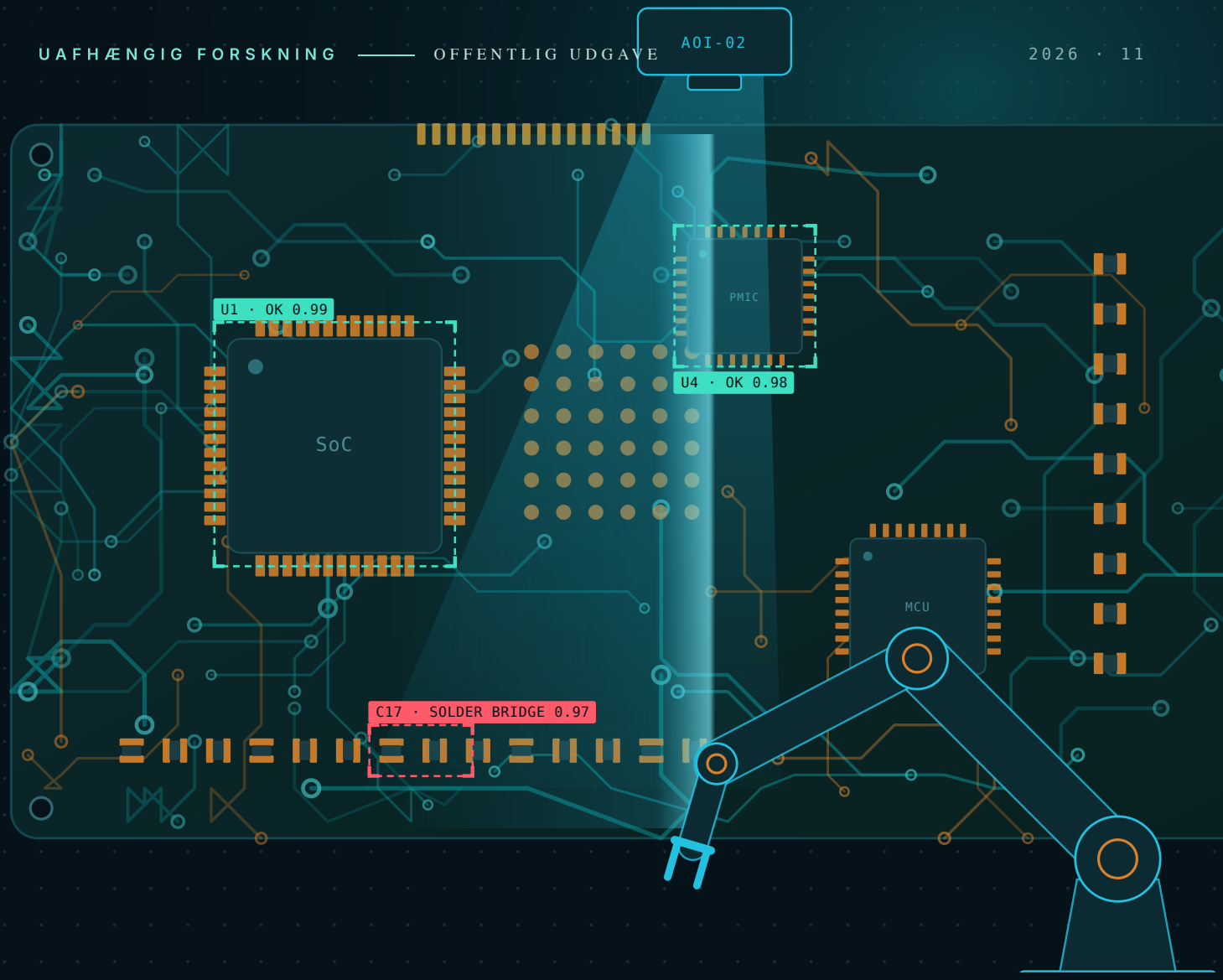


ELEKTRONIKFREMSTILLING · AI-INSPEKTION × ROBOTIK

AI-synsinspektion og robotarme i elektronikfabrikker

Vurdering 2026

Gængse konfigurationer · Regional gennemførlighed · On-premise-implementering · Lærer-elev-modeller
Investeringsafkast · Konkurrence · Hvad kun et pilotprojekt kan besvare

NØGLEKONKLUSION

AI-gennemgang er værd at gøre, snævert. Embodied robotarme kan vente.

Tan Shuai
tanshuai.comi@tanshuai.com

Data pr. 11. november 2026
Fastlands-Kina · Taiwan · Sydøstasien · USA

INDHOLD

Indhold



◆ **RESUMÉ** — Værd at gøre, men snævert, nu og med partnere — 5

01 **Linjen i dag** — Hvad elektronikfabrikker kører i dag, og hvad der er teknisk gennemførligt — 9

02 **On-premises** — At køre AI inde i fabrikken: renrum, luftgab, hardware og kvantisering — 17

03 **Lærer og elev** — Lærer og elev: frontmodeller kan præ-annotere inspektionsdata; for robotarme er de virkelige lærere stadig mennesker — 26

04 **Omkostninger og tilbagebetaling** — Omkostninger og tilbagebetaling: AI-gennemgang tjener sig hjem på måneder, maskinbetjening på år, udplukningsarme aldrig — 35

05 **Marked og konkurrence** — Marked, konkurrence og hvordan købere betaler — 43

06 **Embodied AI** — Embodied AI: først ændrer den, hvordan robotter læres, senere hvordan de bevæger sig — 52

07 **Åbne spørgsmål** — Hvad de offentlige kilder ikke kan besvare, og hvordan et pilotprojekt afgør det — 58

◆ **Om forfatteren** — Tan Shuai — 61

Konfidensgradsmærker i teksten

- **verified** — to eller flere uafhængige kilder, eller to uafhængige primærdokumenter, er enige
- **single source** — kun én kilde, herunder primærdokumenter
- **vendor claim** — offentliggjort af sælgeren af produktet eller tjenesten
- **estimate** — forfatterens egen beregning eller udledning af de citerede data

Sådan læser du denne rapport

Nøgletal bærer et konfidensgradsmærke. **verified** betyder, at to eller flere uafhængige kilder er enige, eller at to uafhængige primærdokumenter er enige; **single source** betyder kun én kilde (herunder primærdokumenter); **vendor claim** betyder offentliggjort af sælgeren af produktet eller tjenesten, selv hvis vidt gentaget; **estimate** betyder min egen beregning eller udledning af de citerede data, med inputs kildehenvist. "Primær" beskriver kildetypen (årsrapport, prospekt, forordning, standard, officiel statistik); det ændrer ikke konfidensklassen. Hvor kilder modsiger hinanden, angives tallene side om side og gennemsnittes ikke. Valutaomregninger bruger People's Bank of Chinas centralparitet fra 9. november 2026, $1 = RMB 6.7188 (Xinhua), og for Taiwan SMBC's fixing samme dag, $1 = NT$32.12 (SMBC). Løn- og subsidiedata for fastlandet bruger Pearl River Delta, hvor elektronikfremstilling er mest koncentreret (officielle løndata fra Shenzhen), med Yangtze River Delta til sammenligning. Påstande fra brancheinterviews behandles som påstande, der skal kontrolleres, aldrig som kendsgerninger. Spørgsmål, som de offentlige kilder ikke kan besvare, er samlet i kapitel 7.

FIGURER

Figurer



01	Hvor AI passer på fabriksgulvet i dag	7
02	En volumen-SMT-linje: menneskene sidder bag AOI'en	10
03	Omkring 97–99% af AOI-alarmerne er falskudkald	11
04	Et generelt kamera er ingen AOI	13
05	En air-gapped udrulning	17
06	Ren- og ESD-zoner afgør, hvor beregning kan stå	18
07	Hastighedsgrænser pr. anvendelse	20
08	On-prem beregning: hukommelse, pris og hvad der lovligt kan købes i fastlands-Kina	21
09	Én stor arbejdsstation giver ingen delt assistent	22
10	Kvantisering: 8-bit er sikkert, 4-bit afhænger af opskriften, 3-bit og lavere ryger ud	23
11	Lærerkandidater: generiske scores fordobledes på en generation	27
12	«Er der en defekt?» er stadig svært for generelle modeller	28
13	To lærer–elev-pipelines	29
14	Hvor stort det gyldne sæt skal være	30
15	Edge-politiklatens	32
16	Robotelever: tæt på produktionsgrad kun når stationen er struktureret	33
17	Hvad én inspektions- eller håndteringsarbejder koster om året	36
18	Prisstigen	37
19	Tilbagebetaling pr. region	39
20	Hvorfor udplukning af dårlige kort ikke betaler sig	40
21	USA-told på robotter, AOI og edge-computere	41
22	Markedsstørrelser adskiller sig mere end tidobbelt	44
23	Bottom-up: 40.000 linjer → 16.000 servicebare → ≈RMB 320M engangs	45
24	Retrofitvinduet er cirka fem år	46
25	Gennemgangskilen er omringet på tre sider	46
26	Marginen sidder i produktificeret hardware	48
27	Hvordan købere betaler	49
28	Days sales outstanding	50
29	Embodied-AI-stacken	52
30	Hvad robotudrulninger i fabrik faktisk offentliggør	53
31	Netto erstattet arbejde pr. robot	54
32	Tidslinje for standarder og regulering	56

RESUMÉ

Værd at gøre, men snævert, nu og med partnere

NØGLEKONKLUSION

Dommen: For elektronikfabrikker er det, der er modent i 2026, og som fabrikker vil betale for, AI-gennemgang og nicheinspektion på AOI-billeder plus programmeret robotmaskinbetjening, ikke "en stor model, der kigger på kort, plus embodied robotter, der plukker dem".

0.46–0.65 år
Tilbagebetaling af AI-gennemgang i USA, Taiwan og Shenzhen, ved RMB 50k pr. linje

≥95% · <20 PPM
Dækning og escapes i en top-EMS' interne AI-gennemgang: barren for at sælge til tier-1 EMS

100–700 mio. RMB
Adresserbart tredjeparts-retrofitmarked i fastlands-Kina, engangs, over et vindue på omkring fem år

0 point
Publicerede industrielle defekt-benchmarkscores for flagskibene fra november 2026: lærere, ikke inspektører

1.1–1.5 år
Tilbagebetaling af en cobot-maskinbetjeningscelle over to skift i USA, Taiwan og Shenzhen

0.003–0.14 FTE
Arbejdskraft en udpluknings-robotarm sparer pr. skift på en SMT-linje: den tjener sig aldrig hjem

Inspektion
ROI

Superviseret deep learning anvendt på de billeder, en eksisterende AOI allerede producerer, skærer falskudkald eller gennemgangstimer ned med 60–90% (vendor claims). DBG, en stor EMS-virksomhed, oplyser i sin årsrapport, at dens interne system dækker mere end 95% af de flagrede komponenter med escapes under 20 PPM og mere end 85% mindre menneskelig gennemgang; det er den bar, et tredjepartsprodukt skal klare for at sælge til tier-1 EMS. Ved RMB 50,000 pr. linje tjener AI-gennemgang sig hjem inden for et år i fastlands-Kina, Taiwan og USA (Shenzhen 0.65 år, Taiwan 0.53, USA 0.46), cobot-maskinbetjeningsceller på 1.1–1.5 år og Sydøstasien på 2–6 år, mens en robotarm, der føjes til en SMT-linje kun for at plukke dårlige kort ud, aldrig tjener sig hjem nogetsteds, fordi udplukning ved over 90% first-pass yield kun tager 0.003–0.14 af en person pr. skift estimate .

Modeller
ROBOTTER

Multimodale store modeller kan stadig ikke være den primære inspektør. Pr. 11. november 2026 har ingen af flagskibene (Claude Fable 5.2, GPT-6.x, Gemini 4, Qwen3.8 og andre) en publiceret score på nogen industri- eller PCB-defektbenchmark, og i ACM MM 2026 PCBA-udfordringen scorede de fem bedste systemer, trænet på 6,600 ægte linjebilleder, kun 77–82 på "er der en defekt". Frontmodeller kan præ-annotere og triagere. For robotarme er de virkelige lærere stadig menneskelige demonstrationer, reinforcement learning på ægte robotter og simulering: elever når 90–99.4% ved fikserede, strukturerede stationer, men kun 30–80% på bevægelige transportbånd, usete dele og nye produkter; en frontagent gennemførte 104 af 108 kvasistatiske ægte robotopgaver, med 280 sekunder hver.

On-premises
IMPLEMENTERING

Luftgappet implementering i rene, kontrollerede fabrikker er fuldt praktisk. En almindelig SMT-hal er som regel et ESD-beskyttet område snarere end et klassificeret renrum: en blæserløs Jetson- eller indenlandsk NPU-boks står ved linjen, og GPU-servere står i serverrummet. Ifølge NVIDIAs egne eksportklassifikationer kan Jetson Thor og DGX Spark sælges til en almindelig fastlandsfabrik uden en USA-licens; DGX Station kan ikke. Præcision er som standard 8-bit; leverandørens kvantiseringsbevidste 4-bit er brugbart med forbehold; 3-bit og derunder er ude.

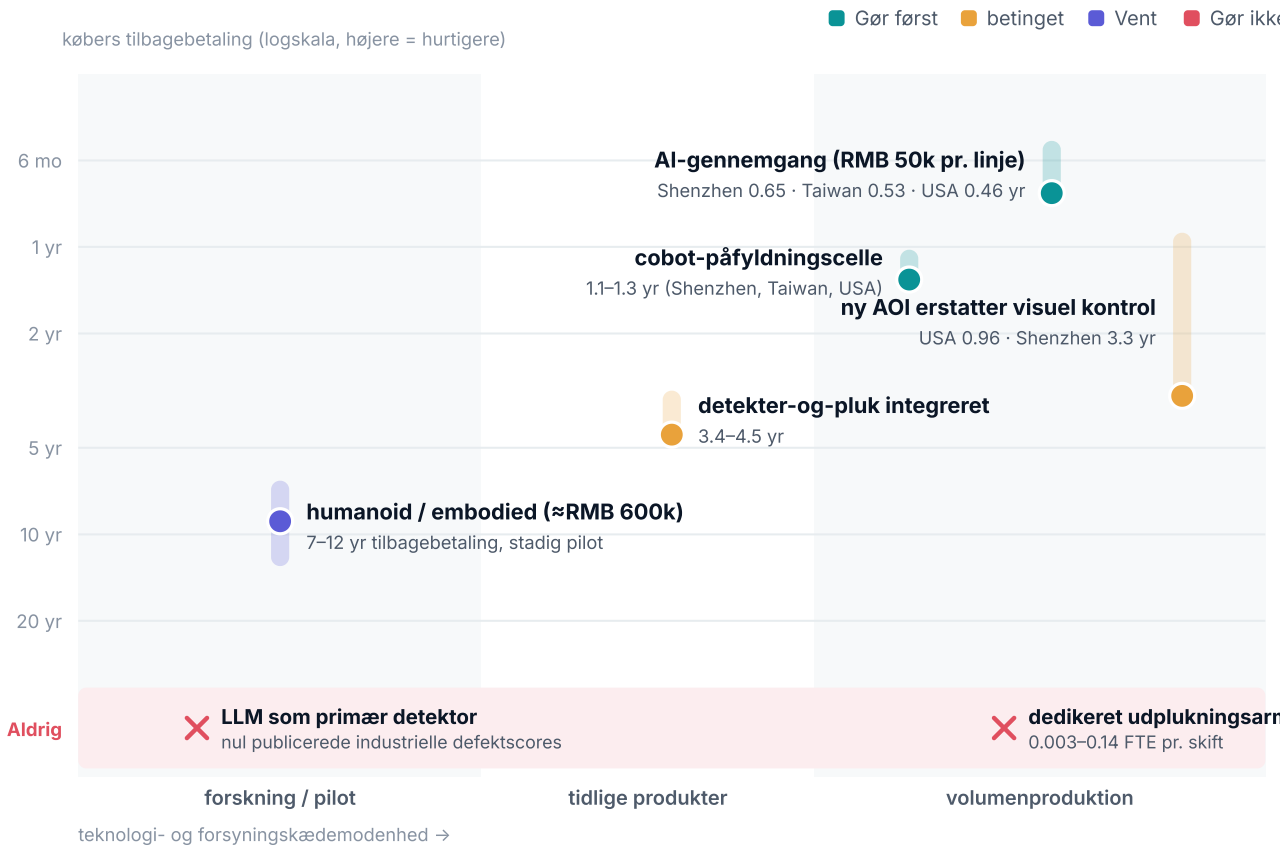
Marked
KONKURRENCE

Markedet er reelt men begrænset. Fastlands-Kina har omkring 35,000–46,000 SMT-linjer og omkring 26,000–66,000 inline-AOI'er, der ikke kører AI; det adresserbare marked for tredjeparts-retrofit er omkring RMB 100–700 mio. engangs, over et vindue på omkring fem år, og AOI-producenter er begyndt at bygge gennemgang ind i nye maskiner estimate. Konkurrencen er allerede stærk: gennemgangskilen er omringet af AOI-producenter, interne EMS-teams og platformleverandører, og marginen sidder i produktiseret hardware.

Spørgsmål	Vurdering	Nøgletal (kilder og konfidensgrad i kapitlerne)
1 **Hvad kører gængse elektronikfabrikker i dag?**	En SMT-volumenlinje er "3D SPI + AOI'er før og efter reflow + stikprøvebaseret røntgen + ICT/FCT + reparation og synsinspektion", med menneskelig gennemgang bag hver AOI. AI kommer ind i gennemgangen på tre måder: indbygget af AOI-producenter, bygget internt af top-EMS eller tilføjet af tredjeparter. Montage og håndtering bruger arme, SCARA'er og AMR'er med indlært perception og programmeret bevægelse; embodied AI er stadig i pilot	Takt ≈13–22 s pr. kort; én ODM budgetterede RMB 30.6M til AOI til 18 linjer, ≈3–5 AOI'er pr. linje; ≈97–99% af AOI-alarmerne er falskudkald
2 **Hvor langt rækker teknologien, kan penge købe nul mennesker, og hvad kan hver region bruge?**	Inspektion er industriel kvalitet som "dedikeret detektor + AI-gennemgang + nogle få mennesker"; selv ubegrænsede penge når ikke nul mennesker. Fastlandet bruger kun indenlandske modeller eller lokale åbne vægte; private fabrikker i Taiwan kan bruge næsten alt, men fabs og kritisk infrastruktur udelukker i praksis fastlandsmærker; Sydøstasien har færrest begrænsninger; USA har hele stakken plus told og certificering	AI-gennemgang skærer falskudkald eller gennemgangstimer ned med 60–90% vendor claims ; et 4K-kamera over et helt panel ≈78 µm pr. pixel mod 8–20 µm for en AOI; fastlands-fremstillede arme betaler ≈40% samlet USA-told, instrumenter af AOI-typen ≈37.5%
3 **Kan det køre on-premises i renrum og luftgappede fabrikker, på hvilken hardware, hurtigt og præcist nok?**	Ja; begrænsningerne er proceduremæssige, ikke beregningsmæssige. En blæserløs boks ≤100 W ved linjen, GPU-servere i serverrummet; Thor til stationer, et Mac Studio eller en indenlandsk enhed til fabriks-Q&A ingen DGX Station på fastlandet. Hastigheden er fin, når modelstørrelsen matcher opgaven; behold 8-bit, brug leverandørens kvantiseringsbevidste 4-bit med forbehold, aldrig 3-bit	≈0.5 s pr. udsnit for en 8B visionsmodel på Thor estimate ; Thor-devkit RMB 59,104 i Kina; DGX Station $123–140k, 139 tokens/s i alt til 32 brugere på en 743B-model
4 **Kan det nyeste flagskib virke som lærer over snesevis af runder og destilleres til lokale elever, også til fleksibel fremstilling og robotarme?**	Inspektion: præ-annotering og triage; auto-acceptér lærerannoteringer kun for klasser med "≥299 defekter, nul misser" på et gyldent sæt. Robotter: lærerne er stadig menneskelige demonstrationer, RL på ægte robotter og simulering, og elever nærmer sig produktionsgrad kun ved strukturerede stationer	Intet november-flagskib har en industriel defektscore; elever 97.8% på trænede stik, 29–67% på usete, 37–59% på bevægelige transportbånd; frontmodeller, der styrer led direkte, 0–5.5%
5 **Hvad koster det, og hvad er ROI pr. region?**	AI-gennemgang tjener sig hjem inden for et år i fastlands-Kina, Taiwan og USA; cobot-betjening 1.1–1.5 år; Sydøstasien 2–6 år; en udplukningsarm på en SMT-linje tjener sig aldrig hjem	RMB 50k gennemgangsboks pr. linje: Shenzhen 0.65 år, Taiwan 0.53, USA 0.46, Vietnam 2.2 estimate ; acceptabel tredjeparts-tillægspris ≈RMB 10–30k pr. linje engangs eller RMB 5–10k om året

Spørgsmål	Vurdering	Nøgletal (kilder og konfidensgrad i kapitlerne)
⑥ **Er konkurrenterne allerede stærke nok, og er det bedre at starte nu eller vente?**	Efterspørgslen er ikke mættet, men udbuddet er overfyldt; OEM-AI begrænser retrofitvinduet til omkring fem år. Gå ind nu med et snævert produkt gennem kanalpartnere, uden at angribe platformleverandører frontalt; vent på 2027–2028 produktionsdata, før du satser på embodied AI	≈26–66k inline-AOI'er uden AI på fastlandet, skrumpende 8–12% om året **estimate** ; den førende AI-vision-leverandør havde i 2025 en omsætning på RMB 1.086 mia. og et nettotab på RMB 991 mio.; den nærmeste børsnoterede sammenlignings 3C-omsætning faldt 79.9% år-til-år i januar–maj 2026

FIG 01 **Hvor AI passer på fabriksgulvet i dag** ***AI-gennemgang af AOI-falskudkald og cobot-påfyldning er modne og betaler sig hurtigt; en dedikeret udplukningsarm og en LLM som primær inspektør er ikke besværet værd.***



Kilde: Tilbagebetaling fra forfatterens model (kapitel 4); modenhedsaksen er forfatterens vurdering.

Hvad man skal bygge, og hvad man ikke skal bygge

	Anbefaling
Byg	AI-gennemgang og defektklassifikation på AOI-billeder, til mellemstore EMS med ældre eller blandede AOI-mærker og ingen AI-opgradering fra AOI-producenten; inspektion, som SMT-AOI'en dækker dårligt (gennemhuls- og håndlodning, konform belægning, kabler, stik og FPC'er, slutudseende og skruer); cobot-ilægnings- og udtagningsceller, hvor hver del håndteres (testfiksturer, bakker, pakning), med inspektion som tillæg; et gyldent sæt og skyggetilstand på en ægte linje som grundlag for enhver påstand
Byg ikke	En generel "AI plus robotarm"-platform; en multimodal LLM som primær inspektør; en robotarm på en SMT-linje kun til at plukke NG-kort ud (en NG-afleder eller -buffer klarer det); humanoider og robots-as-a-service; frontmodeller inde i en styreløkke eller en sikkerhedsfunktion; bunker af GPU-hukommelse ved stationer for at køre "fuldskala"-modeller; direkte salg til fabs, taiwansk kritisk infrastruktur eller USA's forsvarselektronikfabrikker

	Anbefaling
Hvem man skal samarbejde med	AOI-producenter og distributører (kanal- og billedgrænseflader; distributører bragte 46.68% af Sinic-Teks omsætning i 2025); cobot-producenter (certificerede arme og sikkerhedsfunktioner); EMS-virksomheder med 10–40 linjer (rullende udvidelse og forudbetalinger); lokale integratorer i Thailand og andre steder (for at opfylde BOI's local-linkage-regler)

Linjen i dag

01 Hvad elektronikfabrikker kører i dag, og hvad der er teknisk gennemførligt

Standardvolumenlinjen: flere AOI'er, stikprøvebaseret røntgen, central gennemgang, 13–22 sekunder pr. kort

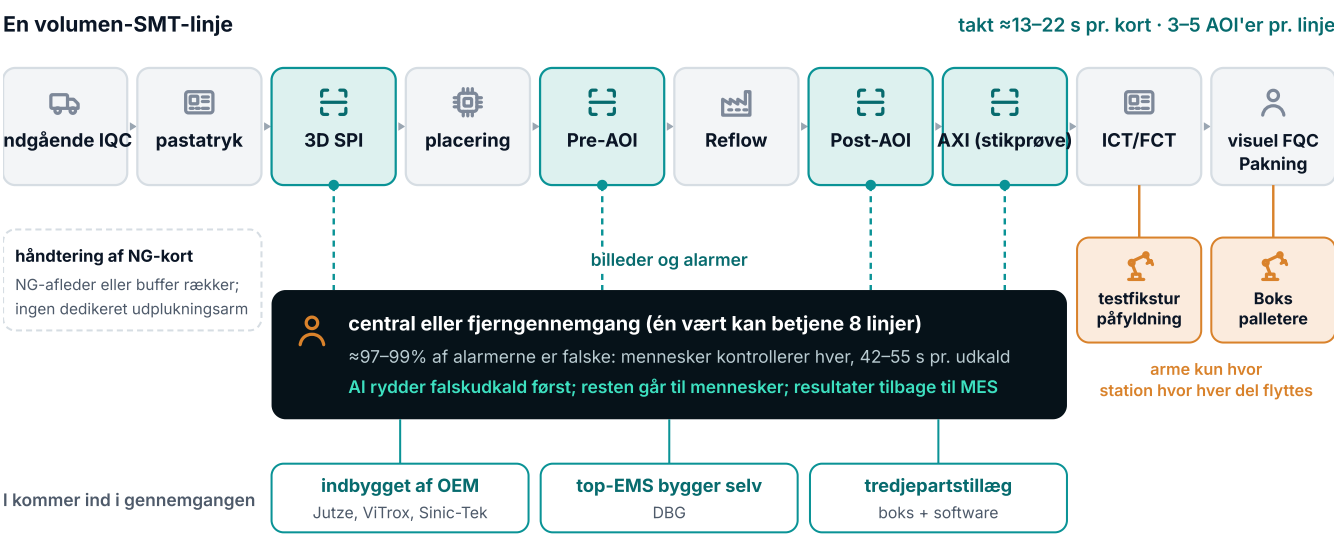
Primære indberetninger fra børsnoterede ODM'er viser, hvordan en volumenlinje i 2025–2026 faktisk ser ud. Longcheers Hongkong-prospekt (data til 30. september 2025) opregner virksomhedens SMT-kvalitetskontrol som "røntgeninspektion, loddepastainspektion (SPI), automatisk optisk inspektion (AOI), PCBA-funktionstest og synsinspektion"; procesflowet kører AOI før reflow, AOI efter reflow, stikprøvebaseret røntgen, NG-kort til en SMT-reparationsstation og derefter et afsluttende synstjek. Dets fabrikker i Huizhou, Nanchang og Vietnam har 34, 24 og 8 SMT-linjer og 67, 60 og 13 slutmontagelinjer (Longcheer H-share-prospekt) single source, primær indberetning. Huaqins 2023-laptopfabrik planlagde 18 SMT-linjer med et AOI-budget på RMB 30.6 mio., omkring RMB 1.7 mio. pr. linje, hvilket ved indenlandske 3D-AOI-priser er tre til fem AOI'er pr. linje til inspektion før og efter reflow af dobbeltsidede kort (Huaqin A-share-prospekt) primær indberetning + estimate. En volumenlinje har derfor normalt to til fire gennemgangspunkter, ikke ét.

Maskinerne er allerede højautomatiserede: Huaqin oplyser 96% automatisering på sine SMT-linjer (Huaqin H-share-prospekt) single source. De resterende mennesker sidder i gennemgang, reparation, synsinspektion, test og slutmontage, og ingen virksomhed offentliggør, hvor mange gennemgangsfolk den har pr. linje (kapitel 7).

Tre uafhængige estimater er enige om en takt på **13–22 sekunder pr. kort** (en miljøindberetning fra 2026 for fem Hebei-linjer (Hebei SMT-projekt EIA), Huaqins laptplinjer og et iFactory-casestudie), eller 7–11 sekunder pr. side for dobbeltsidede kort estimate. Det svære ved en high-mix-linje er ikke volumen, men introduktion af nye produkter og omstilling, og hver omstilling er endnu en koldstart for en AI-model.

På softwaresiden er AOI-producenternes egne gennemgangsstationer allerede centraliserede: Sinic-Teks reparationsstationssoftware lader "én host styre udstyret på otte SMT-linjer" (Sinic-Tek-prospekt) single source, primær, 2022, og efter at Jabil implementerede ViTrox V-ONE fjern-AOI-styring i 2026, kunne én ingeniør styre 5–8 AOI'er eksternt mod 1–2 lokalt (AtlasPCB) vendor claim. Et tredjeparts AI-gennemgangsprodukt kobler sig på præcis denne centrale station gennem AOI'ens billed- og resultateksport (IPC-CFX, IPC-HERMES-9852 eller en producents eget format). AOI'erne selv tilføjer AI hurtigt: Yamaha siger, at dets AI-vurdering tager omkring ét sekund pr. kald, hjælper inspektører efter 30 NG-billeder og bliver "meget høj" kvalitet efter 1,000 (DirectIndustry/Yamaha) vendor claim.

