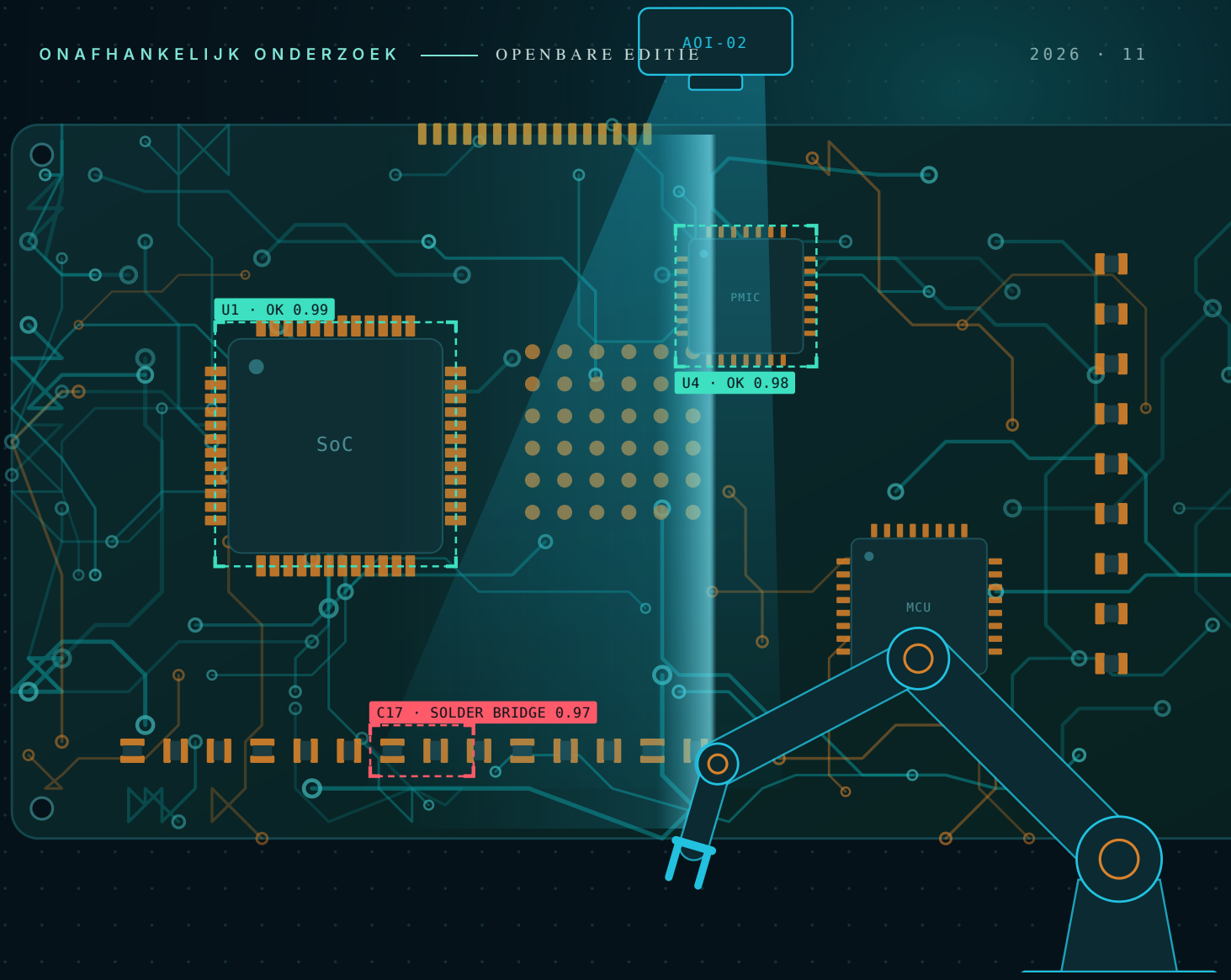


ELEKTRONICAPRODUCTIE · AI-INSPECTIE × ROBOTICA

AI-visuele inspectie en robotarmen in elektronicafabrieken

Beoordeling 2026

Gangbare configuraties · Regionale haalbaarheid · On-premises implementatie · Leraar-leerlingmodellen
Rendement op investering · Concurrentie · Wat alleen een pilot kan beantwoorden

KERNBEVINDING

AI-review loont, in beperkte zin. Belichaamde robotarmen kunnen wachten.

