בית הספר
למנהל עסקים
הפקולטה למדעי החברה
אוניברסיטת בר־אילן

תאוריה ופרקטיקה בניהול
כתב עת בהוצאת בית הספר למנהל עסקים
יוני 2026, סיון תשפ״ו | גיליון מס׳ 11

# בינה מלאכותית לניהול סיכונים בזמן אמת: ״אי-ודאות נלמדת״ כמגשר בין מודלים הסתברותיים להחלטות עסקיות

**ליאור טובלי**
בית הספר למנהל עסקים
אוניברסיטת בר-אילן

**פרופ׳ איל יניב**
בית הספר למנהל עסקים
אוניברסיטת בר-אילן

**פרופ׳ זאב זלבסקי**
הפקולטה להנדסה
אוניברסיטת בר-אילן

## תקציר

בארגונים, החלטות מתקבלות כמעט תמיד תחת אי-ודאות: נתונים חלקיים, מקורות מידע שלא תמיד אמינים, שינויי סביבה מהירים ומודלים שלא תמיד משקפים את המציאות. במצבים כאלה הבעיה אינה רק דיוק החיזוי, אלא אמינות מדד הביטחון שעל בסיסו מחליטים: האם לאשר פעולה אוטומטית, להסלים לבקרה אנושית, להקצות משאב נדיר, או לדחות החלטה עד לקבלת מידע נוסף. כאשר מדד הביטחון אינו מותאם לתנאי השטח מתקבלות שתי תקלות סימטריות עם מחיר ארגוני ברור: ביטחון יתר שמייצר תגובת יתר והתראות שווא, או ביטחון חסר שמייצר שמרנות, איטיות ופספוס אירועים.

מאמר זה מציג מסגרת מאוחדת לניהול אי-ודאות כמשאב ארגוני: שומרים על ליבה הסתברותית שקופה וברת ביקורת (מנגנון רקורסיבי בסגנון מסנן קלמן לא ליניארי), ובמקביל מוסיפים רכיב בינה מלאכותית שמכוונן בזמן אמת את ״מדיניות האמון״ של המערכת, כלומר לקבוע בכל רגע מה מידת האמון במודל לעומת מידת האמון במדידות, על סמך ההקשר התפעולי וסימני קוהרנטיות פנימית. המאמר הנוכחי מבוסס על שלושה מחקרים קודמים של הכותבים, שכל אחד מהם מציג יישום משלים של אותה תפיסה בהקשר שונה. כאן מתבצעת אינטגרציה של שלושת הנדבכים למסגרת אחת, עם ניסוח ניהולי אחיד שמאפשר לתרגם את העקרונות למדיניות ספים, להסלמה מבוקרת, ולהקצאת משאבים תחת סיכון. ההתאמה מתבצעת תחת אילוצי יציבות וחוקיות, באופן שמונע ״הסתגלות אגרסיבית״. המסגרת הודגמה בשלושה יישומים משלימים (ניווט, סינון רב־אובייקטים מודע־קטגוריה וחיזוי מוקדם תומך החלטה), ומודגשת בה ההבחנה הניהולית הקריטית בין שיפור דיוק ממוצע לבין שיפור אמינות שניתן לתרגם למדיניות סף, תיעדוף משאבים והפחתת סיכון.

**מילות מפתח**: ניהול סיכונים בזמן אמת, אי-ודאות תפעולית, מדדי ביטחון וכיול הסתברויות, הקצאת משאבים ותיעדוף, אוטומציה מול בקרה אנושית, מערכות תומכות החלטה

## מבוא

בארגונים מודרניים, כמעט כל החלטה חשובה מתקבלת כשחלק מהתמונה חסר: נתונים מגיעים באיחור, מקורות מידע אינם אחידים, והסביבה העסקית משתנה מהר יותר מהיכולת לעדכן נהלים. לכן, השאלה הניהולית אינה רק ״מה המודל מנבא?״, אלא ״עד כמה אפשר לסמוך על מדד הביטחון שמלווה את ההמלצה, ומה עושים כשאי הוודאות עולה?״. בשנים האחרונות מתנסחת סביב

סוגיה זו מסגרת מדיניות ותפעול ל AI, המדגישה אחריותיות, שקיפות, ניהול סיכונים ובקרות לאורך מחזור החיים. לפי מדריכים וסטנדרטים, דיוק כשלעצמו אינו תנאי מספק; יש להגדיר אחריות לסיכונים, לתעד הנחות, ולהקים מנגנוני ניטור ושיפור שיטתיים (National Institute of Standards and Technology (NIST), 2023; Office of the Privacy Commissioner for Personal Data (PCPD), Hong Kong, 2024).

ברמה הארגונית, גם רגולציה מתהווה וגם תקנים חדשים מכוונים את ההטמעה לכיוון של ניהול סיכונים פרגמטי, מה נחשב שימוש "סביר", מה חייב בקרה, ואיך מתעדים כדי שניתן יהיה להסביר החלטות בדיעבד (International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC), 2023b, 2023a).