FIG 02 **En volumen-SMT-linje: menneskene sidder bag AOI'en** ***Arbejdet er koncentreret på gennemgangspladserne efter automatisk optisk inspektion. Det er der, AI kommer ind.***



Kilde: Offentlige årsregnskaber og leverandørmateriale (Huaqin, DBG, Sinic-Tek med flere); sammensat af forfatteren.

Trin	Gængse produkter og teknologi	Repræsentative leverandører	Hvor langt AI er nået
Efter tryk	3D SPI (struktureret lys, moiré, noget lasertriangulering)	Koh Young, Sinic-Tek, TRI	OEM AI-pastavurdering i produktion; Jutze siger, at nogle modeller er "helt AI"
Efter placering og reflow	2D/3D-AOI med flervinklede, flerfarvede LED'er og struktureret lys, 8–20 µm pr. pixel, 8–12 MP-kameraer	Koh Young, Omron, Saki, TRI, ViTrox, Mirtec, Jutze, Sinic-Tek og andre	OEM AI-gennemgang (Jutze i produktion, ViTrox V510Ai, Sinic-Tek med Tencent); top-EMS bygger selv (DBG); tredjeparts-tillæg
Bundterminerede og skjulte lodninger	Stikprøvebaseret eller inline røntgen (AXI)	—	Hikvision hævder, at dets Guanlan-model fejlvurderer røntgenhulrum i under 1% af tilfældene vendor claim
Gennemgang og reparation	Central gennemgangsstation (én host, otte linjer), fjerngennemgang, reparationsstation	AOI-producenters software, ViTrox V-ONE	AI lukker falskudkald, folk tager resten
Synsinspektion, IQC, FQC	For det meste manuel; nogle AI-udseendemaskiner med robotpåfyldning	SmartMore og andre	Udseendemaskiner i produktion; ingen offentlige bemandingstal for IQC/FQC
Håndtering og maskinbetjening	Seksaksede arme og cobotter, SCARA, AMR'er, 2D/3D visionsstyring	FANUC, ABB, Yaskawa, KUKA, UR, Dobot, JAKA, Elite; Mech-Mind, Hikrobot	Indlært perception plus programmeret bevægelse i produktion; VLA'er og humanoider stadig i pilot
Data og systemer	MES, IPC-CFX, IPC-HERMES-9852	—	Resultater skrives tilbage til MES; understøttelse i ældre udstyr er ujævn

Kilder til tabellen: indberetningerne ovenfor og (Koh Young, Jutze årsrapport 2025, Sinic-Tek årsrapport 2025, ViTrox, Sohu om Hengtianyi, Xinhua om Hikvision).

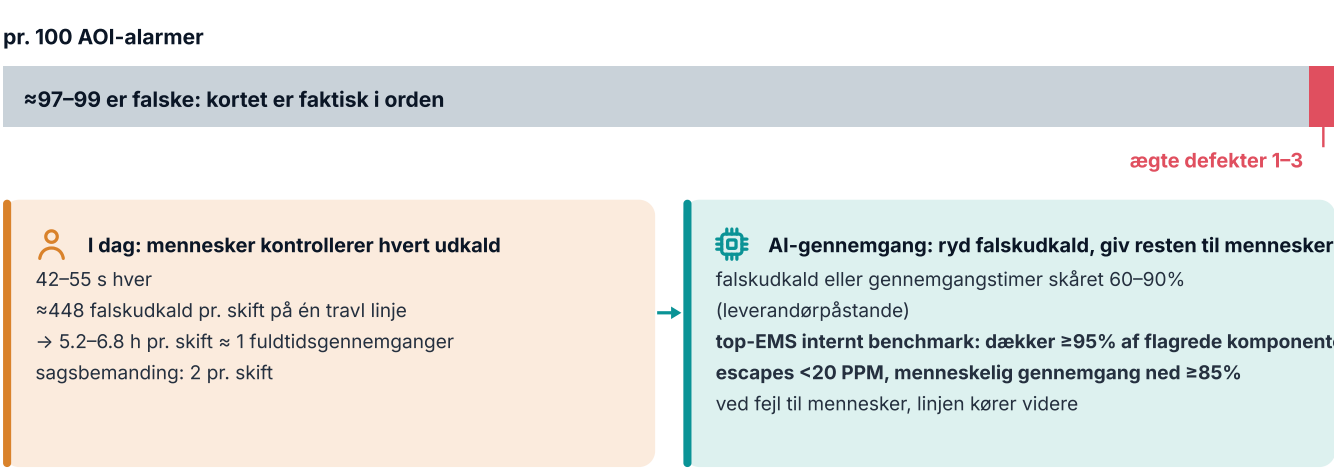
Folkene sidder bag AOI'en, fordi de fleste alarmer er falske

Inspektionskæden på en gængs SMT/PCBA-linje kører: 3D SPI efter tryk for manglende, utilstrækkelig eller overskydende pasta, brodannelse og offset; AOI efter placering (valgfrit) og efter reflow for manglende, forskudte, rote-

rede, vendte eller tombstonede dele og loddemængde; røntgen, når der er BGA'er eller andre bundterminerede dele; derefter in-circuit- eller flying-probe-test, funktionstest og manuel synsinspektion. Hver AOI har en gennemgangsplads og en reparationsstation bag sig (Sinic-Tek-prospekt, JLCPCB) verified. AOI efter reflow er den mest almindelige og bedste værdi, men den ser kun form og areal; lodninger, hulrum og befugtning under en BGA kræver røntgen eller elektrisk test (JLCPCB).

Arbejdskraften sidder i gennemgang, fordi de fleste alarmer er falske. En britisk EMS registrerede **54,019 falskudkald mod 544 ægte defekter på tværs af 917 artikelnumre, omkring 99% af alarmerne** (arXiv 2505.03576) single source, preprint; på Siemens' Rastatt-fabrik var **97.5% af de kort, AOI'en afviste, gode** (Siemens-casestudie) vendor claim. En travl linje ser omkring 448 falskudkald pr. skift på 42–55 sekunder hver, 5.2–6.8 timer pr. skift, omtrent én fuldtidsgennemgangsperson (iFactory) vendor claim; timerne er et estimate.

FIG 03 **Omkring 97–99% af AOI-alarmerne er falskudkald*I dag bekræfter en person hvert udkald; AI-gennemgang rydder først de falske og sender resten til mennesker.***



Kilde: Siemens casestudie, iFactory (leverandørpåstand), DBGs årsrapport 2025; gennemgangstimer estimeret af forfatteren.

Ud fra "gennemgangstimer pr. skift = kort × falskudkald pr. kort × sekunder pr. kald" kan én gennemgangsperson dække omkring 0.7–1 volumenlinje med en regelbaseret AOI. Når AI fjerner 60–90% af falskudkaldene, ville vurderingstiden alene tillade 2–14 linjer pr. person, men gang, NG-bufferhåndtering og reparationsafleveringer begrænser lokal dækning til to til fire tilstødende linjer; kun central eller fjerngennemgang nærmer sig det teoretiske loft estimate; sekunder pr. kald kommer fra én kilde. Da gængse AOI'er scanner 60–80 cm² pr. sekund (Yunheng AOI-guide 2026), tager en side på 20×15 cm 4–5 sekunder; AI'en skal kun behandle de få flagrede områder inden for én takt, ikke genscanne kortet.

Hvad der allerede virker i produktion: superviseret klassifikation på AOI'ens egne billeder

Det beviste mønster er superviseret deep-learning-klassifikation på billeder fra den AOI, en fabrik allerede ejer. Fujitsus tillæg på ADLINKs AOI i Taiwan skar falskudkald ned med **mere end 90%** (Fujitsu-casestudie); OKI har brugt AI siden juli 2026 til at skære manuel bekræftelse efter AOI ned med omkring **80%** i sin egen EMS-forretning (Circuits Assembly); Siemens oplyser 60% mindre manuel analyse, €160,287 sparet i det første år og en tilbagebetaling på otte måneder (Siemens-casestudie) alle vendor claims. To ting er fælles for disse casestudier: næsten ingen offentliggør en escape-rate, og alle beholder folk på de resterende alarmer. Det ene casestudie, der giver et escape-tal (0.8% ned til 0.3%, mod en kontraktlig 0.5%), er internt inkonsistent (iFactory). AI skærer gennemgangstimer, ikke selve gennemgangsjobbet.

AI-gennemgang er bevæget sig fra tredjeparts-tillæg til OEM-funktioner og interne systemer, hvilket ændrer, hvad en nykommer måles op imod. Jutzes årsrapport 2025 opregner dets "3D-AOI-opgradering" som i volumenproduktion og siger, at den "bruger AI-algoritmer til at fjerne manuel gennemgang ved inspektionsenden" (Jutze årsrap-

port 2025); Sinic-Tek har udviklet et "multimodalt AI-assisteret manuelt gennemgangssystem" siden 2019 og samudvikler en top-og-bund-3D-AOI med Tencent, målrettet 99.5%+ defektdækning og 98% genkendelsesnøjagtighed (Sinic-Tek årsrapport 2025) vendor claims i primære indberetninger. Store EMS-virksomheder bygger selv. DBGs årsrapport 2025 opregner "AI-vurdering til synsinspektion" som et afsluttet projekt, der "dækker mere end 95% af de AOI-flagrede komponenter", med en "escape-rate under 20 PPM" og "mere end 85% mindre manuel gennemgang", fuldt sporbar (DBG årsrapport 2025) single source, primær. Det er det eneste elektronikfabrikstal, jeg fandt, som giver dækning, escapes og reduktion samlet. **Et tredjepartsprodukt solgt til tier-1 EMS skal matche DBGs 95% dækning, 20 PPM escapes og 85% reduktion**; under tier 1 er benchmarken den 50–85% reduktion, leverandører reklamerer med.

Anomalidetektion, der kun er trænet på gode dele, er hurtig (EfficientAD-S tager 2.2 ms pr. billede (EfficientAD, WACV 2024)), men svag på "logiske" anomalier som manglende eller forkerte dele, og de bedste metoder ligger i gennemsnit under 60% AU-PRO på det sværere MVTec AD 2 (MVTec AD 2) single source. Manglende og forkerte dele kontrolleres bedst mod CAD og BOM.

Loftet ved enhver pris: lag af maskiner, meget få mennesker, men ikke nul

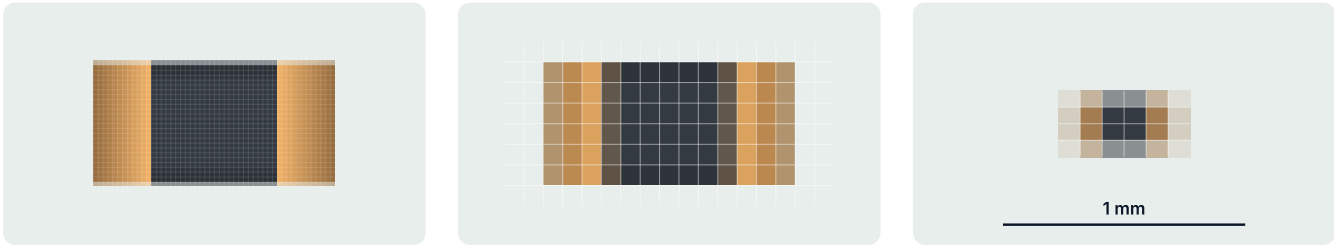
Jeg fandt ingen elektronikfabrik, der fjernede manuel gennemgang helt og nåede nul escapes. Praksis for IPC-A-610 Class 3-produkter (høj pålidelighed) lader stadig trænede inspektører afgøre de grænsetilfælde, som automatiske tærskler ikke kan klassificere, og én PCB-montør kalder "automatiseret inspektion kan erstatte menneskelig sign-off" en "almindelig misforståelse" (PCBCart) single source. Den menneskelige baseline skal også korrigeres: med henvisning til et Sandia-studie fra 2012 fanger en enkelt synsinspektør i bedste fald omkring **80%** af defekterne, to i serie omkring 96% (Instrumental) single source, citeret af en leverandør. Den forsvarlige påstand er "AI plus en lille menneskelig gennemgangsløkke slår helt manuel inspektion", ikke "AI erstatter inspektion".

Uden budgetgrænse er loftet 3D SPI, 3D-AOI, røntgen, AI-gennemgang og AI-udseendestationer stablet sammen, med folk kun tilbage til grænsetilfælde, nye produkter og omstillinger, nye defekttyper og kundesign-off. Flagskibsfabrikker oplyser 85–92% mindre manuel inspektion (I-Connect007, 2021, WiAdvance) vendor claims; mit estimat af loftet uden budget er gennemgangsarbejde på 8–15% af dagens niveau estimate. Slutmontage er stadig arbejdsintensiv: automatiseringen af iPhone 16-knapperne blev droppet på grund af høje defektrater (9to5Mac) single source.

Et generelt højopløsningskamera er ikke en AOI

Produktions-AOI'er afbilder flise for flise ved 8–20 µm pr. pixel, med 8–12 MP-kameraer, programmerbar flervinklet flerfarvet belysning og struktureret-lys-3D (Koh Young, Saki) verified. Branchen bemærkede for længe siden, at en 01005-del kun dækker omkring 8×16 pixels ved 25 µm pr. pixel, "muligvis ikke nok" til pålidelig vurdering (Circuits Assembly, 2007). Ved samme regnestykke estimate får et 4K-kamera over et 300 mm-panel omkring **78 µm pr. pixel**: en imperial 0402-del er omkring 13×6 pixels, nok til tilstedeværelse og grov offset, men ikke til lodningsform, og en 01005 er omkring 5×3 pixels, reelt usynlig. At nå 15 µm pr. pixel over et helt panel ville kræve omkring 330 megapixel, hvilket er grunden til, at AOI'er flytter et kamera hen over kortet. Når nogen siger "0402", så tjek, hvilken de mener: metrisk 0402 er imperial 01005, langt ud over ethvert vidvinkelkamera.

FIG 04 **Et generelt kamera er ingen AOI***Et 4K-kamera over et 300 mm-panel giver cirka 78 µm pr. pixel: en 0402-del krymper til ≈13×6 pixels og en 01005 forsvinder.*



AOI: 8–20 µm pr. pixel
0402 ≈ 100×50 px

4K-kamera over et 300 mm-panel: ≈78 µm/px
0402 krymper til ≈13×6 px: nok til tilstedeværelse og grov forskydning

samme kamera på en 01005
≈5×3 px: i praksis usynlig

Kilde: Circuits Assembly; forfatterens beregning for et 300 mm-panel og en 4K-sensor. Pixelnettene er illustrative.

På linjer, der allerede har AOI'er, er det bedste sted for AI oven på AOI'ens billeder. Kun linjer uden AOI, eller end-of-line-stationer, retfærdiggør en indkapslet kamerastation med kontrolleret belysning; et åbent højopløsningskamera over linjen er mest udsat for omgivende lys og skygger (CSDN).

Montage og håndtering: indlært perception plus programmeret bevægelse er i produktion

IFR's 2026-rapport siger, at "maskinsyn til inspektion og robotstyring er modent og industrielt", mens "generel fysisk AI og læring fra simulering og erfaring endnu ikke er industriel kvalitet" (IFR World Robotics 2026-resumé) single source, primær . Elektronik er den største robotkøber (142,235 enheder installeret globalt i 2025, 59% af alle installationer i Kina (IFR, IFR Kina-pressemeddelelse)), men Kina har kun 166 robotter pr. 10,000 produktionsarbejdere mod 307 i USA (IFR, april 2026). Mønsteret i produktion er indlært perception plus deterministisk bevægelse; end-to-end indlært bevægelse er tæt på produktionsgrad kun ved strukturerede stationer (kapitel 3 og 6).

Hvad hver region kan bruge

Fastlands-Kina er afhængigt af indenlandske modeller og lokale vægte; i Taiwan afgør kundens identitet; i USA afgør told og certificering.

Element	Fastlands-Kina	Taiwan (private fabrikker)	Sydøstasien	USA
Claude (enhver kanal)	● Ikke tilgængelig	● Betinget: leverandøren må ikke være majoritetsejet eller kontrolleret af personer eller enheder i fastlands-Kina, Hongkong eller Macau	● Samme	● Samme
GPT- og Gemini-cloud	● Ikke tilgængelig	● Tilgængelig	● Tilgængelig (lister ikke kontrolleret land for land)	● Tilgængelig
Kinesiske cloudmodeller (Qwen, DeepSeek, Doubao, GLM, Kimi)	● Tilgængelig (registrerede modeller; at udbyde en tjeneste kræver registrering)	● Lovlig, men fastlandshosting passerer sjældent kundens sikkerhedsgennemgang; forbudt i den offentlige sektor, statsvirksomheder og kritisk infrastruktur	● Lovlig (internationale noder)	● Lovlig, streng gennemgang

Element	Fastlands-Kina	Taiwan (private fabrikker)	Sydøstasien	USA
Lokale åbne vægte af kinesisk oprindelse (Qwen, GLM, DeepSeek)	🟢 Førstevalg; privat fabriksimplementering kræver ingen model- eller algoritmeregistrering (advokatfirma-læsning)	🟢 Tilgængelig (offentlig sektor begrænset)	🟢 Tilgængelig	🟡 Betinget: oprindelsesgennemgang
Lokale åbne vægte af ikke-kinesisk oprindelse (f.eks. Gemma 4)	🟢 Tilgængelig	🟢 Tilgængelig	🟢 Tilgængelig	🟢 Tilgængelig
Klassiske ikke-generative visionsmodeller	🟢 Tilgængelig (uden for de generative AI-regler)	🟢 Tilgængelig	🟢 Tilgængelig	🟢 Tilgængelig
Jetson Thor/T5000/T4000, DGX Spark	🟢 Tilgængelig: NVIDIA klassificerer dem 5A992.c, ingen USA-licens nødvendig for en almindelig civil fabrik; Spark har en Kina-SKU; kanalpriser omkring 40–60% over USA-liste	🟢 Tilgængelig	🟢 Tilgængelig	🟢 Tilgængelig
RTX 5090, RTX PRO 6000, DGX Station	🔴 Ikke tilgængelig: de to første er 4A090.a, DGX Station 5A992.z; lovlige alternativer er 4090D, 5090 D V2, RTX PRO 5000 72GB, RTX 6000D, L20 (4A994.l)	🟢 Tilgængelig	🟢 Tilgængelig, men købere med hovedsæde i eller i sidste ende ejet fra Kina tæller som kinesiske; Malaysia kræver tilladelser til eksport og omladning	🟢 Tilgængelig
Mac Studio M5 Ultra 512 GB	🟢 Tilgængelig (Apple China-butik)	🟢 Tilgængelig	🟡 Formodet tilgængelig	🟢 Tilgængelig
Huawei Ascend og andre kinesiske acceleratorer	🟢 Tilgængelig	🔴 Høj risiko	🔴 Høj risiko	🔴 Ikke brugbar i praksis
Fastlands-fremstillede faste arme, AOI'er, industrikameraer	🟢 Tilgængelig	🟡 Kan importeres; siden 1. december 2025 kan myndighederne handle direkte mod fastlandsmærker, og styrelser, statsvirksomheder, visse fonde og kritisk infrastruktur-udbydere kan begrænses eller udelukkes fra at bruge dem	🟢 Tilgængelig	🟡 Lovlig, men samlet told er omkring 40% på arme, 37.5% på instrumenter af AOI-typen og kameraer, 25% på edge-computere
Udenlandsk fremstillede AMR'er og mobile manipulatorer	🟢 Tilgængelig	🟢 Tilgængelig	🟢 Tilgængelig	🟡 Nye modeller kræver betinget godkendelse fra USA's War Department
Hikvision- og Dahua-kameraer	🟢 Tilgængelig	🟡 Privat brug tilladt, offentlig sektor forbudt	🟢 Tilgængelig	🔴 Nye FCC-autorisationer blokeret; forbudt for føderale entreprenører

Element	Fastlands-Kina	Taiwan (private fabrikker)	Sydøstasien	USA
Robotsikkerhed og certificering	● GB 11291.1-2011 og GB 11291.2-2013 obligatoriske, under revision siden april 2026; CR-certificering frivillig; GB/T 39404-2026 dækker controller-cybersikkerhed	● Indhegning som standard; hegnløst samarbejde skal opfylde CNS 14490 eller ISO 10218 med en skriftlig vurdering	● Pr. kunde og eksportdestination, bekræftet skriftligt før forsendelse	● Elektrisk udstyr skal være "godkendt" (i praksis et NRTL-mærke eller feltevaluering); UL 1740 kræver, at robotceller opfylder ISO 10218-2 siden 11. august 2025
Fabs og fabrikker med kontrollerede produkter	—	● TSMC's indkøbskontrakter kræver SEMI E187; fastlandsudstyr fjernes fra 2 nm-fabs	—	● ITAR tekniske data omfatter "fotografier"; afvisningspolitik for Kina; fastlandskontrollerede leverandører kan i praksis ikke betjene forsvarselektronikfabrikker

Kilder til tabellen: (Anthropic understøttede regioner, OpenAI understøttede lande, Gemini-tilgængelighed, CAC foreløbige foranstaltninger om generativ AI, NVIDIA eksportklassifikationsdata, eCFR 15 CFR 774, NVIDIA Kina DGX Spark, Apple Kina, Covington om nye BIS-regler, Greenberg Traurig, juni 2026, Fulcrum, Tom's Hardware om Ascend-vejledning, Taiwan BOFT-importliste, Taiwans historie for Cyber Security Management Act, Taiwans gennemgangs-regler for produkter, der bringer national cybersikkerhed i fare, Taipei Times, USITC HTS kapitel 99, Latham & Watkins, IPVM, Kinas nationale standardplat-form, GB 11291.2, Intertek UL 1740-opdatering, eCFR 22 CFR 126.1).

Fastlands-Kina: en komplet stak, men hverken lukkede frontmodeller eller de stærkeste enkeltmaskiner kan komme ind. Registrerede indenlandske cloudmodeller, lokale åbne vægte (Qwen under Apache-2.0, GLM og DeepSeek under MIT), klassiske visionsmodeller og indenlandske AOI'er, kameraer og cobotter er alle brugbare. NVIDIAs egne eksportklassifikationer opregner Jetson AGX Thor-kittet, T5000/T4000-modulerne og DGX Spark som 5A992.c, så en almindelig EMS- eller PCBA-fabrik ikke behøver nogen USA-licens, mens RTX 5090 og RTX PRO 6000 (4A090.a) og DGX Station (5A992.z, TPP 58,484) ikke kan købes på fastlandet (NVIDIA eksportklassifikationer, eCFR) [primær; den juridiske læsning er min og ikke juridisk rådgivning]. RS China opregner Thor-kittet til RMB 59,104 inklusive skat, omkring 1.4 gange USA-listeprisen (RS China).

Fire advokatfirmaer er enige om, at en privat fabriksimplementering til fabrikkens eget brug ikke kræver nogen model- eller algoritmeregistrering, mens en hostet multi-tenant-tjeneste ligger i en gråzone (Covington, FPF, JunHe) verified: konsistente læsninger, ingen officiel afgørelse. GB 11291.1-2011 og GB 11291.2-2013 forbliver den obligatoriske robotsikkerhedsbaseline og har været under revision siden 28. april 2026; GB/T 39404-2026 om controller-cybersikkerhed trådte i kraft den 1. november 2026 (revisionsplan 20262542, GB/T 39404-2026) single source, primær. Indtil den nye GB 11291 kommer (2028–2029 estimate), skal man designe efter ISO 10218:2025 og beholde GB 11291-overensstemmelsesfiler.

Taiwan: private fabrikker kan bruge næsten alt; kundens identitet afgør. Front-API'er, åbne vægte og fastlandsudstyr er alle lovlige (data.gov.tw). Men siden den ændrede Cyber Security Management Act trådte i kraft den **1. december 2025**, kan myndighederne handle direkte mod produkter, der er "fastlandsmærker", og begrænse dem ved myndighedsorganer, statsvirksomheder og kritisk infrastruktur-udbydere (Laws & Regulations Database, gennemgangsregler) verified; listen over kritisk infrastruktur-udbydere er fortrolig; og TSMC har skrevet SEMI E187 ind i sine indkøbskontrakter (CNA) og siges at fjerne fastlandsudstyr fra sine 2 nm-fabs (TrendForce) single source. Private EMS-, montage- og små PCBA-fabrikker er åbne; fabs, parkforsyning og regeringsrelaterede kunder er lukkede i praksis. En fastlandskontrolleret leverandør bør designe efter E187, sende en software bill of materials, tilbyde offlineopdateringer og aldrig oprette fjernforbindelse fra fastlandet.

Sydøstasien: færrest begrænsninger; problemerne er langsom tilbagebetaling og kinesiske moderselskaber. Alt er brugbart, men USA's vejledning af 31. maj 2026 kræver en licens til at eksportere avancerede computerprodukter til enheder med hovedsæde i, eller i sidste ende ejet fra, Kina, selv når modtageren er i et tredjeland (Greenberg Traurig) single source, og Malaysia har sine egne tilladelser til avancerede chips (Fulcrum). Efterspørgslen flytter hertil: TrendForce forventede, at Thailand, Vietnam og Malaysia ville nærme sig 50% af server-ODM-SMT-kapaciteten i 2026 (TrendForce) single source, en prognose fra 2023.

USA: hele stakken, til prisen af told, certificering og kontrollerede data. Under kapitel 99 i USA's toldskema betaler en fastlands-fremstillet robotarm omkring 40% i samlet told, instrumenter af AOI-typen og kameraer omkring 37.5% og edge-computere 25%; fremstillet i Malaysia eller Indonesien omkring 10–12.5%, i Vietnam, Thailand eller Filippinerne 12.5–15%, i Taiwan 10%, med computere på 0% uden for fastlandet (USITC HTS kapitel 99, Great Lakes Customs Law) primært skema + estimate. USTR's meddelelse af 10. november forlængede 178 gamle Section 301-undtagelser til 9. januar 2027, hvoraf ingen dækker robotter, inspektionsinstrumenter, kameraer eller computere (Federal Register 2026-22925) verified.

OSHA kræver, at elektrisk udstyr på arbejdspladsen er "godkendt" (i praksis et NRTL-mærke eller en feltevaluering på omkring $3,000 til $20,000+ pr. celle), og siden 11. august 2025 bringer UL 1740 hele robotceller ind under ISO 10218-2 (eCFR 1910.399, Intertek, Industrial Monitor Direct). Forsvarselektronik er lukket for fastlandskontrollerede leverandører i praksis: ITAR behandler produktionsfotografier som tekniske data og anvender en afvisningspolitik over for Kina (eCFR 22 CFR 120.33, eCFR 22 CFR 126.1). En USA-bestemt celle bør bruge en fast arm, ubegrænsede kameraer og lokale eller USA-cloudmodeller, med slutmontage i Sydøstasien, hvor det reelt ændrer oprindelsen.

Ved ethvert budget adskiller loftet sig pr. region, men når ingen steder nul mennesker estimate. Fastlandet kan stable indenlandsk 3D SPI, 3D-AOI, røntgen, AI-gennemgang og et internt AI-team for at bringe gennemgangsarbejde ned til 8–15%, men uden lukkede frontmodeller eller DGX Station skal fabriksdækkende LLM-Q&A køre på lokale åbne vægte og indenlandsk computerkraft (en fuldskalamodel til mange brugere kræver mindst to Ascend-servere med otte kort). Taiwan og Sydøstasien kan teknisk bruge det bedste af alt, herunder front-API'er og DGX Station, begrænset af kundeidentitet, fastlandsmærker og licenser til kinesiske moderselskaber. USA har ingen tekniske grænser, til prisen af told, NRTL-certificering og personadskillelse for kontrollerede data. Overalt forbliver folk til grænsetilfælde, nye produkter og omstillinger, nye defekttyper og kundesign-off.

Et produkt, der sælges på tværs af regioner, har derfor brug for to modelstakke: en fastlandsstak på registrerede indenlandske modeller eller lokale åbne vægte, og en oversøisk stak på front-API'er plus en lokal model af ikke-kinesisk oprindelse. Anthropics politik med virkning fra 12. november 2026 vurderer efter ejerskab og kontrol, så en fastlandskontrolleret leverandør ikke kan bygge på Claude, selv ikke gennem et oversøisk datterselskab, en cloudplatform eller en forhandler (Anthropic) single source, primær.