Tan Shuai
tanshuai.comi@tanshuai.com

Gegevens per 11 november 2026
Vasteland-China · Taiwan · Zuidoost-Azië · Verenigde Staten

INHOUD

Inhoud



◆ **MANAGEMENTSAMENVATTING** 5
De moeite waard, maar smal, nu en met partners

01 **De lijn vandaag** 9
Wat elektronicafabrieken vandaag draaien, en wat technisch haalbaar is

02 **On-premises** 17
AI binnen de fabriek draaien: cleanrooms, luchtspleten, hardware en kwantisering

03 **Leraar en leerling** 26
Leraar en leerling: frontier-modellen kunnen inspectiedata pre-labelen; voor robotarmen zijn de echte leraren nog altijd mensen

04 **Kosten en terugverdientijd** 35
Kosten en terugverdientijd: AI-review verdient zich in maanden terug, machinebediening in jaren, uithaalarmen nooit

05 **Markt en concurrentie** 43
Markt, concurrentie en hoe kopers betalen

06 **Belichaamde AI** 52
Belichaamde AI: eerst verandert het hoe robots worden aangeleerd, later hoe ze bewegen

07 **Open vragen** 58
Wat de openbare documentatie niet kan beantwoorden, en hoe een pilot het beslecht

◆ **Over de auteur** 61
Tan Shuai

Betrouwbaarheidstags in de tekst

- **verified** twee of meer onafhankelijke bronnen, of twee onafhankelijke primaire documenten, komen overeen
- **single source** slechts één bron, inclusief primaire documenten
- **vendor claim** gepubliceerd door de verkoper van het product of de dienst
- **estimate** de eigen berekening of afleiding van de auteur uit geciteerde gegevens

Hoe dit rapport te lezen

Kerncijfers dragen een betrouwbaarheidstag. **verified** betekent dat twee of meer onafhankelijke bronnen overeenkomen, of twee onafhankelijke primaire documenten overeenkomen; **single source** betekent slechts één bron (inclusief primaire documenten); **vendor claim** betekent gepubliceerd door de verkoper van het product of de dienst, zelfs wanneer wijd verbreid; **estimate** betekent mijn eigen berekening of afleiding uit geciteerde gegevens, met gedocumenteerde invoer. “Primair” beschrijft het type bron (jaarverslag, prospectus, verordening, norm, officiële statistiek); het verandert de betrouwbaarheidsklasse niet. Waar bronnen elkaar tegenspreken, worden de cijfers naast elkaar gegeven en niet gemiddeld. Valutaomrekeningen gebruiken de centrale koers van de People’s Bank of China van 9 november 2026, $1 = RMB 6.7188 (Xinhua), en voor Taiwan de fixing van SMBC op dezelfde dag, $1 = NT$32.12 (SMBC). Loon- en subsidiegegevens voor het vasteland gebruiken de Parelrivierdelta, waar de elektronicaproductie het sterkst geconcentreerd is (officiële loongegevens uit Shenzhen), met de Jangtsekiangdelta ter vergelijking. Beweringen uit branche-interviews worden behandeld als te controleren beweringen, nooit als feiten. Vragen die de openbare documentatie niet kan beantwoorden, zijn verzameld in hoofdstuk 7.

FIGUREN

Figuren



01	Waar AI vandaag past op de fabrieksvloer	7
02	Een volume-SMT-lijn: de mensen zitten achter de AOI	10
03	Ongeveer 97–99% van de AOI-alarmen zijn false calls	11
04	Een universele camera is geen AOI	13
05	Een air-gapped uitrol	17
06	Clean- en ESD-zones bepalen waar rekenkracht mag staan	18
07	Snelheidsgrenzen per toepassing	20
08	On-prem rekenkracht: geheugen, prijs en wat legaal te koop is op het Chinese vasteland	21
09	Eén groot workstation maakt geen gedeelde assistent	22
10	Quantisatie: 8-bit is veilig, 4-bit hangt van het recept af, 3-bit en lager valt af	23
11	Leraarkandidaten: generieke scores verdubbelden in een generatie	27
12	"Is er een defect?" blijft moeilijk voor algemene modellen	28
13	Twee leraar–leerling-pipelines	29
14	Hoe groot de gouden set moet zijn	30
15	Edge-policylatentie	32
16	Robotleerlingen: bijna productiekwaliteit alleen als het station gestructureerd is	33
17	Wat één inspectie- of handlingmedewerker per jaar kost	36
18	De prijsladder	37
19	Terugverdientijd per regio	39
20	Waarom slechte boards uitnemen zich niet terugbetaalt	40
21	VS-heffing op robots, AOI en edge-computers	41
22	Marktomvang verschilt meer dan tienvoudig	44
23	Bottom-up: 40.000 lijnen → 16.000 serviceerbaar → ≈RMB 320M eenmalig	45
24	Het retrofitvenster is ongeveer vijf jaar	46
25	De herbeoordelingswig is aan drie zijden omsingeld	46
26	De marge zit in geproductiseerde hardware	48
27	Hoe kopers betalen	49
28	Days sales outstanding	50
29	De embodied-AI-stack	52
30	Wat in-plant robotuitrollen echt openbaar maken	54
31	Netto vervangen arbeid per robot	54
32	Tijdlijn van normen en regelgeving	56

De moeite waard, maar smal, nu en met partners

KERNBEVINDING

Het oordeel: voor elektronicafabrieken is wat in 2026 volwassen is en waarvoor fabrieken zullen betalen AI-review en niche-inspectie op AOI-beelden, plus geprogrammeerde robot-machinebediening, niet “een groot model dat naar platen kijkt plus belichaamde robots die ze pakken”.

