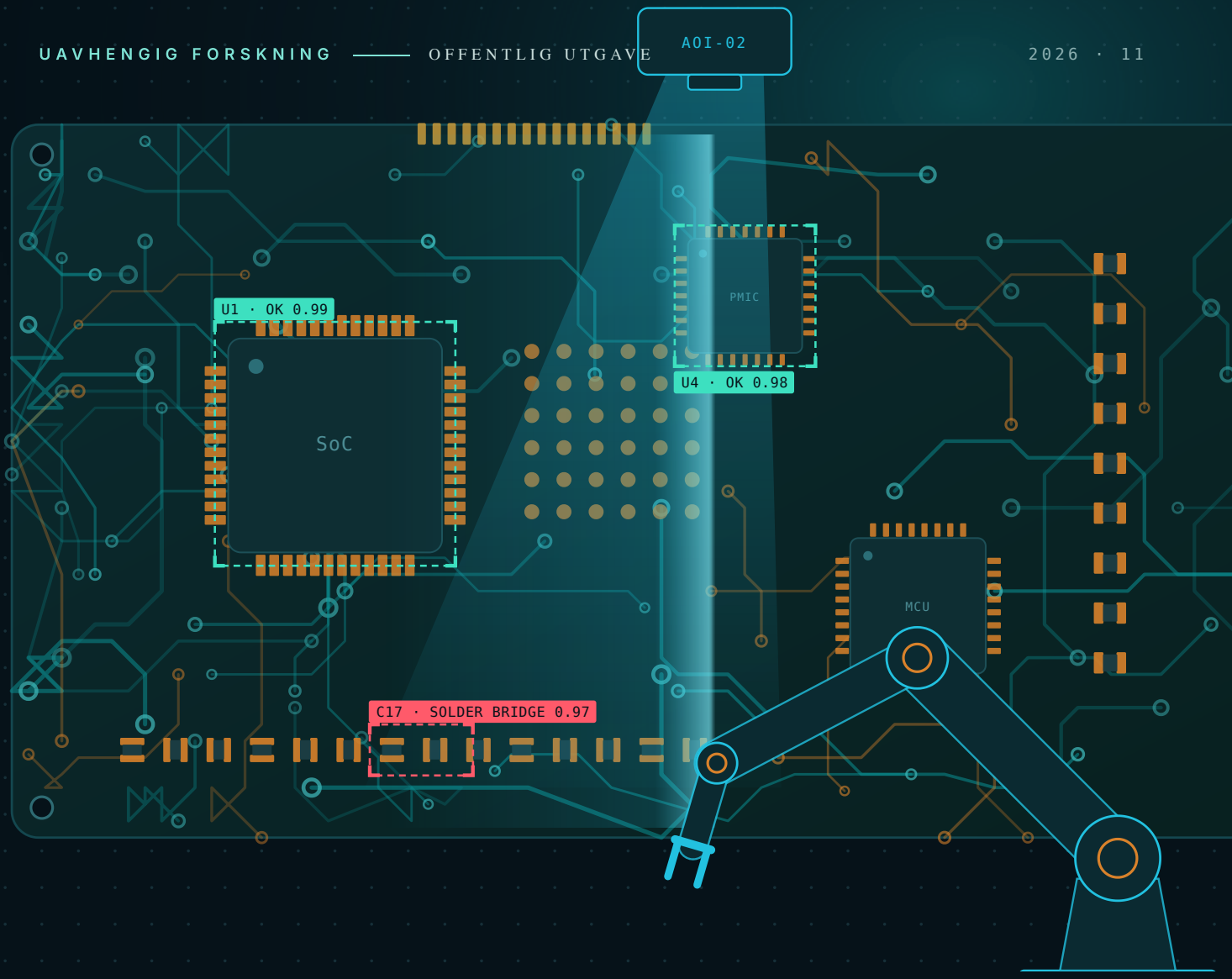


ELEKTRONIKKPRODUKSJON · AI-INSPEKSJON × ROBOTIKK

AI-visuell inspeksjon og robotarmer i elektronikkfabrikker

Vurdering 2026

Vanlige konfigurasjoner · Regional gjennomførbarhet · Lokal utrulling · Lærer–elev-modeller
Avkastning på investering · Konkurranse · Hva bare en pilot kan svare på

HOVEDFUNN

AI-gjennomgang er verdt å gjøre, i snever forstand. Embodied robotarmer kan vente.

Tan Shuai
tanshuai.comi@tanshuai.com

Data per 11. november 2026
Fastlands-Kina · Taiwan · Sørøst-Asia · USA

INNHOLD

Innhold



◆ **SAMMENDRAG** — Verdt å gjøre, men snevert, nå og med partnere — 4

01 **Linjen i dag** — Hva elektronikkfabrikker kjører i dag, og hva som er teknisk gjennomførbart — 8

02 **Lokal utrulling** — Å kjøre AI inne i fabrikken: renrom, luftgap, maskinvare og kvantisering — 16

03 **Lærer og elev** — Lærer og elev: frontmodeller kan forhåndsmerke inspeksjonsdata; for robotarmer er de virkelige lærerne fortsatt mennesker — 25

04 **Kostnader og tilbakebetaling** — Kostnader og tilbakebetaling: AI-granskning betaler seg på måneder, maskinmating på år, utplukkingsarmer aldri — 34

05 **Marked og konkurranse** — Marked, konkurranse og hvordan kjøpere betaler — 42

06 **Embodied AI** — Embodied AI: først endrer den hvordan roboter læres, senere hvordan de beveger seg — 51

07 **Åpne spørsmål** — Hva den offentlige dokumentasjonen ikke kan svare på, og hvordan en pilot avgjør det — 57

◆ **Om forfatteren** — Tan Shuai — 60

Konfidensmerker i teksten

verified to eller flere uavhengige kilder, eller to uavhengige primærdokumenter, stemmer overens

single source kun én kilde, inkludert primærdokumenter

vendor claim publisert av selgeren av produktet eller tjenesten

estimate forfatterens egen beregning eller slutning fra siterte data

Slik leser du denne rapporten

Nøkkeltall bærer et konfidensmerke. **verified** betyr at to eller flere uavhengige kilder stemmer overens, eller at to uavhengige primærdokumenter stemmer overens; **single source** betyr bare én kilde (inkludert primærdokumenter); **vendor claim** betyr publisert av selgeren av produktet eller tjenesten, selv om den er mye gjentatt; **estimate** betyr min egen beregning eller slutning fra siterte data, med dokumenterte inndata. «Primær» beskriver kildetypen (årsmelding, prospekt, forskrift, standard, offisiell statistikk); det endrer ikke konfidensklassen. Der kilder er i konflikt, oppgis tallene side om side og blir ikke midlet. Valutaomregning bruker People's Bank of Chinas sentralkurs av 9. november 2026, $1 = RMB 6.7188 (Xinhua), og for Taiwan SMBCs fiksering samme dag, $1 = NT$32.12 (SMBC). Lønns- og subsidiedata for fastlandet bruker Perlefloddeltaet, der elektronikkproduksjonen er mest konsentrert (offisielle lønnsdata fra Shenzhen), med Yangtsefloddeltaet til sammenligning. Påstander fra bransjeintervjuer behandles som påstander som skal kontrolleres, aldri som fakta. Spørsmål den offentlige dokumentasjonen ikke kan svare på, er samlet i kapittel 7.

FIGURER

Figurer



01	Hvor AI passer på fabrikkgulvet i dag	6
02	En volum-SMT-linje: menneskene sitter bak AOI-en	9
03	Omtrent 97–99% av AOI-alarmene er falskutrop	10
04	Et generelt kamera er ingen AOI	12
05	En air-gapped utrulling	16
06	Ren- og ESD-soner avgjør hvor beregning kan stå	17
07	Hastighetsgrenser per bruk	19
08	On-prem beregning: minne, pris og hva som lovlig kan kjøpes i fastlands-Kina	20
09	Én stor arbeidsstasjon gir ingen delt assistent	21
10	Kvantisering: 8-bit er trygt, 4-bit avhenger av resepten, 3-bit og lavere ryker ut	22
11	Lærerkandidater: generiske poeng doblet seg på en generasjon	26
12	«Er det en defekt?» er fortsatt vanskelig for generelle modeller	27
13	To lærer–elev-pipelines	28
14	Hvor stort det gylne settet må være	29
15	Edge-policylatens	31
16	Robotelever: nær produksjonsgrad bare når stasjonen er strukturert	32
17	Hva én inspeksjons- eller håndteringsarbeider koster per år	35
18	Prisstigen	36
19	Tilbakebetaling per region	38
20	Hvorfor plukking av dårlige kort ikke betaler seg	39
21	USA-toll på roboter, AOI og edge-datamaskiner	40
22	Markedsstørrelser skiller seg mer enn tidobbelt	43
23	Bottom-up: 40 000 linjer → 16 000 servicebare → ≈RMB 320M engangs	44
24	Retrofitvinduet er omtrent fem år	45
25	Gjennomgangskilen er omgitt på tre sider	45
26	Marginen sitter i produktifisert maskinvare	47
27	Hvordan kjøpere betaler	48
28	Days sales outstanding	49
29	Embodied-AI-stacken	51
30	Hva robotutrullinger i fabrikk faktisk offentliggjør	52
31	Netto erstattet arbeid per robot	53
32	Tidslinje for standarder og regulering	55

SAMMENDRAG

Verdt å gjøre, men snevert, nå og med partnere

HOVEDFUNN

Konklusjonen: For elektronikkfabrikker er det som er modent i 2026, og som fabrikker vil betale for, AI-gjennomgang og nisjeinspeksjon på AOI-bilder, pluss programmert robot-mating, ikke «en stor modell som ser på kort pluss embodied roboter som plukker dem».

0.46–0.65 år

Tilbakebetaling for AI-gjennomgang i USA, Taiwan og Shenzhen, ved RMB 50k per linje

≥95% · <20 PPM

Dekning og escapes i en topp-EMS' interne AI-gjennomgang: listen for salg til tier-1 EMS

100–700 RMB M

Servicebart tredjeparts retrofitmarked i Fastlands-Kina, engangs, over et vindu på omtrent fem år

0 poeng

Publiserte industrielle defektbenchmarkpoeng for november 2026-flaggskipene: lærere, ikke inspektører

1.1–1.5 år

Tilbakebetaling for en cobot-matingcelle over to skift i USA, Taiwan og Shenzhen

0.003–0.14 FTE

Arbeid en utplukkingsrobotarm sparer per skift på en SMT-linje: den betaler seg aldri

Inspeksjon
ROI

Veiledet dyp læring brukt på bildene en eksisterende AOI allerede produserer, kutter falskutrop eller granskningstimer med 60–90% (vendor claims). DBG, et stort EMS-selskap, rapporterer i årsmeldingen sin at det interne systemet dekker mer enn 95% av flaggede komponenter med escapes under 20 PPM og mer enn 85% mindre menneskelig granskning; det er listen et tredjepartsprodukt må klare for å selge til tier-1 EMS. Ved RMB 50 000 per linje betaler AI-gjennomgang seg innen ett år i Fastlands-Kina, Taiwan og USA (Shenzhen 0.65 år, Taiwan 0.53, USA 0.46), cobot-matingceller på 1.1–1.5 år og Sørøst-Asia på 2–6 år, mens en robotarm som legges til en SMT-linje bare for å plukke ut dårlige kort aldri betaler seg noe sted, fordi plukking over 90% first-pass yield bare tar 0.003–0.14 av en person per skift estimate.

Modeller
ROBOTER

Multimodale store modeller kan fortsatt ikke være den primære inspektøren. Per 11. november 2026 har ingen av flaggskipene (Claude Fable 5.2, GPT-6.x, Gemini 4, Qwen3.8 og andre) en publisert poeng på noen industriell defekt- eller PCB-benchmark, og i ACM MM 2026 PCBA-utfordringen oppnådde de fem beste systemene, trent på 6 600 ekte linjebilder, bare 77–82 på «er det en defekt». Frontmodeller kan forhåndsmerke og triagere. For robotarmer er de virkelige lærerne fortsatt menneskelige demonstrasjoner, forsterkende læring på ekte roboter og simulering: elever når 90–99.4% ved fikserte, strukturerte stasjoner, men bare 30–80% på bevegelige samlebånd, usette deler og nye produkter; en frontagent fullførte 104 av 108 kvasistatiske oppgaver på ekte robot, med 280 sekunder hver.

Lokal
UTRULLING

Air-gapped utrulling i rene, kontrollerte fabrikker er fullt praktisk. En vanlig SMT-hall er vanligvis et ESD-beskyttet område snarere enn et klassifisert renrom: en vifteløs Jetson- eller innenlandsk NPU-boks står ved linjen, og GPU-servere står i serverrommet. Under NVIDIAs egne eksportklassi-

fiseringer kan Jetson Thor og DGX Spark selges til en vanlig fastlandsfabrikk uten US-lisens; DGX Station kan ikke. Presisjon er som standard 8-bit; leverandørens kvantiseringsbevisste 4-bit er brukbart med forbehold; 3-bit og lavere er uaktuelt.

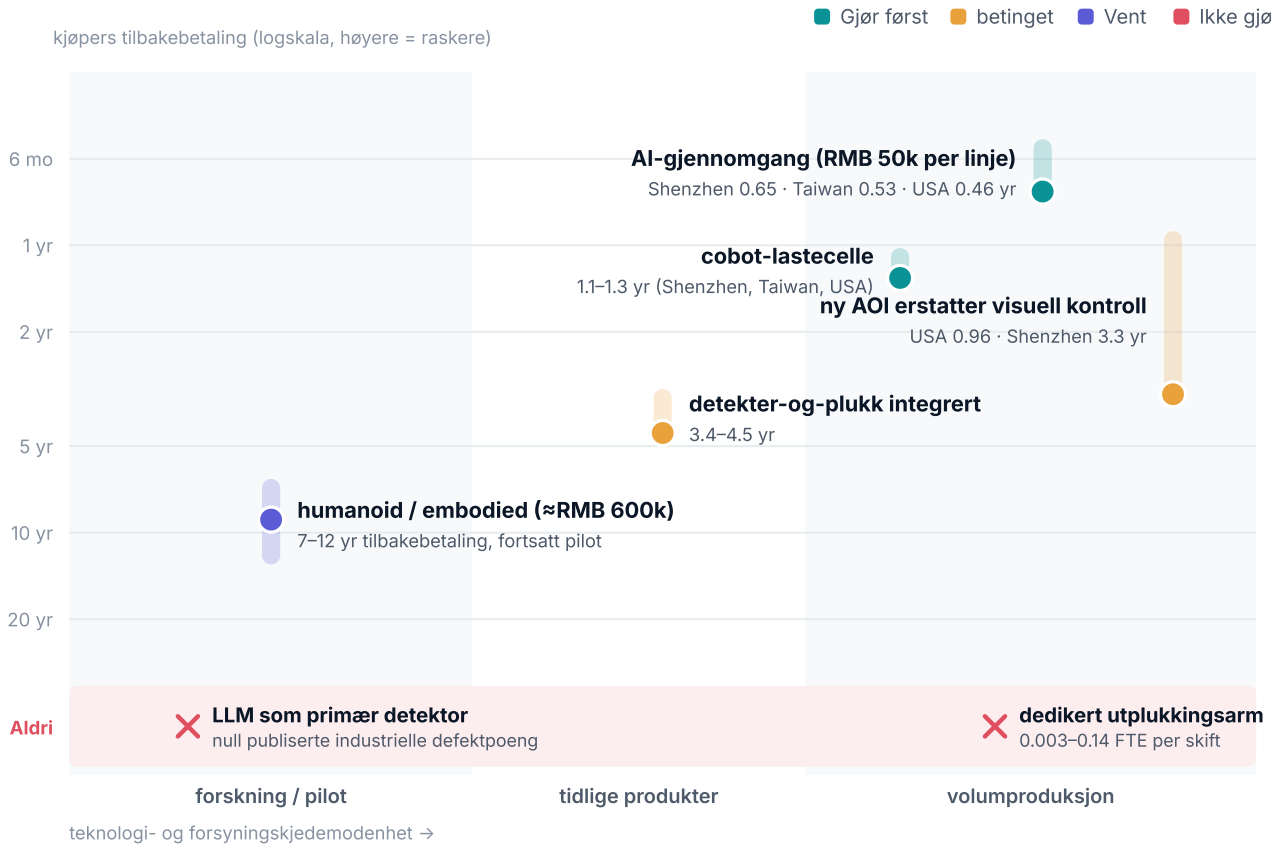
Marked
KONKURRANSE

Markedet er reelt, men begrenset. Fastlands-Kina har omtrent 35 000–46 000 SMT-linjer og omtrent 26 000–66 000 inline-AOI-er som ikke kjører AI; det servicebare markedet for tredjeparts retrofit er omtrent RMB 100–700 millioner engangs, over et vindu på omtrent fem år, og AOI-produsenter har begynt å bygge granskning inn i nye maskiner estimate. Konkurransen er allerede sterk: granskningskilen er omringet av AOI-produsenter, interne EMS-team og plattformleverandører, og marginen sitter i produktifisert maskinvare.

Spørsmål	Vurdering	Nøkkeltall (kilder og konfidens i kapitlene)
1 **Hva kjører vanlige elektronikkfabrikker i dag?**	En volum-SMT-linje er «3D SPI + AOI-er før og etter reflow + stikkprøvebasert røntgen + ICT/FCT + reparasjon og visuell inspeksjon», med menneskelig granskning bak hver AOI. AI kommer inn i granskningen på tre måter: innebygd av AOI-produsenter, bygd internt av topp-EMS, eller lagt til av tredjeparter. Montasje og håndtering bruker armer, SCARA-er og AMR-er med innlært persepsjon og programmert bevegelse; embodied AI er fortsatt i pilotfase	Takt ≈13–22 s per kort; én ODM budsjetterte RMB 30.6M i AOI for 18 linjer, ≈3–5 AOI-er per linje; ≈97–99% av AOI-alarmene er falskutrop
2 **Hvor langt rekker teknologien, kan penger kjøpe null mennesker, og hva kan hver region bruke?**	Inspeksjon er industriell kvalitet som «dedikert detektor + AI-gjennomgang + noen få mennesker»; selv ubegrensede penger når ikke null mennesker. Fastlandet bruker bare innenlandske modeller eller lokale åpne vekter; private fabrikker i Taiwan kan bruke nesten hva som helst, men fabber og kritisk infrastruktur utelukker fastlandsmerker i praksis; Sørøst-Asia har færrest restriksjoner; USA har hele stakken pluss toll og sertifisering	AI-gjennomgang kutter falskutrop eller granskningstimer 60–90% vendor claims; et 4K-kamera over et helt panel ≈78 µm per piksel mot 8–20 µm for en AOI; fastlandslagde armer betaler ≈40% kombinert US-toll, AOI-lignende instrumenter ≈37.5%
3 **Kan den kjøre lokalt i renrom og air-gapped fabrikker, på hvilken maskinvare, raskt og nøyaktig nok?**	Ja; begrensningene er prosessuelle, ikke beregningsmessige. En vifteløs boks ≤100 W ved linjen, GPU-servere i serverrommet; Thor for stasjoner, en Mac Studio eller innenlandsk enhet for fabrikk-Q&A ingen DGX Station på fastlandet. Hastighet er god når modellstørrelsen passer oppgaven; behold 8-bit, bruk leverandørens kvantiseringsbevisste 4-bit med forbehold, aldri 3-bit	≈0.5 s per beskjæring for en 8B synsmodell på Thor estimate; Thor-devkit RMB 59 104 i Kina; DGX Station $123–140k, 139 tokens/s totalt for 32 brukere på en 743B-modell
4 **Kan det nyeste flaggskipet fungere som lærer over dusinvis av runder og destilleres til lokale elever, også for fleksibel produksjon og robotarmer?**	Inspeksjon: forhåndsmerking og triage; auto-aksepter læreretiketter bare for klasser med «≥299 defekter, null misser» på et gyllent sett. Roboter: lærerne er fortsatt menneskelige demonstrasjoner, RL på ekte roboter og simulering, og elever nærmer seg produksjonsgrad bare ved strukturerte stasjoner	Ingen november-flaggskip har en industriell defektpoeng; elever 97.8% på trente kontakter, 29–67% på usette, 37–59% på bevegelige samlebånd; frontmodeller som driver ledd direkte 0–5.5%
5 **Hva koster det, og hva er ROI per region?**	AI-gjennomgang betaler seg innen ett år i Fastlands-Kina, Taiwan og USA; cobot-mating 1.1–1.5 år; Sørøst-Asia 2–6 år; en utplukkingsarm på en SMT-linje betaler seg aldri	RMB 50k granskningsboks per linje: Shenzhen 0.65 år, Taiwan 0.53, USA 0.46, Vietnam 2.2 estimate; akseptabel tredjeparts tilleggspris ≈RMB 10–30k per linje engangs eller RMB 5–10k i året

Spørsmål	Vurdering	Nøkkeltall (kilder og konfidens i kapitlene)
⑥ **Er konkurrentene allerede sterke nok, og er det bedre å starte nå eller vente?**	Etterspørselen er ikke mettet, men tilbudet er overfylt; OEM-AI begrenser retrofitvinduet til omtrent fem år. Gå inn nå med et snevert produkt gjennom kanalpartnere, uten å kjempe direkte mot plattformleverandører; vent på 2027–2028 produksjonsdata før du satser på embodied AI	≈26–66k inline-AOI-er uten AI på fastlandet, som krymper 8–12% i året estimate ; den ledende AI-synsleverandøren hadde i 2025 en omsetning på RMB 1.086B og et netto tap på RMB 991M; den nærmeste børsnoterte sammenlignbare selskapets 3C-omsetning falt 79.9% år over år i januar–mai 2026

FIG 01 **Hvor AI passer på fabrikkgulvet i dag** ***AI-gjennomgang av AOI-falskutrop og cobot-mating er modne og betaler seg raskt; en dedikert utplukkingsarm og en LLM som primær inspektør er ikke verdt bryet.***



Kilde: Tilbakebetaling fra forfatterens modell (kapittel 4); modenhetsaksen er forfatterens vurdering.

Hva man bør bygge og ikke bygge

	Anbefaling
Bygg	AI-gjennomgang og defektklassifisering på AOI-bilder, for mellomstore EMS med eldre eller blandet-merkede AOI-er og ingen AI-oppgradering fra AOI-produsenten; inspeksjon SMT-AOI-en dekker dårlig (gjennomhulls- og håndlodding, konform belegg, kabler, kontakter og FPC-er, sluttutseende og skruer); cobot-laste- og -losseceller der hver del håndteres (testfiksturer, brett, pakking), med inspeksjon som et tillegg; et gyllent sett og skyggemodus på en ekte linje som grunnlag for enhver påstand
Ikke bygg	En generell «AI pluss robotarm»-plattform; en multimodal LLM som primær inspektør; en robotarm på en SMT-linje bare for å plukke ut NG-kort (en NG-avleder eller -buffer gjør det); humanoider og robots-as-a-service; frontmodeller inne i en kontrolløkke eller en sikkerhetsfunksjon; hauger av GPU-minne ved stasjoner for å kjøre «fullverdige» modeller; direkte salg til fabber, taiwansk kritisk infrastruktur eller USAs forsvarselektronikkfabrikker

	Anbefaling
Hvem man bør samarbeide med	AOI-produsenter og distributører (kanal- og bildegrensesnitt; distributører brakte inn 46.68% av Sinic-Teks omsetning i 2025); cobot-produsenter (sertifiserte armer og sikkerhetsfunksjoner); EMS-selskaper med 10–40 linjer (rullende ekspansjon og forskudd); lokale integratører i Thailand og andre steder (for å møte BOI-reglene om lokal tilknytning)

Linjen i dag

01 Hva elektronikkfabrikker kjører i dag, og hva som er teknisk gjennomførbart

Standard volumlinje: flere AOI-er, stikkprøvebasert røntgen, sentral granskning, 13–22 sekunder per kort

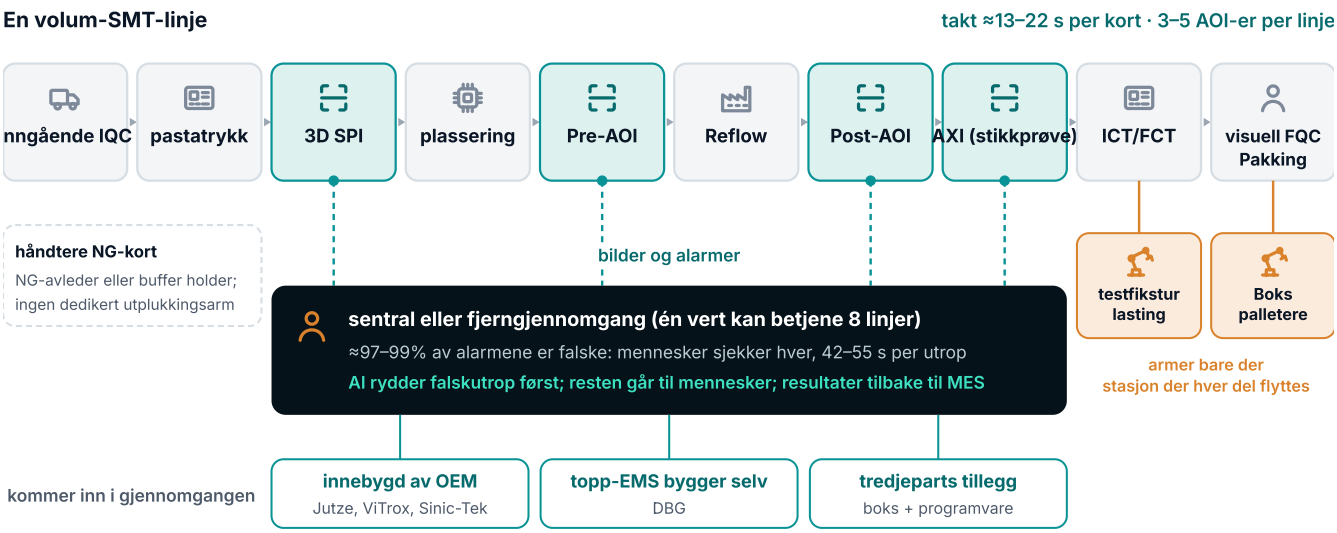
Primærregistreringer fra børsnoterte ODM-er viser hvordan en volumlinje i 2025–2026 faktisk ser ut. Longcheers Hongkong-prospekt (data til 30. september 2025) lister SMT-kvalitetskontrollene sine som «røntgeninspeksjon, loddepastainspeksjon (SPI), automatisert optisk inspeksjon (AOI), PCBA-funksjonstest og visuell inspeksjon»; prosessflyten går pre-reflow-AOI, post-reflow-AOI, stikkprøvebasert røntgen, NG-kort til en SMT-reparasjonsstasjon, og deretter en endelig visuell kontroll. Fabrikkene i Huizhou, Nanchang og Vietnam har 34, 24 og 8 SMT-linjer og 67, 60 og 13 sluttmontasjelinjer (Longcheer H-share prospectus) single source, primary filing. Huaqins bærbare-fabrikk fra 2023 planla 18 SMT-linjer med et AOI-budsjett på RMB 30.6 millioner, omtrent RMB 1.7 millioner per linje, noe som ved innenlandske 3D-AOI-priser er tre til fem AOI-er per linje for pre- og post-reflow-inspeksjon av dobbeltsidige kort (Huaqin A-share prospectus) primary filing + estimate. En volumlinje har derfor vanligvis to til fire granskningspunkter, ikke ett.

Maskinene er allerede høyt automatisert: Huaqin rapporterer 96% automatisering på SMT-linjene sine (Huaqin H-share prospectus) single source. De gjenværende menneskene sitter i granskning, reparasjon, visuell inspeksjon, test og sluttmontasje, og ingen selskap offentliggjør hvor mange granskere det har per linje (kapittel 7).

Tre uavhengige estimater stemmer overens om en takt på **13–22 sekunder per kort** (en miljøregistrering fra 2026 for fem Hebei-linjer (Hebei SMT project EIA), Huaqins bærbare-linjer og et iFactory-tilfelle), eller 7–11 sekunder per side for dobbeltsidige kort estimate. Det vanskelige med en high-mix-linje er ikke volum, men introduksjon av nye produkter og omstilling, og hver omstilling er nok en kaldstart for en AI-modell.

På programvaresiden er AOI-produsentenes egne granskningsstasjoner allerede sentralisert: Sinic-Teks reparasjonsstasjonsprogramvare lar «én vert kontrollere utstyret på åtte SMT-linjer» (Sinic-Tek prospectus) single source, primary, 2022, og etter at Jabil i 2026 tok i bruk ViTrox V-ONE fjernstyring av AOI-er, kunne én ingeniør styre 5–8 AOI-er eksternt mot 1–2 lokalt (AtlasPCB) vendor claim. Et tredjeparts AI-granskningsprodukt kobles direkte til denne sentralstasjonen, gjennom AOI-ens bilde- og resultateksport (IPC-CFX, IPC-HERMES-9852 eller et leverandøreget format). AOI-ene selv legger til AI raskt: Yamaha sier AI-vurderingen tar omtrent ett sekund per kall, hjelper inspektører etter 30 NG-bilder og blir «svært høy» kvalitet etter 1 000 (DirectIndustry/Yamaha) vendor claim.

FIG 02 **En volum-SMT-linje: menneskene sitter bak AOI-en** ***Arbeidet er konsentrert på granskningsplassene etter automatisk optisk inspeksjon. Det er der AI kommer inn.***



Kilde: Offentlige årsmeldinger og leverandørmateriale (Huaqin, DBG, Sinic-Tek med flere); sammenstilt av forfatteren.