הצורך הזה אינו תאורטי. בסייבר למשל, סף גבוה מדי מציף צוותי SOC בהתראות שווא, וסף נמוך מדי מאפשר חדירה שקטה למערכות המחשוב; בהונאה, חסימת יתר פוגעת בהכנסות ובחוויית לקוח וחסימת חסר מגדילה הפסדים; בשרשרת אספקה, ביטחון יתר בתחזיות ביקוש גורר מלאי עודף או חוסרים יקרים; ובמוקד שירות, ההחלטה אם להמשיך באוטומציה או להעביר לנציג תלויה בשאלה האם ההמלצה "בטוחה מספיק". כאן נכנסת הספרות על אמינות הסתברויות ואי-ודאות: מודלים רבים נוטים להיות לא מכוילים ולהציג ביטחון יתר, כך שמספר כמו 0.8 אינו בהכרח "80% אמינות". מעבר לכך, גם כשמודל היה טוב ביום ההשקה, הוא נשחק עם הזמן בגלל שינויי שוק, מוצר ולקוחות ולכן נדרש ממד תפעולי של ניטור, זיהוי סטיות ותיקון מבוקר. בהיבט הניהולי הרחב, מחקרי ממשל מצביעים על פער חוזר בין עקרונות אתיים/רגולטוריים לבין פרקטיקות שניתן למדוד, לנהל ולהטמיע(Birkstedt et al., 2023).

על רקע זה, מאמר זה מציע למסגר את אי-הוודאות כמשאב ניהולי: גודל שניתן למדוד, לכייל ולעדכן באופן מבוקר, ובכך לחבר בין ביצועי מודל לבין מדיניות סיכון ארגונית. בהקשר זה חשוב להדגיש את החידוש התיאורטי של המאמר: בעוד שבמחקרים קודמים אמון (trust) מוסגר במידה רבה כמרכיב התנהגותי-ארגוני המסביר אימוץ, שימוש והסתמכות על טכנולוגיה, במיוחד בהרחבות של מסגרות קבלת טכנולוגיה ובהקשרים של מסחר אלקטרוני, המוקד היה לרוב בזיהוי גורמים המסבירים כוונות שימוש, תפיסות משתמשים ותוצאות התנהגותיות (Gefen, 2000; Gefen et al., 2000, 2003). לעומת זאת, במאמר הנוכחי אמון/אי-ודאות אינם ממוסגרים רק כעמדה תפיסתית של משתמש כלפי מערכת, אלא כמנגנון ניהולי-תפעולי דינמי בתוך תהליך קבלת החלטות, שניתן למדוד, לכייל ולעדכן בזמן אמת. במובן זה, התרומה התיאורטית שלנו היא בהעברת מושג האמון מרמת ההסבר ההתנהגותי לרמת מנגנון בקרה ארגוני; והתרומה הטכנית היא בהצעת מסגרת חישובית שבה רמת האמון משפיעה בפועל על מדיניות פעולה (למשל ספים, בקרה, הסלמה או דחיית החלטה), תוך עדכון מקוון על בסיס אינדיקציות פנימיות של המערכת.

תרומתנו מבוססת על גישה היברידית: לשמר ליבה בייסיאנית שקופה וברת ביקורת שמייצרת אומדן יחד עם רמת אמון, ולצידה רכיב למידת מכונה שמכוונן בזמן אמת את "רמת האמון" בהתאם להקשר התפעולי, על בסיס אינדיקציות פנימיות שהמערכת מפיקה ממילא. בעולמות אמידת מצב אנו מדגימים זאת באמצעות התאמה מקוונת של פרמטרי אי-ודאות במסנן לא ליניארי, כך שהמערכת מגיבה אחרת כשהחיישן משתבש לעומת כשהדינמיקה משתנה, ומקטינה תגובת יתר לחריגות לצד התאוששות מהירה יותר (Tobaly et al., 2025).

בעולמות תמיכה בהחלטות ארגוניות אנו מדגימים כיצד כיול הסתברויות, מדיניות ספים ובקרות מפחיתים ביטחון יתר ומאפשרים תיעדוף משאבים "לפי אמון" במקום לפי תחושת בטן (Tobaly & Zalevsky, 2025).

המסר לקורא הניהולי הוא פשוט: כאשר הופכים את האמון למשתנה מנוהל, עם מדדים, גבולות, תיעוד ומנגנוני בלימה, אפשר להרחיב אוטומציה היכן שהיא בטוחה, ולהדק בקרה היכן שהסיכון גבוה. במילים אחרות, "אי-ודאות נלמדת" מאפשרת לארגון לעבור מניהול אינטואיטיבי של

ספים לניהול מדיניות: להגדיר מראש מהו מחיר טעות בכל תחום (הפסד כספי, חיכוך לקוח, חשיפה רגולטורית), לקבוע רמות פעולה מדורגות (אוטומטי/בדיקה/אישור), ולמדוד באופן רציף האם רמת האמון שהמערכת מציגה באמת תואמת את הסיכון בפועל. כך למשל ניתן להוריד עומס במוקדים, לצמצם חקירות הונאה מיותרות, לייצב החלטות מלאי, ולשפר תגובה לאירועי סייבר, לא באמצעות "עוד מודל", אלא באמצעות משמעת ניהולית סביב אמינות.

## רקע תיאורטי

במבט ניהולי, ארגון הוא מערכת שמקבלת החלטות תחת מגבלות מידע. חלק מהמידע מגיע מאנשים (מוקד שירות, אנשי שטח, מנהלים), חלק ממערכות (ERP/CRM, רשת, חיישנים תפעוליים), וחלק ממודלים אנליטיים ובינה מלאכותית. מה שמשותף לכל ההקשרים הללו הוא שהחלטה מתקבלת לעיתים קרובות לפני שהאמת מתבררת: האם לאשר עסקה, האם לחסום פעולה חשודה, האם להזרים מלאי, האם לתעדף ליד, האם להסלים אירוע לצוות מומחה.