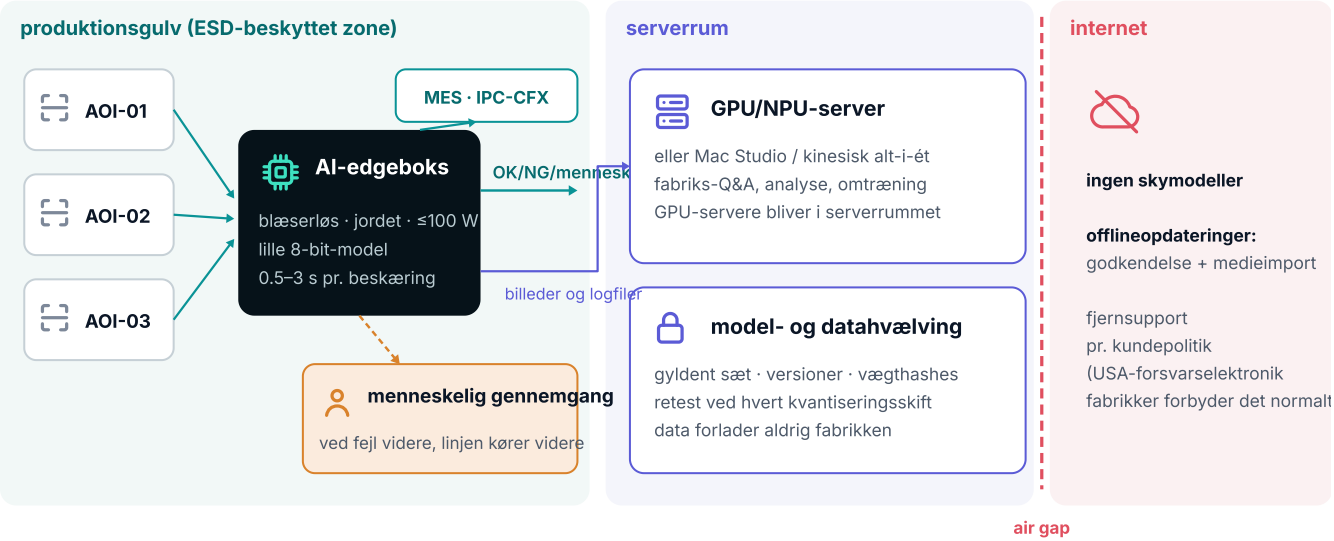
On-premises

02 At køre AI inde i fabrikken: renrum, luftgab, hardware og kvantisering

I en luftgappet fabrik er begrænsningen proces, ikke computerkraft

Fuldt offline synsinspektion er rutine; begrænsningerne er proceduremæssige. Kinas MIIT-vejledning fra 2024 om sikkerhed i industrielle kontrolsystemer beder om zonering, firewalls eller datadioder, auditeret tidsbegrænset fjernvedligeholdelse, scannede medier og testede patches (vejledningstekst) single source ; halvlederkunder tilføjer SEMI E187 og E188, som TSMC skrev ind i sine indkøbskontrakter i 2023 (SEMI, TXOne) verified . Resultater går tilbage til MES over IPC-CFX, og kortidentitet bevæger sig over IPC-HERMES-9852 (Hermes-standard). Den almindelige fælde er licensering, der forudsætter et netværk: Roboflows offlinetilstand cacher vægte på en 30-dages lease (Roboflow). Et offlineprodukt har brug for offlineaktivering, signerede opdateringer på scannede medier, en software bill of materials, USB-kontrol, lokale revisionslogfiler og en CFX-adapter.

FIG 05 **En air-gapped udrulning** ***En lille boks ved linjen beslutter; serverrummet svarer på spørgsmål; ved fejl falder de tilbage på mennesker og data forlader aldrig fabrikken.***



Kilde: Forfatterens syntese (kapitel 2).

Lukkede frontmodeller kommer i praksis ikke ind i en fabriks serverrum. Gemini udbydes kun luftgappet gennem Google Distributed Cloud for regeringer og store virksomheder (Google Cloud); Microsofts Foundry Local åbner kun store modeller for "kvalificerede kunder" og nævner ingen GPT-model (Microsoft); Claude udbydes kun i cloud-regeringsregioner uden on-premise-implementering (Claude for government); ingen af de tre udbydes i fastlands-Kina. Fabriksudstyr skal designes omkring åbne vægte; frontmodeller kan højst annotere eller forklare billeder i hovedkvarteret, offline, når kunden tillader anonymiserede billeder ud.

Serverrum-hardware bør behandles som kontorklasse: Mac Studio er klassificeret til 10–35 °C og DGX Spark til 5–30 °C, med rapporter om termiske nedlukninger under vedvarende inferens (Apple, NVIDIA-forum) verified . Linjen får blæserløse industribokse, såsom en Jetson T4000-industri-PC klassificeret til −25 til 60 °C (AAEON) vendor claim . Fordi en times nedetid koster en stor fabrik $36,000 til $2.3 mio. (Siemens True Cost of Downtime 2024) [le-

verandørundersøgelse], skal hvert AI-trin på linjen **fejle åbent**: ved enhver fejl falder det tilbage til den eksisterende AOI-plus-folk-proces.

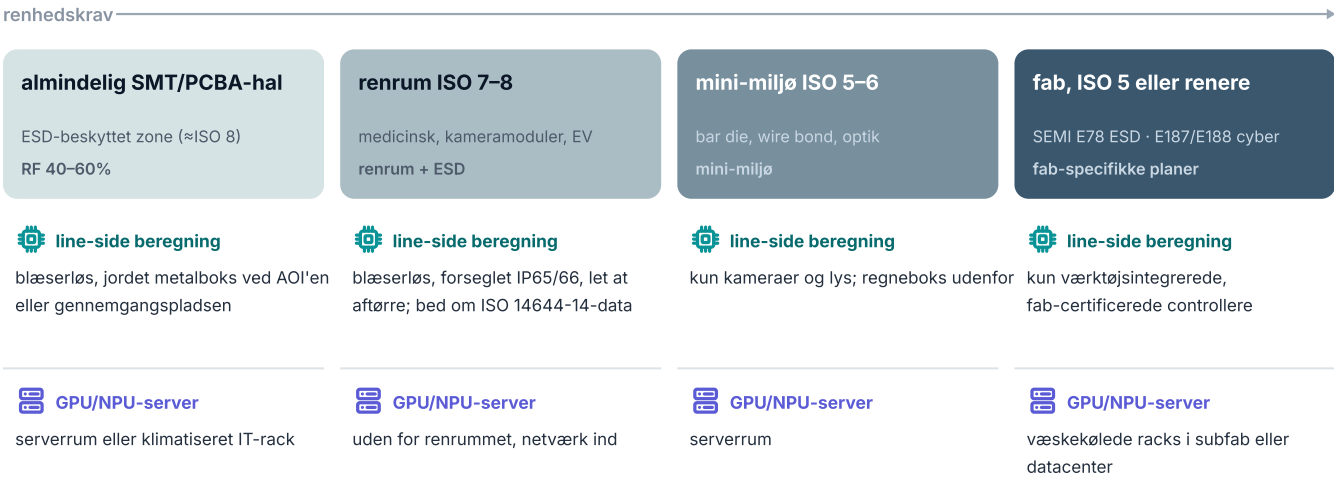
Ren- og ESD-zoner: en almindelig SMT-hal er et ESD-beskyttet område, ikke et klassificeret renrum

Det, der begrænser, hvor computerkraft kan stå, er statisk elektricitet og luftstrøm, ikke renhedsklasse. De fleste SMT-montageområder er "ikke formelt partikelkontrollerede", og fabrikker i verdensklasse "kan nå eller nærme sig ISO 8" (Circuitnet-ekspertpanel) single source. Medicinsk PCBA og kameramoduler ligger typisk i ISO 7, bare die og optisk justering i lokale ISO 5–6 (PCBSync) single source; renrumsbyggere opgiver ISO 7–8 for SMT generelt (Angstrom Technology) vendor claim.

Det universelle krav er ESD-kontrol: ANSI/ESD S20.20-2021 kræver isolerede ledere under 35 V og holder stærkfeltsisolatorer mindst 30 cm fra følsomme enheder (ANSI, ESDA, Desco) verified. Fabs placerer computerkraft i sub-fab'en eller datacenteret, i væskekølede racks, der opfylder SEMI S2/S8/F47 (Prodrive) vendor claim.

Hvorvidt udstyr kan komme ind i et rent område, testes for partikelemission under ISO 14644-14; robotter har sådanne certifikater (Fraunhofer IPA), men jeg fandt ingen industri-PC, edge-AI-boks eller GPU-server med ét. CE, FCC og IEC 60068 viser, at en boks tåler sit miljø, ikke at den må komme ind i et ESD-beskyttet område; kundens ESD-koordinator kontrollerer det. GPU-servere kommer aldrig ud på linjen: en DGX Station trak 1,292 W ved GPU'en i test (StorageReview).

FIG 06 **Ren- og ESD-zoner afgør, hvor beregning kan stå** ***En almindelig SMT-hal er en ESD-beskyttet zone, ikke et renrum. GPU- og NPU-servere hører hjemme i serverrummet.***



Kilde: ANSI/ESD S20.20, IEC 61340-5-1, ISO 14644, SEMI-standarder og leverandørmateriale; sammensat af forfatteren.

Zone	Typisk miljø	Computerkraft ved linjen	GPU/NPU-servere
Almindelig SMT/PCBA-hal (forbruger, industriel)	ESD-beskyttet område under S20.20 eller IEC 61340-5-1, normalt ikke et klassificeret renrum (≈ISO 8-ækvivalent); 40–60% RF	Blæserløs boks i et jordet metalkabinet (Jetson T4000/Thor, kinesisk NPU-boks) ved siden af AOI'en eller gennemgangspladsen; plastmonitorer og -kabinetter uden for området eller ≥30 cm fra kort	Serverrum eller klimatiseret IT-rack
Medicinsk, kameramoduler, EV-elektronik (ISO 7–8 renrum)	Renrum plus ESD-beskyttelse	Blæserløs, IP65/66-forseglet, glat og let at aftørre; bed om ISO 14644-14-data eller et IPA-certifikat	Uden for renrummet, netværket føres ind

Zone	Typisk miljø	Computerkraft ved linjen	GPU/NPU-servere
Bare die, wire bond, optiske stationer (lokale ISO 5–6)	Mini-miljø	Kun kameraer og lys indenfor; regneboks udenfor	Serverrum
Fab-renrum (ISO 5 og renere)	Fab-specifikke ESD- og elektrostatisk-tiltrækningsplaner, udstyr efter SEMI E78; cybersikkerhed efter SEMI E187/E188	Kun værktøjsintegrerede, fab-kvalificerede controllere; ingen ekstra PC'er med blæsere	Væskekølede racks i sub-fab'en eller fab-datacenteret

Tabellen er min syntese af kilderne ovenfor estimate. For et produkt betyder det en blæserløs linjeside-boks under omkring 100 W, et metalkabinet forbundet til beskyttelsesjord, 12–24 V DC-input, bredtemperaturdrift ved reel kabinetluftstrøm og en ISO 14644-14-partikelemissionsrapport klar til kunder ved ISO 7 og renere. Ethvert design, hvor "én GPU-server betjener fire linjer", placerer serveren i serverrummet og efterlader kun en lille boks eller et kamera ved linjen.

Hastighedsgrænser pr. anvendelse: nøjagtighed er sværere end hastighed

Fabrikker og leverandører offentliggør næsten aldrig AI-latens- eller tilgængelighedsmål. Tærsklerne nedenfor er udledt af takt, menneskelig gennemgangstid, robotstyringsrater og interaktionsforskning estimate:

Anvendelse	Hastighedsgrænse	Nøjagtighedsgrænse	Tilgængelighed
Inline-inspektion af hver del	Inden for ½–1 takt: 240–1,000 udsnit i en 10–30 s takt, 8–100 udsnit pr. sekund	Escapes ikke værre end dagens "AOI plus folk" (kontrakteksempel <0.5%)	Fejl åbent, stop aldrig linjen; mål ≥99.9% (ingen offentlig SLA)
Inline second opinion kun på AOI-flagrede udsnit	Flag (normalt ≤5–20 pr. kort) inden for ½ takt: 0.5–3 s pr. udsnit	Samme; trænet på linjens egne billeder	Falder tilbage til menneskelig gennemgang
AI-gennemgang af falskudkald	2–5 s pr. kald, mere end ti gange hurtigere end en persons 42–55 s	Escapes ikke værre end en menneskelig gennemgangsperson	Degradering tilføjer folk, stopper aldrig linjen
Robotstyring (VLA)	100–300 ms pr. handlingschunk med realtidschunking, 10–50 Hz output	≥99% opgavesucces for at slå programmeret bevægelse	Sikkerhedssystem uafhængigt af modellen
Robotplanlægning og færdighedsvalg	1–3 s, når det kræves hver cyklus; ≤1 min, hvis kun ved undtagelser eller omstillinger	Output skal være verificerbar kode eller et færdigheds-ID	Falder tilbage til det faste program
Q&A (SOP'er, MES-forespørgsler)	Første token ideelt ≤1–2 s, højst 10 s; ≥15–20 tokens/s pr. bruger; 5–20 samtidige brugere	Præcise citater; tro mod lang kontekst	Kontor-IT-grad (≈99–99.5%)
Analyse- og rodårsagsrapporter	Minutter er fint (batch)	Højest; tro mod lang kontekst	Kontor-IT-grad

Physical Intelligence målte 97 ms inferens for en VLA i 3B-klassen (omkring 108–139 ms med netværk og forbehandling), fandt, at realtidschunking holder gennemstrømningen uændret op til 200 ms ekstra forsinkelse, og gennemførte finmotoriske opgaver som at tænde en tændstik eller sætte et netværkskabel i ved mere end 300 ms (Physical Intelligence) [leverandørforskning]; interaktionsforskning sætter grænsen for at holde en tankerække ved ét sekund og for opmærksomhed ved ti (NN/g). Nøjagtighed er den svære del: åben Qwen3-VL-8B opnår i gennemsnit kun 56.8 zero-shot på tværs af fem industrielle anomalidatasæt (IndusAgent) single source. Ingen hardware redder en vision-language-model, der ikke er tunet på linjens egne billeder.

FIG 07 **Hastighedsgrænser pr. anvendelse***Inline-inspektion måles i millisekunder, Q&A i sekunder, analyse i minutter. Mennesker bruger 42–55 s pr. udkald.*



Kilde: Forfatterens syntese (kapitel 1–2).

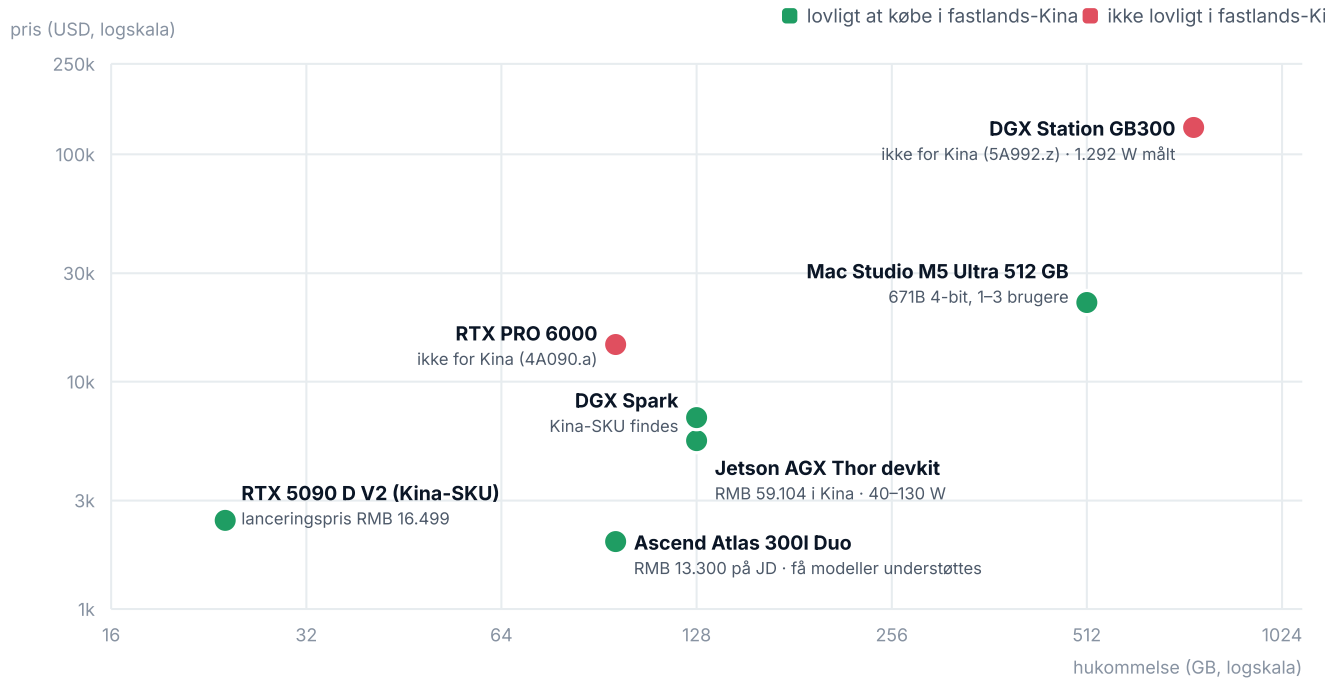
Hardware: Thor til stationen, et Mac Studio eller en indenlandsk enhed til fabriks-Q&A, og ingen DGX Station på fastlandet

Enhed	Hukommelse og båndbredde	Pris (november 2026)	Hvad den er god til	Fastlands-Kina
Jetson AGX Thor-devkit / T5000 / T4000	128 GB LPDDR5X, 273 GB/s; 40–130 W	$3,499 ved lancering → $5,499 (juli 2026); T5000 $4,999, T4000 $2,999; RS China-devkit RMB 59,104 inkl. skat	2–8B visionsmodeller, der gentjekker udsnit; realtids-VLA'er i 3B-klassen	Ja (5A992.c, ingen licens for civile fabrikker)
DGX Spark	128 GB, 273 GB/s	$3,999 → $4,699 → $6,950; samme-chip ASUS GX10 Kina-model RMB 32,999 på JD (okt 2025)	Multi-bruger 8–35B MoE	Ja (5A992.c, Kina-model, autoriserede kanaler)
RTX PRO 6000	96 GB GDDR7	≈$8,565 → $13,250 (juni 2026) → ≈$16,005 gennemsnitlig eBay-notering i november	Multi-bruger 27–35B FP8; 120B-klasse MoE ved 4-bit	Nej (4A090.a)
Mac Studio M5 Ultra 512 GB	512 GB unified, 1.2 TB/s; 480 W maks. vedvarende	Fra $22,299 i USA; ≈RMB 173–186k estimeret i Kina	DeepSeek 671B ved 4-bit til 1–3 brugere	Ja
DGX Station GB300	748 GB (252 GB HBM3e ved 7.1 TB/s + 496 GB LPDDR5X ved 396 GB/s); 1,292 W målt ved GPU'en	$123–140k gade pris i november (≈RMB 830–940k før told); $95–117k i august	Høj samtidighed på 120–235B MoE, der passer i HBM; kun få brugere på 433–743B-modeller	Nej (5A992.z, TPP 58,484)
Huawei Atlas 300I Duo	96 GB LPDDR4X, 408 GB/s	RMB 13,300 fra en JD-forhandler	Lavvolumen batchgennemgang; få understøttede modeller	Ja (indenlandsk)
Huawei Atlas 800I A2 (8 × Ascend 910B 64 GB)	8 × 64 GB HBM	Ikke indhentet; fuld DeepSeek kræver to ved W8A8, fire ved BF16	Fabriksdækkende multi-bruger Q&A og analyse	Ja (indenlandsk)

Enhed	Hukommelse og båndbredde	Pris (november 2026)	Hvad den er god til	Fastlands-Kina
RTX 5090 D V2 (Kina-model)	24 GB	RMB 16,499 ved lancering (aug 2025), steget siden med hukommelsespriser	8B visionsmodeller, små VLA'er	● Ja

Kilder: (NVIDIA Kina Thor, NVIDIA Thor-lancering, CNX om prisstigninger, NVIDIA eksportklassifikationsdata, RS China, Implicator, IThome om ASUS GX10, Tom's Hardware RTX PRO 6000, GPU Poet, Apple-specifikationer, TechRadar, IThome om M5 Ultra Kina-priser, Wccftech, StorageReview, CyberPowerPC, monitors.com, Tom's Hardware DGX Station, vLLM-Ascend-understøttelsesmatrix, China Telecom Cloud, Huawei Atlas 800I A2, TweakTown). Fastlandspriser for 512 GB Mac Studio er ekstrapoleret fra USA-konfigurationstillæg og er lavkonfidens estimate .

FIG 08 **On-prem beregning: hukommelse, pris og hvad der lovligt kan købes i fastlands-Kina*Eksportklassifikationer afgør shortlisten før benchmarks.***



Kilde: NVIDIAs eksportklassifikationer, Apple, RS China, JD.com og anmeldelsespriser (november 2026).

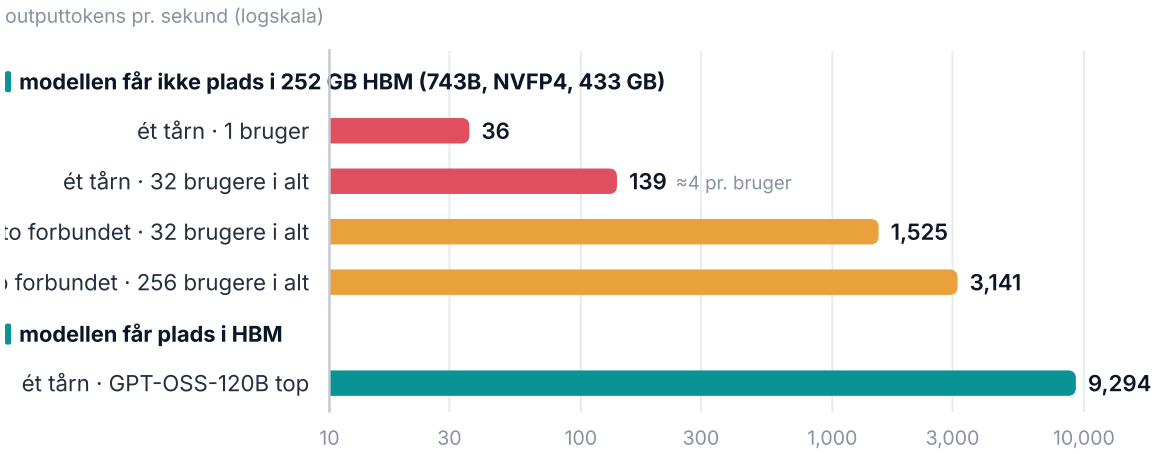
Jetson Thor får stationsniveau-implementering til at fungere. NVIDIAs TensorRT Edge-LLM-test viser Qwen3-VL-8B (NVFP4) præfylde 2,048 tokens på 104.5 ms og afkode 44.4 tokens/s på Thor (TensorRT Edge-LLM) vendor claim ; for et AOI-udsnit og et 20-token-svar er det omkring 0.5 s med 8B-modellen og 0.2 s med 2B, eller 8.5–21 udsnit pr. sekund batchet estimate . Outputlængde, ikke parameterantal, sætter latensen, så begrænset afkodning til et klassetoken eller kort JSON er en del af designet. GR00T N1.7 (3B) kører ved 25.1 Hz på Thor (Jetson AI Lab) vendor claim . En Orin NX tager omkring 1.8 s pr. udsnit, og en Atlas 300I Duo har en p95 over 23 s, kun egnet til batcharbejde (CSDN) single source . På fastlandet skal man budgettere NVIDIA-produkter til 40–60% over USA-listeprisen estimate .

Et 512 GB Mac Studio gør en fuldskalamodel på én maskine mulig, for nogle få brugere. M5 Ultra (august 2026) tilbyder op til 512 GB ved 1.2 TB/s fra omkring $22,299 (Apple, TechRadar); DeepSeek 671B ved 4-bit (≈380–404 GB) afkoder 17–20 tokens/s på en M3 Ultra (Pinggy) verified og estimeret 25–29 på M5 Ultra. Lange prompts er det svage punkt: en 10,000-token retrieval-prompt tager omtrent 17–70 s estimate , ud over 10-sekunders interaktionsgrænsen.

DGX Station er den stærkeste enkeltmaskine, men ikke tilgængelig på fastlandet og fem til seks gange prisen på en Mac. Den leverede GB300 har **748 GB**, hvoraf kun 252 GB er hurtig HBM (Wccftech, Tom's Hardware, aug 2026) verified . Uafhængige test viser omkring 9,300 outputtokens/s på GPT-OSS-120B, som passer i HBM, men en 743B-model, der flyder over i LPDDR5X, gav 32 brugere kun 139 tokens/s i alt, omkring 4 hver; to forbundne enhe-

der nåede 1,525 (StorageReview, aug 2026, StorageReview, okt 2026) single source. Gade priser steg til $123–140k i november (CyberPowerPC, monitors.com). På fastlandet er de realistiske muligheder for samme job Mac Studio, KTransformers-agtige CPU+GPU-servere (KTransformers) [leverandørtest] eller indenlandske Ascend-servere.

FIG 09 **Én stor arbejdsstation giver ingen delt assistent** ***Når modellen ikke får plads i højkapacitetshukommelse, får 32 samtidige brugere cirka 4 tokens pr. sekund hver.***



Kilde: StorageReview (aug og okt 2026); ServeTheHome. Uafhængige tests fra én kilde.

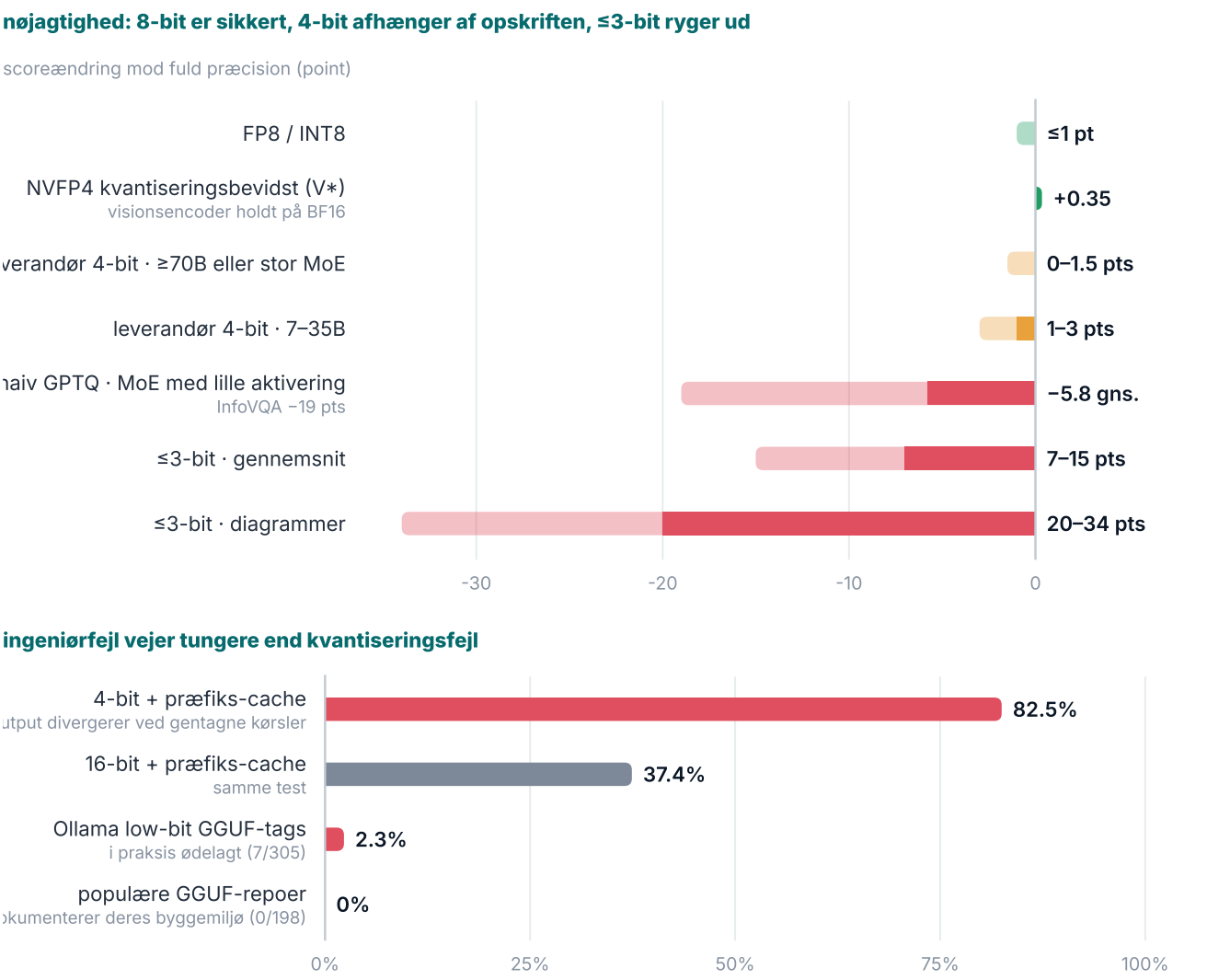
Hukommelsesmangel flyttede de samme dele 20–100% på et år (DGX Spark gik fra $3,999 til $6,950 (TechRadar)), så udstyrstilbud bør indekseres frem for fastlåses.

Indenlandske acceleratorer kører fuldskalamodeller; softwarevedligeholdelse er flaskehalsen. Ascend 910B har 64 GB HBM, men vLLM-Ascend markerer DeepSeek V3/R1 som "ikke længere vedligeholdt" og understøtter kun få modeller på mindre kort (vLLM-Ascend) single source, primær, og hver ny model tager én til to måneder at tilpasse (Caijing/Anue); send en frossen, valideret modelversion. At betjene fuld DeepSeek 671B til 20–100 brugere kræver mindst to Atlas 800I A2-servere med otte kort (China Telecom Cloud) verified, men offentlige udbud for LLM-enheder i Kina udgjorde i alt kun RMB 27.5 mio. i januar–september 2025 (C114) single source: fabrikker betaler for integration og service, ikke for en boks computerkraft.