Trinn	Vanlige produkter og teknologi	Representative leverandører	Hvor langt AI har kommet
Etter trykking	3D SPI (strukturert lys, moiré, noe lasertriangulering)	Koh Young, Sinic-Tek, TRI	OEM-AI-vurdering av pasta i produksjon; Jutze sier noen modeller er «helt AI»
Etter plassering og reflow	2D/3D AOI med flervinklet, flerfargede LED-er og strukturert lys, 8–20 µm per piksel, 8–12 MP-kameraer	Koh Young, Omron, Saki, TRI, ViTrox, Mirtec, Jutze, Sinic-Tek og andre	OEM-AI-granskning (Jutze i produksjon, ViTrox V510Ai, Sinic-Tek med Tencent); topp-EMS bygger eget (DBG); tredjeparts tillegg
Bunnavsluttede og skjulte skjøter	Stikkprøvebasert eller inline røntgen (AXI)	—	Hikvision hevder Guanlan-modellen feilvurderer røntgenhulrom under 1% av tiden vendor claim
Granskning og reparasjon	Sentral granskningsstasjon (én vert, åtte linjer), fjerngranskning, reparasjonsstasjon	AOI-produsentenes programvare, ViTrox V-ONE	AI lukker falskutrop, folk tar resten
Visuell inspeksjon, IQC, FQC	Mest manuelt; noen AI-utseendemaskiner med robotlasting	SmartMore og andre	Utseendemaskiner i produksjon; ingen offentlige bemanningstall for IQC/FQC
Håndtering og maskinmating	Seksaksearmer og coboter, SCARA, AMR-er, 2D/3D synsstyring	FANUC, ABB, Yaskawa, KUKA, UR, Dobot, JAKA, Elite; Mech-Mind, Hikrobot	Innlært persepsjon pluss programmert bevegelse i produksjon; VLA-er og humanoider fortsatt i pilot
Data og systemer	MES, IPC-CFX, IPC-HERMES-9852	—	Resultater skrives tilbake til MES; støtte i eldre utstyr er ujevn

Kilder for tabellen: registreringene ovenfor og (Koh Young, Jutze 2025 annual report, Sinic-Tek 2025 annual report, ViTrox, Sohu on Hengtianyi, Xinhua on Hikvision).

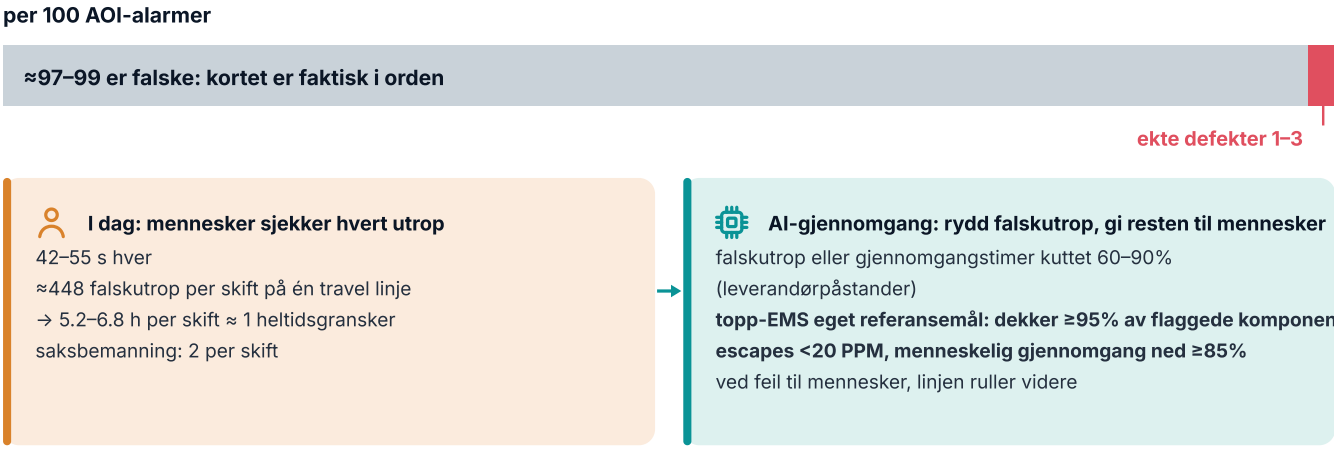
Menneskene sitter bak AOI-en, fordi de fleste alarmene er falske

Inspeksjonskjeden på en vanlig SMT/PCBA-linje går: 3D SPI etter trykking for manglende, utilstrekkelig eller overskytende pasta, brodannelse og forskyvning; AOI etter plassering (valgfritt) og etter reflow for manglende, forskjø-

vede, roterte, reverserte eller «tombstone»-deler og loddevolum; røntgen når det er BGA-er eller andre bunnavsluttede deler; deretter in-circuit- eller flying-probe-test, funksjonstest og manuell visuell inspeksjon. Hver AOI har en granskningsplass og en reparasjonsstasjon bak seg (Sinic-Tek prospectus, JLCPCB) verified. Post-reflow-AOI er den vanligste og beste verdien, men den ser bare form og areal; skjøter, hulrom og fukting under en BGA krever røntgen eller elektrisk test (JLCPCB).

Arbeidskraften sitter i granskningen fordi de fleste alarmene er falske. Én britisk EMS registrerte **54 019 falskutrop mot 544 reelle defekter over 917 delenumre, omtrent 99% av alarmene** (arXiv 2505.03576) single source, preprint; ved Siemens' Rastatt-fabrikk var **97.5% av kortene AOI-en avviste gode** (Siemens case study) vendor claim. En travel linje ser omtrent 448 falskutrop per skift ved 42–55 sekunder hver, 5.2–6.8 timer per skift, omtrent én heltidsgransker (iFactory) vendor claim; hours are an estimate.

FIG 03 **Omtrent 97–99% av AOI-alarmene er falskutrop*I dag bekrefter en person hvert utrop; AI-gjennomgang rydder først de falske og sender resten til mennesker.***



Kilde: Siemens casestudie, iFactory (leverandørpåstand), DBGs årsmelding 2025; gjennomgangstimer estimert av forfatteren.

Fra «granskningstimer per skift = kort × falskutrop per kort × sekunder per utrop» kan én gransker dekke omtrent 0.7–1 volumlinje med en regelbasert AOI. Når AI fjerner 60–90% av falskutropene, ville vurderingstiden alene tillate 2–14 linjer per person, men gange, NG-bufferhåndtering og reparasjonsoverleveringer begrenser lokal dekning til to til fire tilstøtende linjer; bare sentral eller fjerngranskning nærmer seg det teoretiske taket estimate; seconds per call come from one source. Siden vanlige AOI-er skanner 60–80 cm² per sekund (Yunheng 2026 AOI guide), tar en 20×15 cm-side 4–5 sekunder; AI-en behøver bare å behandle de få flaggede områdene innen én takt, ikke skanne kortet på nytt.

Det som allerede virker i produksjon: veiledet klassifisering på AOI-ens egne bilder

Det dokumenterte mønsteret er veiledet dyp-læring-klassifisering på bilder fra en AOI fabrikken allerede eier. Fujitsus tillegg på ADLINKs AOI i Taiwan kuttet falskutrop med **mer enn 90%** (Fujitsu case); OKI har brukt AI siden juli 2026 for å kutte manuell bekreftelse etter AOI med omtrent **80%** i sin egen EMS-virksomhet (Circuits Assembly); Siemens rapporterer 60% mindre manuell analyse, €160 287 spart det første året og en tilbakebetaling på åtte måneder (Siemens case study) all vendor claims. To ting er felles for disse tilfellene: nesten ingen publiserer en escape-rate, og alle beholder folk på de gjenværende alarmene. Det ene tilfellet som oppgir en escape-tall (0.8% ned til 0.3%, mot en kontraktsfestet 0.5%) er internt inkonsistent (iFactory). AI kutter granskningstimer, ikke selve granskningsjobben.

AI-granskning har beveget seg fra tredjeparts tillegg til OEM-funksjoner og interne systemer, noe som endrer hva en nykommer måles mot. Jutzes årsmelding for 2025 lister «3D AOI-oppgradering» som i volumproduksjon og sier den «bruker AI-algoritmer for å fjerne manuell granskning ved inspeksjonssiden» (Jutze 2025 annual report); Sinic-

Tek har utviklet et «multimodalt AI-assistert manuelt granskningssystem» siden 2019 og samutvikler en topp-og-bunn 3D AOI med Tencent, med mål om 99.5%+ defektdekning og 98% gjenkjenningsnøyaktighet (Sinic-Tek 2025 annual report) vendor claims in primary filings. Store EMS-selskaper bygger sitt eget. DBGs årsmelding for 2025 lister «AI-vurdering for visuell inspeksjon» som et fullført prosjekt som «dekker mer enn 95% av AOI-flaggede komponenter», med en «escape-rate under 20 PPM» og «mer enn 85% mindre manuell granskning», fullt sporbar (DBG 2025 annual report) single source, primary. Det er det eneste elektronikkfabrikktallet jeg fant som oppgir dekning, escapes og reduksjon sammen. **Et tredjepartsprodukt solgt til tier-1 EMS må matche DBGs 95% dekning, 20 PPM escapes og 85% reduksjon**; under tier 1 er benchmarken den 50–85% reduksjonen leverandører reklamerer med.

Anomalideteksjon trent bare på gode deler er rask (EfficientAD-S tar 2.2 ms per bilde (EfficientAD, WACV 2024)) men svak på «logiske» anomalier som manglende eller feil deler, og de beste metodene snitter under 60% AU-PRO på den vanskeligere MVTec AD 2 (MVTec AD 2) single source. Manglende og feil deler kontrolleres bedre mot CAD og BOM.

Taket uansett pris: lag av maskiner, svært få mennesker, men ikke null

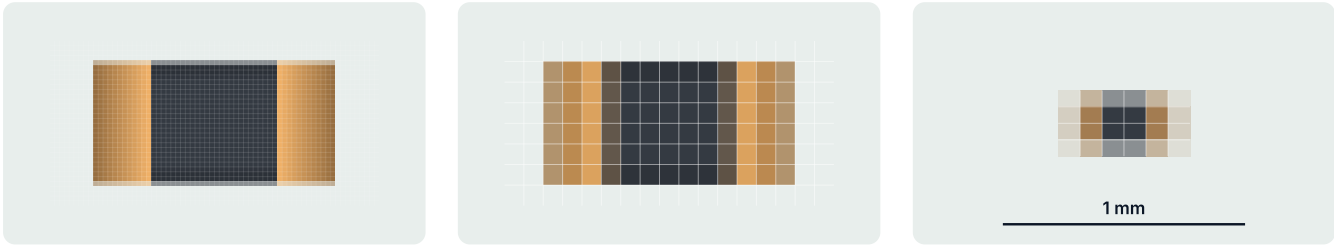
Jeg fant ingen elektronikkfabrikk som fjernet manuell granskning helt og nådde null escapes. Praksis for IPC-A-610 klasse 3 (høypålitelighets-) produkter har fortsatt trente inspektører som avgjør grensetilfellene automatiske terskler ikke kan klassifisere, og én PCB-montør kaller «automatisert inspeksjon kan erstatte menneskelig signering» en «vanlig misforståelse» (PCBCart) single source. Den menneskelige grunnlinjen må også korrigeres: med henvisning til en Sandia-studie fra 2012 fanger én enkelt visuell inspektør i beste fall omtrent **80%** av defektene, to i serie omtrent 96% (Instrumental) single source, cited by a vendor. Den forsvarbare påstanden er «AI pluss en liten menneskelig granskningsløkke slår helmanuell inspeksjon», ikke «AI erstatter inspeksjon».

Uten budsjettgrense er taket 3D SPI, 3D AOI, røntgen, AI-granskning og AI-utseendestasjoner stablet sammen, med folk igjen bare for grensetilfeller, nye produkter og omstillinger, nye defekttyper og kundesignering. Flaggskipfabrikker rapporterer 85–92% mindre manuell inspeksjon (I-Connect007, 2021, WiAdvance) vendor claims; mitt estimat av taket uten budsjett er granskningsarbeid på 8–15% av dagens estimate. Sluttmontasje er fortsatt arbeidsintensiv: automatiseringen av iPhone 16-knappene ble droppet på grunn av høye defektrater (9to5Mac) single source.

Et generelt høyoppløselig kamera er ikke en AOI

Produksjons-AOI-er tar bilde flis for flis ved 8–20 µm per piksel, med 8–12 MP-kameraer, programmerbar flervinklet flerfarget belysning og strukturert-lys 3D (Koh Young, Saki) verified. Bransjen bemerket for lenge siden at en 01005-del bare dekker omtrent 8×16 piksler ved 25 µm per piksel, «muligens ikke nok» for pålitelig vurdering (Circuits Assembly, 2007). Ved samme regnestykke estimate får et 4K-kamera over et 300 mm-panel omtrent **78 µm per piksel**: en imperial 0402-del er omtrent 13×6 piksler, nok for tilstedeværelse og grov forskyvning men ikke skjøteform, og en 01005 er omtrent 5×3 piksler, i praksis usynlig. Å nå 15 µm per piksel over et helt panel ville ta omtrent 330 megapiksler, noe som er grunnen til at AOI-er flytter et kamera over kortet. Når noen sier «0402», sjekk hvilken de mener: metrisk 0402 er imperial 01005, langt utenfor ethvert vidvinkelkamera.

FIG 04 **Et generelt kamera er ingen AOI***Et 4K-kamera over et 300 mm-panel gir omtrent 78 µm per piksel: en 0402-del krymper til ≈13×6 piksler og en 01005 forsvinner.*



AOI: 8–20 µm per piksel
0402 ≈ 100×50 px

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

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. Pikselnettene er illustrative.

På linjer som allerede har AOI-er, er det beste stedet for AI oppå AOI-ens bilder. Bare linjer uten AOI, eller end-of-line-stasjoner, rettferdiggjør en lukket kamerastasjon med kontrollert belysning; et åpent høyoppløselig kamera over linjen er mest utsatt for omgivelseslys og skygger (CSDN).

Montasje og håndtering: innlært persepsjon pluss programmert bevegelse er i produksjon

IFRs rapport for 2026 sier «maskinsyn for inspeksjon og robotstyring er modent og industrielt», mens «generell fysisk AI og læring fra simulering og erfaring ennå ikke er industrielt» (IFR World Robotics 2026 executive summary) single source, primary. Elektronikk er den største robotkjøperen (142 235 enheter installert globalt i 2025, 59% av alle installasjoner i Kina (IFR, IFR China release)), likevel har Kina bare 166 roboter per 10 000 produksjonsarbeidere mot 307 i USA (IFR, April 2026). Mønsteret i produksjon er innlært persepsjon pluss deterministisk bevegelse; end-to-end innlært bevegelse er nær produksjonsgrad bare ved strukturerte stasjoner (kapittel 3 og 6).

Hva hver region kan bruke

Fastlands-Kina støtter seg på innenlandske modeller og lokale vekter; i Taiwan avgjør kundens identitet; i USA avgjør toll og sertifisering.

Element	Fastlands-Kina	Taiwan (private fabrikker)	Sørøst-Asia	USA
Claude (enhver kanal)	Ikke tilgjengelig	Betinget: leverandøren må ikke være majoritetseid eller kontrollert av personer eller enheter i Fastlands-Kina, Hongkong eller Macau	Samme	Samme
GPT og Gemini sky	Ikke tilgjengelig	Tilgjengelig	Tilgjengelig (lister ikke sjekket land for land)	Tilgjengelig
Kinesiske skymodeller (Qwen, DeepSeek, Doubao, GLM, Kimi)	Tilgjengelig (registrerte modeller; å tilby en tjeneste krever registrering)	Lovlig, men fastlandshosting passerer sjelden kundens sikkerhetsgjennomgang; forbudt i offentlig sektor, statlige foretak og kritisk infrastruktur	Lovlig (internasjonale noder)	Lovlig, streng gjennomgang

Element	Fastlands-Kina	Taiwan (private fabrikker)	Sørøst-Asia	USA
Lokale åpne vekter av kinesisk opphav (Qwen, GLM, DeepSeek)	● Førstevalg; privat in-fabrikk-utrulling krever ingen modell- eller algoritmeregistrering (advokatlesing)	● Tilgjengelig (offentlig sektor begrenset)	● Tilgjengelig	● Betinget: opphavsgjennomgang
Lokale åpne vekter av ikke-kinesisk opphav (f.eks. Gemma 4)	● Tilgjengelig	● Tilgjengelig	● Tilgjengelig	● Tilgjengelig
Klassiske ikke-generative synsmodeller	● Tilgjengelig (utenfor de generative AI-reglene)	● Tilgjengelig	● Tilgjengelig	● Tilgjengelig
Jetson Thor/T5000/T4000, DGX Spark	● Tilgjengelig: NVIDIA klassifiserer dem 5A992.c, ingen US-lisens nødvendig for en vanlig sivil fabrikk; Spark har en Kina-SKU; kanalpriser omtrent 40–60% over US-liste	● Tilgjengelig	● Tilgjengelig	● Tilgjengelig
RTX 5090, RTX PRO 6000, DGX Station	● Ikke tilgjengelig: 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)	● Tilgjengelig	● Tilgjengelig, men kjøpere med hovedkontor i eller ultimat eid fra Kina regnes som kinesiske; Malaysia krever tillatelser for eksport og omlasting	● Tilgjengelig
Mac Studio M5 Ultra 512 GB	● Tilgjengelig (Apple China-butikk)	● Tilgjengelig	● Antas tilgjengelig	● Tilgjengelig
Huawei Ascend og andre kinesiske akseleratorer	● Tilgjengelig	● Høy risiko	● Høy risiko	● Ikke brukbar i praksis
Fastlandslagde faste armer, AOI-er, industrikameraer	● Tilgjengelig	● Importerbar; siden 1. desember 2025 kan myndighetene handle direkte mot fastlandsmerker, og etater, statlige foretak, visse stiftelser og kritisk-infrastrukturleverandører kan begrenses eller utestenges fra å bruke dem	● Tilgjengelig	● Lovlig, men kombinert toll er omtrent 40% på armer, 37.5% på AOI-lignende instrumenter og kameraer, 25% på edge-datamaskiner
Utenlandske AMR-er og mobile manipulatorer	● Tilgjengelig	● Tilgjengelig	● Tilgjengelig	● Nye modeller krever betinget godkjenning fra US Department of War
Hikvision- og Dahua-kameraer	● Tilgjengelig	● Privat bruk tillatt, offentlig sektor forbudt	● Tilgjengelig	● Nye FCC-godkjenninger blokkert; forbudt for føderale entreprenører
Robotsikkerhet og sertifisering	● GB 11291.1-2011 og GB 11291.2-2013 obligatoriske, under revisjon siden april 2026; CR-sertifisering frivillig; GB/T 39404-2026 dekker kontrollerens cybersikkerhet	● Inngjerding som standard; inngjerdsløst samarbeid må møte CNS 14490 eller ISO 10218 med skriftlig vurdering	● Per kunde og eksportdestinasjon, bekreftet skriftlig før forsendelse	● Elektrisk utstyr må være «godkjent» (i praksis et NRTL-merke eller feltvurdering); UL 1740 krever at robotceller møter ISO 10218-2 siden 11. august 2025

Element	Fastlands-Kina	Taiwan (private fabrikker)	Sørøst-Asia	USA
Fabber og kontrollert-produkt-fabrikker	—	🟡 TSMCs innkjøpskontrakter krever SEMI E187; fastlandsutstyr fjernes fra 2 nm-fabber	—	🔴 ITAR tekniske data inkluderer «fotografier»; avslagspolicy for Kina; fastlandskontrollerte leverandører kan i praksis ikke betjene forsvarelektronikkfabrikker

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

Fastlands-Kina: en komplett stakk, men verken lukkede frontmodeller eller de sterkeste enkeltmaskinene kan komme inn. Registrerte innenlandske skymodeller, lokale åpne vekter (Qwen under Apache-2.0, GLM og DeepSeek under MIT), klassiske synsmodeller og innenlandske AOI-er, kameraer og coboter er alle brukbare. NVIDIAs egne eksportklassifiseringer lister Jetson AGX Thor-settet, T5000/T4000-modulene og DGX Spark som 5A992.c, så en vanlig EMS- eller PCBA-fabrikk trenger ingen US-lisens, mens RTX 5090 og RTX PRO 6000 (4A090.a) og DGX Station (5A992.z, TPP 58 484) ikke kan kjøpes på fastlandet (NVIDIA export classifications, eCFR) primary; the legal reading is mine and not legal advice. RS China lister Thor-settet til RMB 59 104 inkludert avgift, omtrent 1.4 ganger US-listeprisen (RS China).

Fire advokatfirmaer er enige om at en privat in-fabrikk-utrulling til fabrikkens eget bruk ikke krever modell- eller algoritmeregistrering, mens en hostet flerleietjeneste befinner seg i en gråsone (Covington, FPF, JunHe) verified: consistent readings, no official ruling. GB 11291.1-2011 og GB 11291.2-2013 forblir den obligatoriske robotsikkerhetsgrunnlinjen og har vært under revisjon siden 28. april 2026; GB/T 39404-2026 om kontrollerens cybersikkerhet trådte i kraft 1. november 2026 (revision plan 20262542, GB/T 39404-2026) single source, primary. Inntil den nye GB 11291 kommer (2028–2029 estimate), design etter ISO 10218:2025 og behold GB 11291-samsvarsfiler.

Taiwan: private fabrikker kan bruke nesten hva som helst; kundens identitet avgjør. Front-API-er, åpne vekter og fastlandsutstyr er alle lovlige (data.gov.tw). Men siden den endrede Cyber Security Management Act trådte i kraft **1. desember 2025**, kan myndighetene handle direkte mot produkter som er «fastlandsmerker» og begrense dem hos statlige organer, statlige foretak og kritisk-infrastrukturleverandører (Laws & Regulations Database, review rules) verified; listen over kritisk-infrastrukturleverandører er konfidensiell; og TSMC har skrevet SEMI E187 inn i innkjøpskontraktene sine (CNA) og rapporteres å fjerne fastlandsutstyr fra sine 2 nm-fabber (TrendForce) single source. Private EMS-, montasje- og små PCBA-fabrikker er åpne; fabber, parkforsyninger og myndighetsrelaterte kunder er lukket i praksis. En fastlandskontrollert leverandør bør designe etter E187, sende en programvareliste (SBOM), tilby offlineoppdateringer og aldri koble til eksternt fra fastlandet.

Sørøst-Asia: færrest restriksjoner; problemene er langsom tilbakebetaling og kinesiske morselskaper. Alt er brukbart, men USAs veiledning av 31. mai 2026 krever lisens for å eksportere avanserte beregningsvarer til enheter med hovedkontor i, eller ultimat eid fra, Kina selv når mottakeren er i et tredjeland (Greenberg Traurig) single source, og Malaysia har egne tillatelser for avanserte chips (Fulcrum). Etterspørselen beveger seg hit: TrendForce forventet at Thailand, Vietnam og Malaysia skulle nærme seg 50% av server-ODM SMT-kapasitet i 2026 (TrendForce) single source, a 2023 forecast.

USA: hele stakken, til prisen av toll, sertifisering og kontrollerte data. Under kapittel 99 i USAs tolltariff betaler en fastlandslaget robotarm omtrent 40% i kombinert toll, AOI-lignende instrumenter og kameraer omtrent 37.5% og edge-datamaskiner 25%; laget i Malaysia eller Indonesia omtrent 10–12.5%, i Vietnam, Thailand eller Filippinene 12.5–15%, i Taiwan 10%, med datamaskiner på 0% utenfor fastlandet (USITC HTS chapter 99, Great Lakes Customs Law) primary schedule + estimate. USTRs melding av 10. november forlenget 178 gamle seksjon 301-unntak til 9. januar 2027, ingen av dem som dekker roboter, inspeksjonsinstrumenter, kameraer eller datamaskiner (Federal Register 2026-22925) verified.

OSHA krever at arbeidsplassens elektriske utstyr er «godkjent» (i praksis et NRTL-merke, eller en feltvurdering på omtrent $3 000 til $20 000+ per celle), og siden 11. august 2025 bringer UL 1740 hele robotceller under ISO 10218-2 (eCFR 1910.399, Intertek, Industrial Monitor Direct). Forsvarselektronikk er lukket for fastlandskontrollerte leverandører i praksis: ITAR behandler produksjonsfotografier som tekniske data og anvender en avslagspolicy overfor Kina (eCFR 22 CFR 120.33, eCFR 22 CFR 126.1). En US-rettet celle bør bruke en fast arm, ubegrensede kameraer og lokale eller US-sky-modeller, med sluttmontasje i Sørøst-Asia der det genuint endrer opphav.

Uansett budsjett skiller taket seg per region, men når ingen steder null mennesker estimate. Fastlandet kan stable innenlandsk 3D SPI, 3D AOI, røntgen, AI-granskning og et internt AI-team for å bringe granskningsarbeidet ned til 8–15%, men uten lukkede frontmodeller eller DGX Station må fabrikkomfattende LLM-Q&A kjøre på lokale åpne vekter og innenlandsk beregning (en fullverdig modell for mange brukere trenger minst to Ascend-servere med åtte kort). Taiwan og Sørøst-Asia kan teknisk bruke det beste av alt, inkludert front-API-er og DGX Station, begrenset av kundeidentitet, fastlandsmerker og kinesiske-morselskapslisensiering. USA har ingen tekniske grenser, til prisen av toll, NRTL-sertifisering og personellseparasjon for kontrollerte data. Overalt blir folk igjen for grensetilfeller, nye produkter og omstillinger, nye defekttyper og kundesignering.

Et produkt som selges på tvers av regioner trenger derfor to modellstakker: en fastlandsstakk på registrerte innenlandske modeller eller lokale åpne vekter, og en oversjøisk stakk på front-API-er pluss en lokal modell av ikke-kinesisk opphav. Anthropics policy gjeldende fra 12. november 2026 vurderer etter eierskap og kontroll, så en fastlandskontrollert leverandør kan ikke bygge på Claude selv gjennom et oversjøisk datterselskap, en skyplattform eller en forhandler (Anthropic) single source, primary.

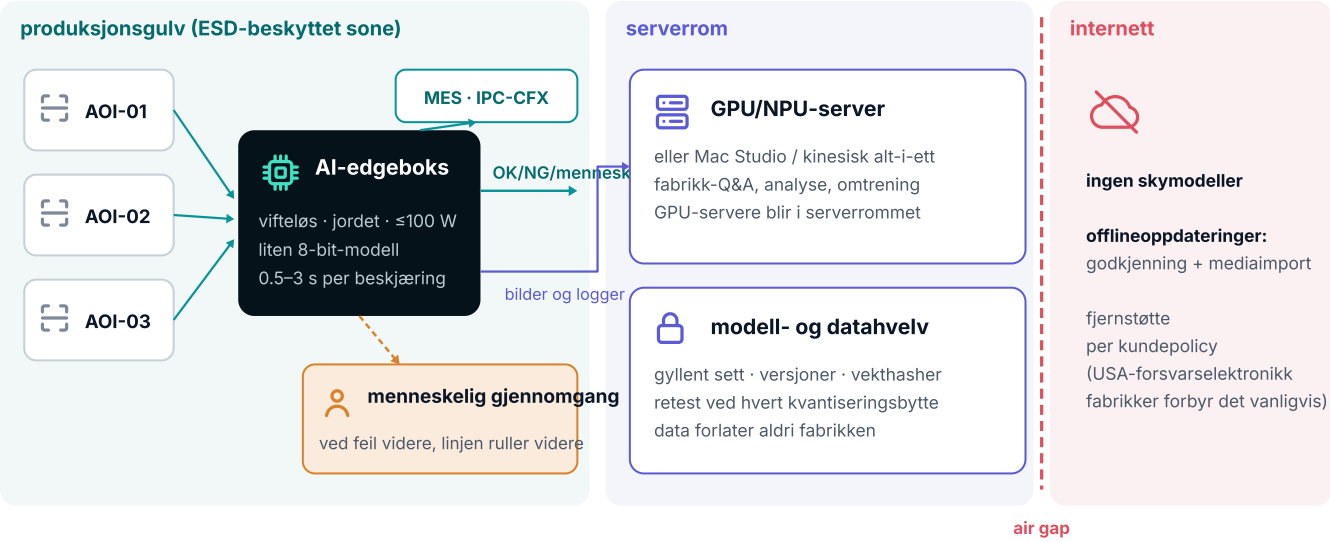
Lokal utrulling

02 Å kjøre AI inne i fabrikken: renrom, luftgap, maskinvare og kvantisering

I en air-gapped fabrikk er begrensningen prosess, ikke beregning

Helt offline synsinspeksjon er rutine; begrensningene er prosessuelle. Kinas MIIT-veiledning fra 2024 om sikkerhet i industrielle kontrollsystemer ber om soning, brannmurer eller datadioder, revidert tidsbegrenset fjernvedlikehold, skannede medier og testede patcher (guide text) single source; halvlederkunder legger til SEMI E187 og E188, som TSMC skrev inn i innkjøpskontraktene sine i 2023 (SEMI, TXOne) verified. Resultater går tilbake til MES over IPC-CFX og kortidentitet reiser over IPC-HERMES-9852 (Hermes standard). Den vanlige fellen er lisensiering som antar et nettverk: Roboflows offline-modus cacher vekter på en 30-dagers leieavtale (Roboflow). Et offline produkt trenger offline-aktivering, signerte oppdateringer på skannede medier, en programvareliste (SBOM), USB-kontroll, lokale revisjonslogger og en CFX-adapter.

FIG 05 **En air-gapped utrulling** ***En liten boks ved linjen bestemmer; serverrommet svarer på spørsmål; ved feil faller de tilbake på mennesker og data forlater aldri fabrikken.***



Kilde: Forfatterens syntese (kapittel 2).