כאן נכנסת נקודת הכשל הנפוצה ביותר בהטמעת AI בארגון: לא בהכרח שהמודל "טועה", אלא שהוא בטוח מדי כשהוא טועה, או לא בטוח כשאפשר לפעול. שתי התקלות הללו אינן "תיאורטיות", הן מתורגמות לעלות: התראות שווא שמעמיסות צוותים, חסימות יתר שפוגעות בהכנסות ובחוויה, או איטיות ושמרנות שמייצרות פספוס אירועים והפסדים. הבעיה מחמירה בתנאי אמת, משום שהסביבה אינה סטטית: בסייבר, דפוסי תעבורה משתנים, עדכונים המוניים מייצרים רעש, ומקור מידע אחד עלול "להרעיש" את המערכת. בשרשרת אספקה, זמני אספקה, ספקים וביקושים משתנים, ולעיתים מידע מגיע באיחור או חלקית. במוקד שירות, תיוגים לא עקביים, עומסים חריגים ואירועי תקלה משנים את התפלגות הפניות.

במצבים כאלה, מדיניות אמון קבועה (ספים קבועים, משקלים קבועים למקורות מידע, או הנחות רעש קבועות) מייצרת פער מבני מול המציאות. המאמר מציע מסגרת היברידית שמטרתה להקטין את הפער הזה: לשמר מנגנון הסתברותי שקוף שניתן לבקרה (כך שהארגון יודע "למה" התקבלה החלטה), ובמקביל להוסיף רכיב בינה מלאכותית שמעדכן בזמן אמת את רמת האמון במודל ובנתונים, לא כדי "לנצח" את המודל הישן, אלא כדי להפוך את האמון למדיד, מנוהל ומבוקר.

## שאלת המחקר והתרומה המרכזית

שאלת המחקר המאחדת היא :כיצד ניתן להפוך את האמון של מערכת החלטה, הביטחון שלה בעצמה ובנתונים, למתואם הקשר ובזמן אמת, כך שהארגון יקבל החלטות טובות יותר תחת שינויי סביבה, מבלי לוותר על שקיפות, יכולת ביקורת והטמעה במערכות קיימות?

התרומה המרכזית היא מסגרת "למידה מסייעת אמון": הליבה ההסתברותית נשמרת כמנגנון שקוף שמייצר החלטות ומדדי אמינות. רכיב ה-AI מעדכן בזמן אמת פרמטרי אמון על בסיס אינדיקציות פנימיות (פערים, עקביות, חריגות). הכול מתבצע תחת מגבלות שמונעות תנודתיות ומקטינות סיכון (החלקה, מגבלות קצב שינוי, בלימה).

ייחוד התרומה אינו רק ב"שיפור דיוק", אלא בהפיכת האמון לגודל שניתן לנהל:

**בעולמות תפעוליים**: אמון עקבי שמפחית התראות שווא ומשפר התאוששות מתקלות מידע.

**בעולמות החלטה עסקיים**: הסתברויות מכוילות שמאפשרות מדיניות סף, תיעדוף משאבים וניהול סיכון.

## עקרונות מתודולוגיים מאחדים ל AI-אמין בארגון

כדי שמערכת AI תעבוד בפרודקשן לאורך זמן, היא צריכה להפסיק להתנהג כמו "מודל סטטיסטי שמחזיר מספר" ולהתחיל להתנהג כמו מערכת ניהול סיכונים ארגונית. ברגע שמכניסים AI לתהליך אמיתי, במוקד שירות, ב-SOC של סייבר, במניעת הונאות, בשרשרת אספקה או במיון לידים, הארגון לא שואל רק "כמה זה מדויק?", אלא "מתי מותר לי לסמוך על זה, ומתי אני חייב לעצור, להסלים או לבקש עוד מידע?". כאן נולדת ההבחנה הקריטית: דיוק ממוצע יכול להיראות מצוין בדו"ח,

ועדיין להיכשל בדיוק ביום שבו הכי צריך את המערכת. לכן, העיקרון המאחד הוא שמדד הביטחון של המערכת הוא מטבע ניהולי, אבל רק אם הוא אמין, מכויל, ומתעדכן לפי הקשר, ולא לפי "מספר יפה" שאינו מחובר למציאות.

בשטח, זה תמיד מתחיל בסיטואציה שמנהלים מכירים: משהו משתנה. בסייבר, למשל, מגיע עדכון המוני למערכות קצה, ופתאום יש גל התראות. התרחיש הזה קלאסי: אם המערכת מתייחסת לכל האותות כאילו העולם יציב, היא תתרגם את השינוי להצפה של התרעות שווא, תשרוף את זמן האנליסטים ותייצר "עייפות התראות". אבל אם המערכת מבינה שהתפקיד שלה אינו להחליף את ההסקה אלא לכוונן אמון באופן מבוקר, היא יכולה לזהות סימנים פנימיים לחריגה, פערים שגדלו, עקביות שנשברה, מקור מידע שנהיה "רועש" ולשנות זמנית את מדיניות האמון: להפחית את המשקל של מקור הנתונים הבעייתי, להעלות סף פעולה, או להעביר יותר מקרים למסלול אנושי. וכשהעקיבות חוזרת לנורמה, האמון חוזר בהדרגה, לא בקפיצה. זה נשמע טכני, אבל המשמעות ניהולית לגמרי: במקום להציף אנשים, אתה מייצר מנגנון שמווסת עומס וסיכון בזמן אמת. ובדיוק כדי שזה לא יהפוך ל"הסתגלות אגרסיבית" שמשנה הכל בלי שליטה, מוסיפים בלמים פשוטים וניהוליים: החלקת עדכונים, מגבלות קצב שינוי, ומנגנון "חזרה לברירת מחדל" כשיש חריגה חריפה במיוחד.