Kvantisering: 8-bit er sikkert, 4-bit afhænger af opskriften, 3-bit og derunder er ude

8-bit (FP8 eller INT8) er stort set tabsfrit for vision-language-modeller (Red Hat, Hugging Face) verified. Ved 4-bit afhænger tabet af størrelse og opskrift: Qwens 72B AWQ-build taber under ét point på MMMU, mens 7B taber 2.8 (Qwen AWQ) verified, og detaljetunge opgaver falder mere (naiv 4-bit GPTQ koster Qwen3-VL-30B-A3B 19 point på InfoVQA (VEQ)). Ved 3-bit og derunder er det gennemsnitlige tab 7–15 point, og diagrammer taber 20–34 (VLMQ) verified; ud over 64k tokens taber 4-bit op til 59%, værre på ikke-engelske input (arXiv 2505.20276). Hold visionsencoderen på FP16/BF16.

FIG 10 **Kvantisering: 8-bit er sikkert, 4-bit afhænger af opskriften, 3-bit og lavere ryger ud*Ingeniørfejl i serving-stacken koster mere nøjagtighed end selve kvantiseringen.***



Kilde: NVIDIA Nemotron-modelkort, VLMQ, VEQ, arXiv 2609.05881, 2610.19498, 2609.04748 med flere.

To advarsler betyder mere. For det første: **pr. 11. november 2026 har ingen undersøgelse eller modelkort målt kvantisering på PCB/SMT-defektbilleder**, så revalider på det gyldne sæt. For det andet er serving-stack-fejl ofte større end kvantiseringsfejl: den samme model scorede omkring 30 point lavere under Ollama end under referenceimplementeringen (SiliconBench); 7 af Ollamas 305 officielle kodemodel-tags var stort set ødelagte, selvom filerne var intakte (arXiv 2609.05881); ingen af 198 populære GGUF-repositorier registrerede deres build (arXiv 2610.19498); og med prefix-cache-genbrug divergerede 82.5% af episoderne ved 4-bit mod 37.4% ved 16-bit (arXiv 2609.04748) single sources. Leverandørens kvantiseringsbevidste 4-bit er tæt på tabsfrit på detaljeopgaver (NVIDIA Nemotron 3.5) vendor claim, men 4-bit VLA'er taber 1.5–4.3 point under forstyrrelse og kører langsommere på Thor (LoopQ) single source. Håndter en kvantiseret model som et nyt måleinstrument: pin vægthashes, registrer bygningen, nulstil prefix-caching, og revalider efter enhver ændring af præcision, motor eller skabelon.

Præcision	≥70B dense eller ≥100B MoE	7–35B	≤4B eller detaljetunge opgaver	Dom
FP8 / INT8	Sikker (≤1 point)	Sikker	Sikker; standarden for inspektion	Brug

Præcision	≥70B dense eller ≥100B MoE	7–35B	≤4B eller detaljetunge opgaver	Dom
4-bit AWQ/GPTQ (leverandøropskrift, visionsencoder FP16)	0–1.5 point	1–3 point; naiv GPTQ på lille-aktiv MoE −5.8 i gennemsnit, −19 på InfoVQA	2–3 point; valider detaljeopgaver først	Betinget
NVFP4 (leverandøropskrift)	≈0.3–1 point under FP8 (leverandørdata)	Kvantiseringsbevidste builds nær tabsfrit på V*, OCRBench	Ingen offentlige data	Betinget (foretræk kvantiseringsbevidste builds)
GGUF Q4_K_M	Utestet	Høj varians på domæneopgaver	Ikke til lille tekst eller defekter	Ikke til detaljevurdering
GGUF Q5_K_M og derover	Formodet sikker	Nær tabsfrit i to test	Gulvet for detaljer på llama.cpp	Brug
≤3-bit (inklusive 1T-modeller presset til 1.8–2 bits på én maskine)	Ikke til beslutninger	Nej	Nej	Nej

Hvilken konfiguration til hvilket job

Konfiguration	Inline-inspektion	AI-gennemgang af falskudkald	Robotplanlægning og -styring	Q&A og analyse
Dedikerede detektorer eller anomalimodeller (YOLO, RF-DETR, PatchCore-klasse), INT8/FP16, enhver NPU- eller GPU-edgeboks	Ja: hoveddommeren, 2–12 ms pr. udsnit	Ja (klassifikator tunet på linjen)	Kun perception	—
2–4B visionsmodeller, NVFP4/AWQ/FP8 med FP16 visionsencoder, på Thor/T4000 eller DGX Spark	Betinget: gentjek AOI-flagrede udsnit ved 0.2–0.3 s hver	Betinget, under 1 s efter tuning	VLA-rygrad (2–3B VLA ved ≈40–130 ms pr. chunk)	Svag
7–9B visionsmodeller (Qwen3-VL-8B, MiniCPM-V 4.5, GLM-4.6V-Flash), FP8 eller leverandør 4-bit, på Thor, Spark eller et 24–32 GB-kort	Betinget: ≈0.5 s pr. flagret udsnit	Ja: 0.5–1 s pr. kald, bedste værdi	Planlægning og færdighedsvalg (1–3 s), ikke i styreløkken	Grundlæggende SOP-Q&A for én eller nogle få brugere
26–35B MoE (Gemma-4-26B-A4B, Qwen3.6-35B-A3B), NVFP4/FP8	Nej (for tung)	Ja, som forklarer eller til svære tilfælde	Planlægning	Ja (8 brugere på en Spark, ≈0.1 s første token)
106–241B MoE ved 4-bit på et Mac Studio, RTX PRO 6000 eller Spark	Nej	Nej	Offlineplanlægning	Ja, for én eller nogle få brugere
DeepSeek 671B eller Kimi K2.x 1T ved 4-bit på en 512 GB Mac, KTransformers-server eller DGX Station	Nej	Nej	Kun offline programgenerering	Betinget: 1–3 brugere og korte prompts; lange prompts tager 16–170 s

Konfiguration	Inline-inspektion	AI-gennemgang af falskudkald	Robotplanlægning og -styring	Q&A og analyse
Alt ved ≤3-bit; GGUF Q4_K_M til småmodel-detaljearbejde; Ollama til mange brugere	● Nej	● Nej	● Nej	● Nej

Kilder: TensorRT Edge-LLM-, Jetson AI Lab-, Apple ML Research- og KTransformers-data ovenfor, og (DGX Spark-test, RTX PRO 6000-test, Hardware Corner om Atlas, Ultralytics YOLO26, RF-DETR) **fællesskabs- og leverandørtest, for det meste single source**.

Så til de to sædvanlige bekymringer: **det behøver ikke at være langsomt**, hvis modellen matcher jobbet (små og mellemstore visionsmodeller på Thor eller Spark opfylder ethvert inline- og gennemgangsbudget; 671B–1T-modeller holdes væk fra linjen), og **kvantisering behøver ikke at gøre det forkert** med FP8/INT8 som standard, eller leverandøropskrift 4-bit over 7B valideret på det gyldne sæt. De større risici er den omtrent 57-point zero-shot industrielle nøjagtighed for generelle modeller og serving-stack-fejl. Giv inspektionsforespørgsler faste batchstørrelser og deres egen kø (temperatur 0 er ikke bit-reproducerbart under batching (vLLM-dokumentation)), behold en N+1 varm standby, og få hvert trin på linjen til at fejle åbent.

Lærer og elev

03 Lærer og elev: frontmodeller kan præ-annotere inspektionsdata; for robotarme er de virkelige lærere stadig mennesker

November 2026-lærerkandidater: generisk syn fordobledes, industrielle defektscores mangler

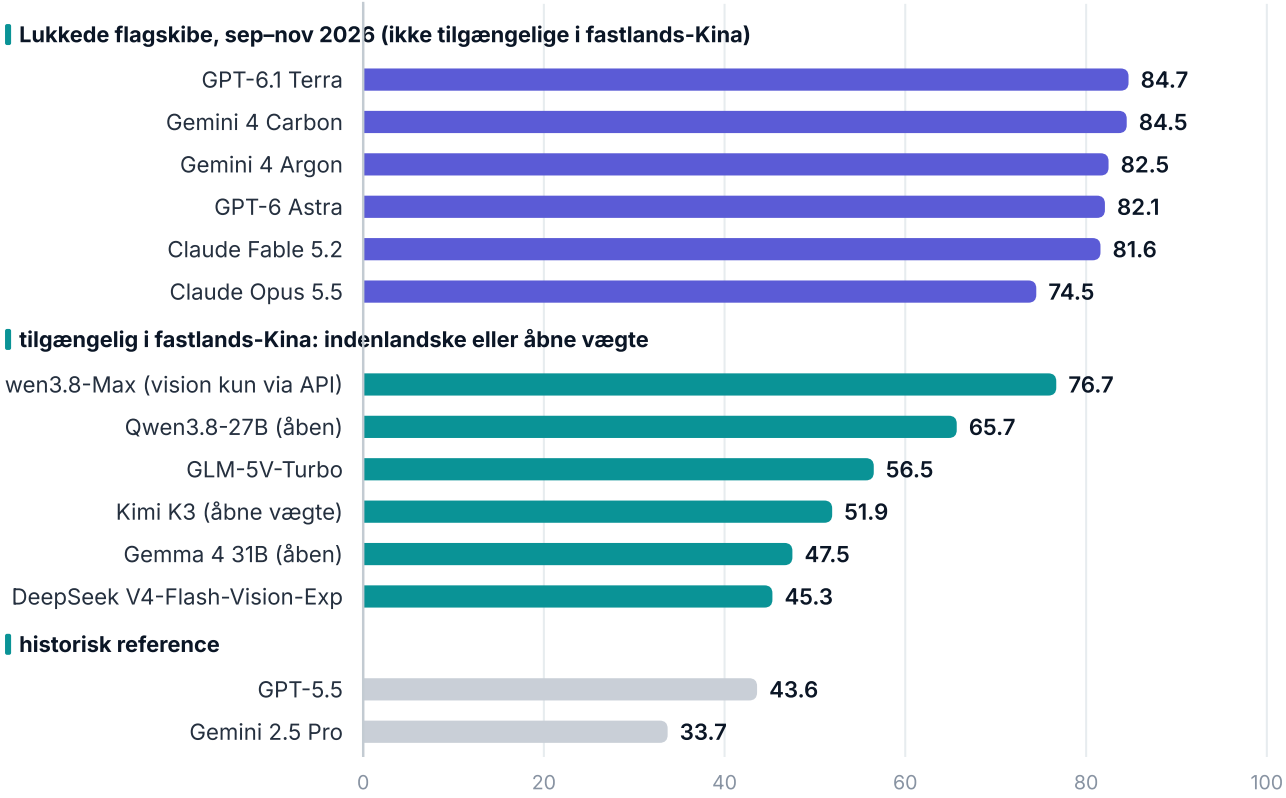
Hvert lukket flagskib, der kunne virke som lærer i november 2026, tager imod billeder: Claude Fable 5.2 (generelt tilgængelig 9. november, $8/$40 pr. mio. tokens) og Opus 5.5 ($4/$20) (Anthropic, Claude-modeloversigt) verified; GPT-6 Astra ($10/$50) og GPT-6.1 Terra ($1/$5) (OpenAI-priser); og Gemini 4 Argon og Carbon, åbnet i etaper (Google Carbon). Brugbare i fastlands-Kina er Qwen3.8-Max (vision via API) og åbne modeller som Qwen3.8-27B, Kimi K3, GLM-5V-Turbo, DeepSeek V4-Flash-Vision-Exp og Gemma 4 (Roboflow-blog, Hugging Face).

Det centrale hul skal siges rent ud: ingen af disse modeller har en publiceret score på MMAD, MVTec AD 2, VisA, Real-IAD, UniPCB eller noget PCB/SMT-datasæt. MMAD har stadig ingen leaderboard (MMAD GitHub), og de to nyeste industrielle anomalipapers (august–september 2026) bruger stadig Gemini 2.5 Pro og Claude Sonnet 4 som deres nyeste generelle baselines (InspectorGPT, Anomaly-LR). En arXiv-fulltekstsøgning den 11. november fandt nul hits for "Fable 5.2" eller "Gemini 4 Carbon" i denne kontekst (arXiv API) single source, fravær af evidens. De nærmeste substitutter:

- Den nyeste frontmodel med en MMAD-score er GPT-5.5: 69.95% på "defekt eller ej" og 76.20% gennemsnit over syv opgaver på et 500-billede-deludsnit, mod 95.24% og 86.65% for menneskelige eksperter (AnomalyClaw) single source; GPT-4o scorede 68.63% og 74.92% i 2024 (MMAD, ICLR 2025), omkring 1.3 points fremskridt på to år.
- På LDU-Bench, ægte SEM-gennemgangsbilleder fra IC-litografi og den eneste elektronikbenchmark med 2026-modeller, er god-eller-dårlig-triage hæderlig (makro F1 0.932 for GPT-5.4, 0.884 for GLM-5V-Turbo, 0.876 for Claude Opus 4.6, 0.504 for Gemma 4 31B), men klassificering af 11 defektmorfologier når kun 0.20–0.41 (LDU-Bench) single source.
- På Roboflows uafhængige vision-evals (opdateret 10. november 2026) er objektdetektions-mAP@50 84.5 for Gemini 4 Carbon, 84.7 for GPT-6.1 Terra og 81.6 for Fable 5.2, mod 33.7, 43.6 og 38.6 for Gemini 2.5 Pro, GPT-5.5 og Claude Opus 4.8 (Roboflow Vision Evals) single source, uafhængig evaluator, private datasæt.

FIG 11 **Lærerkandidater: generiske scores fordobledes på en generation*Men ingen af 2026-flagskibene har en publiceret score på en industriel defektbenchmark.***

Roboflow generisk detektion mAP@50



industrielle defektbenchmarks (MMAD, MVTec AD 2, VisA, Real-IAD, UniPCB): nul publicerede scores for disse 2026-flagskibe
generiske detektionsscores fordobledes på en generation, men det er ingen defektvurdering; test på dit eget gyldne sæt

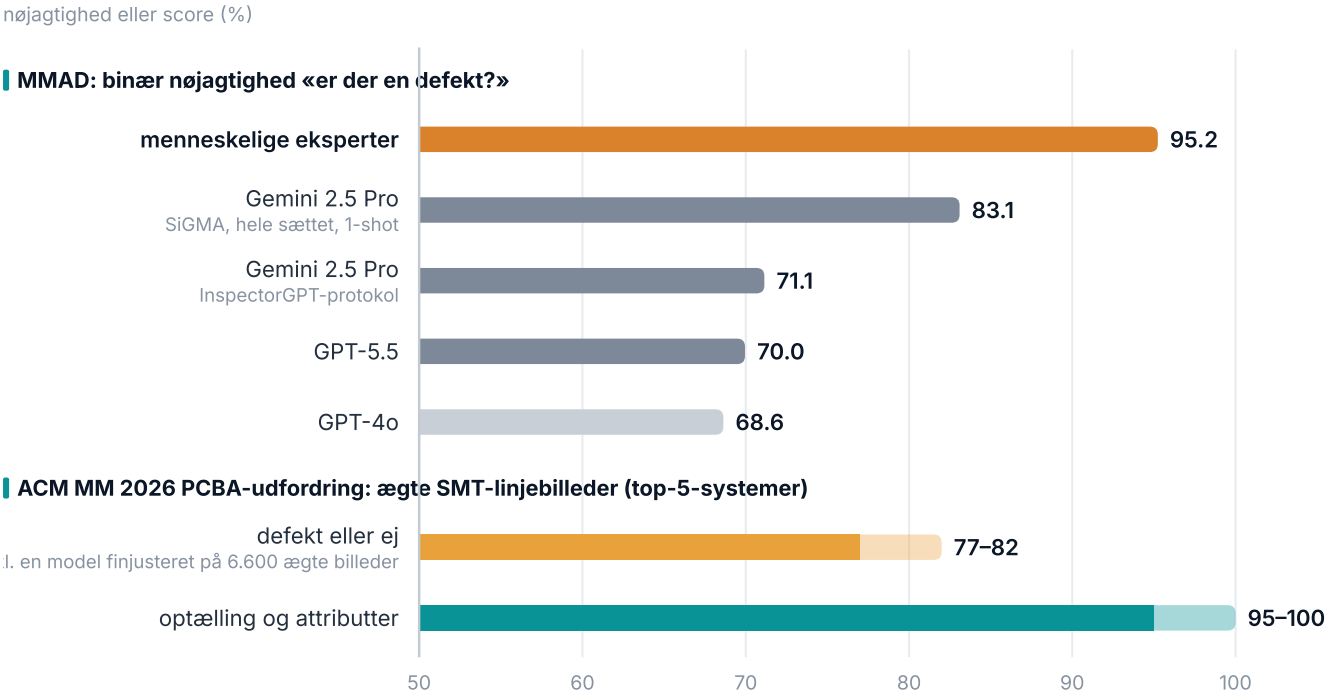
Kilde: Roboflow Vision Evals; modeludgivelsesnoter; MMAD, MVTec AD 2 og andre benchmarksider (pr. 11. nov 2026).

Lærerkandidat (nov 2026)	Pris pr. M tokens (ind/ud)	Roboflow-detektion mAP@50 / optælling	Industrielle defektbenchmarks	Brugbar i fastlands-Kina
Claude Fable 5.2	$8 / $40	81.6 / 86.5	Ingen	Nej
Claude Opus 5.5	$4 / $20	74.5 / 80.6	Ingen	Nej
GPT-6 Astra	$10 / $50	82.1 / 80.2	Ingen	Nej
GPT-6.1 Terra	$1 / $5	84.7 / 80.2	Ingen	Nej
Gemini 4 Carbon	$2 / $10 (kampagne)	84.5 / 90.1	Ingen	Nej
Qwen3.8-Max (vision kun via API)	$2 / $6	76.7 / 81.1	Ingen	Ja
Qwen3.8-27B (åben)	Lokal	65.7 / 64.9	Ingen	Ja
GLM-5V-Turbo	$1.20 / $4.00	56.5 / 48.6	LDU-Bench-triage 0.884	Ja
Gemma 4 31B (åben)	Lokal	47.5 / 51.4	LDU-Bench-triage 0.504	Ja (lokale vægte)
Reference: GPT-5.5 / GPT-4o	—	43.6 / 68.0 (GPT-5.5)	MMAD binær 69.95% / 68.63%	Nej

Lærerkandidat (nov 2026)	Pris pr. M tokens (ind/ud)	Roboflow-detektion mAP@50 / optælling	Industrielle defektbenchmarks	Brugbar i fastlands-Kina
Menneskelige eksperter	—	—	MMAD binær 95.24%	—

Nyere papers tester stadig den foregående generation, men peger samme vej. På det fulde MMAD-sæt scorer menneskelige eksperter 95.2% på defekt-eller-ej, og Gemini 2.5 Pro 83.1% (SiGMA), men den samme model scorede kun 71.14% i et andet paper (InspectorGPT): testprotokollen flytter scoren lige så meget som en modelgeneration gør. I CVPR 2026 VAND 4.0-udfordringen toppede hylde-vision-language-modeller ved 61.88 AP mod 90.84 for superviserede metoder, og arrangørerne konkluderede, at de "ikke i øjeblikket kan erstatte finetunede klassifikatorer" (VAND 4.0) single source. I ACM MM 2026 PCBA-udfordringen scorede de fem bedste 76.92–82.47 på defekt-eller-ej, mens komponenttype og pin-optælling nåede 95–99.5 (arXiv 2609.21276) single source: at genkende er meget lettere end at vurdere.

FIG 12 **«Er der en defekt?» er stadig svært for generelle modeller** ***Den bedste generelle model ligger mere end 12 point bag menneskelige eksperter på MMAD; tunede systemer gør det bedre på ægte linjebilleder.***



Kilde: MMAD; SiGMA; InspectorGPT; ACM MM 2026 PCBA-udfordring.

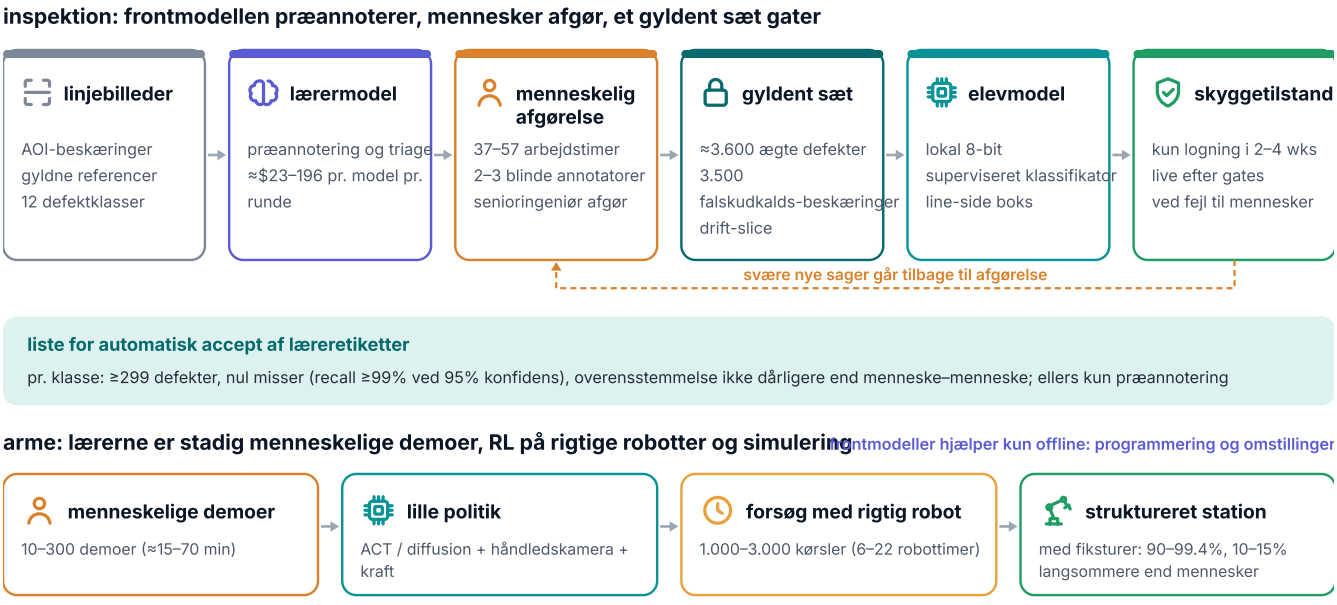
Tre konklusioner følger. At se og tælle fordobledes på en generation, men at vurdere abnormitet bevæger sig langsomt, afhænger kraftigt af protokollen og kommer med systematiske skævheder; hvor meget de nye modeller vinder, kan kun måles på en fabriks egne billeder. Inden for deres træningsfordeling slår finetunede 7B-specialister stadig generelle modeller (Anomaly-LR, UniPCB). Og ægte linjedata får scorer til at kollapse: på samlede elektroniske paneler faldt billedniveau-AUROC fra 97–98 på MVTec AD til 31.5–78.6 (arXiv 2606.01992) single source. For fastlandsvirksomheder ligger den stærkeste tilgængelige lærer, Qwen3.8-Max, omkring 8 point under den globale front.

At køre snesevis eller hundredvis af runder: penge er ikke problemet, annoteringskvalitet er

Penge er ikke problemet. Ved november 2026-listepriser koster 1,000 én-megapixel-billeder omkring $3 (GPT-6.1 Terra) til $32 (GPT-6 Astra) pr. runde, og hundrede runder over 100,000 billeder omkring $32,000–135,000 før rabatter på volumen estimate (Claude vision-dokumentation, OpenAI vision-guide, Roboflow Vision Evals). For simple annoteringer koster en person omtrent det samme som Terra; lærerens fordel er hastighed og skala, ikke pris. Kundernes NDA'er holder normalt produktbilleder inde i fabrikken, så det gyldne sæt annoteres på stedet.

Den virkelige hindring er kvalitet, og flere runder løser ikke systematiske fejl. Roboflow kører hver opgave tre gange, og scorer flytter sig kun ±0.1–1.4 point (Roboflow), så en model gentager for det meste sine fejl, og flertalsafstemning kan ikke redde den. Runder hjælper kun, når hver tilføjer information: ét eller to gode referencebilleder tilføjer omkring 3–4 point, flere tilføjer intet (AD-Copilot); at lade en model argumentere mod en god reference over flere ture tilføjer 6.23 point cross-domain AUROC for GPT-5.5 og 2.95 på MMAD, toppende efter omkring tre kald (AnomalyClaw) single sources . "En elev trænet på lærerannoteringer nærmer sig menneskelige annoteringer" er kun vist på naturlige billeder: en ResNet-101 trænet på Qwen3-VL-annoteringer af COCO og Objects365 nåede 95.7–99.4% af resultatet med menneskelige annoteringer, og forfatterne siger, at domænedata er utestede (arXiv 2602.20972) single source .

FIG 13 **To lærer–elev-pipelines** ***For inspektion præannoterer en frontmodel og mennesker afgør; for robotarme er de virkelige lærere stadig menneskelige demonstrationer.***



Kilde: Forfatterens syntese (kapitel 3); omkostninger og stikprøvestørrelser er estimater.

En brugbar protokol estimate :

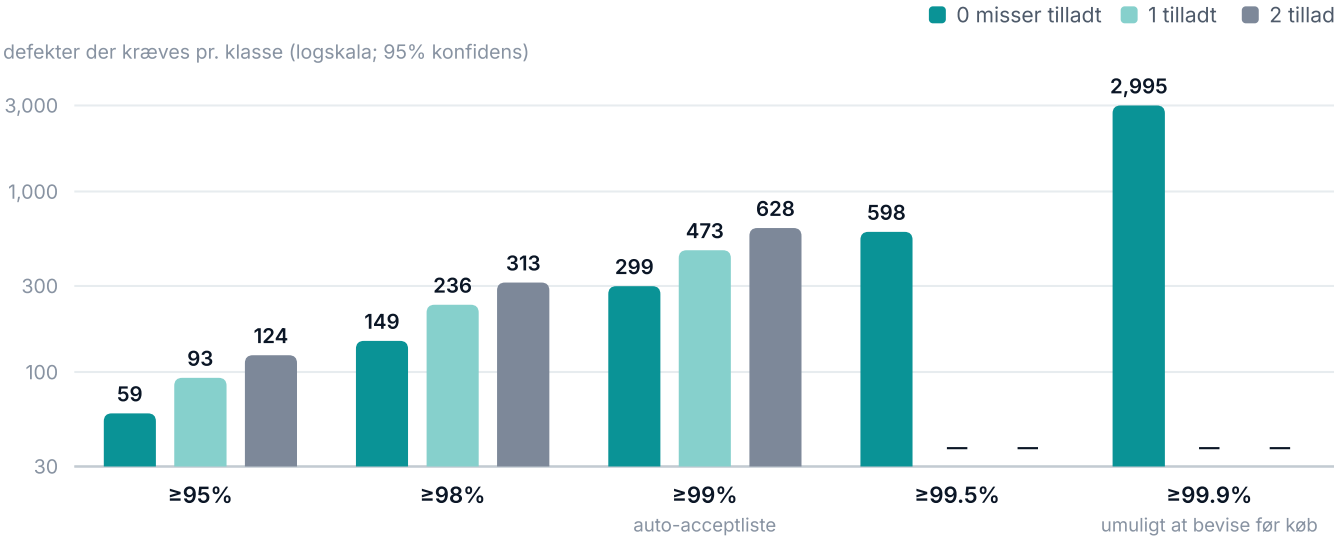
1. **Indsaml det gyldne sæt fra fabrikkens eget AOI-arkiv og gemte NG-kort**, stratificeret i 12 ægte defektklasser (manglende, offset, tombstone, polaritet, forkert del, bro/kortslutning, utilstrækkelig lodning, overskydende lodning, loddekugle, løftet ben/åben, beskadiget del, fremmedmateriale) og fire falskudkalds-årsager (skygge, flusmiddel, serigrafi, tolerance), plus en "drift-slice" på mindst 300 defekter fra en anden linje eller anden belysning, kortfarve eller artikelnumre.
2. **Annotér blindt med to eller tre IPC-A-610-trænede inspektører**, med en senioringeniør, der afgør, og røntgen, elektrisk test eller tværsnit, der afgør tvetydige prøver. Acceptér kun en klassedefinition, hvis inspektørenigheden når κ ≥0.75 (≥0.90 for kritiske klasser); menneskelig enighed om 11 elektroniske defektmorfologier i LDU-Bench var kun 0.727, så nogle klasser skal slås sammen eller omdefineres. AIAG-agtig

attributenighed (effektivitet ≥90%, miss ≤2%, falsk alarm ≤5%, κ >0.75; >0.90 for release) er et fornuftigt minimum (Fabrico, iFactory).

3. **Dimensionér det med eksakte binomiale (Clopper–Pearson) ensidede 95%-grænser** (binomiale konfidensintervaller, tre-reglen:

Recall der skal bevises	0 misser tilladt	1 tilladt	2 tilladt
≥95%	59 defekter	93	124
≥98%	149	236	313
≥99%	299	473	628
≥99.5%	598	—	—
≥99.9%	2,995	—	—

FIG 14 **Hvor stort det gyldne sæt skal være*At bevise 99% recall med nul misser kræver 299 defekter pr. klasse; 99.9% kan ikke bevises før køb.***



Kilde: Eksakte binomiale konfidensintervaller (95%).