Lukkede frontmodeller kommer i praksis ikke inn i en fabrikks serverrom. Gemini tilbys air-gapped bare gjennom Google Distributed Cloud for myndigheter og store foretak (Google Cloud); Microsofts Foundry Local åpner store modeller bare for «kvalifiserte kunder» og nevner ingen GPT-modell (Microsoft); Claude tilbys bare i skybaserte myndighetsregioner, uten lokal utrulling (Claude for government); ingen av de tre tilbys i Fastlands-Kina. Fabrikkutstyr må designes rundt åpne vekter; frontmodeller kan i høyden merke eller forklare bilder ved hovedkontoret, offline, når kunden tillater at anonymiserte bilder forlater fabrikken.

Serverrommaskinvare bør behandles som kontorklasse: Mac Studio er klassifisert for 10–35 °C og DGX Spark for 5–30 °C, med rapporter om termiske nedstengninger under vedvarende inferens (Apple, NVIDIA forum) verified. Linjen får vifteløse industribokser, som en Jetson T4000 industri-PC klassifisert for −25 til 60 °C (AAEON) vendor claim. Fordi én times nedetid koster en stor fabrikk $36 000 til $2.3 millioner (Siemens True Cost of Downtime 2024)

vendor survey, må hvert AI-trinn på linjen **feile åpent (fail open)**: ved enhver feil faller det tilbake til den eksisterende AOI-pluss-mennesker-prosessen.

Ren- og ESD-soner: en vanlig SMT-hall er et ESD-beskyttet område, ikke et klassifisert renrom

Det som begrenser hvor beregning kan stå, er statisk elektrisitet og luftstrøm, ikke renhetsklasse. De fleste SMT-montasjeområder er «ikke formelt partikkelkontrollerte», og verdensklassefabrikker «kan nå eller nærme seg ISO 8» (Circuitnet expert panel) single source. Medisinsk PCBA og kameramoduler ligger typisk i ISO 7, bar die og optisk justering i lokale ISO 5–6 (PCBSync) single source; renromsbyggere oppgir ISO 7–8 for SMT generelt (Angstrom Technology) vendor claim.

Det universelle kravet er ESD-kontroll: ANSI/ESD S20.20-2021 krever isolerte ledere under 35 V og holder sterkt-felt-isolatorer minst 30 cm fra sensitive enheter (ANSI, ESDA, Desco) verified. Fabber plasserer beregning i sub-fab eller datasenteret, i væskekjølte rack som møter SEMI S2/S8/F47 (Prodrive) vendor claim.

Hvorvidt utstyr kan gå inn i et rent område, testes for partikkelutslipp under ISO 14644-14; roboter har slike sertifikater (Fraunhofer IPA), men jeg fant ingen industri-PC, edge-AI-boks eller GPU-server med ett. CE, FCC og IEC 60068 viser at en boks tåler miljøet sitt, ikke at den får gå inn i et ESD-beskyttet område; kundens ESD-koordinator sjekker det. GPU-servere kommer aldri på linjen: en DGX Station trakk 1 292 W ved GPU-en i testing (StorageReview).

FIG 06 **Ren- og ESD-soner avgjør hvor beregning kan stå** ***En vanlig SMT-hall er en ESD-beskyttet sone, ikke et renrom. GPU- og NPU-servere hører hjemme i serverrommet.***



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

Sone	Typisk miljø	Beregning ved linjen	GPU/NPU-servere
Vanlig SMT/PCBA-hall (forbruker, industri)	ESD-beskyttet område under S20.20 eller IEC 61340-5-1, vanligvis ikke et klassifisert renrom (≈ISO 8-ekvivalent); 40–60% RF	Vifteløs boks i jordet metallkabinett (Jetson T4000/Thor, kinesisk NPU-boks) ved AOI-en eller granskningsplassen; plastskjermer og -kabinetter ut av området eller ≥30 cm fra kort	Serverrom eller klimatisert IT-rack
Medisinsk, kameramoduler, EV-elektronikk (ISO 7–8 renrom)	Renrom pluss ESD-beskyttelse	Vifteløs, IP65/66-forseglet, glatt og lett å tørke av; be om ISO 14644-14-data eller et IPA-sertifikat	Utenfor renrommet, nettverk inn

Sone	Typisk miljø	Beregning ved linjen	GPU/NPU-servere
Bar die, wire bond, optiske stasjoner (lokal ISO 5–6)	Mini-miljø	Bare kameraer og lys inni; beregningsboks utenfor	Serverrom
Fab-renrom (ISO 5 og renere)	Fab-spesifikke ESD- og elektrostatisk-tiltrekningsplaner, utstyr til SEMI E78; cybersikkerhet til SEMI E187/E188	Bare verktøyintegrerte, fab-kvalifiserte kontrollere; ingen ekstra PC-er med vifter	Væskekjølte rack i sub-fab eller fabdatasenteret

Tabellen er min syntese av kildene ovenfor estimate. For et produkt betyr det en vifteløs line-side boks under omtrent 100 W, et metallkabinett bundet til beskyttelsesjord, 12–24 V DC-inngang, vidtemperaturdrift ved reell kabinett-luftstrøm, og en ISO 14644-14 partikkelutslippsrapport klar for kunder på ISO 7 og renere. Ethvert design der «én GPU-server betjener fire linjer» plasserer serveren i serverrommet og etterlater bare en liten boks eller et kamera ved linjen.

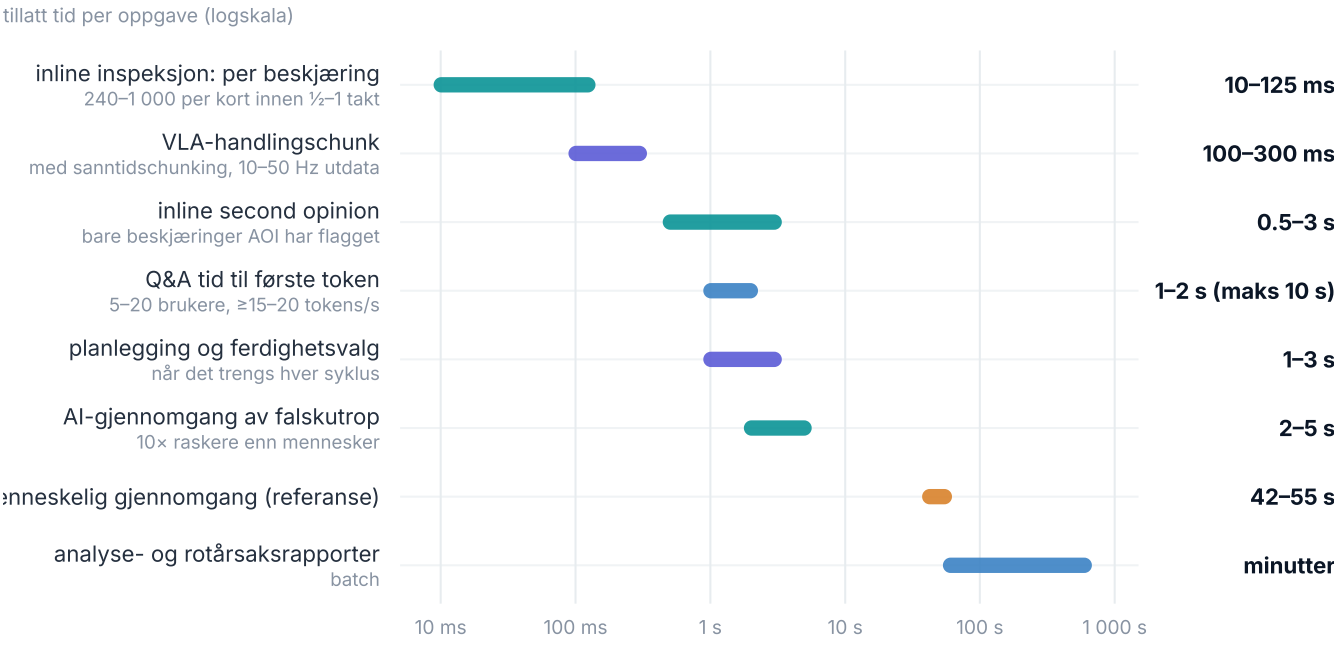
Hastighetsgrenser per bruk: nøyaktighet er vanskeligere enn hastighet

Fabrikker og leverandører offentliggjør nesten aldri AI-latens eller tilgjengelighetsmål. Tersklene nedenfor er utledet fra takt, menneskelig granskningstid, robotkontrollrater og interaksjonsforskning estimate:

Bruk	Hastighetsgrense	Nøyaktighetsgrense	Tilgjengelighet
Inline inspeksjon av hver del	Innen ½–1 takt: 240–1 000 beskjæringer i en 10–30 s takt, 8–100 beskjæringer per sekund	Escapes ikke verre enn dagens «AOI pluss mennesker» (kontrakteksempel <0.5%)	Feil åpent, aldri stopp linjen; mål ≥99.9% (ingen offentlig SLA)
Inline second opinion bare på AOI-flaggede beskjæringer	Flag (vanligvis ≤5–20 per kort) innen ½ takt: 0.5–3 s per beskjæring	Samme; trent på linjens egne bilder	Faller tilbake til menneskelig granskning
AI-granskning av falskutrop	2–5 s per utrop, mer enn ti ganger raskere enn en persons 42–55 s	Escapes ikke verre enn en menneskelig gransker	Degradering legger til folk, stopper aldri linjen
Robotkontroll (VLA)	100–300 ms per handlingschunk med sanntidschunking, 10–50 Hz utdata	≥99% oppgavesuksess for å slå programmert bevegelse	Sikkerhetssystem uavhengig av modellen
Robotplanlegging og ferdighetsvalg	1–3 s når nødvendig hver syklus; ≤1 min hvis bare ved unntak eller omstillinger	Utdata må være verifiserbar kode eller en ferdighets-ID	Faller tilbake til det faste programmet
Q&A (SOP-er, MES-spørringer)	Første token ideelt ≤1–2 s, maks 10 s; ≥15–20 tokens/s per bruker; 5–20 samtidige brukere	Nøyaktige sitater; trofast over lang kontekst	Kontor-IT-klasse (≈99–99.5%)
Analyse- og rotårsaksrapporter	Minutter er greit (batch)	Høyest; trofast over lang kontekst	Kontor-IT-klasse

Physical Intelligence målte 97 ms inferens for en VLA i 3B-klassen (omtrent 108–139 ms med nettverk og forbehandling), fant at sanntidschunking holder gjennomstrømningen uendret opp til 200 ms ekstra forsinkelse, og fullførte fine oppgaver som å tenne en fyrstikk eller koble en nettverkskabel på mer enn 300 ms (Physical Intelligence) vendor research; interaksjonsforskning setter grensen for å holde en tankerekke ved ett sekund og for oppmerksomhet ved ti (NN/g). Nøyaktighet er den vanskelige delen: åpen Qwen3-VL-8B snitter bare 56.8 zero-shot over fem industrielle anomalidatasett (IndusAgent) single source. Ingen maskinvare redder en visjon-språk-modell som ikke er tunet på linjens egne bilder.

FIG 07 **Hastighetsgrenser per bruk** ***Inline-inspeksjon måles i millisekunder, Q&A i sekunder, analyse i minutter. Mennesker bruker 42–55 s per utrop.***



Kilde: Forfatterens syntese (kapittel 1–2).

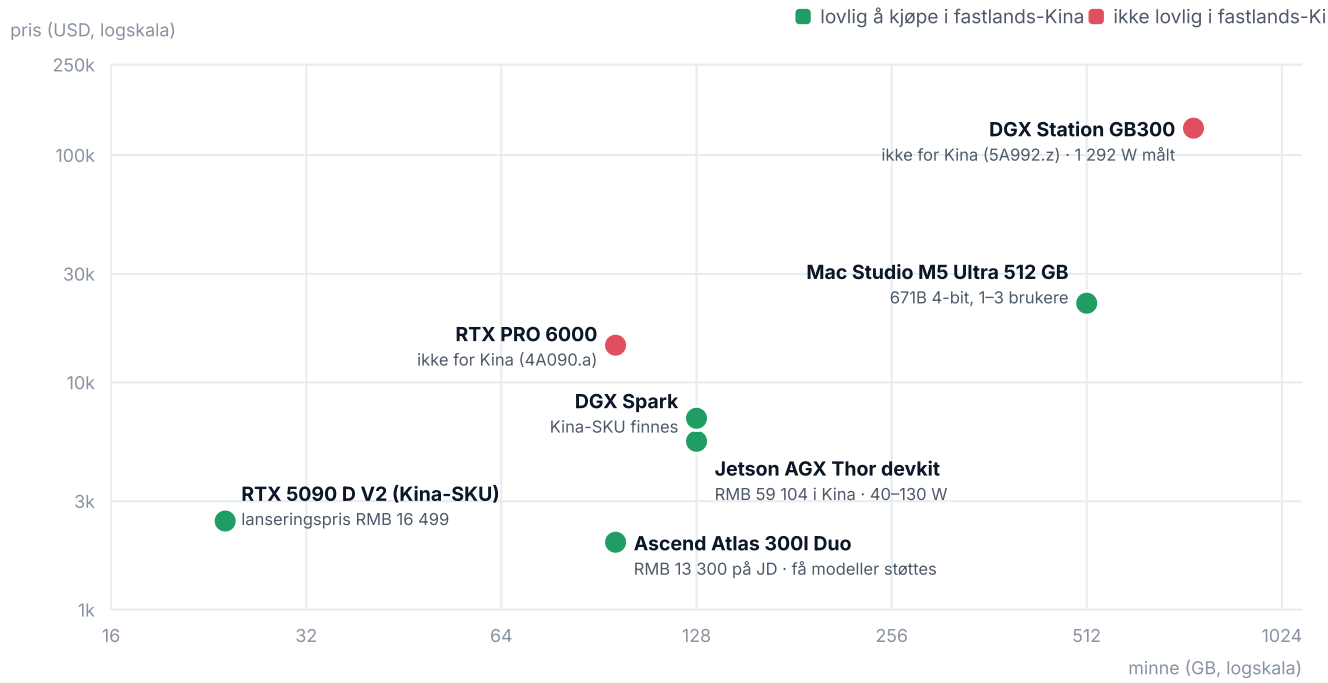
Maskinvare: Thor for stasjonen, en Mac Studio eller innenlandsk enhet for fabrikk-Q&A, og ingen DGX Station på fastlandet

Enhet	Minne og båndbredde	Pris (november 2026)	Hva den er god for	Fastlands-Kina
Jetson AGX Thor devkit / T5000 / T4000	128 GB LPDDR5X, 273 GB/s; 40–130 W	$3 499 ved lansering → $5 499 (juli 2026); T5000 $4 999, T4000 $2 999; RS China devkit RMB 59 104 inkl. avgift	2–8B synsmodeller som sjekker beskjæringer på nytt; sanntids VLA-er i 3B-klassen	● Ja (5A992.c, ingen lisens for sivile fabrikker)
DGX Spark	128 GB, 273 GB/s	$3 999 → $4 699 → $6 950; samme-chip ASUS GX10 Kina-modell RMB 32 999 på JD (okt 2025)	Flerbruker 8–35B MoE	● Ja (5A992.c, Kina-modell, autoriserte kanaler)
RTX PRO 6000	96 GB GDDR7	≈$8 565 → $13 250 (juni 2026) → ≈$16 005 gjennomsnittlig eBay-oppføring i november	Flerbruker 27–35B FP8; 120B-klasse MoE ved 4-bit	● Nei (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 estimert i Kina	DeepSeek 671B ved 4-bit for 1–3 brukere	● 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 gatepris i november (≈RMB 830–940k før toll); $95–117k i august	Høy samtidighet på 120–235B MoE som passer i HBM; bare noen få brukere på 433–743B-modeller	● Nei (5A992.z, TPP 58 484)
Huawei Atlas 300I Duo	96 GB LPDDR4X, 408 GB/s	RMB 13 300 fra en JD-forhandler	Lavvolums batchgranskning; få støttede modeller	● Ja (innenlandsk)
Huawei Atlas 800I A2 (8 × Ascend 910B 64 GB)	8 × 64 GB HBM	Ikke innhentet; full DeepSeek trenger to ved W8A8, fire ved BF16	Fabrikkomfattende flerbruker Q&A og analyse	● Ja (innenlandsk)

Enhet	Minne og båndbredde	Pris (november 2026)	Hva den er god for	Fastlands-Kina
RTX 5090 D V2 (Kina-modell)	24 GB	RMB 16 499 ved lansering (aug 2025), opp siden med minnepriser	8B synsmodeller, små VLA-er	● Ja

Kilder: (NVIDIA China Thor, NVIDIA Thor launch, CNX on price rises, NVIDIA export classification data, RS China, Implicator, IThome on ASUS GX10, Tom's Hardware RTX PRO 6000, GPU Poet, Apple specifications, TechRadar, IThome on M5 Ultra China pricing, Wccftech, StorageReview, CyberPowerPC, monitors.com, Tom's Hardware DGX Station, vLLM-Ascend support matrix, China Telecom Cloud, Huawei Atlas 800I A2, TweakTown). Fastlandspriser for 512 GB Mac Studio er ekstrapolert fra US-konfigurasjonspåslag og har lav konfidens estimate .

FIG 08 **On-prem beregning: minne, pris og hva som lovlig kan kjøpes i fastlands-Kina*Eksportklassifiseringer avgjør shortlisten før benchmarks.***



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

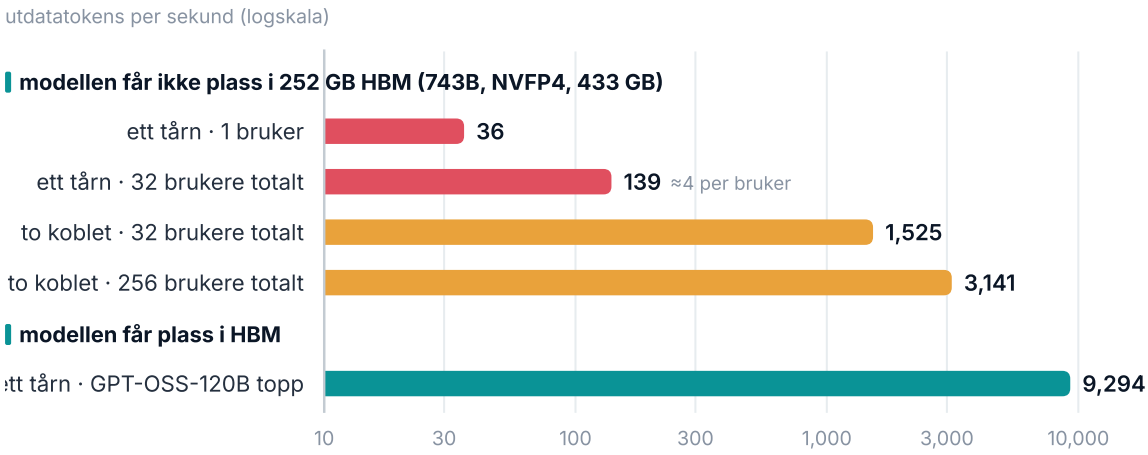
Jetson Thor får stasjonsnivå-utrulling til å virke. NVIDIAs TensorRT Edge-LLM-tester viser Qwen3-VL-8B (NVFP4) som pre-fyller 2 048 tokens på 104.5 ms og dekoder 44.4 tokens/s på Thor (TensorRT Edge-LLM) vendor claim ; for en AOI-beskjæring og et 20-tokens svar er det omtrent 0.5 s med 8B-modellen og 0.2 s med 2B, eller 8.5–21 beskjæringer per sekund batchet estimate . Utlengde, ikke parameterantall, setter latensen, så begrenset dekoding til et klassetoken eller kort JSON er en del av designet. GR00T N1.7 (3B) kjører på 25.1 Hz på Thor (Jetson AI Lab) vendor claim . En Orin NX tar omtrent 1.8 s per beskjæring og en Atlas 300I Duo har en p95 over 23 s, bare egnet for batcharbeid (CSDN) single source . På fastlandet bør NVIDIA-produkter budsjetteres til 40–60% over US-listepris estimate .

En 512 GB Mac Studio gjør en fullverdig modell på én maskin mulig, for noen få brukere. M5 Ultra (august 2026) tilbyr opptil 512 GB ved 1.2 TB/s fra omtrent $22 299 (Apple, TechRadar); DeepSeek 671B ved 4-bit (≈380–404 GB) dekoder 17–20 tokens/s på en M3 Ultra (Pinggy) verified og anslått 25–29 på M5 Ultra. Lange prompt er svakheten: en 10 000-tokens retrieval-prompt tar omtrent 17–70 s estimate , utenfor 10-sekunders interaksjonsgrensen.

DGX Station er den sterkeste enkeltmaskinen, men ikke tilgjengelig på fastlandet og fem til seks ganger prisen på en Mac. Den sendte GB300 har **748 GB**, hvorav bare 252 GB er raskt HBM (Wccftech, Tom's Hardware, Aug 2026) verified . Uavhengige tester viser omtrent 9 300 utdatatokens/s på GPT-OSS-120B, som passer i HBM, men en 743B-modell som renner over i LPDDR5X ga 32 brukere bare 139 tokens/s totalt, omtrent 4 hver; to koblede enheter nådde 1 525 (StorageReview, Aug 2026, StorageReview, Oct 2026) single source . Gateprisene steg til $123–

140k innen november (CyberPowerPC, monitors.com). På fastlandet er de realistiske alternativene for samme jobb Mac Studio, KTransformers-stil CPU+GPU-servere (KTransformers) vendor test eller innenlandske Ascend-servere.

FIG 09 **Én stor arbeidsstasjon gir ingen delt assistent** ***Når modellen ikke får plass i høykapasitetsminne, får 32 samtidige brukere omtrent 4 tokens per sekund hver.***



Kilde: StorageReview (aug og okt 2026); ServeTheHome. Uavhengige tester fra én kilde.

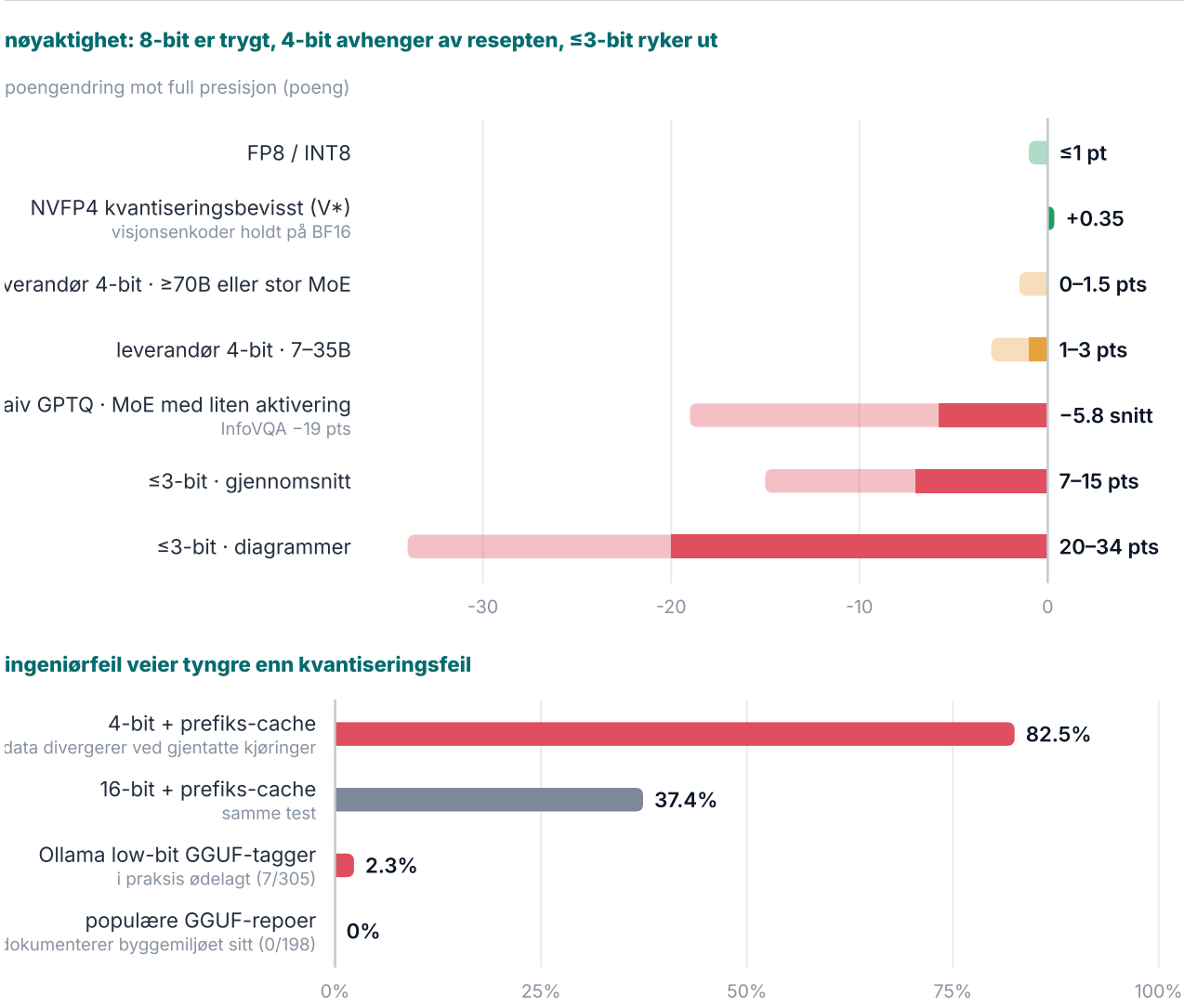
Minnemangel flyttet de samme delene 20–100% på ett år (DGX Spark gikk fra $3 999 til $6 950 (TechRadar)), så utstyrspriser bør indekseres snarere enn fastsettes.

Innenlandske akseleratorer kjører fullverdige modeller; programvarevedlikehold er flaskehalsen. Ascend 910B har 64 GB HBM, men vLLM-Ascend merker DeepSeek V3/R1 som «ikke lenger vedlikeholdt» og støtter bare noen få modeller på mindre kort (vLLM-Ascend) single source, primary, og hver ny modell tar én til to måneder å tilpasse (Caijing/Anue); send en frosset, validert modellversjon. Å betjene full DeepSeek 671B til 20–100 brukere trenger minst to Atlas 800I A2-servere med åtte kort (China Telecom Cloud) verified, likevel utgjorde offentlige anbud for LLM-enheter i Kina bare RMB 27.5 millioner i januar–september 2025 (C114) single source: fabrikker betaler for integrasjon og tjeneste, ikke for en boks med beregning.

Kvantisering: 8-bit er trygt, 4-bit avhenger av resepten, 3-bit og lavere er uaktuelt

8-bit (FP8 eller INT8) er i hovedsak tapsfritt for visjon-språk-modeller (Red Hat, Hugging Face) verified. Ved 4-bit avhenger tapet av størrelse og resept: Qwens 72B AWQ-bygg taper under ett poeng på MMMU mens 7B taper 2.8 (Qwen AWQ) verified, og detaljtunge oppgaver faller mer (naiv 4-bit GPTQ koster Qwen3-VL-30B-A3B 19 poeng på InfoVQA (VEQ)). Ved 3-bit og lavere er gjennomsnittstapet 7–15 poeng og diagrammer taper 20–34 (VLMQ) verified; utover 64k tokens taper 4-bit opptil 59%, verre på ikke-engelske inndata (arXiv 2505.20276). Hold synsenkoderen på FP16/BF16.

FIG 10 **Kvantisering: 8-bit er trygt, 4-bit avhenger av resepten, 3-bit og lavere ryker ut** ***Ingeniørfeil i serving-stacken koster mer nøyaktighet enn selve kvantiseringen.***



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

To advarsler betyr mer. For det første, **per 11. november 2026 har ingen studie eller modellkort målt kvantisering på PCB/SMT-defektbilder**, så revalider på det gylne settet. For det andre er serving-stakk-feil ofte større enn kvantiseringsfeil: samme modell skåret omtrent 30 poeng lavere under Ollama enn under referanseimplementasjonen (SiliconBench); 7 av Ollamas 305 offisielle kodemodell-tagger var i hovedsak ødelagt selv om filene var intakte (arXiv 2609.05881); ingen av 198 populære GGUF-repositorier registrerte bygget sitt (arXiv 2610.19498); og med prefiks-cache-gjenbruk divergerte 82.5% av episodene ved 4-bit mot 37.4% ved 16-bit (arXiv 2609.04748) single sources. Leverandørens kvantiseringsbevisste 4-bit er nær tapsfritt på detaljoppgaver (NVIDIA Nemotron 3.5 vendor claim, men 4-bit VLA-er taper 1.5–4.3 poeng under perturbasjon og kjører tregere på Thor (LoopQ) single source. Håndter en kvantisert modell som et nytt måleinstrument: pin vekthasher, registrer bygget, tilbakestill prefiks-caching og revalider etter ethvert bytte av presisjon, motor eller mal.