אותו עיקרון עובד גם בהונאה, רק שהמחיר הניהולי מתהפך. בתקופות חגים, התנהגות לגיטימית נראית פתאום חריגה: יותר עסקאות, שעות לא שגרתיות, דפוסים שונים. מודל שלא יודע לעדכן אמון לפי הקשר ייטה לחסום יותר מדי, ואז המחיר הוא אובדן הכנסות, פגיעה בלקוח, וערוץ שירות שמתמלא בתלונות. עדכון אמון דינמי מאפשר להימנע מחסימות יתר בלי לפתוח את השער להפסדי הונאה: כשמדדי הקוהרנטיות מאותתים "זה שינוי עונתי צפוי ולא מתקפה", הארגון מפעיל מדיניות שמרנית אחרת, לא ביטול ההגנה, אלא התאמה של ספים ומסלולי טיפול.

אחרי שמבינים שהנושא הוא אמון לפי הקשר, מגיעה הבעיה השנייה שמכריעה מערכות בפרודקשן: הערכה לא מציאותית. כמעט כל ארגון יכול להראות KPI יפה אם הוא בוחן את המודל בתנאים נקיים, אבל הפרודקשן הוא לא מעבדה. במוקד שירות, למשל, ביום תקלה יש עומס חריג, תיוגים לא עקביים, שינוי בשפה של הלקוחות, וגם שינוי בהתנהגות של הנציגים. אם המודל נבחן על "ממוצע" שמערבב שגרה עם חריג, הוא נראה טוב, ואז נכשל בדיוק בזמן משבר, כשהוא אמור להציל את המערכת. לכן משמעת נתונים היא לא שלב טכני אלא החלטה ניהולית: מפרידים בין תרחישים, בונים סגמנטציה של שגרה מול חריג, מזהים חריגים, מנקים נתונים, ומוודאים שהתרחישים שנבדקים באמת מייצגים את היום שבו הכל "לא עובד חלק". רק אז אפשר לטעון שה-AI מוסיף תועלת אמיתית ולא אשליה סטטיסטית.

ומכאן קל לראות למה ניהול אמון לפי סגמנטים הוא תנאי ל-AI אמין. בשרשרת אספקה, קטגוריות שונות מתנהגות אחרת לגמרי: מוצר עונתי, מוצר עם זמן אספקה גבוה, ומוצר עם ביקוש יציב, לא יכולים לשבת תחת אותה מדיניות. אם מיישמים סף אחד לכולם או הסתברות אחת שמפורשת באותה דרך, מקבלים עודפים במוצרים מסוימים וחוסרים באחרים. לכן בונים מדיניות אמון מותאמת-קטגוריה, ושומרים מאגר ניסיון שמכיל גם מקרי קצה: עיכובים חריגים, שיבושים, שינויי ספק. זה ההבדל בין "מודל שמנבא" לבין מערכת שמנהלת סיכון תפעולי.

אבל כדי שספים באמת יעבדו, ההסתברויות חייבות להיות מכוילות. בארגונים קל להגיד "אם זה 0.8 נעשה אוטומציה", אבל המספר הזה שווה משהו רק אם 0.8 מתנהג לאורך זמן כמו "אמון גבוה" ולא כמו מספר שמשתנה במשמעותו בין חודשים. בתיעדוף לידים, אם 0.8 לא יציב, אי אפשר לבנות מדיניות: מה עולה לנציג בכיר, ומה נשאר אוטומטי. באשראי ובסיכון, אותו חוסר כיול הופך מהר מאוד לחוסר יציבות ניהולי: אותה "הסתברות" תוביל בחודש אחד לאישור ובחודש אחר לסירוב, מה שיוצר תנודתיות, רעש בתהליכים, ואובדן אמון במערכת. כיול הוא השלב שבו AI מפסיק להיות "ציון" והופך להיות כלי החלטה.

וכאן מגיעה נקודת השבר הריאלית ביותר: נתונים חסרים ומקורות לא אמינים. במציאות, מקור נתונים יכול ליפול, להיות חלקי, או להפוך רועש. מערכת AI אמינה לא קורסת במצב כזה, ולא ממשיכה כרגיל כאילו כלום לא קרה. היא מתרגמת

חוסר נתונים לשינוי במדיניות אמון. בהונאה, אם מקור IP/location נופל, המערכת לא עוברת מיד לחסימה גורפת (פגיעה עסקית) וגם לא מתעלמת (סיכון). היא עוברת למדיניות "אימות נוסף" או למסלול טיפול אחר. במוקד שירות, אם התיוגים משתבשים, המערכת מורידה את רמת האוטומציה ומעלה Man-in-the-loop עד שהמידע מתייצב.

הארגון צריך זרימת עבודה ברורה שאפשר להסביר במילה אחת: תפעול. מפיקים תחזית או המלצה, מודדים פערים ועקביות, מעדכנים אמון באופן מבוקר, ואז מחליטים, אוטומציה, הסלמה, או דחייה עד לקבלת מידע נוסף. ב-SOC/NOC זו הדרך לתרגם AI למדיניות: מה נסגר אוטומטית, מה עולה לצוות, ומה נכנס להשהיה. זה מגדיר גבולות תקפות, מפחית רעש, ומייצר אחריותיות אמיתית.

