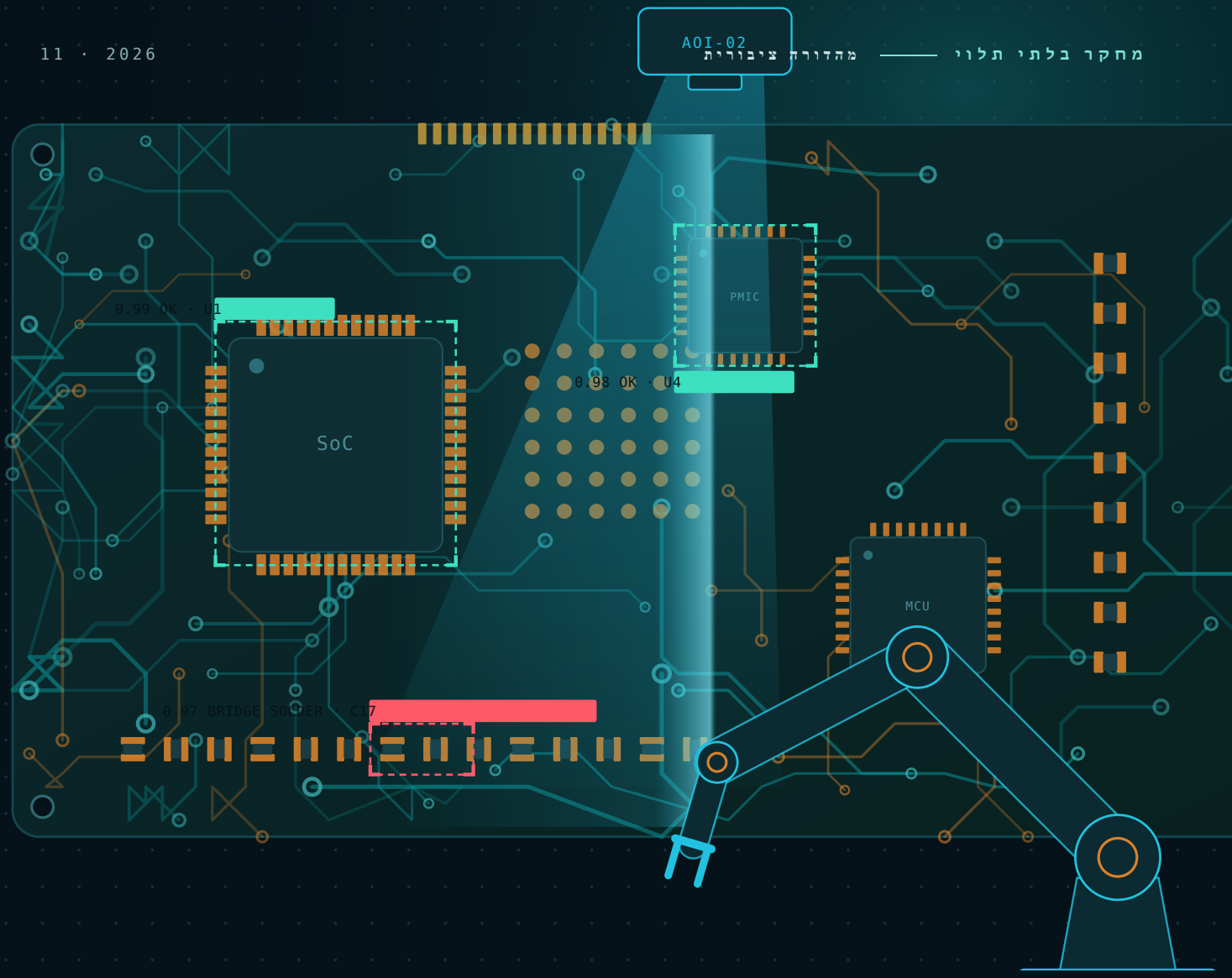


ייצור אלקטרוניקה · בדיקת AI × רובוטיקה

בדיקה חזותית מבוססת AI וזרועות רובוטיות במפעלי אלקטרוניקה

הערכת 2026

תצורות נפוצות · היתכנות אזורית · פריסה מקומית · מודלי מורה–תלמיד
החזר השקעה · תחרות · מה שרק פיילוט יכול להכריע

ביקורת AI שווה לעשות, בצורה ממוקדת. זרועות רובוטיות מגולמות יכולות לחכות.

ממצא מרכזי
Tan Shuai
tanshuai.comi@tanshuai.com

נתונים נכון ל-11 בנובמבר 2026
Mainland China · Taiwan · Southeast Asia · United States

תוכן

תוכן העניינים



◆ **תקציר מנהלי** — שווה לעשות, אבל בצורה ממוקדת, עכשיו ועם שותפים — 4

01 **הקו היום** — מה מפעלי אלקטרוניקה מריצים היום, ומה ישים מבחינה טכנית — 7

02 **פריסה מקומית** — הרצת AI בתוך המפעל: חדרים נקיים, ניתוק אוויר, חומרה וקוונטיזציה — 14

03 **מורה ותלמיד** — מורה ותלמיד: מודלי חלוץ יכולים לתייג מראש נתוני בדיקה; עבור זרועות רובוטיות המורים האמיתיים הם עדיין אנשים — 23

04 **עלויות והחזר השקעה** — עלויות והחזר השקעה: ביקורת AI מחזירה את עצמה תוך חודשים, טיפול במכונות תוך שנים, זרועות הוצאה לעולם לא — 31

05 **שוק ותחרות** — שוק, תחרות ואיך קונים משלמים — 38

06 **AI מגולם** — AI מגולם: קודם הוא משנה איך מלמדים רובוטים, אחר כך איך הם נעים — 47

07 **שאלות פתוחות** — מה שהרשומה הציבורית אינה יכולה לענות עליו, ואיך פיילוט מכריע זאת — 52

◆ **אודות המחבר** — Tan Shuai — 54

תגי ביטחון בטקסט

- **verified** שני מקורות בלתי תלויים או יותר, או שני מסמכים ראשוניים בלתי תלויים, מסכימים
- **single source** מקור אחד בלבד, כולל מסמכים ראשוניים
- **vendor claim** פורסם על ידי מוכר המוצר או השירות
- **estimate** חשבון או היסק של המחבר מתוך נתונים מצוטטים

כיצד לקרוא דוח זה

מספרים מרכזיים נושאים תג ביטחון. **verified** פירושו ששני מקורות בלתי תלויים או יותר מסכימים, או ששני מסמכים ראשוניים בלתי תלויים מסכימים; **single source** פירושו מקור אחד בלבד (כולל מסמכים ראשוניים); **vendor claim** פירושו שפורסם על ידי מוכר המוצר או השירות, גם אם חוזר על עצמו בהרחבה; **estimate** פירושו חשבון או היסק שלי מנתונים מצוטטים, עם תיעוד הקלטים. "ראשוני" מתאר את סוג המקור (דוח שנתי, תשקיף, תקנה, תקן, סטטיסטיקה רשמית); הוא אינו משנה את דרגת הביטחון. היכן שמקורות סותרים זה את זה, הנתונים מוצגים זה לצד זה ללא ממוצע. המרות מטבע משתמשות בשער המרכזי של בנק העם הסיני מ-9 בנובמבר 2026, $1 = RMB 6.7188 (Xinhua), ועבור טאיוואן שער הקביעה של SMBC באותו יום, $1 = NT$32.12 (SMBC). נתוני שכר וסבסוד בסין היבשתית משתמשים בדלתת נהר הפנינה, שם ייצור האלקטרוניקה מרוכז ביותר (נתוני שכר רשמיים משנזין), עם דלתת נהר היאנגצה להשוואה. טענות מראיונות בתעשייה מטופלות כטענות שיש לבדוק, לעולם לא כעובדות. שאלות שהרשומה הציבורית אינה יכולה לענות עליהן מרוכזות בפרק 7.

איורים

איורים



01	היכן AI משתלב ברצפת המפעל כיום	6
02	קו SMT בייצור המוני: האנשים יושבים מאחורי ה-AOI	7
03	כ-97–99% מהתראות ה-AOI הן קריאות שווא	9
04	מצלמה לשימוש כללי אינה AOI	10
05	פריסה מנותקת-אוויר	14
06	אזורים נקיים ואזורי ESD קובעים היכן יכול לשבת החישוב	15
07	מגבלות מהירות לפי שימוש	17
08	חישוב מקומי: זיכרון, מחיר ומה ניתן לרכוש באופן חוקי בסין היבשתית	18
09	תחנת עבודה גדולה אחת אינה יוצרת עוזר משותף	19
10	קוונטיזציה: bit-8 בטוח, bit-4 תלוי במתכון, bit-3 ומטה בחוץ	20
11	מועמדי מורה: ציונים גנריים הוכפלו בדור	24
12	"האם יש פגם?" עדיין קשה למודלים כלליים	25
13	שני צינורות מורה–תלמיד	26
14	כמה גדול חייב להיות סט הזהב	27
15	השהיית מדיניות קצה	28
16	תלמידי רובוט: קרוב לדרגת ייצור רק כשהעמדה מובנית	29
17	כמה עולה עובד בדיקה או טיפול אחד בשנה	32
18	סולם המחירים	33
19	החזר לפי אזור	34
20	מדוע הוצאת לוחות גרועים אינה משתלמת	35
21	מכס ארה"ב על רובוטים, AOI ומחשבי קצה	36
22	גדלי השוק נבדלים ביותר מפי עשרה	39
23	מלמטה-למעלה: 40,000 קווים → 16,000 ניתנים לשירות → ≈RMB 320M חד-פעמי	40
24	חלון השדרוג לאחור הוא כחמש שנים	41
25	טריז הבדיקה מכותר משלושה צדדים	41
26	המרווח יושב בחומרה ממוצרת	43
27	כיצד הקונים משלמים	44
28	ימי מכירות פתוחות	45
29	ערימת ה-AI המגולם	47
30	מה פריסות רובוט פנים-מפעליות חושפות בפועל	48
31	כוח עבודה שהוחלף נטו לרובוט	49
32	ציר זמן תקנים ורגולציה	50

שווה לעשות, אבל בצורה ממוקדת, עכשיו ועם שותפים

ממצא מרכזי

פסק הדין: עבור מפעלי אלקטרוניקה, מה שבשל ב-2026 ועל מה שמפעלים ישלמו הוא ביקורת AI ובדיקת נישות על תמונות AOI, בתוספת טיפול מתוכנת במכונות ברובוטים, לא "מודל גדול שמביט בלוחות בתוספת רובוטים מגולמים שמרימים אותם".

0.65–0.46 שנה
החזר השקעה של ביקורת AI ב-the US, Taiwan ו-Shenzhen, ב-RMB 50k לקו

≤95% · >20 PPM
כיסוי וברירות של ביקורת AI פנימית ב-EMS מובילה: הרף למכירה ל-EMS מדרגה ראשונה

700–100 מיליון RMB
שוק שדרוג לאחור ניתן לשירות בידי צד שלישי ב-mainland China, חד-פעמי, על פני חלון של כחמש שנים

0 נקודות
ציוני מדד פגמים תעשייתיים שפורסמו לדגלים של נובמבר 2026: מורים, לא מפקחים

1.5–1.1 שנה
החזר השקעה של תא קובוט לטיפול במכונות על פני שתי משמרות ב-the US, Taiwan ו-Shenzhen

0.14–0.003 FTE
כוח העבודה שזרוע רובוטית להוצאה חוסכת למשמרת על קו SMT: היא לעולם לא מחזירה את עצמה

בדיקה
ROI

למידה עמוקה מפוקחת המיושמת על התמונות ש-AOI קיים כבר מייצר מקצצת קריאות שווא או שעות ביקורת ב-60–90% (טענות ספק). DBG, חברת EMS גדולה, מדווחת בדוח השנתי שלה שהמערכת הפנימית שלה מכסה יותר מ-95% מהרכיבים המסומנים עם בריחות מתחת ל-20 PPM ועם הפחתה של יותר מ-85% בביקורת האנושית; זהו הרף שמוצר צד שלישי חייב לעבור כדי למכור ל-EMS מדרגה ראשונה. ב-RMB 50,000 לקו, ביקורת AI מחזירה את עצמה תוך שנה בסין היבשתית, בטאיוואן ובארה"ב (שנזין 0.65 שנים, טאיוואן 0.53, ארה"ב 0.46), תאי טיפול קובוט במכונות ב-1.1–1.5 שנים ודרום-מזרח אסיה ב-2–6 שנים, בעוד שזרוע רובוטית שמוסיפים לקו SMT רק כדי להוציא לוחות פגומים לעולם לא מחזירה את עצמה בשום מקום, כי מעל 90% תפוקת מעבר ראשון ההוצאה לוקחת רק 0.003–0.14 של אדם למשמרת estimate.

מודלים
רובוטים

מודלים גדולים רב-מודאליים עדיין אינם יכולים לשמש כמפקח הראשי. נכון ל-11 בנובמבר 2026, לאף אחד מהדגלים (Claude Fable 5.2, GPT-6.x, Gemini 4, Qwen3.8 ואחרים) אין ציון שפורסם על שום מדד פגמים תעשייתיים או PCB, ובאתגר PCBA של ACM MM 2026 חמש המערכות המובילות, שאומנו על 6,600 תמונות קו אמיתיות, קיבלו רק 77–82 בשאלת "האם יש פגם". מודלי חלוץ יכולים לתייג מראש ולמיין. עבור זרועות רובוטיות, המורים האמיתיים הם עדיין הדגמות אנושיות, למידת חיזוק על רובוט אמיתי וסימולציה: תלמידים מגיעים ל-90–99.4% בתחנות מובנות ומקובעות, אבל רק 30–80% על מסועים נעים, חלקים שלא נראו ומוצרים חדשים; סוכן חלוץ השלים 104 מתוך 108 משימות רובוטיות מעין-סטטיות על רובוט אמיתי, ב-280 שניות כל אחת.

מקומי
פריסה

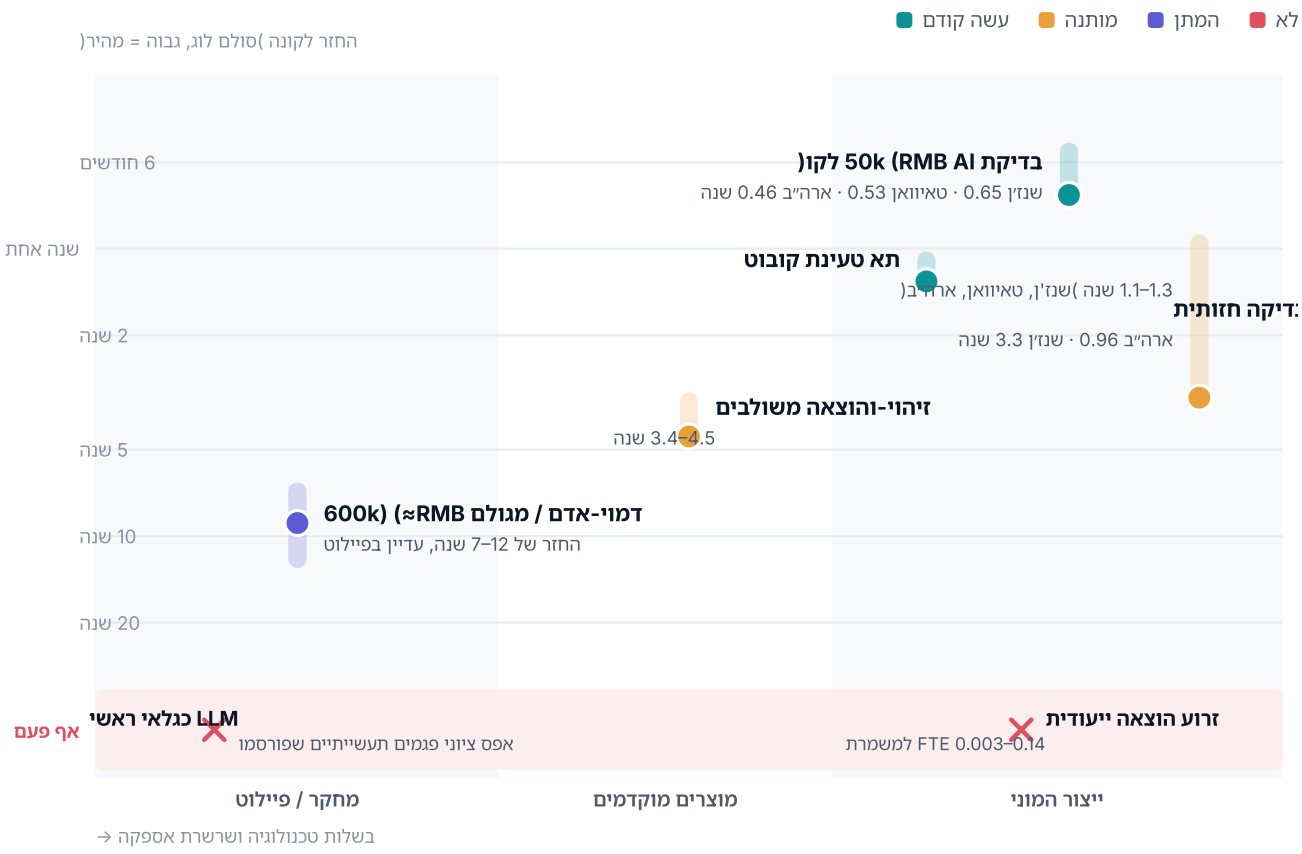
פריסה מנותקת-אוויר במפעלים נקיים ומבוקרים היא מעשית לחלוטין. אולם SMT רגיל הוא בדרך כלל אזור מוגן ESD ולא חדר נקי מדורג: תיבת Jetson ללא מאוורר או תיבת NPU מקומית עולה ליד הקו ושרתי GPU עולים לחדר השרתים. לפי סיווגי הייצוא של NVIDIA עצמה, ניתן למכור Jetson Thor ו-DGX Spark למפעל רגיל בסין היבשתית ללא רישיון אמריקאי; את DGX Station לא ניתן. הדיוק מוגדר כברירת מחדל ל-4-bit; 8-bit מודע-לקוונטיזציה של הספק שמיש בתנאים; 3-bit ומטה פסול.

שוק
תזרות

השוק אמיתי אך מוגבל. לסין היבשתית יש כ-35,000–46,000 קווי SMT וכ-26,000–66,000 AOI בקו שאינם מריצים AI; השוק הניתן לשירות לשדרוגי צד שלישי הוא כ-700–100 RMB מיליון חד-פעמי, על פני חלון של כחמש שנים, ויצרני AOI החלו לבנות ביקורת לתוך מכונות חדשות estimate . התחרות כבר חזקה: טריז הביקורת מכותר על ידי יצרני AOI, צוותי EMS פנימיים וספקי פלטפורמה, והמרווח יושב בחומרה ממוצרת.

שאלה	הערכה	מספרים מרכזיים (מקורות ורמת ביטחון בפרקים)
1 **מה מפעלי אלקטרוניקה מהזרם המרכזי מריצים היום?**	קו SMT בהיקף גדול הוא "3D SPI + AOI לפני ואחרי הזרמה + X-ray בדגימה + ICT/FCT + תיקון ובדיקה חזותית", עם ביקורת אנושית מאחורי כל AOI. נכנס לביקורת בשלוש דרכים: מובנה על ידי יצרני AOI, נבנה פנימי על ידי EMS מובילים, או נוסף על ידי צדדים שלישיים. הרכבה וטיפול משתמשים בזרועות SCARA ו-AMR עם תפיסה נלמדת ותנועה מתוכנתת; AI מגולם עדיין בפיילוט	טקט ≈13–22 שניות ללוח; ODM אחד תקצב RMB 30.6M של AOI עבור 18 קווים, ≈3–5 AOI לקו; ≈97–99% מהתראות ה-AOI הן קריאות שווא
2 **עד כמה הטכנולוגיה מגיעה, האם כסף יכול לקנות אפס אנשים, ומה כל אזור יכול להשתמש בו?**	הבדיקה היא ברמה תעשייתית כ"גלאי ייעודי + ביקורת AI + כמה אנשים"; אפילו כסף בלתי מוגבל אינו מגיע לאפס אנשים. סין היבשתית משתמשת רק במודלים מקומיים או במשקלים פתוחים מקומיים; מפעלים פרטיים בטאיוואן יכולים להשתמש כמעט בכל דבר, אבל מפעלי שבבים ותשתית קריטית מוציאים בפועל מותגים מסין היבשתית; לדרום-מזרח אסיה יש הכי מעט הגבלות; לארה"ב יש את הערימה המלאה בתוספת מכסים והסמכה	ביקורת AI מקצצת קריאות שווא או שעות ביקורת ב-60–90% vendor claims ; מצלמת 4K מעל לוח שלם ≈78 מיקרון לפיקסל לעומת 8–20 מיקרון ל-AOI; זרועות מתוצרת סין היבשתית משלמות ≈40% מכס אמריקאי משולב, מכשירים מסוג AOI ≈37.5%
3 **האם זה יכול לרוץ מקומית בחדרים נקיים ובמפעלים מנותקי-אוויר, על איזו חומרה, מהר ומדויק מספיק?**	כן; האילוצים הם פרוצדורליים, לא חישוביים. תיבה ללא מאוורר ≥100 W ליד הקו, שרתי GPU בחדר השרתים; Thor לתחנות, Mac Studio או מכשיר מקומי לשאלות-ותשובות במפעל; אין DGX Station בסין היבשתית. המהירות טובה כשגודל המודל תואם למשימה; שמור על bit-8, השתמש ב-4-bit מודע-לקוונטיזציה של הספק בתנאים, לעולם לא bit-3	≈0.5 שניות לחיתוך עבור מודל ראייה 8B על Thor estimate ; ערכת פיתוח Thor ב-RMB 59,104 בסין; DGX Station ב-$123–140k, 139 tokens/s בסך הכול ל-32 משתמשים על מודל 743B
4 **האם הדגל החדש ביותר יכול לשמש כמורה על פני עשרות סבבים ולהתזקק לתלמידים מקומיים, כולל לייצור גמיש ולזרועות רובוטיות?**	בדיקה: תיוג מראש ומיון; קבל תיוגי מורה אוטומטית רק למחלקות עם "≤299 פגמים, אפס החטאות" על סט זהב. רובוטים: המורים הם עדיין הדגמות אנושיות, RL על רובוט אמיתי וסימולציה, ותלמידים מתקרבים לדרגת ייצור רק בתחנות מובנות	לאף דגל מנובמבר אין ציון פגמים תעשייתי; תלמידים 97.8% על מחברים שאומנו עליהם, 29–67% על כאלה שלא נראו, 37–59% על מסועים נעים; מודלי חלוץ שמניעים מפרקים ישירות 0–5.5%
5 **מה העלות, ומה ה-ROI לפי אזור?**	ביקורת AI מחזירה את עצמה תוך שנה בסין היבשתית, בטאיוואן ובארה"ב; טיפול קובוט 1.1–1.5 שנים; דרום-מזרח אסיה 2–6 שנים; זרוע להוצאה על קו SMT לעולם לא מחזירה את עצמה	תיבת ביקורת RMB 50k לקו: שנזין 0.65 שנים, טאיוואן 0.53, ארה"ב 0.46, וייטנאם 2.2 estimate ; מחיר תוסף צד שלישי מקובל ≈RMB 10–30k לקו חד-פעמי או RMB 5–10k בשנה
6 **האם המתחרים כבר חזקים מספיק, והאם עדיף להתחיל עכשיו או לחכות?**	הביקוש אינו רווי, אבל ההיצע צפוף; AI של יצרני ציוד מקורי מגביל את חלון השדרוג לאחור לכחמש שנים. היכנס עכשיו עם מוצר צר דרך שותפי ערוץ, בלי להילחם מול ספקי פלטפורמה חזיתית; חכה לנתוני ייצור של 2027–2028 לפני הימור על AI מגולם	≈26–66k AOI בקו ללא AI בסין היבשתית, מתכווץ 8–12% בשנה estimate ; לספק ראיית ה-AI המוביל היו הכנסות 2025 של RMB 1.086B והפסד נקי של RMB 991M; הכנסות ה-3C של החברה הנסחרת הקרובה ביותר להשוואה ירדו 79.9% שנה-על-שנה בינואר–מאי 2026

FIG 01 **היכן AI משתלב ברצפת המפעל כיום*בדיקת AI של קריאות שווא מ-AOI וטיפול מכונה על ידי קובוט בוגרים ומחזירים מהר; זרוע הוצאה ייעודית ו-LLM כמפקח ראשי אינם משתלמים.***



מקור: החזר ממודל המחבר (פרק 4); ציר הבשלות הוא שיקול דעתו של המחבר.

מה לבנות ומה לא לבנות

	המלצה
בנה	ביקורת AI וסיווג פגמים על תמונות AOI, עבור EMS בינוני עם AOI ישן או רב-מותגי וללא שדרוג AI מיצרן ה-AOI; בדיקה שה-AOI של ה-SMT מכסה בצורה חלשה (הלחמת THT והלחמה ידנית, ציפוי קונפורמלי, כבלים, מחברים ו-FPC, מראה סופי וברגים); תאי טעינה ופריקה של קובוט שבהם כל חלק מטופל (מתקני בדיקה, מגשים, אריזה), עם בדיקה כתוסף; סט זהב ומצב צל על קו אמיתי כבסיס לכל טענה
אל תבנה	פלטפורמה כללית של "AI פלוס זרוע רובוטית"; LLM רב-מודאלי כמפקח הראשי; זרוע רובוטית על קו SMT רק כדי להוציא לוחות NG (מפנה NG או חוצץ עושים זאת); רובוטים דמויי אדם ורובוטים-כשירות; מודלי חלוץ בתוך לולאת בקרה או פונקציית בטיחות; ערמות של זיכרון GPU בתחנות כדי להריץ מודלים "בגודל מלא"; מכירות ישירות למפעלי שבבים, לתשתית קריטית טאיוואנית או למפעלי אלקטרוניקת הגנה אמריקאית
עם מי לעבוד	יצרני AOI ומפיצים (ממשקי ערוץ ותמונות; מפיצים הביאו 46.68% מהכנסות Sinic-Tek ב-2025); יצרני קובוטים (זרועות מוסמכות ופונקציות בטיחות); חברות EMS עם 10–40 קווים (הרחבה מתגלגלת ומקדמות); אינטגרטורים מקומיים בתאילנד ובמקומות אחרים (כדי לעמוד בכללי הקישור המקומי של BOI)

הקו היום

01 מה מפעלי אלקטרוניקה מריצים היום, ומה ישים מבחינה טכנית

קו ההיקף התקני: כמה AOI, X-ray בדגימה, ביקורת מרכזית, 13–22 שניות ללוח

תשקיפים ראשוניים של ODM נסחרים מראים איך נראה בפועל קו היקף של 2025–2026. התשקיף של Longcheer בהונג קונג (נתונים עד 30 בספטמבר 2025) מפרט את בדיקות האיכות של קווי ה-SMT שלו כ"בדיקת X-ray, בדיקת משחת הלחמה (SPI), בדיקה אופטית אוטומטית (AOI), בדיקה תפקודית של PCBA ובדיקה חזותית"; זרימת התהליך שלו מריצה AOI לפני הזרמה, AOI אחרי הזרמה, X-ray בדגימה, לוחות NG לתחנת תיקון SMT, ולאחר מכן בדיקה חזותית סופית. למפעליו בהוויז'ואו, נאנצ'אנג ווייטנאם יש 34, 24 ו-8 קווי SMT ו-67, 60 ו-13 קווי הרכבה סופית (Longcheer H-share prospectus) single source, primary filing. מפעל המחשבים הניידים של Huaqin מ-2023 תכנן 18 קווי SMT עם תקציב AOI של RMB 30.6 מיליון, כ-RMB 1.7 מיליון לקו, שבמחירי 3D AOI מקומיים הם שלושה עד חמישה AOI לקו לבדיקה לפני ואחרי הזרמה של לוחות דו-צדדיים (Huaqin A-share prospectus) primary filing + estimate. לקו היקף יש לכן בדרך כלל שתיים עד ארבע נקודות ביקורת, לא אחת.

המכונות כבר אוטומטיות מאוד: Huaqin מדווחת על 96% אוטומציה בקווי ה-SMT שלה (Huaqin H-share prospectus) single. האנשים הנותרים נמצאים בביקורת, תיקון, בדיקה חזותית, בדיקה והרכבה סופית, ואף חברה אינה מפרסמת כמה מבקרים יש לה לקו (פרק 7) source.

שלוש הערכות בלתי תלויות מסכימות על טקט של **13–22 שניות ללוח** (תיק סביבתי מ-2026 לחמישה קווים בחביי Hebei SMT project EIA), קווי המחשבים הניידים של Huaqin ומקרה של iFactory), או 7–11 שניות לצד עבור לוחות דו-צדדיים estimate. החלק הקשה של קו בתמהיל גבוה אינו הנפח אלא הכנסת מוצר חדש והחלפות, וכל החלפה היא עוד התנעה קרה למודל AI.

בצד התוכנה, תחנות הביקורת של יצרני ה-AOI עצמם כבר מרוכזות: תוכנת תחנת התיקון של Sinic-Tek מאפשרת "מחשב מארח אחד לשלוט בציוד של שמונה קווי SMT" (Sinic-Tek prospectus) single source, primary, 2022, ולאחר ש-Jabil פרסה שליטה מרחוק של ViTrox V-ONE ב-AOI ב-2026, מהנדס אחד יכול היה לנהל 5–8 AOI מרחוק לעומת 1–2 מקומית (AtlasPCB) vendor claim. מוצר ביקורת AI של צד שלישי מתחבר בדיוק לתחנה המרכזית הזו, דרך ייצוא התמונה והתוצאה של ה-AOI (IPC-CFX, IPC-HERMES-9852 או פורמט משל הספק). מכשירי ה-AOI עצמם מוסיפים AI במהירות: Yamaha אומרת שהשיפוט של ה-AI שלה לוקח כשנייה לקריאה, עוזר למפקחים אחרי 30 תמונות NG והופך לאיכות "גבוהה מאוד" אחרי 1,000 (DirectIndustry/Yamaha) vendor claim.

FIG 02 **קו SMT בייצור המוני: האנשים יושבים מאחורי ה-AOI*כוח העבודה מרוכז בעמדות הבדיקה לאחר בדיקה אופטית אוטומטית. שם נכנס ה-AI.***



מקור: דיווחים ציבוריים וחומר ספקים (Huaqin, DBG, Sinic-Tek ואחרים); קובץ על ידי המחבר.