Presisjon	≥70B tett eller ≥100B MoE	7–35B	≤4B eller detaljtunge oppgaver	Konklusjon
FP8 / INT8	Trygt (≤1 poeng)	Trygt	Trygt; standard for inspeksjon	Bruk

Presisjon	≥70B tett eller ≥100B MoE	7–35B	≤4B eller detaljtunge oppgaver	Konklusjon
4-bit AWQ/GPTQ (leverandørresept, synsenkoder FP16)	0–1.5 poeng	1–3 poeng; naiv GPTQ på liten-aktivering MoE −5.8 i snitt, −19 på InfoVQA	2–3 poeng; valider detaljoppgaver først	Betinget
NVFP4 (leverandørresept)	≈0.3–1 poeng under FP8 (leverandørdata)	Kvantiseringsbevisste bygg nær tapsfritt på V*, OCRBench	Ingen offentlige data	Betinget (foretrekk kvantiseringsbevisste bygg)
GGUF Q4_K_M	Utestet	Høy varians på domeneoppgaver	Ikke for liten tekst eller defekter	Ikke for detaljvurdering
GGUF Q5_K_M og høyere	Antas trygt	Nær tapsfritt i to tester	Gulvet for detalj på llama.cpp	Bruk
≤3-bit (inkludert 1T-modeller presset til 1.8–2 bits på én maskin)	Ikke for beslutninger	Nei	Nei	Nei

Hvilken konfigurasjon for hvilken jobb

Konfigurasjon	Inline inspeksjon	AI-granskning av falskutrop	Robotplanlegging og -kontroll	Q&A og analyse
Dedikerte detektorer eller anomalimodeller (YOLO, RF-DETR, PatchCore-klassen), INT8/FP16, enhver NPU- eller GPU-edgeboks	● Ja: hoveddommeren, 2–12 ms per beskjæring	● Ja (klassifikator tunet på linjen)	● Bare persepsjon	—
2–4B synsmodeller, NVFP4/AWQ/FP8 med FP16 synsenkoder, på Thor/T4000 eller DGX Spark	● Betinget: sjekk AOI-flaggede beskjæringer på nytt ved 0.2–0.3 s hver	● Betinget, under 1 s etter tuning	● VLA-ryggrad (2–3B VLA ved ≈40–130 ms per chunk)	● Svak
7–9B synsmodeller (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 per flagget beskjæring	● Ja: 0.5–1 s per utrop, best verdi	● Planlegging og ferdighetsvalg (1–3 s), ikke i kontrolløkken	● Grunnleggende SOP-Q&A for én eller noen få brukere
26–35B MoE (Gemma-4-26B-A4B, Qwen3.6-35B-A3B), NVFP4/FP8	● Nei (for tung)	● Ja, som forklarer eller for vanskelige tilfeller	● Planlegging	● Ja (8 brukere på en Spark, ≈0.1 s første token)
106–241B MoE ved 4-bit på en Mac Studio, RTX PRO 6000 eller Spark	● Nei	● Nei	● Offline planlegging	● Ja, for én eller noen få brukere
DeepSeek 671B eller Kimi K2.x 1T ved 4-bit på en 512 GB Mac, KTransformers-server eller DGX Station	● Nei	● Nei	● Bare offline programgenerering	● Betinget: 1–3 brukere og korte prompt; lange prompt tar 16–170 s

Konfigurasjon	Inline inspeksjon	AI-granskning av falskutrop	Robotplanlegging og -kontroll	Q&A og analyse
Alt ved ≤3-bit; GGUF Q4_K_M for småmodell-detaljarbeid; Ollama for mange brukere	● Nei	● Nei	● Nei	● Nei

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

Altså, til de to vanlige bekymringene: **det behøver ikke være langsomt** hvis modellen passer jobben (små og mellomstore synsmodeller på Thor eller Spark møter hvert inline- og granskningsbudsjett; 671B–1T-modeller holdes av linjen), og **kvantisering behøver ikke gjøre det feil** med FP8/INT8 som standard, eller leverandørresept 4-bit over 7B validert på det gylne settet. De større risikoene er de omtrent 57-poeng zero-shot industrielle nøyaktigheten til generelle modeller og serving-stakk-feil. Gi inspeksjonsforespørsler faste batchstørrelser og sin egen kø (temperatur 0 er ikke bit-reproduserbar under batching (vLLM documentation)), hold en N+1 varm standby, og få hvert trinn på linjen til å feile åpent.

Lærer og elev

03 Lærer og elev: frontmodeller kan forhåndsmerke inspeksjonsdata; for robotarmer er de virkelige lærerne fortsatt mennesker

November 2026 lærerkandidater: generisk syn doblet seg, industrielle defektpoeng mangler

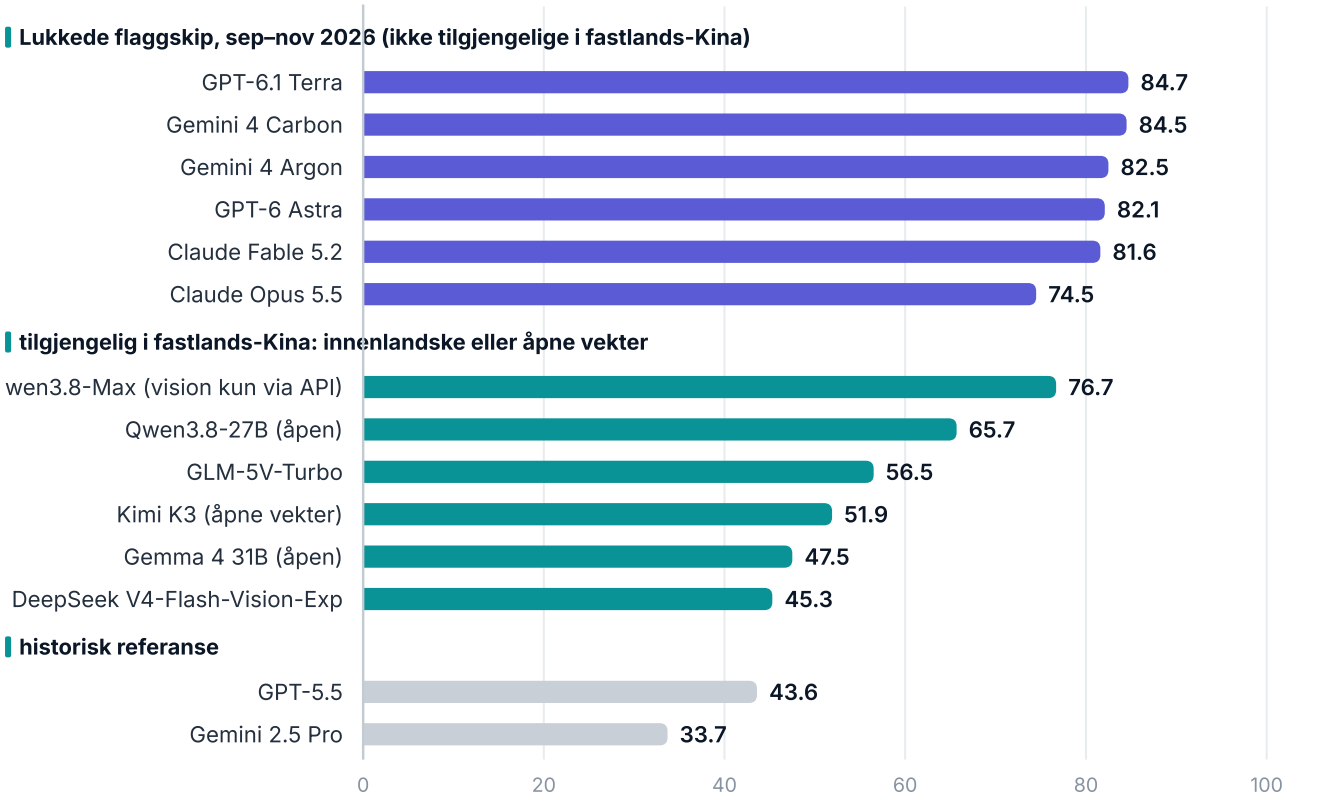
Hvert lukket flaggskip som kunne fungere som lærer i november 2026 aksepterer bilder: Claude Fable 5.2 (generelt tilgjengelig 9. november, $8/$40 per million tokens) og Opus 5.5 ($4/$20) (Anthropic, Claude models overview) verified; GPT-6 Astra ($10/$50) og GPT-6.1 Terra ($1/$5) (OpenAI pricing); og Gemini 4 Argon og Carbon, åpnet i etapper (Google Carbon). Brukbare i Fastlands-Kina er Qwen3.8-Max (syn via API) og åpne modeller som Qwen3.8-27B, Kimi K3, GLM-5V-Turbo, DeepSeek V4-Flash-Vision-Exp og Gemma 4 (Roboflow blog, Hugging Face).

Det sentrale gapet må sies tydelig: ingen av disse modellene har en publisert poeng på MMAD, MVTec AD 2, VisA, Real-IAD, UniPCB eller noe PCB/SMT-datasett. MMAD har fortsatt ingen leaderboard (MMAD GitHub), og de to nyeste industrielle anomalipapirene (august–september 2026) bruker fortsatt Gemini 2.5 Pro og Claude Sonnet 4 som sine nyeste generelle grunnlinjer (InspectorGPT, Anomaly-LR). Et arXiv-fulltekstsøk 11. november fant null treff for «Fable 5.2» eller «Gemini 4 Carbon» i denne konteksten (arXiv API) single source, absence of evidence. De nærmeste erstatningene:

- Den nyeste frontmodellen med en MMAD-poeng er GPT-5.5: 69.95% på «defekt eller ikke» og 76.20% snitt over syv oppgaver på et 500-bilders delsett, mot 95.24% og 86.65% for menneskelige eksperter (AnomalyClaw) single source; GPT-4o skåret 68.63% og 74.92% i 2024 (MMAD, ICLR 2025), omtrent 1.3 poeng fremgang på to år.
- På LDU-Bench, ekte SEM-granskningsbilder fra IC-litografi og den eneste elektronikkbenchmarken med 2026-modeller, er god-eller-dårlig-triage grei (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 klassifisering av 11 defektmorfologier når bare 0.20–0.41 (LDU-Bench) single source.
- På Roboflows uavhengige synsevalueringer (oppdatert 10. november 2026), er objektdeteksjon mAP@50 84.5 for Gemini 4 Carbon, 84.7 for GPT-6.1 Terra og 81.6 for Fable 5.2, mot 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, independent evaluator, private datasets.

FIG 11 **Lærerkandidater: generiske poeng doblet seg på en generasjon*Men ingen av 2026-flaggskipene har en publisert poeng på en industriell defektbenchmark.***

Roboflow generisk deteksjon mAP@50



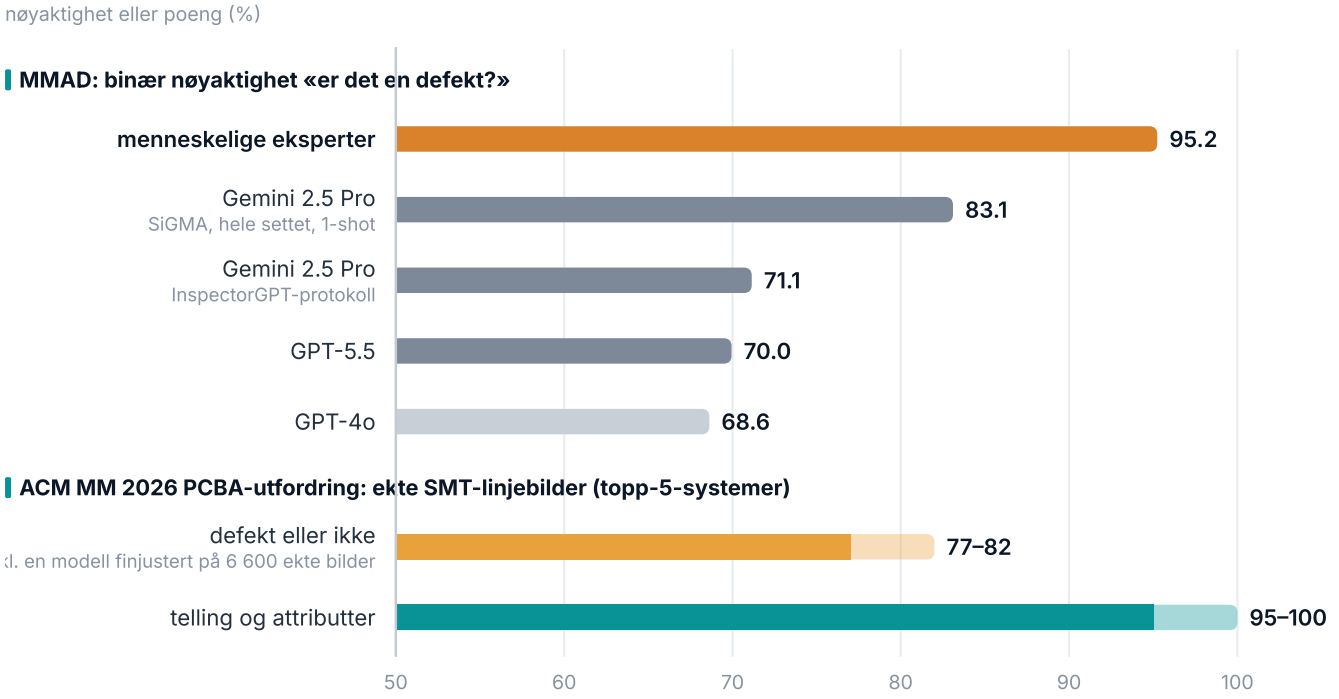
industrielle defektbenchmarks (MMAD, MVTec AD 2, VisA, Real-IAD, UniPCB): null publiserte poeng for disse 2026-flaggskipe
generiske deteksjonspoeng doblet seg på en generasjon, men det er ingen defektvurdering; test på ditt eget gylne sett

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

Lærerkandidat (nov 2026)	Pris per M tokens (inn/ut)	Roboflow deteksjon mAP@50 / telling	Industrielle defektbenchmarks	Brukbar i Fastlands-Kina
Claude Fable 5.2	$8 / $40	81.6 / 86.5	Ingen	Nei
Claude Opus 5.5	$4 / $20	74.5 / 80.6	Ingen	Nei
GPT-6 Astra	$10 / $50	82.1 / 80.2	Ingen	Nei
GPT-6.1 Terra	$1 / $5	84.7 / 80.2	Ingen	Nei
Gemini 4 Carbon	$2 / $10 (kampanje)	84.5 / 90.1	Ingen	Nei
Qwen3.8-Max (syn kun via API)	$2 / $6	76.7 / 81.1	Ingen	Ja
Qwen3.8-27B (åpen)	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 (åpen)	Lokal	47.5 / 51.4	LDU-Bench triage 0.504	Ja (lokale vekter)
Referanse: GPT-5.5 / GPT-4o	—	43.6 / 68.0 (GPT-5.5)	MMAD binær 69.95% / 68.63%	Nei
Menneskelige eksperter	—	—	MMAD binær 95.24%	—

Nyere papirer tester fortsatt den forrige generasjonen, men peker samme vei. På hele MMAD-settet skårer menneskelige eksperter 95.2% på defekt-eller-ikke og Gemini 2.5 Pro 83.1% (SiGMA), likevel skåret samme modell bare 71.14% i et annet papir (InspectorGPT): testprotokollen flytter poengsummen like mye som en modellgenerasjon gjør. I CVPR 2026 VAND 4.0-utfordringen toppet hyllevare visjon-språk-modeller på 61.88 AP mot 90.84 for veiledede metoder, og arrangørene konkluderte at de «for øyeblikket ikke kan erstatte finjusterte klassifikatorer» (VAND 4.0) single source . I ACM MM 2026 PCBA-utfordringen skåret de fem beste 76.92–82.47 på defekt-eller-ikke mens komponenttype og pin-telling nådde 95–99.5 (arXiv 2609.21276) single source : å gjenkjenne er mye lettere enn å vurdere.

FIG 12 **«Er det en defekt?» er fortsatt vanskelig for generelle modeller** ***Den beste generelle modellen ligger mer enn 12 poeng bak menneskelige eksperter på MMAD; tunede systemer gjør det bedre på ekte linjebilder.***



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

Tre konklusjoner følger. Å se og telle doblet seg på en generasjon, men å vurdere avvik beveger seg langsomt, avhenger sterkt av protokollen og kommer med systematiske skjevheter; hvor mye de nye modellene vinner kan bare måles på en fabrikks egne bilder. Innenfor treningsdistribusjonen sin slår finjusterte 7B-spesialister fortsatt generelle modeller (Anomaly-LR, UniPCB). Og ekte linjedata kollapser poengsummer: på monterte elektroniske paneler falt bilde-nivå AUROC fra 97–98 på MVTec AD til 31.5–78.6 (arXiv 2606.01992) single source . For fastlandsselskaper ligger den sterkeste tilgjengelige læreren, Qwen3.8-Max, omtrent 8 poeng under den globale fronten.

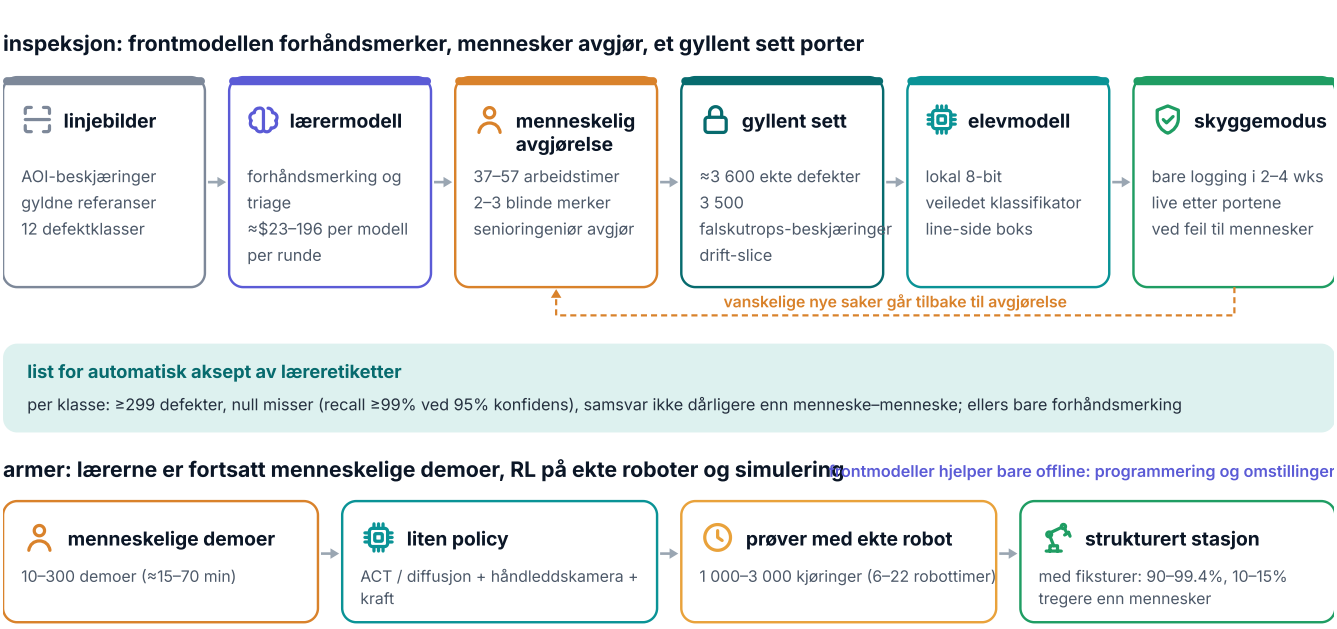
Å kjøre dusinvis eller hundrevis av runder: penger er ikke problemet, etikettkvalitet er

Penger er ikke problemet. Ved november 2026 listepriser koster 1 000 én-megapiksel-bilder omtrent $3 (GPT-6.1 Terra) til $32 (GPT-6 Astra) per runde, og hundre runder over 100 000 bilder omtrent $32 000–135 000 før batchrabatter estimate (Claude vision docs, OpenAI vision guide, Roboflow Vision Evals). For enkle etiketter koster en person omtrent det samme som Terra; lærerens fordel er hastighet og skala, ikke pris. Kundenes NDA-er holder vanligvis produktbilder inne i fabrikken, så det gylne settet merkes på stedet.

Den virkelige hindringen er kvalitet, og flere runder fikser ikke systematiske feil. Roboflow kjører hver oppgave tre ganger og poeng beveger seg bare ±0.1–1.4 poeng (Roboflow), så en modell gjentar stort sett feilene sine og flertallsavstemning kan ikke redde den. Runder hjelper bare når hver legger til informasjon: ett eller to gode referansebil-

der legger til omtrent 3–4 poeng, mer legger ikke til noe (AD-Copilot); å la en modell argumentere mot en god referanse over flere runder legger til 6.23 poeng kryssdomene AUROC for GPT-5.5 og 2.95 på MMAD, med topp etter omtrent tre kall (AnomalyClaw) single sources. «En elev trent på læreretiketter nærmer seg menneskelige etiketter» er bare vist på naturlige bilder: en ResNet-101 trent på Qwen3-VL-etiketter av COCO og Objects365 nådde 95.7–99.4% av menneskelig-etikett-resultatet, og forfatterne sier domenedata er utestet (arXiv 2602.20972) single source.

FIG 13 **To lærer–elev-pipelines** ***For inspeksjon forhåndsmerker en frontmodell og mennesker avgjør; for robotarmer er de virkelige lærerne fortsatt menneskelige demonstrasjoner.***



Kilde: Forfatterens syntese (kapittel 3); kostnader og utvalgsstørrelser er estimater.

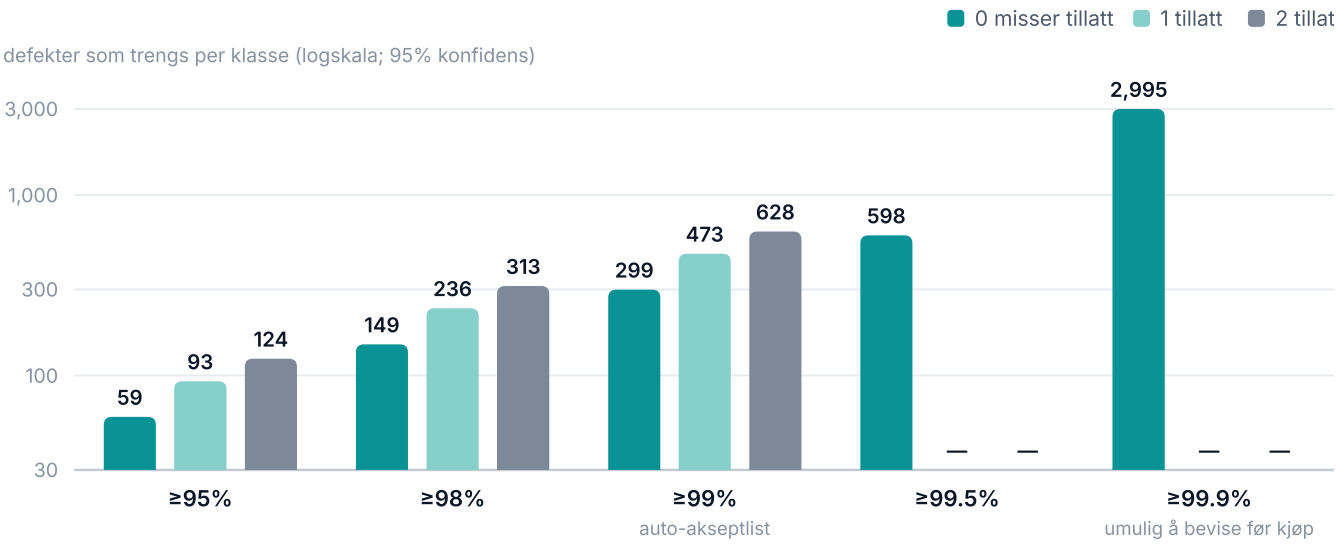
En brukbar protokoll estimate:

1. **Samle det gylne settet fra fabrikkens eget AOI-arkiv og beholdte NG-kort**, stratifisert i 12 sanne defektklasser (manglende, forskyvning, tombstone, polaritet, feil del, bro/kortslutning, utilstrekkelig lodd, overskytende lodd, loddkule, løftet ledning/åpen, skadet del, fremmedmateriale) og fire falskutrops-årsaker (skygge, fluss, silketrykk, toleranse), pluss en «drift-slice» på minst 300 defekter fra en annen linje eller annen belysning, kortfarge eller delenumre.
2. **Blindmerke med to eller tre IPC-A-610-trente inspektører**, med en senioringeniør som avgjør og røntgen, elektrisk test eller tverrsnitt som avgjør tvetydige prøver. Aksepter en klassedefinisjon bare hvis inspektørenighet når κ ≥0.75 (≥0.90 for kritiske klasser); menneskelig enighet om 11 elektronikkdefektmorfologier i LDU-Bench var bare 0.727, så noen klasser trenger sammenslåing eller omdefinering. AIAG-stil attributtenighet (effektivitet ≥90%, miss ≤2%, falsk alarm ≤5%, κ >0.75; >0.90 for release) er et fornuftig minimum (Fabrico, iFactory).
3. **Dimensjoner det med eksakte binomiale (Clopper–Pearson) ensidige 95%-grenser** (binomial confidence intervals, rule of three):

Recall å bevise	0 misser tillatt	1 tillatt	2 tillatt
≥95%	59 defekter	93	124
≥98%	149	236	313
≥99%	299	473	628
≥99.5%	598	—	—

Recall å bevise	0 misser tillatt	1 tillatt	2 tillatt
≥99.9%	2 995	—	—

FIG 14 **Hvor stort det gylne settet må være*Å bevise 99% recall med null misser krever 299 defekter per klasse; 99.9% kan ikke bevises før kjøp.***



Kilde: Eksakte binomiale konfidensintervaller (95%).

Seksti defekter per klasse er nok for en bestått/ikke-bestått-screening; 300 trengs for å avgjøre automatisk aksept. Hele settet, omtrent 3 600 defekter pluss 3 500 falskutrops-beskjæringer, tar omtrent 37–57 arbeidstimer å merke og avgjøre, og å kjøre en lærerkandidat over det koster $23–196; den virkelige kostnaden er to til seks ukers innsamling, satt av den sjeldneste klassen estimate. Send enkeltkomponent AOI-beskjæringer, ikke hele kort, med målet minst 42–48 piksler på tvers (arXiv 2610.09313).

1. **Aksepter læreretiketter automatisk bare for en klasse med null misser på ≥299 gylne defekter**, modell–sannhet-enighet ikke dårligere enn menneske–menneske-enighet, og minst 50% av falskutropene ryddet. Kritiske klasser som kortslutninger, manglende deler, polaritet og feil deler trenger 99.5–99.9% (598–2 995 defekter); ellers går alt AI-en ikke kan rydde med høy konfidens til en person. En escape-rate under 0.1% av defektene kan ikke bevises før kjøp (2 995 defekter med null misser; 0.01% ville ta 29 956), så linjen beholder menneskelig stikkprøve av AI-ryddede utrop, nedstrøms ICT/FCT som sikkerhetsnett og SPC på escapes. Verifiserte etiketter destilleres deretter til en konvensjonell detektor eller klassifikator og revalideres på hvert nytt produkt.

Vilkår og dataregler avgjør også læreren. Anthropic forbyr å bruke tjenestene sine til å trene konkurrerende modeller (Anthropic commercial terms); OpenAI tillater at utdata trener interne klassifikatorer som ikke selges til tredjeparter (OpenAI business terms); DeepSeek tillater eksplisitt destillasjon (DeepSeek terms) primary terms. Ingen definerer en «konkurrerende modell», så kommersiell bruk krever juridisk gjennomgang, og kundenes dataregler holder vanligvis bilder på stedet. Det tryggeste standardvalget er en lokalt kjørt lærer med åpne vekter, med skymodeller i fronten bare når kunden tillater det skriftlig.

Inspeksjonselever: små og raske; flaskehalsene er optikk, data og validering

Eleven er ikke flaskehalsen: YOLO26n tar 3.80 ms per bilde i INT8 på en $249–399 Orin Nano Super (Ultralytics) vendor test, NVIDIA rapporterer 98.51% på tre PCB-defektklasser etter å ha tilpasset en grunnmodell på omtrent 600 bilder (NVIDIA) vendor claim, og syntetiske defekter legger bare til 0.9–2.5 mAP-poeng (arXiv 2605.04635). Be-

grensningene er optikk, linjedata, validering og lisenser: Ultralytics YOLO er AGPL-3.0 eller bedriftslisensiert (Ultralytics), så lukkede produkter trenger en lisens eller en Apache-2.0-detektor som RF-DETR.

Robotelever: strukturerte stasjoner er nær produksjonsgrad; fleksibelt og dynamisk arbeid er ikke

I november 2026 er det tre slags robotlærere, og bare én kan bli en lokal elev: åpne, finjusterbare VLA-er (π0/π0.5, GR00T N1.7, X-VLA, RDT2, SmolVLA). Lukkede robotmodeller (π0.6/π0.7 gjennom en beta-API, Gemini Robotics 2 for betrodde testere) kan ikke rulles ut på stedet (The AI Insider, DeepMind). Frontens generelle modeller er langsomme og dårlige på fin manipulasjon: i Artificial Analysis' test skåret de 46–69 med 28–45% grepssuksess og omtrent 30 minutter per oppgave, «altfor langsomt og dyrt for enhver produksjonsbruk» (AA-Robotics); GPT-6 Astra klarte et presisjonsspor bare 2 ganger av 20 (RoboCurve); og å drive ledd direkte gir 0–5.5% på LIBERO ved 0.2–0.4 Hz mot omtrent 83 Hz nødvendig (Anthropic) verified across three sources. Fable 5.2s 64.9% på den simulerte Robo-Dojo-benchmarken er en intern kjøring (Anthropic) vendor claim.