ולפני שמגיעים לפריסה, חייבים להוכיח שהתרומה אמינה ותישאר אמינה. כאן נכנסים שני כלים שמנהלים מבינים היטב: Stress-tests ו- Ablation. Stress-tests שואלים מה קורה ביום שבו מקור נתונים נופל, ביום חריג, או כשיש שינוי התפלגות. Ablation שואל שאלה ניהולית עוד יותר חדה: מה תרומת כל רכיב? האם השיפור בא מהתאמת האמון או מ"טריק מדידה"? ומכאן מגיע עיקרון קריטי נוסף: מניעת דליפות באמצעות פיצול זמן וקבוצות, כדי לא לייצר אשליית ביצועים שמתפרקת בשטח. זה למעשה תרגום ישיר של AI Governance לשפת ביצוע: ארגון שלא עושה את זה, מסתכן בהטמעה שמצליחה בפיילוט ונכשלת במציאות.

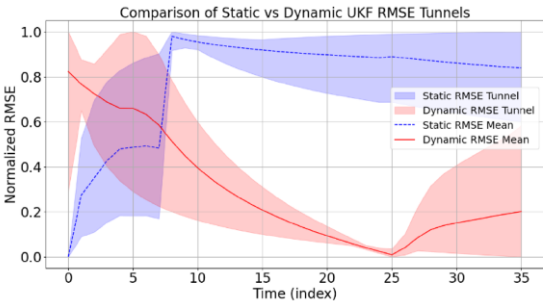
ואז אפשר לתרגם את הכל לשאלה הפרקטית שהכי מעניינת: **מה מנהל עושה מחר בבוקר?**. מנהל שמטמיע AI אמין פועל כמו מנהל סיכונים: הוא מגדיר מדיניות ספים (אוטומציה/הסלמה/דחייה) הנשענת על ביטחון מכויל ולא על מספר שרירותי; הוא קובע כללי תפעול לחריגות, נפילת דאטה, מקור רועש, אירועי קצה, כולל מצב fallback ברור; הוא מיישם ניטור שוטף של אמינות ולא רק של דיוק, כלומר התראות שווא, החמצות, יציבות ספים לאורך זמן והבדלים בין סגמנטים; ולבסוף הוא מגדיר תהליך עדכון מבוקר, שינויים הדרגתיים עם בלמים, ומבצע Stress-tests תקופתיים כדי לוודא שהמערכת נשארת בטוחה גם כשהמציאות משתנה. כך AI מפסיק להיות "מודל שמרשים בדו"ח" והופך להיות רכיב תפעולי יציב שאפשר לסמוך עליו לאורך זמן.

## ממצאים מרכזיים

### "ניווט אווירי" כמשל תפעולי: שיפור ביצועים דווקא כשיש שינוי, רעש וחריגות

הממצאים מהדגמת הניווט מציגים תופעה ניהולית מוכרת: מערכות מתפקדות טוב בשגרה, אך נבחנות באמת במעבר בין מצבים, כשיש שינוי דפוס, ירידה באיכות המידע, או אירוע קצה. כאשר פרמטרי האמון "קפואים" (כיול סטטי), המערכת נוטה ליפול לאחת משתי קצוות: תגובה חזקה מדי למידע שגוי (ביטחון יתר), או תגובה חלשה מדי לשינוי אמיתי (ביטחון חסר).

לעומת זאת, כאשר האמון מתעדכן בזמן אמת בהתאם לאותות פנימיים של עקביות (פער חיזוי–תצפית, זיהוי חריגות), מתקבלת התנהגות ארגונית רצויה: המערכת "משנה הילוך" עם ההקשר. כשמדידות פחות אמינות, היא מפחיתה את משקלן; כשהדינמיקה חורגת מהמודל, היא מגדילה זמנית את אי הוודאות התהליכית. התוצאה היא פחות סטיות מצטברות, יציבות גבוהה יותר, והתאוששות מהירה יותר לאחר חלונות קצרי טווח של מידע פגום. כפי שניתן לראות בתרשים 1, העדכון הדינמי מצמצם לאורך זמן הן את ה RMSE-הממוצע והן את רוחב "מנהרת" השגיאה, ובכך משקף אמינות תפעולית גבוהה יותר לעומת כיול סטטי.




**תרשים 1**. השוואת מנהרות RMSE מנורמלות בין UKF סטטי ל UKF-דינמי לאורך זמן.

**תרגום עסקי (סייבר/הונאה/מוקד שירות/שרשרת אספקה):**

**סייבר:** בזמן "רעש" (עדכונים המוניים, סריקות, תקלה בסנסור), מערכת אמון סטטית מציפה התראות שווא או מפספסת התקפה אמיתית. מנגנון אמון דינמי מפחית זמנית הסתמכות על מקור בעייתי ומונע עומס יתר על האנליסטים.

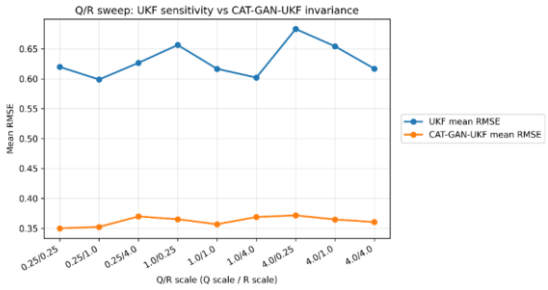
**הונאה:** בתקופות חריגות (חגים, קמפיינים), התנהגות לגיטימית נראית חשודה. עדכון אמון לפי עקביות מאפשר להפחית חסימות יתר בלי להוריד את רמת ההגנה.