0.46–0.65 jaar

Terugverdientijd van AI-review in de VS, Taiwan en Shenzhen, bij RMB 50k per lijn

≥95% · <20 PPM

Dekking en doorglippers van de interne AI-review van een top-EMS: de lat voor verkoop aan tier-1-EMS

100–700 mln RMB

Adresseerbare derden-retrofitmarkt in vasteland-China, eenmalig, over een venster van ongeveer vijf jaar

0 scores

Gepubliceerde benchmark-scores voor industriedefecten van de vlaggenschepen van november 2026: leraren, geen inspecteurs

1.1–1.5 jaar

Terugverdientijd van een cobot-machinebedieningscel over twee ploegen in de VS, Taiwan en Shenzhen

0.003–0.14 FTE

Arbeid die een uithaal-robotarm per ploeg bespaart op een SMT-lijn: verdient zich nooit terug

Inspectie
ROI

Gesuperviseerd deep learning toegepast op de beelden die een bestaande AOI al produceert, verlaagt valse alarmen of review-uren met 60–90% (vendorclaims). DBG, een groot EMS-bedrijf, meldt in zijn jaarverslag dat zijn interne systeem meer dan 95% van de gemarkeerde componenten dekt, met doorglippers onder 20 PPM en meer dan 85% minder menselijke beoordeling; dat is de lat die een product van derden moet halen om aan tier-1-EMS te verkopen. Bij RMB 50,000 per lijn verdient AI-review zich binnen een jaar terug in vasteland-China, Taiwan en de VS (Shenzhen 0.65 jaar, Taiwan 0.53, VS 0.46), cobot-machinebedieningscellen in 1.1–1.5 jaar en Zuidoost-Azië in 2–6 jaar, terwijl een robotarm die aan een SMT-lijn wordt toegevoegd alleen om slechte platen eruit te halen zich nergens terugverdient, omdat boven 90% first-pass-opbrengst het uithalen slechts 0.003–0.14 van een persoon per ploeg kost estimate .

Modellen
ROBOTS

Multimodale grote modellen kunnen nog altijd niet de primaire inspecteur zijn. Per 11 november 2026 heeft geen van de vlaggenschepen (Claude Fable 5.2, GPT-6.x, Gemini 4, Qwen3.8 en andere) een gepubliceerde score op een benchmark voor industriedefecten of PCB, en in de ACM MM 2026 PCBA-uitdaging scoorden de vijf beste systemen, getraind op 6.600 echte lijnbeelden, slechts 77–82 op “is er een defect”. Frontier-modellen kunnen pre-labelen en triëren. Voor robotarmen zijn de echte leraren nog altijd menselijke demonstraties, reinforcement learning op echte robots en simulatie: leerlingen halen 90–99.4% op gefixeerde, gestructureerde stations, maar slechts 30–80% op bewegende transportbanden, ongeziene onderdelen en nieuwe producten; een frontier-agent voltooide 104 van 108 quasistatische taken op echte robots, elk in 280 seconden.

On-premises
IMPLEMENTATIE

Luchtgescheiden implementatie in schone, gecontroleerde fabrieken is volledig praktisch. Een gewone SMT-hal is meestal een ESD-beschermde zone en geen geclassificeerde cleanroom: een ventilatorloze Jetson- of binnenlandse NPU-box staat aan de lijn en GPU-servers staan in de serverruimte. Volgens de eigen exportclassificaties van NVIDIA kunnen Jetson Thor en DGX Spark zonder Amerikaanse licentie aan een gewone vasteland-fabriek worden verkocht; DGX Station niet. De standaardprecisie is 8-bit; quantization-aware 4-bit van de leverancier is onder voorwaarden bruikbaar; 3-bit en lager is uitgesloten.