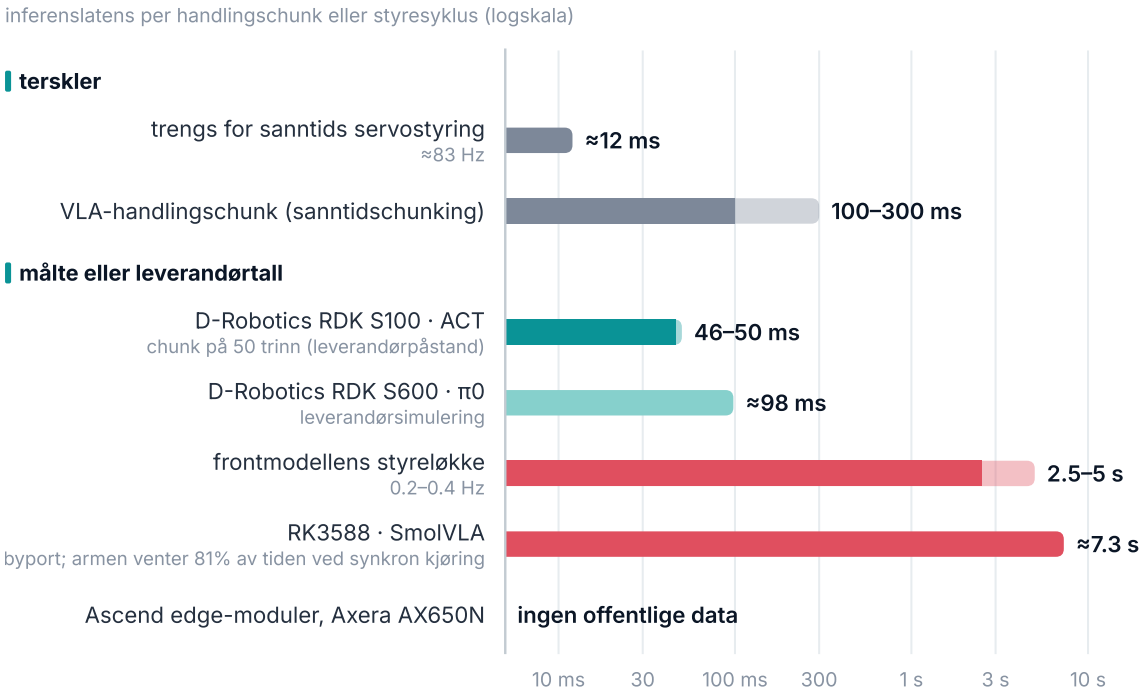
Tredjepartsstudier siden oktober er enige: omtrent 33–44% på LIBERO-10 for de tre ledende agentene (arXiv 2610.30051); agenter som feilaktig hevder suksess på 24–39% av deloppgavene (arXiv 2611.08209); og 104 av 108 kvasistatiske ekte-robot-oppgaver for Opus 5.5, men ved 280 sekunder hver og uten innsetting eller montasje (AMP1) single sources. Frontmodeller hører hjemme offline: å skrive programmer, orkestrere oppgaver og utarbeide suksesskriterier.

Elevkvalitet har en klar kant: god innenfor treningsdistribusjonen, kollapser utenfor den. π0.7 «overstiger ofte 90%» på sette oppgaver og når 60–80% på usette (π0.7 paper) vendor research; ContactDP når 97.8% på trente kontakter men 28.6–66.7% på usette (ContactDP); små kamera- eller posisjonsendringer faller 95% under 30% i LIBERO-Plus (LIBERO-Plus, CVPR 2026); innlært grep fra bevegelige samlebånd topper på 58.89% (ForeTime-VLA) single sources, en jobb klassisk synssporing har gjort i produksjon i årevis.

Bevisene for nær-produksjonsgrad kommer alle fra strukturerte stasjoner, menneskelige demonstrasjoner og ekte-robot-korreksjon. Neuromeka beholdt en tilstandsmaskin og forhåndslært bevegelse og lærte bare grep, kabelinnsetting og lodding, hver fra under 20 minutters data: 322 av 324 i produksjon, 12.8% tregere enn en person (Neuromeka) single source, vendor. Fraunhofer nådde 92.8% over fire kontakter fra 68 minutters demonstrasjoner (Fraunhofer); Berkeleys HIL-SERL lyktes 100 ganger av 100 på RAM-, SSD- og USB-innsetting etter 1–2.5 timers RL på ekte robot (HIL-SERL); AgiBot kjørte 18–20 sekunder per trinn innenfor en 20-sekunders takt på en nettbrett-testlinje (IT-home) vendor claim.

Destillasjon kjøper hastighet, ikke generalisering: π0.5 kjører på 0.70 Hz på Jetson Orin rå og 10.30 Hz etter asynkron planlegging (EagleVLA); lagdestillasjon tar den fra omtrent 3 Hz til omtrent 10 Hz (Shallow-π) verified. En elev arver bare det læreren dekket; hvert nytt produkt trenger fortsatt oppgavedata.

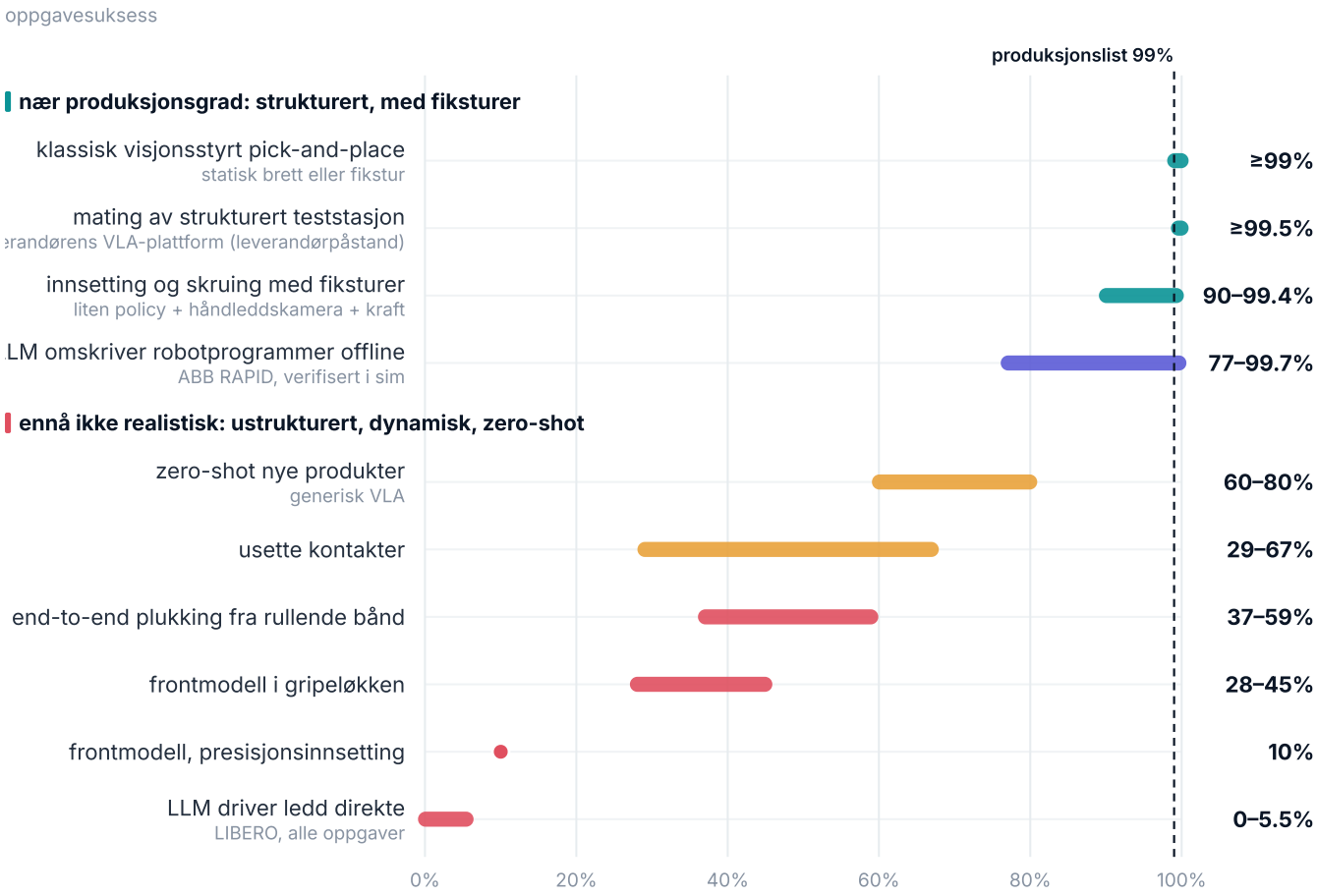
FIG 15 **Edge-policylatens***En liten ACT-policy kjører på omtrent 50 ms på et edge-kort; en VLA portert til en hobbychip tar 7.3 s.*



Kilde: D-Robotics dokumentasjon og forum (leverandørtall); community-porter på GitHub; Anthropic-forskning.

På kinesiske edge-chips er den eneste målte innlærte policyen en ACT-policy på D-Robotics' RDK S100 ved 46–50 ms per 50-trinns chunk, uten publisert suksessrate (D-Robotics RDK S docs) vendor claim ; en hobbyport av SmolVLA på en RK3588 tar omtrent 7.3 s per chunk (GitHub). Kvantisering kjøper ikke hastighet gratis: ingen 4-bit π0.5-bygg var raskere enn 16-bit ved batchstørrelse 1 (LoopQ). Demonstrasjoner er dyre (Reuters rapporterte omtrent 300 forsøk per brukbar bevegelse for en nybegynnerteleoperatør (Notebookcheck)), og validering koster mer: å bevise en feilrate ≤0.1% ved 95% konfidens tar omtrent 3 000 påfølgende suksesser, 17–22 robottimer estimate . Ingen innlært manipulasjonspolicy har sikkerhetssertifisering.

FIG 16 **Robotelever: nær produksjonsgrad bare når stasjonen er strukturert** ***Stasjoner med fiksturer når 90–99%+; ustrukturerte, dynamiske og zero-shot-oppgaver er ennå ikke realistiske.***



Kilde: Artikler og leverandørmateriale sitert i kapittel 3; noen tall er leverandørpåstander.

Elektronikkoppgave	Brukbar elev	Forventet suksess	Hastighet vs en person	Kostnad per omstilling	Konklusjon 2026–2028
High-mix pick-and-place, kitting, maskinmating (statisk brett eller fikstur)	Innlært persepsjon pluss programmert eller planlagt bevegelse	≥99% med klassisk synsstyring	Samme som konvensjonelle stasjoner	Timer	Gjennomførbart nå
Fiksert kontakt-, FFC- og kabelinnsetting, skruing, lodding	Tilstandsmaskin + liten BC/diffusjon/ACT-policy + håndleddskamera + kraft/moment (ideelt berøring) + menneskelig korreksjon	90–99.4%	10–15% tregere	10–300 demoer (≈15–70 min) + 1 000–3 000 valideringskjøringer (6–22 robottimer)	Gjennomførbart, R&D-tungt måneder for første stasjon
Strukturert teststasjons-mating og OK/NG-sortering	Leverandør VLA-plattform + digital tvilling + RL på ekte robot	≥99.5% vendor claim, conflicting	18–20 s per trinn, innenfor en 20 s takt	≤4 t omtrening + 15 min kalibrering	Gjennomførbart på en leverandørplattform

Elektronikkoppgave	Brukbar elev	Forventet suksess	Hastighet vs en person	Kostnad per omstilling	Konklusjon 2026–2028
Omstillingsprogrammering	LLM skriver eller redigerer robotprogrammer offline, verifisert i simulering	77–99.7% på ABB RAPID-omskrivninger	Ingen effekt på takt	Minutter til timer per program	**Gjennomførbart som ingeniørverktøy**
End-to-end innlært plukking fra bevegelige samlebånd	VLA-elev	37–59%; ≤37% på raske bånd	—	≥458 demoer per oppsett	**Ikke før 2028; bruk klassisk sporing pluss innlært persepsjon**
Nye produkter uten nye data (zero-shot)	Generisk VLA	60–80%; usette kontakter 29–67%	—	—	**Ikke realistisk**
Front-LLM i løkken som driver armen	GPT-6, Claude, Gemini-agenter	Grep 28–45%; presisjonsinnsetting 10%	≈30 min eller 280 s per oppgave vs ≈20 s	—	**Ikke realistisk**
Fritt rutede ledningsnett og flertrinns fleksibel montasje	—	Ingen 2026 produksjonsbevis	—	—	**Ikke før 2028**

Kilder: studiene ovenfor og (University of Passau on RAPID rewriting, GraspVLA, Komax) **mostly single source and vendor data** . I 2026–2028 er en stor modells verdi for roboter som en forhåndstrent prior og en offline-programmerer, ikke som noe som «lærer opp» en elev. Hvert nær-produksjonstilfelle jeg fant (Neuromeka, AgiBot hos Longcheer, Fraunhofer, ContactDP) støtter seg på menneskelige demonstrasjoner, kraft- eller berøringssensorer, simulering eller digitale tvillinger og veiledet RL. Ingen støtter seg på etiketter fra en frontmodell.

Kostnader og tilbakebetaling

04 Kostnader og tilbakebetaling: AI-granskning betaler seg på måneder, maskinmating på år, utplukkingsarmer aldri

Arbeidskostnad: USA er omtrent 5.5 ganger Shenzhen, Taiwan omtrent 1.5 ganger, det meste av Sørøst-Asia under halvparten

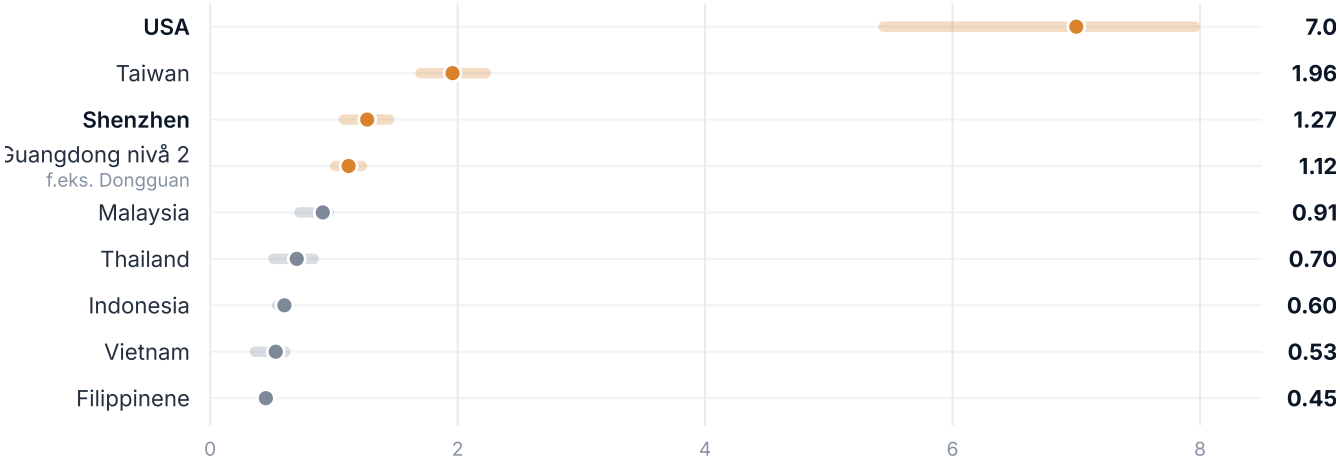
Avkastningen på å erstatte arbeidskraft starter med hva arbeidskraften koster. Shenzhens offisielle lønnsveiledning setter 2024-medianen for en produksjonskvalitetsinspektør til **RMB 74 074** i året; lønn er 86.92% av den totale arbeidskostnaden i elektronikkproduksjon, så fullt belastet er kostnaden omtrent RMB 85 000. JETROs 2025-undersøkelse setter årskostnaden for en kinesisk fabrikkarbeider til $12 667, omtrent RMB 85 100, innenfor 0.2% (Shenzhen wage guidance, JETRO 2025) verified .

Region	Fullt belastet årskostnad for én inspektør eller håndterer (lav / base / høy)	Grunnlag
Fastlands-Kina, Shenzhen	RMB 70k / **85k** / 100k ($10.4k / **12.7k** / 14.9k)	Offisiell median × 1.15; matcher JETRO verified
Guangdong nivå-2 lønnsområder (Dongguan etc.)	RMB 65k / **75k** / 85k	Guangdong 2023-median × 1.15 pluss to års lønnsøkning estimate
Yangtsefloddeltaet (Suzhou, Kunshan, Wuxi)	Som Shenzhen	Intet grunnlag for billigere inspeksjonsarbeid; Jiangsus høyeste minstelønn er omtrent 5.6% over Shenzhens estimate
Taiwan	$16.6k / **19.6k** / 22.7k (NT$534k / **630k** / 729k)	Arbeidsdepartementets yrkeslønnsundersøkelse pluss ≈1.18 lovpålagte tillegg; JETRO som høyre ende estimate
Malaysia	$6.8k / **9.1k** / 10.0k	JETRO og minstelønner estimate
Thailand	$4.7k / **7.0k** / 8.8k	Samme
Indonesia	$5.0k / **6.0k** / 6.3k	Samme
Vietnam	$3.2k / **5.3k** / 6.5k	Samme
Filippinene	$4.0k / **4.5k** / 4.9k	Samme
USA	$54k / **70k** / 80k	BLS mai 2025 inspektørmedian $48 570 ÷ lønn ved 65.5% av kompensasjon estimate

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

FIG 17 **Hva én inspeksjons- eller håndteringsarbeider koster per år** ***Fullt belastet er USA omtrent 5.5× Shenzhen og Taiwan omtrent 1.5×.***

årlig full kostnad per QC- eller håndteringsarbeider ($10k; punkt = base, søyle = lav–høy)



Kilde: Shenzhens offisielle lønnsveiledning, Taiwans arbeidsdepartement, JETRO, BLS; forfatterens estimater.

Tre korreksjoner setter grunnlaget. For det første hevet Shenzhen pensjonsbidragsgulvet fra RMB 4 775 til 5 521 i måneden i juli 2026 (Shenzhen Bendibao) single source, så en Shenzhen-inspektør i 2026 koster mer sannsynlig RMB 80–90k. For det andre er en stilling ikke en person: to 12-timers skift trenger omtrent 2.0 personer (seks dager i uken) til 2.4 (hele året) i Perlefloddeltaet, omtrent 2.7–3.1 i Sørøst-Asia under overtidstak, og 4.2–4.5 for 24×7 i USA og Taiwan estimate. AI-granskning lar vanligvis én gransker dekke to til fire linjer, så den realistiske besparelsen er **0.5–1.5 heltidsansatte per linje, ikke én person per linje per skift**. For det tredje jager kjøpere i Taiwan, Vietnam og Malaysia ikke bare kostnad: i JETROs undersøkelse sa 61.1% av firmaene i Taiwan, 48.2% i Vietnam og 44.9% i Malaysia at ansettelse hadde blitt vanskeligere de siste to årene, mot 18.6% i Fastlands-Kina (JETRO 2025) single source.

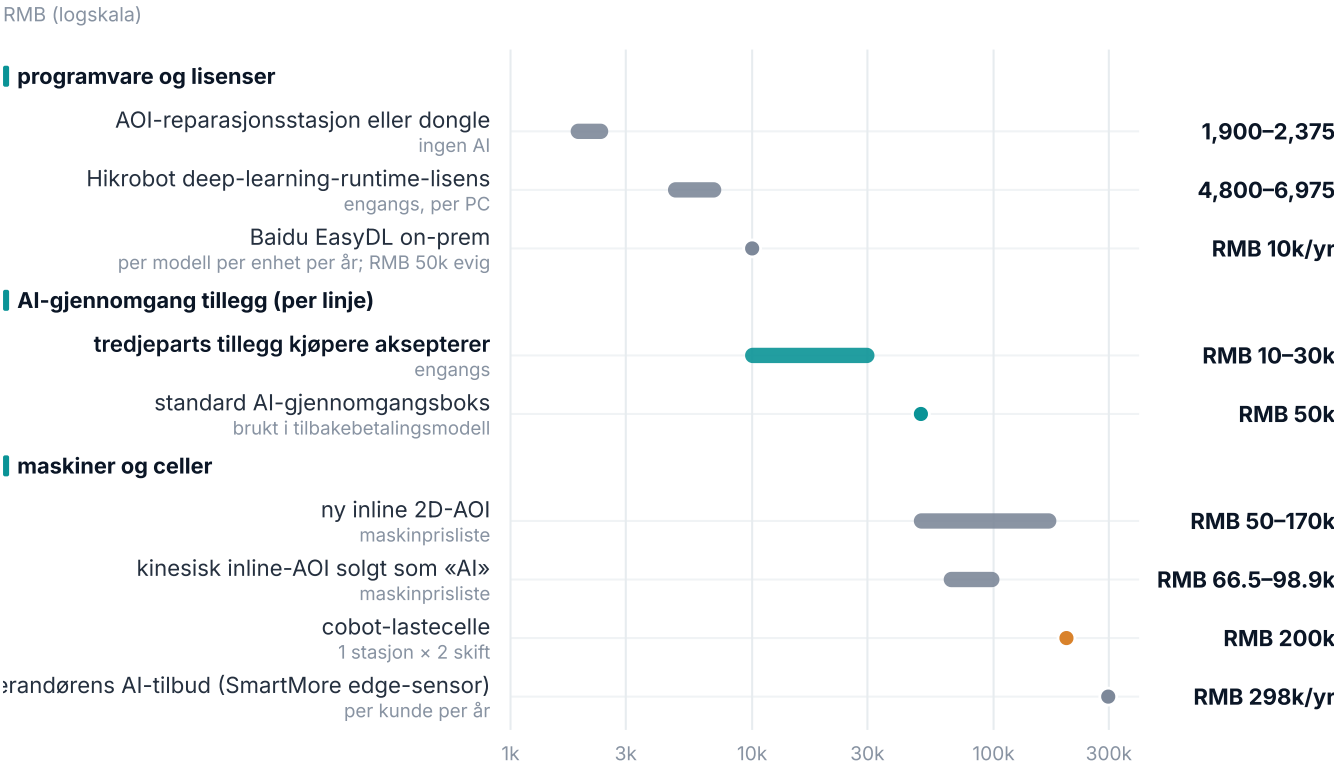
Utstyrspriser: fire grunnlag som ikke bør midles

AOI-priser er uenige hovedsakelig fordi de måler forskjellige ting. Nettannonser (november 2026, nye, brukte og leide blandet) ligger omtrent RMB 17–80k for offline eller stasjonær 2D, RMB 50–170k for inline 2D og RMB 80–475k for inline 3D, opptil RMB 700k for koreansk 3D (1688 inline AOI, 1688 offline AOI, 1688 3D AOI) single source, list prices. Et anbud fra november 2025 satte tak på én 3D AOI pluss én 3D SPI til RMB 800k (tender notice). Sinic-Teks prospekt antyder RMB 428–512k per 3D AOI i 2019–2022, omtrent RMB 340–470k i 2025-rapporten (Sinic-Tek prospectus, Sinic-Tek 2025 annual report) primary + estimate. En oversjøisk leverandør anslår $3 000–20 000 for stasjonær 2D, $30–70k inline 2D, $50–90k entry 3D og $100–200k+ high-end 3D, med samme inline-maskin til $30–50k i Kina (Averroes) vendor estimate from a seller of competing AI software.

AI-granskning i seg selv har ingen offentlig RMB-pris; leverandører oppgir pris fra sak til sak etter en gratis evaluering på stedet (Sohu on Hengtianyi), og søk i e-handelsannonser, anbud, leverandørartikler og taiwanske medier fant ingen offentlig pris (1688, 51SMT). Ankrene: omtrent €27k per linje antydet av Siemens-tilfellet estimate; $21 924 listepris for en US-distributørs fjerngranskings-nettverksmodul (Manncorp) vendor list price; omtrent RMB 298k per kunde per år for SmartMores edge-AI-sensorer i 2025 (36Kr) single source. En innenlandsk inline 2D-AOI solgt som «AI» listes til RMB 66.5k (1688 inline AOI), og en NG-buffer for manuell re-inspeksjon omtrent RMB 50k (1688 AOI review). Så taket for et tredjeparts granskningstillegg er omtrent RMB 10–30k per linje, eller en tilsvarende årsavgift; over det vil kjøperen heller erstatte maskinen med en innenlandsk AOI som har AI innebygd estimate.

Anker	Pris (RMB)	Hva det kjøper	Kilde
AOI-reparasjonsstasjon eller programvaredongle (eldre TRI, Zhenhuaxing)	≈1 900–2 375	Lisensnøkkel, ingen AI	1688 list price
Generell dyp-læring synsruntime (Hikrobot VisionMaster-dongle)	≈4 800–6 975 engangs per PC	Plattformen; kunden bygger og vedlikeholder modeller	1688 list price
Baidu EasyDL lokal utrulling	10 000 per modell per enhet per år, eller 50 000 evigvarende	Pakket inferenstjeneste	Baidu AI docs 2022 price page
Innenlandsk inline AOI solgt som «AI»	66 500–98 900 per maskin	Maskinerstatning	1688 list price
SmartMore edge-AI-sensorer	≈298 000 per kunde per år	Leverandørløsning over flere maskiner	36Kr single source

FIG 18 **Prisstigen***Fra noen få tusen RMB programvarelisens til RMB 200–300k for en robotcelle eller et AI-tilbud fra en leverandør.*



Kilde: 1688-oppføringer, Baidu AI-dokumentasjon (prisliste 2022), 36Kr, et bransjeintervju (ikke verifisert).

En Baidu-stil årsavgift, med én inferensserver for én til fire linjer og én til fire modeller hver, antyder et abonnement på RMB 2.5–40k per linje per år, realistisk omtrent RMB 5–10k; ved 20% av en evigvarende lisens brytes abonnement og kjøp like i år fem, en struktur verdt å kopiere estimate. Kjøpere er vant til nær null programvareavgifter: hele Sinic-Teks «øvrige» omsetning (reservedeler, service, leie, tilbehør) var 7.65% av 2025-salget, omtrent RMB 2 300–3 100 per installert maskin per år (Sinic-Tek 2025 annual report) primary + estimate. Et AI-abonnement på RMB 5–10k per linje per år er to til fem ganger det kjøpere nå bruker på AOI-programvare, så det må rettferdiggjøres av granskningstimene det frigjør (omtrent RMB 75–90k per heltidsansatt i året), ikke av sammenligning med eksisterende programvareavgifter.

For robotarmer falt Kinas gjennomsnittlige cobotpris til omtrent RMB 58 600 i 2025 (GGII) estimate, og Dobots seksakse-snitt falt 19.1% det året da billigere modeller tok andeler (Huxiu) verified, selv om listeprisene for de samme

modellene steg 5–15% i 2026 (Sina Finance). I USA listet en FANUC CRX-10iA til $51 434 (Vention) og sammenstillinger satte en UR7e til $30–48k (cobotcost) estimate. Armen er bare 40–50% av en US-celle (Standard Bots); i Kina løper kameraer, gripere, sikkerhet og ingeniørarbeid lett to til fire ganger en RMB 40k-arm, så budsjetter absolutt ingeniørkostnad snarere enn et multiplum estimate.

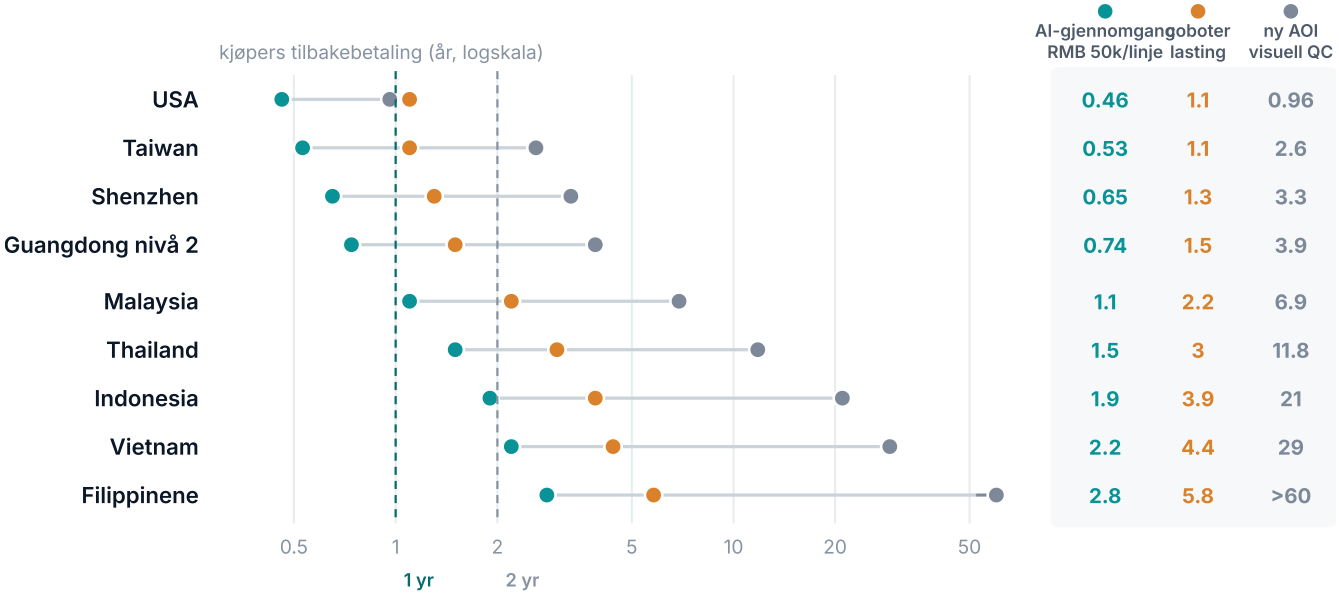
Tilbakebetaling: granskning på måneder, mating på år, utplukking aldri

Enkel tilbakebetaling = investering ÷ (sparte personer × belastet arbeidskostnad − årlig driftskostnad), for en to-skiftslinje uten subsidier eller toll. Taiwan-utstyr prises til 1.3 ganger Kina, Vietnam, Malaysia og Thailand til 1.15 ganger, Indonesia og Filippinene til 1.2 ganger, og USA til oversjøiske priser estimate.