**מוקד שירות:** באירוע תקלה יש שינוי חד בהתפלגות הפניות. עדכון אמון דינמי מפחית החלטות אוטומטיות שגויות ומעלה Man-in-the-loop עד שהמערכת חוזרת ליציבות.

**שרשרת אספקה:** כאשר ספק "מזייף" זמני אספקה או מידע מתעכב, עדכון אמון במקור הנתונים מפחית החלטות מלאי שגויות (עודף/חוסר).

**למה "מדיניות אחת לכולם" נכשלת ?**

הממצא השני מדגיש נקודה ניהולית קריטית: הטרוגניות. כשיש קטגוריות שונות (לקוחות שונים, מוצרים שונים, אזורים שונים, מקורות מידע שונים), מדיניות אמון גלובלית אחת עלולה לשפר ממוצע אך לפגוע היכן שהעלות גבוהה, בזנבות ההתפלגות ובמקרי הקצה. המבנה המודע קטגוריה, שבו לכל "סגמנט" יש מדיניות אמון "יעודית, מצמצם העברה שלילית: שיפור במוצר/קטגוריה אחת אינו "נקנה" בירידה באחרת. כפי שממחיש תרשים 2, הUKF-מציג תנודתיות מהותית בביצועים לאורך סריקת Q/R (רגישות לאתחול), בעוד שהגישה המודעת קטגוריה שומרת על יציבות, כלומר מדיניות אמון עקבית יותר בין סגמנטים ותנאים.




**תרשים 2**. סריקת Q/R המדגימה רגישות ביצועי UKF לאתחול לעומת יציבות של CAT-GAN-UKF לאורך קני מידה שונים.

**תרגום עסקי:**

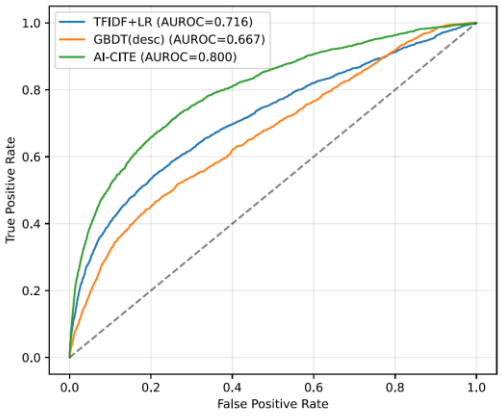
**הונאה :**קטגוריות תשלום שונות (כרטיס, ארנק דיגיטלי) דורשות מדיניות אמון שונה; אחרת תחסום יתר קטגוריה אחת ותחשוף אחרת.

**שרשרת אספקה :**מוצרים עונתיים אינם מתנהגים כמו "הממוצע". מדיניות לפי קטגוריות מצמצמת פגיעה במוצרים קריטיים/רגישים.

**מוקד שירות :**סגמנטים של לקוחות צריכים ספים שונים, כי עלות הטעות עבור כל לקוח היא שונה.

**חיזוי השפעה מדעית כמשל לארגון**

ביישום של חיזוי השפעה, הערך נמדד פחות ב"עוד נקודות דיוק" ויותר ביכולת להפיק מדד אמון אמין בזמן שהאמת תתברר מאוחר. זהו מצב טיפוסי בארגון: החלטה מתקבלת היום, בעוד מדד התוצאה (הכנסה, נטישה, אירוע סיכון) יתברר שבועות או חודשים קדימה. מכאן שהשאלה אינה רק "מה המודל חושב", אלא "האם אפשר לסמוך על הביטחון שלו כדי להפעיל מדיניות סף". כפי שמוצג בתרשים 3, מודל AI-CITE משיג עקומת ROC גבוהה יותר (AUROC≈0.80) לעומת קווי הבסיס, ובכך מחזק את ההצדקה הניהולית להשתמש בו כבסיס למדיניות סף, כל עוד הביטחון שהוא מפיק נשמר מכויל ועקבי לאורך זמן.




**תרשים 3**. עקומות ROC על סט הבדיקה H=3 להשוואת מודלי הבסיס מול AI-CITE; הקו המקווקו מייצג קו בסיס אקראי.

### אמינות, רגישות והחלטות סף: הממצא החוצה יישומים

הממצא העקבי בכל היישומים הוא שדיוק ממוצע לבדו אינו מספיק. הערך העסקי נוצר כאשר יש אמון שניתן לתרגם למדיניות פעולה: אוטומציה מלאה כאשר האמון גבוה, הסלמה לבקרה אנושית כאשר האמון בינוני/לא יציב, דחייה או איסוף מידע נוסף כאשר האמון נמוך.

ביישום הניווט הדבר התבטא בכך שהמערכת נשארה "מיושרת" עם המציאות גם בפרקי זמן חריגים. ביישום הרב־סגמנטלי נראתה ירידה באירועי קצה שבהם סגמנט מסוים קיבל אמון לא תואם. וביישום ההסתברותי השיפור הופיע בצורה שמנהל מבין מיד: סף פעולה הפך להיות עקבי יותר לאורך זמן ובהקשרים שונים.

במקביל, ניתוחי רגישות מדגישים נקודת ממשל: התאמה אגרסיבית מדי עלולה לייצר תנודתיות, כלומר מערכת שמחליפה החלטות במהירות, מבלבלת צוותים ומייצרת "רעש החלטתי". לכן מגבלות קצב שינוי, החלקת עדכונים ומנגנוני בלימה אינם תוספת אקדמית; הם רכיבי בטיחות ארגוניים שמאפשרים להפעיל התאמה בזמן אמת בלי לאבד שליטה.

