שער האימות: שיטת עבודה לבינה מלאכותית גנרטיבית בעבודה הנדסית שבה טעות עולה כסף או בטיחות

מסמך עבודה, גרסה 0.91 · לא עבר שיפוט עמיתים · The Verification Gate — Working paper v0.91, not peer-reviewed

אינג' מטי מלכיאור · Mati Melchior
חוקר עצמאי, בטיחות Physical AI · מהנדס מכונות בעל רישיון, mati@physical-ai-safety.com · ORCID 0009-0003-9647-1498 · MBA

ספטמבר 2026 · Concept DOI: 10.5281/zenodo.22271339 · עמוד השיטה: physical-ai-safety.com/method/verification-gate
מסמך זה: CC BY-NC-ND 4.0 · הנספחים (המכשירים): CC BY 4.0

תקציר

כלי בינה מלאכותית גנרטיבית מנסחים, מתרגמים, מסכמים ומארגנים היטב. הם נופלים בנקודות שבהן מהנדס או הנדסאי חייב לחתום. חמש נקודות כאלה חוזרות: טבלה שנקראת מתמונה, מד אנלוגי, סעיף תקן, קוד תקלה של יצרן, וחישוב אריתמטי ישיר. מחקרים שפיטים בהנדסת בניין מתעדים את הראשון ואת האחרון. תיעוד יצרנים, preprint ודיווחי שטח מתעדים את השאר. בכל מקרה הפלט שוטף וללא סימן. שום דבר בטקסט אינו מסמן את השגיאה. אנו מציגים את שער האימות, שיטת עבודה לשימוש בכלים האלה במקום שבו טעות עולה כסף או בטיחות. לשיטה אקסיומה אחת ושלושה שלבים בסדר קבוע. היא מציבה נקודת עצירה מחייבת אחת בין פלט הכלי לכל התחייבות. היא מוסיפה ארבעה מכשירים במבנה קבוע, סולם בן חמש רמות לתיוג משימות, מבחן כשירות ללא כלים, ונוהל התקנה בן חמישה צעדים. היא מגדירה שבעה סוגי כשל (G1–G7) וממפה כל אחד לבקרה שתופסת אותו. היא אינה מלמדת כלים. היא קובעת את הסדר שבו המהנדס והכלי עובדים. מאמר זה מקבע את אוצר המילים של השיטה, מגדיר את גבולותיה, ומנסח פרוטוקול הערכה שנרשם מראש. אין כאן תוצאות, ולא נאספו נתונים. הפרוטוקול מגדיר את שדותיו, את רשימותיו הסגורות, ואת גודל המדגם הנדרש. כל קבוצה שעומדת בקריטריונים שהוגדרו ומגישה יומני אימות לפי שדות אלה היא יישום תקף. גרסה 1.0 תדווח תוצאות מדודות אם וכאשר יומנים כאלה יתקיימו. כל המכשירים משוחררים ברישיון CC BY 4.0.

מילות מפתח: בינה מלאכותית גנרטיבית · אימות · עבודה הנדסית · מיזוג אוויר · הנדסת מכונות · שיטת עבודה · הזיות (hallucination) · פיקוח אנושי

1. מבוא

יחידת מיזוג מפוצלת מציגה קוד תקלה E3. טכנאי שואל כלי בינה מלאכותית גנרטיבית מה פירוש E3. טבלאות היצרנים נותנות שלוש תשובות. דייקין מציינת E3 כהפעלת מפסק לחץ גבוה (Daikin n.d.). GREE מציינת E3 כלחץ נמוך (GREE n.d.). מיצובישי אלקטריק מציינת E3 כשגיאת תקשורת של השלט (Mitsubishi Electric n.d.). השניים הראשונים הם תקלות הפוכות. טכנאי שפועל לפי הפירוש השגוי מוסיף קרר למערכת שכבר עמוסה מדי, או מרוקן מערכת שכבר חסרה. כלי שעונה מהזיכרון, בלי המדריך של היחידה הספציפית לפניו, אינו יכול להראות באיזו טבלה השתמש. תשובתו אינה נושאת שום סימן מוצא.

זו הצורה האופיינית של הנזק מבינה מלאכותית גנרטיבית בעבודה הנדסית. הכלי אינו חלש. GPT-4o עבר שני מבחני רישוי לאומיים בדרום קוריאה למהנדסי ציוד מכני בבניינים, כולל מכונות מיזוג וקירור, בכל חמשת הניסיונות (Choi, Lee, and Kim 2025). הציונים הממוצעים שלו היו 80.6 ו-81.25 מתוך 100. אותו מחקר מזהה שלוש מגבלות: הסקה מתקדמת חלשה, קושי בשאלות משפטיות, ופירוש גרוע של תרשימים מדעיים. אלה שלושה סוגי פריטים שטכנאי חותם עליהם: חישוב, דרישה רגולטורית, שרטוט.

אז הפער אינו יכולת. שום שלב בזרימת העבודה אינו עוצר את הפלט לפני החתימה. להנדסה יש כבר שלבים כאלה. חתימת בודק קודמת לשחרור שרטוט. מבחן לחץ קודם למסירה. לבינה מלאכותית גנרטיבית אין שלב מקביל, ולכן פלט יכול להיכנס לדוח חתום בלי שום בדיקה.

מסמך זה מספק את נקודת העצירה החסרה כשיטה שלמה וקבועה. אנו קוראים לה שער האימות. השיטה נכנסת למפגש הדרכה אחד. ליומן שלה 13 שדות קבועים, ולכן מבקר יכול לספור עמידה בדרישות.

תרומות המסמך.

1. אוצר מילים של אחד-עשר מונחים מוגדרים, כך ש"אומת" יאמר דבר אחד (סעיף 3).
2. סיווג של שבעה סוגי כשל, G1–G7, ייחודיים לתוצרים הנדסיים, כל אחד קשור לראיה מפורסמת (סעיף 4).
3. השיטה עצמה: אקסיומה אחת, שבעה עקרונות, שלושה שלבים, שער אחד, וכלל החלטה לשער (סעיף 5).
4. ארבעה מכשירים במבנה קבוע, בהם יומן אימות בן 13 שדות עם רשימות סגורות, משוחררים כתבניות (סעיף 6).
5. סולם תיוג משימות בן חמש רמות ונוהל התקנה ארגוני בן חמישה צעדים (סעיפים 7 ו-8).
6. פרוטוקול הערכה שנרשם מראש, עם שאלות מחקר שנקבעו לפני איסוף הנתונים (סעיף 10).