Scenario	Investering (Kina / USA)	Sparte personer	Årlig driftskostnad
(a) AI-granskningsboks eller AI-kameraboks	RMB 50k / $30k per linje	1.0 per linje	15%
(b) Ny inline 2D-AOI som erstatter manuell visuell kontroll (0.5 gransker per skift gjenstår)	RMB 200k / $60k	1.0	12%
(c) Cobot-lastecelle (én stasjon × to skift)	RMB 200k / $140k	2.0	8%
(d) Integrert detekter-og-plukk-celle (inspeksjon 1.0 + plukking 0.07)	RMB 300k / $200k	1.07	8%
(e) En robotarm lagt til bare for å plukke ut NG-kort	RMB 150k / $100k	0.07	8%

Region	(a) AI-granskning	(b) Ny AOI	(c) Cobot-celle	(d) Detekter-og-plukk	(e) Utplukkingsarm
Shenzhen	0.65 yr (0.54–0.80)	3.3 (2.6–4.4)	**1.3** (1.1–1.6)	4.5 (3.6–5.9)	Aldri
Guangdong nivå-2	0.74	3.9	1.5	5.3	Aldri
Taiwan	0.53	2.6	1.1	3.6	Aldri
Malaysia	1.1	6.9	2.2	9.2	Aldri
Thailand	1.5	11.8	3.0	15	Aldri
Indonesia	1.9	21	3.9	25	Aldri
Vietnam	2.2 (1.6–4.5)	29	4.4	34	Aldri
Filippinene	2.8	>60 (driftskostnad ≈ besparelse)	5.8	≈100	Aldri
USA	0.46	**0.96**	**1.1**	3.4	Aldri

FIG 19 **Tilbakebetaling per region** ***AI-gjennomgang betaler seg på måneder i USA, Taiwan og Shenzhen og på 1–3 år i Sørøst-Asia; cobot-mating på år.***



en robotarm kun for å plukke ut NG-kort betaler seg aldri, noe sted
ved first-pass yield over 90% tar utplukking bare 0.003–0.14 FTE per skift

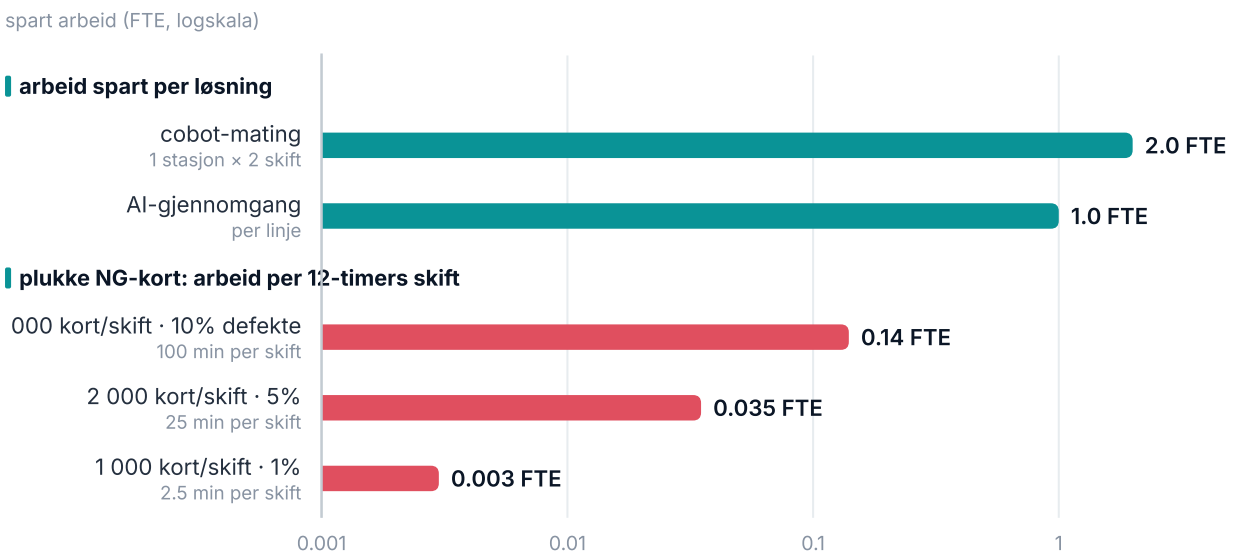
Kilde: Forfatterens modell: RMB 50k AI-gjennomgangsboks per linje; cobot-matingcelle på RMB 200k eller $140k, to skift.

Alle verdier er estimater fra inndataene ovenfor. Følsomheter: halver de sparte personene og hver tilbakebetaling omtrent dobles; kjør ett skift og (b)–(d) dobles; 100% førsteårsavskrivning i USA forkorter tilbakebetaling med omtrent 20%, mens toll på kinesisk, japansk eller koreansk utstyr forlenger den med omtrent 10–40%; et 15% lokalt subsidie i Kina forkorter den med omtrent 15%. Jeg fant ingen uavhengig undersøkelse av kjøperes tilbakebetalingsterskler; leverandører sier AOI betaler seg «typisk på 9–18 måneder» (Averroes) og Siemens-tilfellet hevder åtte måneder vendor claims. Jeg bruker ett og to år som screeninglinjer. Modellen utelater kvalitetsgevinster (garanti, returer) og turnover- og opplæringskostnader, så den er konservativ.

Den kontraintuitive delen av «detekter og plukk» er at arbeidet ligger i granskningen, ikke i å ta de få ekte NG-kortene av linjen. Ved 15 sekunders håndtering per NG-kort estimate:

Kort per skift	Defektrate	Plukketid per 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 plukking av dårlige kort ikke betaler seg** ***Over 90% first-pass yield tar utplukking bare 0.003–0.14 av en person per skift.***



Kilde: Forfatterens estimater.

Basistilfellet (2 000 kort per skift, 5% defekte, to skift) er 0.07 av en person, verdt omtrent RMB 6 000 i året i Shenzhen, mindre enn armens omtrent RMB 12 000 årlige driftskostnad, så den betaler seg aldri noe sted; selv verste tilfellet (4 000 kort per skift, 10% defekte) tar omtrent 13 år i Shenzhen og 9 i USA estimate. Hvis bekymringen er å stoppe linjen for å plukke, løser en standard NG-avleder eller buffertransportør det uten en arm. Folk i bransjen sier «over 90% first-pass yield, stopp linjen og ta kortet ut for hånd», og **det er riktig for plukking, men ikke for inspeksjon**. Produktdesignet på en SMT-linje er derfor «inspiser pluss merk eller avled», med robotarmer holdt for steder der hver del håndteres (testfiksturlasting, brett- og magasinhåndtering, pakking), der inspeksjon blir et nærmest gratis tillegg og scenario (c) gjelder.

Subsidier, skatt og toll flytter tilbakebetaling med 15–25%, men ikke rangeringen

Fastlands-Kina: støtte senker mest finansieringskostnaden; enkeltmaskiner kvalifiserer sjelden direkte. Et nasjonalt lånesubsidie betaler 1.5 rentepoeng i opptil to år i 14 prioriterte kjeder (Yicai) verified, omtrent RMB 9 000 på en RMB 300k-maskin; Guangdongs oppgraderingsbelønning på opptil 20% krever et registrert prosjekt med minst RMB 7 millioner nytt utstyr og ekskluderer Shenzhen (Guangdong rules), og Dongguan betalte faktisk 4.42% i juli 2026 (Dongguan allocation) single source, primary. En AI-inspeksjonsenhet godkjennes bare som utstyr «med støtteprogramvare» inne i kjøperens registrerte prosjekt, så basistilfellet antar ingen subsidie.

Taiwan: insentivene går til kjøperen. Under paragraf 10-1 i Statute for Industrial Innovation, endret i mai 2025 til å inkludere AI, kan en kjøper kreditere 5% av 2025–2029-utgiftene til nytt smart maskineri eller AI (inkludert CNN/DNN-basert inspeksjon) mot det årets inntektsskatt, eller 3% i året i tre år (Statute, credit rules) verified; leverandørsidige tilskudd er lukket for fastlandskapital (MOEA) single source.

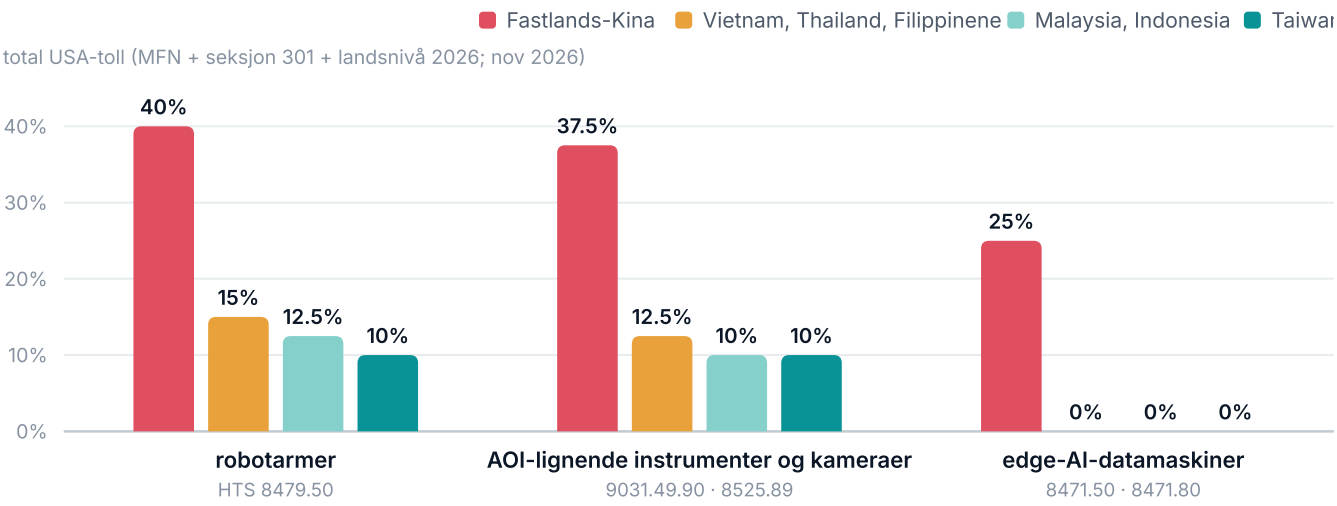
Sørøst-Asia: bare Malaysia og Thailand belønner kjøp av automatisering. Malaysias 200% automatiseringskapitalfradrag på de første RM10 millionene er verdt omtrent 24% av capex for en lønnsom kjøper (MIDA, PwC Malaysia) verified; Thailands BOI gir tre års skattefritak opptil 50–100% av en oppgraderingsinvestering, det høyere taket når thailandske automatiseringsleverandører er involvert (BOI) single source, primary, så inngang i Thailand fungerer best med en lokal integratør. Vietnam, Indonesia og Filippinene belønner bygging av fabrikker snarere enn kjøp av automatisering (Vietnam Law Magazine, PwC Philippines).

USA: avskrivning forkorter tilbakebetaling; toll forlenger den. Eiendeler kjøpt etter 19. januar 2025 får permanent 100% førsteårsavskrivning, omtrent et 21% føderalt skatteskjold for en lønnsom kjøper (IRS Notice 2026-11) single

source, primary ; seksjon 232-undersøkelsen av roboter og industrimaskineri ble avsluttet 28. oktober 2026 med forhandlinger snarere enn ny toll, som skal revurderes innen 180 dager (White House, Mayer Brown) verified . Sluttmontasje i Sørøst-Asia kutter påslaget på armer og AOI-lignende instrumenter fra omtrent 37.5–40% til 10–15% (kapittel 1).

Region	Hovedinstrumenter	Effekt på kjøpers tilbakebetaling estimate
Fastlands-Kina	1.5-poeng lånesubsidie i opptil to år; provinsielle og kommunale oppgraderingsbelønninger (Guangdong krever ≥RMB 7M registrerte prosjekter; Dongguan betalte faktisk 4.42%); Shenzhen 2026–2029-tiltak (satser ikke offentlige)	Omtrent 0–5% for en enkeltmaskin; 4–15% inne i et stort prosjekt
Taiwan	Statute §10-1-kreditt, 5% eller 3% × 3 år	Omtrent 5% kortere, hvis kjøperen betaler nok skatt
Malaysia	200% automatiseringskapitalfradrag (første RM10M)	Omtrent 15–25% kortere for lønnsomme kjøpere
Thailand	BOI tre års skattefritak opptil 50–100% av investeringen	Omtrent 15–25% kortere for lønnsomme kjøpere utenfor skatteferie
Vietnam, Indonesia, Filippinene	Inntektsskatteinsentiver for bygging av fabrikker; ekstra 20% utstyrsavskrivning i Filippinene	Nesten ingenting for enkeltstående automatiseringskjøp
USA	100% førsteårsavskrivning; ≈37.5–40% kombinert toll på fastlandslaget utstyr	Avskrivning ≈20% kortere; toll ≈10–40% lengre

FIG 21 **USA-toll på roboter, AOI og edge-datamaskiner** ***Armer laget i fastlands-Kina betaler omtrent 40%; sendt fra Sørøst-Asia eller Taiwan, 10–15%.***



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

Per region: i Fastlands-Kina betaler en AI-granskningsboks seg på måneder og en cobot-celle på 1.3–1.5 år, mens en ny AOI kjøpt bare for å kutte arbeid tar 3–4 år og integrert detekter-og-plukk 4.5–5.3, begge utenfor en toårs-screening. Taiwan gjør det litt bedre enn Shenzhen (0.53 år for granskning, 1.1 for en cobot-celle), og ansettelse der er vanskeligere. I Malaysia og Thailand tar en AI-boks 1–1.5 år og en cobot-celle 2–3, omtrent 15–25% raskere med lokale skatteinsentiver for lønnsomme kjøpere. I Indonesia, Vietnam og Filippinene betaler bare en svært billig AI-boks seg innen 2–3 år; etterspørselen der kommer mer fra kundenes kvalitetskrav og arbeidskraftmangel. I USA be-

taler hvert alternativ unntatt en ren utplukkingsarm seg på 0.2–3.4 år, med toll, seksjon 232-revurderingen og NRTL-sertifisering som hovedvariablene.

05 Marked og konkurranse

Marked, konkurranse og hvordan kjøpere betaler

Markedsstørrelse: publiserte tall skiller seg med multipler; bare en nedenfra-og-opp-telling er nyttig

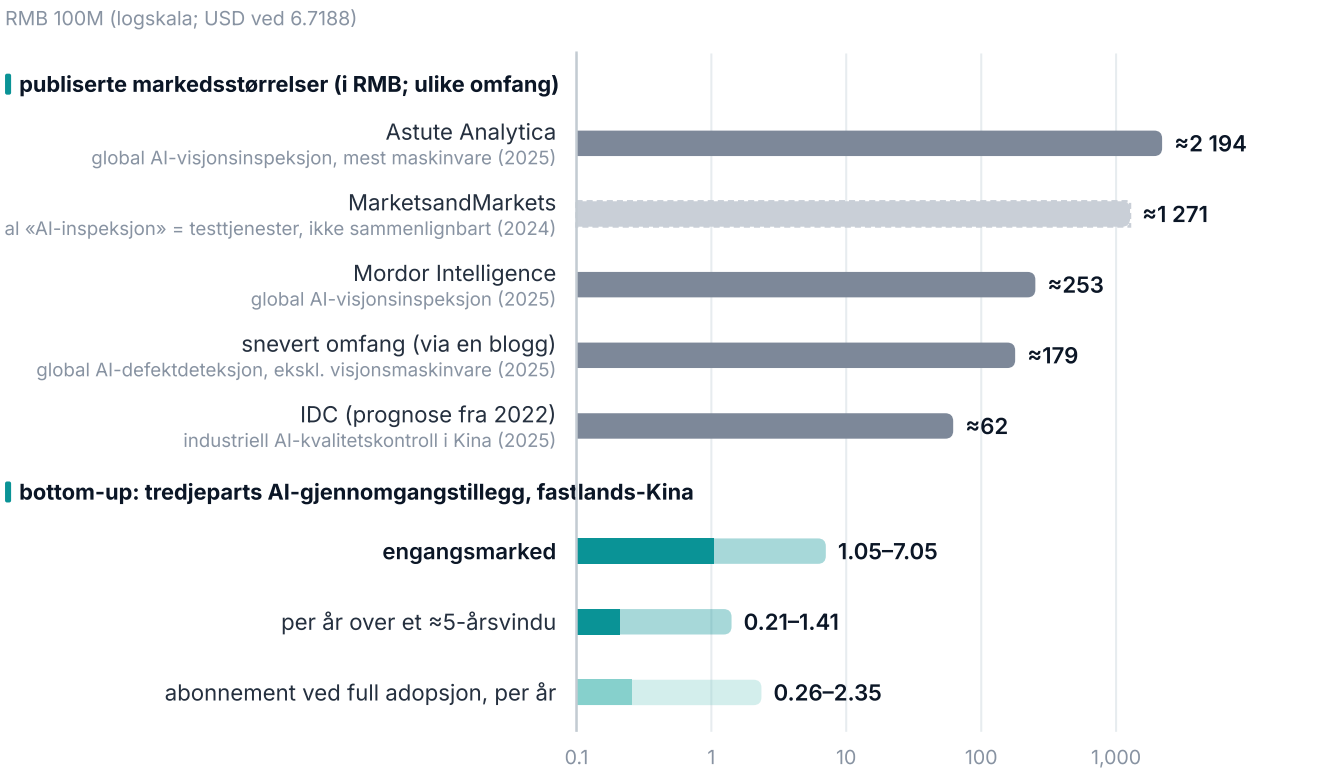
Hvert markedstall må leses med sitt omfang. Kinas maskinsynsmarked var RMB 18.15 milliarder i 2024 (Zhiyan) verifiied, med 3C-elektronikk som største kjøper; globale estimater for maskinsyn skiller seg omtrent tre ganger og for AOI omtrent 6.5 ganger (MarketsandMarkets, Vision Systems Design/Interact Analysis, Zhiyan). Nedenfra og opp fra omsetningen til Koh Young, TRI, Jutze, Sinic-Tek og ViTrox er globalt SMT-inspeksjonsutstyr i størrelsesorden $1–2 milliarder estimate.

Det finnes intet selvstendig marked for «AI visuell inspeksjonsprogramvare». De publiserte tallene skiller seg med mer enn ti ganger og inkluderer alle maskinvare eller tjenester:

Omfang	Utgiver	Tall
Global «AI industriell defektdeteksjon» (snevert, ekskludert synsmaskinvare)	Unavngitt firma, via en iFactory-blogg	$2.66 milliarder i 2025, 8.6% CAGR til 2035, elektronikk 34%
Global «AI visuell inspeksjon»	Mordor Intelligence	$3.76 milliarder i 2025, $12.8 milliarder i 2031, 22.67% CAGR
Global «AI visuell inspeksjon» (mest maskinvare)	Astute Analytica	≈$32.66 milliarder i 2025
Global «AI-inspeksjon» (testing- og sertifiseringstjenester; ikke sammenlignbart)	MarketsandMarkets	$18.92 milliarder i 2024
Kina «industriell AI-kvalitetsinspeksjon»	IDC, prognose laget i 2022	$958 millioner i 2025 (≈RMB 6.2 milliarder), 91.5% i kommunikasjon, elektronikk og to andre bransjer

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

FIG 22 **Markedsstørrelser skiller seg mer enn tidobbelt** ***Bare det snevre omfanget og en bottom-up-telling er nyttige for beslutninger.***



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

Nedenfra og opp finnes det bare én offentlig telling av SMT-linjer: New Venture Research estimerte omtrent 73 231 linjer globalt i 2020, 47% av dem i Kina (19% i 2000), omtrent 70% high-end og 30% low-end (EMSNow, NVR SM-T2020 summary) single source, 2020 ; NVR tilhører nå IPC og nyere data er betalte (NVR). Å vokse 3–5% i året med en 40–47% Kina-andel gir omtrent 87 000–98 000 linjer globalt i 2026 og 35 000–46 000 i Fastlands-Kina (base 40 000) estimate .

Inndata	Lav	Base	Høy	Grunnlag
Fastlands-Kina SMT-linjer (2026)	35 000	40 000	46 000	NVR 2020-base × vekst × andel
Andel uten AI-granskning	75%	80%	85%	Retrofitvindu, nedenfor
Andel en tredjepart kan nå (ikke intern, ikke knyttet til OEM-AI, kan betale)	40%	50%	60%	Antakelse: ≈30% low-end-linjer pluss noen blandet-merkede high-end-linjer, ekskludert DBG-stil interne bygg
Servicebare linjer	**10 500**	**16 000**	**23 500**	Produkt
Engangspris per linje	RMB 10k	RMB 20k	RMB 30k	Prisstige og tak (kapittel 4)
Engangs servicebart marked (installert base)	**RMB 105M**	**RMB 320M**	**RMB 705M**	Produkt
Per år over et ≈5-års vindu	RMB 21M	RMB 64M	RMB 141M	Vindu krymper 8–12% i året
Abonnement (RMB 2.5k / 5k / 10k per linje per år, full adopsjon)	RMB 26M i året	RMB 80M i året	RMB 235M i året	Baidu-stil årsavgift-anker

FIG 23 **Bottom-up: 40 000 linjer → 16 000 servicebare → ≈RMB 320M engangs *Tredjeparts-AI-gjennomgangsnisjen i fastlands-Kina er reell, men liten.***



en nykommer som tar 2–5% av servicebare linjer på 3 år: ≈320–800 linjer, ≈RMB 2–5M per år
som abonnement (RMB 5k per linje per år): RMB 1.6–4M gjentakende, bare 0.3–2.3% av IDCs prognose for industriell AI-QC i Kina

Kilde: NVR (base 2020); forfatterens estimater.

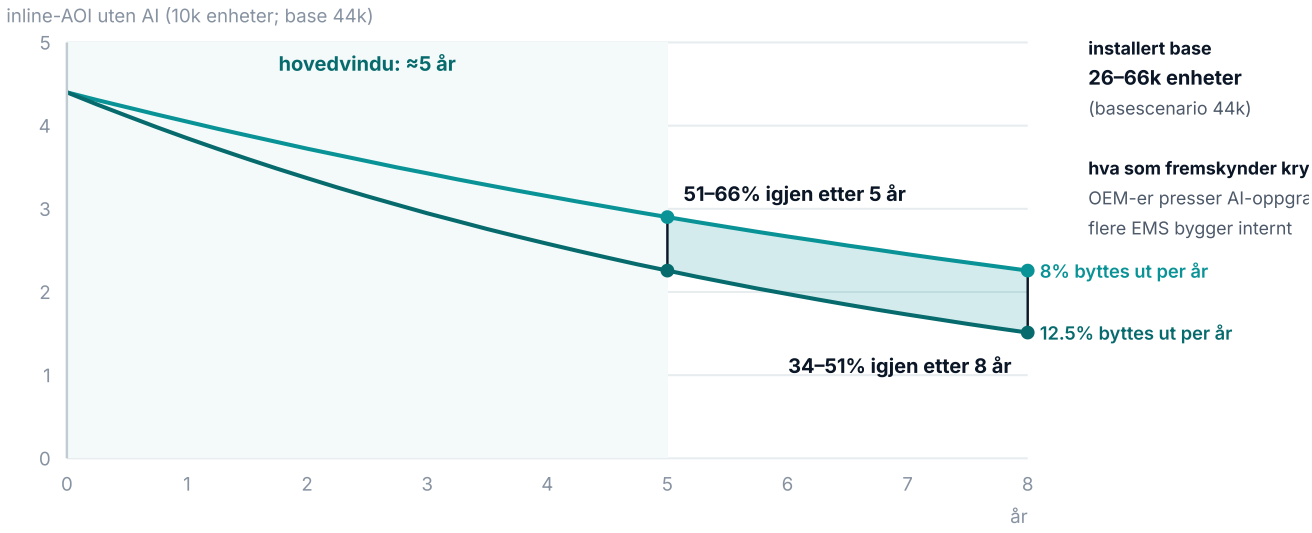
Alt dette er estimater; verken andel eller pris er målt offentlig. Nisjen er verdt bare 0.3–2.3% i året av IDCs prognose for hele Kinas industrielle AI-kvalitetsinspeksjon i 2025 (omtrent RMB 6.2 milliarder). En nykommer som vinner 2–5% av base-servicebare linjer på tre år ville ha omtrent 320–800 linjer: RMB 6.4–16 millioner totalt ved RMB 20 000 per linje, omtrent RMB 2–5 millioner i året, eller RMB 1.6–4 millioner i året i gjentakende omsetning ved RMB 5 000 per linje estimate. Det er nok for en forretningsenhet inne i et selskap, ikke for et venture-skala selvstendig selskap. Hver 1 000 installerte linjer i Taiwan, Sørøst-Asia eller USA er verdt omtrent RMB 3–15 millioner engangs til fastlandspriser; prisene er høyere i Taiwan og USA, men kjøpere der bruker oftere OEM 3D-AOI-er med AI-alternativer estimate.

Retrofitvinduet er omtrent fem år

Ingen AOI-produsent offentliggjør maskinlivslengde; EMS-selskaper avskriver utstyr over 3–10 år (DBG 2025 annual report), og eldre AOI-er mister programvareoppdateringer og reservedeler etter end of life (I-Connect007) single source. Jeg bruker 8–12 år (base 10).

Ved 1.0–1.65 inline-AOI-er per linje har fastlandet omtrent 35 000–78 000 inline-AOI-er i bruk (base omtrent 55 000), i samsvar med kjente forsendelser (Circuits Assembly, Sinic-Tek 2025 annual report). Ingen kilde måler hvor mange som kjører AI-granskning; hvis man antar innenlandske maskiner har tilbudt det siden omtrent 2023 og 40–60% av nylige leveranser bruker det, **kjører omtrent 26 000–66 000 inline-AOI-er i Fastlands-Kina (base omtrent 44 000) ikke AI** estimate. Hvis hver nye maskin nå leveres med AI, krymper den poolen 8–12.5% i året: 51–66% gjenstår etter fem år og 34–51% etter åtte, raskere hvis AOI-produsenter presser AI-oppgraderinger til nyere maskiner. **Mesteparten av muligheten faller i de neste fem årene eller så**, med mellomstore EMS som kjører eldre innenlandske 2D-AOI-er, blandede merker og ingen AI-oppgraderingsvei i kjernen.

FIG 24 **Retrofitvinduet er omtrent fem år** ***Inline-AOI uten AI krymper med 8–12.5% per år etter hvert som de byttes ut eller oppgraderes.***

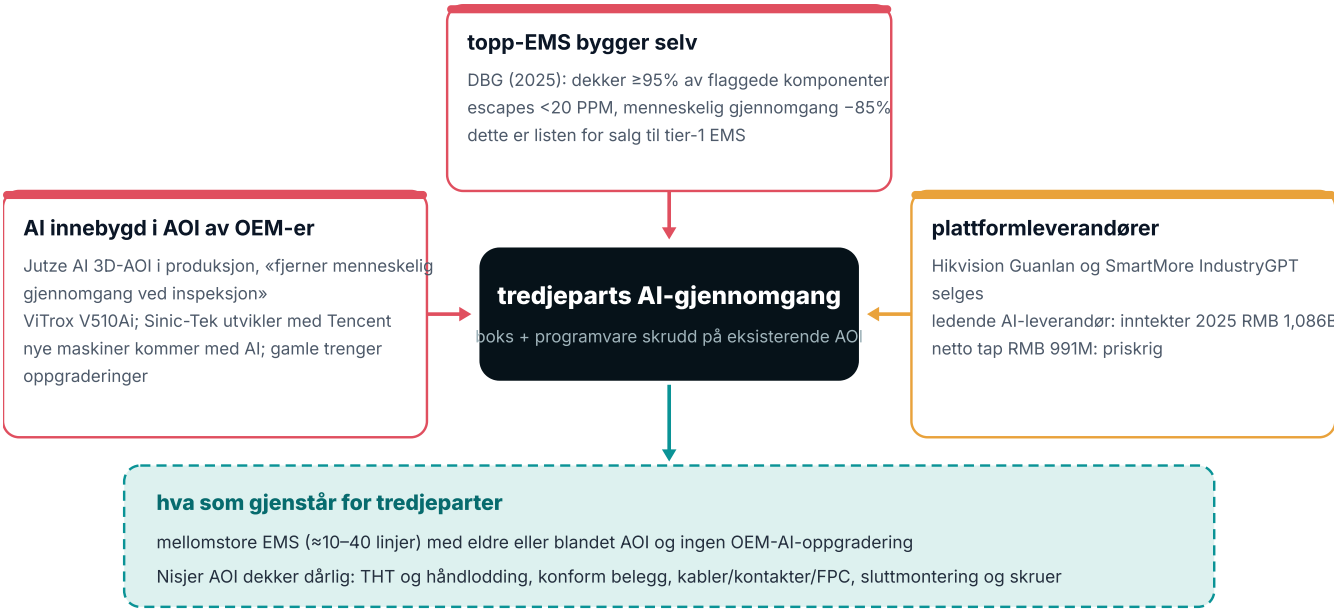


Kilde: Forfatterens estimater; installert base og andeler ikke uavhengig verifisert.

Konkurrentene er allerede sterke: granskningskilen er omringet

AI-granskningskilen smalner fra tre retninger. AOI- og synsledere bygger AI inn i nye plattformer: Jutze har «fullt tatt i bruk AI-algoritmer» (Jutze investor relations, June 2026), OPTs industrielle AI-omsetning vokste 96.77% i 2025 (OPT 2025 annual report) single source, primary, og ViTrox hevder opptil 95% nøyaktighet for autonom granskning (ViTrox) vendor claim. Store produsenter bygger sitt eget (iThome). Og venturefinansierte AI-selskaper selger i skala: SmartMores edge-sensorkunder vokste fra 65 til 251 på to år (36Kr) single source, Hikvision selger Guanlan-modellen sin for inspeksjon (Xinhua), og «øye, hjerne og hånd»-integrasjon tilbys allerede fra SmartMore, Hikrobot og Mech-Mind (36Kr, Securities Star).