שלב	מוצרים וטכנולוגיה נפוצים	ספקים מייצגים	עד כמה ה-AI הגיע
אחרי הדפסה	3D SPI (אור מובנה, מוארה, חלקית טריאנגולציית לייזר)	Koh Young, Sinic-Tek, TRI	שיפוט AI של יצרן ציוד מקורי על המשחה בייצור; Jutze אומרת שחלק מהדגמים הם "AI מלא"
אחרי מיקום והזרמה	2D/3D AOI עם נורות LED רב-זוויתיות רב-צבעיות ואור מובנה, 8–20 מיקרון לפיקסל, מצלמות 8–12 MP	Koh Young, Omron, Saki, TRI, ViTrox, Mirtec, Jutze, Sinic-Tek ואחרים	ביקורת AI של יצרן ציוד מקורי (Jutze בייצור, ViTrox V510Ai, Sinic-Tek עם Tencent); EMS מובילים בונים משלהם (DBG); תוספים של צד שלישי
חיבורים תחתיים ונסתרים	X-ray בדגימה או בקו (AXI)	—	Hikvision טוענת שהמודל Guanlan שלה שוגה בשיפוט חללי X-ray פחות מ-1% מהזמן vendor claim
ביקורת ותיקון	תחנת ביקורת מרכזית (מארח אחד, שמונה קווים), ביקורת מרחוק, תחנת תיקון	תוכנות יצרני AOI, ViTrox V-ONE	AI סוגר קריאות שווא, האנשים מקבלים את השאר
בדיקה חזותית, IQC, FQC	בעיקר ידנית; כמה מכונות מראה עם AI עם טעינת רובוט	SmartMore ואחרים	מכונות מראה בייצור; אין נתוני כח אדם ציבוריים ל-IQC/FQC
טיפול וטיפול במכונות	זרועות שישה צירים וקובוטים, SCARA, AMR, הנחיית ראייה 2D/3D	FANUC, ABB, Yaskawa, KUKA, UR, Dobot, JAKA, Elite; Mech-Mind, Hikrobot	תפיסה נלמדת בתוספת תנועה מתוכנתת בייצור; VLA ורובוטים דמויי אדם עדיין בפיילוט
נתונים ומערכות	MES, IPC-CFX, IPC-HERMES-9852	—	תוצאות נכתבות בחזרה ל-MES; התמיכה בציוד ישן אינה אחידה

מקורות לטבלה: התיקים לעיל ו-(Koh Young, Jutze 2025 annual report, Sinic-Tek 2025 annual report, ViTrox, Sohu on Hengtianyi, Xinhua on Hikvision).

האנשים יושבים מאחורי ה-AOI, כי רוב ההתראות הן שווא

שרשרת הבדיקה בקו SMT/PCBA מהזרם המרכזי רצה כך: 3D SPI אחרי הדפסה עבור משחה חסרה, לא מספקת או עודפת, גישור והיסט; AOI אחרי מיקום (אופציונלי) ואחרי הזרמה עבור חלקים חסרים, מוזזים, מסובבים, הפוכים או "מצבות" ונפח הלחמה; X-ray כשיש BGA או חלקים תחתיים אחרים; ולאחר מכן בדיקת in-circuit או בדיקת מחט מעופפת, בדיקה תפקודית ובדיקה חזותית ידנית. לכל AOI יש עמדת ביקורת ותחנת תיקון מאחוריו (Sinic-Tek prospectus, JLCPCB) . verified אחרי הזרמה הוא הנפוץ ביותר ובעל הערך הטוב ביותר, אבל הוא רואה רק צורה ושטח; חיבורים, חללים והרטבה מתחת ל-BGA דורשים X-ray או בדיקה חשמלית (JLCPCB).

כוח העבודה יושב בביקורת כי רוב ההתראות הן שווא. EMS בריטי אחד רשם **54,019 קריאות שווא מול 544 פגמים אמיתיים על פני 917 מספרי חלק, כ-99% מההתראות** (arXiv 2505.03576) single source, preprint ; במפעל ראשטאט של Siemens **97.5%** **מהלוחות שה-AOI דחה היו תקינים** (Siemens case study) vendor claim . קו עמוס רואה כ-448 קריאות שווא למשמרת ב-42–55 שניות כל אחת, 6.8–5.2 שעות למשמרת, בערך מבקר במשרה מלאה אחד (iFactory) vendor claim; hours are an estimate .

FIG 03 **כ-97–99% מהתראות ה-AOI הן קריאות שוואהיום אדם מאשר כל קריאה; בדיקת AI מנקה תחילה את השגויות ומעבירה את השאר לאנשים.**



מקור: מקרה בוחן של סימנס, iFactory (טענת ספק), דוח שנתי 2025 של DBG; שעות בדיקה הוערכו על ידי המחבר.

מתוך "שעות ביקורת למשמרת = לוחות × קריאות שווא ללוח × שניות לקריאה", מבקר אחד יכול לכסות כ-0.7–1 קו היקף עם AOI מבוסס-כללים. ברגע ש-AI מסיר 60–90% מקריאות השווא, זמן השיפוט בלבד היה מאפשר 2–14 קווים לאדם, אבל הליכה, טיפול במאגר NG ומסירות תיקון מגבילים את הכיסוי המקומי לשניים עד ארבעה קווים סמוכים; רק ביקורת מרכזית או מרחוק מתקרבת לתקרה התיאורטית estimate; seconds per call come from one source . מאחר ש-AOI מהזרם המרכזי סורקים 60–80 סמ"ר לשנייה (Yunheng 2026 AOI guide), צד של 15×20 ס"מ לוקח 4–5 שניות; ה-AI צריך לעבד רק את האזורים המסומנים המעטים בתוך טקט אחד, לא לסרוק מחדש את הלוח.

מה שכבר עובד בייצור: סיווג מפוקח על התמונות של ה-AOI עצמו

הדפוס המוכח הוא סיווג למידה עמוקה מפוקחת על תמונות מה-AOI שמפעל כבר מחזיק. התוסף של Fujitsu על ה-AOI של ADLINK בטאיוואן קיצץ קריאות שווא ב-**יותר מ-90%** (Fujitsu case); OKI משתמשת ב-AI מאז יולי 2026 כדי לקצץ אישור ידני אחרי AOI בכ-**80%** בעסקי ה-EMS שלה (Circuits Assembly); Siemens מדווחת על 60% פחות ניתוח ידני, €160,287 שנחסכו בשנה הראשונה והחזר של שמונה חודשים (Siemens case study) all vendor claims . שני דברים משותפים למקרים האלה: כמעט אף אחד לא מפרסם שיעור בריחות, וכולם משאירים אנשים על ההתראות הנותרות. המקרה היחיד שנותן נתון בריחות (0.8% ירד ל-0.3%, מול 0.5% חוזי) אינו עקבי פנימית (iFactory). AI מקצץ שעות ביקורת, לא את עבודת הביקורת עצמה.

ביקורת AI עברה מתוספים של צד שלישי לתכונות של יצרני ציוד מקורי ולמערכות פנימיות, וזה משנה את מה שמתחרה חדש נמדד מולו. הדוח השנתי של Jutze ל-2025 מפרט את "שדרוג ה-3D AOI" שלה כבייצור המוני ואומר שהוא "משתמש באלגוריתמי AI כדי להסיר ביקורת ידנית בקצה הבדיקה" (Jutze 2025 annual report); Sinic-Tek פיתחה "מערכת ביקורת ידנית מסייעת AI רב-מודאלית" מאז 2019 ומפתחת יחד עם Tencent 3D AOI עליון-תחתון, במטרה ל-99.5%+ כיסוי פגמים ו-98% דיוק זיהוי (Sinic-Tek 2025 annual report) vendor claims in primary filings . חברות EMS גדולות בונות משלהן. הדוח השנתי של DBG ל-2025 מפרט "שיפוט AI לבדיקה חזותית" כפרויקט שהושלם ש"מכסה יותר מ-95% מהרכיבים המסומנים על ידי AOI", עם "שיעור בריחות מתחת ל-20 PPM" ו"יותר מ-85% פחות ביקורת ידנית", ניתן למעקב מלא (DBG 2025 annual report) single source, primary . זהו הנתון היחיד ממפעל אלקטרוניקה שמצאתי שנותן כיסוי, בריחות והפחתה יחד. **מוצר צד שלישי שנמכר ל-EMS מדרגה ראשונה חייב להשתוות ל-95% כיסוי, 20 PPM בריחות ו-85% הפחתה של DBG**; מתחת לדרגה הראשונה, המדד הוא ההפחתה של 50–85% שספקים מפרסמים.

זיהוי אנומליות שאומן רק על חלקים תקינים הוא מהיר (EfficientAD-S לוקח 2.2 אלפיות שנייה לתמונה EfficientAD, WACV) 2024)) אבל חלש באנומליות "לוגיות" כמו חלקים חסרים או שגויים, והשיטות הטובות ביותר בממוצע מתחת ל-60% AU-PRO על ה-MVTec AD 2 הקשה יותר (MVTec AD 2) single source . חלקים חסרים ושגויים עדיף לבדוק מול CAD ו-BOM.

התקרה בכל מחיר: שכבות של מכונות, מעט מאוד אנשים, אבל לא אפס

לא מצאתי מפעל אלקטרוניקה שהסיר ביקורת ידנית לחלוטין והגיע לאפס בריחות. הפרקטיקה למוצרי IPC-A-610 Class 3 (אמינות גבוהה) עדיין משאירה למפקחים מאומנים להכריע את קריאות הגבול שספים אוטומטיים אינם יכולים לסווג, ומרכיב PCB אחד מכנה

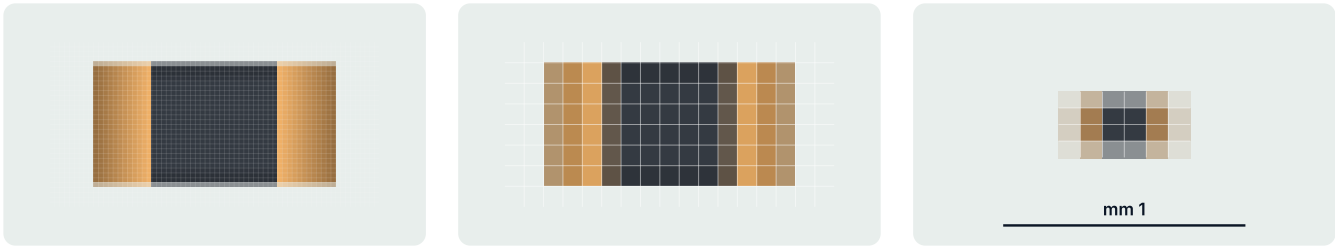
"בדיקה אוטומטית יכולה להחליף חתימה אנושית" כ"תפיסה שגויה נפוצה" (PCBCart) single source . גם קו הבסיס האנושי צריך תיקון: בהסתמך על מחקר של Sandia מ-2012, מפקח חזותי יחיד תופס במקרה הטוב כ-**80%** מהפגמים, ושניים בטור כ-96% (Instrumental) single source, cited by a vendor . הטענה הניתנת להגנה היא "AI בתוספת לולאת ביקורת אנושית קטנה מנצחת בדיקה ידנית מלאה", לא "AI מחליף בדיקה".

ללא מגבלת תקציב, התקרה היא 3D SPI, 3D AOI, X-ray, ביקורת AI ותחנות מראה עם AI בערימה אחת, כשאנשים נותרים רק לקריאות גבול, מוצרים חדשים והחלפות, סוגי פגמים חדשים וחתימת לקוח. מפעלי דגל מדווחים על 85–92% פחות בדיקה ידנית (I-Connect007, 2021, WiAdvance) vendor claims ; ההערכה שלי לתקרה ללא תקציב היא כוח עבודת ביקורת של 8–15% מהיום estimate . הרכבה סופית עדיין עתירת עבודה: אוטומציה של כפתורי ה-iPhone 16 בוטלה בגלל שיעורי פגמים גבוהים (9to5Mac) single source .

מצלמה כללית ברזולוציה גבוהה אינה AOI

AOI ייצוריים מצלמים אריח אחר אריח ב-8–20 מיקרון לפיקסל, עם מצלמות 8–12 MP, תאורה רב-זוויתית רב-צבעית ניתנת לתכנות ואור מובנה תלת-ממדי (Koh Young, Saki) verified . התעשייה ציינה מזמן שחלק 01005 מכסה רק כ-8×16 פיקסלים ב-25 מיקרון לפיקסל, "אולי לא מספיק" לשיפוט אמין (Circuits Assembly, 2007). באותו חשבון estimate , מצלמת 4K על פני לוח 300 מ"מ מקבלת כ-**78 מיקרון לפיקסל**: חלק 0402 אימפריאלי הוא כ-6×13 פיקסלים, מספיק לנוכחות והיסט גס אבל לא לצורת החיבור, וחלק 01005 הוא כ-3×5 פיקסלים, בלתי נראה למעשה. הגעה ל-15 מיקרון לפיקסל על פני לוח שלם הייתה דורשת כ-330 מגה-פיקסל, וזו הסיבה ש-AOI מזיזים מצלמה מעל הלוח. כשמישהו אומר "0402", בדוק לאיזה הוא מתכוון: 0402 מטרי הוא 01005 אימפריאלי, הרבה מעבר לכל מצלמה רחבה-זווית.

FIG 04 מצלמה לשימוש כללי אינה AOIמצלמת 4K מעל לוח 300 מ"מ נותנת כ-78 מיקרון לפיקסל: רכיב 0402 מצטמצם ל-≈6×13 פיקסלים ו-01005 נעלם.



מקור: Circuits Assembly; חישוב המחבר ללוח 300 מ"מ וחיישן 4K. רשתות הפיקסלים הן להמחשה.

בקווים שכבר יש בהם AOI, המקום הטוב ביותר ל-AI הוא על גבי התמונות של ה-AOI. רק קווים ללא AOI, או תחנות קצה קו, מצדיקים תחנת מצלמה סגורה עם תאורה מבוקרת; מצלמת רזולוציה גבוהה פתוחה מעל הקו היא החשופה ביותר לאור סביבה ולצללים (CSDN).

הרכבה וטיפול: תפיסה נלמדת בתוספת תנועה מתוכנתת נמצאים בייצור

הדוח של IFR ל-2026 אומר ש"ראיית מכונה לבדיקה ולהנחיית רובוטים היא בשלה וברמה תעשייתית", בעוד "AI פיזי למטרות כלליות ולמידה מסימולציה ומניסיון עדיין אינם ברמה תעשייתית" (IFR World Robotics 2026 executive summary) single source, . אלקטרוניקה היא קונה הרובוטים הגדול ביותר (142,235 יחידות שהותקנו בעולם ב-2025, 59% מכלל ההתקנות בסין (IFR, IFR China release) primary), ובכל זאת לסין יש רק 166 רובוטים לכל 10,000 עובדי ייצור לעומת 307 בארה"ב (IFR, April 2026). הדפוס בייצור הוא תפיסה נלמדת בתוספת תנועה דטרמיניסטית; תנועה נלמדת מקצה-לקצה קרובה לדרגת ייצור רק בתחנות מובנות (פרקים 3 ו-6).

מה כל אזור יכול להשתמש בו

סין היבשתית נשענת על מודלים מקומיים ומשקלים מקומיים; בטאיוואן הזהות של הלקוח מחליטה; בארה״ב, מכסים והסמכה מחליטים.

רכיב	סין היבשתית	טאיוואן (מפעלים פרטיים)	דרום-מזרח אסיה	ארה״ב
Claude (כל ערוץ)	לא זמין	מותנה: הספק לא חייב להיות בבעלות רוב או בשליטה של אנשים או ישויות בסין היבשתית, הונג קונג או מקאו	זהה	זהה
GPT ו-Gemini בענן	לא זמין	זמין	זמין (רשימות לא נבדקו מדינה אחר מדינה)	זמין
מודלי ענן סיניים (Qwen, DeepSeek, Doubao, GLM, Kimi)	זמין (מודלים מוגשים; מתן שירות דורש רישום)	חוקי, אבל אירוח מסין היבשתית לעיתים רחוקות עובר את בדיקת האבטחה של הלקוח; אסור במגזר הציבורי, במפעלים ממשלתיים ובתשתית קריטית	חוקי (צמתים בינלאומיים)	חוקי, בדיקה מחמירה
משקלים פתוחים מקומיים ממוצא סיני (Qwen, GLM, DeepSeek)	בחירה ראשונה; פריסה פרטית בתוך המפעל אינה דורשת הגשת מודל או אלגוריתם (קריאה של משרד עורכי דין)	זמין (מגזר ציבורי מוגבל)	זמין	מותנה: בדיקת מקור
משקלים פתוחים מקומיים ממוצא לא-סיני (למשל Gemma 4)	זמין	זמין	זמין	זמין
מודלי ראייה קלאסיים לא-גנרטיביים	זמין (מחוץ לכללי ה-AI הגנרטיבי)	זמין	זמין	זמין
Jetson Thor/T5000/T4000, DGX Spark	זמין: NVIDIA מסווגת אותם 5A992.c, אין צורך ברישיון אמריקאי למפעל אזרחי רגיל; ל-Spark יש SKU סיני; מחירי ערוץ גבוהים בכ-40–60% מהמחירון האמריקאי	זמין	זמין	זמין
RTX 5090, RTX PRO 6000, DGX Station	לא זמין: השניים הראשונים הם 4A090.a, DGX Station 5A992.z; חלופות חוקיות הן 4090D, 5090 D V2, RTX PRO 5000 72GB, RTX 6000D, L20 (4A994.l)	זמין	זמין, אבל קונים שמשרדיהם הראשיים בסין או בבעלות סופית ממנה נחשבים סיניים; מלזיה דורשת היתרים ליצוא ולהעברה	זמין
Mac Studio M5 Ultra 512 GB	זמין (חנות Apple סין)	זמין	ככל הנראה זמין	זמין
Huawei Ascend ומאיצים סיניים אחרים	זמין	סיכון גבוה	סיכון גבוה	לא שמיש בפועל
זרועות קבועות, AOI, מצלמות תעשייתיות מתוצרת סין היבשתית	זמין	ניתן לייבוא; מאז 1 בדצמבר 2025 הרשויות יכולות לפעול ישירות על מותגים מסין היבשתית, וגופים, מפעלים ממשלתיים, קרנות מסוימות וספקי תשתית קריטית יכולים להיות מוגבלים או מנועים משימוש בהם	זמין	חוקי, אבל המכס המשולב הוא כ-40% על זרועות, 37.5% על מכשירים ומצלמות מסוג AOI, 25% על מחשבי קצה
AMR ומניפולטורים ניידים מתוצרת חוץ	זמין	זמין	זמין	מודלים חדשים דורשים אישור מותנה ממשרד המלחמה האמריקאי

רכיב	סין היבשתית	טאיוואן (מפעלים פרטיים)	דרום-מזרח אסיה	ארה"ב
מצלמות Hikvision ו-Dahua	🟢זמין	🟡שימוש פרטי מותר, מגזר ציבורי אסור	🟢זמין	🔴אישורי FCC חדשים חסומים; אסור לקבלנים פדרליים
בטיחות רובוט והסמכה	🟡GB 11291.1-2011 ו-GB 11291.2-2013 חובה, בעדכון מאז אפריל 2026; הסמכת CR וולונטרית; GB/T 39404-2026 מכסה אבטחת סייבר של בקרים	🟡גידור כברירת מחדל; שיתוף פעולה ללא גדר חייב לעמוד ב-CNS 14490 או ISO 10218 עם הערכה בכתב	🟡לפי לקוח ויעד יצוא, מאושר בכתב לפני משלוח	🟡ציוד חשמלי חייב להיות "מאושר" (בפועל סימון NRTL או הערכת שטח); UL 1740 דורש מתאי רובוט לעמוד ב-ISO 10218-2 מאז 11 באוגוסט 2025
מפעלי שבבים ומפעלי מוצרים בפיקוח	—	🟡חוזי הרכישה של TSMC דורשים SEMI E187; ציוד מסין היבשתית מוסר ממפעלי 2 ננומטר	—	🔴נתונים טכניים של ITAR כוללים "צילומים"; מדיניות שלילה לסין; ספקים שבשליטת סין היבשתית אינם יכולים בפועל לשרת מפעלי אלקטרוניקת הגנה

מקורות לטבלה: Anthropic supported regions, OpenAI supported countries, Gemini availability, CAC interim measures on generative AI, NVIDIA export classification data, eCFR 15 CFR 774, NVIDIA China DGX Spark, Apple China, Covington on new BIS rules, Greenberg Traurig, June 2026, Fulcrum, Tom's Hardware on Ascend guidance, Taiwan BOFT import list, Taiwan Cyber Security Management Act history, Taiwan review rules for products endangering national cybersecurity, Taipei Times, USITC HTS chapter 99, Latham & Watkins, IPVM, China national standards platform, GB 11291.2, Intertek UL 1740 update, eCFR 22 CFR 126.1).

סין היבשתית: ערימה שלמה, אבל לא מודלי חלוץ סגורים ולא המכונות הבודדות החזקות ביותר יכולים להיכנס. מודלי ענן מקומיים מוגשים, משקלים פתוחים מקומיים (Qwen תחת Apache-2.0, GLM ו-DeepSeek תחת MIT), מודלי ראייה קלאסיים ו-AOI, מצלמות וקובוטים מקומיים — כולם שמישים. סיווגי הייצוא של NVIDIA עצמה מפרטים את ערכת Jetson AGX Thor, את מודולי T5000/T4000 ואת DGX Spark כ-5A992.c, כך שמפעל EMS או PCBA רגיל אינו זקוק לרישיון אמריקאי, בעוד RTX 5090 ו-RTX PRO 6000 (4A090.a) אינם ניתנים לרכישה בסין היבשתית (NVIDIA export, eCFR classifications, RS China) . primary; the legal reading is mine and not legal advice מפרטת את ערכת Thor ב-RMB 59,104 כולל מס, בערך פי 1.4 ממחיר המחירון האמריקאי (RS China).

ארבעה משרדי עורכי דין מסכימים שפריסה פרטית בתוך המפעל לשימוש עצמי של המפעל אינה דורשת הגשת מודל או אלגוריתם, בעוד שירות מתארח רב-דיירים יושב באזור אפור (Covington, FPF, JunHe) . verified: consistent readings, no official ruling GB 11291.1-2011 ו-GB 11291.2-2013 נותרו קו הבסיס המחייב לבטיחות רובוטים ונמצאים בעדכון מאז 28 באפריל 2026; GB/T 39404-2026 על אבטחת סייבר של בקרים נכנס לתוקף ב-1 בנובמבר 2026 (GB/T 39404-2026, revision plan 20262542) . single source, primary עד שה-GB 11291 החדש מגיע (2028–2029 estimate), תכנן ל-ISO 10218:2025 ושמור תיקי תאימות ל-GB 11291.

טאיוואן: מפעלים פרטיים יכולים להשתמש כמעט בכל דבר; הזהות של הלקוח מחליטה. ממשקי API של מודלי חלוץ, משקלים פתוחים וציוד מסין היבשתית — כולם חוקיים (data.gov.tw). אבל מאז שחוק ניהול אבטחת הסייבר המתוקן נכנס לתוקף ב-**1 בדצמבר 2025**, הרשויות יכולות לפעול ישירות על מוצרים שהם "מותגי היבשת" ולהגביל אותם בגופים ממשלתיים, מפעלים ממשלתיים וספקי תשתית קריטית (Laws & Regulations Database, review rules) verified; רשימת ספקי התשתית הקריטית חסויה; ו-TSMC כתבה את SEMI E187 לתוך חוזי הרכישה שלה (CNA) ודווח שהיא מסירה ציוד מסין היבשתית ממפעלי ה-2 ננומטר שלה (TrendForce) . single source מפעלי EMS, הרכבה ו-PCBA קטנים פרטיים פתוחים; מפעלי שבבים, שירותי פארק ולקוחות הקשורים לממשלה סגורים בפועל. ספק שבשליטת סין היבשתית צריך לתכנן ל-E187, לספק רשימת חומרי תוכנה, להציע עדכונים במצב לא מקוון ולעולם לא להתחבר מרחוק מסין היבשתית.

דרום-מזרח אסיה: הכי מעט הגבלות; הבעיות הן החזר איטי וחברות אם סיניות. הכול שמיש, אבל הנחיית ארה"ב מ-31 במאי 2026 דורשת רישיון לייצוא פריטי מחשוב מתקדמים לישויות שמשרדיהן הראשיים בסין, או בבעלות סופית ממנה, גם כשהנמען נמצא במדינה שלישית (Greenberg Traurig) single source, ולמלזיה יש היתרים משלה לשבבים מתקדמים (Fulcrum). הביקוש נע לכאן: TrendForce צפה שתאילנד, וייטנאם ומלזיה יתקרבו ל-50% מכושר ה-SMT של שרתי ODM ב-2026 (TrendForce) single source, a 2023 forecast .

ארה"ב: הערימה המלאה, במחיר מכסים, הסמכה ונתונים בפיקוח. תחת פרק 99 של לוח התעריפים האמריקאי, זרוע רובוטית מתוצרת סין היבשתית משלמת כ-40% במכס משולב, מכשירים ומצלמות מסוג AOI כ-37.5% ומחשבי קצה 25%; מתוצרת מלזיה או אינדונזיה כ-10–12.5%, מווייטנאם, תאילנד או הפיליפינים 12.5–15%, מטאיוואן 10%, עם מחשבים ב-0% מחוץ לסין היבשתית

(USITC HTS chapter 99, Great Lakes Customs Law) primary schedule + estimate . ההודעה של USTR מ-10 בנובמבר האריכה 178 פטורים ישנים של סעיף 301 עד 9 בינואר 2027, אף אחד מהם לא מכסה רובוטים, מכשירי בדיקה, מצלמות או מחשבים (Federal Register 2026-22925) verified .

OSHA דורשת שציוד חשמלי במקום העבודה יהיה "מאושר" (בפועל סימון NRTL, או הערכת שטח של כ-$3,000 עד $20,000+ לתא), ומאז 11 באוגוסט 2025 UL 1740 מכניס תאי רובוט שלמים תחת ISO 10218-2 (eCFR 1910.399, Intertek, Industrial Monitor Direct). אלקטרוניקת הגנה סגורה בפועל לספקים שבשליטת סין היבשתית: ITAR מתייחס לצילומי ייצור כנתונים טכניים ומפעיל מדיניות שלילה לסין (eCFR 22 CFR 120.33, eCFR 22 CFR 126.1). תא המיועד לארה"ב צריך להשתמש בזרוע קבועה, מצלמות בלתי מוגבלות ומודלים מקומיים או בענן אמריקאי, עם הרכבה סופית בדרום-מזרח אסיה היכן שזה באמת משנה מקור.

בכל תקציב, התקרה שונה לפי אזור אבל בשום מקום אינה מגיעה לאפס אנשים estimate . סין היבשתית יכולה לערום 3D SPI מקומי, 3D AOI, X-ray, ביקורת AI וצוות AI פנימי כדי להוריד את כוח עבודת הביקורת ל-8–15%, אבל ללא מודלי חלוץ סגורים או DGX Station, שאלות-ותשובות LLM מפעליות חייבות לרוץ על משקלים פתוחים מקומיים ומחשוב מקומי (מודל בגודל מלא למשתמשים רבים דורש לפחות שני שרתי Ascend בני שמונה כרטיסים). טאיוואן ודרום-מזרח אסיה יכולות מבחינה טכנית להשתמש בטוב מכל דבר, כולל ממשקי API של מודלי חלוץ ו-DGX Station, מוגבלים על ידי זהות הלקוח, מותגי היבשת ורישוי חברות אם סיניות. לארה"ב אין מגבלות טכניות, במחיר מכסים, הסמכת NRTL והפרדת כוח אדם לנתונים בפיקוח. בכל מקום, אנשים נותרים לקריאות גבול, מוצרים חדשים והחלפות, סוגי פגמים חדשים וחתימת לקוח.

מוצר שנמכר על פני אזורים זקוק לכן לשתי ערימות מודלים: ערימת היבשת על מודלים מקומיים מוגשים או משקלים פתוחים מקומיים, וערימה לחו"ל על ממשקי API של מודלי חלוץ בתוספת מודל מקומי ממוצא לא-סיני. המדיניות של Anthropic, בתוקף מ-12 בנובמבר 2026, שופטת לפי בעלות ושליטה, כך שספק שבשליטת סין היבשתית אינו יכול להיבנות על Claude אפילו דרך חברת בת בחו"ל, פלטפורמת ענן או משווק (Anthropic) single source, primary .

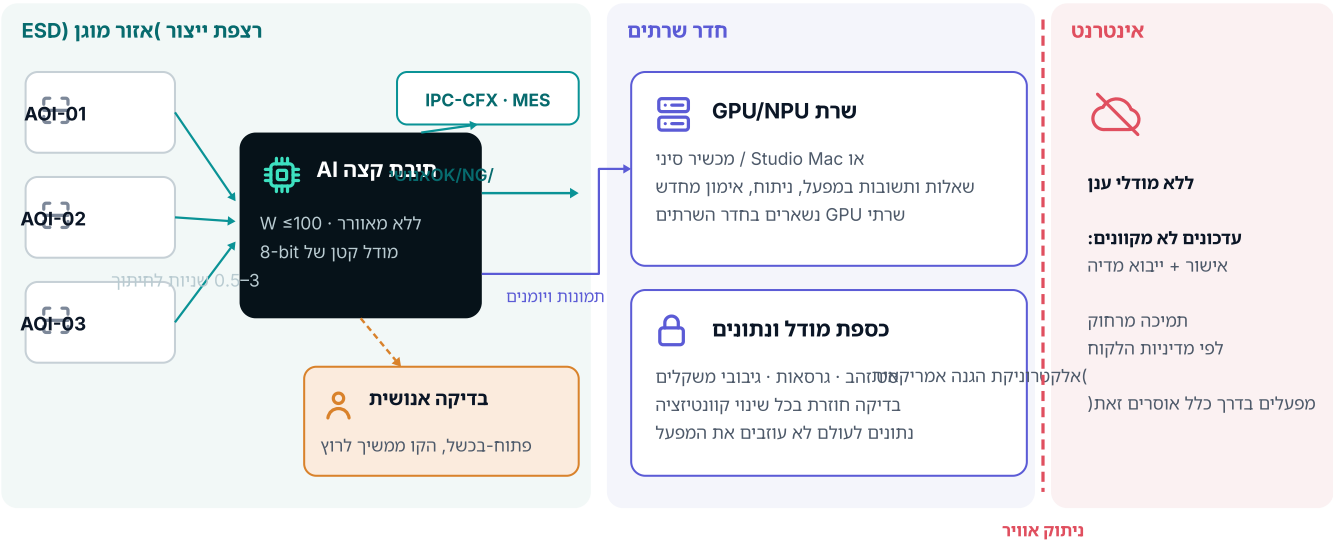
פריסה מקומית

02 הרצת AI בתוך המפעל: חדרים נקיים, ניתוק אוויר, חומרה וקוונטיזציה

במפעל מנותק-אוויר האילוץ הוא תהליך, לא חישוב

בדיקת ראייה מנותקת לגמרי מהרשת היא עניין שבשגרה; האילוצים הם פרוצדורליים. המדריך של MIIT בסין מ-2024 על אבטחת מערכות בקרה תעשייתיות מבקש אזוריות, חומות אש או דיודות נתונים, תחזוקה מרחוק מבוקרת ומוגבלת בזמן, מדיה סרוקה ותיקונים שנבדקו (guide text) ; לקוחות מוליכים למחצה מוסיפים SEMI E187 ו-E188, ש-TSMC כתבה לחוזי הרכישה שלה ב-2023 (SEMI, TXOne) . תוצאות חוזרות ל-MES דרך IPC-CFX וזהות הלוח נעה דרך Hermes) IPC-HERMES-9852 (standard. המלכודת הנפוצה היא רישוי שמניח רשת: מצב ה-offline של Roboflow שומר משקלים במטמון על חוזה של 30 יום (Roboflow). מוצר לא מקוון צריך הפעלה לא מקוונת, עדכונים חתומים על מדיה סרוקה, רשימת חומרי תוכנה, בקרת USB, יומני ביקורת מקומיים ומתאם CFX.