Tres defekter pr. klasse er nok til en bestå/fejl-screening; 300 er nødvendige for at beslutte automatisk accept. Det fulde sæt, omkring 3,600 defekter plus 3,500 falskudkalds-udsnit, tager omkring 37–57 arbejdstimer at annotere og afgøre, og at køre en lærerkandidat over det koster $23–196; den reelle omkostning er to til seks ugers indsamling, sat af den sjældneste klasse estimate. Send enkeltkomponent-AOI-udsnit, ikke hele kort, med målet mindst 42–48 pixels på tværs (arXiv 2610.09313).

1. **Acceptér lærerannoteringer automatisk kun for en klasse med nul misser på ≥299 gyldne defekter**, model–sandhed-enighed ikke dårligere end menneske–menneske-enighed, og mindst 50% af falskudkaldene ryddet. Kritiske klasser som kortslutninger, manglende dele, polaritet og forkerte dele kræver 99.5–99.9% (598–2,995 defekter); ellers går alt, som AI'en ikke kan rydde med høj konfidens, til en person. En escape-rate under 0.1% af defekterne kan ikke bevises før køb (2,995 defekter med nul misser; 0.01% ville kræve 29,956), så linjen beholder menneskelig stikprøvekontrol af AI-ryddede kald, downstream ICT/FCT som backstop og SPC på escapes. Verificerede annoteringer destilleres derefter til en konventionel detektor eller klassifikator og revalideres på hvert nyt produkt.

Vilkår og dataregler afgør også læreren. Anthropic forbyder at bruge sine tjenester til at træne konkurrerende modeller (Anthropic kommercielle vilkår); OpenAI tillader outputs til at træne interne klassifikatorer, der ikke sælges til tredjeparter (OpenAI-forretningsvilkår); DeepSeek tillader eksplicit destillation (DeepSeek-vilkår) [primære vilkår]. Ingen definerer en "konkurrerende model", så kommerciel brug kræver juridisk gennemgang, og kundernes dataregler holder normalt billeder på stedet. Det sikreste standardvalg er en lokalt kørt lærer med åbne vægte, med cloud-frontmodeller kun, når kunden tillader det skriftligt.

Inspektionselever: små og hurtige; flaskehalsene er optik, data og validering

Eleven er ikke flaskehalsen: YOLO26n tager 3.80 ms pr. billede i INT8 på en $249–399 Orin Nano Super (Ultralytics) [leverandørtest], NVIDIA oplyser 98.51% på tre PCB-defektklasser efter at have tilpasset en foundation-model på omkring 600 billeder (NVIDIA) vendor claim, og syntetiske defekter tilføjer kun 0.9–2.5 mAP-point (arXiv 2605.04635). Begrænsningerne er optik, linjedata, validering og licenser: Ultralytics YOLO er AGPL-3.0 eller enterprise-licenseret (Ultralytics), så lukkede produkter kræver en licens eller en Apache-2.0-detektor som RF-DETR.

Robotelever: strukturerede stationer er tæt på produktionsgrad; fleksibelt og dynamisk arbejde er det ikke

I november 2026 findes der tre slags robotlærere, og kun én kan blive en lokal elev: åbne, finetunbare VLA'er (π0/π0.5, GR00T N1.7, X-VLA, RDT2, SmolVLA). Lukkede robotmodeller (π0.6/π0.7 gennem en beta-API, Gemini Robotics 2 for betroede testere) kan ikke implementeres på stedet (The AI Insider, DeepMind). Frontgenerelle modeller er langsomme og dårlige til finmanipulation: i Artificial Analysis' test scorede de 46–69 med 28–45% grebssucces og omkring 30 minutter pr. opgave, "alt for langsomt og dyrt til nogen produktionsbrug" (AA-Robotics); GPT-6 Astra klarede en præcisionsslot kun 2 gange ud af 20 (RoboCurve); og at drive led direkte giver 0–5.5% på LIBERO ved 0.2–0.4 Hz mod omkring 83 Hz nødvendigt (Anthropic) verified på tværs af tre kilder. Fable 5.2's 64.9% på den simulerede RoboDojo-benchmark er et internt run (Anthropic) vendor claim.

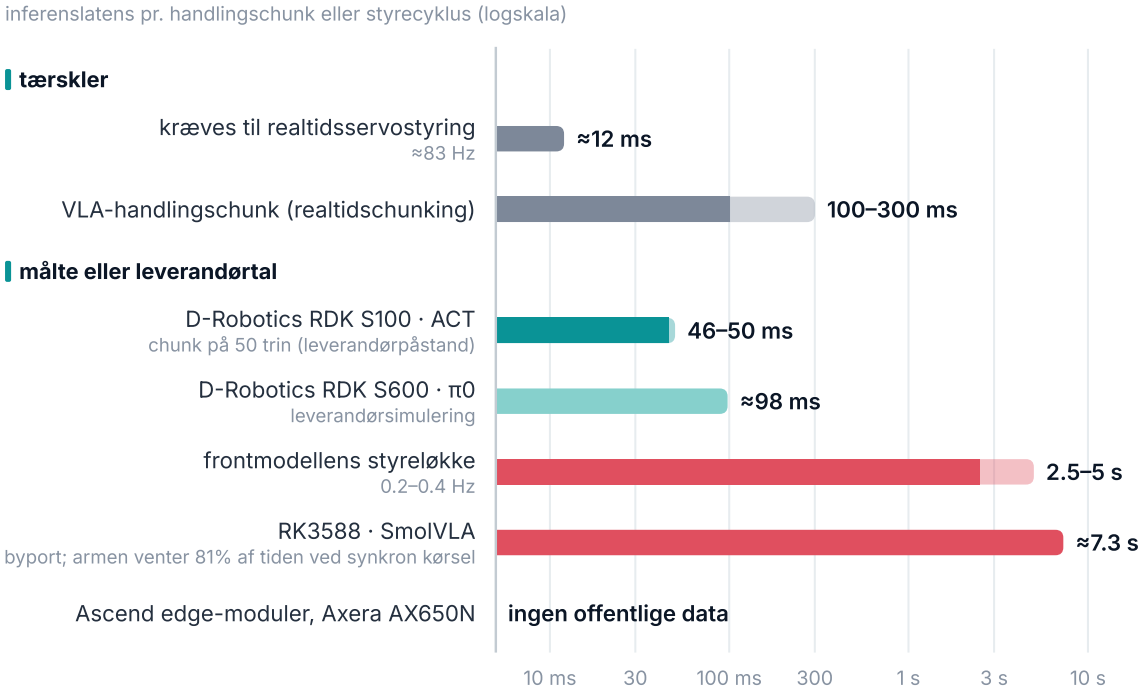
Tredjepartsstudier siden oktober er enige: omkring 33–44% på LIBERO-10 for de tre førende agenter (arXiv 2610.30051); agenter, der fejlagtigt hævder succes på 24–39% af delopgaverne (arXiv 2611.08209); og 104 af 108 kvasistatiske ægte robotopgaver for Opus 5.5, men ved 280 sekunder hver og uden indsætning eller montage (AMP1) single sources. Frontmodeller hører til offline: at skrive programmer, orkestrere opgaver og udarbejde succeskriterier.

Elevkvalitet har en klar kant: god inden for træningsfordelingen, kollapsende uden for den. π0.7 "overstiger ofte 90%" på sete opgaver og når 60–80% på usete (π0.7-paper) [leverandørforskning]; ContactDP når 97.8% på trænede stik, men 28.6–66.7% på usete (ContactDP); små kamera- eller poseændringer får 95% til at falde under 30% i LIBERO-Plus (LIBERO-Plus, CVPR 2026); indlært gribning fra bevægelige transportbånd topper ved 58.89% (ForeTime-VLA) single sources, et job klassisk visionssporing har udført i produktion i årevis.

Den nær-produktionsklare evidens kommer alle fra strukturerede stationer, menneskelige demonstrationer og ægte robotkorrektion. Neuromeka beholdt en tilstandsmaskine og præindlært bevægelse og lærte kun gribning, kabelindsætning og lodning, hver fra under 20 minutters data: 322 af 324 i produktion, 12.8% langsommere end en person (Neuromeka) single source, leverandør. Fraunhofer nåede 92.8% på tværs af fire stik fra 68 minutters demonstrationer (Fraunhofer); Berkeleys HIL-SERL lykkedes 100 gange ud af 100 på RAM-, SSD- og USB-indsætning efter 1–2.5 timers ægte robot-RL (HIL-SERL); AgiBot kørte 18–20 sekunder pr. trin inden for en 20-sekunders takt på en tablettestlinje (IThome) vendor claim.

Destillation køber hastighed, ikke generalisering: π0.5 kører ved 0.70 Hz på Jetson Orin rå og 10.30 Hz efter asynkron planlægning (EagleVLA); lagdestillation tager den fra omkring 3 Hz til omkring 10 Hz (Shallow-π) verified. En elev arver kun det, læreren dækkede; hvert nyt produkt kræver stadig opgavedata.

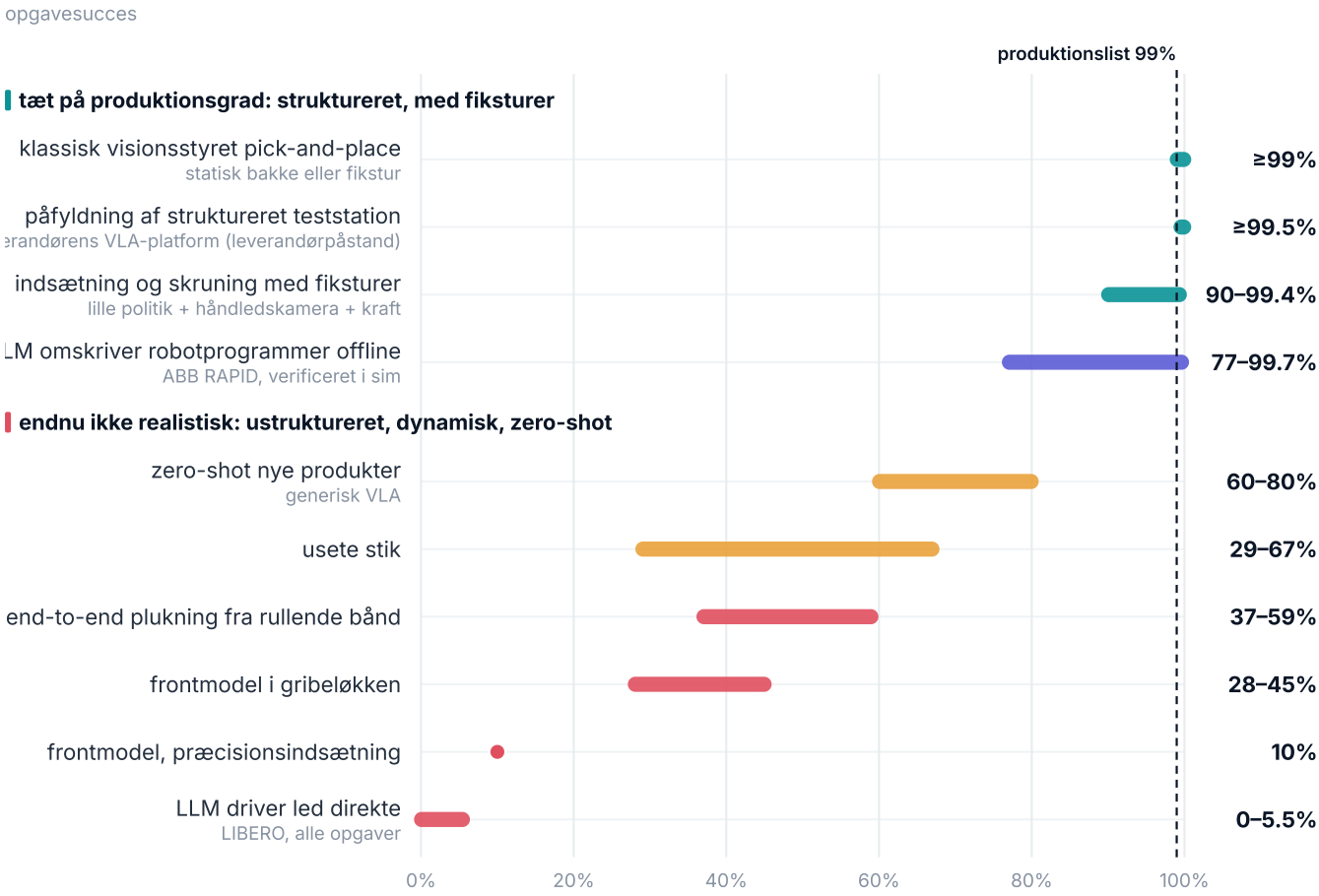
FIG 15 **Edge-politiklatens***En lille ACT-politik kører på cirka 50 ms på et edge-kort; en VLA porteret til en hobbychip tager 7.3 s.*



Kilde: D-Robotics dokumentation og forum (leverandørtal); community-porte på GitHub; Anthropic-forskning.

På kinesiske edge-chips er den eneste målte indlærte politik en ACT-politik på D-Robotics' RDK S100 ved 46–50 ms pr. 50-trins chunk, uden publiceret succesrate (D-Robotics RDK S-dokumentation) vendor claim; en hobbyport af SmolVLA på en RK3588 tager omkring 7.3 s pr. chunk (GitHub). Kvantisering køber ikke hastighed gratis: ingen 4-bit π0.5-build var hurtigere end 16-bit ved batchstørrelse 1 (LoopQ). Demonstrationer er dyre (Reuters oplyste omkring 300 forsøg pr. brugbar bevægelse for en novice-teleoperatør (Notebookcheck)), og validering koster mere: at bevise en fejlrate ≤0.1% ved 95% konfidens kræver omkring 3,000 på hinanden følgende succeser, 17–22 robottimer estimate. Ingen indlært manipulationspolitik har sikkerhedscertificering.

FIG 16 **Robotelever: tæt på produktionsgrad kun når stationen er struktureret** ***Stationer med fiksturer når 90–99%+; ustrukturerede, dynamiske og zero-shot-opgaver er endnu ikke realistiske.***



Kilde: Artikler og leverandørmateriale citeret i kapitel 3; nogle tal er leverandørpåstande.

Elektronikopgave	Brugbar elev	Forventet succes	Hastighed vs. en person	Omkostning pr. omstilling	Dom 2026–2028
High-mix pick-and-place, kitning, maskinbetjening (statisk bakke eller fikstur)	Indlært perception plus programmeret eller planlagt bevægelse	≥99% med klassisk visionsstyring	Samme som konventionelle stationer	Timer	Gennemførlig nu
Fikseret stik-, FFC- og kabelindsætning, skruning, lodning	Tilstandsmaskine + lille BC/diffusion/ACT-politik + håndledskamera + kraft/moment (ideelt berøring) + menneskelig korrektion	90–99.4%	10–15% langsommere	10–300 demoer (≈15–70 min) + 1,000–3,000 valideringskørsler (6–22 robottimer)	Gennemførlig, R&D-tung måneder for den første station
Struktureret teststationsbetjening og OK/NG-sortering	Leverandør VLA-platform + digital tvilling + ægte robot-RL	≥99.5% vendor claim, modstridende	18–20 s pr. trin, inden for en 20 s takt	≤4 t omtræning + 15 min kalibrering	Gennemførlig på en leverandørplatform

Elektronikopgave	Brugbar elev	Forventet succes	Hastighed vs. en person	Omkostning pr. omstilling	Dom 2026–2028
Omstillingsprogrammering	LLM skriver eller redigerer robotprogrammer offline, verificeret i simulering	77–99.7% på ABB RAPID-omskrivninger	Ingen effekt på takt	Minutter til timer pr. program	**Gennemførlig som ingeniørværktøj**
End-to-end indlært plukning fra bevægelige transportbånd	VLA-elev	37–59%; ≤37% på hurtige bånd	—	≥458 demoer pr. opsætning	**Ikke før 2028; brug klassisk sporing plus indlært perception**
Nye produkter uden nye data (zero-shot)	Generisk VLA	60–80%; usete stik 29–67%	—	—	**Ikke realistisk**
Front-LLM i løkken, der driver armen	GPT-6-, Claude-, Gemini-agenter	Greb 28–45%; præcisionsindsætning 10%	≈30 min eller 280 s pr. opgave vs. ≈20 s	—	**Ikke realistisk**
Frit rutede ledningsseler og flertrins fleksibel montage	—	Ingen 2026-produktionsevidens	—	—	**Ikke før 2028**

Kilder: studierne ovenfor og (University of Passau om RAPID-omskrivning, GraspVLA, Komax) **for det meste single source og leverandørdata**. I 2026–2028 er en stor models værdi for robotter som en fortrænet prior og en offlineprogrammør, ikke som noget, der "underviser" en elev. Hvert nær-produktions-tilfælde, jeg fandt (Neuromeka, AgiBot hos Longcheer, Fraunhofer, ContactDP), bygger på menneskelige demonstrationer, kraft- eller berøringssensorer, simulering eller digitale tvillinger og superviseret RL. Ingen bygger på annoteringer fra en frontmodel.

Omkostninger og tilbagebetaling

04 Omkostninger og tilbagebetaling: AI-gennemgang tjener sig hjem på måneder, maskinbetjening på år, udplukningsarme aldrig

Arbejdskraftomkostning: USA er omkring 5.5 gange Shenzhen, Taiwan omkring 1.5 gange, det meste af Sydøstasien under halvdelen

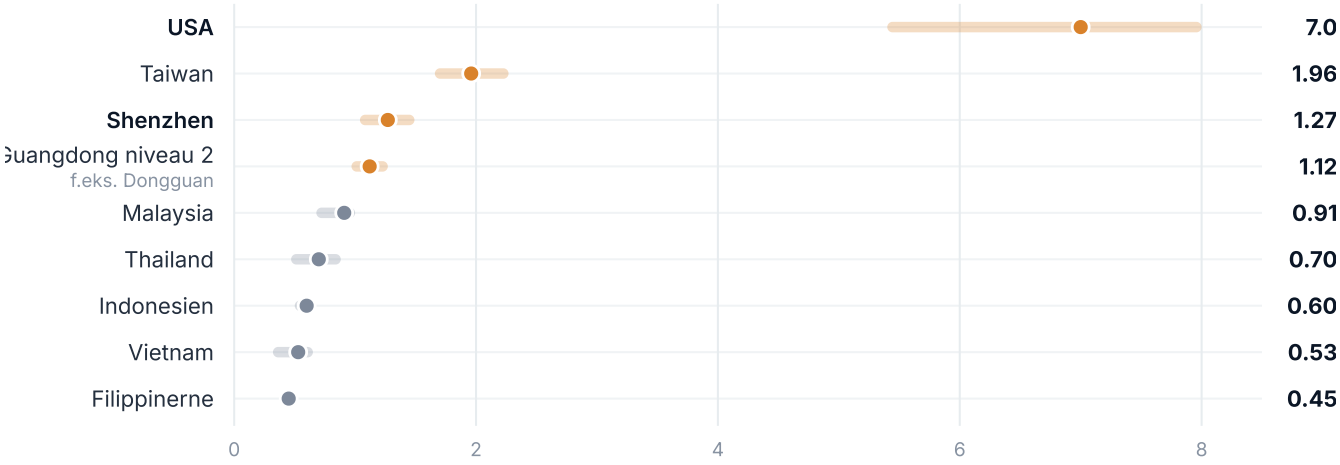
Afkastet ved at erstatte arbejdskraft starter med, hvad arbejdskraften koster. Shenzhens officielle lønvejledning sætter 2024-medianen for en kvalitetsinspektør i fremstilling til **RMB 74,074** om året; løn udgør 86.92% af de samlede arbejdskraftomkostninger i elektronikfremstilling, så den fuldt belastede omkostning er omkring RMB 85,000. JETRO's 2025-undersøgelse sætter den årlige omkostning for en kinesisk fabriksarbejder til $12,667, omkring RMB 85,100, inden for 0.2% (Shenzhen lønvejledning, JETRO 2025) verified .

Region	Fuldt belastet årlig omkostning for én inspektør eller håndteringsarbejder (lav / base / høj)	Grundlag
Fastlands-Kina, Shenzhen	RMB 70k / **85k** / 100k ($10.4k / **12.7k** / 14.9k)	Officiel median × 1.15; matcher JETRO verified
Guangdong tier-2-lønområder (Dongguan m.fl.)	RMB 65k / **75k** / 85k	Guangdong 2023-median × 1.15 plus to års stigninger estimate
Yangtze River Delta (Suzhou, Kunshan, Wuxi)	Som Shenzhen	Intet grundlag for billigere inspektionsarbejde; Jiangsus højeste mindsteløn er omkring 5.6% over Shenzhens estimate
Taiwan	$16.6k / **19.6k** / 22.7k (NT$534k / **630k** / 729k)	Arbejdsministeriets erhvervslønundersøgelse plus ≈1.18 lovpligtige tillæg; JETRO som den høje ende estimate
Malaysia	$6.8k / **9.1k** / 10.0k	JETRO og mindstelønninger estimate
Thailand	$4.7k / **7.0k** / 8.8k	Samme
Indonesien	$5.0k / **6.0k** / 6.3k	Samme
Vietnam	$3.2k / **5.3k** / 6.5k	Samme
Filippinerne	$4.0k / **4.5k** / 4.9k	Samme
USA	$54k / **70k** / 80k	BLS maj 2025-inspektørmedian $48,570 ÷ løn på 65.5% af kompensation estimate

Kilder: (JETRO 2025, Guangdong lønvejledning, Wuxi HR-bureau, Taiwan erhvervslønundersøgelse, PTS News, BLS OEWS, BLS ECEC).

FIG 17 **Hvad én inspektions- eller håndteringsarbejder koster om året*Fuldt belastet er USA cirka 5.5× Shenzhen og Taiwan cirka 1.5×.***

årlig fuld omkostning pr. QC- eller håndteringsarbejder ($10k; punkt = base, søjle = lav–høj)



Kilde: Shenzhens officielle lønvejledning, Taiwans arbejdsministerium, JETRO, BLS; forfatterens estimater.

Tre korrektioner sætter grundlaget. For det første hævede Shenzhen sit pensionsbidrags-gulv fra RMB 4,775 til 5,521 om måneden i juli 2026 (Shenzhen Bendibao) single source, så en Shenzhen-inspektør koster i 2026 mere sandsynligt RMB 80–90k. For det andet er en stilling ikke en person: to 12-timers skift kræver omkring 2.0 personer (seks dage om ugen) til 2.4 (hele året) i Pearl River Delta, omkring 2.7–3.1 i Sydøstasien under overtidslofter, og 4.2–4.5 for 24×7 i USA og Taiwan estimate. AI-gennemgang lader normalt én gennemgangsperson dække to til fire linjer, så den realistiske besparelse er **0.5–1.5 fuldtidspersoner pr. linje, ikke én person pr. linje pr. skift**. For det tredje jagter købere i Taiwan, Vietnam og Malaysia ikke kun omkostninger: i JETRO's undersøgelse sagde 61.1% af virksomhederne i Taiwan, 48.2% i Vietnam og 44.9% i Malaysia, at rekruttering var blevet sværere de seneste to år, mod 18.6% i fastlands-Kina (JETRO 2025) single source.

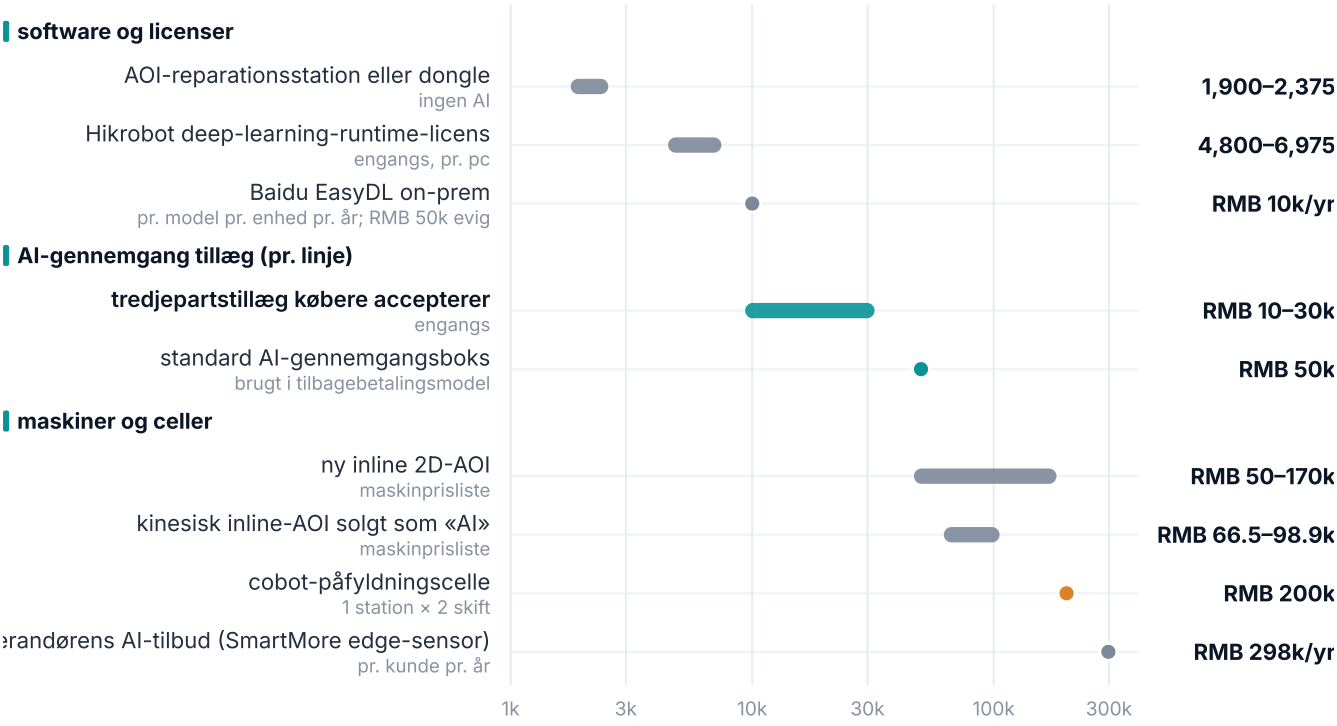
Udstyrspriser: fire grundlag, der ikke bør gennemsnittes

AOI-priser er uenige, hovedsageligt fordi de måler forskellige ting. Onlinenoteringer (november 2026, nyt, brugt og udlejning blandet) ligger omkring RMB 17–80k for offline eller desktop 2D, RMB 50–170k for inline 2D og RMB 80–475k for inline 3D, op til RMB 700k for koreansk 3D (1688 inline AOI, 1688 offline AOI, 1688 3D AOI) single source, liste-priser. Et november 2025-udbud satte et loft på én 3D-AOI plus én 3D SPI til RMB 800k (udbudsmeddelelse). Sinic-Teks prospekt antyder RMB 428–512k pr. 3D-AOI i 2019–2022, omkring RMB 340–470k ifølge dets 2025-rapport (Sinic-Tek-prospekt, Sinic-Tek årsrapport 2025) primær + estimate. En oversøisk leverandør estimerer $3,000–20,000 for desktop 2D, $30–70k inline 2D, $50–90k entry 3D og $100–200k+ high-end 3D, med den samme inline-maskine til $30–50k i Kina (Averroes) [leverandørestimat fra en sælger af konkurrerende AI-software].

AI-gennemgang selv har ingen offentlig RMB-pris; leverandører opgiver case-for-case efter en gratis evaluering på stedet (Sohu om Hengtianyi), og søgninger i e-handelsnoteringer, udbud, leverandørartikler og taiwanske medier fandt ingen offentlig pris (1688, 51SMT). Ankerpunkterne: omkring €27k pr. linje antydet af Siemens-casestudiet estimate; $21,924 listepris for en USA-distributørs fjerngennemgangs-netværksmodul (Manncorp) [leverandørliste-pris]; omkring RMB 298k pr. kunde pr. år for SmartMores edge-AI-sensorer i 2025 (36Kr) single source. En indenlandsk inline 2D-AOI solgt som "AI" koster RMB 66.5k på listen (1688 inline AOI), og en NG-buffer til manuel genin-spektion omkring RMB 50k (1688 AOI-gennemgang). Så loftet for et tredjeparts gennemgangs-tillæg er omkring RMB 10–30k pr. linje, eller et tilsvarende årligt gebyr; over det ville køberen hellere udskifte maskinen med en indenlandsk AOI med AI indbygget estimate.

Ankerpunkt	Pris (RMB)	Hvad det køber	Kilde
AOI-reparationsstation eller software-dongle (ældre TRI, Zhenhuaxing)	≈1,900–2,375	Licensnøgle, ingen AI	1688 [listepris]
Generel deep-learning-vision-runtime (Hikrobot VisionMaster-dongle)	≈4,800–6,975 engangs pr. PC	Platformen; kunden bygger og vedligeholder modeller	1688 [listepris]
Baidu EasyDL on-premise-implementering	10,000 pr. model pr. enhed pr. år, eller 50,000 evig	Pakket inferenstjeneste	Baidu AI-dokumentation [2022-prisside]
Indenlandsk inline AOI solgt som "AI"	66,500–98,900 pr. maskine	Maskinudskiftning	1688 [listepris]
SmartMore edge-AI-sensorer	≈298,000 pr. kunde pr. år	Leverandørløsning på tværs af flere maskiner	36Kr single source

FIG 18 **Prisstigen***Fra nogle få tusinde RMB softwarelicens til RMB 200–300k for en robotcelle eller et AI-tilbud fra en leverandør.*



Kilde: 1688-opslag, Baidu AI-dokumentation (prisliste 2022), 36Kr, et brancheinterview (ikke verificeret).