אנו לא טוענים ששבעת הסוגים ממצים, או שהשיטה עושה את הפלט נכון, ואנו אומרים זאת במפורש. הטענה שאנו מגנים עליה צרה יותר, וניתנת להפרכה. תחת השיטה, כל פריט בפלט הוא אחד משניים: מקושר למקור מוסמך, או חסום. והקישור הזה ניתן לספירה.

2. רקע

2.1 איפה הכלים נופלים במשימות הנדסיות

מקורות: שני מדדים שפיטים, מחקר אחד על מבחן רישוי לאומי, פוסטר אחד של ACM, preprint אחד, ועורך דין אחד בתחום הבנייה. תרשים 2 מציג את המספרים המרכזיים.

טבלאות בתוך תקנים. AECBench בוחן תשעה מודלי שפה גדולים על 4,800 שאלות שכתבו מהנדסים בתחומי האדריכלות, ההנדסה והבנייה (Liang et al. 2026). שתי משימות דורשות קריאת טבלה מתוך תקן בנייה. הדיוק הבסיסי היה 45.27 ו-40.65 אחוז. כשאותה טבלה סופקה כטקסט רץ שכתב מומחה, הדיוק עלה ל-98.94 ו-87.27 אחוז. המודלים ידעו את התוכן. הם נכשלו בקריאתו מתוך טבלה. המחקר מסכם שהמודלים מציגים חולשה "בפירוש ידע מטבלאות בתקני בנייה, בביצוע הסקה וחישוב מורכבים, ובהפקת מסמכים מקצועיים."

שגיאות במבחן רישוי לפי סוג. Choi, Lee, and Kim (2025) הריצו את GPT-4o חמש פעמים על שני מבחני רישוי לאומיים למהנדסים בקוריאה. הם מסווגים 172 שגיאות לשלושה סוגים. שגיאות ידע (104) מופיעות בשאלות ללא חישוב. שגיאות הסקה (48) מופיעות בשאלות חישוב. שגיאות הקשר (20) מופיעות בשאלות שדורשות קריאת תרשים או שרטוט. עקביות התשובות הממוצעת של המודל הייתה 97 אחוז. אז תשובה שגויה נשארת שגויה גם בניסיון הבא. העקביות הזאת היא ממצא בפני עצמו. שאילה חוזרת אינה חושפת את השגיאה.

שרטוטים. מחקר על תרשימי צנרת ומכשור (P&ID) במיזוג אוויר מסכם ש"יישום ישיר של מודלי שפה גדולים לדיגיטציה של P&ID הוא מאתגר מאוד" (Lu et al. 2025). תוצאות שמישות דרשו צנרת עיבוד מהונדסת, לא תמונה שהועלתה ושאלה.

מכשירים אנלוגיים. מדד לקריאת מדידה חזותית מדווח שמודלי ראייה-שפה מובילים "מתקשים בקריאת מדידות באופן כללי" (Lin et al. preprint, 2026). המחברים מייחסים את הקושי לעיגון מרחבי עדין: המודלים קוראים ספרות מודפסות ומטעים במיקום המחוג. אנו מצטטים את הדפוס ולא את המספרים, כי העבודה עדיין בשיפוט.

סימוכין מומצאים. הזיות במודלי שפה הן סוג מתועד עם סיווג משלו (Huang et al. 2025). במסמכי בנייה הן לובשות צורה ספציפית. עורך דין בתחום הבנייה מתאר "תקני רפאים": סימוני תקן שעוקבים אחר מוסכמת המספור האמיתית ואינם קיימים (Thomas 2026). Thomas מדווח שקודים כאלה עוברים את בדיקת המסמכים ומתגלים מאוחר יותר, כשמעבדת בדיקה אינה מוצאת את השיטה. העלות אז היא פקודת שינוי או תביעה.

2.2 למה סדר השימוש חשוב

ניסוי שדה עם כמעט אלף תלמידי תיכון בחן שני עיצובי כלי (Bastani et al. 2025). גישה חופשית לכלי צ'אט רגיל העלתה ציוני תרגול ב-48 אחוז. אחרי שהחוקרים הסירו את הגישה, אותם תלמידים קיבלו ציון נמוך ב-17 אחוז מתלמידים שמעולם לא קיבלו את הכלי. גרסה עם מעקות בטיחות שמנעה תשובות ישירות העלתה ציונים ב-127 אחוז, והמחברים מדווחים שהנפילה המאוחרת נמנעה ברובה. הם מייחסים את ההבדל לאופן שבו הכלי נכנס למשימה. המחקר עוסק במתמטיקה תיכונית, לא בעבודה הנדסית. אנו לוקחים אותו כמוטיבציה לעיקרון 1, וסעיף 10 מגדיר פרוטוקול לבחינת העיקרון בסביבה שלנו.

2.3 גופים מקצועיים

ASHRAE אימצה מדיניות בנושא בינה מלאכותית בפברואר 2025 ומחילה אותה על תהליך התקינה שלה (ASHRAE 2025). ל-ASTM International מדיניות דומה (ASTM 2025). שתיהן, כפי שאנו קוראים אותן, רואות בפלט AI חומר שאדם חייב לבדוק ולקחת עליו אחריות. אך אף אחת מהן אינה מפרטת שיטה למי שעובד בשטח. מסמך זה מספק אחת.

2.4 מה השיטה אינה

השיטה היא נוהל עבודה, במסורת נקודת העצירה של הבקרה ההנדסית, המופעל על מקור חדש של מספרים לא מאומתים. היא אינה מזהה הזיות, אינה מכוונת פרומפטים, ואינה מדרגת כלים.

3. הגדרות

השיטה משתמשת במונחים בטבלה 1 במשמעות הקבועה שם בלבד.

טבלה 1. הגדרות.

מונח	הגדרה
כלי	כל מערכת בינה מלאכותית גנרטיבית, טקסטואלית, חזותית או משולבת, המקבלת פנייה ומחזירה פלט. השיטה אינה נוקבת בשמות מוצרים.
פלט	כל מה שהכלי מחזיר, בכל פורמט, לפני שעבר את השער.
מקור מוסמך	מסמך שיש לו אחראי ומעמד: מדריך יצרן, תקן בנוסחו הרשמי, נוהל ארגוני מאושר, שרטוט חתום, מדידה ישירה, טבלת נתוני יצרן. כלי אינו מקור מוסמך לעולם. גם פלט של כלי שני אינו כזה.
פריט	יחידת מידע ברת-אימות: מספר, מזהה, סעיף, טענה עובדתית, פעולה מדווחת.
אימות	השוואת פריט אחד מול מקור מוסמך אחד ורישום התוצאה.