FIG 05 ***פריסה מנותקת-אוויר: תיבה קטנה בקו מחליטה; חדר השרתים עונה על שאלות; תקלות נופלות חזרה לאנשים והנתונים לעולם לא עוזבים את המפעל.***



מקור: סינתזת המחבר (פרק 2).

מודלי חלוץ סגורים אינם נכנסים בפועל לחדר השרתים של מפעל. Gemini מוצע מנותק-אוויר רק דרך Google Distributed Cloud לממשלות וארגונים גדולים (Google Cloud); Foundry Local של Microsoft פותח מודלים גדולים רק ל"לקוחות מוסמכים" ואינו נוקב בשם אף מודל Claude); GPT (Microsoft) מוצע רק באזורי ענן ממשלתיים, ללא פריסה מקומית (Claude for government); אף אחד מהשלושה אינו מוצע בסין היבשתית. ציוד מפעל חייב להיות מתוכנן סביב משקלים פתוחים; מודלי חלוץ יכולים לכל היותר לתייג או להסביר תמונות במטה, לא מקוון, כשהלקוח מאפשר הוצאת תמונות אנונימיות.

חומרת חדר שרתים צריכה להיות מטופלת כדרגת משרד: ה-Mac Studio מדורג ל-10–35 מעלות צלזיוס ו-DGX Spark ל-5–30 מעלות, עם דיווחים על כיבוי תרמי תחת הסקה מתמשכת (Apple, NVIDIA forum) . הקו מקבל תיבות תעשייתיות ללא מאוורר, כמו מחשב תעשייתי Jetson T4000 המדורג ל-25- עד 60 מעלות צלזיוס (AAEON) . מכיוון ששעת השבתה עולה למפעל גדול $36,000 עד $2.3 מיליון (Siemens True Cost of Downtime 2024) , כל צעד AI על הקו חייב **להיכשל-ולפתוח**: בכל תקלה הוא נופל חזרה לתהליך ה-AOI-בתוספת-אנשים הקיים.

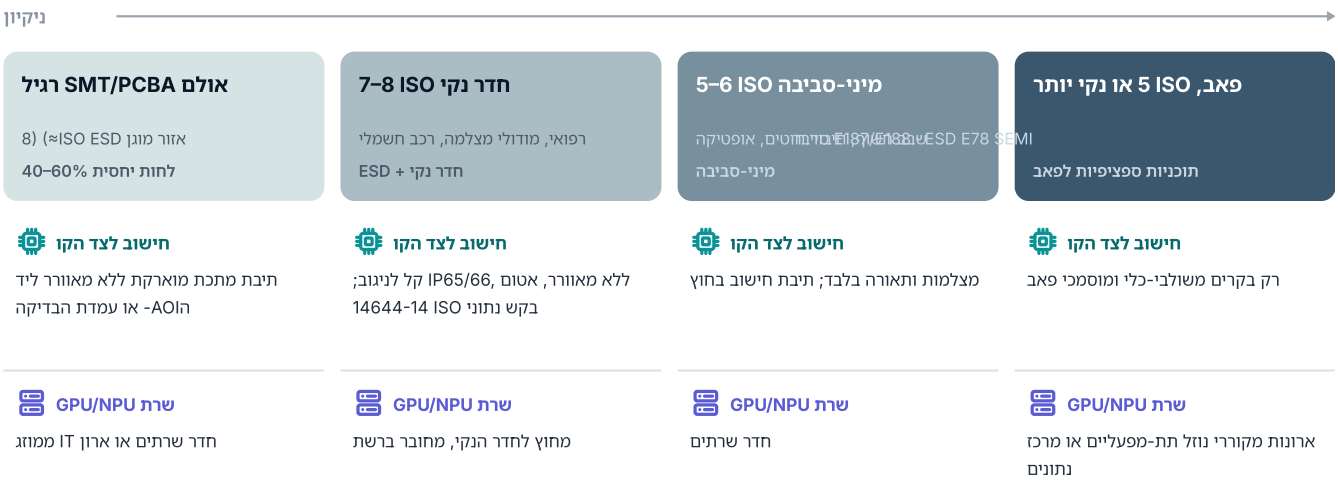
אזורים נקיים ואזורי ESD: אולם SMT רגיל הוא אזור מוגן ESD, לא חדר נקי מדורג

מה שמגביל היכן החישוב יכול לשבת הוא חשמל סטטי וזרימת אוויר, לא דרגת ניקיון. רוב אזורי הרכבת ה-SMT הם "לא מבוקרי חלקיקים פורמלית", ומפעלים ברמה עולמית "עשויים להגיע או להתקרב ל-PCBA ISO 8" (Circuitnet expert panel) single source. רפואי ומודולי מצלמה יושבים בדרך כלל ב-ISO 7, שבב חשוף ויישור אופטי ב-ISO 5–6 מקומי (PCBSync) single source; בוני חדרים נקיים מצטטים ISO 7–8 ל-SMT בכלל (Angstrom Technology) vendor claim.

הדרישה האוניברסלית היא בקרת ESD: ANSI/ESD S20.20-2021 דורש מוליכים מבודדים מתחת ל-35 וולט ומרחיק מבודדי שדה חזק לפחות 30 ס"מ מהתקנים רגישים (ANSI, ESDA, Desco) verified. מפעלי שבבים שמים חישוב בתת-הפאב או במרכז הנתונים, בארונות מקוררי נוזל שעומדים ב- SEMI S2/S8/F47 (Prodrive) vendor claim.

האם ציוד יכול להיכנס לאזור נקי נבדק על פליטת חלקיקים תחת ISO 14644-14; לרובוטים יש תעודות כאלה (Fraunhofer IPA), אבל לא מצאתי מחשב תעשייתי, תיבת edge AI או שרת GPU עם אחת. CE, FCC ו-IEC 60000 מראים שתיבה סובלת את הסביבה שלה, לא שהיא רשאית להיכנס לאזור מוגן ESD; רכז ה-ESD של הלקוח בודק את זה. שרתי GPU לעולם לא עולים על הקו: DGX Station משך 1,292 וואט ב-GPU בבדיקה (StorageReview).

FIG 06 **אזורים נקיים ואזורי ESD קובעים היכן יכול לשבת החישוב*אולם SMT רגיל הוא אזור מוגן ESD, לא חדר נקי. שרתי GPU ו-NPU שייכים לחדר השרתים.***



מקור: ISO 14644, IEC 61340-5-1, ANSI/ESD S20.20, תקני SEMI וחומר ספקים; קובץ על ידי המחבר.

אזור	סביבה אופיינית	חישוב ליד הקו	שרתי GPU/NPU
אולם SMT/PCBA רגיל (צרכני, תעשייתי)	אזור מוגן ESD תחת S20.20 או IEC 61340-5-1, בדרך כלל לא חדר נקי מדורג (≈ISO 8 שווה-ערך); לחות יחסית 40–60%	תיבה ללא מאוורר במארז מתכת מוארק (Jetson T4000/Thor, תיבת NPU סינית) ליד ה-AOI או עמדת הביקורת; צגי פלסטיק ומארזים מחוץ לאזור או ≥30 ס"מ מהלוחות	חדר שרתים או ארון IT ממוזג
רפואי, מודולי מצלמה, אלקטרוניקת רכב חשמלי (חדר נקי ISO 7–8)	חדר נקי בתוספת הגנת ESD	ללא מאוורר, אטום IP65/66, חלק וקל לניגוב; בקש נתוני ISO 14644-14 או תעודת IPA	מחוץ לחדר הנקי, מחובר ברשת פנימה
שבב חשוף, חיבור חוטים, תחנות אופטיות (ISO 5–6 מקומי)	מיני-סביבה	רק מצלמות ותאורה בפנים; תיבת החישוב בחוץ	חדר שרתים
חדר נקי של פאב (ISO 5 ונקי יותר)	תוכניות ESD ומשיכה אלקטרוסטטית ספציפיות לפאב, ציוד ל-SEMI E78; אבטחת סייבר ל-SEMI E187/E188	רק בקרים משולבי-כלי ומוסמכי פאב; אין מחשבים נוספים עם מאוררים	ארונות מקוררי נוזל תת-מפעליים או מרכז הנתונים של הפאב

הטבלה היא הסינתזה שלי של המקורות לעיל estimate. עבור מוצר, זה אומר תיבת צד-קו ללא מאוורר מתחת לכ-100 וואט, מארז מתכת מחובר להארקת מגן, כניסת 12–24 וולט DC, פעולה בטווח טמפרטורות רחב בזרימת אוויר אמיתית של ארון, ודוח פליטת חלקיקים ISO 14644-14 מוכן ללקוחות ב-7 ISO ונקי יותר. כל תכנון שבו "שרת GPU אחד משרת ארבעה קווים" שם את השרת בחדר השרתים ומשאיר רק תיבה קטנה או מצלמה ליד הקו.

מגבלות מהירות לפי שימוש: דיוק קשה יותר ממהירות

מפעלים וספקים כמעט אף פעם לא מפרסמים יעדי השהיית AI או זמינות. הספים למטה נגזרים מטקט, זמן ביקורת אנושי, קצבי בקרת רובוט ומחקר אינטראקציה estimate:

שימוש	מגבלת מהירות	מגבלת דיוק	זמינות
בדיקה בקו של כל חלק	בתוך ½–1 טקט: 240–1,000 חיתוכים בטקט של 30–10 שניות, 8–100 חיתוכים לשנייה	בריחות לא גרועות יותר מ"AOI בתוספת אנשים" של היום (דוגמת חוזה <0.5%)	פתוח-בכשל, לעולם לא עוצר את הקו; יעד ≥99.9% (אין SLA ציבורי)
חוות דעת שנייה בקו רק על חיתוכים שמסומנים על ידי AOI	סימונים (בדרך כלל ≥5–20 ללוח) בתוך ½ טקט: 0.5–3 שניות לחיתוך	זהה; מאומן על התמונות של הקו עצמו	נופל חזרה לביקורת אנושית
ביקורת AI של קריאות שווא	2–5 שניות לקריאה, מהיר פי עשרה ויותר מ-42–55 השניות של אדם	בריחות לא גרועות יותר ממבקר אנושי	התדרדרות מוסיפה אנשים, לעולם לא עוצרת את הקו
בקרת רובוט (VLA)	100–300 אלפיות שנייה למקטע פעולה עם חלוקה בזמן אמת, פלט 10–50 הרץ	≥99% הצלחת משימה כדי לנצח תנועה מתוכנתת	מערכת בטיחות בלתי תלויה במודל
תכנון רובוט ובחירת מיומנות	1–3 שניות כשצריך בכל מחזור; ≥1 דקה אם רק בחריגים או בהחלפות	הפלט חייב להיות קוד בר-אימות או מזהה מיומנות	נופל חזרה לתוכנית הקבועה
שאלות-ותשובות (SOP, שאילתות MES)	טוקן ראשון באופן אידיאלי ≥1–2 שניות, לכל היותר 10 שניות; ≥15–20 tokens/s למשתמש; 5–20 משתמשים מקבילים	ציטוטים מדויקים; נאמן על פני הקשר ארוך	דרגת IT משרדי (≈99–99.5%)
ניתוח ודוחות סיבת שורש	דקות הן בסדר (אצווה)	הגבוה ביותר; נאמן על פני הקשר ארוך	דרגת IT משרדי

Physical Intelligence מדדה 97 אלפיות שנייה של הסקה ל-VLA מסוג 3B (כ-108–139 אלפיות שנייה עם רשת ועיבוד מקדים), מצאה שחלוקה בזמן אמת שומרת על התפוקה ללא שינוי עד 200 אלפיות שנייה של עיכוב נוסף, והשלימה משימות עדינות כמו הדלקת גפרור או חיבור כבל רשת ביותר מ-300 אלפיות שנייה (Physical Intelligence) vendor research; מחקר אינטראקציה מציב את הגבול לשמירת רצף מחשבה בשנייה אחת ולקשב בעשר (NN/g). הדיוק הוא החלק הקשה: Qwen3-VL-8B פתוח עומד בממוצע על 56.8 zero-shot בלבד על פני חמישה מערכי נתוני אנומליה תעשייתיים (IndusAgent) single source. אין חומרה שמצילה מודל ראייה-שפה שלא כוונן על התמונות של הקו עצמו.

FIG 07 ***מגבלות מהירות לפי שימושבדיקה בקו נמדדת באלפיות שנייה, שאלות-ותשובות בשניות, ניתוח בדקות. אנשים לוקחים 42–55 שניות לקריאה.***



מקור: סינתזת המחבר (פרקים 1–2).

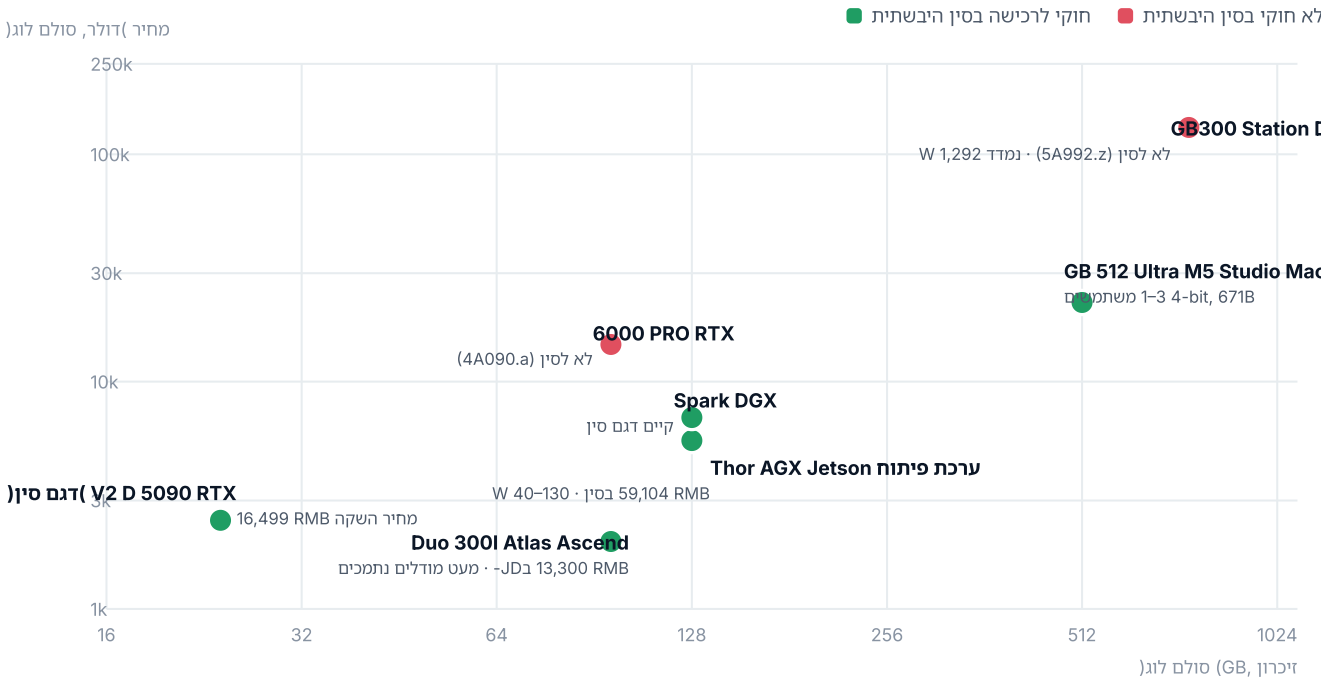
חומרה: Thor לתחנה, Mac Studio או מכשיר מקומי לשאלות-ותשובות במפעל, ואין DGX Station בסין היבשתית

מכשיר	זיכרון ורוחב פס	מחיר (נובמבר 2026)	למה הוא טוב	סין היבשתית
ערכת פיתוח Jetson AGX Thor / T5000 / T4000	128 GB LPDDR5X, 273 GB/s; 40–130 W	$3,499 בהשקה → $5,499 (יולי 2026); T5000 $2,999, T4000 $4,999; ערכת פיתוח RS China ב-59,104 RMB כולל מס	מודלי ראייה 2–8B שבודקים מחדש חיתוכים; VLA מסוג 3B בזמן אמת	כן (5A992.c, אין רישיון למפעלים אזרחיים)
DGX Spark	128 GB, 273 GB/s	$3,999 → $4,699 → $6,950; דגם סין ASUS GX10 על אותו שבב ב-RMB 32,999 ב-JD (אוק' 2025)	MoE רב-משתמשים 8–35B	כן (5A992.c, דגם סין, ערוצים מורשים)
RTX PRO 6000	96 GB GDDR7	≈$8,565 → $13,250 (יוני 2026) → ≈$16,005 מחירון eBay ממוצע בנובמבר	27–35B FP8 רב-משתמשים; MoE מסוג 120B ב-4-bit	לא (4A090.a)
Mac Studio M5 Ultra 512 GB	512 GB מאוחד, 1.2 TB/s; 480 W מרבי מתמשך	מ-$22,299 בארה"ב; ≈RMB 173–186k מוערך בסין	DeepSeek 671B ב-4-bit עבור 1–3 משתמשים	כן
DGX Station GB300	748 GB (252 GB HBM3e ב-7.1 TB/s + 496 GB LPDDR5X ב-396 GB/s); 1,292 W נמדדו ב-GPU	≈RMB 123k–140k מחיר רחוב בנובמבר; 95$–117k לפני מכס; 830–940k באוגוסט	עומס גבוה על 120–235B MoE שנכנס ב-HBM; רק כמה משתמשים על מודלי 433–743B	לא (5A992.z, TPP 58,484)
Huawei Atlas 300I Duo	96 GB LPDDR4X, 408 GB/s	13,300 RMB מסוחר JD	ביקורת אצווה בנפח נמוך; מעט מודלים נתמכים	כן (מקומי)
Huawei Atlas 800I A2 (8 × Ascend 910B 64 GB)	8 × 64 GB HBM	לא הושג; DeepSeek מלא דורש שניים ב-W8A8, ארבעה ב-BF16	שאלות-ותשובות וניתוח רב-משתמשים מפעליים	כן (מקומי)

מכשיר	זיכרון ורוחב פס	מחיר (נובמבר 2026)	למה הוא טוב	סין היבשתית
RTX 5090 D V2 (דגם סין)	24 GB	16,499 RMB בהשקה (אוג׳ 2025), עלה מאז עם מחירי הזיכרון	מודלי ראייה 8B, VLA קטנים	כן

מקורות: NVIDIA China Thor, NVIDIA Thor launch, CNX on price rises, NVIDIA export classification data, RS China, Implicator, IThome on ASUS GX10, Tom's Hardware RTX PRO 6000, GPU Poet, Apple specifications, TechRadar, IThome on M5 Ultra China pricing, Wccftech, StorageReview, CyberPowerPC, monitors.com, Tom's Hardware DGX Station, vLLM-Ascend support matrix, China Telecom Cloud, Huawei Atlas 800I A2, TweakTown). מחירי סין היבשתית ל-Mac Studio בנפח 512 GB מוקצנים מפרמיית התצורה האמריקאית והם בעלי ביטחון נמוך estimate.

FIG 08 **חישוב מקומי: זיכרון, מחיר ומה ניתן לרכוש באופן חוקי בסין היבשתית*סיווגי יצוא קובעים את הרשימה הקצרה לפני המדדים.***



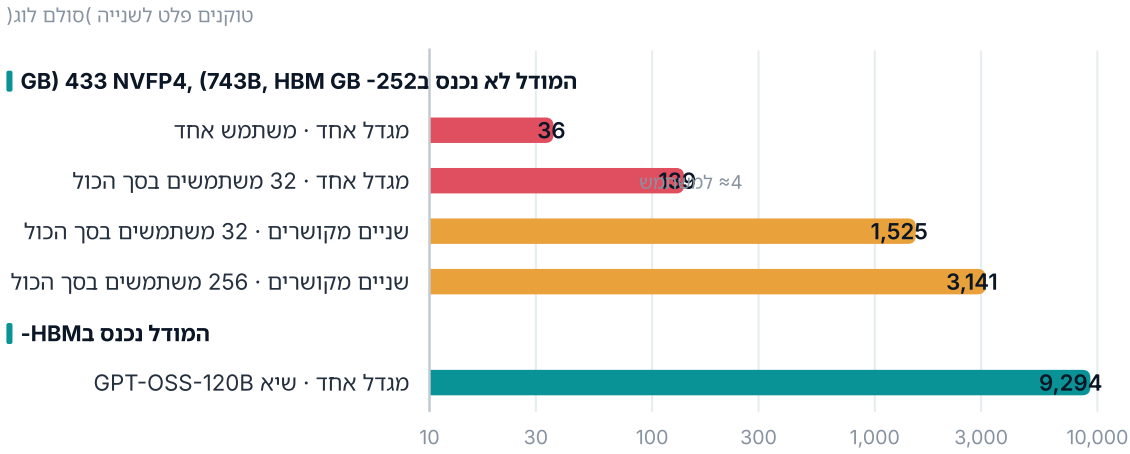
מקור: סיווגי יצוא של NVIDIA, Apple, RS China, JD.com ומחירי ביקורת (נובמבר 2026).

Jetson Thor הופך פריסה ברמת תחנה למעשית. בדיקות TensorRT Edge-LLM של NVIDIA מראות Qwen3-VL-8B (NVFP4) שממלא מראש 2,048 טוקנים ב-104.5 אלפיות שנייה ומפענח 44.4 tokens/s על Thor (TensorRT Edge-LLM) vendor claim; עבור חיתוך AOI ותשובה של 20 טוקנים זה כ-0.5 שניות עם המודל 8B ו-0.2 שניות עם ה-2B, או 8.5–21 חיתוכים לשנייה באצווה estimate. אורך הפלט, לא מספר הפרמטרים, קובע את ההשהיה, כך שפענוח מוגבל לטוקן מחלקה או JSON קצר הוא חלק מהתכנון. Thor (Jetson AI Lab) vendor claim לוקח כ-1.8 שניות לחיתוך ו-Atlas 300I Duo עם GR00T N1.7 (3B) רץ ב-25.1 הרץ על Orin NX. p95 מעל 23 שניות, מתאים רק לעבודת אצווה (CSDN) single source. בסין היבשתית, תקצב מוצרי NVIDIA ב-40–60% מעל המחירון האמריקאי estimate.

Mac Studio של 512 GB מאפשר מודל בגודל מלא על מכונה אחת, לכמה משתמשים. ה-M5 Ultra (אוגוסט 2026) מציע עד 512 GB ב-1.2 TB/s החל מ-$22,299 (Apple, TechRadar); DeepSeek 671B ב-4-bit (≈380–404 GB) מפענח 17–20 tokens/s על M3 Ultra (Pinggy) verified ומוערך ב-25–29 על ה-M5 Ultra. הנחיות ארוכות הן נקודת החולשה: הנחיית שליפה של 10,000 טוקנים לוקחת בערך 17–70 שניות estimate, מעבר לגבול האינטראקציה של 10 שניות.

DGX Station הוא המכונה הבודדת החזקה ביותר, אבל לא זמין בסין היבשתית ופי חמישה עד שישה ממחירו של Mac. ה-GB300 הנשלח כולל **748 GB**, שמתוכם רק 252 GB הם HBM מהיר (Wccftech, Tom's Hardware, Aug 2026) verified. בדיקות בלתי תלויות מראות כ-9,300 טוקני פלט לשנייה על GPT-OSS-120B, שנכנס ב-HBM, אבל מודל 743B שגולש ל-LPDDR5X נתן ל-32 משתמשים רק 139 tokens/s בסך הכול, כ-4 לכל אחד; שתי יחידות מקושרות הגיעו ל-1,525 (StorageReview, Aug 2026, StorageReview, Oct 2026). מחירי הרחוב עלו ל-$123k–140k עד נובמבר single source. בסין היבשתית, האפשרויות הריאליות לאותה עבודה הן ה-Mac Studio, שרתי CPU+GPU (CyberPowerPC, monitors.com) בסגנון KTransformers (KTransformers) vendor test או שרתי Ascend מקומיים.

FIG 09 **תחנת עבודה גדולה אחת אינה יוצרת עוזר משותף כשהמודל לא נכנס בזיכרון ברוחב פס גבוה, *32 משתמשים מקבילים מקבלים כ-4 טוקנים לשנייה כל אחד.***



מקור: StorageReview (אוג׳ ואוק׳ 2026); ServeTheHome. בדיקות עצמאיות ממקור יחיד.

מחסור בזיכרון הזיז את אותם חלקים ב-20–100% בשנה (DGX Spark עלה מ-$3,999 ל-$6,950 (TechRadar)), כך שהצעות מחיר לציוד צריכות להיות צמודות למדד ולא קבועות.

מאיצים מקומיים מריצים מודלים בגודל מלא; תחזוקת התוכנה היא צוואר הבקבוק. ל-Ascend 910B יש 64 GB של HBM, אבל vLLM-Ascend מסמן את DeepSeek V3/R1 כ"לא מתוחזק עוד" ותומך רק בכמה מודלים על כרטיסים קטנים יותר (vLLM-Ascend) single source, primary , וכל מודל חדש לוקח חודש עד חודשיים להתאמה (Caijing/Anue); שלח גרסת מודל קפואה ומאומתת. הגשת DeepSeek 671B מלא ל-20–100 משתמשים דורשת לפחות שני שרתי Atlas 800I A2 בני שמונה כרטיסים (China Telecom Cloud) verified , ובכל זאת מכרזים ציבוריים למכשירי LLM בסין הסתכמו רק ב-RMB 27.5 מיליון בינואר–ספטמבר 2025 (C114) single source : מפעלים משלמים על אינטגרציה ושירות, לא על קופסת חישוב.

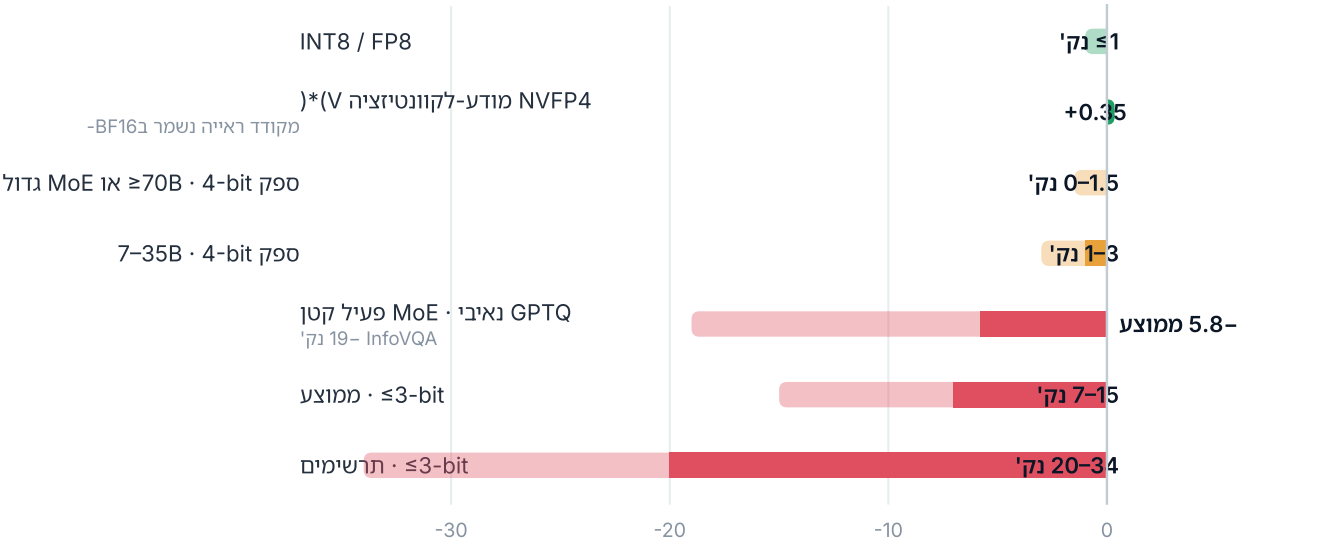
קוונטיזציה: bit-8 בטוח, bit-4 תלוי במתכון, bit-3 ומטה פסול

bit-8 (FP8 או INT8) הוא למעשה ללא אובדן עבור מודלי ראייה-שפה (Red Hat, Hugging Face) verified . ב-bit-4, האובדן תלוי בגודל ובמתכון: בניית ה-AWQ של Qwen 72B מאבדת פחות מנקודה על MMMU בעוד ה-7B מאבד 2.8 (Qwen AWQ) verified , ומשימות עתירות פרטים נופלות יותר (GPTQ נאיבי ב-bit-4 עולה ל-19 Qwen3-VL-30B-A3B נקודות על (VEQ) InfoVQA). ב-3-bit ומטה, האובדן הממוצע הוא 7–15 נקודות ותרשימים מאבדים 20–34 (VLMQ) verified ; מעבר ל-64k טוקנים, bit-4 מאבד עד 59%, גרוע יותר בקלט לא-אנגלי (arXiv 2505.20276). שמור את מקודד הראייה ב-FP16/BF16.

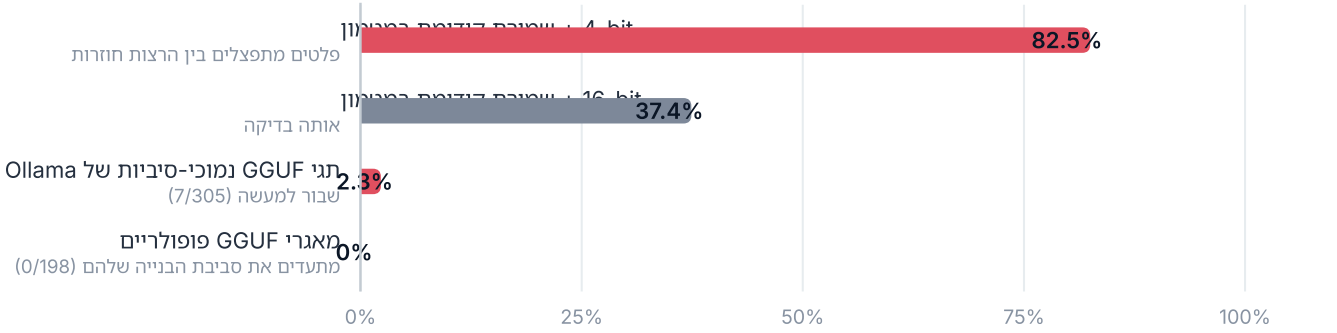
FIG 10 **קוונטיזציה: bit-8 בטוח, bit-4 תלוי במתכון, bit-3 ומטה בחוץ*טעויות הנדסה בערימת ההגשה עולות יותר דיוק מהקוונטיזציה עצמה.***

דיוק: bit-8 בטוח, bit-4 תלוי במתכון, ≤bit-3 בחוץ

שינוי ציון לעומת דיוק מלא)נקודות(



כשלי הנדסה עולים על שגיאת קוונטיזציה



מקור: כרטיס מודל NVIDIA Nemotron, VLMQ, VEQ, arXiv 2609.05881, 2610.19498, 2609.04748 ואחרים.