Et Baidu-agtigt årligt gebyr, med én inferensserver til én til fire linjer og én til fire modeller hver, antyder et abonnement på RMB 2.5–40k pr. linje pr. år, realistisk omkring RMB 5–10k; ved 20% af en evig licens går abonnement og køb lige op i år fem, en struktur værd at kopiere estimate. Købere er vant til næsten nul softwaregebyrer: Sinic-Teks samlede "anden" omsætning (reservedele, service, udlejning, tilbehør) var 7.65% af 2025-salget, omkring RMB 2,300–3,100 pr. installeret maskine pr. år (Sinic-Tek årsrapport 2025) primær + estimate. Et AI-abonnement på RMB 5–10k pr. linje pr. år er to til fem gange, hvad købere nu bruger på AOI-software, så det skal retfærdiggøres af de gennemgangstimer, det frigør (omkring RMB 75–90k pr. fuldtidsperson om året), ikke ved sammenligning med eksisterende softwaregebyrer.

For robotarme faldt Kinas gennemsnitlige cobot-pris til omkring RMB 58,600 i 2025 (GGII) estimate, og Dobots seksaksede gennemsnit faldt 19.1% det år, da billigere modeller vandt andele (Huxiu) verified, selvom listepriserne for

de samme modeller steg 5–15% i 2026 (Sina Finance). I USA kostede en FANUC CRX-10iA $51,434 på listen (Vention), og opgørelser sætter en UR7e til $30–48k (cobotcost) estimate. Armen er kun 40–50% af en USA-celle (Standard Bots); i Kina løber kameraer, gribere, sikkerhed og ingeniørarbejde nemt op i to til fire gange en RMB 40k-arm, så budgettér absolut ingeniøromkostning frem for et multiplum estimate.

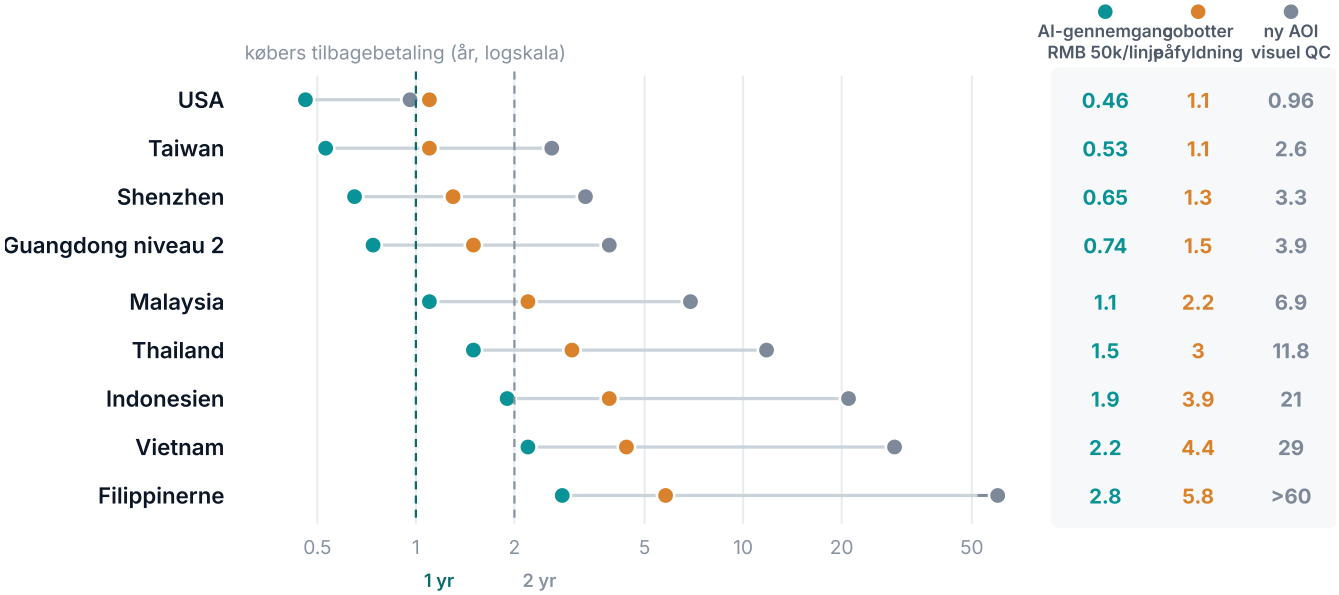
Tilbagebetaling: gennemgang på måneder, betjening på år, udplukning aldrig

Simpel tilbagebetaling = investering ÷ (sparede personer × belastet arbejdskraftomkostning − årlig driftsomkostning), for en to-skifts linje uden subsidier eller told. Taiwan-udstyr prissættes til 1.3 gange Kina, Vietnam, Malaysia og Thailand til 1.15 gange, Indonesien og Filippinerne til 1.2 gange, og USA til oversøiske priser estimate.

Scenarie	Investering (Kina / USA)	Sparrede personer	Årlig driftsomkostning
(a) AI-gennemgangsboks eller AI-kameraboks	RMB 50k / $30k pr. linje	1.0 pr. linje	15%
(b) Ny inline 2D-AOI, der erstatter manuelle synstjek (0.5 gennemgangsperson pr. skift forbliver)	RMB 200k / $60k	1.0	12%
(c) Cobot-påfyldningscelle (én station × to skift)	RMB 200k / $140k	2.0	8%
(d) Integreret detektér-og-pluk-celle (inspektion 1.0 + plukning 0.07)	RMB 300k / $200k	1.07	8%
(e) En robotarm kun tilføjet for at plukke NG-kort ud	RMB 150k / $100k	0.07	8%

Region	(a) AI-gennemgang	(b) Ny AOI	(c) Cobot-celle	(d) Detektér-og-pluk	(e) Udplukningsarm
Shenzhen	0.65 år (0.54–0.80)	3.3 (2.6–4.4)	**1.3** (1.1–1.6)	4.5 (3.6–5.9)	Aldrig
Guangdong tier-2	0.74	3.9	1.5	5.3	Aldrig
Taiwan	0.53	2.6	1.1	3.6	Aldrig
Malaysia	1.1	6.9	2.2	9.2	Aldrig
Thailand	1.5	11.8	3.0	15	Aldrig
Indonesien	1.9	21	3.9	25	Aldrig
Vietnam	2.2 (1.6–4.5)	29	4.4	34	Aldrig
Filippinerne	2.8	>60 (driftsomkostning ≈ besparelse)	5.8	≈100	Aldrig
USA	0.46	**0.96**	**1.1**	3.4	Aldrig

FIG 19 **Tilbagebetaling pr. region** ***AI-gennemgang betaler sig på måneder i USA, Taiwan og Shenzhen og på 1–3 år i Sydøstasien; cobot-påfyldning på år.***



en robotarm kun til at plukke NG-kort ud betaler sig aldrig, nogen steder
ved first-pass yield over 90% tager udplukning kun 0.003–0.14 FTE pr. skift

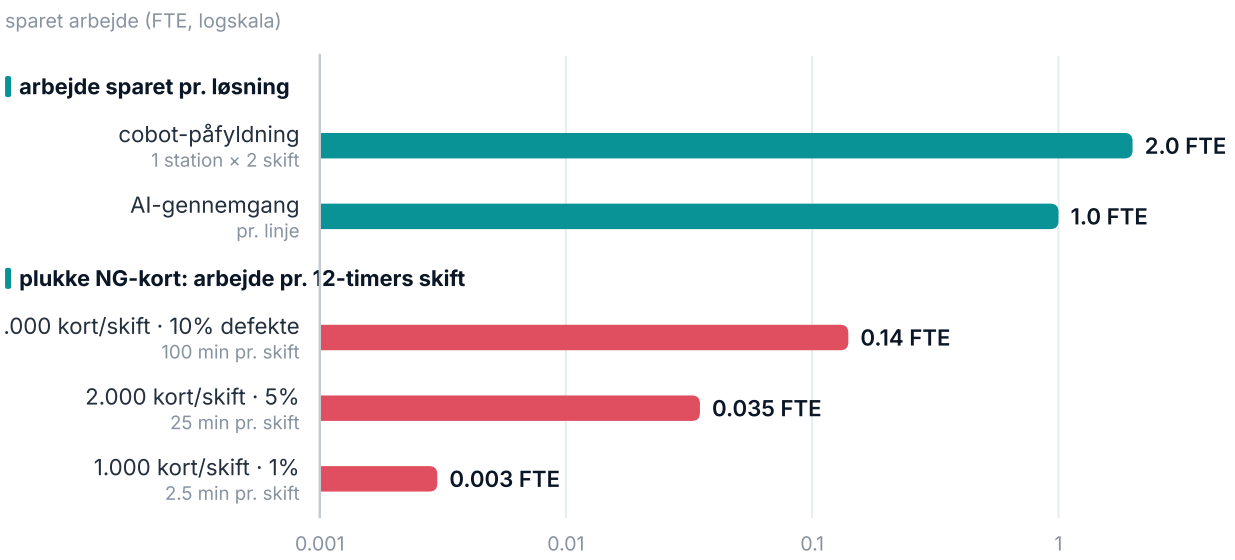
Kilde: Forfatterens model: RMB 50k AI-gennemgangsboks pr. linje; cobot-påfyldningscelle på RMB 200k eller $140k, to skift.

Alle værdier er estimater fra inputene ovenfor. Følsomheder: halvér de sparede personer, og enhver tilbagebetaling fordobles omtrent; kør ét skift, og (b)–(d) fordobles; 100% førsteårsafskrivning i USA forkorter tilbagebetalingen med omkring 20%, mens told på kinesisk, japansk eller koreansk udstyr forlænger den med omkring 10–40%; et 15% lokalt subsidie i Kina forkorter den med omkring 15%. Jeg fandt ingen uafhængig undersøgelse af køberes tilbagebetalingstærskler; leverandører siger, at AOI tjener sig hjem "typisk på 9–18 måneder" (Averroes), og Siemens-casestudiet hævder otte måneder vendor claims. Jeg bruger ét og to år som screeningslinjer. Modellen udelader kvalitetsgevinster (garanti, returneringer) og personaleomsætnings- og træningsomkostninger, så den er konservativ.

Den kontraintuitive del af "detektér og pluk" er, at arbejdskraften ligger i gennemgang, ikke i at tage de få ægte NG-kort af linjen. Ved 15 sekunders håndtering pr. NG-kort estimate:

Kort pr. skift	Defektrate	Plukketid pr. 12-timers skift	Personer
1,000	1%	2.5 min	0.003
2,000	5%	25 min	0.035
4,000	10%	100 min	0.14

FIG 20 **Hvorfor udplukning af dårlige kort ikke betaler sig** ***Over 90% first-pass yield tager udplukning kun 0.003–0.14 af en person pr. skift.***



Kilde: Forfatterens estimater.

Basis-casen (2,000 kort pr. skift, 5% defekte, to skift) er 0.07 af en person, omkring RMB 6,000 om året i Shenzhen, mindre end armens omtrent RMB 12,000 årlige driftsomkostning, så den tjener sig aldrig hjem nogetsteds; selv worst case (4,000 kort pr. skift, 10% defekte) tager omkring 13 år i Shenzhen og 9 i USA estimate. Hvis bekymringen er at stoppe linjen for at plukke, løser en standard NG-afleder eller -buffertransportør det uden en arm. Folk i branchen siger "over 90% first-pass yield, stop linjen og tag kortet ud i hånden", og **det er rigtigt for plukning, men ikke for inspektion**. Produktdesignet på en SMT-linje er derfor "inspicér plus markér eller afled", med robotarme gemt til steder, hvor hver del håndteres (testfikstur-påfyldning, bakke- og magasinhåndtering, pakning), hvor inspektion bliver et næsten gratis tillæg, og scenarie (c) gælder.

Subsidier, skat og told flytter tilbagebetalingen med 15–25%, men ikke rangeringen

Fastlands-Kina: støtte sænker for det meste finansieringsomkostningen; enkeltmaskiner kvalificerer sjældent direkte. Et nationalt lånesubsidie betaler 1.5 rentepoints i op til to år i 14 prioriterede kæder (Yicai) verified, omkring RMB 9,000 på en RMB 300k-maskine; Guangdongs opgraderingsbelønning på op til 20% kræver et registreret projekt med mindst RMB 7 mio. nyt udstyr og udelukker Shenzhen (Guangdong-regler), og Dongguan betalte faktisk 4.42% i juli 2026 (Dongguan-tildeling) single source, primær. En AI-inspektionsenhed godkendes kun som udstyr "med understøttende software" inde i køberens registrerede projekt, så basis-casen antager intet subsidie.

Taiwan: incitamenter går til køberen. Under artikel 10-1 i Statute for Industrial Innovation, ændret i maj 2025 for at omfatte AI, kan en køber kreditere 5% af 2025–2029-udgifter til nyt smart maskineri eller AI (inklusive CNN/DNN-baseret inspektion) mod det års indkomstskat, eller 3% om året i tre år (Statute, kreditregler) verified; leverandørsidetilskud er lukket for fastlandskapital (MOEA) single source.

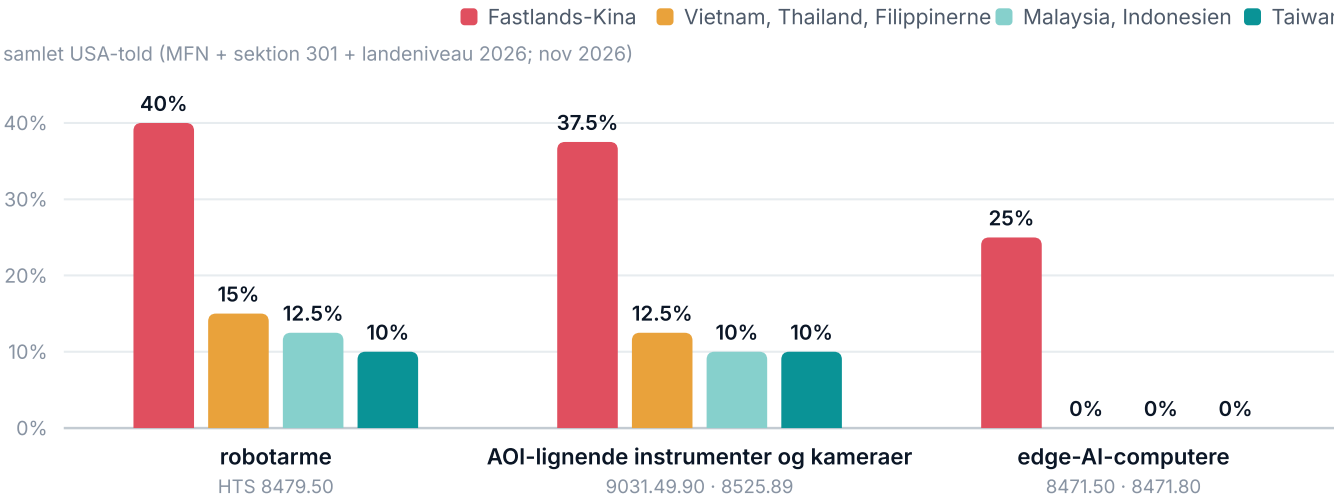
Sydøstasien: kun Malaysia og Thailand belønner køb af automatisering. Malaysias 200% automatiseringskapitalfradrag på de første RM10 mio. er omkring 24% af capex for en rentabel køber (MIDA, PwC Malaysia) verified; Thailands BOI giver en treårig skattefritagelse op til 50–100% af en opgraderingsinvestering, den højere sats når thailandske automatiseringsleverandører er involveret (BOI) single source, primær, så at gå ind i Thailand fungerer bedst med en lokal integrator. Vietnam, Indonesien og Filippinerne belønner at bygge fabrikker frem for at købe automatisering (Vietnam Law Magazine, PwC Philippines).

USA: afskrivning forkorter tilbagebetalingen; told forlænger den. Aktiver købt efter 19. januar 2025 får permanent 100% førsteårsafskrivning, omkring et 21% føderalt skatteskjold for en rentabel køber (IRS Notice 2026-11) single

source, primær ; Section 232-undersøgelsen af robotter og industrimaskiner sluttede den 28. oktober 2026 med forhandlinger frem for ny told, der skal revurderes inden for 180 dage (White House, Mayer Brown) verified . Slutmontage i Sydøstasien skærer tillægget på arme og instrumenter af AOI-typen fra omkring 37.5–40% til 10–15% (kapitel 1).

Region	Hovedinstrumenter	Effekt på købers tilbagebetaling estimate
Fastlands-Kina	1.5-points lånesubsidie i op til to år; provins- og byopgraderingsbelønninger (Guangdong kræver ≥RMB 7M registrerede projekter; Dongguan betalte faktisk 4.42%); Shenzhen 2026–2029-foranstaltninger (satser ikke offentlige)	Omkring 0–5% for en enkeltmaskine; 4–15% inde i et stort projekt
Taiwan	Statute §10-1-kredit, 5% eller 3% × 3 år	Omkring 5% kortere, hvis køberen betaler nok skat
Malaysia	200% automatiseringskapitalfradrag (første RM10M)	Omkring 15–25% kortere for rentable købere
Thailand	BOI treårig skattefritagelse op til 50–100% af investeringen	Omkring 15–25% kortere for rentable købere uden for en skatteferie
Vietnam, Indonesien, Filippinerne	Indkomstskat-incitamenter for at bygge fabrikker; ekstra 20% udstyrsafskrivning i Filippinerne	Næsten intet for enkelt automatiseringskøb
USA	100% førsteårsafskrivning; ≈37.5–40% samlet told på fastlands-fremstillet udstyr	Afskrivning ≈20% kortere; told ≈10–40% længere

FIG 21 **USA-told på robotter, AOI og edge-computere** ***Arme lavet i fastlands-Kina betaler cirka 40%; sendt fra Sydøstasien eller Taiwan, 10–15%.***



Kilde: USITC Harmonized Tariff Schedule kapitel 99; Federal Register; USTR (november 2026).

Pr. region: i fastlands-Kina tjener en AI-gennemgangsboks sig hjem på måneder og en cobot-celle på 1.3–1.5 år, mens en ny AOI købt kun for at skære arbejdskraft tager 3–4 år og integreret detektér-og-pluk 4.5–5.3, begge ud over en toårs-screening. Taiwan klarer sig lidt bedre end Shenzhen (0.53 år for gennemgang, 1.1 for en cobot-celle), og rekruttering der er sværere. I Malaysia og Thailand tager en AI-boks 1–1.5 år og en cobot-celle 2–3, omkring 15–25% hurtigere med lokale skatteincitamenter for rentable købere. I Indonesien, Vietnam og Filippinerne tjener kun en meget billig AI-boks sig hjem inden for 2–3 år; efterspørgslen der kommer mere fra kundernes kvalitets-

krav og arbejdskraftmangel. I USA tjener enhver mulighed undtagen en ren udplukningsarm sig hjem på 0.2–3.4 år, med told, Section 232-revurderingen og NRTL-certificering som hovedvariablerne.

05

Marked og konkurrence

Marked, konkurrence og hvordan købere betaler

Markedsstørrelse: publicerede tal adskiller sig med multipler; kun en bottom-up-opgørelse er brugbar

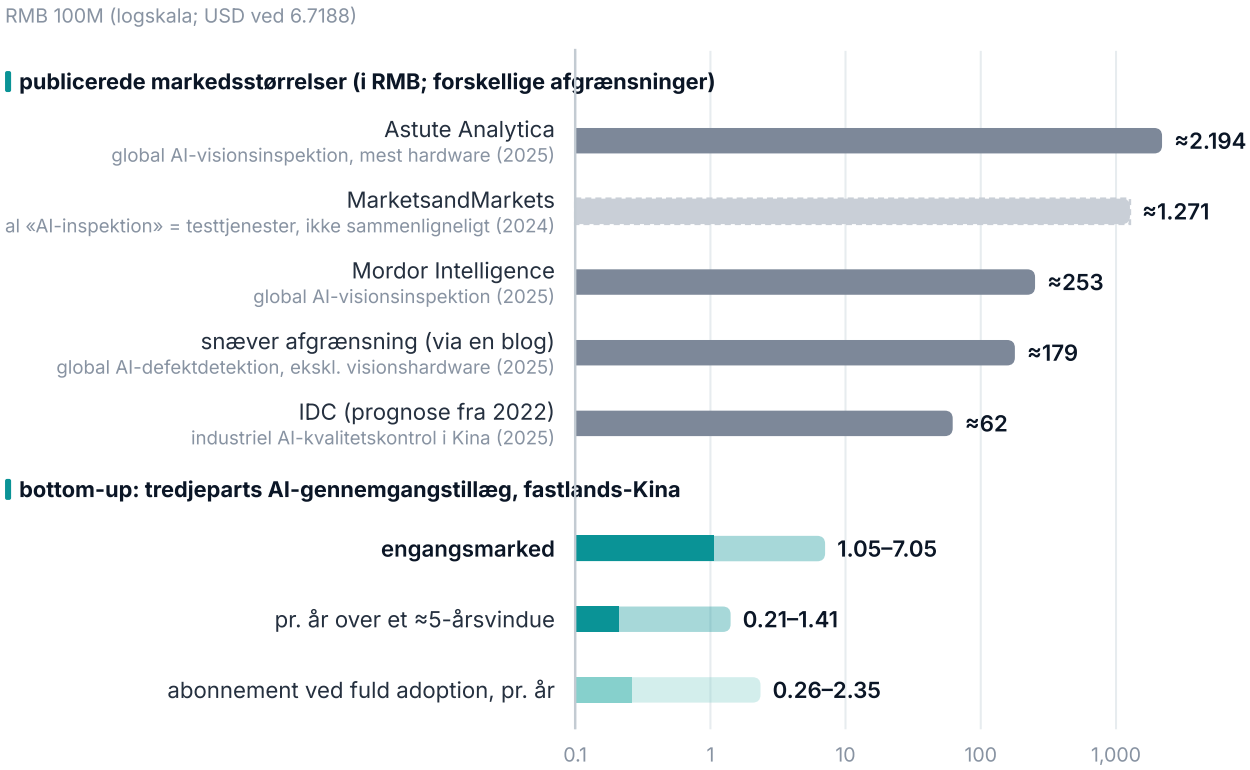
Ethvert markedstal skal læses med sit omfang. Kinas maskinsynsmarked var RMB 18.15 mia. i 2024 (Zhiyan) verified, med 3C-elektronik som den største køber; globale estimater for maskinsyn adskiller sig omkring tre gange og for AOI omkring 6.5 gange (MarketsandMarkets, Vision Systems Design/Interact Analysis, Zhiyan). Bottom-up fra omsætningen hos Koh Young, TRI, Jutze, Sinic-Tek og ViTrox er det globale SMT-inspektionsudstyr i størrelsesordenen $1–2 mia. estimate.

Der er intet selvstændigt marked for "AI-synsinspektionssoftware". De publicerede tal adskiller sig med mere end ti gange og omfatter alle hardware eller tjenester:

Omfang	Udgiver	Tal
Global "AI industriel defektdetektion" (snævert, eksklusive visionshardware)	Unavngivet firma, via en iFactory-blog	$2.66 mia. i 2025, 8.6% CAGR til 2035, elektronik 34%
Global "AI-synsinspektion"	Mordor Intelligence	$3.76 mia. i 2025, $12.8 mia. i 2031, 22.67% CAGR
Global "AI-synsinspektion" (mest hardware)	Astute Analytica	≈$32.66 mia. i 2025
Global "AI-inspektion" (test- og certificeringstjenester; ikke sammenlignelig)	MarketsandMarkets	$18.92 mia. i 2024
Kina "industriel AI-kvalitetsinspektion"	IDC, prognose lavet i 2022	$958 mio. i 2025 (≈RMB 6.2 mia.), 91.5% i kommunikation, elektronik og to andre brancher

Kilder: (iFactory, GII/Mordor, GII/Astute, MarketsandMarkets, IDC) hver single source, for det meste betalte rapportresuméer.

FIG 22 **Markedsstørrelser adskiller sig mere end tidobbelt** ***Kun den snævre afgrænsning og en bottom-up-optælling er brugbare for beslutninger.***



Kilde: Mordor Intelligence, Astute Analytica, MarketsandMarkets, IDC (2022), iFactory; forfatterens bottom-up-estimat.

Bottom-up findes der kun én offentlig opgørelse af SMT-linjer: New Venture Research estimerede omkring 73,231 linjer globalt i 2020, 47% af dem i Kina (19% i 2000), omkring 70% high-end og 30% low-end (EMSNow, NVR SMT2020-resumé) single source, 2020 ; NVR tilhører nu IPC, og nyere data er betalte (NVR). At vokse 3–5% om året med en 40–47% Kina-andel giver omkring 87,000–98,000 linjer globalt i 2026 og 35,000–46,000 i fastlands-Kina (base 40,000) estimate .

Input	Lav	Base	Høj	Grundlag
Fastlands-Kina SMT-linjer (2026)	35,000	40,000	46,000	NVR 2020-base × vækst × andel
Andel uden AI-gennemgang	75%	80%	85%	Retrofitvindue, nedenfor
Andel en tredjepart kan nå (ikke intern, ikke bundet til OEM-AI, kan betale)	40%	50%	60%	Antagelse: ≈30% low-end-linjer plus nogle blandede high-end-linjer, eksklusive DBG-agtige interne builds
Adresserbare linjer	**10,500**	**16,000**	**23,500**	Produkt
Engangspris pr. linje	RMB 10k	RMB 20k	RMB 30k	Pristrappe og loft (kapitel 4)
Engangs adresserbart marked (installeret base)	**RMB 105M**	**RMB 320M**	**RMB 705M**	Produkt
Pr. år over et ≈5-årsvindue	RMB 21M	RMB 64M	RMB 141M	Vinduet skrumper 8–12% om året
Abonnement (RMB 2.5k / 5k / 10k pr. linje pr. år, fuld adoption)	RMB 26M om året	RMB 80M om året	RMB 235M om året	Baidu-agtigt årligt gebyr-anker

FIG 23 **Bottom-up: 40.000 linjer → 16.000 servicebare → ≈RMB 320M engangs *Tredjeparts-AI-gennemgangsnichen i fastlands-Kina er reel, men lille.***



en nykommer, der tager 2–5% af servicebare linjer på 3 år: ≈320–800 linjer, ≈RMB 2–5M om året
som abonnement (RMB 5k pr. linje pr. år): RMB 1.6–4M tilbagevendende, kun 0.3–2.3% af IDC's prognose for industriel AI-QC i Kina

Kilde: NVR (base 2020); forfatterens estimater.

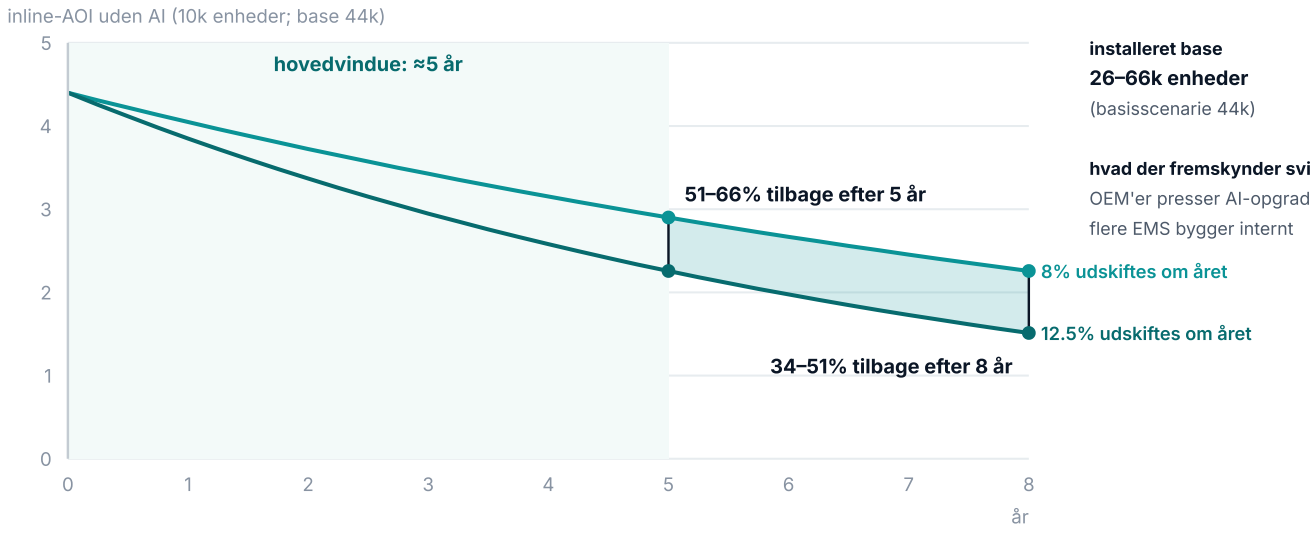
Alt dette er estimater; hverken andelen eller prisen er målt offentligt. Nicheen er kun 0.3–2.3% om året af IDC's prognose for hele Kinas industrielle AI-kvalitetsinspektion i 2025 (omkring RMB 6.2 mia.). En nykommer, der vinder 2–5% af de base-adresserbare linjer på tre år, ville have omkring 320–800 linjer: RMB 6.4–16 mio. i alt ved RMB 20,000 pr. linje, omkring RMB 2–5 mio. om året, eller RMB 1.6–4 mio. om året i tilbagevendende omsætning ved RMB 5,000 pr. linje estimate. Det er nok til en forretningsenhed inde i en virksomhed, ikke til en venture-skala selvstændig virksomhed. Hver 1,000 installerede linjer i Taiwan, Sydøstasien eller USA er omtrent RMB 3–15 mio. engangs til fastlandspriser; priserne er højere i Taiwan og USA, men købere der bruger oftere OEM-3D-AOI'er med AI-muligheder estimate.