מונח	הגדרה
פער	כל הבדל בין הפלט למקור, בכל גודל.
תוצר	פלט שעבר את השער במלואו ויש לו יומן אימות מלא. רק תוצר רשאי להיחתם, להיגש, או לשמש להחלטה.
התחייבות	כל פעולה שמעבירה תוצר מטיוטה למציאות: חתימה, הגשה, הזמנת רכש, הוראת עבודה, דוח ללקוח, הזנה למערכת ארגונית.
השער	נקודת העצירה המחייבת בין פלט להתחייבות. שלב ג' בתהליך.
יומן אימות	המכשיר שמתעד את האימות. מפורט בסעיף 6.1.
רמה	דרגת השימוש המותרת בכלי למשימה נתונה, לפי סולם חמש הרמות בסעיף 7.

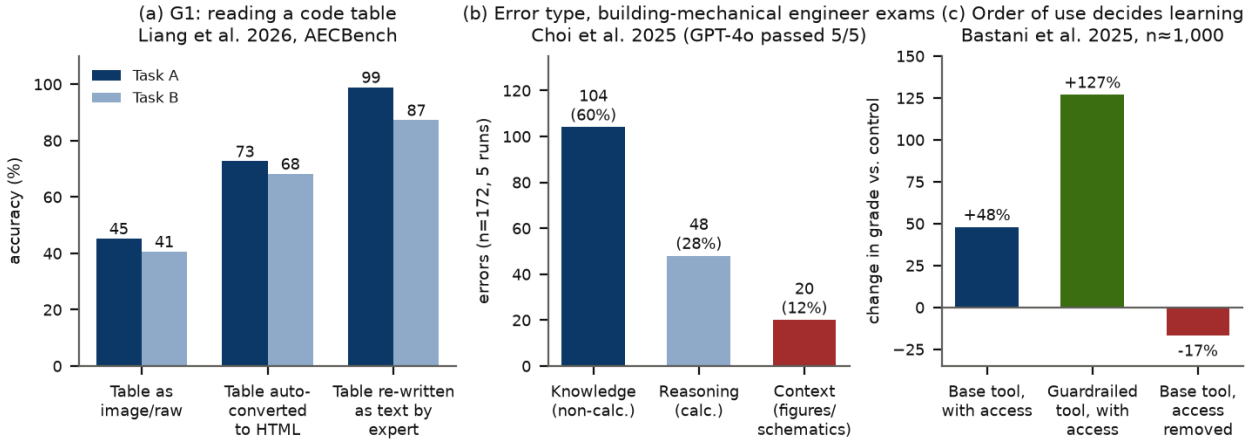
4. סיווג הכשלים

אנו מגדירים שבעה סוגי כשל. כל אחד הוא דפוס חוזר בפלט של כלים על תוצרים הנדסיים. טבלה 2 נותנת את הסוג, הדפוס, ההשלכה ההנדסית, ומעמד הראיות. אנו מתארים דפוסים, לא אחוזי דיוק, אלא כאשר מקור שפיט מדווח את המספר. אחוזים תלויים בדור הכלי ומתיישנים בתוך חודשים. האם הדפוסים יציבים יותר היא שאלה לגרסה 1.0 (סעיף 11).

טבלה 2. שבעה סוגי כשל.

#	סוג	הדפוס	למה זה חשוב בהנדסה	ראיות
G1	טבלה מתמונה	שאילתה על טבלה מצולמת מדויקת הרבה פחות מאותה שאילתה על הטבלה כטקסט	טבלות יחידות, טבלות תקן, לוחות נתונים, כולם מגיעים כתמונה	שפיט (Liang et al. 2026)
G2	מד אנלוגי	ספרות ותוויות נקראות היטב. מיקום המחוג נקרא שגוי	לחץ, טמפרטורה, זרם. הקריאה נכנסת לדוח חתום	preprint (Lin et al. 2026)
G3	סעיף תקן	מספרי סעיפים ותכנים מומצאים במבנה אמין	תדירות בדיקה חוקית או דרישת בטיחות שאינה קיימת	מקצועי (Thomas 2026); מנגנון (Huang et al. 2025)
G4	קוד תקלה	פירוש מהזיכרון. אותו קוד אומר דברים שונים, ואף הפוכים, בין יצרנים	אבחון שגוי, חלף שגוי, פעולה מסוכנת	טבלאות היצרנים מראות את הסתירה (סעיף 1). שגיאת הכלי מוסקת, טרם נמדדה
G5	חישוב ישיר	שגיאות בחישוב שמתבצע בתוך המודל, בבחירת נוסחה ובהמרת יחידות	עומס, ספיקה, מאמץ. המספר נכנס להחלטת תכן	שפיט (Choi, Lee, and Kim 2025; Liang et al. 2026)
G6	השלמה שקטה	מק"ט, תאריך, דגם או פעולה שלא נמסרו מתמלאים בביטחון מלא	דוח מדווח על פעולה שלא בוצעה	מנגנון (Huang et al. 2025)
G7	הסבר נכון-שגוי	תוצאה נכונה עם הנמקה שגויה	המהנדס סופג הנמקה שגויה שנראית תקינה	תצפית שדה. טרם נמדד

לשבעת הסוגים תכונה משותפת אחת. בכל מקרה הפלט אינו נושא שום סימן חיצוני לשגיאה. לכן אימות אינו יכול להישען על חשד. הוא רץ על כל פריט ומשאיר רשומה.



תרשים 2. ראיות מפורסמות מאחורי שלושה סוגי כשל. (א) G1: דיוק בשתי משימות קריאת טבלה מתוך תקני בנייה, לפי פורמט הטבלה, תשעה מודלים (Liang et al. 2026). (ב) סוג השגיאה בשני מבחני רישוי לאומיים למהנדסי ציוד מכני בבניינים, שהמודל עבר 5 מתוך 5 פעמים (Choi, Lee, and Kim 2025). כל סוג שגיאה ממופה לסוג שאלה. (ג) השפעת עיצוב הכלי והסרת הכלי על ציוני לומדים (Bastani et al. 2025). הערכים כפי שדווחו בתקצירי המקורות.