FIG 25 **Gjennomgangskilen er omgitt på tre sider** ***Rom for tredjeparter finnes i mellomstore EMS og i inspeksjonsnisjer AOI dekker dårlig.***



Kilde: Jutze, Sinic-Tek og DBGs årsmeldinger; Hikvision og SmartMore offentlig materiale; sammenstilt av forfatteren.

Plassen som er igjen for en nykommer er smal: mellomstore EMS med eldre 2D/3D-AOI-er og ingen AI-oppgradering fra produsenten; linjer som blander flere AOI-merker (én retrofit-leverandør støtter allerede Omron, VI, Koh Young og Mirtec (Sohu on Hengtianyi)); nisjer SMT-AOI-er dekker dårlig, som gjennomhulls- og håndloddede skjøter, konformt belegg, kabler, kontakter og FPC-er, og sluttutseende og skruer; og små fabrikker som ikke har råd til en RMB 230–280k-AOI.

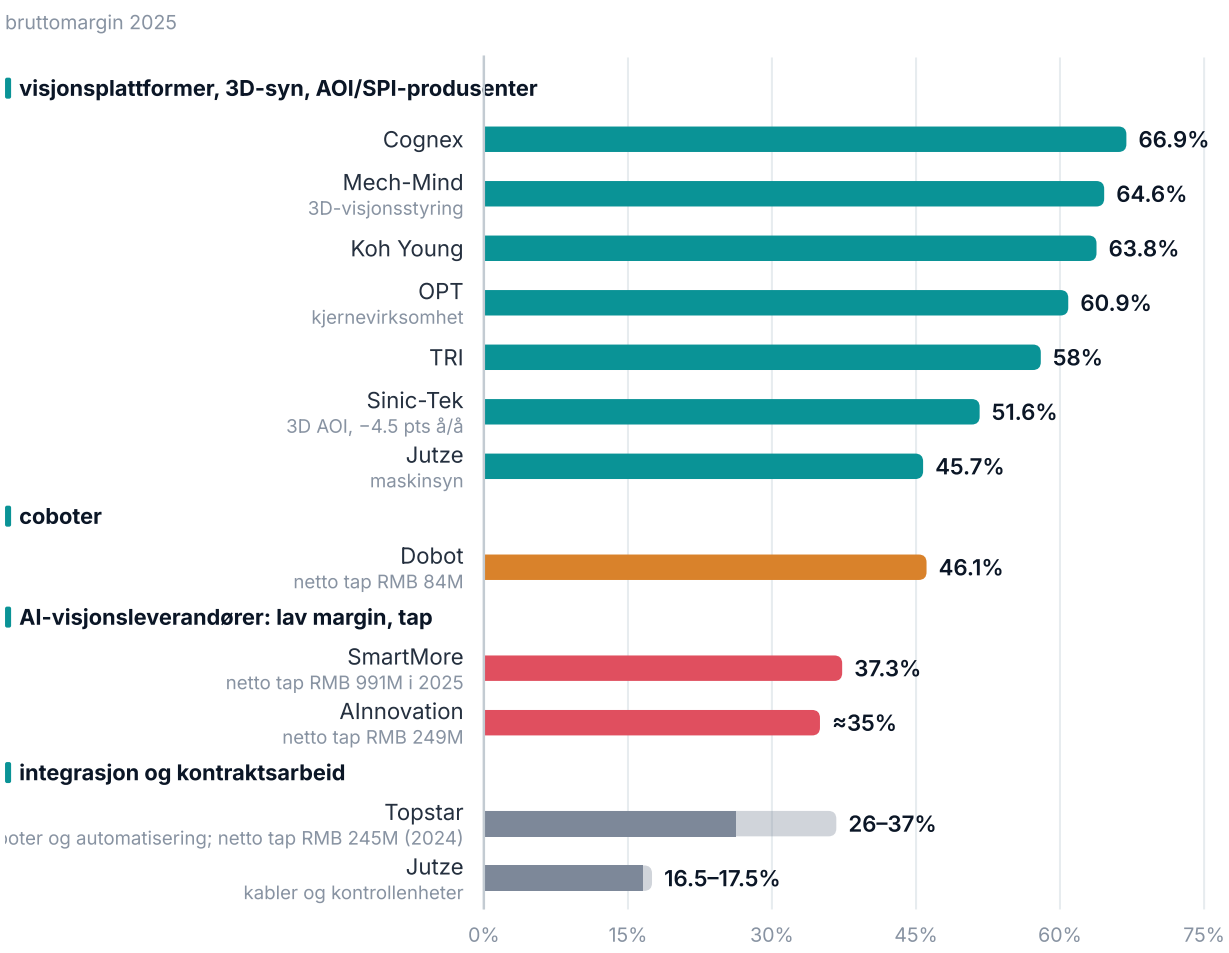
Det nærmeste børsnoterte sammenlignbare gir en advarsel. Weiyi Zhizao, som selger AI-inspeksjonsutstyr og roboter, så 3C-omsetning falle 79.9% år over år i januar–mai 2026, anerkjente «intensiverende konkurranse og prispress», utsatte Hongkong-noteringen 25. oktober 2026, og offentliggjorde 15 prosjekter avsluttet før fullføring, 12 av dem embodied-robot-prosjekter (Weiyi Zhizao prospectus, postponement notice) single source, primary. AInnovation avsluttet 2025 med 29.8% av fordringene avsatt (AInnovation 2025 annual report). 3C-inspeksjonskilen er overfylt, og priskraften per linje er svak.

Marginen sitter i produktifisert maskinvare

Posisjon i kjeden	Representative selskaper og 2025 bruttomargin	Resultat og inndrivelse
Synsplattformer	Cognex 66.9%	Lønnsomt; avhender ≈$22M lavmargin-omsetning
AOI/SPI-utstyr	Koh Young 63.8%; TRI 58%; OPT kjernevirksomhet 60.85%; Sinic-Tek 3D AOI 51.62% (−4.53 poeng YoY); Jutze maskinsyn 45.74%	Lønnsomt; fordringsdager 139 (Jutze) til 247–274 (OPT)
3D synsstyring	Mech-Mind 64.6%	—
AI synsløsningsleverandører	SmartMore 37.3%; AInnovation ≈35%	Tapsbringende: SmartMore netto tap RMB 991M i 2025, AInnovation RMB 249M
Coboter	Dobot 46.1%	Dobot netto tap RMB 84M; UR-omsetning fra $326M (2022) til ≈$232M (2025, prognose)
Systemintegrasjon	Topstar robotikk og automatisering 26.2–36.7%	Netto tap RMB 245M (2024); fordringsdager 180–240
Kontraktsmaskinvare	Jutze kabler og kontrollenheter 16.5–17.5%	—

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

FIG 26 **Marginen sitter i produktifisert maskinvare** ***Visjonsplattformer og utstyrsprodusenter tjener 46–67% bruttomargin; AI-leverandører 35–37% og går med tap.***



Kilde: Årsmeldinger eller prospekter 2025.

En nykommer beholder marginen sin bare ved å gjøre inspeksjon til et produkt, utstyr eller programvare; prosjekt-for-prosjekt-tuning presser marginene til omtrent 35% med vedvarende tap. Mekanikken virker også mot små selskaper: omsetning innregnes først etter aksept på stedet, forskuddene er små (Jutze, OPT), piloter er vanligvis gratis (A3/Mindtrace), og merkeeiere spesifiserer ofte inspeksjonsutstyret på kontraktsprodusentenes linjer.

Hvordan kjøpere betaler: 30% ved signering, 60% etter aksept, 10% holdt i et år

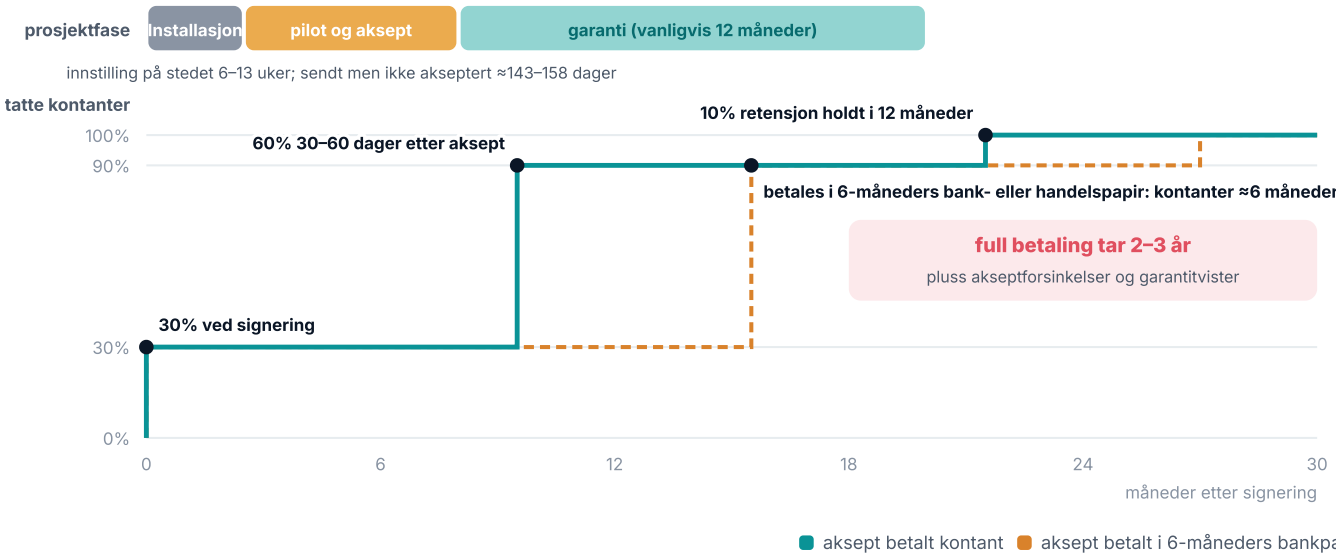
Betalingsvilkårene fastlandselektronikkfabrikker bruker for inspeksjonsutstyr er beskrevet i 13 navngitte kjøperkontrakter offentliggjort i Sinic-Teks noteringssvar (Sinic-Tek inquiry reply 8-1-1) single source: one primary filing, several independent buyers, 2020–2022 :

Kjøper (år)	Signering	Levering	Aksept	Tilbakeholdelse eller saldo
En Hikvision-enhet (2021)	30% innen 15 dager	—	65% innen 30 dager etter endelig aksept	5% garantitilbakeholdelse betalt 12 måneder etter aksept; del betalt i 6-måneders+ bankakseptsedler
Nanchang Huaqin (2020–2021)	30% innen 10 virkedager	—	60% ved 60 dager etter aksept	10% etter ett års garanti
Guangdong Yinuo (2021, 3D AOI)	30% innen 10 virkedager	—	60% innen 30 dager	10% etter tre måneders feilfri kjøring

Kjøper (år)	Signering	Levering	Aksept	Tilbakeholdelse eller saldo
Xiamen Hongxin, Jingmen Hongyi (2022)	30% innen 10 virkedager	—	60% innen 30 virkedager etter endelig aksept, 6-måneders banksedler	10% etter ett års garanti
AKM Panyu, Suzhou (2021–2022)	30% innen 7–10 virkedager	—	60% en måned etter aksept og fakturering, 180-dagers banksedler	10% etter 360 dager
BYD Precision, Shenzhen, Huizhou (2020)	—	30% ved ankomst	70% etter aksept og fakturering	Alt i 6-måneders kommersielle sedler
Shenzhen H&T (2020)	—	—	90% etter aksept og fakturering	10% etter seks måneders feilfri bruk

Andre primærregistreringer viser samme struktur: Mech-Minds store prosjekter betaler 30% ved signering, 30% før levering, 30% etter aksept og 10% garantitilbakeholdelse over 12–15 måneder (Mech-Mind prospectus); GKG sier tilbakeholdelse i utstyrsbransjen vanligvis er 5–10% av kontrakten, betalt et år etter aksept (GKG inquiry reply) single sources, primary. Normen er da omtrent 30% ved signering (20–30%, null ved noen fabrikker som BYD), 60–65% etter aksept (betalt 30–60 dager senere, ofte i seksmåneders bank- eller kommersielle sedler, som forsinker brukbar kontant med ytterligere et halvt år), og 5–10% tilbakeholdelse i 3–12 måneder (12 vanligst) verified across several contracts.

FIG 27 **Hvordan kjøpere betaler *30% ved signering, 60% etter aksept, 10% holdt i ett år: full betaling tar 2–3 år.***



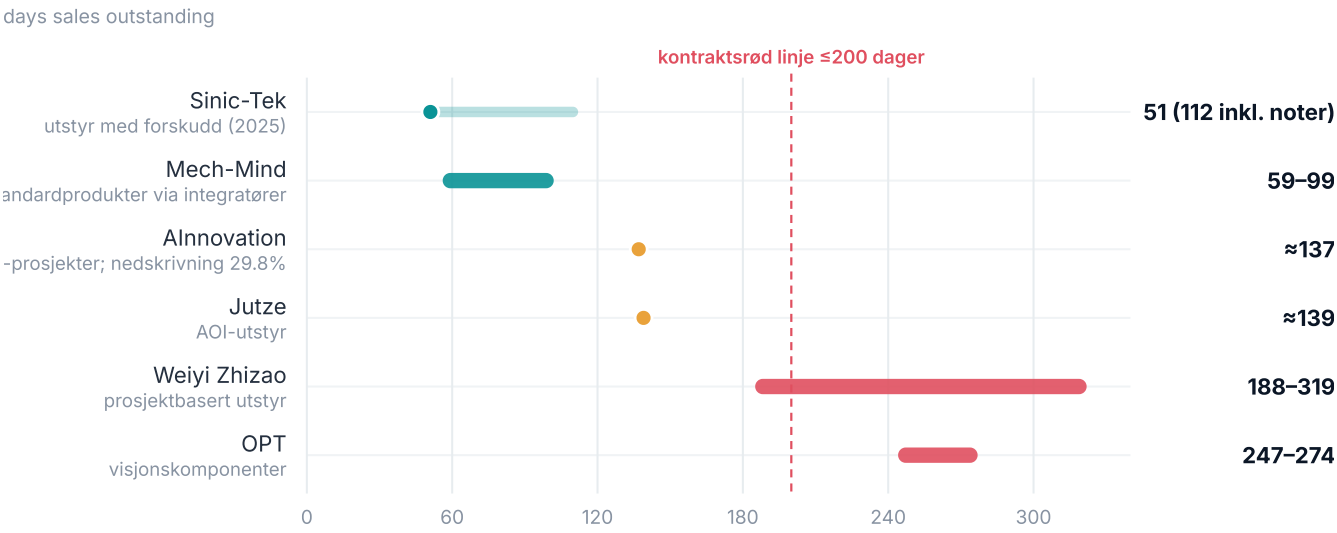
Kilde: Sinic-Teks svar på forespørsel, Mech-Mind, GKG, JT og andre dokumenter; skjema på sedvanlige vilkår, tidspunkt estimert.

Aksept og prøver er langsomme og betales av selgeren. Sinic-Teks sendte-men-ikke-aksepterte varer tok omtrent 143–158 dager å snu, fordi store EMS- og merkevarekunder trenger hel-linje-integrasjonstesting og en utvidet prøvekjøring (Sinic-Tek inquiry reply 8-1-2); Weiyis AI-prosjekter tar 4–8 måneder fra evaluering til utrulling (Weiyi Zhizao prospectus) single sources, primary. Det finnes ingen konverteringsdata for AI-granskningspiloter; Sinic-Tek konverterte bare omtrent 18–19% av demomaskinene sine direkte til salg. Jeg planlegger én av tre til én av fem for gratispiloter og omtrent 80% for betalte piloter etter shortlisting estimate.

Selskap	Forretningsmodell	Dager utestående salg
Sinic-Tek (2025)	Utstyr med forskudd	≈51 (≈112 inkludert sedler)
Mech-Mind	Standardprodukter gjennom integratører	59–99
Jutze	AOI-utstyr	≈139
AInnovation (2025)	AI-prosjekter	≈137 ved årsslutt, med 29.8% nedskrivning
Weiyi Zhizao	Prosjektbasert utstyr	188–319
OPT	Synskomponenter	247–274

Kilder: årsmeldingene og prospektene ovenfor; fordringer (inkludert kontraktsmidler) ÷ omsetning × 365 primary + estimate .

FIG 28 **Days sales outstanding** ***Standardprodukter 51–99 dager; prosjektbasert utstyr 188–319 dager.***



Kilde: Selskapsdokumenter; fordringer (inkl. kontraktsaktiva) ÷ omsetning × 365.

AI-granskningslisenser solgt gjennom AOI-produsenter eller integratører bør inndrives som Mech-Minds 60–100 dager; direkte prosjektsalg som Weiyis 190–320 dager. Satt sammen som en kontanttidslinje: gratis pilot én til tre måneder, kvalifisering ved en stor ODM opptil omtrent tolv måneder, bygging, levering og aksept tre til syv måneder; 30% kommer ved bestillingen, 60% omtrent fire til ni måneder etter den og de siste 10% omtrent 16–21 måneder etter den. **Fra første pilot hos en stor EMS til full betaling kan det ta to til tre år**; mellomstore EMS uten formelle anbud kan forkorte kvalifiseringen til uker estimate . Akseptkriteriene i de offentliggjorte kontraktene dekker bare hellinje-integrasjon, gjennomstrømning (ett taiwansk prosjekt krevde 100 enheter i timen) og måneder med feilfri kjøring; ingen spesifiserer escape- eller falskutrops-rater (Anda inquiry reply) single source, primary . Skriftlige akseptkriterier for AI-granskning må avtales før piloten.

Fire feilmønstre

Dokumenterte feil faller i fire mønstre: å sende før produktet var pålitelig (Rethink (The Robot Report)); et programvarelag på en leders maskinvare som lederens økosystem senere erstattet (READY Robotics (The Robot Report)); å gå tom for penger på bunnen av en syklus (Dextrous Robotics (Robots & Startups)); og modellselskaper som ender i acqui-hires (Covariant (GeekWire)). Lærdommene: produktifiser snevert i stedet for å tune prosjekt for prosjekt; hold arbeidskapital for akseptbasert omsetning og lange inndrivelser; tilby lokal service for fabrikker som kjører 24 timer; og ikke skaler før produktet er bevist.

Timing: inspeksjon vinner ingenting på å vente; arm-maskinvare blir billigere

Argumentene for å vente er svake. Maskinvare ble ikke billigere i 2026: robotprodusenter hevet prisene 5–15%, noen produkter 25–50% (Sina Finance) single source, og NVIDIA-kit steg en tredje gang (36Kr). Produksjonsgradige VLA-er kommer ikke innenfor vinduet (IFR, Instrumental). Imens fjerner hver AOI-erstatning et retrofitmål og hver OEM-AI-utrulling låser inn en linjes data og prosess; modellkapasitet kommodifiseres, og det som er knapt er tilgang til fabrikker og merkede fabrikkdata. Lokal støtte har også vinduer: Shenzhens 2026 «AI-pionérby»-regler, med opptil RMB 2 millioner for AI-applikasjonsdemonstrasjonsprosjekter og modellkuponger som dekker 30% av kjøp, varer til **30. juni 2027** (Yicai, May 2026) terms single source.

Det som fortsatt blir billigere raskt er humanoider og fingerferdige hender (hender falt fra omtrent RMB 50 000 til under 10 000 på to år (Sina Finance)), og entry-coboter ettersom produktmiksen beveger seg ned; det taler for å utsette embodied-delen, ikke inspeksjonsdelen. **Samlet: inspeksjon nå, som et snevert produkt; robotceller nå, gjennom partnere og integrasjon; embodied AI senere, med bare en R&D-stasjon i dag.**

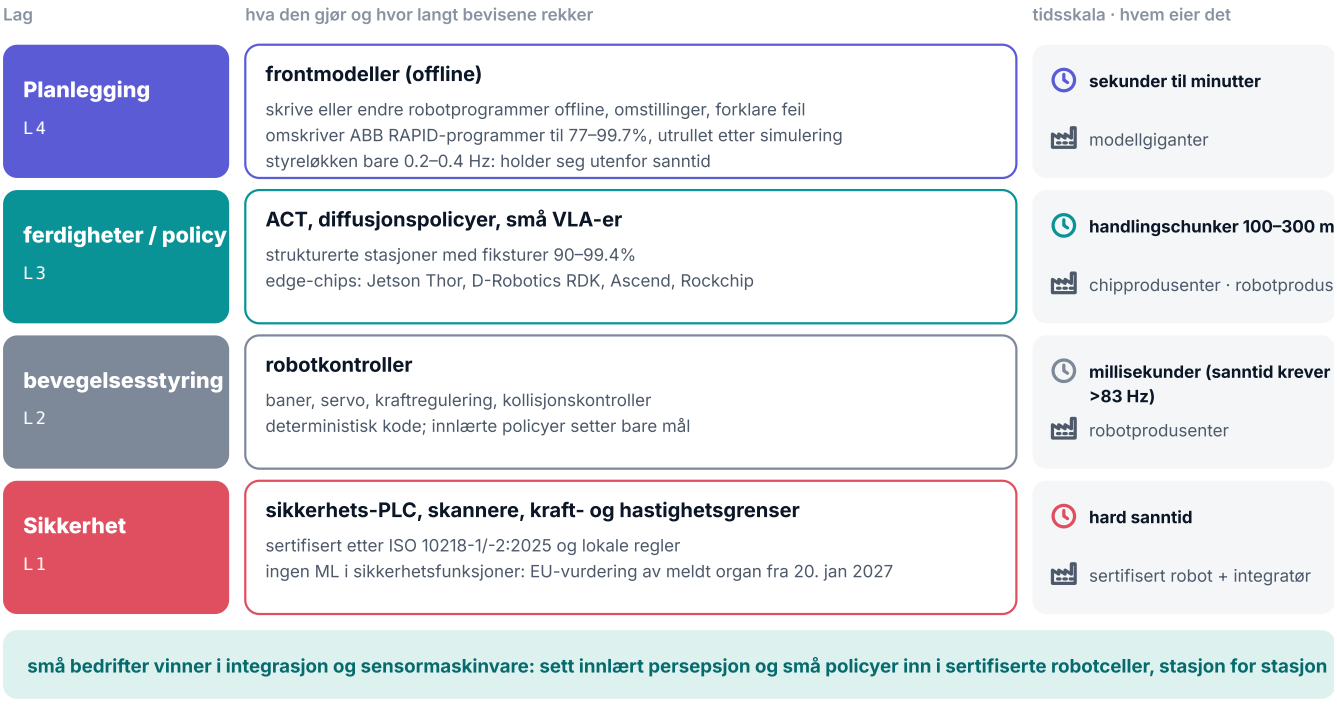
Embodied AI

06 Embodied AI: først endrer den hvordan roboter læres, senere hvordan de beveger seg

Det som endrer seg er persepsjon og programmering; bevegelse forblir deterministisk

Hver robotarm-leder markedsfører nå «fysisk AI», men det som er sendt faller i tre lag: innlært persepsjon uten opplæring (FANUCs AI Masterless Vision, som sendes i januar 2027 (FANUC); KUKAs nevrale grepvalg (KUKA)); store modeller som skriver programmer og planlegger oppgaver (KUKAs iiQWorks.Copilot, som advarer om at AI-generert kode bør verifiseres i en digital tvilling (KUKA); Yaskawa med Gemini Robotics-ER, der «generativ AI bestemmer hva som skal gjøres» og roboten bestemmer «hvordan den skal bevege seg» (Yaskawa)); og beregnings- og dataplattformer (URs AI Accelerator med en Jetson Orin (Universal Robots)) vendor claims. Ingen leder publiserer produksjonssuksessrater for AI-funksjonene sine, og de japanske lederne samarbeider med Google, SoftBank, Microsoft og andre i stedet for å bygge front-VLA-er (FANUC news index): det innlærte policylaget vil komme fra modellselskaper, de sertifiserte armene og kontrollerne fra armprodusentene.

FIG 29 **Embodied-AI-stacken** ***Store modeller planlegger offline, små policyer kjører i cellen, bevegelse og sikkerhet forblir deterministiske.***



Kilde: Forfatterens syntese (kapittel 3 og 6); ISO 10218:2025; EUs maskinforordning.

Hvor langt fra ustrukturert, fleksibelt småbatcharbeid: strukturerte stasjoner om 1–3 år, fleksibel montasje 5+

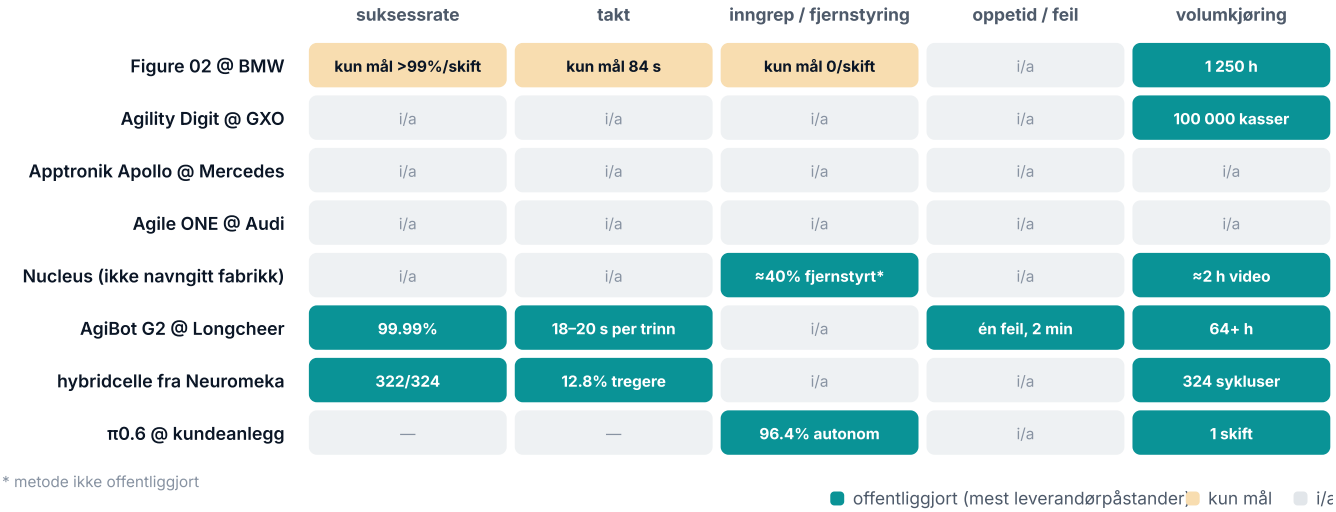
Ekte fabrikktall har begynt å dukke opp, men de er leverandørrapporterte, smale og fra strukturerte stasjoner. AgiBots åtte humanoider på Longcheers nettbrett slutt-testlinje loggførte 64 828 oppgaver ved 99.99% over mer enn 64 timer (AgiBot, Yicai Global) vendor and customer, men dette er mating-arbeid en konvensjonell arm med syn kan gjøre. UBTech sier Walker S2 arbeider på 30–50% av menneskelig effektivitet (Irish Times/FT) single source. Figure

02s 11 måneder ved BMW Spartanburg produserte volumtall (1 250+ timer, 90 000+ deler), mens syklus-, suksess- og intervensjonstallene er mål, ikke resultater (Figure, BMW Group) verified, numerators only. Nucleus sier omtrent 40% av fabrikkvideoen er teleoperert (Interesting Engineering) vendor claim, og en Reuters-undersøkelse siterte omtrent 20% av en arbeiders produksjon for noen kinesiske humanoider (Notebookcheck) single source. Ingen utrulling publiserer suksessrate, syklustid og intervensjonsrate sammen, og ingen er revidert, likevel er nettopp de tre det en tilbakebetalingsberegning trenger.

Utrulling	Suksessrate	Takt	Intervensjoner eller teleoperasjon	Oppetid eller feilintervall	Volumkjøring
Figure 02 @ BMW Spartanburg	Bare mål (>99% per skift)	Bare mål (84 s)	Bare mål (0 per skift)	Ikke offentliggjort (underarm toppfeilpunkt)	1 250 h
Agility Digit @ GXO	Ikke offentliggjort	Ikke offentliggjort	Ikke offentliggjort	Ikke offentliggjort	100 000 kasser
Apptronik Apollo @ Mercedes-Benz	Ikke offentliggjort	Ikke offentliggjort	Ikke offentliggjort	Ikke offentliggjort	Ikke offentliggjort
Agile ONE @ Audi	Ikke offentliggjort	Ikke offentliggjort	Ikke offentliggjort	Ikke offentliggjort	Ikke offentliggjort
Nucleus (unavngitt fabrikk)	Ikke offentliggjort	Ikke offentliggjort	≈40% teleoperert (metode ikke offentliggjort)	Ikke offentliggjort	≈2 h video
AgiBot G2 @ Longcheer	99.99% av 64 828 oppgaver	18–20 s per trinn	Ikke offentliggjort	Én kommunikasjonsfeil, ≈2 min å gjenopprette	64+ h
Neuromeka hybridcelle	322/324	12.8% tregere enn en person	Ikke offentliggjort	Ikke offentliggjort	324 sykluser
Physical Intelligence π0.6 ved et kundested	—	—	96.4% autonomt over et skift	Ikke offentliggjort	Ett skift

Kilder: dette kapittelet og kapittel 3 mostly vendor claims.