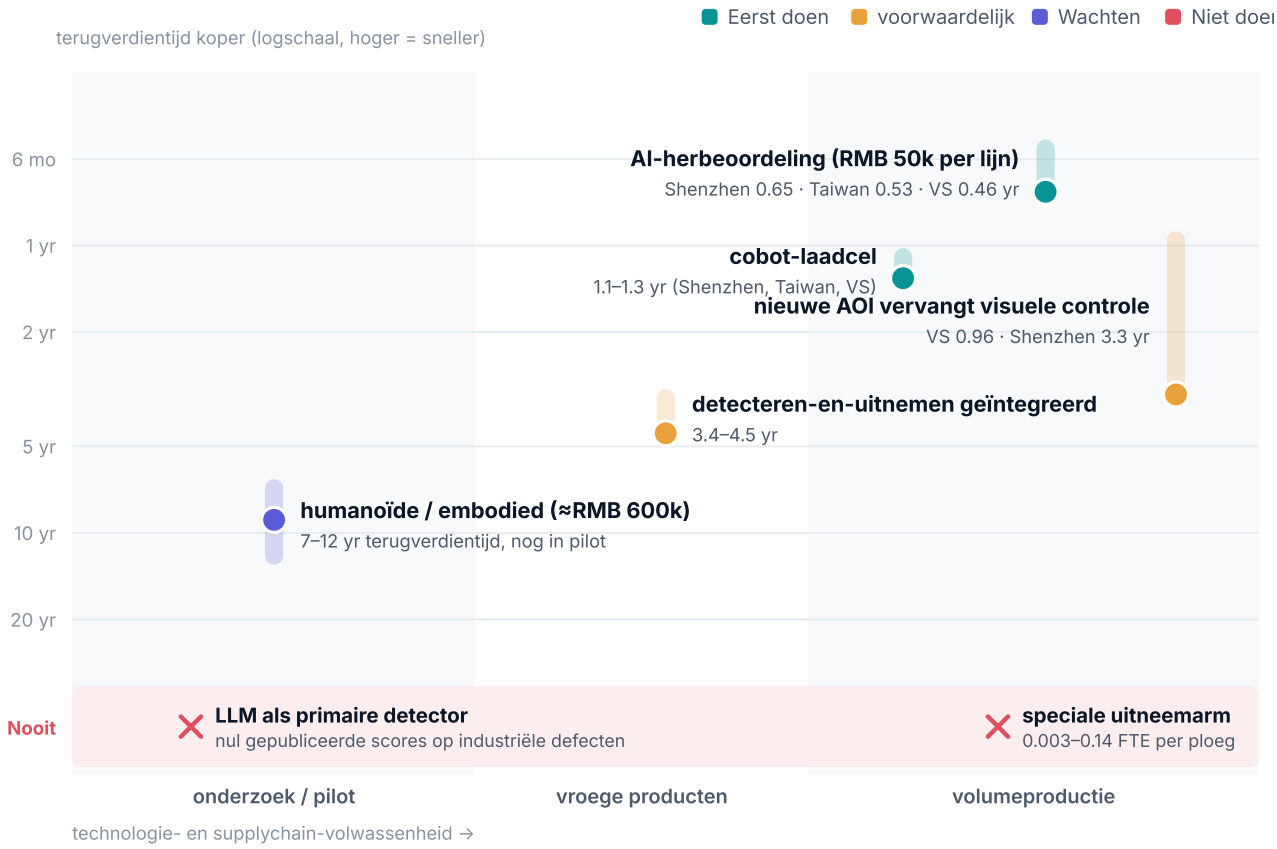
Markt
CONCURRENTIE

De markt is reëel maar beperkt. Vasteland-China heeft ongeveer 35,000–46,000 SMT-lijnen en ongeveer 26,000–66,000 inline-AOI's die niet met AI draaien; de adresseerbare markt voor retrofits door derden bedraagt ongeveer RMB 100–700 miljoen eenmalig, over een venster van ongeveer vijf jaar, en AOI-makers zijn begonnen de review in nieuwe machines in te bouwen **estimate**. De concurrentie is al sterk: de reviewwig wordt omsingeld door AOI-makers, interne EMS-teams en platformleveranciers, en de marge zit bij geproductiseerde hardware.

Vraag	Oordeel	Kerncijfers (bronnen en betrouwbaarheid in de hoofdstukken)
① **Wat draaien gangbare elektronicafabrieken vandaag?**	Een SMT-volumelijn is "3D SPI + AOI's voor en na reflow + steekproefsgewijze röntgeninspectie + ICT/FCT + reparatie en visuele inspectie", met menselijke beoordeling achter elke AOI. AI komt op drie manieren de review binnen: ingebouwd door AOI-makers, intern gebouwd door top-EMS, of toegevoegd door derden. Assemblage en handling gebruiken armen, SCARA's en AMR's met geleerde perceptie en geprogrammeerde beweging; belichaamde AI draait nog in pilot	Takt ≈13–22 s per plaat; één ODM begrootte RMB 30.6M aan AOI voor 18 lijnen, ≈3–5 AOI's per lijn; ≈97–99% van de AOI-alarmen zijn vals
② **Hoe ver reikt de technologie, kan geld nul mensen kopen, en wat kan elke regio gebruiken?**	Inspectie is industrieklasse als "toegewijde detector + AI-review + een paar mensen"; zelfs onbeperkt geld bereikt geen nul mensen. Het vasteland gebruikt alleen binnenlandse modellen of lokale open gewichten; private fabrieken in Taiwan kunnen vrijwel alles gebruiken, maar fabs en kritieke infrastructuur sluiten vasteland-merken in de praktijk uit; Zuidoost-Azië heeft de minste beperkingen; de VS hebben de volledige stack plus invoerrechten en certificering	AI-review verlaagt valse alarmen of review-uren met 60–90% **vendor claims**; een 4K-camera over een heel paneel ≈78 µm per pixel tegen 8–20 µm voor een AOI; vasteland-armen betalen ≈40% gecombineerd Amerikaans invoerrecht, instrumenten van het AOI-type ≈37.5%
③ **Kan het on-premises draaien in cleanrooms en luchtgescheiden fabrieken, op welke hardware, snel en nauwkeurig genoeg?**	Ja; de beperkingen zijn procedureel, niet rekenkundig. Een ventilatorloze box ≤100 W aan de lijn, GPU-servers in de serverruimte; Thor voor stations, een Mac Studio of een binnenlands apparaat voor fabrieks-Q&A geen DGX Station op het vasteland. Snelheid is prima wanneer de modelgrootte bij de taak past; houd 8-bit aan, gebruik quantization-aware 4-bit van de leverancier onder voorwaarden, nooit 3-bit	≈0.5 s per uitsnede voor een 8B-vision-model op Thor **estimate**; Thor-devkit RMB 59,104 in China; DGX Station $123–140k, 139 tokens/s in totaal voor 32 gebruikers op een 743B-model
④ **Kan het nieuwste vlaggenschip over tientallen rondes als leraar fungeren en in lokale leerlingen worden gedestilleerd, ook voor flexibele productie en robotarmen?**	Inspectie: pre-labelen en triage; accepteer leraarlabels alleen automatisch voor klassen met "≥299 defecten, nul missers" op een golden set. Robots: de leraren zijn nog altijd menselijke demonstraties, RL op echte robots en simulatie, en leerlingen naderen productiekwaliteit alleen op gestructureerde stations	Geen november-vlaggenschip heeft een industriedefect-score; leerlingen 97.8% op getrainde connectoren, 29–67% op ongeziene, 37–59% op bewegende transportbanden; frontier-modellen die gewrichten direct aansturen 0–5.5%