5. השיטה

5.1 האקסיומה

> **מספר שלא אומת אינו קיים.**

משמעות תפעולית: לערך שמקורו בפלט כלי אין מעמד של נתון עד שמישהו בודק אותו מול מקור מוסמך. עד שנבדק, הערך הוא טיוטה. הוא אינו רשאי להיכנס למסמך חתום או מוגש, או להחלטה. סוגי הפריטים המכוסים הם אלה של טבלה 4, שדה 6.

כל כלל בסעיפים 5 עד 8 נגזר מהאקסיומה הזאת.

5.2 עקרונות

ע'1 סדר קבוע. שלושת השלבים מבוצעים בסדר שנקבע, בלי דילוג. אומדן שנעשה אחרי שראית את התשובה מעוגן אליה. לכן שלב א' קודם לשלב ב', בלי חריגים (מוטיבציה: Bastani et al. 2025).

ע'2 הכלי מנסח. הוא לא מחשב. לכל חישוב, הכלי רשאי להציע נוסחה ולזהות משתנים. האריתמטיקה רצה בגיליון אלקטרוני או בכלי חישוב דטרמיניסטי, לא בתוך המודל.

ע'3 המקור קודם לזיכרון. לכל פריט נורמטיבי, המהנדס מצרף את המקור המוסמך לפנייה או בודק אותו אחריה. פירוש מזיכרון הכלי אינו קביל, גם כשהוא נכון.

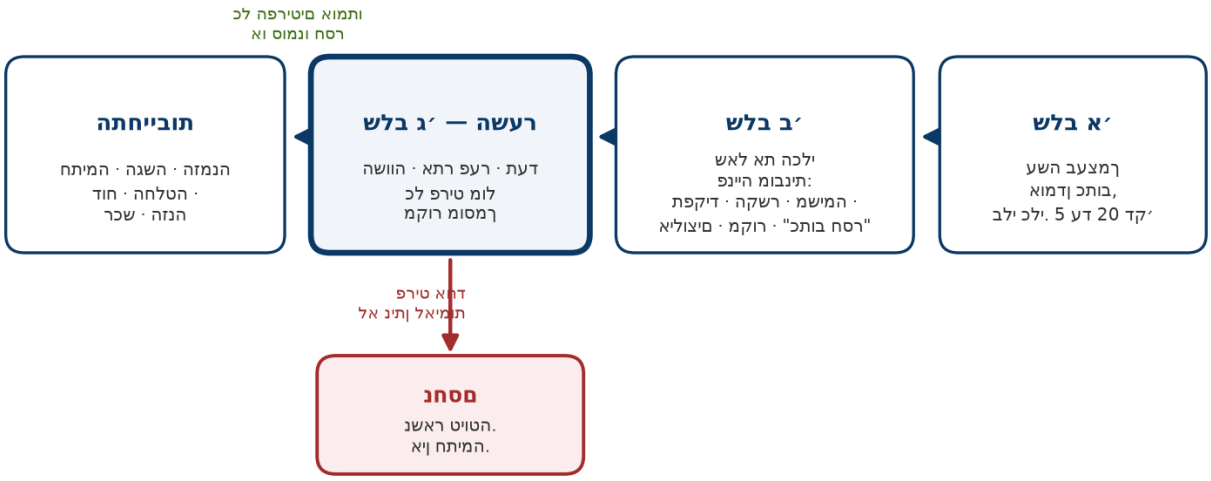
ע'4 הכרז על החסר. כל פנייה כוללת את ההנחיה: "אל תמציא. כתוב חסר." פריט שסומן חסר הוא תוצאה נכונה. פריט שהושלם בלי מקור הוא כשל.

ע'5 התיעוד הוא חלק מהתוצר. תוצר בלי יומן אימות אינו תוצר. אין חריגים.

ע'6 מדוד כשירות בלי כלים. לפחות רכיב הערכה אחד בכל הכשרה רץ בלי גישה לכלי. סעיף 2.2 נותן את הנימוק.

ע'7 עצמאות מכלי. השיטה מדברת על משימות, לעולם לא על מוצרים. שם מוצר אינו מופיע בשיטה. שמות מוצרים מופיעים רק בחומרי הדרכה, מסומנים כדוגמה.

5.3 שלושה שלבים ושער אחד



תרשים 1. התהליך. שלושה שלבים בסדר קבוע. השער עומד בין פלט הכלי לכל התחייבות. פריט אחד שאינו ניתן לאימות חוסם את הפלט כולו.

שלב א'. עשה בעצמך. מעבר ראשון על המשימה בלי כלי. אומדן, קריאה ידנית, השערה כתובה, טיוטה גולמית. שלב א' לוקח חמש עד עשרים דקות. מטרתו שיפוט עצמאי מתועד, לא שיפוט מדויק. המהנדס כותב את האומדן לפני שהוא פותח את הכלי. אומדן שלא נכתב אינו נחשב שלב א'.

שלב ב'. שאל את הכלי. פנייה מובנית עם שישה חלקים חובה:

1. תפקיד: את מי הכלי מדמה (למשל, הנדסאי אחזקה ותיק).
2. הקשר: סוג הציוד, המתקן, הסביבה, אילו ערכים באים ממדידה ואילו לא.
3. משימה: מה להפיק, בפורמט מוגדר.
4. אילוצים: יחידות, תקן מחייב, אורך.

5. מקור: המסמך המוסמך, מצורף כשרלוונטי.
6. הנחיית החסר: "אל תמציא. כתוב חסר. אם חסר לך נתון, שאל לפני שאתה כותב."

פנייה בלי חלקים 4 ו-6 אינה פנייה לפי השיטה.

שלב ג'. השער: השווה, אתר את הפער, תעד. המהנדס בודק כל פריט בפלט מול מקור מוסמך, רושם את הפער, וכותב את התוצאה ביומן האימות. טבלה 3 נותנת את כלל ההחלטה.

טבלה 3. כלל ההחלטה בשער.