Retrofitvinduet er omkring fem år

Ingen AOI-producent offentliggør maskinlevetid; EMS-virksomheder afskriver udstyr over 3–10 år (DBG årsrapport 2025), og ældre AOI'er mister softwareopdateringer og reservedele efter end-of-life (I-Connect007) single source. Jeg bruger 8–12 år (base 10).

Ved 1.0–1.65 inline-AOI'er pr. linje har fastlandet omkring 35,000–78,000 inline-AOI'er i drift (base omkring 55,000), i overensstemmelse med kendte forsendelser (Circuits Assembly, Sinic-Tek årsrapport 2025). Ingen kilde måler, hvor mange der kører AI-gennemgang; hvis man antager, at indenlandske maskiner har tilbudt det siden omkring 2023, og at 40–60% af de seneste leverancer bruger det, **kører omkring 26,000–66,000 inline-AOI'er i fastlands-Kina (base omkring 44,000) ikke AI** estimate. Hvis hver ny maskine nu leveres med AI, skrumper den pulje 8–12.5% om året: 51–66% forbliver efter fem år og 34–51% efter otte, hurtigere hvis AOI-producenter skubber AI-opgraderinger til nyere maskiner. **Det meste af muligheden falder i de næste fem år eller deromkring**, med mellemstore EMS, der kører ældre indenlandske 2D-AOI'er, blandede mærker og ingen AI-opgraderingsvej, som kernen.

FIG 24 **Retrofitvinduet er cirka fem år** ***Inline-AOI uden AI skrumper med 8–12.5% om året, efterhånden som de udskiftes eller opgraderes.***



Kilde: Forfatterens estimater; installeret base og andele ikke uafhængigt verificeret.

Konkurrenterne er allerede stærke: gennemgangskilen er omringet

AI-gennemgangskilen indsnævres fra tre retninger. AOI- og visionsledere bygger AI ind i nye platforme: Jutze har "fuldt adopteret AI-algoritmer" (Jutze investor relations, juni 2026), OPT's industrielle AI-omsætning voksede 96.77% i 2025 (OPT årsrapport 2025) single source, primær, og ViTrox hævder op til 95% nøjagtighed for autonom gennemgang (ViTrox) vendor claim. Store producenter bygger selv (iThome). Og venture-støttede AI-virksomheder sælger i skala: SmartMores edge-sensorkunder voksede fra 65 til 251 på to år (36Kr) single source, Hikvision sælger sin Guanlan-model til inspektion (Xinhua), og "øje, hjerne og hånd"-integration udbydes allerede af SmartMore, Hikrobot og Mech-Mind (36Kr, Securities Star).

FIG 25 **Gennemgangskilen er omringet på tre sider** ***Rum for tredjeparter findes i mellemstore EMS og i inspektionsnicher, AOI dækker dårligt.***



Kilde: Jutze, Sinic-Tek og DBGs årsrapporter; Hikvision og SmartMore offentligt materiale; sammensat af forfatteren.

Pladsen, der er tilbage til en nykommer, er snæver: mellemstore EMS med ældre 2D/3D-AOI'er og ingen AI-opgradering fra producenten; linjer, der blander flere AOI-mærker (én retrofit-leverandør understøtter allerede Omron, VI, Koh Young og Mirtec (Sohu om Hengtianyi)); nicher, som SMT-AOI'er dækker dårligt, såsom gennemhuls- og håndloddede samlinger, konform belægning, kabler, stik og FPC'er, og slutudseende og skruer; og små fabrikker, der ikke har råd til en RMB 230–280k-AOI.

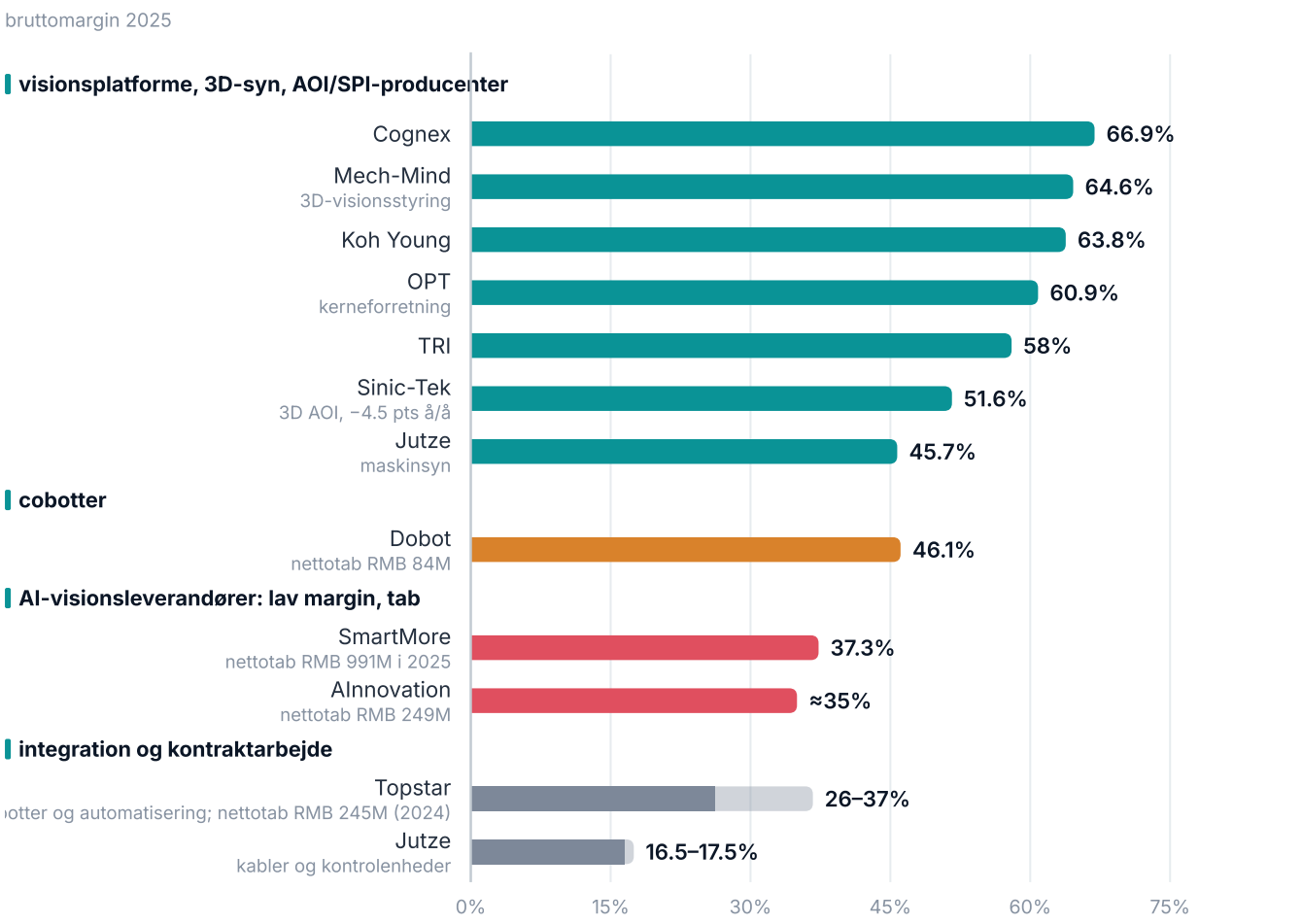
Den nærmeste børsnoterede sammenligning giver en advarsel. Weiyi Zhizao, der sælger AI-inspektionsudstyr og robotter, så 3C-omsætningen falde 79.9% år-til-år i januar–maj 2026, anerkendte "tiltagende konkurrence og prispres", udskød sin Hongkong-notering den 25. oktober 2026 og oplyste 15 projekter afsluttet før færdiggørelse, 12 af dem embodied-robot-projekter (Weiyi Zhizao-prospekt, udskydelsesmeddelelse) single source, primær . AInnovation sluttede 2025 med 29.8% af sine tilgodehavender hensat (AInnovation årsrapport 2025). 3C-inspektionskilen er overfyldt, og prissætningskraften pr. linje er svag.

Marginen sidder i produktiseret hardware

Position i kæden	Repræsentative virksomheder og 2025-bruttomargin	Overskud og inddrivelser
Visionsplatforme	Cognex 66.9%	Rentabel; frasælger ≈$22M lavmarginomsætning
AOI/SPI-udstyr	Koh Young 63.8%; TRI 58%; OPT kerneforretning 60.85%; Sinic-Tek 3D-AOI 51.62% (−4.53 point år-til-år); Jutze maskinsyn 45.74%	Rentabel; tilgodehavendedage 139 (Jutze) til 247–274 (OPT)
3D-visionsstyring	Mech-Mind 64.6%	—
AI-vision-løsningsleverandører	SmartMore 37.3%; AInnovation ≈35%	Tabsgivende: SmartMore nettotab RMB 991M i 2025, AInnovation RMB 249M
Cobotter	Dobot 46.1%	Dobot nettotab RMB 84M; UR-omsætning fra $326M (2022) til ≈$232M (2025, prognose)
Systemintegration	Topstar robotik og automatisering 26.2–36.7%	Nettotab RMB 245M (2024); tilgodehavendedage 180–240
Kontrakthardware	Jutze kabler og styreenheder 16.5–17.5%	—

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

FIG 26 **Marginen sidder i produktificeret hardware** ***Visionsplatforme og udstyrsproducenter tjener 46–67% bruttomargin; AI-leverandører 35–37% og taber penge.***



Kilde: Årsrapporter eller prospekter 2025.

En nykommer beholder kun sin margin ved at gøre inspektion til et produkt, udstyr eller software; projekt-for-projekt-tuning presser marginen ned til omkring 35% med vedvarende tab. Mekanikken virker også imod små virksomheder: omsætning indregnes først efter accept på stedet, forudbetalinger er små (Jutze, OPT), piloter er normalt gratis (A3/Mindtrace), og brandejere specificerer ofte inspektionsudstyret på deres kontraktproducenters linjer.

Hvordan købere betaler: 30% ved underskrift, 60% efter accept, 10% tilbageholdt i et år

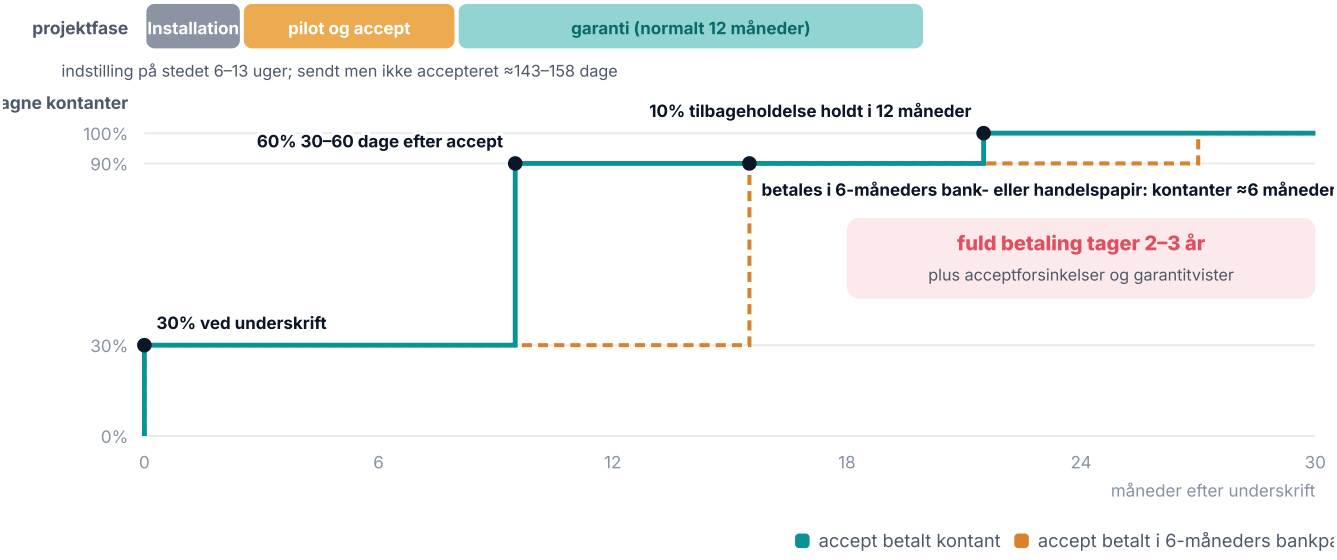
Betalingsvilkårene, som fastlandselektronikfabrikker bruger for inspektionsudstyr, er udfoldet i 13 navngivne køberkontrakter oplyst i Sinic-Teks noteringsforespørgselssvar (Sinic-Tek forespørgselssvar 8-1-1) single source: én primær indberetning, flere uafhængige købere, 2020–2022:

Køber (år)	Underskrift	Levering	Accept	Tilbageholdelse eller saldo
En Hikvision-enhed (2021)	30% inden for 15 dage	—	65% inden for 30 dage efter endelig accept	5% garantitilbageholdelse betalt 12 måneder efter accept; del betalt i 6-måneders+ bankacceptnotaer
Nanchang Huaqin (2020–2021)	30% inden for 10 arbejdsdage	—	60% ved 60 dage efter accept	10% efter ét års garanti
Guangdong Yinuo (2021, 3D-AOI)	30% inden for 10 arbejdsdage	—	60% inden for 30 dage	10% efter tre måneders fejlfri drift

Køber (år)	Underskrift	Levering	Accept	Tilbageholdelse eller saldo
Xiamen Hongxin, Jingmen Hongyi (2022)	30% inden for 10 arbejdsdage	—	60% inden for 30 arbejdsdage efter endelig accept, 6-måneders banknotaer	10% efter ét års garanti
AKM Panyu, Suzhou (2021–2022)	30% inden for 7–10 arbejdsdage	—	60% en måned efter accept og fakturering, 180-dages banknotaer	10% efter 360 dage
BYD Precision, Shenzhen, Huizhou (2020)	—	30% ved ankomst	70% efter accept og fakturering	Alt i 6-måneders kommercielle notaer
Shenzhen H&T (2020)	—	—	90% efter accept og fakturering	10% efter seks måneders fejlfri brug

Andre primære indberetninger viser samme struktur: Mech-Minds store projekter betaler 30% ved underskrift, 30% før levering, 30% efter accept og 10% garantitilbageholdelse over 12–15 måneder (Mech-Mind-prospekt); GKG siger, at tilbageholdelse i udstyrsbranchen normalt er 5–10% af kontrakten, betalt et år efter accept (GKG-forespørgselssvar) single sources, primære. Normen er altså omkring 30% ved underskrift (20–30%, nul hos nogle fabrikker som BYD), 60–65% efter accept (betalt 30–60 dage senere, ofte i seksmåneders bank- eller kommercielle notaer, hvilket forsinker brugbare kontanter med endnu et halvt år), og 5–10% tilbageholdelse i 3–12 måneder (12 mest almindeligt) verified på tværs af flere kontrakter.

FIG 27 **Hvordan købere betaler *30% ved underskrift, 60% efter accept, 10% holdt i et år: fuld betaling tager 2–3 år.***



Kilde: Sinic-Teks svar på forespørgsel, Mech-Mind, GKG, JT og andre dokumenter; skema på sædvanlige vilkår, tidspunkt estimeret.

Accept og prøvekørsler er langsomme og betales af sælgeren. Sinic-Teks afsendte-men-ikke-accepterede varer tog omkring 143–158 dage at omsætte, fordi store EMS- og brandkunder kræver integrationstest af hele linjen og en forlænget prøvekørsel (Sinic-Tek forespørgselssvar 8-1-2); Weiyis AI-projekter tager 4–8 måneder fra evaluering til implementering (Weiyi Zhizao-prospekt) single sources, primære. Der findes ingen konverteringsdata for AI-gennemgangspiloter; Sinic-Tek konverterede kun omkring 18–19% af sine demomaskiner direkte til salg. Jeg planlægger med én ud af tre til én ud af fem for gratis piloter og omkring 80% for betalte piloter efter shortlist estimate.

Virksomhed	Forretningsmodel	Days sales outstanding
Sinic-Tek (2025)	Udstyr med forudbetalinger	≈51 (≈112 inklusive notaer)
Mech-Mind	Standardprodukter gennem integratorer	59–99
Jutze	AOI-udstyr	≈139
AInnovation (2025)	AI-projekter	≈137 ved årets udgang, med 29.8% nedskrivning
Weiyi Zhizao	Projektbaseret udstyr	188–319
OPT	Visionskomponenter	247–274

Kilder: årsrapporterne og prospekterne ovenfor; tilgodehavender (inklusive kontraktaktiver) ÷ omsætning × 365 primær + estimate.

FIG 28 **Days sales outstanding** ***Standardprodukter 51–99 dage; projektbaseret udstyr 188–319 dage.***



Kilde: Selskabsdokumenter; tilgodehavender (inkl. kontraktaktiver) ÷ omsætning × 365.

AI-gennemgangslicenser solgt gennem AOI-producenter eller integratorer bør inddrive som Mech-Minds 60–100 dage; direkte projektsalg som Weiyis 190–320 dage. Samlet som en kontanttidslinje: gratis pilot én til tre måneder, kvalificering hos en stor ODM op til omkring tolv måneder, bygning, levering og accept tre til syv måneder; 30% kommer ved ordren, 60% omkring fire til ni måneder efter den, og de sidste 10% omkring 16–21 måneder efter den. **Fra det første pilotprojekt hos en stor EMS til fuld betaling kan der gå to til tre år**; mellemstore EMS uden formelle udbud kan forkorte kvalificeringen til uger estimate. Acceptkriterierne i de oplyste kontrakter dækker kun integration af hele linjen, gennemløb (ét taiwansk projekt krævede 100 enheder pr. time) og måneders fejlfri drift; ingen specificerer escape- eller falskudkaldsrater (Anda-forespørgselssvar) single source, primær. Skriftlige acceptkriterier for AI-gennemgang skal aftales før piloten.

Fire fejlmønstre

Dokumenterede fejl falder i fire mønstre: at sende før produktet var pålideligt (Rethink (The Robot Report)); et softwarelag på en leders hardware, som lederens økosystem senere erstattede (READY Robotics (The Robot Report)); at løbe tør for penge i bunden af en cyklus (Dextrous Robotics (Robots & Startups)); og modelvirksomheder, der ender i acqui-hires (Covariant (GeekWire)). Læringen: produktisér snævert i stedet for at tune projekt for projekt; hold arbejdskapital til acceptbaseret omsætning og lange inddrivelser; lever lokal service til fabrikker, der kører 24 timer; og skalér ikke, før produktet er bevist.

Timing: inspektion vinder intet ved at vente; armhardware bliver billigere

Argumenterne for at vente er svage. Hardware blev ikke billigere i 2026: robotproducenter hævede priserne 5–15%, nogle produkter 25–50% (Sina Finance) single source, og NVIDIA-kits steg en tredje gang (36Kr). Produktionsklare VLA'er kommer ikke inden for vinduet (IFR, Instrumental). Imens fjerner hver AOI-udskiftning et retrofit-mål, og hver OEM-AI-implementering låser en linjes data og proces; modelkapacitet kommoditiseres, og det, der er knapt, er adgang til fabrikker og annoterede fabriksdata. Lokal support har også vinduer: Shenzhens 2026 "AI-pionérby"-regler, med op til RMB 2 mio. til AI-applikationsdemonstrationsprojekter og modelvouchere, der dækker 30% af køb, løber til **30. juni 2027** (Yicai, maj 2026) vilkår single source.

Det, der stadig bliver hurtigere billigere, er humanoider og fingerfærdige hænder (hænder faldt fra omkring RMB 50,000 til under 10,000 på to år (Sina Finance)), og entry-cobotter, efterhånden som produktmikset bevæger sig nedad; det taler for at udskyde den embodied del, ikke inspektionsdelen. **Samlet: inspektion nu, som et snævert produkt; robotceller nu, gennem partnere og integration; embodied AI senere, med kun en R&D-station i dag.**

Embodied AI

06 Embodied AI: først ændrer den, hvordan robotter læres, senere hvordan de bevæger sig

Det, der ændrer sig, er perception og programmering; bevægelse forbliver deterministisk

Enhver robotarm-leder markedsfører nu "fysisk AI", men det, der er leveret, falder i tre lag: indlært perception uden undervisning (FANUCs AI Masterless Vision, leveres i januar 2027 (FANUC); KUKAs neurale grebsudvælgelse (KUKA)); store modeller, der skriver programmer og planlægger opgaver (KUKAs iiQWorks.Copilot, som advarer om, at AI-genereret kode bør verificeres i en digital tvilling (KUKA); Yaskawa med Gemini Robotics-ER, hvor "generativ AI beslutter, hvad der skal gøres", og robotten beslutter "hvordan den skal bevæge sig" (Yaskawa)); og computerkraft- og dataplatforme (URs AI Accelerator med en Jetson Orin (Universal Robots)) vendor claims. Ingen leder offentliggør produktionssuccesrater for sine AI-funktioner, og de japanske ledere samarbejder med Google, SoftBank, Microsoft og andre frem for at bygge front-VLA'er (FANUC-nyhedsindeks): det indlærte politiklag kommer fra modelvirksomheder, de certificerede arme og controllere fra armproducenterne.

FIG 29 **Embodied-AI-stacken** ***Store modeller planlægger offline, små politikker kører i cellen, bevægelse og sikkerhed forbliver deterministiske.***

Lag	hvad den gør, og hvor langt beviserne rækker	tidsskala · hvem ejer det
Planlægning L4	**frontmodeller (offline)** skrive eller ændre robotprogrammer offline, omstillinger, forklare fejl omskriver ABB RAPID-programmer til 77–99.7%, udrullet efter simulering styreløkken kun 0.2–0.4 Hz: holder sig uden for realtid	**sekunder til minutter** modelgiganter
færdigheder / politi L3	**ACT, diffusionspolitikker, små VLA'er** strukturerede stationer med fiksturer 90–99.4% edge-chips: Jetson Thor, D-Robotics RDK, Ascend, Rockchip	**handlingschunks 100–300 ms** chipproducenter · robotproduc
bevægelsesstyring L2	**robotcontroller** baner, servo, kraftregulering, kollisionskontrol deterministisk kode; indlærte politikker sætter kun mål	**millisekunder (realtid kræver >83 Hz)** robotproducenter
Sikkerhed L1	**sikkerheds-PLC, scannere, kraft- og hastighedsgrænser** certificeret efter ISO 10218-1/-2:2025 og lokale regler ingen ML i sikkerhedsfunktioner: EU-vurdering af bemyndiget organ fra 20. jan 2027	**hård realtid** certificeret robot + integrator

små virksomheder vinder i integration og sensorhardware: sæt indlært perception og små politikker ind i certificerede robotceller, station for sta

Kilde: Forfatterens syntese (kapitel 3 og 6); ISO 10218:2025; EU's maskinforordning.

Hvor langt fra ustruktureret, fleksibelt, småbatch-arbejde: strukturerede stationer på 1–3 år, fleksibel montage 5+

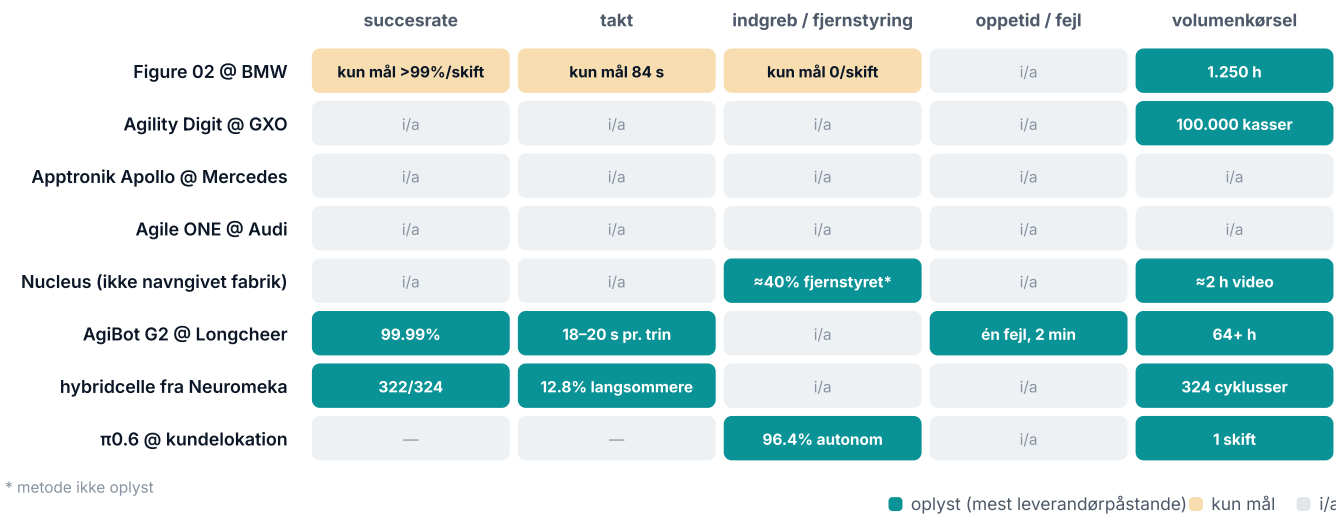
Ægte fabrikstal er begyndt at dukke op, men de er leverandørrapporterede, snævre og fra strukturerede stationer. AgiBots otte humanoider på Longcheers tablettestlinje loggede 64,828 opgaver ved 99.99% over mere end 64 timer (AgiBot, Yicai Global) [leverandør og kunde], men det er betjeningsarbejde, en konventionel arm med vision kan udføre. UBTech siger, at dens Walker S2 arbejder ved 30–50% af menneskelig effektivitet (Irish Times/FT) sing-

le source. Figure 02's 11 måneder hos BMW Spartanburg producerede volumental (1,250+ timer, 90,000+ dele), mens dens cyklus-, succes- og indgrebstal er mål, ikke resultater (Figure, BMW Group) verified, kun tællere. Nucleus siger, at omkring 40% af dens fabriksvideo er teleopereret (Interesting Engineering) vendor claim, og en Reuters-undersøgelse citerede omkring 20% af en arbejders output for nogle kinesiske humanoider (Notebookcheck) single source. Ingen implementering offentliggør succesrate, cyklustid og indgrebsrate samlet, og ingen er auditeret, men netop de tre er, hvad en tilbagebetalingsberegning kræver.

Implementering	Succesrate	Takt	Indgreb eller teleoperation	Oppetid eller fejlinterval	Volumenrun
Figure 02 @ BMW Spartanburg	Kun mål (>99% pr. skift)	Kun mål (84 s)	Kun mål (0 pr. skift)	Ikke oplyst (underarmen øverste fejlpunkt)	1,250 t
Agility Digit @ GXO	Ikke oplyst	Ikke oplyst	Ikke oplyst	Ikke oplyst	100,000 kasser
Apptronik Apollo @ Mercedes-Benz	Ikke oplyst	Ikke oplyst	Ikke oplyst	Ikke oplyst	Ikke oplyst
Agile ONE @ Audi	Ikke oplyst	Ikke oplyst	Ikke oplyst	Ikke oplyst	Ikke oplyst
Nucleus (unavngivet fabrik)	Ikke oplyst	Ikke oplyst	≈40% teleopereret (metode ikke oplyst)	Ikke oplyst	≈2 t video
AgiBot G2 @ Longcheer	99.99% af 64,828 opgaver	18–20 s pr. trin	Ikke oplyst	Én kommunikationsfejl, ≈2 min at komme sig	64+ t
Neuromeka hybrid-celle	322/324	12.8% langsommere end en person	Ikke oplyst	Ikke oplyst	324 cyklusser
Physical Intelligence π0.6 hos en kunde	—	—	96.4% autonom over et skift	Ikke oplyst	Ét skift

Kilder: dette kapitel og kapitel 3 for det meste vendor claims.