### "אי-ודאות נלמדת" כמתודולוגיה כללית: כלל האצבע הניהולי

מן הסינתזה של שלושת המחקרים עולה כלל אצבע פרקטי: כאשר החלטות מתקבלות באמצעות ספים, השאלה איננה רק "מה המודל מנבא", אלא בעיקר "האם רמת האמון שלו מתורגמת בפועל לרמת סיכון".

בדיקות ההסרה (ablation test) מראות כי התאמת רעש/אמון במקורות המידע מספקת לרוב את השיפור הראשוני הגדול ביותר, משום שהיא מטפלת ישירות בפער בין הרעש התצפיתי בפועל לבין רעש המודל. יחד עם זאת, התאמות עדינות יותר במדיניות הקירוב והיציבות מוסיפות ערך כשיש אי ליניאריות גבוהה או אזורי קצה שבהם המערכת רגישה במיוחד.

במילים ניהוליות: יש כאן שילוב בין "כיוון מדיניות אמון" לבין "הנדסת בטיחות" שמונעת מהאמון להפוך לתנודתי.

## דיון

שלושת המחקרים מתכנסים לטענה אחת: במערכות תפעוליות, פיזיקליות או ארגוניות, אי-ודאות אינה "רעש רקע", אלא רכיב מבני שמכתיב ביצועים, יציבות ויכולת קבלת החלטות. כאשר הסביבה חורגת מהנחות סטטיות (עומסים משתנים, מקורות מידע שונים, דפוסים חדשים), כיול ידני וקבוע של אמון הופך לנקודת תורפה מרכזית ומוביל לשני מצבי כשל סימטריים: ביטחון יתר או ביטחון חסר. בשני המקרים המחיר תפעולי: התראות שווא, פספוס אירועים, והעמסת יתר על בקרה אנושית.

המסגרת המוצעת מתייחסת לכך כמו לתהליך ניהולי: שומרים על ליבה שקופה שמאפשרת ביקורת והסבר. מוסיפים מנגנון למידה שמעדכן אמון לפי אותות פנימיים. מפעילים מעטפת בקרות שמבטיחה חוקיות ויציבות: מגבלות שינוי, החלקה, בלימה וחזרה לערכי בסיס בעת חריגה.

ביישום הניווט הודגם היתרון הבסיסי: התאמת אמון בזמן אמת מאפשרת למערכת "להחליף מצב" עם ההקשר ולשמור על התאוששות מהירה מאירועים קצרים של מידע פגום, תרגום ישיר להפחתת רעש החלטתי. ביישום הרב־סגמנטלי הודגם מימד קריטי להטמעה ארגונית: ניהול לפי ממוצע אינו מספיק; חייבים ניהול לפי סגמנטים

ותרחישי קיצון כדי להימנע מפגיעה בקטגוריות יקרות. וביישום של חיזוי השפעה, הודגש הערך ההחלטי: הסתברות מכוילת היא תנאי להחלטות סף אחראיות, היא הופכת ל"מטבע החלטתי" ולא למספר שמייצר ביטחון יתר.

המסקנה המאחדת היא מעבר מהיגיון של "שיפור דיוק" להיגיון של יכולת פעולה: מערכת טובה היא זו שמייצרת אמינות מדידה שממופה לסיכון בפועל ומאפשרת מדיניות פעולה עקבית.

## סיכום

המחקר הנוכחי יוצא מנקודת מוצא שמנהלים פוגשים שוב ושוב בתפעול היומיומי: רוב ההחלטות הארגוניות מתקבלות לא כשהמידע "מושלם", אלא דווקא כשיש חלקיות, רעש, סתירות ושינוי מתמשך. בין אם מדובר באירוע סייבר שמתפתח תוך דקות, בקמפיין שיווק שמייצר התנהגות לקוחות חריגה, בתקלה בשרשרת אספקה, או בעומס פתאומי במוקד שירות, הארגון נדרש לפעול בזמן אמת. במצבים האלה, השאלה הקריטית אינה רק "מה המערכת מנבאת", אלא עד כמה אפשר לסמוך על מדד הביטחון שלה כדי להפעיל אוטומציה, להעביר לבקרה אנושית, או לעכב החלטה עד שיושג מידע נוסף.

כאן נמצא הכשל הנפוץ של מערכות AI ומודלים תפעוליים: הן עשויות להיראות מצוינות בממוצע, אך להיות מסוכנות או לא יעילות בדיוק במקומות היקרים ביותר, במעברים בין מצבי פעולה, באירועי קצה, ובסגמנטים שבהם עלות טעות גבוהה. כאשר מדד הביטחון של המודל אינו מותאם לתנאי השטח, מתקבלות שתי תקלות סימטריות עם מחיר עסקי ברור:

**ביטחון יתר** שמייצר החלטות אוטומטיות שגויות, תגובת יתר והתראות שווא.

**ביטחון חסר** שמייצר שמרנות, איטיות, עודף "יד אדם" ובסוף פספוס אירועים או הזדמנויות.

התרומה המרכזית של קו המחקר היא הצעה למסגרת שמטפלת בדיוק בנקודה הזאת: להפוך את האמון/אי הוודאות מ"כיול חד־פעמי" למשאב נלמד ומבוקר. במקום להחליף את המנגנון ההסתברותי בקופסה שחורה, נשמרת ליבה מתמטית שקופה וברת ביקורת **UKF**, כשלד רקורסיבי ובמקביל נוסף רכיב למידת מכונה שמעדכן בזמן אמת את רמת האמון במודל ובנתונים בהתאם להקשר התפעולי. העדכון אינו נשען על היוריסטיקה אד־הוק של המערכת, אלא על אותות פנימיים שהמערכת ממילא מפיקה בכל צעד: פער חיזוי–תצפית, עקביות השגיאות, ומדדי חריגה.