Vraag	Oordeel	Kerncijfers (bronnen en betrouwbaarheid in de hoofdstukken)
⑤ **Wat kost het, en wat is de ROI per regio?**	AI-review verdient zich binnen een jaar terug in vasteland-China, Taiwan en de VS; cobotbediening 1.1–1.5 jaar; Zuidoost-Azië 2–6 jaar; een uithaalarm op een SMT-lijn verdient zich nooit terug	RMB 50k reviewbox per lijn: Shenzhen 0.65 jaar, Taiwan 0.53, VS 0.46, Vietnam 2.2 estimate ; acceptabele derden-add-onprijs ≈RMB 10–30k per lijn eenmalig of RMB 5–10k per jaar
⑥ **Zijn concurrenten al sterk genoeg, en is het beter om nu te starten of te wachten?**	De vraag is niet verzadigd, maar het aanbod is overvol; OEM-AI beperkt het retrofitvenster tot ongeveer vijf jaar. Stap nu in met een smal product via kanaalpartners, zonder platformleveranciers frontaal te bestrijden; wacht op productiedata van 2027–2028 alvorens op belichaamde AI te wedden	≈26–66k inline-AOI's zonder AI op het vasteland, jaarlijks 8–12% krimpend estimate ; de leidende AI-visieleverancier had in 2025 een omzet van RMB 1.086B en een nettoverlies van RMB 991M; de 3C-omzet van het dichtstbijzijnde beursgenoteerde vergelijkingsbedrijf daalde in januari–mei 2026 met 79.9% op jaarbasis

FIG 01 **Waar AI vandaag past op de fabrieksvloer** ***AI-herbeoordeling van AOI-false calls en cobot-belading zijn volwassen en betalen zich snel terug; een speciale uitneemarm en een LLM als primaire inspecteur zijn de moeite niet waard.***



Bron: Terugverdientijd uit het model van de auteur (hoofdstuk 4); de volwassenheidsas is het oordeel van de auteur.

Wat te bouwen en wat niet te bouwen

	Aanbeveling
Bouwen	AI-review en defectclassificatie op AOI-beelden, voor middelgrote EMS met oudere of gemengde AOI-merken en geen AI-upgrade van de AOI-maker; inspectie die de SMT-AOI slecht dekt (through-hole en handmatig solderen, conformal coating, kabels, connectoren en FPC's, einduiterlijk en schroeven); cobot-laad- en -loscellen waar elk onderdeel wordt gehanteerd (testopstellingen, trays, verpakking), met inspectie als add-on; een golden set en schaduwmodus op een echte lijn als basis van elke claim

	Aanbeveling
Niet bouwen	Een algemeen “AI-plus-robotarm”-platform; een multimodaal LLM als primaire inspecteur; een robotarm op een SMT-lijn alleen om NG-platen eruit te halen (een NG-afleider of -buffer doet dat); humanoïden en robots-as-a-service; frontier-modellen binnen een regelkring of een veiligheidsfunctie; bergen GPU-geheugen op stations om “volledige” modellen te draaien; directe verkoop aan fabs, Taiwanese kritieke infrastructuur of Amerikaanse defensie-elektronicafabrieken
Met wie samenwerken	AOI-makers en distributeurs (kanaal- en beeldinterfaces; distributeurs brachten 46.68% van de omzet van Sinic-Tek in 2025 binnen); cobotmakers (gecertificeerde armen en veiligheidsfuncties); EMS-bedrijven met 10–40 lijnen (uitrollende expansie en aanbetalingen); lokale integratoren in Thailand en elders (om aan de BOI-local-linkage-regels te voldoen)

De lijn vandaag

01 Wat elektronicafabrieken vandaag draaien, en wat technisch haalbaar is

De standaard volumelijn: meerdere AOI's, steekproefsgewijze röntgeninspectie, centrale review, 13–22 seconden per plaat