מצב הפריט	פעולה
אומת, תואם למקור	עובר. רשום מקור ומיקום.
אומת, נמצא פער	מתוקן לפי המקור. רשום פלט מקורי, מקור, תיקון, סוג הפער.
לא ניתן לאמת, אין מקור זמין	**נחסם.** הפריט מסומן "לא אומת". אינו נכנס לתוצר. אם התוצר תלוי בו, התוצר נשאר טיוטה.
הכלי כתב חסר	התנהגות נכונה. נרשם ככזה.

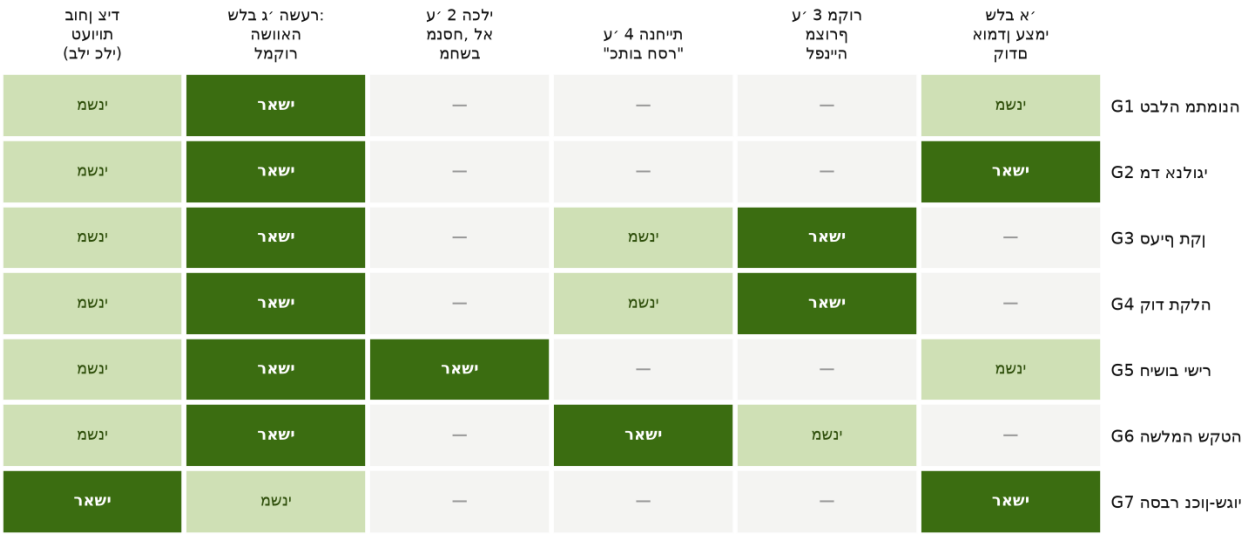
כלל השער. פלט שבו פריט אחד נחסם אינו תוצר. אימות חלקי אינו מצב בטבלה 3. פריט חסום אחד מחזיק את הפלט כולו כטיוטה.

5.4 מעבר לשער: התחייבות

רק תוצר רשאי לעבור להתחייבות. תוצר הוא פלט שכל פריט בו עבר אימות או נושא את סימון החסר, עם יומן אימות מלא.

5.5 איזו בקרה תופסת איזה כשל

תרשים 3 ממפה את שבעת הסוגים לבקרות שבשיטה. לכל סוג יש לפחות בקרה ראשית אחת. השער הוא הבקרה הראשית לשישה מתוך שבעה. G7 הוא החריג. תשובה נכונה עם הנמקה שגויה עוברת בדיקת מקור. אותה תופסים שלב א', שבו קיימת הנמקה עצמאית של המהנדס להשוואה, ובוחן ציד הטעויות.



תרשים 3. מטריצת כיסוי. השורות הן סוגי כשל. העמודות הן בקרות בשיטה. ראשי: הבקרה מתוכננת לתפוס את הסוג הזה. משני: היא תופסת אותו כתוצאת לוואי. זהו מיפוי תכנוני, לא מדידה. אנו מתכננים לבחון אותו מול הפערים שנרשמו בגרסה 1.0.

6. המכשירים

ארבעה מסמכים במבנה קבוע. המבנה קבוע כדי שכל אחד יוכל לספור את התוצאות (סעיף 10).

6.1 יומן האימות (נספח א')

יומן אחד מצורף לכל תוצר. טבלה 4 נותנת את 13 שדות החובה לפי הסדר. שדות 3, 4, 6, 10 ו-11 הם רשימות סגורות. רשימות סגורות הופכות את היומן לניתן לספירה (סעיף 10).

טבלה 4. שדות יומן האימות.

#	שדה	סוג	ערכים
1	מזהה תוצר	טקסט	קוד המשימה או המפגש
2	תאריך	תאריך	

#	שדה	סוג	ערכים
3	דיסציפלינה	רשימה	מכונות · מיזוג · אחר
4	סוג התוצר	רשימה	דוח · נוהל · חישוב · מפרט · טבלה · תרגום · אחר
5	הפריט שנבדק	טקסט	תיאור קצר
6	סוג הפריט	רשימה	מספר · מק"ט/דגם · סעיף תקן · קוד תקלה · תדירות · פעולה מדווחת · תרגום מונח · אחר
7	מה הכלי אמר	טקסט	הפלט המקורי, מילה במילה
8	מקור האימות	טקסט	שם המסמך המוסמך
9	מיקום במקור	טקסט	עמוד / סעיף / טבלה
10	תוצאה	רשימה	**תואם** · **פער** · **לא אומת** · **חסר (סומן ע"י הכלי)**
11	סוג הפער	רשימה	ריק אם תואם. אחרת **G1–G7** או "אחר"
12	מה תוקן	טקסט	הערך הסופי
13	זמן אימות (דק')	מספר	

6.2 הצהרת שימוש ב-AI (נספח ב')

מצורפת לכל הגשה ולכל תוצר ארגוני. שבעה שדות: הכלים שבהם נעשה שימוש. החלקים בתוצר שהם נגעו בהם. החלקים שנוצרו ידנית. הפריטים שאומתו, והמקור לכל אחד. הפריטים שנמצאו שגויים ותוקנו. הרמה בסולם (סעיף 7). חתימה ותאריך.

6.3 בוחן ציד הטעויות (נספח ג')

מודד כשירות בלי כלים. שישה פלטי כלי מתחום העבודה, בכל אחד שגיאה אחת מסוג G1–G7. הנבחן מזהה את השגיאה, מסווג אותה, ומציין את האימות שהיה תופס אותה. הבוחן רץ בלי גישה לכלי, ולכן הוא רכיב ההערכה היחיד ששום כלי אינו יכול לעקוף.