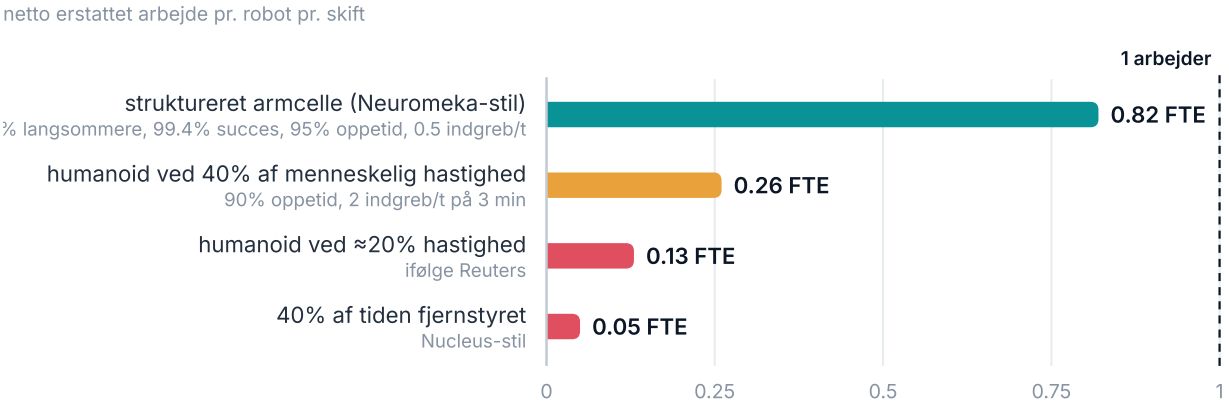
FIG 30 **Hvad robotudrulninger i fabrik faktisk offentliggør** ***De fleste offentliggør kun volumen; ingen offentliggør succesrate, takt og interventionsfrekvens sammen.***



Kilde: Offentligt materiale fra BMW, Figure, Agility, Mercedes-Benz, Audi, AgiBot, Neuromeka, Physical Intelligence med flere.

Teleoperation betyder mere for økonomien end succesraten. Netto erstattede personer pr. robot pr. skift ≈ (menneskelig cyklus ÷ robotcyklus) × oppetid × succes, minus indgrebstid og teleoperation estimate, min formel. En Neuromeka-agtig struktureret station (12.8% langsommere, 99.4% succes, 95% oppetid, 0.5 indgreb på 2 minutter i timen) erstatter omkring 0.82 personer; en humanoid ved 40% af menneskelig hastighed, 90% oppetid og to 3-minutters indgreb i timen omkring 0.26; ved Reuters' 20%, omkring 0.13; og hvis, som Nucleus, én operatør teleopererer den 40% af tiden, omkring 0.05, næsten ingenting. Ved omkring RMB 600k pr. humanoid og RMB 85k pr. person pr. år på fastlandet tjener de sidste tre sig hjem over årtier; kun den strukturerede armstation har en realistisk tilbagebetaling (kapitel 4) estimate.

FIG 31 **Netto erstattet arbejde pr. robot** ***En struktureret armcelle erstatter cirka 0.82 af en arbejder; en humanoid fjernstyret 40% af tiden, cirka 0.05.***



Kilde: Forfatterens formelbaserede estimater.

Væk fra strukturerede stationer scorer generelle modeller stadig 40–90% på nye opgaver: Gemini Robotics 2 når 74.2% på lab-pick-and-place, 78.9% på værktøjskitning og 44% på at binde en pose (eesel/DeepMind); på RoboChallenge Table30, 30 ægte robotopgaver, scorer π0.5 43.7%, og den bedste åbne model 50.3% (RoboChallenge, QbitAI).

Eksterne prognoser er uenige, men ingen understøtter udbredt brug inden for to år. IFR siger, at generel fysisk AI "endnu ikke er industriel kvalitet", og at de fleste humanoider stadig er "prototyper eller piloter" (IFR); Interact Analysis estimerer, at kun omkring 10% af de humanoider, der blev produceret i 2025, gik i reel brug, med det kommercielle vendepunkt "efter 2032" (automation.com); UBS ser et humanoid-"EV-øjeblik" som usandsynligt før 2030 (Yicai Global); kinesiske insidere estimerer to til tre år mere før fabriksskala-brug (Gasgoo) single sources. Min vurdering estimate: i velresourcede elektronik- og bilfabrikker med leverandøren på stedet vil **strukturerede stationsopgaver** (betjening, OK/NG-sortering, fikseret indsætning) nå ≥99% ved næsten menneskelig hastighed på omkring 1–3 år; **blandet-dele-kitning** med generelle VLA'er på omkring 3–5; **frit rutede seler og flertrins fleksibel montage** på mere end 5. Humanoid tilbagebetaling virker ikke i dag: en Walker S2 ved omkring RMB 600k (10jqka) single source, ved 30–50% af menneskelig effektivitet over to skift, svarer til omkring 0.6–1.0 arbejdere, en simpel tilbagebetaling på 7–12 år ved fastlandsarbejdskraftomkostninger før integration, vedligeholdelse og nedetid estimate.

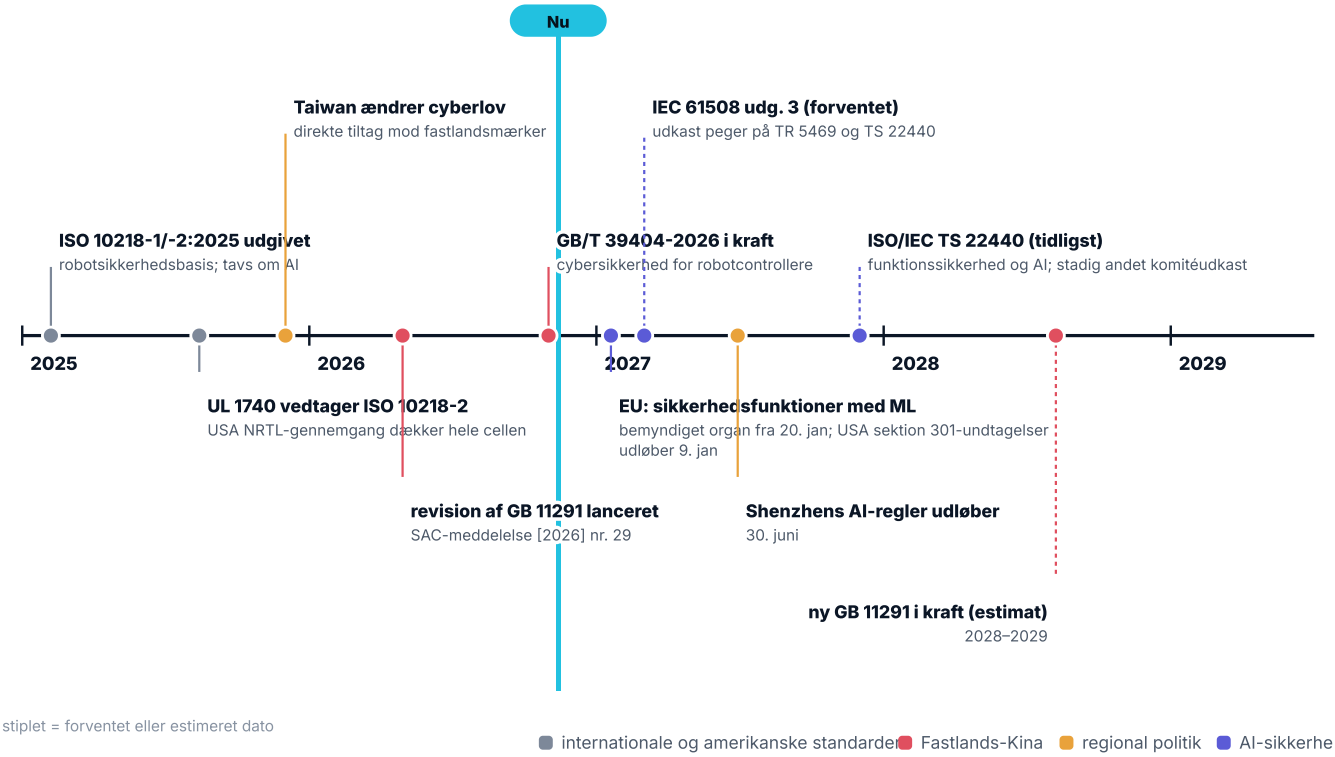
Hvor det sidder fast

Led	Evidens i 2026
Data	Branchen estimerer mindst 10 mio. timers højkvalitets ægte data nødvendige, og kun omkring 500,000 var indsamlet i begyndelsen af 2026 (Gasgoo); Reuters oplyser omkring 500,000 timer mod omkring 100 mio. nødvendige (Notebookcheck); en novice-teleoperatør kræver omkring 300 forsøg pr. brugbar bevægelse; fabrikker deler ikke procesdata; simuleringstrænede politikker falder skarpt på ægte robotter

Led	Evidens i 2026
Pålidelighed	Strukturerede stationer 98–99.99% (vendor claims, fejl udefineret); generelle modeller 40–90% på nye opgaver; usete stik 29–67%
Hastighed	Humanoider ved 30–50% af menneskelig hastighed (UBTech), nogle ved omkring 20% (Reuters); frontagenter omkring 280 s pr. kvasistatisk ægte robotopgave; LLM-styreløkker ved 0.2–0.4 Hz mod omkring 83 Hz nødvendigt (Anthropic)
Omkostning	Walker S2 koster i gennemsnit omkring RMB 600k; 7–12-års tilbagebetaling estimate
Sikkerhedscertificering	ISO 10218-1/-2:2025 gør funktionel sikkerhed mere eksplicit og absorberer kollaborative krav, men siger intet om AI (The Robot Report); ISO/IEC TS 22440 (funktionel sikkerhed og AI) var stadig en anden komitékladd i november 2026 (ISS, ISO Open Data); IEC 61508-3-vejledning fraråder AI over SIL 1 (Analog Devices); under EU's maskinforordning kræver sikkerhedsfunktioner med machine learning vurdering af bemyndiget organ fra 20. januar 2027 (EUR-Lex)
Hænder og værktøjer	Fingerfærdige hænder klarer omkring 30 N (≈3 kg) pr. finger mod omkring 10 kg for en menneskehånd (Gasgoo); værktøjsskift er stadig manuelle
Integration	Sensorfejllæsninger, protokolmismatch og MES-links på stedet; indgrebsrater næsten aldrig oplyst

Sikkerhedsimplikationen er direkte: standarderne hverken forbyder indlærte politikker eller anerkender dem. ISO 10218-1/-2:2025 siger intet om AI; ISO/IEC TS 22440 (funktionel sikkerhed og AI) er stadig ved komitékladd og usandsynlig før slutningen af 2027 eller 2028 estimate ; ISO/IEC TR 5469:2024 er kun vejledning, der anbefaler ikke-AI-fallbacks, superviserende begrænsninger og redundant afstemning; IEC 61508 edition 3 forventes i begyndelsen af 2027 (ISO Open Data, BSI, 61508 Association) verified . Ingen publiceret standard tilbyder endnu en accepteret vej til at certificere machine-learning-perception som en sikkerhedsfunktion. Regionale regler bevæger sig også: Kina reviderer GB 11291 (i kraft omkring 2028–2029) (revisionsplan 20262543); UL 1740 har bragt ISO 10218-2 ind i USA's NRTL-gennemgang siden august 2025; EU kræver vurdering af bemyndiget organ for sikkerhedsfunktioner med machine learning fra 20. januar 2027; og ingen humanoid er certificeret til hegnløs fabriksbrug, mens ISO 25785-1 stadig er i kladd (i-SCOOP). Den forsvarlige arkitektur: machine learning træffer kun procesbeslutninger ("hvilken del", "er dette kort godt"), mens et certificeret deterministisk sikkerhedslag (typisk PL d / Kategori 3) og en sikkerheds-PLC beslutter hastighed, kraft og stop. Behandl enhver "kamera-AI til hegnløs sikkerhed"-påstand som ubevist, indtil et PL d- eller SIL 2-certifikat navngiver machine-learning-komponenten.

FIG 32 **Tidslinje for standarder og regulering** ***En certificeringsvej for AI inden for sikkerhedsfunktioner afklares ikke før 2027–2029.***



Kilde: ISO, IEC, UL, EUR-Lex, SAC og nationale reguleringsdatabaser.

Hvilket lag er muligheden: modeller tilhører giganterne, edge-bokse til chipproducenterne, små firmaer til integration og sensorhardware

Modellaget kræver kapital og data og tilhører giganterne (AgiBot trænede GO-2 på "titusindvis af timer" (iFeng Tech); de stærkeste modeller er lukkede eller cloud-only). Edge-bokslaget tilhører chipproducenterne: Jetson Thor er de facto-standarden, og på fastlandet konkurrerer D-Robotics' RDK-kort (D-Robotics) og Rockchip-co-processorer (CSDN), for det meste uden publicerede VLA-resultater.

Små firmaer har to åbninger. Én er sensorhardware, der direkte køber pålidelighed: berøring løfter ContactDPs indsætningssucces fra 58.7% til 97.8%, og kraft/moment-sensorer og håndledskameraer gør lignende arbejde; dataindsamlingskits sælger også (10jqka), selvom hænder, sensorer og arme alle falder i pris. Den anden er integrationslaget mellem armproducenter og modelvirksomheder, som ejer hver stations opgavedata, sikkerhedskonvolut og MES-link, hvilket armproducenter ikke sælger som et produkt i dag. Konkurrenter falder i fire grupper: armledere (FANUC, Yaskawa, KUKA, ABB, UR, JAKA, Dobot og andre), vision-plus-arm-virksomheder (Mech-Mind, Hikrobot, SmartMore), embodied-AI-virksomheder (AgiBot, UBTech, Unitree, Galbot, Figure, Physical Intelligence, Google DeepMind, NVIDIA) og bil-chip-entranter.

Køberaccept: intet tal uden en nævner

Da ingen leverandør offentliggør indgrebsrater, fejlintervaller eller langsigtede succesoptællinger, hører disse til i kontraktens acceptvilkår. Statistisk beviser 3,000 på hinanden følgende produktionscyklusser med nul fejl en fejlrate på højst 0.1% ved 95% konfidens; 10 fejl i 3,000 beviser kun ≤0.57%; at bevise ≤1%, ≤0.5%, ≤0.1% og ≤0.01% med nul fejl tager 299, 598, 2,995 og 29,956 cyklusser estimate, eksakt binomial. Efter den standard har Neuromekas 322/324 et tosidet 95%-interval på 97.8–99.9% og beviser ikke ≥99%; AgiBots omtrent 6 fejl i 64,828 opgaver antyder 99.98–100%, men kun hvis "fejl" er defineret, som køberen definerer det estimate. Dette er de acceptkrav, jeg

ville stille til enhver robotarm- eller embodied-celle; tærsklerne trækker på Figures egne mål hos BMW og på Neuromeka- og AgiBot-casene, og de er vurdering, ikke en standard estimate:

Punkt	Kræv	Foreslået bestå-tærskel
Definitioner	Underskrevne definitioner af cyklus, succes, fejl (delvis succes, gen-grib, tab, kvalitetsescape), indgreb (berøring, fjern-teleoperation, reset, e-stop) og hvordan autonomi beregnes	Definitioner kommer før enhver metrik; ingen procent uden en formel
Indgreb og teleoperation	Pr. robottime, efter type og varighed; middeltid mellem indgreb; teleoperationsminutter som andel af køretid; operatører pr. robot	Strukturerede stationer ≤0.5 indgreb pr. robottime (≥2 t fra hinanden); nul teleoperation under accept
Langsigtet succes	≥3,000 på hinanden følgende produktionscyklusser over ≥3 skift, ≥2 artikelnumre og normal belysning og forsyningsvariation; fejl klassificeret; eksakt nedre grænse	Nedre grænse ≥99.5% for linjestoppende inline-stationer, ≥99.0% offline; kvalitetsescapes rapporteret separat
Cyklustid	P50/P90/P99 pr. del inklusive genopretning, mod ≥30 timede menneskecyklusser ved samme station	≤1.15 × menneskelig P50, eller inden for linjetakt
Oppetid og fejl	Køretimer, nedetidslog, fejl efter komponent, delskiftetid	≥95% oppetid over ≥2 uger; MTTR ≤20 min med reservedele på stedet
Genopretning	Scriptet fejlinjektion: tab, okklusion, lysændring, transportbåndstop, MES- eller netværkstab, strømtab, e-stop	≥90% autonom genopretning, tilbage i produktion inden for 2 min
Omstilling	Timer, demonstrationer, operatørfærdighed og dataindsamlingsomkostning pr. nyt artikelnummer eller fikstur	Strukturerede stationer: ≤1 skift og ≤100 demoer pr. nyt artikelnummer
Sikkerhed	Arm og celle efter ISO 10218-1/-2:2025, cellerisikovurdering, lokale regler (GB 11291 på fastlandet; CNS 14490 plus skriftlig vurdering i Taiwan; ANSI/A3 R15.06-2025 og UL 1740 i USA)	Ingen sikkerhedsfunktion implementeret af machine learning; et certificeret sikkerhedslag (typisk PL d / Kat 3) begrænser hastighed og kraft; ingen hegnløse gående robotter, før ISO 25785-1 er publiceret og vurderet
Computerkraft og offlineadfærd	Inferenshardware, P99-latens, adfærd uden netværk, modelversionering, revalidering efter hver opdatering	Taktkritisk inferens afhænger ikke af skyen; fuld genaccept efter enhver model-, kvantiserings- eller motorændring
Evidens	Uredigeret video og rå logfiler for hele acceptvinduet; bevidnet af kundens kvalitetsteam eller en tredjepart	Ingen highlight-reels eller leverandør-only-resuméer
Kommercielle vilkår	Betalingsmilepæle bundet til acceptmetrikker; tilgængelighedsbøder ved fakturering pr. robottime; dataejerskab og fjernadgangsregler	Endelig betaling og tilbageholdelse frigives først efter 3,000-cyklus-accepten og ≥4 ugers stabil drift
Økonomi	Netto erstattede personer = hastighedsforhold × oppetid × succes − indgrebstid − teleoperationsandel	Afvis ethvert forslag, hvis nettoerstatning ikke kan understøtte dens hævdede tilbagebetaling

Kilder: (Figure, Neuromeka, Yicai Global, i-SCOOP); tærsklerne er mine forslag.

Åbne spørgsmål

07 Hvad de offentlige kilder ikke kan besvare, og hvordan et pilotprojekt afgør det

De spørgsmål, offentlige kilder ikke kan besvare, falder i fem grupper: hvad AI-gennemgang faktisk sælges for, og hvad købere vil betale; hvordan AOI-producenter bundter AI og opgraderer gamle maskiner; hvor gamle de installerede AOI'er er, og hvor mange der allerede kører AI; hvor mange personer og timer gennemgangen reelt tager pr. linje; og hvordan piloter konverterer, og hvordan købere betaler. Oven på dem kommer, hvordan de nyeste modeller, kvantiserede builds og robotarme faktisk præsterer under en given fabriks forhold. Hvert kan omdannes til en test med en klar tærskel. Tærsklerne nedenfor er beslutningsregler, jeg udledte af evidensen i denne rapport estimate; genberegn dem, når de første fabriksdata kommer ind.

Åbent spørgsmål	Hvordan man afgør det	Hvordan et godt resultat ser ud
Pris og betalingsvilje for AI-gennemgang	En standard forespørgsel (for eksempel fire linjer med blandede ældre 2D- og nyere 3D-AOI'er, to skift, omkring 2,000 kort pr. skift, 12 defektklasser, kun on-premises, MES- eller IPC-CFX-integration) sendt til flere tredjepartsleverandører; en fire-spørgsmåls prisfølsomhedsundersøgelse af 20–30 EMS-fabrikker	De fleste tilbud inden for RMB 40–120k engangs eller 20–40k om året for fire linjer; mindst halvdelen af fabrikkerne accepterer RMB 10–30k pr. linje med tilbagebetaling inden for 12 måneder
AOI-producenters AI-bundling og opgraderinger til ældre maskiner	Korte strukturerede interviews med omkring et dusin AOI-producenter og distributører	Hvorvidt AI-gennemgang er standard, valgfri eller fraværende på 2025–26-modeller; hvorvidt før-2020-maskiner kan opgraderes og til hvilken pris
Alder på installerede AOI'er og andel, der kører AI	On-site-undersøgelse af omkring 30 EMS-fabrikker, stratificeret efter størrelse og region, med udstyrslister gennemgået under NDA	Median AOI-alder ≥5 år og ≥60% uden AI betyder, at retrofit-puljen er reel
Gennemgangsarbejde og -timer	Interviews og 2–4 ugers AOI-alarm-, gennemgangs-, reparations- og skiftlogfiler på to eller flere fabrikker, med arbejdsstikprøver	≥1 dedikeret gennemgangsperson pr. linje pr. skift og ≥3 gennemgangstimer pr. linje pr. skift, med mindst 0.5 FTE pr. linje, som fabrikken er villig til at omplacere
Ydeevne for selve AI-gennemgangen	Gyldent sæt plus skyggetilstand på en ægte linje (kapitel 3): dataaftale i uge 0, baseline og annotering i uge 1–2, skyggekørsel i uge 2–4, valgfri assisteret kørsel i uge 4–8	Falskudkald eller gennemgangstimer skåret ≥50–60%; pr.-klasse escape-grænser ikke værre end den menneskelige baseline; nul misser på kritiske klasser; p95-latens ≤½ takt; automatisk fail-open til folk
Lærerannotering-kvalitet for frontmodeller	Kør to eller tre kandidater på det samme gyldne sæt, tre promptformuleringer hver, ét godt referencebillede, enkeltkomponent-udsnit ≥42–48 px	Automatisk accept kun for klasser med nul misser på ≥299 defekter og menneskeniveau-enighed; ellers kun præ-annotering og triage
Kvantiserede builds	Skift kun præcisionen; pin vægthashes, skabelon, forbehandling, afkodning, motor og batchstørrelse; deaktiver eller nulstil prefix-caching; kontrollér, at referencebuildet reproducerer leverandørresultater; score par på det gyldne sæt og drift-slice	Ingen nye misser på kritiske klasser; ≤1 nyt misser i 1,000 defekter; rydning ned ≤2 point; ≥99.5% identiske beslutninger over tre kørsler; ellers fald tilbage til FP8/INT8

Åbent spørgsmål	Hvordan man afgør det	Hvordan et godt resultat ser ud
Linjeside-hardware, renhed og ESD	≥2 enheder af hver kandidatboks stresstestet ved dobbelt den forventede alarmrate; 72-timers soak ved reel kabinetluftstrøm; ESD-tjek af kundens koordinator; ISO 14644-14-test for ISO 7- og renere kunder	p95 ≤½ takt; ingen nedbrud eller throttling over 72 timer; ESD-sign-off; partikelrapport ved den rette klasse
Betalingsvilkår, pilotregler og konvertering	Betalingsklausuler fra nylige inspektionsudstyrskontrakter hos de undersøgte fabrikker; interviews om pilotregler, acceptkriterier og kvalificeringstid	Omkring 30/60/10 med kontanter inden for 60 dage efter accept; piloter ≤8–12 uger med en navngivet sign-off-ejer; ≥1 ud af 3 prøvekørsler konverterer; en ramme, der dækker ≥5–10 linjer i år ét
Data, IT og ansvar	Interviews og kontraktgennemgang med juridisk rådgiver	En lokal boks må forbinde og tilgå data; ansvar begrænset til softwarefejl
Regional overholdelse før udlandsgang	Taiwan: kundens skriftlige status under Cyber Security Management Act, E187-selvvurdering; USA: toldafgørelse om, hvordan en celle klassificeres, NRTL-tilbud, ITAR-tjek; Sydøstasien: BOI local-linkage og malaysisk fradragsberettigelse, kundens moderselskab; fastlandet: CCC-omfang for bundtede PC'er og strømforsyninger	Kunden er ikke et kontrolleret organ, og tilbagebetalingen forbliver ≤2 år med told og certificering inkluderet
Accept af en robot- eller embodied-celle	12-punkts-accepttabellen i kapitel 6: underskrevne definitioner, ≥3,000 på hinanden følgende produktionscyklusser over ≥3 skift og ≥2 artikelnumre, scriptet fejlinjektion, ≥30 timede menneskecyklusser, bevidnet	Nedre grænse ≥99.5% inline (≥99.0% offline); cyklus ≤1.15 × menneskelig; ≤0.5 indgreb pr. time og nul teleoperation; ≥95% oppetid; MTTR ≤20 min; ≥90% genopretning
ACT- og VLA-politikker på kinesiske edge-chips	Fire uger: portér og mikro-benchmark (≥1,000 inferenser, latensfordeling, effekt, 60-minutters soak); closed-loop-simulering med målt latens injiceret; ægte robot-A/B mod en GPU; 8–24 timers udholdenhed (≥3,000 cyklusser)	ACT-chunks p95 ≤100 ms, VLA ≤300 ms; ægte robotsucces inden for 3 point af GPU'en og cyklus ≤1.2 × menneskelig; ellers behold politikken på Jetson eller en GPU
Cobot-, AMR- og integrationspriser	Model-for-model-tilbud fra flere cobot-distributører, og fuldcelle-tilbud fra to eller tre integratorer til testfikstur-påfyldning (én station, to skift)	Fuld celle ≤RMB 200k og tilbagebetaling ≤1.5 år over to skift i Shenzhen
SMT-linjetællinger i Taiwan, Sydøstasien og USA	IPC-data (tidligere New Venture Research) eller lokale udstyrsdistributørers installerede base	Nok adresserbare linjer pr. region til at retfærdiggøre lokal applikationsingeniørarbejde

Af alle disse afgør fire de mest følsomme tal i denne rapport: den pris, købere vil betale, den arbejdskraft, AI-gennemgang faktisk frigør, hvorvidt AI-gennemgang virkelig misser lige så lidt som folk, og hvor hurtigt kontanterne kommer tilbage. De er værd at afgøre før nogen væsentlig forpligtelse, af en fabrik, der overvejer et køb, eller af en leverandør, der overvejer et produkt.

En pilotsekvens, der virker for begge sider

For en fabrik og for en leverandør, der ønsker en troværdig reference, er rækkefølgen den samme:

1. **Uge 0–2: bliv enige om data og definitioner.** Underskriv dataaftalen, definér defektklasserne og hvad der tæller som en escape, og registrér en baseline: falskudkald pr. kort, gennemgangssekunder pr. kald, escapes fundet nedstrøms.
2. **Uge 1–4: byg det gyldne sæt** fra AOI-arkivet og gemte NG-kort, blindt annoteret og afgjort (kapitel 3).
3. **Uge 2–4 (eller længere): skyggetilstand.** Modellen logger sine beslutninger, mens folk beslutter som før; sammenlign pr. klasse.

4. **Uge 4–8: assisteret kørsel på tilfældige skift,** med modellen, der kun rydder klasser, som bestod gaten, og alt andet går til folk.
5. **Accept, skriftligt, før skalering:** pr.-klasse-resultaterne, latens, fail-open-adfærd, og hvem der ejer modeller og annoteringer. Skalér linje for linje, og revalider ved hvert nyt produkt, modelversion, kvantiserings- eller motorændring.

Begrænsninger i denne rapport

Dette er desk research. Jeg gennemførte ingen primære interviews med fabrikker til denne udgave; meget linjedata kommer fra leverandører og er ikke uafhængigt auditeret. Flere regeringswebsites, udbudsdatabaser og rekrutteringsplatforme kunne ikke tilgås, så Shenzhens 2026-opgraderingssubsidiesatser, Yangtze River Delta-lønvejledning, transaktionspriser for AI-gennemgang, gennemgangspersoner pr. linje og pilotkonverteringsrater er estimater eller analogier; hver optræder ovenfor som et åbent spørgsmål. Valutakurser er People's Bank of Chinas centralparitet den 9. november 2026 ($1 = RMB 6.7188) og, for Taiwan, SMBC's fixing samme dag ($1 = NT$32.12); den vietnamesiske dong bruger maj 2026-data. Alle tærskler er beslutningsregler udledt af den citerede evidens, ikke standarder.

Om forfatteren

Tan Shuai

Tan Shuai er teknolog og grundlægger med mere end tyve års arbejde på tværs af chips, hardware og systemsoftware samt 20 tildelte patenter. Hans arbejde har ført ham gennem Silicon Valley, New York, Chicago, Taiwan, Greater Bay Area (Shenzhen og Hongkong), Vietnam og Malaysia, med et mangeårigt fokus på avanceret fremstilling og forsyningskæder. I dag arbejder han som AI-forsker og -udvikler med fokus på, hvordan AI faktisk bruges på fabriksgulvet.

Han tager gerne imod samtaler med folk, der arbejder med AI i fremstilling, uanset om de bygger et team, leder efter en partner eller udveksler erfaringer: i@tanshuai.com · tanshuai.com

Noter og ansvarsfraskrivelse

Denne rapport er uafhængig desk research baseret på offentlige dokumenter tilgængelige frem til 11. november 2026. Den leveres kun til orientering og er ikke investerings-, juridisk, skatte- eller ingeniørrådgivning. Produktnavne, priser, modellanceringer, eksportklassifikationer, told og regler ændrer sig hurtigt; verificér alt, du stoler på, mod primærkilden. Tal markeret vendor claim eller estimate bør læses som sådan. Virksomheds- og produktnavne tilhører deres ejere og bruges kun til identifikation; ingen virksomhed har gennemgået eller sponsoreret denne rapport.

© 2026 Tan Shuai. Du må dele denne rapport uændret med kildeangivelse.

ELEKTRONIKFREMSTILLING · AI-INSPEKTION × ROBOTIK

AI-synsinspektion og robotarme i elektronikfabrikker: Vurdering 2026

Hvert nøgletal bærer sin kilde og et konfidensgradsmærke: verified, single source, vendor claim eller estimate. Det, som de offentlige kilder ikke kan besvare, er samlet i kapitel 7, med hvordan et pilotprojekt afgør hvert spørgsmål.

Tan Shuai
tanshuai.comi@tanshuai.com

OFFENTLIG UDGAVE · DATA PR. 11. NOVEMBER 2026