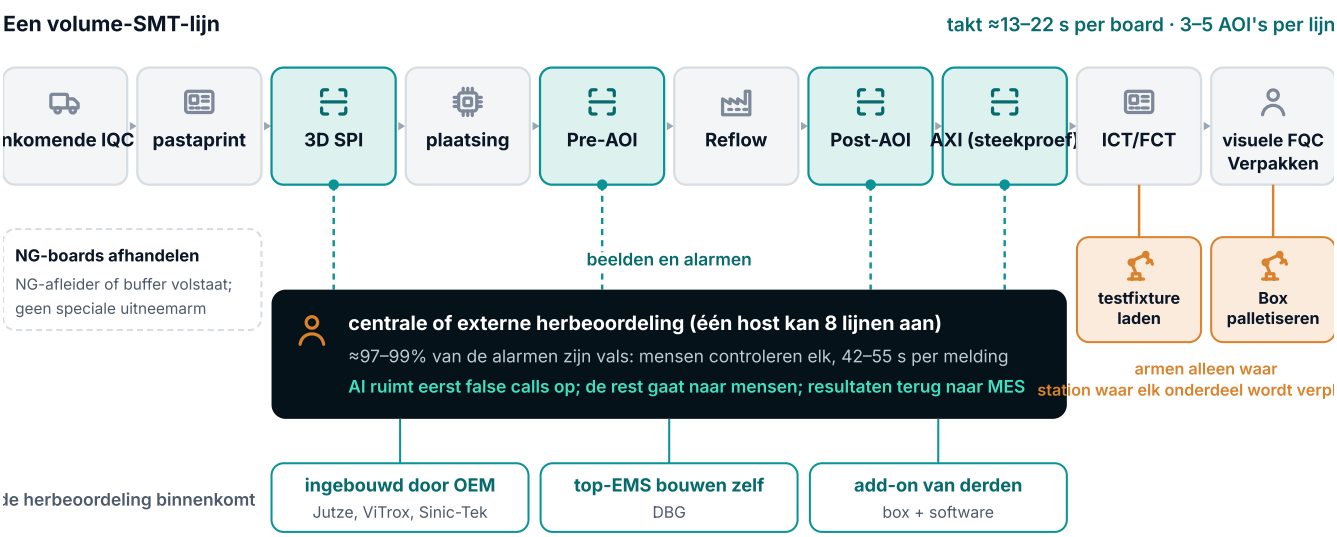
Primaire deponeringen van beursgenoteerde ODM's tonen hoe een volumelijn van 2025–2026 er werkelijk uitziet. Het Hongkongse prospectus van Longcheer (gegevens tot 30 september 2025) vermeldt zijn SMT-kwaliteitscontroles als "röntgeninspectie, soldeerpastinspectie (SPI), geautomatiseerde optische inspectie (AOI), PCBA-functionele test en visuele inspectie"; zijn processtroom verloopt via AOI vóór reflow, AOI na reflow, steekproefsgewijze röntgeninspectie, NG-platen naar een SMT-reparatiestation en daarna een laatste visuele controle. Zijn fabrieken in Huizhou, Nanchang en Vietnam hebben 34, 24 en 8 SMT-lijnen en 67, 60 en 13 eindassemblagelijnen (Longcheer H-share prospectus) single source, primaire deponering. De laptopfabriek van Huaqin uit 2023 plande 18 SMT-lijnen met een AOI-budget van RMB 30.6 miljoen, ongeveer RMB 1.7 miljoen per lijn, wat bij binnenlandse 3D-AOI-prijzen neerkomt op drie tot vijf AOI's per lijn voor inspectie vóór en na reflow van dubbelzijdige platen (Huaqin A-share prospectus) primaire deponering + estimate. Een volumelijn heeft dus gewoonlijk twee tot vier reviewpunten, niet één.

De machines zijn al sterk geautomatiseerd: Huaqin meldt 96% automatisering op zijn SMT-lijnen (Huaqin H-share prospectus) single source. De resterende mensen zitten in review, reparatie, visuele inspectie, test en eindassemblage, en geen enkel bedrijf publiceert hoeveel reviewers het per lijn heeft (hoofdstuk 7).

Drie onafhankelijke schattingen komen overeen op een takt van **13–22 seconden per plaat** (een milieudeponering uit 2026 voor vijf Hebei-lijnen (Hebei SMT-project MER), de laptoplijnen van Huaqin en een iFactory-case), ofwel 7–11 seconden per zijde voor dubbelzijdige platen estimate. Het lastige deel van een high-mix-lijn is niet volume maar de introductie van nieuwe producten en ombouw, en elke ombouw is opnieuw een koude start voor een AI-model.

Aan de softwarekant zijn de eigen reviewstations van de AOI-makers al gecentraliseerd: de reparatiestationsoftware van Sinic-Tek laat "één host de apparatuur van acht SMT-lijnen aansturen" (Sinic-Tek prospectus) single source, primair, 2022, en nadat Jabil in 2026 ViTrox V-ONE remote-AOI-besturing inzette, kon één engineer 5–8 AOI's op afstand beheren tegen 1–2 lokaal (AtlasPCB) vendor claim. Een AI-reviewproduct van derden sluit precies op dit centrale station aan, via de beeld- en resultatexport van de AOI (IPC-CFX, IPC-HERMES-9852 of het eigen formaat van een leverancier). De AOI's zelf voegen snel AI toe: Yamaha zegt dat zijn AI-beoordeling ongeveer één seconde per oproep kost, inspecteurs na 30 NG-beelden helpt en na 1.000 beelden "zeer hoge" kwaliteit bereikt (DirectIndustry/Yamaha) vendor claim.