6.4 טבלת תיוג המשימות (נספח ד')

מיישמת את סולם חמש הרמות על המשימות של ארגון או של קורס. שדות: משימה · תוצר · רמה 0–4 · נימוק · מאשר · תאריך.

7. סולם חמש הרמות

הארגון מתייג כל משימה לפני שהעבודה מתחילה. התיוג הוא רשומה כתובה בטבלת תיוג המשימות (נספח ד'). תרשים 4 וטבלה 5 נותנים את הסולם.



.ישוקה יפל אל ,תועט לש הכלשהה יפל תעבקנ המרה .0 המר :תגיותמ אל המישמל לדחמ תרירב

תרשים 4. סולם חמש הרמות. הרמה נקבעת לפי ההשלכה של טעות, לא לפי הקושי של המשימה.

טבלה 5. רמות ודרישות.

רמה	שם	מה מותר	נדרש
0	ללא כלי	אין שימוש בכלי	ללא
1	רעיונות בלבד	הכלי מציע כיוונים, מבנה, שאלות. כל התוכן נכתב ידנית	הצהרה
2	עריכה	הכלי משפר טקסט שנכתב ידנית: ניסוח, תרגום, סיכום, שפה	הצהרה

רמה	שם	מה מותר	נדרש
3	טיוטה מאומתת	הכלי מנסח את התוצר. כל פריט עובר את השער (סעיף 5.3)	הצהרה + יומן מלא
4	מלא, באישור	שימוש נרחב כולל ניתוח וארגון. רק באישור מפורש של הממונה	הצהרה + יומן מלא + אישור כתוב

שלושה כללי תיוג חלים. ברירת המחדל למשימה לא מתויגת היא רמה 0, עד שמישהו מתייג אחרת. הרמה נקבעת לפי ההשלכה של טעות, לא לפי הקושי. חישוב פשוט שנכנס לתסקיר חתום הוא רמה 0. ארגון רשאי להקשיח רמה. הוא אינו רשאי להרפות רמה מעבר למה שהשיטה מתירה.

8. התקנה בארגון

השיטה מותקנת בחמישה צעדים, בסדר קבוע (תרשים 5, טבלה 6). ההכשרה היא צעד 4.



תרשים 5. התקנה בחמישה צעדים. ההכשרה היא צעד 4, לא צעד 1.

טבלה 6. צעדי ההתקנה.

צעד	שם	מה נעשה	תוצר
1	מיפוי	אילו תוצרים נחתמים או מובילים להתחייבות. מי חותם. אילו מסמכים אסור להעלות לכלי ענן (סודיות, סיווג, קניין רוחני)	רשימת תוצרים + רשימת "אסור לענן"
2	תיוג	כל תוצר ברשימה מקבל רמה, בכתב, עם נימוק ומאשר	טבלת תיוג (נספח ד')
3	הטמעה	היומן והצהרה נכנסים לתבניות הקיימות של הארגון. לא טופס חדש	תבניות מעודכנות
4	הכשרה	סדנה על התוצרים של הארגון עצמו, בשיטת שלושת השלבים. מסתיימת בבוחן ציד טעויות, בלי כלי	עובדים שהוכשרו · תוצאות הבוחן
5	מדידה	איסוף היומנים. אחת לרבעון: אילו סוגי כשל חוזרים, באילו תוצרים, ומה משתנה בתיוג	דוח רבעוני · תיוג מעודכן

כלל ההתקנה. צעד 4 דורש את תוצרי צעדים 1 עד 3: רשימת התוצרים וטבלת התיוג. פורמט הסדנה פעיל, והמשתתפים עובדים על התוצרים של הארגון עצמו. פורמטים פעילים עדיפים על הרצאות בלימודי STEM לתואר ראשון (Freeman et al. 2014). אנו מניחים, ומתכננים לבדוק, שהתוצאה עוברת גם לכאן.

9. גבולות התחולה

השיטה חלה בשלושה תנאים. מחוץ להם היא אינה נותנת שום ביטחון.

9.1 השיטה אינה חלה כאשר:

- אין מקור מוסמך. אם לא קיים מסמך שיש לו אחראי, אין מול מה לאמת. השער חוסם את הפריט. זו התוצאה שהשיטה מגדירה, לא כשל שלה.
- המשימה יצירתית או פתוחה ואינה מובילה להתחייבות. דוגמאות: סיעור מוחות ולמידה אישית. אלה משימות רמה 1, והשער אינו חל.
- ה"כלי" הוא מערכת דטרמיניסטית מאומתת. תוכנת חישוב הנדסי מאושרת אינה כלי במובן של שיטה זו.

9.2 השיטה אינה מבטיחה:

- שהתוצר נכון. היא מבטיחה שמישהו בדק כל פריט מול מקור ורשם את התוצאה. אם המקור שגוי, התוצר שגוי, והיומן מראה בדיוק איפה.
- מהירות. שלב ג' מוסיף זמן לכל תוצר. שדה 13 רושם אותו, ואנו מתכננים לדווח עליו (סעיף 10). האם הוא קטן מעלות של טעות חתומה היא שאלה אמפירית.

9.3 השיטה אינה תחליף ל:

- כשירות מקצועית. אימות דורש כשירות לקרוא את המקור. השיטה מניחה את הכשירות הזאת ואינה מספקת אותה.

• אחריות מקצועית. האחריות נשארת אצל החותם. (2025) ASHRAE ו-(2025) ASTM נוקטות את אותה עמדה.

10. פרוטוקול ההערכה לגרסה 1.0

10.1 מי כשיר להגיש יומן

קבוצה מיישמת את הפרוטוקול כשהיא עומדת בשלושה קריטריוני הכללה ומגישה יומני אימות לפי שדות נספח א'. ראשון, כל משתתף בקבוצה חותם על עבודה בהנדסת מכונות או במיזוג אוויר. שני, הקבוצה משתמשת בשיטה כשלד שלה ובמכשירים של סעיף 6 בעבודה שלה. שלישי, כל משתתף מגיש יומנים בקצב קבוע ולאורך פרק זמן קבוע. הקבוצה היא לכן גם מכשיר המדידה. תרשים 6 נותן את ציר הזמן.