שתי אזהרות חשובות יותר. ראשית, **נכון ל-11 בנובמבר 2026, אין מחקר או כרטיס מודל שמדד קוונטיזציה על תמונות פגמי PCB/SMT**, אז אמת מחדש על סט הזהב. שנית, טעויות בערימת ההגשה הן לרוב גדולות משגיאת הקוונטיזציה: אותו מודל קיבל ציון נמוך בכ-30 נקודות תחת Ollama לעומת מימוש הייחוס (SiliconBench); 7 מ-305 תגי המודלים הרשמיים של Ollama היו למעשה שבורים אף שהקבצים שלמים (arXiv 2609.05881); אף אחד מ-198 מאגרי GGUF פופולריים לא תיעד את בנייתו (arXiv 2610.19498); ועם שימוש חוזר במטמון קידומת, 82.5% מהמקרים התפצלו ב-4-bit לעומת 37.4% ב-16-bit (arXiv 4-bit . single sources) מודע-לקוונטיזציה של ספק קרוב ללא אובדן במשימות פרטים (NVIDIA Nemotron 3.5 vendor claim), אבל VLA ב-4-bit מאבדים 4.3–1.5 נקודות תחת הפרעה ורצים לאט יותר על single source (LoopQ). Thor. נהל מודל מקוונטז כמו מד חדש: נעל גיבובי משקלים, תעד את הבנייה, אפס את מטמון הקידומת ואמת מחדש אחרי כל שינוי דיוק, מנוע או תבנית.

דיוק	70B≤ צפוף או 100B MoE≤	7–35B	4B≥ או משימות עתירות פרטים	פסק דין
FP8 / INT8	בטוח (≤1 נקודה)	בטוח	בטוח; ברירת המחדל לבדיקה	שימוש
AWQ/GPTQ-4 bit (מתכון ספק, מקודד ראייה FP16)	0–1.5 נקודות	1–3 נקודות; GPTQ נאיבי על MoE פעיל-קטן −5.8 בממוצע, −19 ב-InfoVQA	2–3 נקודות; אמת משימות פרטים קודם	מותנה

דיוק	70B≤ צפוף או 100B MoE≤	35B–7	4B≥ או משימות עתירות פרטים	פסק דין
NVFP4 (מתכון ספק)	≈1–0.3 נקודה מתחת ל-FP8 (נתוני ספק)	בניות מודעות-לקוונטיזציה כמעט ללא אובדן על V*, OCRBench	אין נתונים ציבוריים	מותנה (העדף בניות מודעות-לקוונטיזציה)
GGUF Q4_K_M	לא נבדק	שונות גבוהה במשימות תחום	לא לטקסט קטן או פגמים	לא לשיפוט פרטים
GGUF Q5_K_M ומעלה	ככל הנראה בטוח	כמעט ללא אובדן בשתי בדיקות	הרצפה לפרטים על llama.cpp	שימוש
bit-3≥ (כולל מודלי 1T שנדחסו ל-1.8–2 ביט על מכונה אחת)	לא להחלטות	לא	לא	לא

איזו תצורה לאיזו משימה

תצורה	בדיקה בקו	ביקורת AI של קריאות שווא	תכנון ובקרת רובוט	שאלות-ותשובות וניתוח
גלאים ייעודיים או מודלי אנומליה (YOLO, RF-DETR, מחלקת PatchCore), INT8/FP16, כל תיבת קצה NPU או GPU	כן: השופט הראשי, 2–12 אלפיות שנייה לחיתוך	כן (מסווג מכוונן על הקו)	תפיסה בלבד	—
מודלי ראייה 4B,–2 NVFP4/AWQ/FP8 עם מקודד ראייה FP16, על Thor/T4000 או DGX Spark	מותנה: בדוק מחדש חיתוכים שמסומנים על ידי AOI ב-0.2–0.3 שניות כל אחד	מותנה, מתחת לשנייה אחרי כוונון	עמוד שדרה ל-VLA (VLA 2–3B ב-≈40–130 אלפיות שנייה למקטע)	חלש
מודלי ראייה 9B (Qwen3-VL-–7 8B, MiniCPM-V 4.5, GLM-4.6V-Flash), FP8 או bit-4 של ספק, על Thor, Spark או כרטיס 32–24 GB	מותנה: ≈0.5 שניות לחיתוך מסומן	כן: 0.5–1 שניות לקריאה, התמורה הטובה ביותר	תכנון ובחירת מיומנות (1–3 שניות), לא בלולאת הבקרה	שאלות-תשובות SOP בסיסיות למשתמש אחד או כמה
MoE 26–35B (Gemma-4-26B-A4B, Qwen3.6-35B-A3B), NVFP4/FP8	לא (כבד מדי)	כן, כמסביר או למקרים קשים	תכנון	כן (8 משתמשים על Spark, ≈0.1 שניות לטוקן ראשון)
MoE 106–241B ב-bit-4 על Mac Studio, RTX PRO 6000 או Spark	לא	לא	תכנון לא מקוון	כן, למשתמש אחד או כמה
DeepSeek 671B או Kimi K2.x 1T ב-bit-4 על Mac של 512 GB, שרת KTransformers או DGX Station	לא	לא	יצירת תוכניות לא מקוונת בלבד	מותנה: 1–3 משתמשים והנחיות קצרות; הנחיות ארוכות לוקחות 16–170 שניות
כל דבר ב-GGUF-3≥ bit; Q4_K_M לעבודת פרטים של מודל קטן; Ollama למשתמשים רבים	לא	לא	לא	לא

מקורות: נתוני TensorRT Edge-LLM, Jetson AI Lab, Apple ML Research ו-KTransformers לעיל, ו-DGX Spark tests, RTX PRO 6000 tests, Hardware Corner on (Atlas, Ultralytics YOLO26, RF-DETR) community and vendor tests, mostly single source.

אז, לשתי הדאגות הרגילות: **זה לא חייב להיות איטי** אם המודל תואם למשימה (מודלי ראייה קטנים ובינוניים על Thor או Spark עומדים בכל תקציב בקו וביקורת; מודלי 671B–1T נשארים מחוץ לקו), ו**הקוונטיזציה לא חייבת להפוך את זה לשגוי** עם FP8/INT8 כברירת מחדל, או bit-4 במתכון ספק מעל 7B שאומת על סט הזהב. הסיכונים הגדולים יותר הם הדיוק התעשייתי zero-shot של

כ-57 נקודות של מודלים כלליים וטעויות בערימת ההגשה. תן לבקשות בדיקה גדלי אצווה קבועים ותור משלהן (טמפרטורה 0 אינה ניתנת לשחזור ביט-אחר-ביט תחת אצווה (vLLM documentation)), שמור גיבוי חם N+1, וגרום לכל צעד על הקו להיכשל-ולפתוח.

מורה ותלמיד

03 מורה ותלמיד: מודלי חלוץ יכולים לתייג מראש נתוני בדיקה; עבור זרועות רובוטיות המורים האמיתיים הם עדיין אנשים

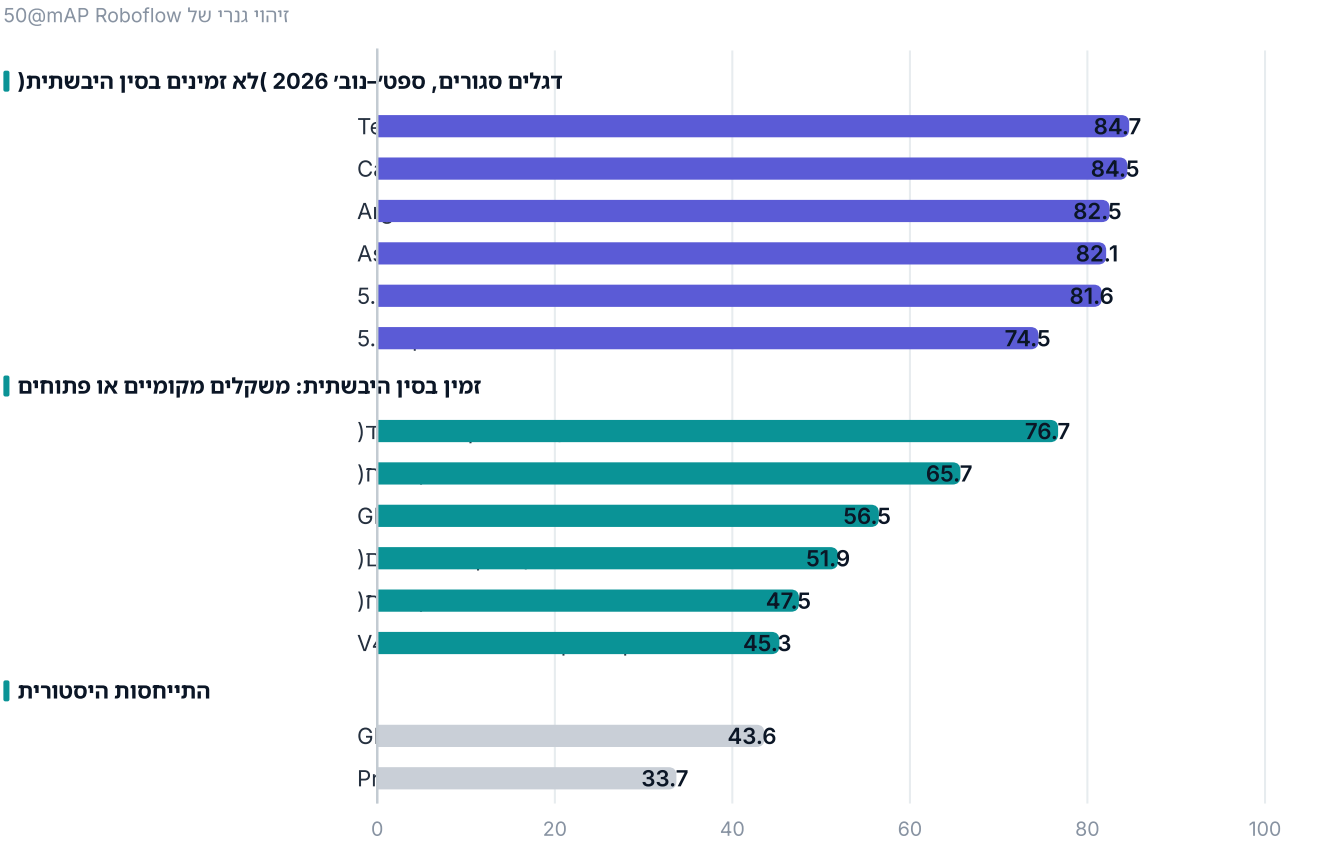
מועמדי מורה של נובמבר 2026: ראייה גנרית הוכפלה, ציוני פגמים תעשייתיים חסרים

כל דגל סגור שיכול לשמש כמורה בנובמבר 2026 מקבל תמונות: Claude Fable 5.2 (זמין כללית ב-9 בנובמבר, $40/$8 למיליון טוקנים) ו-GPT-6 Astra ($10/$50) verified ; GPT-6.1 Terra ו-Opus 5.5 ($4/$20) (Anthropic, Claude models overview); Gemini 4 Argon ו-Carbon, שנפתחו בשלבים ($1/$5); (OpenAI pricing). שמישים בסין היבשתית הם Qwen3.8-Max (ראייה דרך API) ומודלים פתוחים כמו Qwen3.8-27B, Kimi K3, GLM-5V-Turbo, DeepSeek V4-Flash ו-Vision-Exp (Roboflow blog, Hugging Face) Gemma 4.

את הפער המרכזי יש לנסח בפשטות: לאף אחד מהמודלים האלה אין ציון שפורסם על MMAD, MVTec AD 2, VisA, Real-IAD, UniPCB או כל מערך נתוני PCB/SMT. ל-MMAD עדיין אין לוח תוצאות (MMAD GitHub), ושני מאמרי האנומליה התעשייתית החדשים ביותר (אוגוסט–ספטמבר 2026) עדיין משתמשים ב-Gemini 2.5 Pro ו-Claude Sonnet 4 כבסיסים הכלליים החדשים ביותר שלהם (InspectorGPT, Anomaly-LR). חיפוש טקסט מלא ב-arXiv ב-11 בנובמבר מצא אפס התאמות ל-"Fable 5.2" או "Gemini 4 Carbon" בהקשר הזה (arXiv API) single source, absence of evidence . התחליפים הקרובים ביותר:

- מודל החלוץ החדש ביותר עם ציון MMAD הוא GPT-5.5: 69.95% ב"פגם או לא" ו-76.20% בממוצע על פני שבע משימות על תת-קבוצה של 500 תמונות, מול 95.24% ו-86.65% למומחים אנושיים (AnomalyClaw) single source ; GPT-4o קיבל 68.63% ו-74.92% ב-2024 (MMAD, ICLR 2025), כ-1.3 נקודות התקדמות בשנתיים.
- ב-LDU-Bench, תמונות SEM אמיתיות מליתוגרפיית IC והמדד האלקטרוני היחיד עם מודלי 2026, מיון טוב-או-רע סביר (מאקרו F1 של 0.932 ל-0.884 ל-GPT-5.4, 0.876 ל-GLM-5V-Turbo, 0.504 ל-Claude Opus 4.6, Gemma 4 31B ל-), אבל סיווג 11 מורפולוגיות פגמים מגיע רק ל-0.20–0.41 (LDU-Bench) single source .
- במדדי הראייה הבלתי תלויים של Roboflow (עודכנו ב-10 בנובמבר 2026), mAP@50 של זיהוי עצמים הוא 84.5 ל-Gemini 4 Carbon, 84.7 ל-GPT-6.1 Terra ו-81.6 ל-Fable 5.2, מול 33.7, 43.6 ו-38.6 ל-GPT-5.5, Gemini 2.5 Pro, Claude Opus 4.8 (Roboflow Vision Evals) single source, independent evaluator, private datasets -1.

FIG 11 **מועמדי מורה: ציונים גנריים הוכפלו בדוראבל לאף אחד מהדגלים של 2026 אין ציון שפורסם במדד פגמים תעשייתי.**



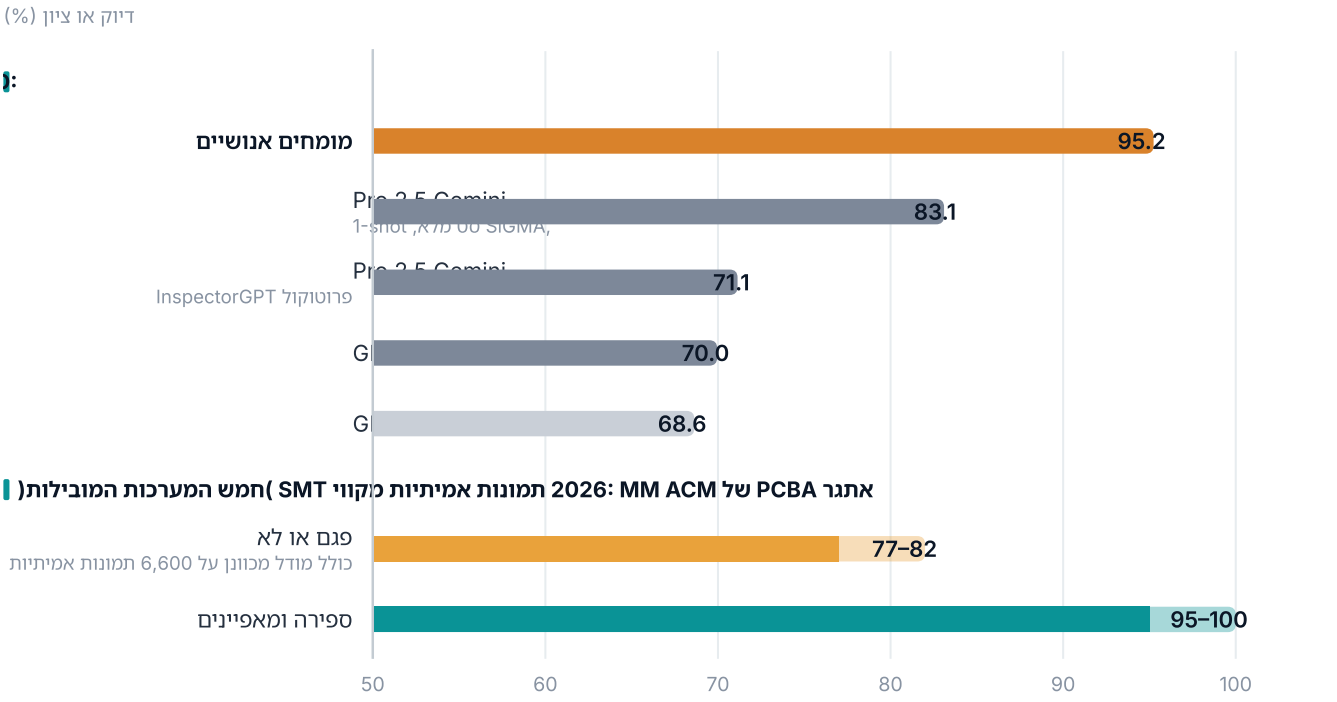
מדדי פגמים תעשייתיים (MMAD, MVTec AD 2, Real-IAD, VisA, UniPCB): אפס ציונים שפורסמו לדגלים אלה של 2026
ציוני זיהוי גנריים הוכפלו בדור, אבל זו לא הערכת פגמים; בדוק על סט הזהב שלך

מקור: Roboflow Vision Evals; הערות שחרור מודלים; עמודי מדדים של MMAD, MVTec AD 2 ואחרים (נכון ל-11 בנוב' 2026).

מועמד מורה (נוב' 2026)	מחיר למיליון טוקנים (in/out)	זיהוי Roboflow mAP@50 / ספירה	מדדי פגמים תעשייתיים	שמיש בסין היבשתית
Claude Fable 5.2	$8 / $40	86.5 / 81.6	אין	לא
Claude Opus 5.5	$4 / $20	80.6 / 74.5	אין	לא
GPT-6 Astra	$10 / $50	80.2 / 82.1	אין	לא
GPT-6.1 Terra	$1 / $5	80.2 / 84.7	אין	לא
Gemini 4 Carbon	$2 / $10 (קידום)	90.1 / 84.5	אין	לא
Qwen3.8-Max (ראייה דרך API בלבד)	$2 / $6	81.1 / 76.7	אין	כן
Qwen3.8-27B (פתוח)	מקומי	64.9 / 65.7	אין	כן
GLM-5V-Turbo	$1.20 / $4.00	48.6 / 56.5	LDU-Bench triage 0.884	כן
Gemma 4 31B (פתוח)	מקומי	51.4 / 47.5	LDU-Bench triage 0.504	כן (משקלים מקומיים)
התייחסות: GPT-5.5 / GPT-4o	—	68.0 / 43.6 (GPT-5.5)	MMAD בינארי 69.95% / 68.63%	לא
מומחים אנושיים	—	—	MMAD בינארי 95.24%	—

מאמרים חדשים עדיין בודקים את הדור הקודם, אבל מצביעים לאותו כיוון. בסט ה-MMAD המלא, מומחים אנושיים קיבלו 95.2% ב"פגם-או-לא" ו-(SiGMA) Gemini 2.5 Pro 83.1%, ובכל זאת אותו מודל קיבל רק 71.14% במאמר אחר (InspectorGPT): פרוטוקול הבדיקה מזיז את הציון כמו דור מודל שלם. באתגר VAND 4.0 של CVPR 2026, מודלי ראייה-שפה מהמדף הגיעו לשיא של AP 61.88 מול 90.84 לשיטות מפוקחות, והמארגנים סיכמו שהם "אינם יכולים כרגע להחליף מסווגים מכוונים" (VAND 4.0) single. באתגר PCBA של ACM MM 2026, חמש המובילות קיבלו 82.47–76.92 ב"פגם-או-לא" בעוד סוג רכיב וספירת פינים הגיעו ל-99.5–95 (arXiv 2609.21276) single source: לזהות הרבה יותר קל מלשפוט.

FIG 12 **"האם יש פגם?" עדיין קשה למודלים כלליים—המודל הכללי הטוב ביותר מפגר אחרי מומחים אנושיים ביותר מ-12 נקודות ב-MMAD; מערכות מכוונות טובות יותר על תמונות קו אמיתיות.**



מקור: MMAD; SiGMA; InspectorGPT; אתגר PCBA של ACM MM 2026.

שלוש מסקנות עולות. ראייה וספירה הוכפלו בדור, אבל שיפוט חריגות נע לאט, תלוי מאוד בפרוטוקול ונושא הטיות שיטתיות; כמה המודלים החדשים משתפרים אפשר למדוד רק על התמונות של המפעל עצמו. בתוך התפלגות האימון שלהם, מומחי 7B מכוונים עדיין מנצחים מודלים כלליים (Anomaly-LR, UniPCB). ונתוני קו אמיתיים ממוטטים ציונים: על לוחות אלקטרוניים מורכבים, AUROC ברמת תמונה נפל מ-97–98 ב-MVTec AD ל-31.5–78.6 (arXiv 2606.01992) single source. עבור חברות סין היבשתית המורה החזק ביותר הזמין, Qwen3.8-Max, יושב כ-8 נקודות מתחת לחלוץ הגלובלי.

הרצת עשרות או מאות סבבים: כסף אינו הבעיה, איכות התיוג היא

כסף אינו הבעיה. במחירי המחירון של נובמבר 2026, 1,000 תמונות של מגה-פיקסל אחד עולות כ-$3 (GPT-6.1 Terra) עד $32 (GPT-6 Astra) לסבב, ומאה סבבים על פני 100,000 תמונות כ-$32,000–135,000 לפני הנחות אצווה (Claude vision estimate) (docs, OpenAI vision guide, Roboflow Vision Evals). לתוויות פשוטות אדם עולה בערך כמו Terra; היתרון של המורה הוא מהירות וקנה מידה, לא מחיר. NDA של לקוחות בדרך כלל שומר תמונות מוצר בתוך המפעל, כך שסט הזהב מתויג באתר.

המכשול האמיתי הוא איכות, ויותר סבבים לא מתקנים שגיאות שיטתיות. Roboflow מריצה כל משימה שלוש פעמים והציונים נעים רק 0.1–1.4± נקודות (Roboflow), כך שמודל בעיקר חוזר על טעויותיו והצבעת רוב לא יכולה להצילו. סבבים עוזרים רק כשכל אחד מוסיף מידע: תמונת ייחוס טובה אחת או שתיים מוסיפות כ-3–4 נקודות, יותר לא מוסיף דבר (AD-Copilot); לגרום למודל להתווכח מול ייחוס טוב על פני כמה פניות מוסיף 6.23 נקודות AUROC חוצה-תחום ל-GPT-5.5 ו-2.95 ב-MMAD, עם שיא אחרי כשלוש קריאות (AnomalyClaw) single sources. "תלמיד שאומן על תיוגי מורה מתקרב לתיוגים אנושיים" הוכח רק על תמונות טבעיות: ResNet-101 שאומן על תיוגי Qwen3-VL של COCO ו-Objects365 הגיע ל-95.7–99.4% מתוצאת התיוג האנושי, והמחברים אומרים שנתוני תחום לא נבדקו (arXiv 2602.20972) single source.

FIG 13 ***שני צינורות מורה–תלמיד לבדיקה, מודל חלוץ מתייג מראש ואנשים מכריעים; לזרועות רובוט המורים האמיתיים הם עדיין הדגמות אנושיות.***

בדיקה: מודל החלוץ מתייג מראש, אנשים מכריעים, סט זהב שומר

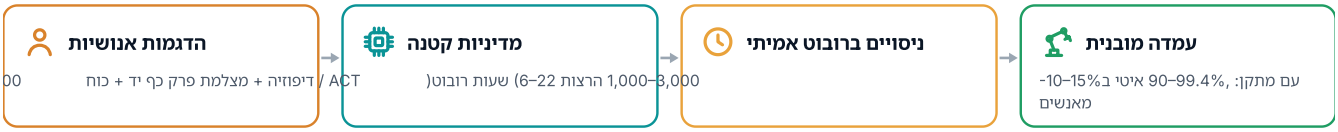


רף לקבלת תיוגי מורה אוטומטית

למחלקה: ≥299 פגמים, אפס החטאות (אחזור ≥99% בביטחון 95%), הסכמה לא גרועה מאדם–אדם; אחרת תיוג מקדים בלבד

זרועות: המורים עדיין הם הדגמות אנושיות, למידת חיזוק על רובוט אמיתי וסימולציה

מודלים חלוצים עוזרים רק לא מקוון: תכנות והחלפות



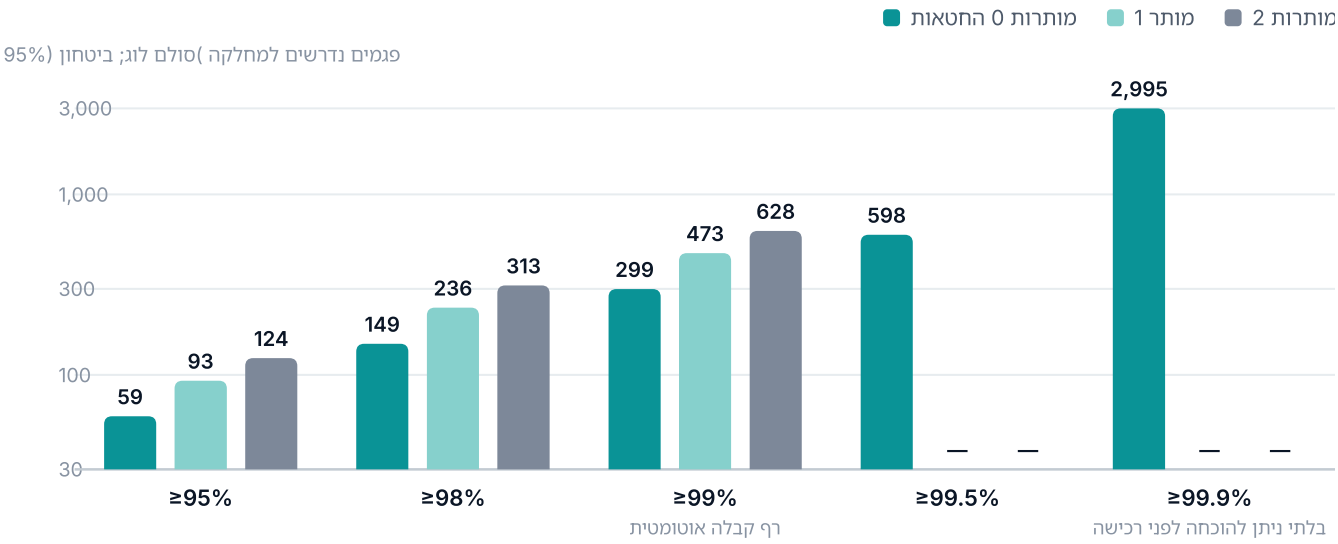
מקור: סינתזת המחבר (פרק 3); עלויות וגדלי מדגם הם הערכות.

פרוטוקול ישים estimate:

1. **אסוף את סט הזהב מהארכיון של ה-AOI של המפעל עצמו ומלוחות NG שנשמרו**, מרובד ל-12 מחלקות פגמים אמיתיים (חסר, היסט, מצבה, קוטביות, חלק שגוי, גשר/קצר, הלחמה לא מספקת, הלחמה עודפת, כדור הלחמה, חיבור מורם/פתוח, חלק פגום, חומר זר) וארבע סיבות לקריאת שווא (צל, שטף, הדפסת משי, סובלנות), בתוספת "פרוסת סחיפה" של לפחות 300 פגמים מקו שני או מתאורה שונה, צבע לוח או מספרי חלק שונים.
2. **תייג בעיוורון עם שניים או שלושה מפקחים מאומנים ל-IPC-A-610**, עם מהנדס בכיר שמכריע ו-X-ray, בדיקה חשמלית או חתכים שמיישבים דגימות לא חד-משמעיות. קבל הגדרת מחלקה רק אם הסכמת המפקחים מגיעה ל-$\kappa \geq 0.75$ ($\kappa \leq 0.90$ למחלקות קריטיות); הסכמה אנושית על 11 מורפולוגיות פגמי אלקטרוניקה ב-LDU-Bench הייתה רק 0.727, כך שחלק מהמחלקות יזדקקו למיזוג או הגדרה מחדש. הסכמת מאפיינים בסגנון AIAG (יעילות ≤90%, החטאה ≤2%, אזעקת שווא ≤5%, κ >0.90 ;>0.75 לשחרור) היא מינימום סביר (Fabrico, iFactory).
3. **קבע גודל עם גבולות בינומיים מדויקים (Clopper–Pearson) חד-צדדיים של 95%** (binomial confidence intervals, rule of three:

אחזור להוכחה	0 החטאות מותרות	1 מותר	2 מותרות
≤95%	59 פגמים	93	124
≤98%	149	236	313
≤99%	299	473	628
≤99.5%	598	—	—
≤99.9%	2,995	—	—

FIG 14 **כמה גדול חייב להיות סט הזהב*הוכחת אחזור של 99% עם אפס החטאות דורשת 299 פגמים למחלקה; לא ניתן להוכיח 99.9% לפני רכישה.***



מקור: רווחי סמך בינומיים מדויקים (95%).

שישים פגמים למחלקה מספיקים למסך עובר/נכשל; 300 נדרשים כדי להחליט על קבלה אוטומטית. הסט המלא, כ-3,600 פגמים בתוספת 3,500 חיתוכי קריאות שווא, לוקח בערך 37–57 שעות עבודה לתיוג ולהכרעה, והרצת מועמד מורה עליו עולה 23$–196; העלות האמיתית היא שבועיים עד שישה שבועות של איסוף, שנקבעים על ידי המחלקה הנדירה ביותר estimate . שלח חיתוכי AOI של רכיב בודד, לא לוחות שלמים, עם מטרה של לפחות 42–48 פיקסלים לרוחב (arXiv 2610.09313).

1. **קבל תיוגי מורה אוטומטית רק למחלקה עם אפס החטאות על ≥299 פגמי זהב**, הסכמת מודל-אמת לא גרועה מהסכמת אדם-אדם, ולפחות 50% מקריאות השווא סווגו. מחלקות קריטיות כמו קצרים, חלקים חסרים, קוטביות וחלקים שגויים זקוקות ל-99.5%–99.9% (598–2,995 פגמים); אחרת כל מה שה-AI אינו יכול לסווג בביטחון גבוה הולך לאדם. שיעור בריחות מתחת ל-0.1% מהפגמים אינו ניתן להוכחה לפני רכישה (2,995 פגמים עם אפס החטאות; 0.01% ידרוש 29,956), כך שהקו שומר על דגימה אנושית של קריאות שסווגו על ידי AI, ICT/FCT במורד הזרם כרשת ביטחון ו-SPC על בריחות. תיוגים מאומתים מזוקקים לאחר מכן לגלאי או מסווג קונבנציונלי ומאומתים מחדש על כל מוצר חדש.

תנאים וכללי נתונים גם מחליטים את המורה. Anthropic אוסר שימוש בשירותיו לאימון מודלים מתחרים (Anthropic commercial terms); OpenAI מאפשר להשתמש בפלטים לאימון מסווגים פנימיים שאינם נמכרים לצדדים שלישיים (OpenAI business terms); DeepSeek מתיר במפורש זיקוק (DeepSeek terms) primary terms . אף אחד אינו מגדיר "מודל מתחרה", כך ששימוש מסחרי זקוק לבדיקה משפטית, וכללי הנתונים של הלקוחות בדרך כלל שומרים תמונות באתר. ברירת המחדל הבטוחה ביותר היא מורה עם משקלים פתוחים שמורצה מקומית, עם מודלי חלוץ בענן רק כשהלקוח מאשר זאת בכתב.