הנקודה הניהולית החשובה היא שהמסגרת אינה מציגה רק "עוד שיפור דיוק", אלא שיפור ביכולת פעולה. כך מתקבל מנגנון שמנהל יכול "לנהל": אמון הופך לישות מדידה שניתן להציב לה ספים, לקשור אותה לעלות טעות, ולתרגם אותה לכללי החלטה ברורים.

הממצאים בשלושת היישומים מחזקים את הטענה הזו מזוויות משלימות. המחקר מראה שהטמעה של אמון דינמי אינה יכולה להיות "חופשית" וחסרת מגבלות ואיזונים. כמו בכל מנגנון אדפטיבי, עדכון אגרסיבי מדי עלול לייצר תנודתיות, חוסר יציבות, ושינויי החלטה מהירים שמבלבלים ארגון ומגדילים סיכון. לכן אחד המסרים המרכזיים של העבודה הוא שהטמעה אחראית של AI דורשת מעטפת הגנה: אילוצי חוקיות (למשל שמירה על מטריצות תקינות), החלקה, מגבלות קצב שינוי, ומנגנוני בלימה וחזרה לערכי בסיס בעת חריגה.

אלו אינם פרטים טכניים בלבד, הם המקבילה הארגונית לבלמים, בקרות ומדיניות אסקלציה במערכות תפעוליות. כלומר :המודל לומד, אבל הארגון שומר על שליטה.

במבט מסכם, התרומה של קו המחקר היא פרדיגמה יישומית לניהול סיכון בזמן אמת: להפוך את מדד הביטחון של המערכת למשתנה שאפשר למדוד, לבקר, ולחבר לעלות. במקום ש AI יהיה "ממליץ" שאי אפשר לדעת מתי לסמוך עליו, מתקבל מנגנון שמאפשר לקבוע מדיניות פעולה מפורשת: מתי לאשר אוטומטית, מתי להסלים לאדם, מתי לדחות החלטה, ומתי להפעיל איסוף מידע נוסף. כך נוצרת יכולת לפתח מערכות תומכות החלטה שמאזנות נכון בין יעילות לבין אחריות, לא כוויתור על ביצועים, אלא כדרך להשיג ביצועים שאפשר להפעיל בבטחה לאורך זמן.

לבסוף, יש כאן מסקנה ניהולית רחבה: בארגון מודרני, אי-ודאות אינה "בעיה טכנית שמטפלים בה פעם אחת". היא דפוס קבוע של המציאות. לכן היתרון התחרותי אינו רק במודלים חכמים יותר, אלא ביכולת לנהל אמון, להפוך את אי הוודאות ממקור סיכון לא מבוקר למנגנון שמייצר תיעדוף, הקצאת משאבים, ויציבות החלטית. במובן הזה, "אי-ודאות נלמדת" אינה רק שיפור אלגוריתמי, אלא כלי ניהולי שמחבר בין דיוק לבין אחריות, ובין אוטומציה לבין בקרה ארגונית.

## רשימה ביבליוגרפית


Birkstedt, T., Kaur, S., & Shell, K. (2023). AI Governance: Themes, Knowledge Gaps and Future Agendas. *Internet Research*, *33*(7), 2612–2641. https://doi.org/10.1108/INTR-01-2022-0042

Gefen, D. (2000). E-commerce: The role of familiarity and trust. *Omega*, *28*(6), 725–737. https://doi.org/10.1016/S0305-0483(00)00021-9

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in Online Shopping: An Integrated Model. *MIS Quarterly*, *27*(1), 51–90. https://doi.org/10.2307/30036519

Gefen, D., Straub, D. W., & Boudreau, M.-C. (2000). Structural Equation Modeling and Regression: Guidelines for Research Practice. *Communications of the Association for Information Systems*, *4*(1), 1–77. https://doi.org/10.17705/1CAIS.00407

International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC). (2023a). *ISO/IEC 23894:2023—Information Technology—Artificial Intelligence—Guidance on Risk Management*. https://www.iso.org/standard/77304.html

International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC). (2023b). *ISO/IEC 42001:2023—Information Technology—Artificial Intelligence—Management System*. https://www.iso.org/standard/81230.html

National Institute of Standards and Technology (NIST). (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (Issue NIST AI 100-1). https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

Office of the Privacy Commissioner for Personal Data (PCPD), Hong Kong. (2024). *Ethical Artificial Intelligence Framework*. https://www.pcpd.org.hk/english/resources_centre/publications/ai/ethical_ai_framework.html

Tobaly, L., Yaniv, E., & Zalevsky, Z. (2025). Adversarial Learning-Based Adaptive Unscented Kalman Filtering for Enhanced Nonlinear State Estimation. *Scientific Reports*, *15*, 42361. https://doi.org/10.1038/s41598-025-26339-9

Tobaly, L., & Zalevsky, Z. (2025). AI-CITE: A Decision-Oriented Framework for Early Scholarly-Influence Forecasting with Calibrated Confidence. *IEEE Access*, *13*, 217675–217691. https://doi.org/10.1109/ACCESS.2025.3646128