תרשים 6. ציר הזמן של ההערכה. הסכמה ותיוג בנקודת יומן 1. מבחן מקדים בלי כלים בכניסה. יומן אימות אחד למשתתף למשימה בנקודות יומן 3 עד 13. מבחן מסכם בלי כלים בסיום.

10.2 הנתונים

כל משתתף מגיש כעשרה יומני אימות (נספח א'). שדות 3, 4, 6, 10 ו-11 הם רשימות סגורות. הנפח הנדרש הוא כ-300 יומנים: כ-30 משתתפים שמגישים כעשרה כל אחד. כל שורת יומן היא פריט מאומת אחד. בוחן ציד הטעויות רץ פעמיים, בכניסה ובסיום, ברמת קושי תואמת ובלי גישה לכלי.

10.3 שאלות המחקר, שנקבעו לפני האיסוף

1. התפלגות סוגי הפער G1–G7 לפי דיסציפלינה ולפי סוג תוצר.
2. שיעור הפריטים שנחסמו (לא ניתן לאמת) לפי סוג פריט.
3. זמן אימות ממוצע לפי סוג תוצר.
4. שינוי לאורך רצף היומנים: האם שיעור הפערים שנתפסו עולה עם התרגול?
5. השוואה בין שתי הדיסציפלינות.
6. שינוי בציון ציד הטעויות מכניסה לסיום.

10.4 הסכמה ופרטיות

הסכמה מדעת בכתב בכניסה (נספח ו'). אנו מנתחים נתונים במצטבר ובאנונימיזציה בלבד. שום משתתף או מעסיק אינו ניתן לזיהוי בשום פרסום. ההשתתפות במחקר נפרדת מהעבודה עצמה. משתתף רשאי לסרב למחקר בלי שום פגיעה.

10.5 הניתוח

סטטיסטיקה תיארית לשאלות 1 עד 3 ו-5. לשאלה 4, שיעור הפערים ליומן לאורך הרצף, עם רווחי סמך. לשאלה 6, השוואה מזווגת של ציוני לפני ואחרי. כל הספירות נגזרות משדות ברשימות סגורות. אנו מתכננים לשחרר את סקריפט הניתוח עם הנתונים.

10.6 מה הנתונים לא יאמרו

הפרוטוקול מקבל יומני אימות ממשתתפים. הנתונים לא אומרים דבר על ההתקנה בארגון. את המקרים הארגוניים (סעיף 8, צעד 5) נתעד בנפרד.

11. מגבלות

בסיס הראיות. לא נאספו נתונים. לא גויסה קבוצה. גרסה 0.91 נשענת על ראיות חיצוניות מפורסמות, ואנו אומרים זאת במפורש. אף אחת מהן אינה מודדת את השיטה הזאת. הן מודדות את הכשלים שהשיטה מכוונת אליהם. המיפוי בתרשים 3 הוא טענה תכנונית, לא תוצאה.

מחבר יחיד. מחבר השיטה כתב את המכשירים, את כללי הקידוד ואת הסיווג. צפוי שההתאמה לכוונת השיטה תהיה גבוהה ממה שמקודד חיצוני היה מפיק. אנו משחררים את המכשירים ואת כללי הקידוד כדי שאחרים יוכלו להריץ את הפרוטוקול באופן עצמאי.

דיווח עצמי ביומן. שדה 13 (זמן אימות) ושדה 7 (הפלט המילולי) תלויים במשתתף. אנו מתכננים לבדוק מדגם אקראי של יומנים מול שיחות הכלי השמורות של המשתתפים.

סחף בין דורות. התנהגות הכלים משתנה בין גרסאות. סוג כשל שכיח ב-2026 עשוי להיות נדיר ב-2028. השיטה מדברת על סוגי משימות כדי לשרוד את הסחף הזה. אנו מתכננים לעדכן את הסיווג בגרסה 1.0 מול הנתונים שנרשמו.

G7 עדיין לא נמדד. תשובה נכונה עם הנמקה שגויה נכנסת לסיווג מתצפית שדה. זה הסוג שסביר ביותר שישתנה בגרסה 1.0.

12. סיכום

כלי בינה מלאכותית גנרטיבית מגיעים לעבודה ההנדסית עם ציונים כלליים גבוהים וכשלים ספציפיים וחוזרים בנקודות החתימה. הגדרנו נקודת עצירה למצב הזה. קבענו אקסיומה אחת וגזרנו ממנה שבעה עקרונות. קבענו שלושה שלבים והצבנו שער אחד בין פלט להתחייבות, עם כלל החלטה בעל ארבע תוצאות ובלי מעבר חלקי. שחררנו ארבעה מכשירים עם שדות ברשימות סגורות וסולם בן חמש רמות הקשור להשלכה של טעות. קבענו את גודל המדגם הנדרש ואת השדות שלו. קבוצה שעומדת בשלושת קריטריוני ההכללה ומגישה יומנים לפי השדות האלה מפיקה נתונים מדודים על השיטה עצמה. אנו משחררים כל מכשיר, כדי שכל אחד יוכל לבחון את השיטה, לערער עליה מקרה-מקרה, או לשפר אותה.

הצהרת שחזוריות ואתיקה

אנו משחררים שישה מכשירים ברישיון CC BY 4.0. הם תבנית יומן האימות (דו-לשונית, רשימות סגורות, גיליון סיכום), הצהרת השימוש ב-AI, ומבנה ומחוון בוחן ציד הטעויות. הם כוללים גם את תבנית תיוג המשימות, תבנית הפנייה, ונוסח ההסכמה מדעת. נספח ז' הוא יומן האימות של המאמר עצמו. הוא כולל 23 פריטים: 14 תואמים, 6 פערים שתוקנו לפני הפרסום, 2 פריטים שהוסרו כלא ניתנים לאימות, ונתון מתוכנן אחד שסומן ככזה. אנו מתכננים להוסיף בגרסה 1.0 את מאגר היומנים המאונונימז, כללי הקידוד, וקוד הניתוח. משתתפים מצטרפים למחקר בהסכמה כתובה ורשאים לפרוש בכל עת. לא נפרסם שום נתון אישי או נתון מעסיק.