תלמידי בדיקה: קטנים ומהירים; צווארי הבקבוק הם אופטיקה, נתונים ואימות

התלמיד אינו צוואר הבקבוק: YOLO26n לוקח 3.80 אלפיות שנייה לתמונה ב-INT8 על Orin Nano Super של 249$–399 (Ultralytics) , NVIDIA vendor test מדווחת על 98.51% על שלוש מחלקות פגמי PCB אחרי התאמת מודל בסיס על כ-600 תמונות (NVIDIA) vendor claim , ופגמים סינתטיים מוסיפים רק 0.9–2.5 נקודות (arXiv 2605.04635) mAP. האילוצים הם אופטיקה, נתוני קו, אימות ורישיונות: YOLO של Ultralytics הוא AGPL-3.0 או רישוי ארגוני (Ultralytics), כך שמוצרים סגורים זקוקים לרישיון או לגלאי Apache-2.0 כמו RF-DETR.

תלמידי רובוט: תחנות מובנות קרובות לדרגת ייצור; עבודה גמישה ודינמית לא

בנובמבר 2026 יש שלושה סוגי מורי רובוט, ורק אחד יכול להפוך לתלמיד מקומי: VLA פתוחים ניתנים לכיוונון (π0/π0.5, GR00T N1.7, X-VLA, RDT2, SmolVLA). מודלי רובוט סגורים (π0.6/π0.7) דרך API בגרסת בטא, Gemini Robotics 2 לבודקים מהימנים) אינם ניתנים לפריסה באתר (The AI Insider, DeepMind). מודלים כלליים של חלוץ איטיים וחלשים במניפולציה עדינה: במבחן של Artificial Analysis הם קיבלו 46–69 עם 28–45% הצלחת אחיזה וכ-30 דקות למשימה, "איטי ויקר מדי לכל שימוש בייצור" (AA-

Robotics); GPT-6 Astra הצליח בחריץ מדויק רק 2 פעמים מתוך 20 (RoboCurve); והנעת מפרקים ישירות נותנת 0–5.5% על LIBERO ב-0.2–0.4 הרץ מול כ-83 הרץ נדרשים (Anthropic) verified across three sources . ה-64.9% של Fable 5.2 במדד RoboDojo המדומה הם הרצה פנימית (Anthropic) vendor claim .

מחקרי צד שלישי מאוקטובר מסכימים: כ-33–44% על LIBERO-10 עבור שלושת הסוכנים המובילים (arXiv 2610.30051); סוכנים שטוענים להצלחה כוזבת ב-24–39% מתתי-המשימות (arXiv 2611.08209); ו-104 מתוך 108 משימות רובוטיות מעין-סטטיות על רובוט אמיתי ל-Opus 5.5, אבל ב-280 שניות כל אחת וללא החדרה או הרכבה (AMP1) single sources . מודלי חלוץ שייכים למצב לא מקוון: כתיבת תוכניות, תיאום משימות וניסוח קריטריוני הצלחה.

לאיכות התלמיד יש יתרון ברור: טובה בתוך התפלגות האימון, מתמוטטת מחוצה לה. π0.7 "לרוב עולה על 90%" במשימות שנראו ומגיע ל-60–80% באלה שלא נראו (π0.7 paper) vendor research ; ContactDP מגיע ל-97.8% על מחברים שאומנו עליהם אבל 28.6–66.7% על כאלה שלא נראו (ContactDP); שינוי מצלמה או תנוחה קל מפיל 95% מתחת ל-30% ב-LIBERO-Plus (LIBERO-Plus, CVPR 2026); אחיזה נלמדת ממסועים נעים מגיעה לשיא של 58.89% (ForeTime-VLA) single sources , עבודה שמעקב ראייה קלאסי עושה בייצור כבר שנים.

הראיות הקרובות לייצור מגיעות כולן מתחנות מובנות, הדגמות אנושיות ותיקון על רובוט אמיתי. Neurometka שמרה על מכונת מצבים ותנועה נלמדת מראש ולמדה רק אחיזה, החדרת כבל והלחמה, כל אחת מפחות מ-20 דקות נתונים: 322 מתוך 324 בייצור, איטית ב-12.8% מאדם (Neurometka) single source, vendor . Fraunhofer הגיעה ל-92.8% על פני ארבעה מחברים מ-68 דקות הדגמות (Fraunhofer); ה-HIL-SERL של Berkeley הצליח 100 פעמים מתוך 100 בהחדרת SSD, RAM ו-USB אחרי 2.5–1 שעות RL על רובוט אמיתי (HIL-SERL); AgiBot רצה 18–20 שניות לצעד בתוך טקט של 20 שניות בקו בדיקת טאבלטים (IThome) vendor claim .

זיקוק קונה מהירות, לא הכללה: π0.5 רץ ב-0.70 הרץ על Jetson Orin גולמי ו-10.30 הרץ אחרי תזמון אסינכרוני (EagleVLA); זיקוק שכבות לוקח אותו מכ-3 הרץ לכ-10 הרץ (Shallow-π) verified . תלמיד יורש רק את מה שהמורה כיסה; כל מוצר חדש עדיין זקוק לנתוני משימה.

FIG 15 **השהיית מדיניות קצה*מדיניות ACT קטנה רצה בכ-50 אלפיות שנייה על לוח קצה; VLA שהועבר לשבב תחביבי לוקח 7.3 שניות.***

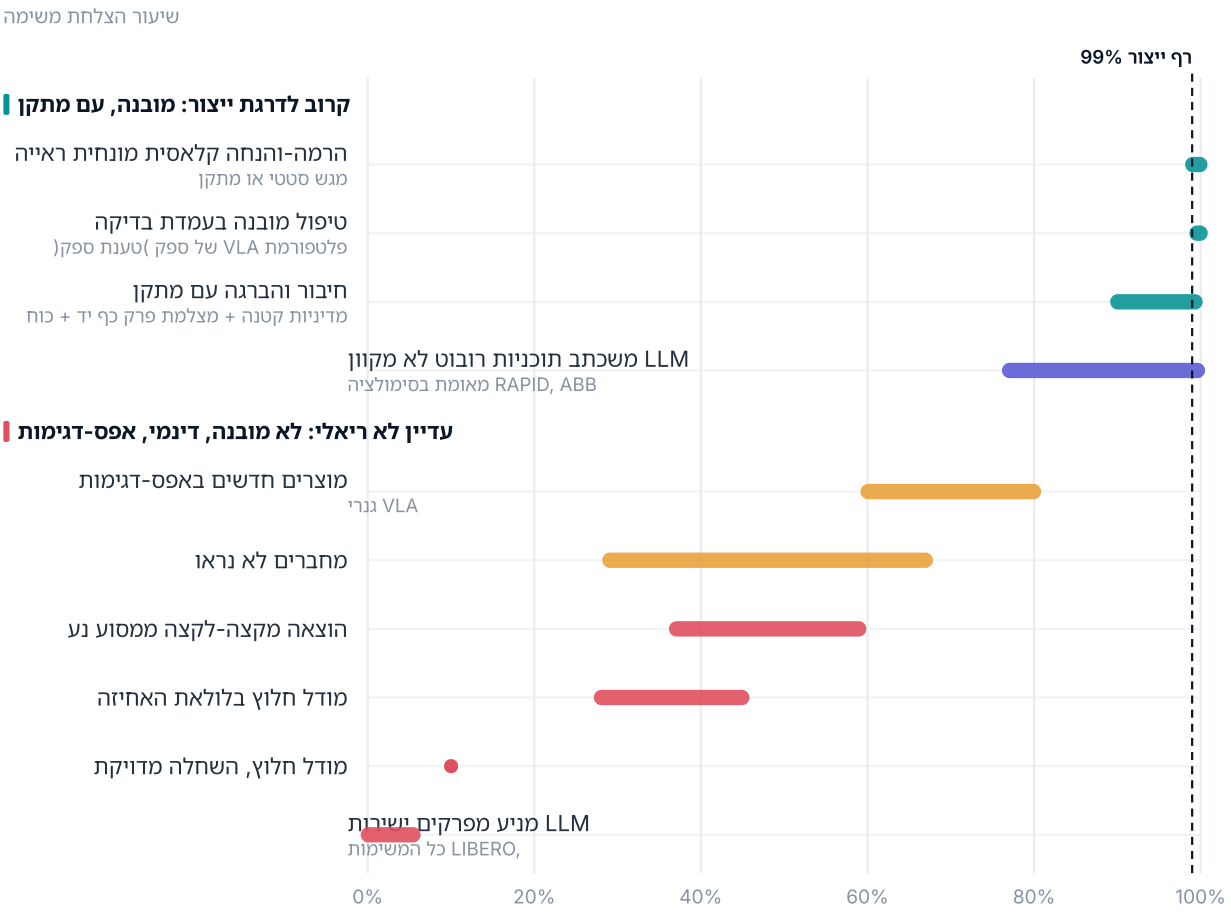


מקור: תיעוד ופורום D-Robotics (נתוני ספק); פורטים קהילתיים ב-GitHub; מחקר של Anthropic.

על שבבי קצה סיניים, המדיניות הנלמדת היחידה שנמדדה היא מדיניות ACT על RDK S100 של D-Robotics ב-46–50 אלפיות שנייה למקטע של 50 צעדים, ללא שיעור הצלחה שפורסם (D-Robotics RDK S docs) vendor claim ; פורט תחביבי של SmolVLA על RK3588 לוקח כ-7.3 שניות למקטע (GitHub). קוונטיזציה לא קונה מהירות בחינם: אין בניית π0.5 ב-4-bit שהייתה מהירה מ-16-bit בגודל אצווה 1 (LoopQ). הדגמות יקרות (רויטרס דיווחה על כ-300 ניסיונות לתנועה שמישה לטלאופרטור מתחיל

(Notebookcheck)), ואימות עולה יותר: הוכחת שיעור כשל ≤0.1% בביטחון 95% לוקחת כ-3,000 הצלחות רצופות, 17–22 שעות רובוט estimate . אין מדיניות מניפולציה נלמדת עם הסמכת בטיחות.

FIG 16 תלמידי רובוט: קרוב לדרגת ייצור רק כשהעמדה מובניתעמדות עם מתקן מגיעות ל-90–99%+; משימות לא מובנות, דינמיות ואפס-דגימות עדיין לא ריאליות.



מקור: מאמרים וחומר ספקים שצוטטו בפרק 3; חלק מהנתונים הם טענות ספק.

משימת אלקטרוניקה	תלמיד ישים	הצלחה צפויה	מהירות מול אדם	עלות להחלפה	פסק דין 2026–2028
Pick-and-place בתמהיל גבוה, קיטים, טיפול במכונות (מגש או מתקן סטטי)	תפיסה נלמדת בתוספת תנועה מתוכנתת או מתוכננת	≥99% עם הנחיית ראייה קלאסית	זהה לתחנות קונבנציונליות	שעות	ישים עכשיו
מחבר מקובע, החדרת FFC וכבל, הברגה, הלחמה	מכונת מצבים + מדיניות ACT/דיפוזיה/BC קטנה + מצלמת פרק כף יד + כוח/מומנט (באופן אידיאלי מגע) + תיקון אנושי	90–99.4%	איטי ב-10–15%	10–300 הדגמות (≈15–70 דקות) + 1,000–3,000 הרצות אימות (6–22 שעות רובוט)	ישים, עתיר מו"פ; חודשים לתחנה הראשונה
טיפול בתחנת בדיקה מובנית ומיון OK/NG	פלטפורמת VLA של ספק + תאום דיגיטלי + RL על רובוט אמיתי	≥99.5% vendor claim, conflicting	18–20 שניות לצעד, בתוך טקט של 20 שניות	≥4 שעות אימון מחדש + 15 דקות כיול	ישים על פלטפורמת ספק
תכנות החלפות	LLM כותב או עורך תוכניות רובוט לא מקוון, מאומת בסימולציה	77–99.7% בשכתובי ABB RAPID	אין השפעה על הטקט	דקות עד שעות לתוכנית	ישים ככלי הנדסי

משימת אלקטרוניקה	תלמיד ישים	הצלחה צפויה	מהירות מול אדם	עלות להחלפה	פסק דין 2026–2028
הרמה נלמדת מקצה-לקצה ממסועים נעים	תלמיד VLA	37%–59%; ≥37% על מסועים מהירים	—	≤458 הדגמות להגדרה	**לא לפני 2028; השתמש במעקב קלאסי בתוספת תפיסה נלמדת**
מוצרים חדשים ללא נתונים חדשים (zero-shot)	VLA גנרי	60%–80%; מחברים לא נראו 29%–67%	—	—	**לא ריאלי**
LLM חלוץ בלולאה שמניע את הזרוע	סוכני GPT-6, Claude, Gemini	אחיזה 28%–45%; החדרה מדויקת 10%	≈30 דקות או 280 שניות למשימה מול ≈20 שניות	—	**לא ריאלי**
רתמות חוטים בנתיב חופשי והרכבה גמישה רב-שלבית	—	אין ראיות ייצור מ-2026	—	—	**לא לפני 2028**

מקורות: המחקרים לעיל ו-(University of Passau on RAPID rewriting, GraspVLA, Komax) **mostly single source and vendor data** . בשנים 2026–2028, הערך של מודל גדול לרובוטים הוא כפריור מאומן מראש ומתכנת לא מקוון, לא כמשהו ש"מלמד" תלמיד. כל מקרה קרוב לייצור שמצאתי (Neuromeka, AgiBot ב-Longcheer, Fraunhofer, ContactDP) נשען על הדגמות אנושיות, חישת כוח או מגע, סימולציה או תאומים דיגיטליים ו-RL מפוקח. אף אחד לא נשען על תיוגים ממודל חלוץ.

עלויות והחזר השקעה

04 עלויות והחזר השקעה: ביקורת AI מחזירה את עצמה תוך חודשים, טיפול במכונות תוך שנים, זרועות הוצאה לעולם לא

עלות עבודה: ארה"ב היא בערך פי 5.5 משנזין, טאיוואן בערך פי 1.5, רוב דרום-מזרח אסיה מתחת לחצי

החזר על החלפת עבודה מתחיל במה שהעבודה עולה. הנחיית השכר הרשמית של שנזין מציבה את החציון של 2024 למפקח איכות בייצור על **RMB 74,074** לשנה; שכר הוא 86.92% מעלות העבודה הכוללת בייצור אלקטרוניקה, כך שהעלות המלאה היא כ-RMB 85,000. סקר JETRO של 2025 מציב את העלות השנתית של עובד מפעל סיני על $12,667, כ-RMB 85,100, בטווח של 0.2% (Shenzhen wage guidance, JETRO 2025) verified.

אזור	עלות שנתית מלאה של מפקח או מטפל אחד (נמוכה / בסיס / גבוהה)	בסיס
סין היבשתית, שנזין	RMB 70k / **85k** / 100k ($10.4k / **12.7k** / 14.9k)	חציון רשמי × 1.15; תואם JETRO verified
אזורי שכר מדרגה 2 בגואנגדונג (דונגגוואן וכו')	RMB 65k / **75k** / 85k	חציון גואנגדונג 2023 × 1.15 בתוספת שנתיים העלאות estimate
דלתת נהר היאנגצה (סוג'ואו, קונשאן, וושי)	כמו שנזין	אין בסיס לעבודת בדיקה זולה יותר; שכר המינימום העליון של ג'יאנגסו גבוה בכ-5.6% משל שנזין estimate
טאיוואן	$16.6k / **19.6k** / 22.7k (NT$534k / **630k** / 729k)	סקר שכר תעסוקתי של משרד העבודה בתוספת עלויות חובה ≈1.18; JETRO כקצה הגבוה estimate
מלזיה	$6.8k / **9.1k** / 10.0k	JETRO ושכר מינימום estimate
תאילנד	$4.7k / **7.0k** / 8.8k	זהה
אינדונזיה	$5.0k / **6.0k** / 6.3k	זהה
וייטנאם	$3.2k / **5.3k** / 6.5k	זהה
הפיליפינים	$4.0k / **4.5k** / 4.9k	זהה
ארה"ב	$54k / **70k** / 80k	חציון מפקח של BLS מאי 2025 $48,570 ÷ שכר ב-65.5% מהפיצוי estimate

מקורות: (JETRO 2025, Guangdong wage guidance, Wuxi HR bureau, Taiwan occupational wage survey, PTS News, BLS OEWS, BLS ECEC).

FIG 17 **כמה עולה עובד בדיקה או טיפול אחד בשנה*בעלות מלאה, ארה"ב היא בערך פי 5.5 משנזין וטאיוואן בערך פי 1.5.***

עלות שנתית כוללת לעובד בקרת איכות או טיפול ($10k); נקודה = בסיס, קו = נמוך–גבוה)



מקור: הנחיית שכר רשמית של שנזין, משרד העבודה של טאיוואן, JETRO, BLS; הערכות המחבר.

שלושה תיקונים קובעים את הבסיס. ראשית, שנזין העלתה את רצפת הפרשת הפנסיה שלה מ-RMB 4,775 ל-5,521 לחודש ביולי 2026 (Shenzhen Bendibao) single source, כך שמפקח בשנזין ב-2026 עולה ככל הנראה RMB 80–90k. שנית, משרה אינה אדם: שתי משמרות של 12 שעות זקוקות לכ-2.0 אנשים (שישה ימים בשבוע) עד 2.4 (כל השנה) בדלתת נהר הפנינה, כ-2.7–3.1 בדרום-מזרח אסיה תחת תקרות שעות נוספות, ו-4.2–4.5 ל-7×24 בארה"ב ובטאיוואן estimate. ביקורת AI בדרך כלל מאפשרת למבקר אחד לכסות שניים עד ארבעה קווים, כך שהחיסכון הריאלי הוא **0.5–1.5 אנשים במשרה מלאה לקו, לא אדם אחד לקו למשמרת.** שלישית, קונים בטאיוואן, וייטנאם ומלזיה לא רודפים רק אחר עלות: בסקר של JETRO, 61.1% מהחברות בטאיוואן, 48.2% בווייטנאם ו-44.9% במלזיה אמרו שגיוס הפך לקשה יותר בשנתיים האחרונות, לעומת 18.6% בסין היבשתית (JETRO 2025) single source.

מחירי ציוד: ארבעה בסיסים שאין למצע

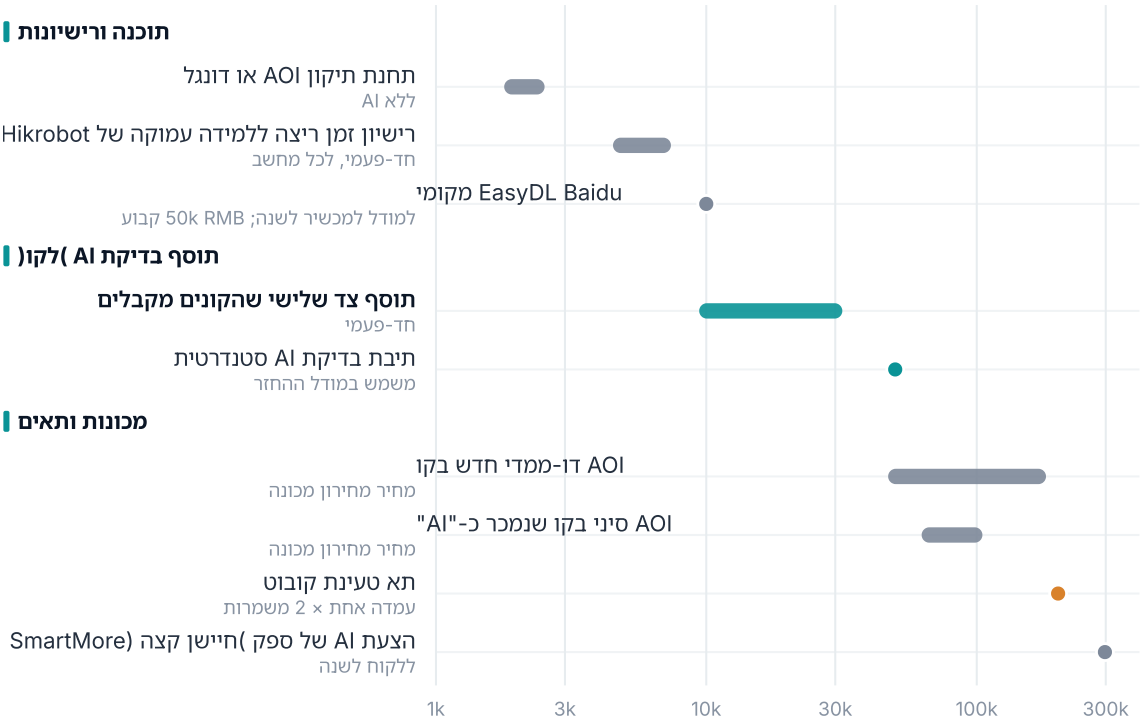
מחירי AOI חלוקים בעיקר כי הם מודדים דברים שונים. רשימות מקוונות (נובמבר 2026, מעורבות חדש, משומש והשכרה) נעות בערך RMB 17–80k עבור 2D שולחני או לא מקוון, RMB 50–170k עבור 2D בקו ו-RMB 80–475k עבור 3D בקו, עד RMB 700k עבור 3D קוריאני (inline AOI, 1688 offline AOI, 1688 3D AOI) single source, list prices. מכרז מנובמבר 2025 הגביל 3D AOI אחד בתוספת 3D SPI אחד ב-(tender notice) RMB 800k. התשקיף של Sinic-Tek מרמז על RMB 428–512k לכל 3D AOI ב-2019–2022, כ-RMB 340–470k לפי דוח 2025 שלה (Sinic-Tek prospectus, Sinic-Tek 2025 annual report) primary + estimate. ספק בחו"ל מעריך 20,000–$3,000 עבור 2D שולחני, $30–70k 2D בקו, $50–90k 3D כניסה ו-$100–200k+ 3D יוקרתי, עם אותה מכונה בקו ב-$30–50k בסין (Averroes) vendor estimate from a seller of competing AI software.

לביקורת AI עצמה אין מחיר RMB ציבורי; ספקים מצטטים מקרה אחר מקרה אחרי הערכה חינמית באתר (Sohu on Hengtianyi), וחיפושים ברשימות מסחר אלקטרוני, מכרזים, מאמרי ספקים ומדיה טאיוואנית לא העלו ציטוט ציבורי (1688, 51SMT). העוגנים: כ-€27k לקו המשתמע ממקרה $21,924 ; Siemens מחירון למודול רשת ביקורת מרחוק של מפיץ אמריקאי (Manncorp) estimate ; כ-RMB 298k ללקוח לשנה עבור חיישני ה-edge AI של SmartMore ב-2025 (36Kr) vendor list price. 2D AOI מקומי בקו שנמכר כ"AI" מפורט ב-(1688 inline AOI) RMB 66.5k, וחוצץ NG לבדיקה ידנית חוזרת כ-1688 AOI) RMB 50k review). כך שהתקרה לתוסף ביקורת של צד שלישי היא כ-RMB 10–30k לקו, או עמלה שנתית דומה; מעבר לזה, הקונה יעדיף להחליף את המכונה ב-AOI מקומי עם AI מובנה estimate.

עוגן	מחיר (RMB)	מה הוא קונה	מקור
תחנת תיקון AOI או דונגל תוכנה (TRI ישן, Zhenhuaxing)	≈1,900–2,375	מפתח רישיון, ללא AI	1688 list price
זמן ריצה כללי של ראייה בלמידה עמוקה (דונגל Hikrobot VisionMaster)	≈4,800–6,975 חד-פעמי למחשב	הפלטפורמה; הלקוח בונה ומתחזק מודלים	1688 list price

עוגן	מחיר (RMB)	מה הוא קונה	מקור
פריסה מקומית של Baidu EasyDL	10,000 למודל למכשיר לשנה, או 50,000 קבוע	שירות הסקה ארוז	Baidu AI docs 2022 price page
AOI מקומי בקו שנמכר כ"AI"	98,900–66,500 למכונה	החלפת מכונה	1688 list price
חיישני edge AI של SmartMore	≈298,000 ללקוח לשנה	פתרון ספק על פני כמה מכונות	36Kr single source

FIG 18 **סולם המחירים*מכמה אלפי RMB של רישיון תוכנה ועד RMB 200–300k לתא רובוט או הצעת AI של ספק.***



מקור: רשימות 1688, תיעוד Baidu AI (רשימת מחירים 2022), 36Kr, ראיון תעשייתי (לא מאומת).

עמלה שנתית בסגנון Baidu, עם שרת הסקה אחד לאחד עד ארבעה קווים ומודל אחד עד ארבעה כל אחד, מרמזת על מנוי של RMB 2.5–40k לקו לשנה, ריאלית כ-RMB 5–10k; ב-20% מרישיון קבוע, מנוי ורכישה משתווים בשנה החמישית, מבנה שכדאי להעתיק estimate. קונים רגילים לעמלות תוכנה של כמעט אפס: כל ההכנסה ה"אחרת" של Sinic-Tek (חלפים, שירות, השכרה, אביזרים) הייתה 7.65% ממכירות 2025, כ-RMB 2,300–3,100 למכונה מותקנת לשנה (Sinic-Tek 2025 annual report) primary + estimate. מנוי AI של RMB 5–10k לקו לשנה הוא פי שניים עד חמישה ממה שקונים מוציאים כעת על תוכנת AOI, כך שהוא חייב להיות מוצדק על ידי שעות הביקורת שהוא משחרר (כ-RMB 75–90k לאדם במשרה מלאה לשנה), לא בהשוואה לעמלות תוכנה קיימות.

עבור זרועות רובוטיות, מחיר הקובוט הממוצע בסין ירד לכ-RMB 58,600 ב-2025 (GGII) estimate, והממוצע של שישה הצירים של Dobot ירד 19.1% באותה שנה כשמודלים זולים יותר צברו נתח (Huxiu) verified, אף שמחירי המחירון של אותם מודלים עלו 5–15% ב-2026 (Sina Finance). בארה"ב, FANUC CRX-10iA הוצג ב-$51,434 (Vention) ואוספים מציבים UR7e ב-$30–48k (cobotcost) estimate. הזרוע היא רק 40–50% מתא אמריקאי (Standard Bots); בסין, מצלמות, תופסנים, בטיחות והנדסה רצים בקלות פי שניים עד ארבעה מזרוע של RMB 40k, אז תקצב עלות הנדסה מוחלטת ולא מכפלה estimate.

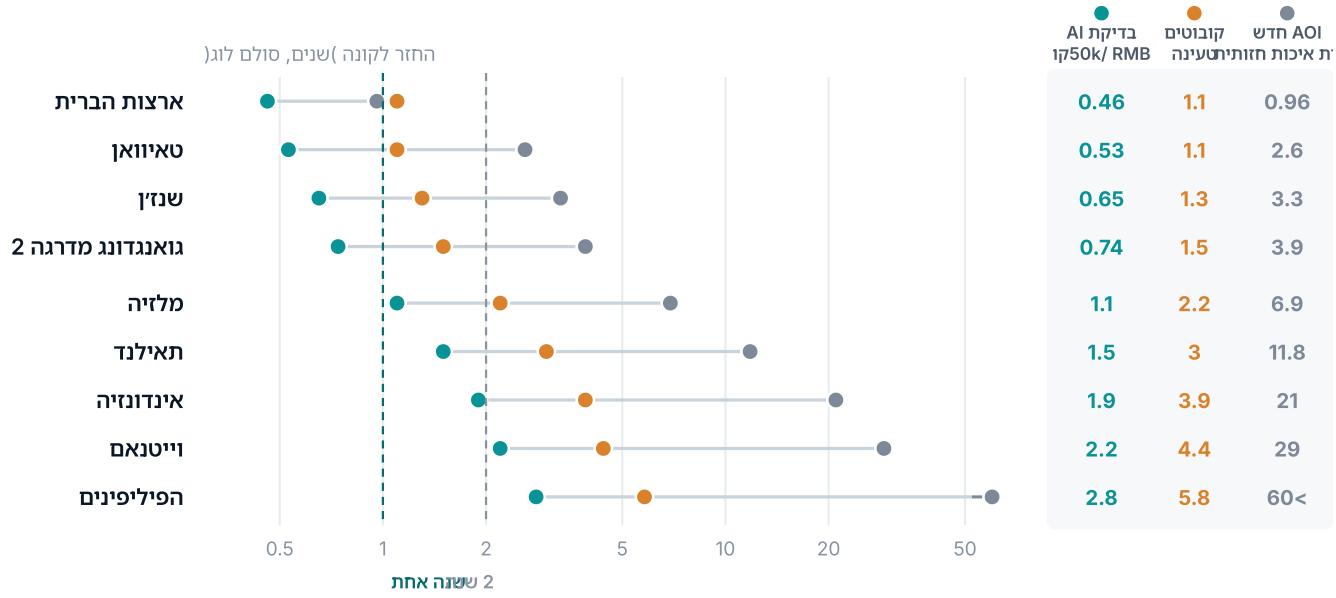
החזר השקעה: ביקורת תוך חודשים, טיפול תוך שנים, הוצאה לעולם לא

החזר פשוט = השקעה ÷ (אנשים שנחסכו × עלות עבודה מלאה − עלות תפעול שנתית), לקו בשתי משמרות ללא סבסוד או מכסים. ציוד טאיוואן מתומחר פי 1.3 מסין, וייטנאם, מלזיה ותאילנד פי 1.15, אינדונזיה והפיליפינים פי 1.2, וארה"ב במחירי חו"ל estimate.

