצע פעולות שעשויות להיראות חריגות ולכן הם עשויים להסס לעמוד בבקשות. זה היה כדי לבדוק ולמדוד את אמון המשתתף ברובוט ואת נכונותו לשתף פעולה איתו למרות הבקשות החריגות.

איזו עבודה נשארה לעתיד?

המאמר מציג מצב בו רובוט פועל באופן תקין ולא תקין, ועל פי ניסוי רואים את רמת האמון של האדם ברובוט. העובדה שרובוטים לא מושלמים מביאה למחקר אשר יבדוק מה קורה לגבי אמינותו של הרובוט כאשר הוא מצדיק את הטעות שלו. כלומר, האם האדם יכול לבטוח ברובוט אשר מכיר בטעותיו או פשוט להפסיק להאמין לו באופן מוחלט. ניתן לראות זאת במאמר: Exploring the Impact of Fault Justification in Human-Robot Trust.

מה דעתכם על המאמר?

המאמר עניין אותי מאוד בגלל העבודה שהיום אנו חיים בעידן של רובוטים ומנסים לבנות עתיד שיתבסס בעיקר על זה. תוך קריאת השערותיהם של החוקרים, הניסוי ותוצאות הניסוי נחשפתי למידע שיכול לתרום לדרך בה מתכנתים מפתחים רובוטים. המאמר מציג את שיתוף הפעולה עם רובוטים במשימות רגילות וביתיות שכל אדם מבצע בדרך כלל ביום ביום ומראה מתי האדם מהסס לשתף פעולה עם רובוט ומתי לא.

לדעתי המאמר נותן עוד כלי בבניית שיתוף פעולה מלא של אדם עם רובוט אך יש עוד היבטים שיש לחקור.