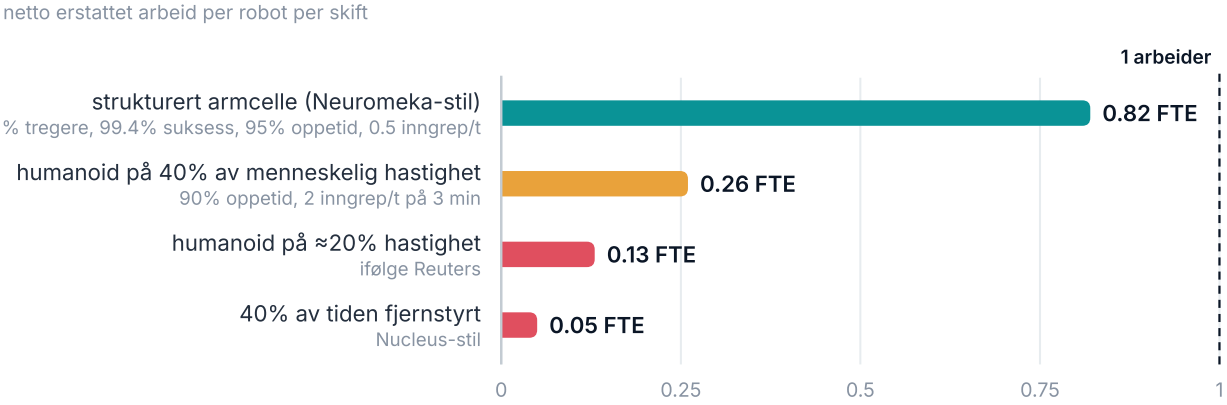
FIG 30 **Hva robotutrullinger i fabrikk faktisk offentliggjør** ***De fleste offentliggjør bare volum; ingen offentliggjør suksessrate, takt og intervensjonsfrekvens sammen.***



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

Teleoperasjon betyr mer for økonomien enn suksessrate. Netto erstattede personer per robot per skift ≈ (menneskelig syklus ÷ robotsyklus) × oppetid × suksess, minus intervensjonstid og teleoperasjon estimate, my formula. En Neuromeka-stil strukturert stasjon (12.8% tregere, 99.4% suksess, 95% oppetid, 0.5 intervensjoner på 2 minutter i timen) erstatter omtrent 0.82 personer; en humanoid ved 40% av menneskelig hastighet, 90% oppetid og to 3-minutters intervensjoner i timen omtrent 0.26; ved Reuters' 20%, omtrent 0.13; og hvis, som Nucleus, én operatør teleopererer den 40% av tiden, omtrent 0.05, nesten ingenting. Ved omtrent RMB 600k per humanoid og RMB 85k per person per år på fastlandet betaler de tre siste seg over tiår; bare den strukturerte armstasjonen har en realistisk tilbakebetaling (kapittel 4) estimate.

FIG 31 **Netto erstattet arbeid per robot** ***En strukturert armcelle erstatter omtrent 0.82 av en arbeider; en humanoid fjernstyrt 40% av tiden, omtrent 0.05.***



Kilde: Forfatterens formelbaserte estimater.

Borte fra strukturerte stasjoner skårer generelle modeller fortsatt 40–90% på nye oppgaver: Gemini Robotics 2 når 74.2% på lab pick-and-place, 78.9% på verktøykitting og 44% på å knytte en pose (eesel/DeepMind); på RoboChallenge Table30, 30 ekte-robot-oppgaver, skårer π0.5 43.7% og den beste åpne modellen 50.3% (RoboChallenge, QbitAI).

Eksterne prognoser skiller seg, men ingen støtter utbredt bruk innen to år. IFR sier generell fysisk AI er «ennå ikke industriell» og de fleste humanoider er «fortsatt prototyper eller piloter» (IFR); Interact Analysis anslår at bare omtrent 10% av humanoidene produsert i 2025 gikk til reell bruk, med det kommersielle vendepunktet «etter 2032» (automation.com); UBS ser et humanoid «EV-øyeblikk» som usannsynlig før 2030 (Yicai Global); kinesiske innsidere anslår to til tre år til før fabrikkskala-bruk (Gasgoo) single sources. Min vurdering estimate: i velutstyrte elektronikk- og bilfabrikker med leverandøren på stedet vil **strukturerte stasjonsoppgaver** (mating, OK/NG-sortering, fiksert innsetting) nå ≥99% ved nær menneskelig hastighet om omtrent 1–3 år; **blandet-del-kitting** med generelle VLA-er om omtrent 3–5; **fritt rutede ledningsnett og flertrinns fleksibel montasje** om mer enn 5. Humanoid tilbakebetaling virker ikke i dag: en Walker S2 til omtrent RMB 600k (10jqka) single source, ved 30–50% av menneskelig effektivitet over to skift, tilsvarer omtrent 0.6–1.0 arbeidere, en enkel tilbakebetaling på 7–12 år ved fastlands arbeidskostnader før integrasjon, vedlikehold og nedetid estimate.

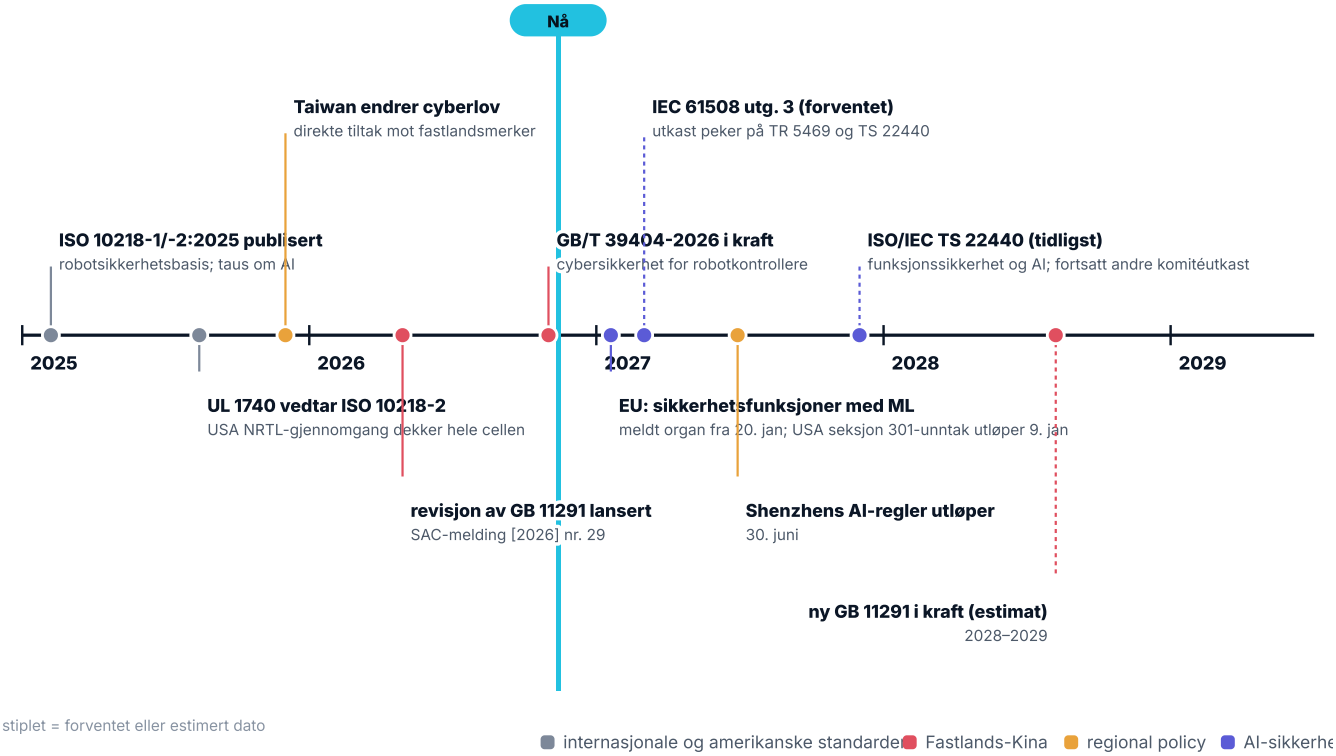
Hvor den sitter fast

Ledd	Bevis i 2026
Data	Bransjen anslår minst 10 millioner timer høykvalitets ekte data er nødvendig og bare omtrent 500 000 var samlet innen tidlig 2026 (Gasgoo); Reuters rapporterer omtrent 500 000 timer mot omtrent 100 millioner nødvendig (Notebookcheck); en nybegynnerteleoperatør trenger omtrent 300 forsøk per brukbar bevegelse; fabrikker vil ikke dele prosessdata; simuleringstrente policyer faller skarpt på ekte roboter
Pålitelighet	Strukturerte stasjoner 98–99.99% (leverandørpåstander, feil udefinert); generelle modeller 40–90% på nye oppgaver; usette kontakter 29–67%

Ledd	Bevis i 2026
Hastighet	Humanoider ved 30–50% av menneskelig hastighet (UBTech), noen ved omtrent 20% (Reuters); frontagenter omtrent 280 s per kvasistatisk ekte-robot-oppgave; LLM-kontrolløkker ved 0.2–0.4 Hz mot omtrent 83 Hz nødvendig (Anthropic)
Kostnad	Walker S2 snitter omtrent RMB 600k; 7–12 års tilbakebetaling estimate
Sikkerhetssertifisering	ISO 10218-1/-2:2025 gjør funksjonell sikkerhet mer eksplisitt og absorberer samarbeidskrav, men sier ingenting om AI (The Robot Report); ISO/IEC TS 22440 (funksjonell sikkerhet og AI) var fortsatt et andre komitéutkast i november 2026 (ISS, ISO Open Data); IEC 61508-3-veiledning fraråder AI over SIL 1 (Analog Devices); under EUs maskinforordning trenger sikkerhetsfunksjoner med maskinlæring vurdering av meldt organ fra 20. januar 2027 (EUR-Lex)
Hender og verktøy	Fingerferdige hender håndterer omtrent 30 N (≈3 kg) per finger mot omtrent 10 kg for en menneskehånd (Gasgoo); verktøybytter er fortsatt manuelle
Integrasjon	Sensorfeillesinger, protokollmismatch og MES-koblinger på stedet; intervensjonsrater offentliggjøres nesten aldri

Sikkerhetsimplikasjonen er direkte: standardene verken forbyr innlærte policyer eller anerkjenner dem. ISO 10218-1/-2:2025 sier ingenting om AI; ISO/IEC TS 22440 (funksjonell sikkerhet og AI) er fortsatt på komitéutkast og usannsynlig før sent 2027 eller 2028 estimate ; ISO/IEC TR 5469:2024 er bare veiledning, som anbefaler ikke-AI-fallbacks, overvåkende begrensninger og redundant avstemning; IEC 61508 utgave 3 forventes tidlig 2027 (ISO Open Data, BSI, 61508 Association) verified . Ingen publisert standard tilbyr ennå en akseptert vei til å sertifisere maskinlæringspersepsjon som en sikkerhetsfunksjon. Regionale regler beveger seg også: Kina reviderer GB 11291 (i kraft omtrent 2028–2029) (revision plan 20262543); UL 1740 har brakt ISO 10218-2 inn i US NRTL-gjennomgang siden august 2025; EU krever vurdering av meldt organ av sikkerhetsfunksjoner med maskinlæring fra 20. januar 2027; og ingen humanoid er sertifisert for inngjerdsløs fabrikkbruk mens ISO 25785-1 fortsatt er i utkast (i-SCOOP). Den forsvarbare arkitekturen: maskinlæring tar bare prosessbeslutninger («hvilken del», «er dette kortet godt»), mens et sertifisert deterministisk sikkerhetslag (typisk PL d / kategori 3) og en sikkerhets-PLC bestemmer hastighet, kraft og stopp. Behandle enhver påstand om «kamera-AI for inngjerdsløs sikkerhet» som ubevist inntil et PL d- eller SIL 2-sertifikat navngir maskinlæringskomponenten.

FIG 32 **Tidslinje for standarder og regulering** ***En sertifiseringsvei for AI innenfor sikkerhetsfunksjoner klargjøres ikke før 2027–2029.***



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

Hvilket lag er muligheten: modeller tilhører gigantene, edge-bokser til chipprodusentene, små firmaer til integrasjon og sensorisk maskinvare

Modellaget trenger kapital og data og tilhører gigantene (AgiBot trente GO-2 på «titusenvis av timer» (iFeng Tech); de sterkeste modellene er lukkede eller skybare). Edge-bokslaget tilhører chipprodusentene: Jetson Thor er de facto-standarden, og på fastlandet konkurrerer D-Robotics' RDK-kort (D-Robotics) og Rockchip-koprosessorer (CSDN), mest uten publiserte VLA-resultater.

Små firmaer har to åpninger. Én er sensorisk maskinvare som direkte kjøper pålitelighet: berøring løfter ContactDPs innsettingssuksess fra 58.7% til 97.8%, og kraft/moment-sensorer og håndleddskameraer gjør tilsvarende arbeid; datainnsamlingskit selger også (10jqka), selv om hender, sensorer og armer alle faller i pris. Den andre er integrasjonslaget mellom armprodusenter og modellselskaper, som eier hver stasjons oppgavedata, sikkerhetshylster og MES-kobling, noe armprodusenter ikke selger som et produkt i dag. Konkurrenter faller i fire grupper: armledere (FANUC, Yaskawa, KUKA, ABB, UR, JAKA, Dobot og andre), syn-pluss-arm-firmaer (Mech-Mind, Hikrobot, Smart-More), embodied AI-selskaper (AgiBot, UBTech, Unitree, Galbot, Figure, Physical Intelligence, Google DeepMind, NVIDIA) og bil-chip-nykommere.

Kjøperaksept: intet tall uten en nevner

Siden ingen leverandør publiserer intervensjonsrater, feilintervaller eller langsiktige suksesstall, hører disse hjemme i akseptvilkårene i kontrakten. Statistisk beviser 3 000 påfølgende produksjonssykluser med null feil en feilrate på høyst 0.1% ved 95% konfidens; 10 feil av 3 000 beviser bare ≤0.57%; å bevise ≤1%, ≤0.5%, ≤0.1% og ≤0.01% med null feil tar 299, 598, 2 995 og 29 956 sykluser `estimate, exact binomial`. Etter den standarden har Neuromekas 322/324 et tosidig 95%-intervall på 97.8–99.9% og beviser ikke ≥99%; AgiBots omtrent 6 feil i 64 828 oppgaver antyder 99.98–100%, men bare hvis «feil» er definert slik kjøperen definerer den `estimate`. Dette er akseptkravene jeg

ville stilt til enhver robotarm- eller embodied-celle; tersklene trekker på Figures egne mål ved BMW og på Neurome-ka- og AgiBot-tilfellene, og de er vurdering, ikke en standard estimate:

Punkt	Krev	Foreslått bestått-terskel
Definisjoner	Signerte definisjoner av syklus, suksess, feil (delvis suksess, omgrep, slipp, kvalitets-escape), intervensjon (berøring, fjern-teleoperasjon, tilbakestilling, e-stop) og hvordan autonomi beregnes	Definisjoner kommer før ethvert mål; ingen prosent uten en formel
Intervensjoner og teleoperasjon	Per robot-time, etter type og varighet; gjennomsnittstid mellom intervensjoner; teleoperasjonsminutter som andel av kjøretid; operatører per robot	Strukturerte stasjoner ≤0.5 intervensjoner per robot-time (≥2 h fra hverandre); null teleoperasjon under aksept
Langsiktig suksess	≥3 000 påfølgende produksjonssykluser over ≥3 skift, ≥2 delenumre og normal belysnings- og forsyningsvariasjon; feil klassifisert; eksakt nedre grense	Nedre grense ≥99.5% for linjestoppende inline-stasjoner, ≥99.0% offline; kvalitets-escapes rapportert separat
Syklustid	P50/P90/P99 per del inkludert gjenoppretting, mot ≥30 tidstakte menneskelige sykluser ved samme stasjon	≤1.15 × menneskelig P50, eller innenfor linjetakt
Oppetid og feil	Kjøretimer, nedetidslogg, feil per komponent, delbyttetid	≥95% oppetid over ≥2 uker; MTTR ≤20 min med reservedeler på stedet
Gjenoppretting	Skriptet feilinjeksjon: slipp, okklusjon, belysningsendring, samlebåndstopp, MES- eller nettverkstap, strømbrudd, e-stop	≥90% autonom gjenoppretting, tilbake i produksjon innen 2 min
Omstilling	Timer, demonstrasjoner, operatørferdighet og datainnsamlingskostnad per nytt delenummer eller fikstur	Strukturerte stasjoner: ≤1 skift og ≤100 demoer per nytt delenummer
Sikkerhet	Arm og celle til ISO 10218-1/-2:2025, celle-risikovurdering, lokale regler (GB 11291 på fastlandet; CNS 14490 pluss skriftlig vurdering i Taiwan; ANSI/A3 R15.06-2025 og UL 1740 i USA)	Ingen sikkerhetsfunksjon implementert av maskinlæring; et sertifisert sikkerhetslag (typisk PL d / kat 3) begrenser hastighet og kraft; ingen inngjerdsløse beinte roboter før ISO 25785-1 er publisert og vurdert
Beregning og offline-atferd	Inferensmaskinvare, P99-latens, atferd uten nettverk, modellversjonering, revalidering etter hver oppdatering	Taktkritisk inferens avhenger ikke av skyen; full re-aksept etter ethvert modell-, kvantiserings- eller motorbytte
Bevis	Uredigert video og rå logger for hele akseptvinduet; bevitnet av kundens kvalitetsteam eller en tredjepart	Ingen highlight-reels eller leverandørbare sammendrag
Kommersielle vilkår	Betalingsmilepæler knyttet til akseptmålinger; tilgjengelighetsstraff når fakturert per robot-time; dataeierskap og fjernaksessregler	Sluttbetaling og tilbakeholdelse frigis først etter 3 000-syklus-aksepten og ≥4 ukers stabil kjøring
Økonomi	Netto erstattede personer = hastighetsforhold × oppetid × suksess − intervensjonstid − teleoperasjonsandel	Avvis ethvert forslag hvis nettoerstatningen ikke kan støtte den påståtte tilbakebetalingen

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

Åpne spørsmål

07 Hva den offentlige dokumentasjonen ikke kan svare på, og hvordan en pilot avgjør det

Spørsmålene offentlige kilder ikke kan svare på faller i fem grupper: hva AI-granskning faktisk selges for og hva kjøpere vil betale; hvordan AOI-produsenter bunter AI og oppgraderer gamle maskiner; hvor gamle de installerte AOI-ene er og hvor mange som allerede kjører AI; hvor mange personer og timer granskning virkelig tar per linje; og hvordan piloter konverterer og hvordan kjøpere betaler. Oppå disse kommer hvordan de nyeste modellene, kvantiserte byggene og robotarmene faktisk presterer under en gitt fabrikks forhold. Hvert av dem kan gjøres om til en test med en klar terskel. Tersklene nedenfor er beslutningsregler jeg utledet fra bevisene i denne rapporten estimate; beregn dem på nytt når de første fabrikkdataene kommer inn.

Åpent spørsmål	Hvordan avgjøre det	Hvordan et godt resultat ser ut
Pris og betalingsvilje for AI-granskning	En standard forespørsel om tilbud (for eksempel fire linjer med blandede eldre 2D- og nyere 3D-AOI-er, to skift, omtrent 2 000 kort per skift, 12 defektklasser, kun lokalt, MES- eller IPC-CFX-integrasjon) sendt til flere tredjepartsleverandører; en firspørsmåls prisfølsomhetsundersøkelse av 20–30 EMS-fabrikker	De fleste tilbud innenfor RMB 40–120k engangs eller 20–40k i året for fire linjer; minst halvparten av fabrikkene aksepterer RMB 10–30k per linje med tilbakebetaling innen 12 måneder
AOI-produsentenes AI-bunting og oppgraderinger for eldre maskiner	Korte strukturerte intervjuer med omtrent et dusin AOI-produsenter og distributører	Hvorvidt AI-granskning er standard, valgfri eller fraværende på 2025–26-modeller; om pre-2020-maskiner kan oppgraderes og til hvilken pris
Alder på installerte AOI-er og andel som kjører AI	Stedsundersøkelse av omtrent 30 EMS-fabrikker, stratifisert etter størrelse og region, med utstyrslister gjennomgått under NDA	Median AOI-alder ≥5 år og ≥60% uten AI betyr retrofit-poolen er reell
Granskningsarbeid og timer	Intervjuer og 2–4 ukers AOI-alarm-, granskings-, reparasjons- og skiftlogger ved to eller flere fabrikker, med arbeidsstikkprøver	≥1 dedikert gransker per linje per skift og ≥3 granskningstimer per linje per skift, med minst 0.5 FTE per linje fabrikken er villig til å omplassere
Ytelsen til selve AI-granskningen	Gyllent sett pluss skyggemodus på en ekte linje (kapittel 3): dataavtale i uke 0, grunnlinje og merking i uke 1–2, skyggekjøring i uke 2–4, valgfri assistert kjøring i uke 4–8	Falskutrop eller granskningstimer kuttet ≥50–60%; per-klasse escape-grenser ikke verre enn menneskelig grunnlinje; null misser på kritiske klasser; p95-latens ≤½ takt; automatisk fail-open til mennesker
Læreretikettkvalitet for frontmodeller	Kjør to eller tre kandidater på samme gylne sett, tre promptformuleringer hver, ett godt referansebilde, enkeltkomponent-beskjæringer ≥42–48 px	Automatisk aksept bare for klasser med null misser på ≥299 defekter og menneskelig-nivå enighet; ellers bare forhåndsmerking og triage
Kvantiserte bygg	Endre bare presisjonen; pin vekthasher, mal, forbehandling, dekoding, motor og batchstørrelse; deaktiver eller tilbakestill prefiks-caching; sjekk at referansebygget reproduserer leverandørresultater; skår par på det gylne settet og drift-slice	Ingen nye misser på kritiske klasser; ≤1 ny miss i 1 000 defekter; rydding ned ≤2 poeng; ≥99.5% identiske beslutninger over tre kjøringer; ellers fall tilbake til FP8/INT8

Åpent spørsmål	Hvordan avgjøre det	Hvordan et godt resultat ser ut
Line-side maskinvare, renhet og ESD	≥2 enheter av hver kandidatboks stresstestet ved dobbelt forventet alarmrate; 72-timers soak ved reell kabinettluftstrøm; ESD-sjekk av kundens koordinator; ISO 14644-14-test for ISO 7 og renere kunder	p95 ≤½ takt; ingen krasj eller throttling over 72 timer; ESD-signering; partikkelrapport i riktig klasse
Betalingsvilkår, pilotregler og konvertering	Betalingsklausuler fra nylige inspeksjonsutstyrskontrakter ved de undersøkte fabrikkene; intervjuer om pilotregler, akseptkriterier og kvalifiseringstid	Omtrent 30/60/10 med kontanter innen 60 dager etter aksept; piloter ≤8–12 uker med en navngitt signeringseier; ≥1 av 3 prøver konverterer; en ramme som dekker ≥5–10 linjer i år én
Data, IT og ansvar	Intervjuer og kontraktsgjennomgang med juridisk rådgiver	En lokal boks kan koble til og få tilgang til data; ansvar begrenset til programvarefeil
Regional samsvar før utenlandsetablering	Taiwan: kundens skriftlige status under Cyber Security Management Act, E187-selvevaluering; USA: tollavgjørelse om hvordan en celle klassifiseres, NRTL-tilbud, ITAR-sjekk; Sørøst-Asia: BOI lokal-tilknytning og malaysisk fradragsberettigelse, kundens morselskap; fastlandet: CCC-omfang for bundne PC-er og strømforsyninger	Kunden er ikke et kontrollert organ, og tilbakebetalingen holder seg ≤2 år med toll og sertifisering inkludert
Aksept av en robot- eller embodied-celle	12-punkts aksepttabellen i kapittel 6: signerte definisjoner, ≥3 000 påfølgende produksjonssykluser over ≥3 skift og ≥2 delenumre, skriptet feilinjeksjon, ≥30 tidstakte menneskelige sykluser, bevitnet	Nedre grense ≥99.5% inline (≥99.0% offline); syklus ≤1.15 × menneskelig; ≤0.5 intervensjoner per time og null teleoperasjon; ≥95% oppetid; MTTR ≤20 min; ≥90% gjenoppretting
ACT- og VLA-policyer på kinesiske edge-chips	Fire uker: port og mikro-benchmark (≥1 000 inferenser, latensfordeling, effekt, 60-minutters soak); lukket-sløyfe simulering med målt latens injisert; ekte-robot A/B mot en GPU; 8–24 timers utholdenhet (≥3 000 sykluser)	ACT-chunks p95 ≤100 ms, VLA ≤300 ms; ekte-robot-suksess innenfor 3 poeng av GPU-en og syklus ≤1.2 × menneskelig; ellers behold policyen på Jetson eller en GPU
Cobot-, AMR- og integrasjonspriser	Modell-for-modell-tilbud fra flere cobot-distributører, og fullcelle-tilbud fra to eller tre integratører for testfiksturlasting (én stasjon, to skift)	Full celle ≤RMB 200k og tilbakebetaling ≤1.5 år over to skift i Shenzhen
SMT-linjetellinger i Taiwan, Sørøst-Asia og USA	IPC (tidligere New Venture Research) data eller lokale utstyrsdistributørers installerte base	Nok servicebare linjer per region til å rettferdiggjøre lokal applikasjonsteknikk

Av alle disse avgjør fire de mest følsomme tallene i denne rapporten: prisen kjøpere vil betale, arbeidet AI-granskning faktisk frigjør, om AI-granskning virkelig misser ikke mer enn mennesker, og hvor raskt kontanter kommer tilbake. De er verdt å avgjøre før enhver betydelig forpliktelse, av en fabrikk som vurderer et kjøp eller av en leverandør som vurderer et produkt.

En pilotsekvens som virker for begge sider

For en fabrikk, og for en leverandør som vil ha en troverdig referanse, er rekkefølgen den samme:

1. **Uke 0–2: enes om data og definisjoner.** Signer dataavtalen, definer defektklassene og hva som teller som en escape, og registrer en grunnlinje: falskutrop per kort, granskningssekunder per utrop, escapes funnet nedstrøms.
2. **Uke 1–4: bygg det gylne settet** fra AOI-arkivet og beholdte NG-kort, blindmerket og avgjort (kapittel 3).

3. **Uke 2–4 (eller lengre): skyggemodus.** Modellen logger beslutningene sine mens folk bestemmer som før; sammenlign per klasse.
4. **Uke 4–8: assistert kjøring på tilfeldige skift,** med modellen som rydder bare klasser som passerte porten og alt annet går til folk.
5. **Aksept, skriftlig, før skalering:** per-klasse-resultatene, latens, fail-open-atferd, og hvem som eier modeller og etiketter. Skaler linje for linje, og revalider ved hvert nytt produkt, modellversjon, kvantisering eller motorbytte.

Begrensninger ved denne rapporten

Dette er skrivebordsforskning. Jeg gjennomførte ingen primærintervjuer med fabrikker for denne utgaven; mye linjedata kommer fra leverandører og er ikke uavhengig revidert. Flere myndighetsnettsteder, anbudsdatabaser og rekrutteringsplattformer kunne ikke nås, så Shenzhens 2026 oppgraderingssubsidiesatser, lønnsveiledning for Yangtsefloddeltaet, transaksjonspriser for AI-granskning, granskere per linje og pilotkonverteringsrater er estimater eller analogier; hvert av dem fremstår ovenfor som et åpent spørsmål. Valutakursene er People's Bank of Chinas sentralkurs 9. november 2026 ($1 = RMB 6.7188) og, for Taiwan, SMBCs fiksering samme dag ($1 = NT$32.12); vietnamesisk dong bruker data fra mai 2026. Alle terskler er beslutningsregler utledet fra de siterte bevisene, ikke standarder.

Om forfatteren

Tan Shuai

Tan Shuai er teknolog og gründer med mer enn tjue års arbeid på tvers av chips, maskinvare og systemprogramvare, og 20 innvilgede patenter. Arbeidet hans har ført ham gjennom Silicon Valley, New York, Chicago, Taiwan, Greater Bay Area (Shenzhen og Hongkong), Vietnam og Malaysia, med et langvarig fokus på avansert produksjon og forsyningskjeder. I dag arbeider han som AI-forsker og -utvikler, med fokus på hvordan AI faktisk brukes på fabrikkgulvet.

Han tar gjerne imot samtaler med folk som arbeider med AI i produksjon, enten de bygger et team, søker en partner eller sammenligner notater: i@tanshuai.com · tanshuai.com

Merknader og ansvarsfraskrivelse

Denne rapporten er uavhengig skrivebordsforskning basert på offentlige dokumenter tilgjengelig frem til 11. november 2026. Den gis kun til informasjon og er ikke investerings-, juridisk, skatte- eller ingeniørrådgivning. Produktnavn, priser, modellutgivelser, eksportklassifiseringer, toll og forskrifter endres raskt; verifiser alt du støtter deg på mot primærkilden. Tall merket vendor claim eller estimate bør leses som sådan. Selskaps- og produktnavn tilhører eierne sine og brukes kun til identifikasjon; intet selskap har gjennomgått eller sponset denne rapporten.

© 2026 Tan Shuai. Du kan dele denne rapporten uendret med kildehenvisning.

ELEKTRONIKKPRODUKSJON · AI-INSPEKSJON × ROBOTIKK

AI-visuell inspeksjon og robotarmer i elektronikkfabrikker: Vurdering 2026

Hvert nøkkeltall har sin kilde og et konfidensmerke: verified, single source, vendor claim eller estimate. Det den offentlige dokumentasjonen ikke kan svare på, er samlet i kapittel 7, sammen med hvordan en pilot avgjør hvert spørsmål.

Tan Shuai
tanshuai.comi@tanshuai.com

OFFENTLIG UTGAVE · DATA PER 11. NOVEMBER 2026