תרחיש	השקעה (סין / ארה"ב)	אנשים שנחסכו	עלות תפעול שנתית
(a) תיבת ביקורת AI או תיבת מצלמת AI	RMB 50k / $30k לקו	1.0 לקו	15%
(b) 2D AOI חדש בקו המחליף בדיקות חזותיות ידניות (0.5 מבקר למשמרת נותר)	RMB 200k / $60k	1.0	12%
(c) תא טעינת קובוט (תחנה אחת × שתי משמרות)	RMB 200k / $140k	2.0	8%
(d) תא זיהוי-והוצאה משולב (בדיקה 1.0 + הוצאה 0.07)	RMB 300k / $200k	1.07	8%
(e) זרוע רובוטית שנוספה רק כדי להוציא לוחות NG	RMB 150k / $100k	0.07	8%

אזור	(a) ביקורת AI	(b) AOI חדש	(c) תא קובוט	(d) זיהוי-והוצאה	(e) זרוע הוצאה
שנזן	0.65 שנה (0.80–0.54)	3.3 (4.4–2.6)	**1.3** (1.6–1.1)	4.5 (5.9–3.6)	אף פעם
גואנגדונג מדרגה 2	0.74	3.9	1.5	5.3	אף פעם
טאיוואן	0.53	2.6	1.1	3.6	אף פעם
מלזיה	1.1	6.9	2.2	9.2	אף פעם
תאילנד	1.5	11.8	3.0	15	אף פעם
אינדונזיה	1.9	21	3.9	25	אף פעם
וייטנאם	2.2 (4.5–1.6)	29	4.4	34	אף פעם
הפיליפינים	2.8	>60 (עלות תפעול ≈ חיסכון)	5.8	≈100	אף פעם
ארה"ב	0.46	**0.96**	**1.1**	3.4	אף פעם

FIG 19 **החזר לפי אזור*בדיקת AI מחזירה תוך חודשים בארה"ב, טאיוואן ושנזן ובתוך 1–3 שנים בדרום-מזרח אסיה; טיפול קובוט תוך שנים.***



זרוע רובוט רק להוצאת לוחות NG אף פעם לא מחזירה את עצמה, בשום מקום
עם תפוקת מעבר ראשון מעל 90%, ההוצאה לוקחת רק 0.003–0.14 FTE למשמרת

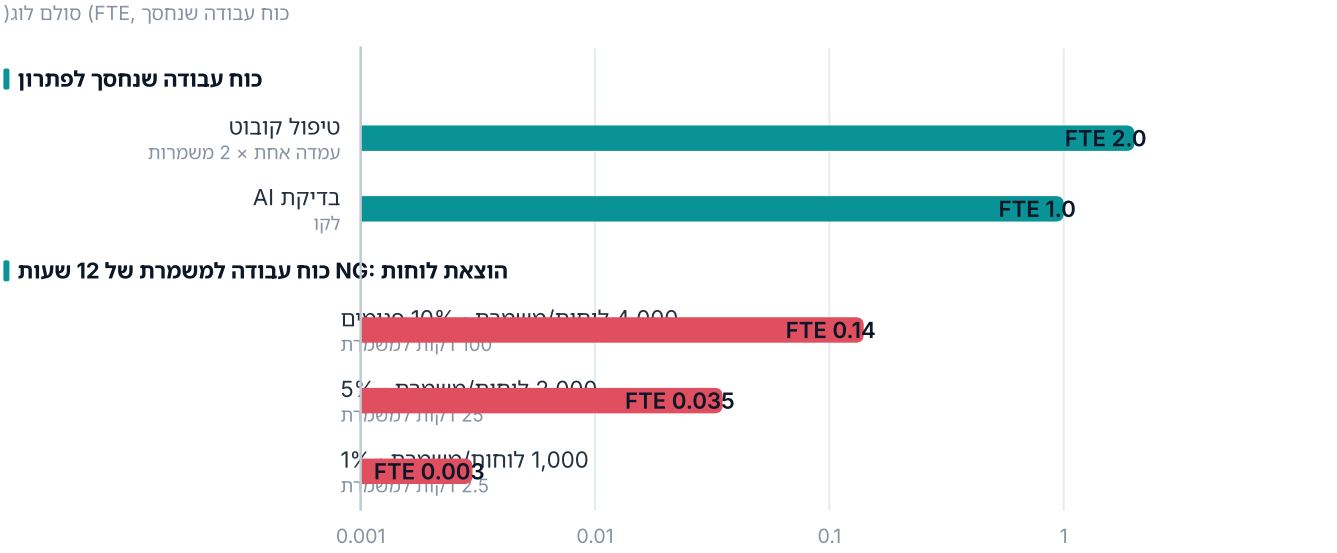
מקור: מודל המחבר: תיבת בדיקת AI של RMB 50k לקו; תא טיפול קובוט ב-RMB 200k או $140k, שתי משמרות.

כל הערכים הם הערכות מהקלטים לעיל. רגישויות: חצה את האנשים שנחסכו וכל החזר בערך מכפיל את עצמו; הרץ משמרת אחת ו-(b)–(d) מכפילים; פחת של 100% בשנה הראשונה בארה"ב מקצר את ההחזר בכ-20%, בעוד מכסים על ציוד סיני, יפני או קוריאני מאריכים אותו בכ-10–40%; סבסוד מקומי של 15% בסין מקצר אותו בכ-15%. לא מצאתי סקר בלתי תלוי של ספי ההחזר של קונים; ספקים אומרים ש-AOI מחזיר "בדרך כלל תוך 9–18 חודשים" (Averroes) ומקרה Siemens טוען שמונה חודשים vendor claims . אני משתמש בשנה ושנתיים כקווי סינון. המודל משמיט רווחי איכות (אחריות, החזרות) ועלויות תחלופה והכשרה, כך שהוא שמרני.

החלק הנגדי-לאינטואיטיבי של "זיהוי והוצאה" הוא שכוח העבודה הוא בביקורת, לא בלקיחת לוחות ה-NG האמיתיים המעטים מהקו. ב-15 שניות טיפול ללוח estimate NG:

לוחות למשמרת	שיעור פגמים	זמן הוצאה למשמרת 12 שעות	אנשים
1,000	1%	2.5 דקות	0.003
2,000	5%	25 דקות	0.035
4,000	10%	100 דקות	0.14

FIG 20 **מדוע הוצאת לוחות גרועים אינה משתלמת *מעל 90% תפוקת מעבר ראשון, ההוצאה לוקחת רק 0.003–0.14 של אדם למשמרת.***



מקור: הערכות המחבר.

מקרה הבסיס (2,000 לוחות למשמרת, 5% פגומים, שתי משמרות) הוא 0.07 של אדם, בשווי כ-RMB 6,000 לשנה בשנז'ן, פחות מעלות התפעול השנתית של הזרוע של כ-RMB 12,000, כך שהוא לעולם לא מחזיר את עצמו בשום מקום; אפילו המקרה הגרוע ביותר (4,000 לוחות למשמרת, 10% פגומים) לוקח כ-13 שנים בשנז'ן ו-9 בארה"ב estimate . אם הדאגה היא עצירת הקו כדי להוציא, מפנה NG או מסוע חוצץ תקני פותר את זה בלי זרוע. אנשים בתעשייה אומרים "מעל 90% תפוקת מעבר ראשון, עצור את הקו והוצא את הלוח ביד", וזה נכון להוצאה, אבל לא לבדיקה. תכנון המוצר על קו SMT הוא לכן "בדיקה בתוספת סימון או הפניה", עם זרועות רובוטיות שמורות למקומות שבהם כל חלק מטופל (טעינת מתקני בדיקה, טיפול במגשים ובמחסניות, אריזה), שם הבדיקה הופכת לתוסף כמעט חינמי ותרחיש (c) חל.

סבסוד, מס ומכסים מזיזים את ההחזר ב-15–25%, אבל לא את הדירוג

סין היבשתית: התמיכה בעיקר מורידה את עלות המימון; מכונות בודדות לעיתים רחוקות זכאיות ישירות. סבסוד הלוואה לאומי משלם 1.5 נקודות ריבית עד שנתיים ב-14 שרשראות עדיפות (Yicai) verified , כ-RMB 9,000 על מכונה של RMB 300k; פרס השדרוג של גואנגדונג עד 20% דורש פרויקט רשום עם לפחות RMB 7 מיליון של ציוד חדש ומוציא את שנז'ן (Guangdong rules), ודונגגוואן שילמה בפועל 4.42% ביולי 2026 (Dongguan allocation) single source, primary . מכשיר בדיקת AI מאושר רק כציוד "עם תוכנה תומכת" בתוך הפרויקט הרשום של הקונה, כך שמקרה הבסיס מניח ללא סבסוד.

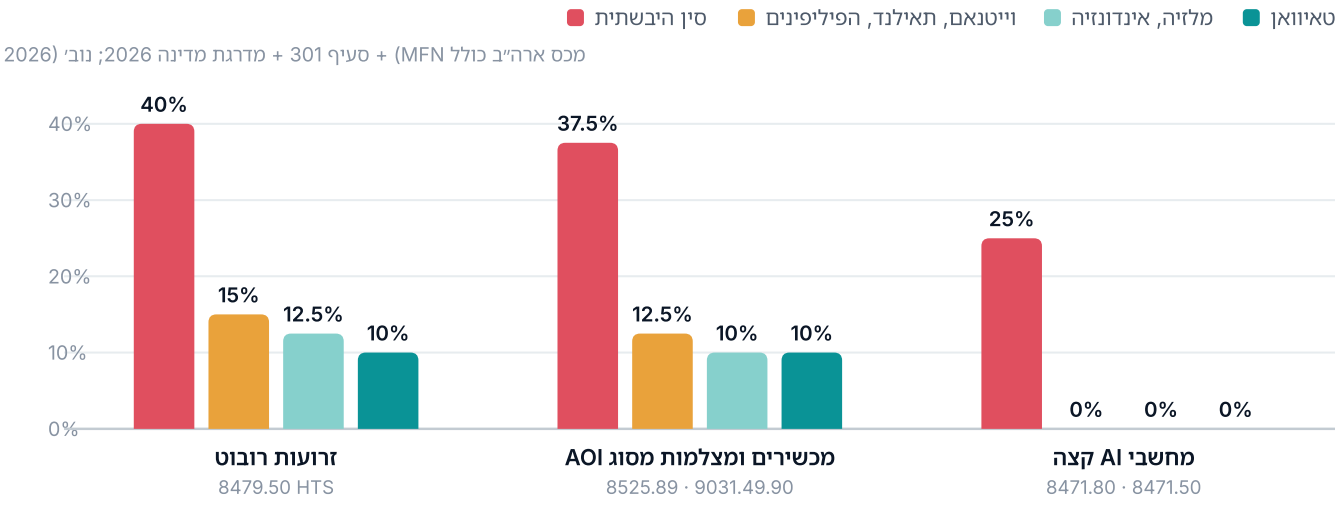
טאיוואן: התמריצים הולכים לקונה. תחת סעיף 1-10 לחוק החדשנות התעשייתית, שתוקן במאי 2025 כדי לכלול AI, קונה יכול לזכות 5% מהוצאות 2025–2029 על מכונות חכמות חדשות או AI (כולל בדיקה מבוססת CNN/DNN) כנגד מס ההכנסה של אותה שנה, או 3% בשנה למשך שלוש שנים (Statute, credit rules) verified ; מענקים בצד הספק סגורים להון סיני (MOEA) single source .

דרום-מזרח אסיה: רק מלזיה ותאילנד מתגמלות קניית אוטומציה. קצבת ההון לאוטומציה של 200% במלזיה על ה-RM10 מיליון הראשונים שווה כ-24% מה-CAPEX לקונה רווחי (MIDA, PwC Malaysia) verified ; ה-BOI של תאילנד נותן פטור ממס לשלוש שנים עד 50–100% מהשקעת השדרוג, התקרה הגבוהה יותר כשספקי אוטומציה תאילנדיים מעורבים (BOI) single source, primary , כך שכניסה לתאילנד עובדת הכי טוב עם אינטגרטור מקומי. וייטנאם, אינדונזיה והפיליפינים מתגמלות בניית מפעלים ולא קניית אוטומציה (Vietnam Law Magazine, PwC Philippines).

ארה"ב: פחת מקצר את ההחזר; מכסים מאריכים אותו. נכסים שנרכשו אחרי 19 בינואר 2025 מקבלים פחת קבוע של 100% בשנה הראשונה, מגן מס פדרלי של כ-21% לקונה רווחי (IRS Notice 2026-11) single source, primary ; חקירת סעיף 232 של רובוטים ומכונות תעשייתיות הסתיימה ב-28 באוקטובר 2026 במשא ומתן ולא במכסים חדשים, להערכה מחדש בתוך 180 יום (White House, Mayer Brown) verified . הרכבה סופית בדרום-מזרח אסיה מקצצת את התוספת על זרועות ומכשירי AOI מכ-37.5–40% ל-10–15% (פרק 1).

אזור	מכשירים עיקריים	השפעה על החזר הקונה estimate
סין היבשתית	סבסוד הלוואה של 1.5 נקודות עד שנתיים; פרסי שדרוג מחוזיים ועירוניים (גואנגדונג דורשת פרויקטים רשומים של ≤RMB 7M; דונגגוואן שילמה בפועל 4.42%); צעדי שנזין 2026–2029 (שיעורים לא ציבוריים)	כ-0–5% למכונה בודדת; 4–15% בתוך פרויקט גדול
טאיוואן	זיכוי סעיף 1-10 לחוק, 5% או 3% × 3 שנים	קצר בכ-5%, אם הקונה משלם מספיק מס
מלזיה	קצבת הון לאוטומציה של 200% (RM10M הראשונים)	קצר בכ-15–25% לקונים רווחיים
תאילנד	פטור ממס BOI לשלוש שנים עד 50–100% מההשקעה	קצר בכ-15–25% לקונים רווחיים מחוץ לחופשת מס
וייטנאם, אינדונזיה, הפיליפינים	תמריצי מס הכנסה לבניית מפעלים; פחת ציוד נוסף של 20% בפיליפינים	כמעט כלום לרכישות אוטומציה בודדות
ארה"ב	פחת של 100% בשנה הראשונה; ≈37.5–40% מכס משולב על ציוד מתוצרת סין היבשתית	פחת מקצר ≈20%; מכסים מאריכים ≈10–40%

FIG 21 ***מכס ארה"ב על רובוטים, AOI ומחשבי קצה שיוצרו בסין היבשתית משלמות כ-40%; שנשלחו מדרום-מזרח אסיה או טאיוואן, 10–15%.***



מקור: לוח תעריפים מותאם של USITC פרק 99; USTR ;Federal Register (נובמבר 2026).

לפי אזור: בסין היבשתית, תיבת ביקורת AI מחזירה תוך חודשים ותא קובוט תוך 1.5–1.3 שנים, בעוד AOI חדש שנרכש רק כדי לקצץ עבודה לוקח 3–4 שנים וזיהוי-והוצאה משולב 5.3–4.5, שניהם מעבר למסך של שנתיים. טאיוואן עושה מעט טוב יותר משנז׳ן (0.53 שנים לביקורת, 1.1 לתא קובוט), והגיוס שם קשה יותר. במלזיה ובתאילנד, תיבת AI לוקחת 1.5–1 שנים ותא קובוט 3–2, מהיר בכ-15–25% עם תמריצי מס מקומיים לקונים רווחיים. באינדונזיה, וייטנאם והפיליפינים רק תיבת AI זולה מאוד מחזירה תוך 3–2 שנים; הביקוש שם מגיע יותר מדרישות האיכות של הלקוחות ומחסור בעבודה. בארה״ב כל אפשרות חוץ מזרוע הוצאה טהורה מחזירה תוך 3.4–0.2 שנים, עם מכסים, ההערכה מחדש של סעיף 232 והסמכת NRTL כמשתנים העיקריים.

שוק ותחרות

05 שוק, תחרות ואיך קונים משלמים

גודל שוק: מספרים שפורסמו שונים בכפולות; רק ספירה מלמטה-למעלה שימושית

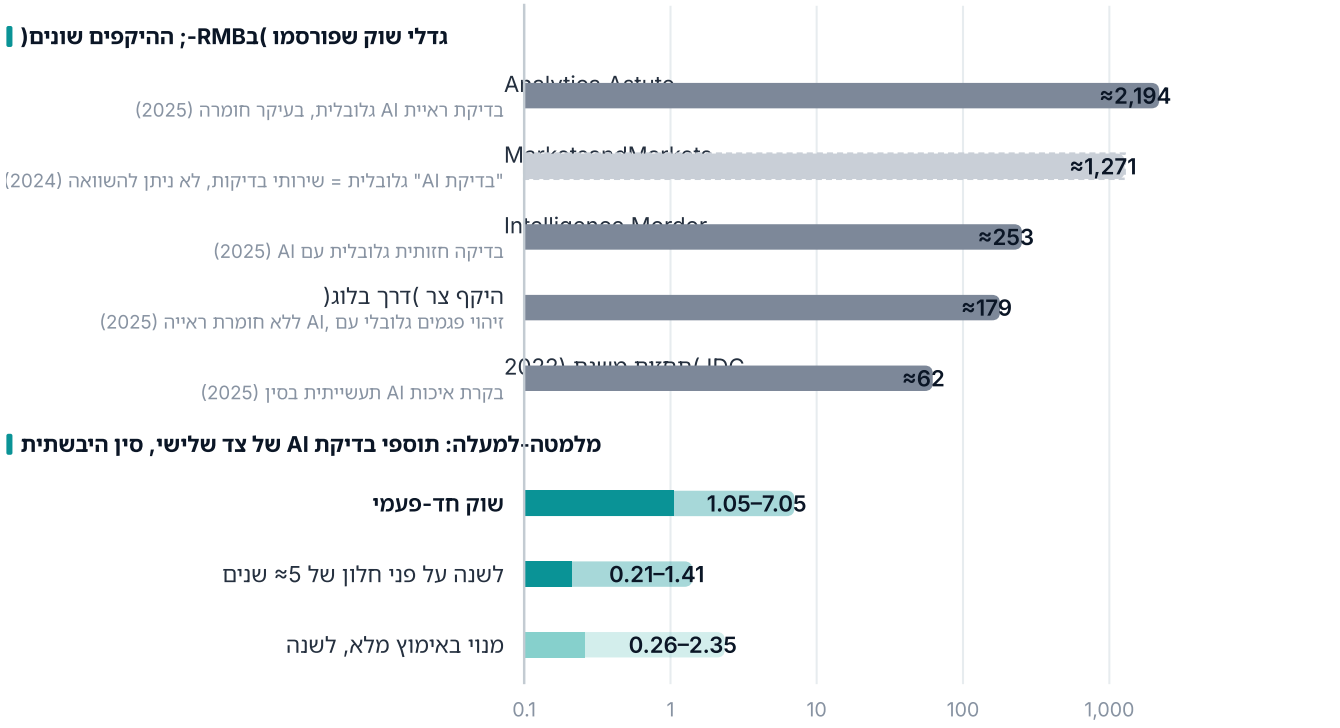
כל מספר שוק חייב להיקרא עם ההיקף שלו. שוק ראיית המכונה של סין היה RMB 18.15 מיליארד ב-2024 (Zhiyan) verified, עם אלקטרוניקת 3C כקונה הגדול ביותר; הערכות גלובליות לראיית מכונה שונות בערך פי שלושה ול-AOI בערך פי 6.5 (MarketsandMarkets, Vision Systems Design/Interact Analysis, Zhiyan). מלמטה-למעלה מההכנסות של Koh Young, TRI, Jutze, Sinic-Tek ו-ViTrox, ציוד בדיקת ה-SMT הגלובלי הוא בסדר גודל של $1–2 מיליארד estimate.

אין שוק עצמאי ל"תוכנת בדיקה חזותית עם AI". המספרים שפורסמו שונים ביותר מפי עשרה וכולם כוללים חומרה או שירותים:

היקף	מפרסם	נתון
גלובלי "זיהוי פגמים תעשייתי עם AI" (צר, ללא חומרת ראייה)	חברה ללא שם, דרך בלוג של iFactory	$2.66 מיליארד ב-2025, CAGR 8.6% עד 2035, אלקטרוניקה 34%
גלובלי "בדיקה חזותית עם AI"	Mordor Intelligence	$3.76 מיליארד ב-2025, $12.8 מיליארד ב-2031, CAGR 22.67%
גלובלי "בדיקה חזותית עם AI" (בעיקר חומרה)	Astute Analytica	≈$32.66 מיליארד ב-2025
גלובלי "בדיקת AI" (שירותי בדיקה והסמכה; לא בר השוואה)	MarketsandMarkets	$18.92 מיליארד ב-2024
סין "בדיקת איכות AI תעשייתית"	IDC, תחזית מ-2022	$958 מיליון ב-2025 (≈RMB 6.2 מיליארד), 91.5% בתקשורת, אלקטרוניקה ושתי תעשיות אחרות

מקורות: (iFactory, GII/Mordor, GII/Astute, MarketsandMarkets, IDC) each single source, mostly paid-report summaries.

FIG 22 **גדלי השוק נבדלים ביותר מפי עשרה*רק ההיקף הצר וספירה מלמטה-למעלה שימושיים להחלטות.***



מקור: Mordor Intelligence, Astute Analytica, MarketsandMarkets, IDC (2022), iFactory; הערכת מלמטה-למעלה של המחבר.

מלמטה-למעלה, יש רק ספירה ציבורית אחת של קווי SMT: New Venture Research העריכה כ-73,231 קווים בעולם ב-2020, 47% מהם בסין (19% ב-2000), כ-70% יוקרתיים ו-30% נמוכים (EMSNow, NVR SMT2020 summary) ; single source, 2020 NVR שייכת כעת ל-IPC ונתונים חדשים הם בתשלום (NVR). צמיחה של 3–5% בשנה עם נתח סין של 40–47% נותנת כ-87,000–98,000 קווים בעולם ב-2026 ו-35,000–46,000 בסין היבשתית (בסיס 40,000) . estimate

קלט	נמוך	בסיס	גבוה	בסיס
קווי SMT בסין היבשתית (2026)	35,000	40,000	46,000	בסיס NVR 2020 × צמיחה × נתח
נתח ללא ביקורת AI	75%	80%	85%	חלון שדרוג לאחור, למטה
נתח שצד שלישי יכול להגיע אליו (לא פנימי, לא קשור ל-AI של יצרן ציוד מקורי, יכול לשלם)	40%	50%	60%	הנחה: ≈30% קווים נמוכים בתוספת כמה קווים יוקרתיים רב-מותגיים, למעט בניות פנימיות בסגנון DBG
קווים ניתנים לשירות	**10,500**	**16,000**	**23,500**	מכפלה
מחיר חד-פעמי לקו	RMB 10k	RMB 20k	RMB 30k	סולם מחירים ותקרה (פרק 4)
שוק חד-פעמי ניתן לשירות (בסיס מותקן)	**RMB 105M**	**RMB 320M**	**RMB 705M**	מכפלה
לשנה על פני חלון של ≈5 שנים	RMB 21M	RMB 64M	RMB 141M	החלון מתכווץ 8–12% בשנה
מנוי (RMB 2.5k / 5k / 10k) לקו לשנה, אימוץ מלא	RMB 26M בשנה	RMB 80M בשנה	RMB 235M בשנה	עוגן עמלה שנתית בסגנון Baidu

FIG 23 **מלמטה-למעלה: 40,000 קווים → 16,000 ניתנים לשירות → ≈RMB 320M חד-פעמיתשת בדיקת ה-AI של *צד שלישי בסין היבשתית היא אמיתית, אך קטנה.***



שחקן חדש שזוכה ב2%–5%- מהקווים הניתנים לשירות ב3- שנים: 800–320≈ קווים, RMB≈ 2–5M בשנה
כמנוי (RMB 5k לקו לשנה:(1.6–4M RMB חוזר, רק 0.3–2.3% מתחזית בקרת איכות האAI- התעשייתית של IDC לסין

מקור: NVR (בסיס 2020); הערכות המחבר.

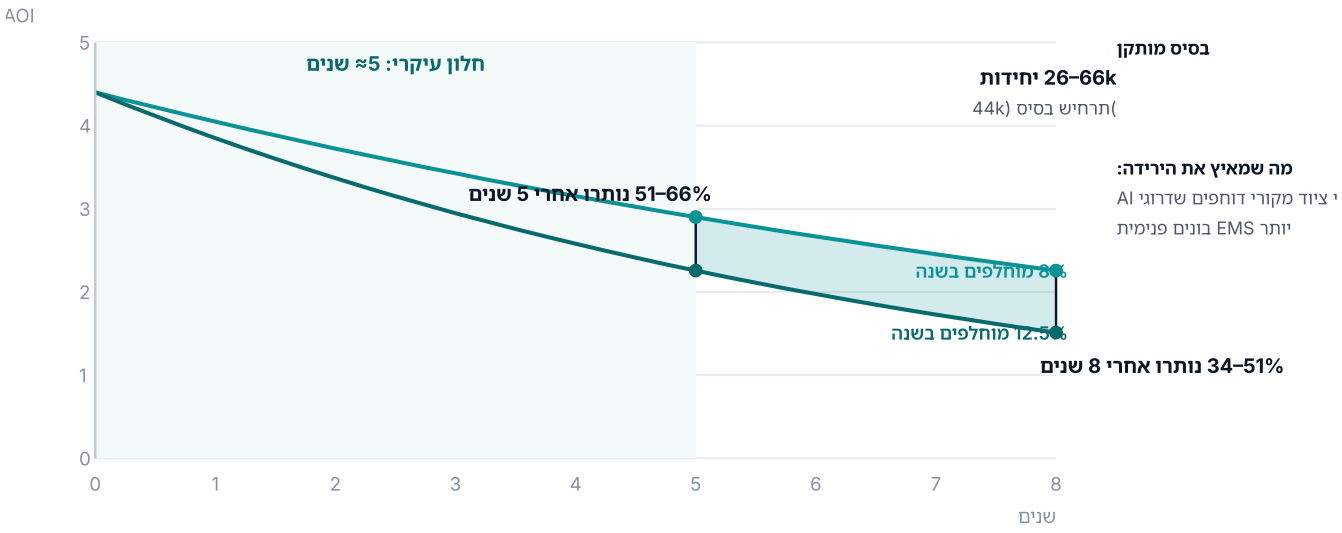
כל אלה הערכות; לא הנתח ולא המחיר נמדדו ציבורית. הנישה שווה רק 0.3–2.3% בשנה מתחזית ה-IDC לכל בדיקת האיכות ה-AI התעשייתית של סין ב-2025 (כ-RMB 6.2 מיליארד). שחקן חדש שזוכה ב-2–5% מהקווים הבסיסיים הניתנים לשירות בשלוש שנים יהיה בעל כ-320–800 קווים: RMB 6.4–16 מיליון בסך הכול ב-RMB 20,000 לקו, כ-2–5 RMB מיליון לשנה, או RMB 1.6–4 מיליון לשנה בהכנסה חוזרת ב-RMB 5,000 לקו estimate . זה מספיק ליחידה עסקית בתוך חברה, לא לחברה עצמאית בקנה מידה של הון סיכון. כל 1,000 קווים מותקנים בטאיוואן, דרום-מזרח אסיה או ארה״ב שווים בערך RMB 3–15 מיליון חד-פעמי במחירי היבשת; המחירים גבוהים יותר בטאיוואן ובארה״ב, אבל קונים שם משתמשים לעיתים קרובות יותר ב-3D AOI של יצרני ציוד מקורי עם אפשרויות AI estimate.

חלון השדרוג לאחור הוא כחמש שנים

אף יצרן AOI לא מפרסם חיי מכונה; חברות EMS מפחיתות ציוד על פני 3–10 שנים (DBG 2025 annual report), ו-AOI ישנים יותר מאבדים עדכוני תוכנה וחלפים אחרי סוף חיים (I-Connect007) single source . אני משתמש ב-8–12 שנים (בסיס 10).

ב-1.0–1.65 AOI בקו לקו, לסין היבשתית יש כ-35,000–78,000 AOI בקו בשירות (בסיס כ-55,000), עקבי עם משלוחים ידועים (Circuits Assembly, Sinic-Tek 2025 annual report). אין מקור שמודד כמה מריצים ביקורת AI; בהנחה שמכונות מקומיות מציעות זאת מאז כ-2023 ו-40–60% מהמשלוחים האחרונים משתמשים בה, **כ-26,000–66,000 AOI בקו בסין היבשתית (בסיס כ-44,000) אינם מריצים AI** estimate. אם כל מכונה חדשה נשלחת כעת עם AI, המאגר הזה מתכווץ 8–12.5% בשנה: 51–66% נותרו אחרי חמש שנים ו-34–51% אחרי שמונה, מהר יותר אם יצרני AOI דוחפים שדרוגי AI למכונות חדשות יותר. **רוב ההזדמנות נופלת בחמש השנים הקרובות בערך**, עם EMS בינוני שמריץ 2D AOI מקומי ישן, מותגים מעורבים וללא נתיב שדרוג AI בליבה שלו.

FIG 24 **חלון השדרוג לאחור הוא כחמש שנים *AOI בקו ללא AI מתכווצים ב-8–12.5% בשנה כשהם מוחלפים או משודרגים.***

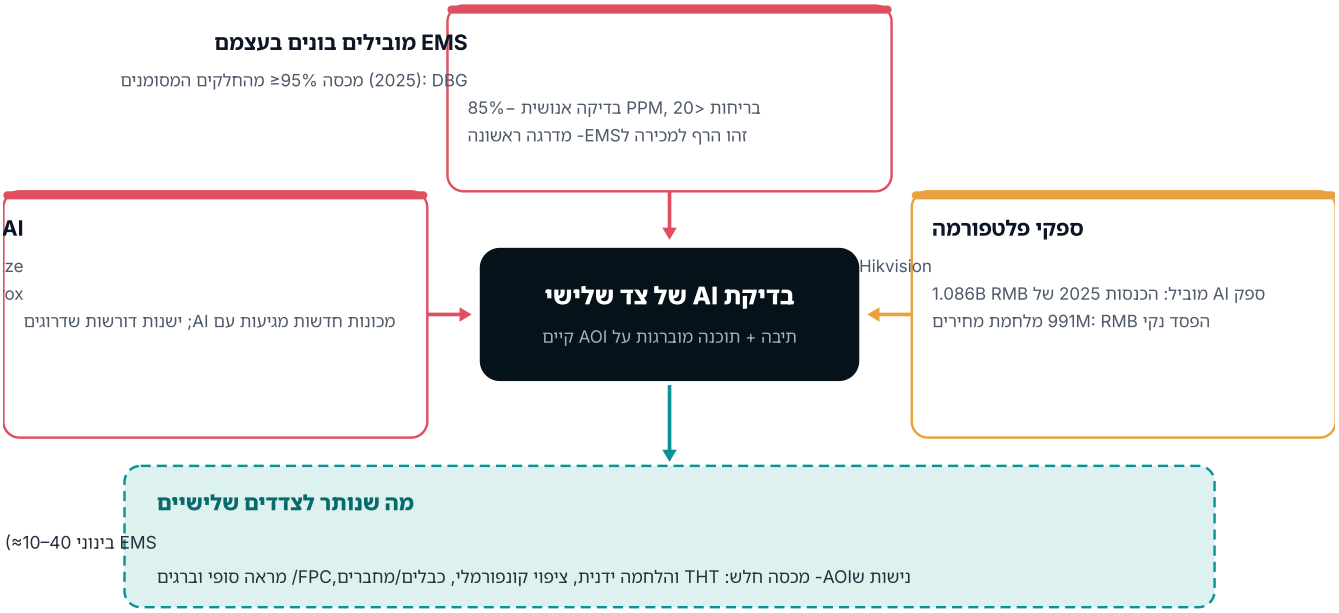


מקור: הערכות המחבר; הבסיס המותקן והנתחים לא אומתו באופן עצמאי.

המתחרים כבר חזקים: טריז הביקורת מכותר

טריז ביקורת ה-AI מצטמצם משלושה כיוונים. מובילי AOI וראייה בונים AI לתוך פלטפורמות חדשות: Jutze "אימצה אלגוריתמי AI במלואם" (Jutze investor relations, June 2026), הכנסות ה-AI התעשייתי של OPT צמחו 96.77% ב-2025 (OPT 2025 annual report) single source, primary , ו-ViTrox טוענת לעד 95% דיוק בביקורת אוטונומית (ViTrox) vendor claim . יצרנים גדולים בונים משלהם (iThome). וחברות AI מגובות הון סיכון מוכרות בקנה מידה: לקוחות חיישני הקצה של SmartMore צמחו מ-65 ל-251 בשנתיים (36Kr), Hikvision single source מוכרת את מודל Guanlan שלה לבדיקה (Xinhua), ושילוב "עין, מוח ויד" כבר מוצע מ-SmartMore, Hikrobot ו-(36Kr, Securities Star) Mech-Mind.

FIG 25 **טריז הבדיקה מכותר משלושה צדדים *מקום לצדדים שלישיים הוא ב-EMS בינוני ובנישות בדיקה ש-AOI מכסה חלש.***



מקור: דוחות שנתיים של Jutze, Sinic-Tek ו-DBG; חומר ציבורי של Hikvision ו-SmartMore; קובץ על ידי המחבר.

המרחב שנותר לשחקן חדש הוא צר: EMS בינוני עם 2D/3D AOI ישן וללא שדרוג AI מהיצרן; קווים שמערבבים כמה מותגי AOI (ספק שדרוג אחד כבר תומך ב-Omron, VI, Koh Young ו-(Sohu on Hengtianyi) Mirtec); נישות ש-AOI של SMT מכסה בצורה חלשה, כמו חיבורי THT והלחמה ידנית, ציפוי קונפורמלי, כבלים, מחברים ו-FPC, ומראה סופי וברגים; ומפעלים קטנים שאינם יכולים להרשות לעצמם AOI של 230–280k RMB.