רישיון וייחוס

מסמך זה נושא רישיון CC BY-NC-ND 4.0. כל אחד רשאי להפיץ אותו בשלמותו עם ייחוס. אף אחד אינו רשאי לשנות אותו או להשתמש בו מסחרית בלי הרשאה. המכשירים בנספחים נושאים רישיון CC BY 4.0. הם מופקדים ברשומת Zenodo נפרדת, המקושרת מרשומה זו. כל אחד רשאי להשתמש בהם, להתאים ולהפיץ אותם עם ייחוס למחבר ולשיטה. ייחוס נדרש לכל שימוש: "שער האימות · The Verification Gate — M. Melchior, v0.91, Concept DOI 10.5281/zenodo.22271339, physical-ai-safety.com". ההפקדה ב-Zenodo עם DOI וחותמת זמן קובעת את תאריך הגרסה הזאת. עמוד השיטה, עדכונים והנספחים נמצאים ב-physical-ai-safety.com/method/verification-gate. שם השיטה אינו רשום כסימן מסחר בגרסה זו. הרישום יבוא כשיהיה גוף מסמיך.

מקורות

ASHRAE. 2025. *ASHRAE Policy on Artificial Intelligence.* February 2025. Atlanta, GA: ASHRAE. https://www.ashrae.org/file%20library/about/position%20documents/ashrae-ai-policy-feb-2025.pdf

ASTM International. 2025. *ASTM Policy on the Use of Artificial Intelligence.* West Conshohocken, PA. https://store.astm.org/media/wysiwyg/ASTM_AI_Policy.pdf

Bastani, H.; Bastani, O.; Sungu, A.; Ge, H.; Kabakcı, Ö.; and Mariman, R. 2025. Generative AI Without Guardrails Can Harm Learning: Evidence from High School Mathematics. *Proceedings of the National Academy of Sciences*, 122(26): e2422633122. https://doi.org/10.1073/pnas.2422633122

Choi, H.; Lee, J.; and Kim, J. 2025. Performance Evaluation of GPT-4o on South Korean National Exams for Building Mechanical Equipment Maintenance. *Scientific Reports*, 15: 30436. https://doi.org/10.1038/s41598-025-16118-x

Daikin. n.d. *Error Code Reference (Service Manual Extract).* Technical document. https://www.daikin.com/-/media/Project/Daikin/daikin_com/products/ac/services/error_codes/pdf/sm-ts3_p1-6_errorcode-pdf.pdf

Freeman, S.; Eddy, S. L.; McDonough, M.; Smith, M. K.; Okoroafor, N.; Jordt, H.; and Wenderoth, M. P. 2014. Active Learning Increases Student Performance in Science, Engineering, and Mathematics. *Proceedings of the National Academy of Sciences*, 111(23): 8410–8415. https://doi.org/10.1073/pnas.1319030111

GREE. n.d. *Understanding GREE Mini-Split Error Codes.* Technical note. https://www.greecomfort.com/news-and-events/understanding-gree-mini-split-error-codes/

Huang, L.; Yu, W.; Ma, W.; Zhong, W.; Feng, Z.; Wang, H.; Chen, Q.; Peng, W.; Feng, X.; Qin, B.; and Liu, T. 2025. A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. *ACM Transactions on Information Systems*, 43(2). https://doi.org/10.1145/3703155

Liang, C.; Huang, Z.; Wang, H.; Chai, F.; Yu, C.; Wei, H.; Liu, Z.; Li, Y.; Wang, H.; Luo, R.; and Zhao, X. 2026. AECBench: A Hierarchical Benchmark for Knowledge Evaluation of Large Language Models in the AEC Field. *Advanced Engineering Informatics*, 71: 104314. https://doi.org/10.1016/j.aei.2026.104314

Lin, F.; Liu, Y.; Xu, H.; Yue, C.; He, Z.; Zhao, M.; Chen, M. H.; Liu, J.; Yao, J. G.; and Yang, X. 2026. Do Vision-Language Models Measure Up? Benchmarking Visual Measurement Reading with MeasureBench. arXiv:2510.26865, v2 (24 March 2026). Preprint, not peer-reviewed. https://arxiv.org/abs/2510.26865

Lu, C.; Walker, S.; Struck, C.; Itard, L.; and Saelens, D. 2025. Poster Abstract: P&ID-to-Graph: LLM-Assisted Digitalization of HVAC Diagrams. In *Proceedings of the 12th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '25)*, 296–297. https://doi.org/10.1145/3736425.3772104

Mitsubishi Electric. n.d. *Mr. Slim Error Code List.* Technical document. https://www.mitsubishielectric.com.sg/getmedia/1a8fe41a-9b17-49f2-a8b5-3e4842aa99ba/MrSlimerrorCode.pdf

Thomas, B. 2026. When "Smart" Tools Make Dumb Mistakes: AI's Hidden Risks in Construction Documents. Gausnell, O'Keefe & Thomas LLC, January 2026. https://gotlawstl.com/when-smart-tools-make-dumb-mistakes-ais-hidden-risks-in-construction-documents/

נספחים (משוחררים בנפרד, CC BY 4.0)

א'. תבנית יומן האימות (xlsx, דו-לשונית, רשימות סגורות, גיליון סיכום). מבנה לפי סעיף 6.1. **ב'.** הצהרת שימוש ב-AI (נוסח קבוע). **ג'.** בוחן ציד הטעויות: מבנה, מחוון, ומחולל פריטים. **ד'.** תבנית טבלת תיוג משימות (xlsx). **ה'.** תבנית פנייה לפי סעיף 5.3, שלב ב'. **ו'.** נוסח הסכמה מדעת לפי סעיף 10.4. **ז'.** יומן האימות של המאמר עצמו: 23 פריטים, כל מספר מקושר למקורו (xlsx).

היסטוריית גרסאות

גרסה	תאריך	שינוי
0.9	09/2026	מסמך יסוד. השיטה מוגדרת במלואה. ראיות ממקורות חיצוניים בלבד.
0.91	09/2026	סוגי הכשל F1 עד F7 שונו ל-G1 עד G7 (G על שם השער). שום טענה, מספר, מקור או מבנה לא השתנו.
1.0	—	+ תוצאות מדודות, רק כשקיימים יומנים רשומים
2.0	—	+ מסלול הסמכה · רק כשקיים גוף מסמיך

ה-Concept DOI 10.5281/zenodo.22271339 מכסה את כל גרסאות המאמר. ה-DOI של גרסה 0.91 הוא 10.5281/zenodo.22844321.

כל מספר במסמך זה מקושר למקור מצוטט. מספרים שלא הצלחנו לקשר הושמטו. יומן האימות של המסמך עצמו משוחרר עם הנספחים.