FIG 02 **Een volume-SMT-lijn: de mensen zitten achter de AOI** ***De arbeid zit geconcentreerd op de herbeoordelingsplekken na automatische optische inspectie. Daar komt AI binnen.***



Bron: Openbare jaarverslagen en leveranciersmateriaal (Huaqin, DBG, Sinic-Tek en anderen); samengesteld door de auteur.

Fase	Gangbare producten en technologie	Representatieve leveranciers	Hoe ver AI is gekomen
Na het printen	3D SPI (gestructureerd licht, moiré, soms lasertriangulatie)	Koh Young, Sinic-Tek, TRI	OEM-AI-pastabeoordeling in productie; Jutze zegt dat sommige modellen “volledig AI” zijn
Na plaatsing en reflow	2D/3D-AOI met meerhoekige, meerkleurige LED’s en gestructureerd licht, 8–20 µm per pixel, 8–12 MP-camera’s	Koh Young, Omron, Saki, TRI, ViTrox, Mirtec, Jutze, Sinic-Tek en andere	OEM-AI-review (Jutze in productie, ViTrox V510Ai, Sinic-Tek met Tencent); top-EMS bouwen zelf (DBG); add-ons van derden
Bodemafgesloten en verborgen verbindingen	Steekproefsgewijze of inline röntgeninspectie (AXI)	—	Hikvision claimt dat zijn Guanlan-model röntgenholtes minder dan 1% van de tijd verkeerd beoordeelt vendor claim
Review en reparatie	Centraal reviewstation (één host, acht lijnen), review op afstand, reparatiestation	Software van AOI-makers, ViTrox V-ONE	AI sluit valse alarmen, mensen doen de rest
Visuele inspectie, IQC, FQC	Meestal handmatig; sommige AI-uiterlijkmachines met robotbelading	SmartMore en andere	Uiterlijkmachines in productie; geen openbare personeelsgegevens voor IQC/FQC
Handling en machinebediening	Zesassige armen en cobots, SCARA, AMR’s, 2D/3D-visiegeleiding	FANUC, ABB, Yaskawa, KUKA, UR, Dobot, JAKA, Elite; Mech-Mind, Hikrobot	Geleerde perceptie plus geprogrammeerde beweging in productie; VLA’s en humanoïden draaien nog in pilot
Data en systemen	MES, IPC-CFX, IPC-HERMES-9852	—	Resultaten teruggeschreven naar MES; ondersteuning in oudere apparatuur is ongelijk

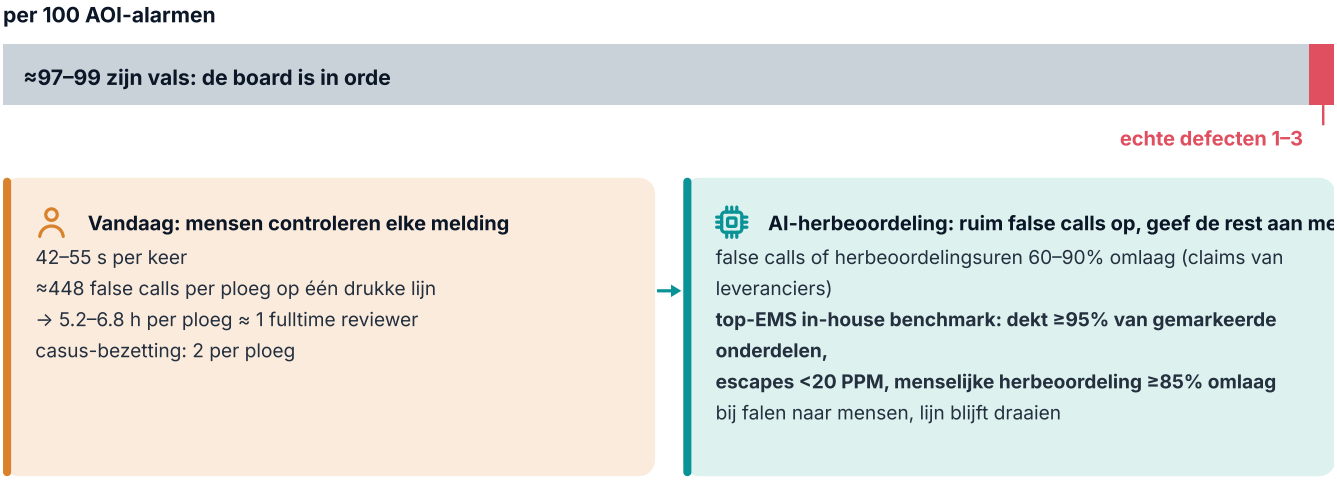
Bronnen bij de tabel: de bovenstaande deponeringen en (Koh Young, Jutze jaarverslag 2025, Sinic-Tek jaarverslag 2025, ViTrox, Sohu over Hengtianyi, Xinhua over Hikvision).