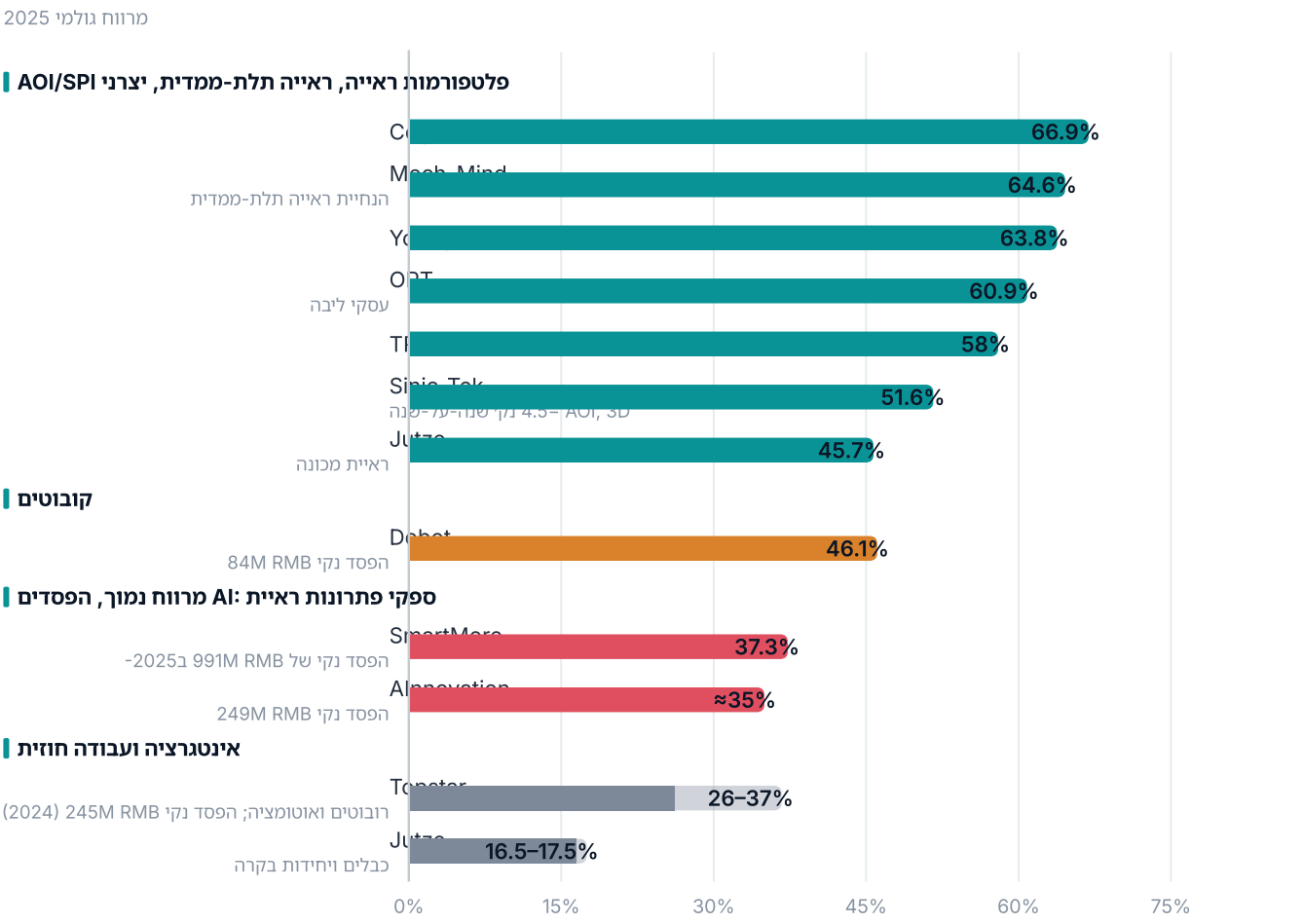
החברה הנסחרת הקרובה ביותר להשוואה נותנת אזהרה. Weiyi Zhizao, שמוכרת ציוד בדיקת AI ורובוטים, ראתה הכנסות 3C נופלות 79.9% שנה-על-שנה בינואר–מאי 2026, הכירה ב"תחרות מתעצמת ולחץ מחירים", דחתה את ההנפקה בהונג קונג ב-25 באוקטובר 2026, וחשפה 15 פרויקטים שהופסקו לפני השלמה, 12 מהם פרויקטי רובוט מגולם (Weiyi Zhizao prospectus, postponement notice) AInnovation . single source, primary סיימה את 2025 עם 29.8% מהחייבים שלה בהפרשה (AInnovation 2025 annual report). טריז בדיקת ה-3C צפוף, וכוח התמחור לקו חלש.

המרווח יושב בחומרה ממוצרת

מיקום בשרשרת	חברות מייצגות ומרווח גולמי 2025	רווח וגבייה
פלטפורמות ראייה	Cognex 66.9%	רווחי; מוכר ≈$22M של הכנסות במרווח נמוך
ציוד AOI/SPI	Koh Young 63.8%; TRI 58%; OPT 60.85%; Sinic-Tek 3D עסקי ליבה 51.62% AOI (−4.53 נקודות שנה-על-שנה); Jutze ראיית מכונה 45.74%	רווחי; ימי חייבים 139 (Jutze) עד 247–274 (OPT)
הנחיית ראייה תלת-ממדית	Mech-Mind 64.6%	—
ספקי פתרונות ראיית AI	SmartMore 37.3%; AInnovation ≈35%	מפסידים: SmartMore הפסד נקי RMB 991M ב-2025, AInnovation RMB 249M
קובוטים	Dobot 46.1%	הפסד נקי של Dobot RMB 84M; הכנסות UR מ-($326M (2022 ל-≈$232M (2025, תחזית)
אינטגרציית מערכות	Topstar רובוטיקה ואוטומציה 26.2–36.7%	הפסד נקי (2024) RMB 245M; ימי חייבים 180–240
חומרה חוזית	Jutze כבלים ויחידות בקרה 16.5–17.5%	—

מקורות: Cognex, StockAnalysis on Koh Young, Fugle on TRI, OPT 2025 annual report, Sinic-Tek 2025 annual report, Jutze 2025 annual report, Jiemian on Mech-Mind, 36Kr on SmartMore, Yicai on SmartMore, TMTPost, DoNews on AInnovation, The Paper on Dobot, The Robot Report on UR, Jiemian on Topstar).

FIG 26 ***המרווח יושב בחומרה ממוצרתפלטפורמות ראייה ויצרני ציוד מרוויחים 46–67% מרווח גולמי; ספקי פתרונות AI 35–37% ומפסידים כסף.***



מקור: דוחות שנתיים או תשקיפים של 2025.

שחקן חדש שומר על המרווח שלו רק על ידי הפיכת בדיקה למוצר, ציוד או תוכנה; כוונון פרויקט-אחר-פרויקט דוחף מרווחים לכ-35% עם הפסדים מתמשכים. המכניקה גם פועלת נגד חברות קטנות: הכנסה מוכרת רק אחרי קבלה באתר, מקדמות קטנות (Jutze, OPT), פיילוטים בדרך כלל חינמיים (A3/Mindtrace), ובעלי מותגים לעיתים קרובות מכתיבים את ציוד הבדיקה בקווי יצרני החוזה שלהם.

איך קונים משלמים: 30% בחתימה, 60% אחרי קבלה, 10% מוחזקים לשנה

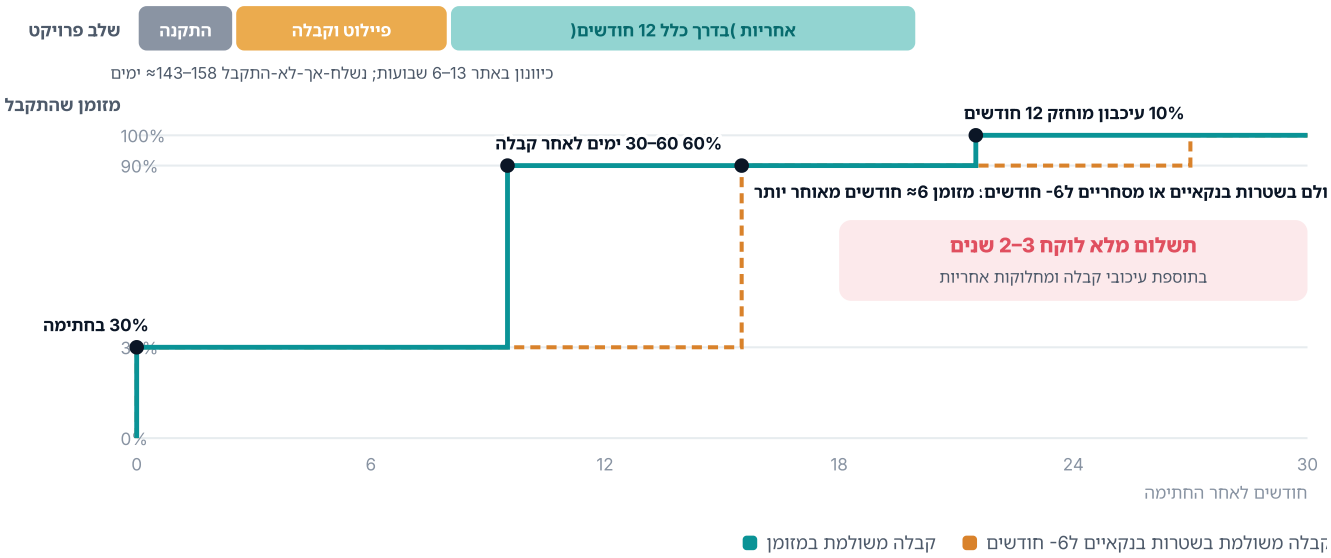
תנאי התשלום שמפעלי אלקטרוניקה בסין היבשתית משתמשים בהם לציוד בדיקה מפורטים ב-13 חוזי קונים נקובים בשם שנחשפו בתשובת בירור ההנפקה של Sinic-Tek (Sinic-Tek inquiry reply 8-1-1) single source: one primary filing, several independent buyers, 2020–2022:

קונה (שנה)	חתימה	אספקה	קבלה	עיכבון או יתרה
יחידת Hikvision (2021)	30% תוך 15 ימים	—	65% תוך 30 ימים מקבלה סופית	עיכבון אחריות 5% משולם 12 חודשים אחרי קבלה; חלק משולם בשטרות בנקאיים של 6+ חודשים
Nanchang Huaqin (2020–2021)	30% תוך 10 ימי עבודה	—	60% ב-60 ימים אחרי קבלה	10% אחרי אחריות של שנה
Guangdong Yinuo (2021, 3D AOI)	30% תוך 10 ימי עבודה	—	60% תוך 30 ימים	10% אחרי שלושה חודשי ריצה ללא תקלות
Xiamen Hongxin, Jingmen Hongyi (2022)	30% תוך 10 ימי עבודה	—	60% תוך 30 ימי עבודה מקבלה סופית, שטרות בנקאיים של 6 חודשים	10% אחרי אחריות של שנה

קונה (שנה)	חתימה	אספקה	קבלה	עיכבון או יתרה
AKM Panyu, Suzhou (2021–2022)	30% תוך 7–10 ימי עבודה	—	60% חודש אחרי קבלה והוצאת חשבונית, שטרות בנקאיים של 180 ימים	10% אחרי 360 ימים
BYD Precision, Shenzhen, Huizhou (2020)	—	30% בהגעה	70% אחרי קבלה והוצאת חשבונית	הכול בשטרות מסחריים של 6 חודשים
Shenzhen H&T (2020)	—	—	90% אחרי קבלה והוצאת חשבונית	10% אחרי שישה חודשי שימוש ללא תקלות

תיקים ראשוניים אחרים מראים אותו מבנה: הפרויקטים הגדולים של Mech-Mind משלמים 30% בחתימה, 30% לפני אספקה, 30% אחרי קבלה ו-10% עיכבון אחריות על פני 12–15 חודשים (Mech-Mind prospectus); GKG אומרת שעיכבון בענף הציוד הוא בדרך כלל 5–10% מהחוזה, משולם שנה אחרי קבלה (GKG inquiry reply) single sources, primary. הנורמה, אם כן, היא כ-30% בחתימה (20–30%, אפס במפעלים מסוימים כמו BYD), 60–65% אחרי קבלה (משולם 30–60 ימים מאוחר יותר, לעיתים קרובות בשטרות בנקאיים או מסחריים של שישה חודשים, מה שמעכב מזומן שמיש בעוד חצי שנה), ו-5–10% עיכבון ל-3–12 חודשים (12 הנפוץ ביותר) verified across several contracts.

FIG 27 כיצד הקונים משלמים*30% בחתימה, 60% לאחר קבלה, 10% מוחזקים לשנה: תשלום מלא לוקח 2–3 שנים.*



מקור: תשובת בירור של Sinic-Tek, Mech-Mind, GKG, JT ודיווחים אחרים; סכמטי על תנאים רגילים, תזמון מוערך.

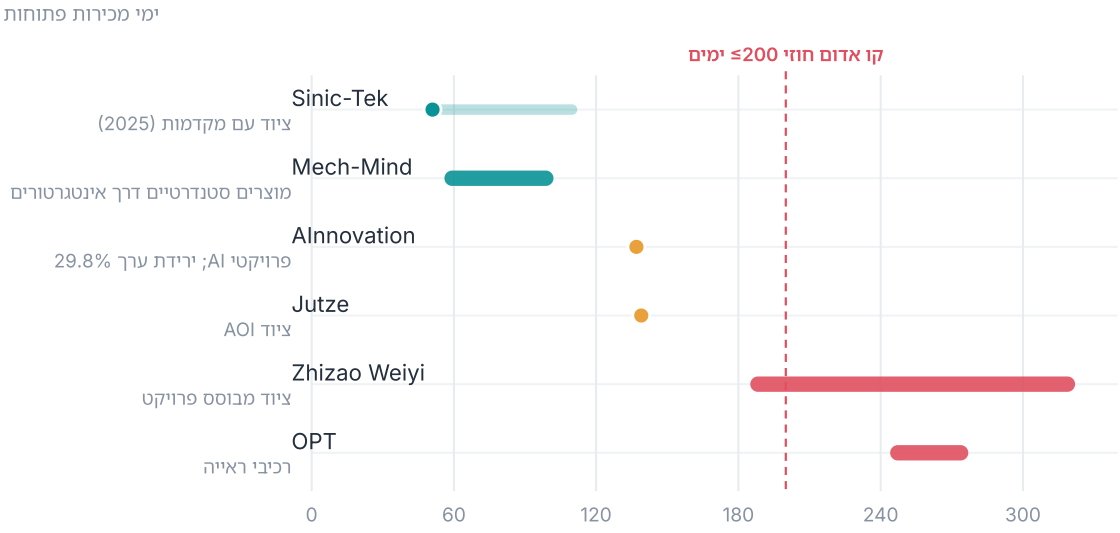
קבלה וניסויים איטיים ומשולמים על ידי המוכר. סחורות שנשלחו-אך-לא-התקבלו של Sinic-Tek לקחו בערך 143–158 ימים להתחלף, כי לקוחות EMS גדולים ומותגים זקוקים לבדיקת אינטגרציה של כל הקו ולהרצת ניסיון ממושכת (Sinic-Tek inquiry reply 8-1-2); פרויקטי ה-AI של Weiyi לוקחים 4–8 חודשים מהערכה לפריסה (Weiyi Zhizao prospectus) single sources, primary. אין נתוני המרה לפיילוטי ביקורת AI; Sinic-Tek המירה רק כ-18–19% ממכונות ההדגמה שלה ישירות למכירות. אני מתכנן אחד לשלושה עד אחד לחמישה לפיילוטים חינמיים וכ-80% לפיילוטים בתשלום אחרי סינון קצר estimate.

חברה	מודל עסקי	ימי מכירות פתוחות
Sinic-Tek (2025)	ציוד עם מקדמות	≈51 (≈112 כולל שטרות)
Mech-Mind	מוצרים תקניים דרך אינטגרטורים	59–99
Jutze	ציוד AOI	≈139
AInnovation (2025)	פרויקטי AI	≈137 בסוף שנה, עם ירידת ערך 29.8%
Weiyi Zhizao	ציוד מבוסס פרויקט	188–319

חברה	מודל עסקי	ימי מכירות פתוחות
OPT	רכיבי ראייה	274–247

מקורות: הדוחות השנתיים והתשקיפים לעיל; חייבים (כולל נכסי חוזה) ÷ הכנסות × 365 primary + estimate .

FIG 28 ימי מכירות פתוחות*מוצרים סטנדרטיים 51–99 ימים; ציוד מבוסס פרויקט 188–319 ימים.*



מקור: דיווחי חברות; חובות (כולל נכסי חוזה) ÷ הכנסות × 365.

רישיונות ביקורת AI שנמכרים דרך יצרני AOI או אינטגרטורים צריכים להיגבות כמו 60–100 הימים של Mech-Mind; מכירות פרויקט ישירות כמו 190–320 הימים של Weiyi. בהרכבה לציר זמן מזומן: פיילוט חינמי חודש עד שלושה חודשים, הסמכה ב-ODM גדול עד כשנים-עשר חודשים, בנייה, אספקה וקבלה שלושה עד שבעה חודשים; 30% מגיע בהזמנה, 60% כארבעה עד תשעה חודשים אחריה וה-10% האחרונים כ-16–21 חודשים אחריה. **מהפיילוט הראשון ב-EMS גדול ועד לתשלום מלא יכולות לחלוף שנתיים עד שלוש**; EMS בינוני ללא מכרזים פורמליים יכול לקצר את ההסמכה לשבועות estimate . קריטריוני הקבלה בחוזים שנחשפו מכסים רק אינטגרציה של כל הקו, תפוקה (פרויקט טאיוואני אחד דרש 100 יחידות לשעה) וחודשי ריצה ללא תקלות; אף אחד לא מפרט שיעורי בריחות או קריאות שווא (Anda inquiry reply) single source, primary . קריטריוני קבלה כתובים לביקורת AI חייבים להיות מוסכמים לפני הפיילוט.

ארבעה דפוסי כשל

כשלים מתועדים נופלים לארבעה דפוסים: משלוח לפני שהמוצר היה אמין (Rethink (The Robot Report); שכבת תוכנה על חומרה של מוביל שהמערכת האקולוגית של המוביל החליפה מאוחר יותר (READY Robotics (The Robot Report); נגמר הכסף בתחתית מחזור (Dextrous Robotics (Robots & Startups); וחברות מודלים שסיימו ברכישות-כישרונות (Covariant (GeekWire). הלקחים: מוצר-בצורה-צרה במקום כוונון פרויקט-אחר-פרויקט; החזק הון חוזר להכנסות מבוססות קבלה וגבייה ארוכה; ספק שירות מקומי למפעלים שרצים 24 שעות; ואל תרחב לפני שהמוצר מוכח.

עיתוי: בדיקה אינה מרוויחה דבר מהמתנה; חומרת זרועות נעשית זולה יותר

הטיעונים להמתנה חלשים. חומרה לא הוזלה ב-2026: יצרני רובוטים העלו מחירים 5–15%, חלק מהמוצרים 25–50% (Sina Finance) single source , וערכות NVIDIA עלו פעם שלישית (36Kr). VLA בדרגת ייצור לא יגיעו בתוך החלון (IFR, Instrumental). בינתיים כל החלפת AOI מסירה יעד שדרוג וכל פריסת AI של יצרן ציוד מקורי נועלת נתוני ותהליך קו; יכולת המודל הופכת למצרך, ומה שחסר הוא גישה למפעלים ולנתוני מפעל מתויגים. לתמיכה מקומית יש גם חלונות: כללי "עיר חלוצת AI" של שנזין 2026, עם עד RMB 2 מיליון לפרויקטי הדגמת יישומי AI ושוברי מודל שמכסים 30% מהרכישות, רצים עד **30 ביוני 2027** (Yicai, May 2026) terms single source .

מה שעדיין נהיה זול מהר הוא רובוטים דמויי אדם וידיים מיומנות (ידיים ירדו מכ-50,000 RMB לפחות מ-10,000 בשנתיים (Sina Finance)), וקובוטי כניסה כשתמהיל המוצר יורד; זה תומך בדחיית החלק המגולם, לא חלק הבדיקה. **בסך הכול: בדיקה עכשיו,**

כמוצר צר; תאי רובוט עכשיו, דרך שותפים ואינטגרציה; AI מגולם מאוחר יותר, עם תחנת מו"פ בלבד היום.

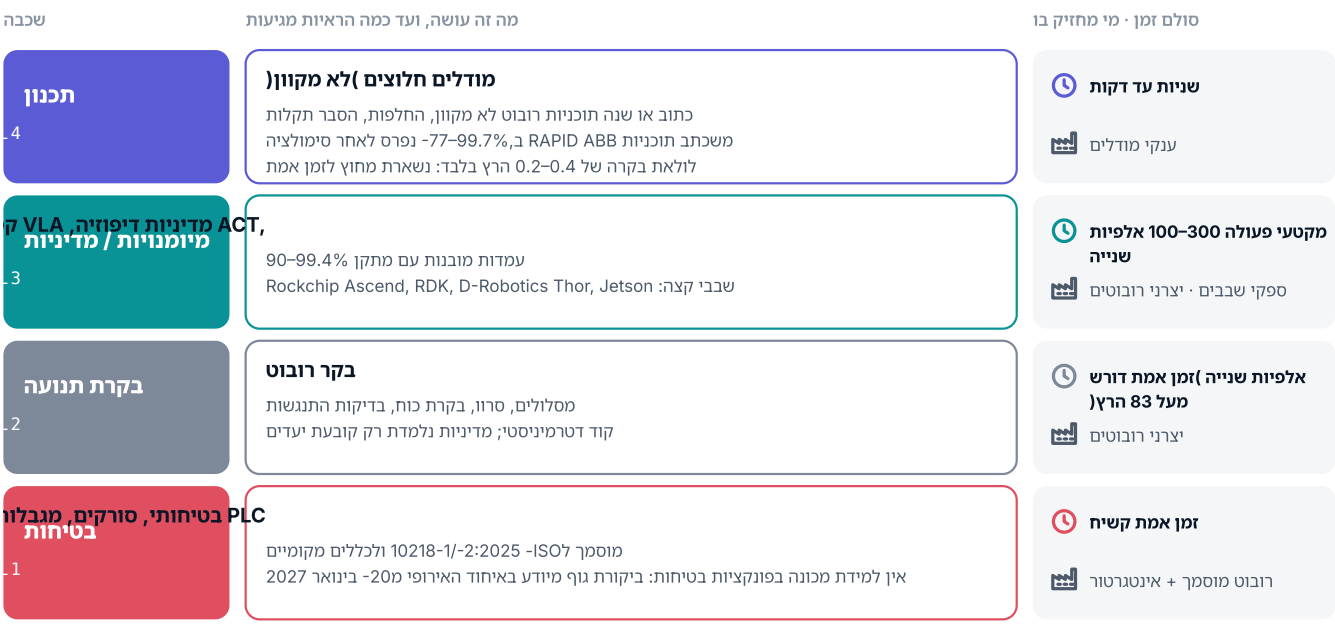
AI מגולם

06 AI מגולם: קודם הוא משנה איך מלמדים רובוטים, אחר כך איך הם נעים

מה שמשתנה הוא תפיסה ותכנות; התנועה נשארת דטרמיניסטית

כל מוביל זרועות רובוטיות משווק כעת "AI פיזי", אבל מה שנשלח בפועל נופל לשלוש שכבות: תפיסה נלמדת ללא לימוד (AI Masterless Vision של FANUC, נשלח בינואר 2027 (FANUC); בחירת אחיזה עצבית של (KUKA) KUKA); מודלים גדולים שכותבים תוכניות ומתכננים משימות (iiQWorks.Copilot של KUKA, שמזהיר שיש לאמת קוד שנוצר על ידי AI בתאום דיגיטלי (KUKA); Yaskawa עם Gemini Robotics-ER, שבו "AI גנרטיבי מחליט מה לעשות" והרובוט מחליט "איך לנוע" (Yaskawa)); ופלטפורמות חישוב ונתונים (AI Accelerator של UR עם (Jetson Orin (Universal Robots vendor claims . אף מוביל לא מפרסם שיעורי הצלחת ייצור לתכונות ה-AI שלו, והמובילים היפניים משתפים פעולה עם Google, SoftBank, Microsoft ואחרים במקום לבנות VLA חלוציים (FANUC news index): שכבת המדיניות הנלמדת תבוא מחברות המודלים, הזרועות והבקרים המוסמכים מיצרני הזרועות.

FIG 29 **ערימת ה-AI המגולם*מודלים גדולים מתכננים לא מקוון, מדיניות קטנה רצה בתא, תנועה ובטיחות נשארות דטרמיניסטיות.***



מקור: סינתזת המחבר (פרקים 3 ו-6); ISO 10218:2025; תקנת המכונות של האיחוד האירופי.

כמה רחוק מעבודה לא מובנית, גמישה, באצוות קטנות: תחנות מובנות תוך 1–3 שנים, הרכבה גמישה +5

מספרי מפעל אמיתיים החלו להופיע, אבל הם מדווחי-ספק, צרים ומתחנות מובנות. שמונת הרובוטים דמויי האדם של AgiBot בקו בדיקת הטאבלטים הסופי של Longcheer רשמו 64,828 משימות ב-99.99% על פני יותר מ-64 שעות (AgiBot, Yicai Global vendor and customer , אבל זו עבודת טיפול שזרוע קונבנציונלית עם ראייה יכולה לעשות. UBTech אומרת שה-Walker S2 שלה עובד ב-30–50% מיעילות אנושית (Irish Times/FT) 11 . single source החודשים של Figure 02 ב-BMW Spartanburg הניבו מספרי נפח (1,250+ שעות, 90,000+ חלקים), בעוד מספרי המחזור, ההצלחה וההתערבות שלו הם יעדים, לא תוצאות (Figure, BMW

Nucleus (Group verified, numerators only . אומרת שכ-40% מווידאו המפעל שלה מופעל מרחוק (Interesting Engineering) vendor claim , וחקירה של רויטרס ציטטה כ-20% מתפוקת עובד עבור כמה רובוטים דמויי אדם סיניים (Notebookcheck) single . אף פריסה לא מפרסמת שיעור הצלחה, זמן מחזור ושיעור התערבות יחד, ואף אחת לא נבדקה, ובכל זאת שלושת אלה הם בדיוק מה שחישוב החזר זקוק לו. source

פריסה	שיעור הצלחה	טקט	התערבויות או הפעלה מרחוק	זמינות או מרווח כשלים	הרצת נפח
Figure 02 @ BMW Spartanburg	יעד בלבד (<99% למשמרת)	יעד בלבד (84 שניות)	יעד בלבד (0 למשמרת)	לא נחשף (נקודת כשל עליונה באמה)	1,250 שעות
Agility Digit @ GXO	לא נחשף	לא נחשף	לא נחשף	לא נחשף	100,000 ארגזים
Apptronik Apollo @ Mercedes-Benz	לא נחשף	לא נחשף	לא נחשף	לא נחשף	לא נחשף
Agile ONE @ Audi	לא נחשף	לא נחשף	לא נחשף	לא נחשף	לא נחשף
Nucleus (מפעל ללא שם)	לא נחשף	לא נחשף	≈40% מופעל מרחוק (שיטה לא נחשפה)	לא נחשף	≈2 שעות וידאו
AgiBot G2 @ Longcheer	99.99% מ-64,828 משימות	18–20 שניות לצעד	לא נחשף	תקלת תקשורת אחת, ≈2 דקות התאוששות	64+ שעות
תא היברידי של Neuromeka	322/324	איטי ב-12.8% מאדם	לא נחשף	לא נחשף	324 מחזורים
Physical Intelligence π0.6 באתר לקוח	—	—	96.4% אוטונומי על פני משמרת	לא נחשף	משמרת אחת

מקורות: פרק זה ופרק 3 mostly vendor claims .

FIG 30 מה פריסות רובוט פנים-מפעליות חושפות בפועל דובן חושפות רק נפח; אף אחת לא חושפת שיעור הצלחה, טקט ושיעור התערבות יחד.



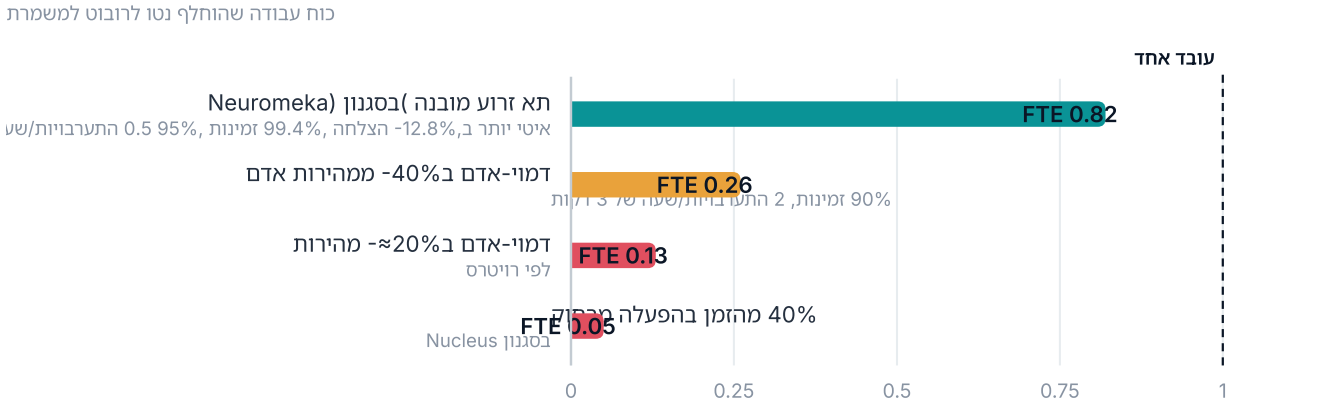
* שיטה לא נחשפה

מקור: חומר ציבורי מ-BMW, Figure, Agility, Mercedes-Benz, Audi, AgiBot, Neuromeka, Physical Intelligence ואחרים.

הפעלה מרחוק חשובה יותר לכלכלה משיעור ההצלחה. אנשים שהוחלפו נטו לרובוט למשמרת ≈ (מחזור אנושי ÷ מחזור רובוט) × זמינות × הצלחה, פחות זמן התערבות והפעלה מרחוק estimate, my formula . תחנה מובנית בסגנון Neuromeka (איטית ב-12.8%, הצלחה 99.4%, זמינות 95%, 0.5 התערבויות של 2 דקות בשעה) מחליפה כ-0.82 אנשים; דמוי-אדם ב-40% ממהירות אנושית, 90% זמינות ושתי התערבויות של 3 דקות בשעה כ-0.26; ב-20% של רויטרס, כ-0.13; ואם, כמו Nucleus, מפעיל אחד מפעיל אותו

מרחוק 40% מהזמן, כ-0.05, כמעט כלום. בכ-RMB 600k לדמוי-אדם ו-RMB 85k לאדם לשנה בסין היבשתית, שלושת האחרונים מחזירים על פני עשורים; רק תחנת הזרוע המובנית היא בעלת החזר ריאלי (פרק 4) estimate .

FIG 31 כוח עבודה שהוחלף נטו לרובוט*תא זרוע מובנה מחליף כ-0.82 עובד; דמוי-אדם המופעל מרחוק 40% מהזמן, כ-0.05.*



מקור: הערכות מבוססות נוסחה של המחבר.

הרחק מתחנות מובנות, מודלים כלליים עדיין קולעים 40–90% במשימות חדשות: Gemini Robotics 2 מגיע ל-74.2% ב-pick-and-place מעבדתי, 78.9% בקיט כלים ו-44% בקשירת שקית (eesel/DeepMind); ב-RoboChallenge Table30, 30- משימות רובוט אמיתיות, π0.5 קולע 43.7% והמודל הפתוח הטוב ביותר 50.3% (RoboChallenge, QbitAI).

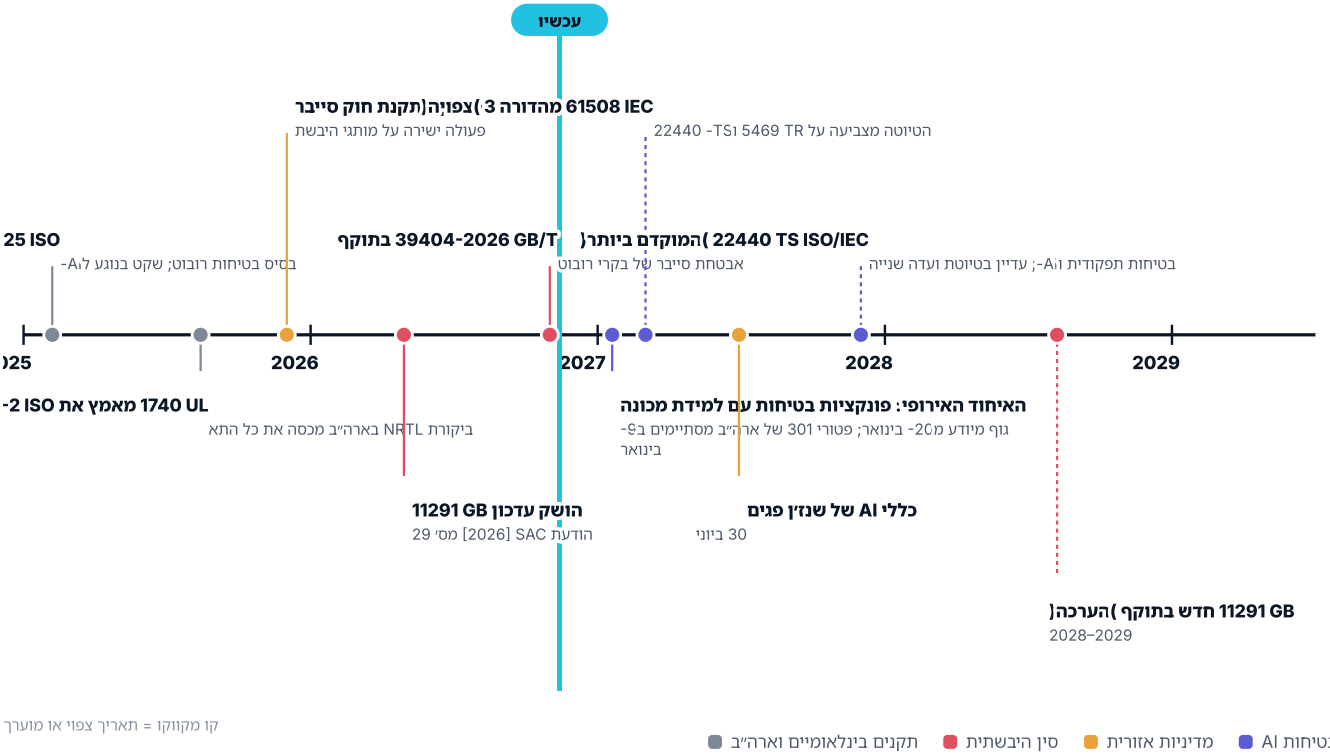
תחזיות חיצוניות שונות, אבל אף אחת לא תומכת בשימוש נרחב בתוך שנתיים. IFR אומר ש-AI פיזי כללי "עדיין לא ברמה תעשייתית" ורוב הרובוטים דמויי האדם "עדיין אבות טיפוס או פיילוטים" (IFR); Interact Analysis מעריכה שרק כ-10% מהרובוטים דמויי האדם שיוצרו ב-2025 נכנסו לשימוש אמיתי, עם נקודת המפנה המסחרית "אחרי 2032" (automation.com); UBS רואה "רגע EV" של דמויי אדם כלא סביר לפני 2030 (Yicai Global); מקורבים סינים מעריכים עוד שנתיים עד שלוש לפני שימוש בקנה מידה מפעלי (Gasgoo) single sources . שיפוטי estimate : במפעלי אלקטרוניקה ורכב מתוקצבים היטב עם הספק באתר, **משימות תחנה מובנית** (טיפול, מיון OK/NG, החדרה מקובעת) יגיעו ל-≥99% בקרוב למהירות אנושית תוך 1–3 שנים; **קיט חלקים מעורבים** עם VLA כלליים תוך 3–5; **רתמות בנתיב חופשי והרכבה גמישה רב-שלבית** ביותר מ-5. ההחזר של דמויי אדם לא עובד היום: Walker S2 בכ-RMB 600k, ב-30–50% מיעילות אנושית על פני שתי משמרות, שווה בערך 0.6–1.0 עובדים, החזר פשוט של 7–12 שנים בעלויות עבודה של סין היבשתית לפני אינטגרציה, תחזוקה והשבתות (10jqka) single source estimate .

איפה זה תקוע

חוליה	ראיות ב-2026
נתונים	התעשייה מעריכה שנדרשות לפחות 10 מיליון שעות של נתונים אמיתיים באיכות גבוהה ורק כ-500,000 נאספו עד תחילת 2026 (Gasgoo); רויטרס מדווחת על כ-500,000 שעות מול כ-100 מיליון נדרשות (Notebookcheck); טלאופרטור מתחיל זקוק לכ-300 ניסיונות לתנועה שמישה; מפעלים לא ישתפו נתוני תהליך; מדיניות שאומנה בסימולציה צונחת בחדות על רובוטים אמיתיים
אמינות	תחנות מובנות 98–99.99% (טענות ספק, כשל לא מוגדר); מודלים כלליים 40–90% במשימות חדשות; מחברים לא נראו 29–67%
מהירות	דמויי אדם ב-30–50% ממהירות אנושית (UBTech), חלקם בכ-20% (רויטרס); סוכני חלוץ כ-280 שניות למשימת רובוט אמיתי מעין-סטטית; לולאות בקרת LLM ב-0.2–0.4 הרץ מול כ-83 הרץ נדרשים (Anthropic)
עלות	Walker S2 בממוצע כ-RMB 600k; החזר של 7–12 שנים estimate
הסמכת בטיחות	ISO 10218-1/-2:2025 הופך בטיחות תפקודית למפורשת יותר וסופג דרישות שיתופיות, אבל לא אומר דבר על AI (The Robot Report); ISO/IEC TS 22440 (בטיחות תפקודית ו-AI) היה עדיין טיוטת ועדה שנייה בנובמבר 2026 (ISS, ISO Open Data); הנחיית IEC 61508-3 מייעצת נגד AI מעל SIL 1 (Analog Devices); תחת תקנת המכונות של האיחוד האירופי, פונקציות בטיחות עם למידת מכונה זקוקות להערכת גוף מיודע מ-20 בינואר 2027 (EUR-Lex)
ידיים וכלים	ידיים מיומנות מנהלות כ-30 N (≈3 ק"ג) לאצבע מול כ-10 ק"ג ליד אנושית (Gasgoo); החלפת כלים עדיין ידנית
אינטגרציה	קריאות חיישנים שגויות, אי-התאמות פרוטוקול וקישורי MES באתר; שיעורי התערבות כמעט אף פעם לא נחשפים

המשמעות הבטיחותית ישירה: התקנים לא אוסרים מדיניות נלמדת ולא מכירים בה. ISO 10218-1/-2:2025 לא אומר דבר על AI; ISO/IEC TS 22440 (בטיחות תפקודית ו-AI) עדיין בטיוטת ועדה ולא סביר לפני סוף 2027 או ISO/IEC TR 2028 ; estimate 5469:2024 הוא הנחיה בלבד, ממליץ על גיבויים שאינם AI, אילוצי פיקוח והצבעה יתירה; IEC 61508 מהדורה 3 צפויה בתחילת 2027 (ISO Open Data, BSI, 61508 Association) verified . אין עדיין תקן שפורסם שמציע מסלול מקובל להסמכת תפיסת למידת מכונה כפונקציית בטיחות. גם כללים אזוריים נעים: סין מעדכנת את GB 11291 (בתוקף כ-2028–2029) (revision plan 20262543); UL 1740 הכניס את ISO 10218-2 לביקורת NRTL האמריקאית מאז אוגוסט 2025; האיחוד האירופי דורש הערכת גוף מיודע של פונקציות בטיחות עם למידת מכונה מ-20 בינואר 2027; ואין דמוי-אדם מוסמך לשימוש מפעלי ללא גדר בעוד ISO 25785-1 עדיין בטיוטה (i-SCOOP). הארכיטקטורה הניתנת להגנה: למידת מכונה מקבלת רק החלטות תהליך ("איזה חלק", "האם הלוח הזה תקין"), בעוד שכבת בטיחות דטרמיניסטית מוסמכת (בדרך כלל PL d / Category 3) ו-PLC בטיחותי מחליטים על מהירות, כוח ועצירות. התייחס לכל טענת "AI מצלמה לבטיחות ללא גדר" כלא מוכחת עד שתעודת PL d או SIL 2 תנקוב בשם רכיב למידת המכונה.

FIG 32 ציר זמן תקנים ורגולציה*מסלול הסמכה ל-AI בתוך פונקציות בטיחות לא יתייצב לפני 2027–2029.*



מקור: ISO, IEC, UL, EUR-Lex, SAC ומאגרי רגולציה לאומיים.

איזו שכבה היא ההזדמנות: המודלים שייכים לענקים, תיבות הקצה ליצרני השבבים, החברות הקטנות לאינטגרציה וחומרת חישה

שכבת המודל זקוקה להון ונתונים ושייכת לענקים (AgiBot אימנה את GO-2 על "עשרות אלפי שעות" (iFeng Tech); המודלים החזקים ביותר סגורים או בענן בלבד). שכבת תיבת הקצה שייכת ליצרני השבבים: Jetson Thor הוא התקן דה-פקטו, ובסין היבשתית לוחות RDK של (D-Robotics) D-Robotics ומעבדי משנה של (CSDN) Rockchip מתחרים, לרוב ללא תוצאות VLA שפורסמו.

לחברות קטנות יש שתי פתיחות. אחת היא חומרת חישה שקונה ישירות אמינות: מגע מרים את הצלחת ההחדרה של ContactDP מ-58.7% ל-97.8%, וחיישני כוח/מומנט ומצלמות פרק כף יד עושים עבודה דומה; ערכות איסוף נתונים נמכרות גם כן (10jqka), אף שידיים, חיישנים וזרועות כולם נופלים במחיר. השנייה היא שכבת האינטגרציה בין יצרני הזרועות לחברות המודלים, בעלות על נתוני המשימה של כל תחנה, מעטפת הבטיחות וקישור ה-MES, שיצרני הזרועות לא מוכרים כמוצר היום. מתחרים נופלים לארבע קבוצות: מובילי זרועות (FANUC, Yaskawa, KUKA, ABB, UR, JAKA, Dobot ואחרים), חברות ראייה-פלוס-זרוע (Mech-Mind, Hikrobot, SmartMore), חברות AI מגולם (AgiBot, UBTech, Unitree, Galbot, Figure, Physical Intelligence, Google DeepMind, NVIDIA) ושחקני שבבי רכב.

קבלת קונה: אין מספר בלי מכנה

מאחר שאף ספק לא מפרסם שיעורי התערבות, מרווחי כשלים או ספירות הצלחה ארוכות טווח, אלה שייכים לתנאי הקבלה של החוזה. סטטיסטית, 3,000 מחזורי ייצור רצופים עם אפס כשלים מוכיחים שיעור כשל של לכל היותר 0.1% בביטחון 95%; 10 כשלים ב-3,000 מוכיחים רק ≥0.57%; הוכחת ≥1%, ≥0.5%, ≥0.1% ו-≥0.01% עם אפס כשלים לוקחת 299, 598, 2,995 ו-29,956 מחזורים estimate, exact binomial . לפי התקן הזה, ה-322/324 של Neuromeka הוא בעל רווח דו-צדדי של 95% של 97.8%–99.9% ואינו מוכיח ≥99%; כ-6 הכשלים של AgiBot ב-64,828 משימות מרמזים על 99.98%–100%, אבל רק אם "כשל" מוגדר כפי שהקונה מגדיר אותו estimate . אלה דרישות הקבלה שהייתי מציב לכל תא זרוע רובוטית או מגולם; הספים שואבים מהיעדים של Figure עצמה ב-BMW וממקרי Neuromeka ו-AgiBot, והם שיפוט, לא תקן estimate :

פריט	דרישה	סף מעבר מוצע
הגדרות	הגדרות חתומות של מחזור, הצלחה, כשל (הצלחה חלקית, אחיזה חוזרת, נפילה, בריחת איכות), התערבות (מגע, הפעלה מרחוק, איפוס, עצירת חירום) ואיך אוטונומיה מחושבת	הגדרות קודמות לכל מדד; אין אחוז בלי נוסחה
התערבויות והפעלה מרחוק	לשעת רובוט, לפי סוג ומשך; זמן ממוצע בין התערבויות; דקות הפעלה מרחוק כנתח מזמן הריצה; מפעילים לרובוט	תחנות מובנות ≥0.5 התערבויות לשעת רובוט (≤2 שעות ביניהן); אפס הפעלה מרחוק במהלך הקבלה
הצלחה ארוכת טווח	≥3,000 מחזורי ייצור רצופים על פני ≥3 משמרות, ≥2 מספרי חלק ותאורה ושונות אספקה רגילות; כשלים מסווגים; גבול תחתון מדויק	גבול תחתון ≥99.5% לתחנות בקו שעוצרות קו, ≥99.0% לא מקוון; בריחות איכות מדווחות בנפרד
זמן מחזור	P50/P90/P99 לחלק כולל התאוששות, מול ≥30 מחזורים אנושיים מתוזמנים באותה תחנה	≥1.15 × P50 אנושי, או בתוך טקט הקו
זמינות ותקלות	שעות ריצה, יומן השבתות, תקלות לפי רכיב, זמן החלפת חלק	≥95% זמינות על פני ≥2 שבועות; MTTR ≤20 דקות עם חלפים באתר
התאוששות	הזרקת תקלות מתוסרטת: נפילות, הסתרה, שינוי תאורה, עצירת מסוע, אובדן MES או רשת, אובדן חשמל, עצירת חירום	≥90% התאוששות אוטונומית, חזרה לייצור תוך 2 דקות
החלפה	שעות, הדגמות, מיומנות מפעיל ועלות איסוף נתונים לכל מספר חלק או מתקן חדש	תחנות מובנות: ≥1 משמרת ו-≤100 הדגמות למספר חלק חדש
בטיחות	זרוע ותא ל-ISO 10218-1/-2:2025, הערכת סיכון תא, כללים מקומיים (GB 11291 בסין היבשתית; CNS 14490 בתוספת הערכה בכתב בטאיוואן; ANSI/A3 R15.06-2025 ו-UL 1740 בארה"ב)	אין פונקציית בטיחות ממומשת בלמידת מכונה; שכבת בטיחות מוסמכת (בדרך כלל PL d / Cat 3) מגבילה מהירות וכוח; אין רובוטים מהלכים ללא גדר עד ש-ISO 25785-1 יפורסם ויוערך
חישוב והתנהגות לא מקוונת	חומרת הסקה, השהיית P99, התנהגות ללא רשת, גרסאות מודל, אימות מחדש אחרי כל עדכון	הסקה קריטית-לטקט אינה תלויה בענן; קבלה מחדש מלאה אחרי כל שינוי מודל, קוונטיזציה או מנוע
ראיות	וידאו לא ערוך ויומנים גולמיים לכל חלון הקבלה; נצפה על ידי צוות האיכות של הלקוח או צד שלישי	אין קטעי שיא או סיכומי ספק בלבד
תנאים מסחריים	אבני תשלום קשורות למדדי קבלה; קנסות זמינות כשחיוב לפי שעת רובוט; כללי בעלות נתונים וגישה מרחוק	תשלום סופי ועיכבון משוחררים רק אחרי קבלת 3,000 המחזורים ו-≤4 שבועות של ריצה יציבה
כלכלה	אנשים שהוחלפו נטו = יחס מהירות × זמינות × הצלחה − זמן התערבות − נתח הפעלה מרחוק	דחה כל הצעה שההחלפה נטו שלה אינה יכולה לתמוך בהחזר הנטען שלה

מקורות: (Figure, Neuromeka, Yicai Global, i-SCOOP); הספים הם הצעות שלי.

שאלות פתוחות

07 מה שהרשומה הציבורית אינה יכולה לענות עליו, ואיך פיילוט מכריע זאת

השאלות שמקורות ציבוריים אינם יכולים לענות עליהן נופלות לחמש קבוצות: מה ביקורת AI באמת נמכרת עבורו ומה קונים ישלמו; איך יצרני AOI מצרפים AI ומשדרגים מכונות ישנות; בני כמה צי ה-AOI המותקנים וכמה כבר מריצים AI; כמה אנשים ושעות ביקורת באמת לוקחת לקו; ואיך פיילוטים מומרים ואיך קונים משלמים. מעל אלה באה השאלה איך המודלים החדשים ביותר, הבניות המקוונטזות וזרועות הרובוטיות באמת מתפקדים תחת התנאים של מפעל נתון. כל אחת ניתנת להפיכה לבדיקה עם סף ברור. הספים למטה הם כללי החלטה שנגזרו מהראיות בדוח זה estimate ; חשב אותם מחדש ברגע שנתוני המפעל הראשונים מגיעים.

שאלה פתוחה	איך להכריע אותה	איך נראית תוצאה טובה
מחיר ונכונות לשלם עבור ביקורת AI	בקשת הצעת מחיר תקנית (למשל ארבעה קווים עם 2D ישן ו-3D חדש מעורבים, שתי משמרות, כ-2,000 לוחות למשמרת, 12 מחלקות פגמים, מקומי בלבד, אינטגרציית MES או IPC-CFX) שנשלחת לכמה ספקי צד שלישי; סקר רגישות מחיר בן ארבע שאלות של 20–30 מפעלי EMS	רוב ההצעות בתוך RMB 40–120k חד-פעמי או 20–40k בשנה לארבעה קווים; לפחות חצי מהמפעלים מקבלים RMB 10–30k לקו עם החזר תוך 12 חודשים
צירוף AI ושדרוגים למכונות ישנות של יצרני AOI	ראיונות מובנים קצרים עם כתריסר יצרני AOI ומפיצים	האם ביקורת AI היא תקנית, אופציונלית או נעדרת במודלי 2025–26; האם מכונות לפני 2020 ניתנות לשדרוג ובאיזה מחיר
גיל צי ה-AOI והנתח שמריץ AI	סקר באתר של כ-30 מפעלי EMS, מרובד לפי גודל ואזור, עם רשימות ציוד שנבדקות תחת NDA	גיל AOI חציוני ≤5 שנים ו-≤60% ללא AI פירושו שמאגר השדרוג לאחור אמיתי
כוח עבודה ושעות ביקורת	ראיונות ו-2–4 שבועות של יומני התראות AOI, ביקורת, תיקון ומשמרות בשני מפעלים או יותר, עם דגימת עבודה	≥1 מבקר ייעודי לקו למשמרת ו-≤3 שעות ביקורת לקו למשמרת, עם לפחות 0.5 FTE לקו שהמפעל מוכן להפנות
ביצועי ביקורת ה-AI עצמה	סט זהב בתוספת מצב צל על קו אמיתי (פרק 3): הסכם נתונים בשבוע 0, קו בסיס ותיוג בשבועות 1–2, ריצת צל בשבועות 2–4, ריצה מסייעת אופציונלית בשבועות 4–8	קריאות שווא או שעות ביקורת נחתכות ב-≤50–60%; גבולות בריחות למחלקה לא גרועים מקו הבסיס האנושי; אפס החטאות במחלקות קריטיות; השהיית p95 ≥½ טקט; פתיחה אוטומטית לאנשים במקרה כשל
איכות תיוג המורה של מודלי חלוץ	הרץ שניים או שלושה מועמדים על אותו סט זהב, שלושה ניסוחי הנחיה כל אחד, תמונת ייחוס טובה אחת, חיתוכי רכיב בודד ≥42–48 פיקסלים	קבלה אוטומטית רק למחלקות עם אפס החטאות על ≥299 פגמים והסכמה ברמה אנושית; אחרת תיוג מראש ומיון בלבד
בניות מקוונטזות	שנה רק את הדיוק; נעל גיבובי משקלים, תבנית, עיבוד מקדים, פענוח, מנוע וגודל אצווה; בטל או אפס מטמון קידומת; בדוק שבניית הייחוס משחזרת תוצאות ספק; דרג זוגות על סט הזהב ופרוסת הסחיפה	אין החטאות חדשות במחלקות קריטיות; ≥1 החטאה חדשה ב-1,000 פגמים; ירידת סיווג ≥2 נקודות; ≤99.5% החלטות זהות על פני שלוש הרצות; אחרת חזור ל-FP8/INT8
חומרה לצד הקו, ניקיון ו-ESD	≥2 יחידות של כל תיבה מועמדת שנבדקות בלחץ בכפול משיעור ההתראות הצפוי; השריה של 72 שעות בזרימת אוויר אמיתית של ארון; בדיקת ESD על ידי רכז הלקוח; בדיקת ISO 14644-14 ללקוחות ISO 7 ונקי יותר	p95 ≥½ טקט; אין קריסות או האטה על פני 72 שעות; אישור ESD; דוח חלקיקים בדרגה הנכונה
תנאי תשלום, כללי פיילוט והמרה	סעיפי תשלום מחוזי ציוד בדיקה אחרונים במפעלים שנסקרו; ראיונות על כללי פיילוט, קריטריוני קבלה וזמן הסמכה	בערך 30/60/10 עם מזומן תוך 60 ימים מקבלה; פיילוטים ≥8–12 שבועות עם בעל אישור נקוב בשם; ≥1 מכל 3 ניסויים מומרים; מסגרת שמכסה ≤5–10 קווים בשנה הראשונה
נתונים, IT ואחריות	ראיונות ובדיקת חוזה עם ייעוץ משפטי	תיבה מקומית רשאית להתחבר ולגשת לנתונים; אחריות מוגבלת לפגמי תוכנה

שאלה פתוחה	איך להכריע אותה	איך נראית תוצאה טובה
תאימות אזורית לפני יציאה לחו״ל	טאיוואן: סטטוס כתוב של הלקוח תחת חוק ניהול אבטחת הסייבר, הערכה עצמית E187; ארה״ב: פסיקת מכס על איך תא מסווג, הצעות NRTL, בדיקת ITAR; דרום-מזרח אסיה: קישור מקומי BOI וזכאות קצבה מלזית, חברת האם של הלקוח; סין היבשתית: היקף CCC למחשבים ארוזים וספקי כוח	הלקוח אינו גוף בפיקוח, וההחזר נשאר ≥2 שנים כולל מכס והסמכה
קבלת תא רובוט או מגולם	טבלת הקבלה בת 12 הפריטים בפרק 6: הגדרות חתומות, ≥3,000 מחזורי ייצור רצופים על פני ≥3 משמרות ו-≥2 מספרי חלק, הזרקת תקלות מתוסרטת, ≥30 מחזורים אנושיים מתוזמנים, נצפה	גבול תחתון ≥99.5% בקו (≤99.0% לא מקוון); מחזור ≥1.15 × אנושי; ≥0.5 התערבויות לשעה ואפס הפעלה מרחוק; ≥95% זמינות; MTTR ≤ 20 דקות; ≥90% התאוששות
מדיניות ACT ו-VLA על שבבי קצה סיניים	ארבעה שבועות: פורט ומיקרו-מדד (≤1,000 הסקות, התפלגות השהיה, הספק, השריה של 60 דקות); סימולציית לולאה סגורה עם השהיה מדודה מוזרקת; A/B על רובוט אמיתי מול GPU; עמידות של 8–24 שעות (≤3,000 מחזורים)	מקטעי ACT p95 ≤100 אלפיות שנייה, VLA ≤300 אלפיות שנייה; הצלחה על רובוט אמיתי בטווח 3 נקודות מה-GPU ומחזור ≥1.2 × אנושי; אחרת השאר את המדיניות על Jetson או GPU
מחירי קובוט, AMR ואינטגרציה	הצעות מודל-אחר-מודל מכמה מפיצי קובוטים, והצעות תא מלא משניים או שלושה אינטגרטורים לטעינת מתקן בדיקה (תחנה אחת, שתי משמרות)	תא מלא ≥RMB 200k והחזר ≥1.5 שנים על פני שתי משמרות בשנ׳ן
ספירות קווי SMT בטאיוואן, דרום-מזרח אסיה וארה״ב	נתוני IPC (לשעבר New Venture Research) או בסיס מותקן של מפיצי ציוד מקומיים	מספיק קווים ניתנים לשירות לאזור כדי להצדיק הנדסת יישום מקומית

מתוך כל אלה, ארבעה מחליטים את המספרים הרגישים ביותר בדוח זה: המחיר שקונים ישלמו, כוח העבודה שביקורת AI באמת משחררת, האם ביקורת AI באמת מחמיצה לא יותר מאנשים, וכמה מהר המזומן חוזר. שווה להכריע אותם לפני כל התחייבות משמעותית, על ידי מפעל ששוקל רכישה או על ידי ספק ששוקל מוצר.

רצף פיילוט שעובד לשני הצדדים

עבור מפעל, ועבור ספק שרוצה אסמכתא אמינה, הסדר זהה:

1. **שבועות 0–2: הסכימו על נתונים והגדרות.** חתמו על הסכם הנתונים, הגדירו את מחלקות הפגמים ומה נחשב בריחה, ורשמו קו בסיס: קריאות שווא ללוח, שניות ביקורת לקריאה, בריחות שנמצאו במורד הזרם.
2. **שבועות 1–4: בנו את סט הזהב** מארכיון ה-AOI ולוחות NG שנשמרו, מתויג בעיוורון ומוכרע (פרק 3).
3. **שבועות 2–4 (או יותר): מצב צל.** המודל רושם את החלטותיו בעוד אנשים מחליטים כבעבר; השוו לפי מחלקה.
4. **שבועות 4–8: ריצה מסייעת במשמרות אקראיות,** כשהמודל מסווג רק מחלקות שעברו את השער וכל השאר הולך לאנשים.
5. **קבלה, בכתב, לפני הרחבה:** תוצאות למחלקה, השהיה, התנהגות פתוח-בכשל, ומי הבעלים של המודלים והתיוגים. הרחב קו אחר קו, ואמת מחדש על כל מוצר חדש, גרסת מודל, קוונטיזציה או שינוי מנוע.

מגבלות דוח זה

זהו מחקר שולחני. לא ערכתי ראיונות ראשוניים עם מפעלים למהדורה זו; חלק ניכר מנתוני הקו מגיע מספקים ולא נבדק בלתי תלוי. כמה אתרי ממשלה, מאגרי מכרזים ופלטפורמות גיוס לא היו נגישים, כך ששיעורי סבסוד השדרוג של שנ׳ן 2026, הנחיית השכר של דלתת נהר היאנגצה, מחירי עסקה לביקורת AI, מבקרים לקו ושיעורי המרת פיילוט הם הערכות או אנלוגיות; כל אחד מופיע לעיל כשאלה פתוחה. שערי חליפין הם השער המרכזי של בנק העם הסיני מ-9 בנובמבר 2026 ($1 = RMB 6.7188) ו, עבור טאיוואן, שער הקביעה של SMBC באותו יום ($1 = NT$32.12); הדונג הוויטנאמי משתמש בנתוני מאי 2026. כל הספים הם כללי החלטה שנגזרו מהראיות המצוטטות, לא תקנים.

אודות המחבר

Tan Shuai

Tan Shuai הוא טכנולוג ומייסד עם יותר מעשרים שנה של עבודה על פני שבבים, חומרה ותוכנת מערכת, ו-20 פטנטים רשומים. עבודתו לקחה אותו דרך עמק הסיליקון, ניו יורק, שיקגו, טאיוואן, אזור המפרץ הגדול (שנז'ן והונג קונג), וייטנאם ומלזיה, עם התמקדות ארוכה בייצור מתקדם ושרשראות אספקה. כיום הוא עובד כחוקר ומפתח AI, ומתרכז באופן שבו AI משמש בפועל על רצפת המפעל.

הוא מזמין שיחות עם אנשים העובדים על AI בייצור, בין אם לבניית צוות, חיפוש שותף או השוואת רשמים: i@tanshuai.com · tanshuai.com

הערות וכתב ויתור

דוח זה הוא מחקר שולחני בלתי תלוי המבוסס על מסמכים ציבוריים שהיו זמינים עד 11 בנובמבר 2026. הוא מסופק למידע בלבד ואינו ייעוץ השקעות, משפטי, מס או הנדסי. שמות מוצרים, מחירים, השקות מודלים, סיווגי יצוא, מכסים ותקנות משתנים במהירות; אמת כל דבר שאתה מסתמך עליו מול המקור הראשוני. נתונים שמסומנים vendor claim או estimate צריכים להיקרא ככאלה. שמות חברות ומוצרים שייכים לבעליהם ומשמשים לזיהוי בלבד; אף חברה לא סקרה או מימנה דוח זה.

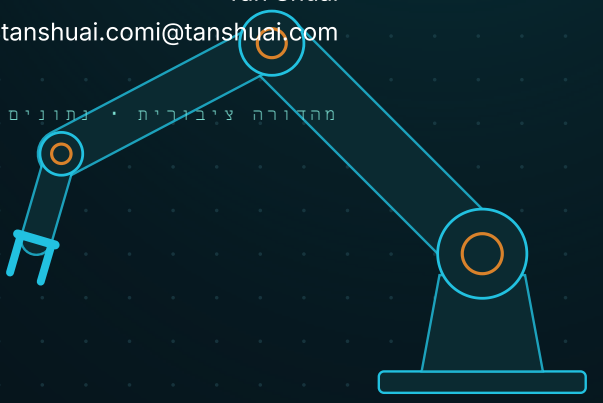
© 2026 Tan Shuai. ניתן לשתף דוח זה ללא שינוי עם ייחוס.

יצור אלקטרוניקה · בדיקת AI × רובוטיקה

בדיקה חזותית מבוססת AI וזרועות רובוטיות במפעלי אלקטרוניקה: הערכת 2026

כל מספר מרכזי נושא את המקור שלו ותג ביטחון: verified, single source, vendor claim או estimate. מה שהרשומה הציבורית אינה יכולה לענות עליו מרוכז בפרק 7, יחד עם איך פיילוט מכריע כל שאלה.

Tan Shuai
tanshuai.comi@tanshuai.com



מהדורה ציבורית · נתונים נכון ל-11 בנובמבר 2026