De mensen zitten achter de AOI, omdat de meeste alarmen vals zijn

De inspectieketen op een gangbare SMT/PCBA-lijn verloopt: 3D SPI na het printen voor ontbrekende, onvoldoende of overmatige pasta, overbrugging en offset; AOI na plaatsing (optioneel) en na reflow voor ontbrekende, verschoven, geroteerde, omgekeerde of tombstoned onderdelen en soldeervolume; röntgeninspectie wanneer er BGA's of andere bodemafgesloten onderdelen zijn; daarna in-circuit- of flying-probe-test, functionele test en handmatige visuele inspectie. Elke AOI heeft een reviewplek en een reparatiestation erachter (Sinic-Tek prospectus, JLCPCB) verified. AOI na reflow is het meest gebruikelijk en levert de beste waarde, maar ziet alleen vorm en oppervlak; verbindingen, holtes en bevochtiging onder een BGA vereisen röntgeninspectie of elektrische test (JLCPCB).

De arbeid zit in review omdat de meeste alarmen vals zijn. Eén Britse EMS noteerde **54,019 valse oproepen tegen 544 echte defecten over 917 artikelnummers, ongeveer 99% van de alarmen** (arXiv 2505.03576) single source, preprint; in de fabriek van Siemens in Rastatt was **97.5% van de platen die de AOI afkeurde goed** (Siemens-casestudy) vendor claim. Een drukke lijn ziet ongeveer 448 valse oproepen per ploeg van elk 42–55 seconden, 5.2–6.8 uur per ploeg, ruwweg één voltijds reviewer (iFactory) vendor claim; uren zijn een schatting.

FIG 03 **Ongeveer 97–99% van de AOI-alarmen zijn false calls *Vandaag bevestigt een persoon elke melding; AI-herbeoordeling ruimt eerst de valse op en geeft de rest aan mensen.***



Bron: Siemens-casestudy, iFactory (claim van leverancier), DBG-jaarverslag 2025; herbeoordelingsuren geschat door de auteur.

Uit "review-uren per ploeg = platen × valse oproepen per plaat × seconden per oproep" kan één reviewer ongeveer 0.7–1 volumelijn dekken met een op regels gebaseerde AOI. Zodra AI 60–90% van de valse oproepen verwijdert, zou alleen al de beoordelingstijd 2–14 lijnen per persoon toelaten, maar loopafstanden, NG-bufferafhandeling en reparatie-overdrachten beperken de lokale dekking tot twee tot vier aangrenzende lijnen; alleen centrale review of review op afstand komt in de buurt van het theoretische plafond estimate; seconden per oproep komen uit één bron. Omdat gangbare AOI's 60–80 cm² per seconde scannen (Yunheng AOI-gids 2026), duurt een zijde van 20×15 cm 4–5 seconden; de AI hoeft alleen de paar gemarkeerde gebieden binnen één takt te verwerken, niet de plaat opnieuw te scannen.

Wat al werkt in productie: gesuperviseerde classificatie op de eigen beelden van de AOI

Het bewezen patroon is gesuperviseerde deep-learning-classificatie op beelden van de AOI die een fabriek al bezit. De add-on van Fujitsu op de AOI van ADLINK in Taiwan verlaagde valse oproepen met **meer dan 90%** (Fujitsu-case); OKI gebruikt sinds juli 2026 AI om de handmatige bevestiging na AOI met ongeveer **80%** te verminderen in zijn eigen EMS-activiteiten (Circuits Assembly); Siemens meldt 60% minder handmatige analyse, €160,287 bespaard in het eerste jaar en een terugverdientijd van acht maanden (Siemens-casestudy) allemaal vendor claims. Twee dingen zijn gemeenschappelijk aan deze cases: vrijwel niemand publiceert een doorglipperpercentage, en allemaal

houden ze mensen op de resterende alarmen. De enige case die een doorglippercijfer geeft (0.8% naar 0.3%, tegen een contractuele 0.5%) is intern inconsistent (iFactory). AI verlaagt review-uren, niet de reviewtaak zelf.

AI-review is verschoven van add-ons van derden naar OEM-functies en interne systemen, wat verandert waartegen een nieuwkomer wordt afgemeten. Het jaarverslag 2025 van Jutze vermeldt zijn "3D-AOI-upgrade" als in volumeproductie en zegt dat het "AI-algoritmen gebruikt om handmatige review aan het inspectie-einde te verwijderen" (Jutze jaarverslag 2025); Sinic-Tek ontwikkelt sinds 2019 een "multimodaal AI-ondersteund handmatig reviewsysteem" en ontwikkelt samen met Tencent een boven-onder-3D-AOI, gericht op 99.5%+ defectdekking en 98% herkenningsnauwkeurigheid (Sinic-Tek jaarverslag 2025) vendor claims in primaire deponeringen. Grote EMS-bedrijven bouwen zelf. Het jaarverslag 2025 van DBG vermeldt "AI-beoordeling voor visuele inspectie" als een voltooid project dat "meer dan 95% van de door de AOI gemarkeerde componenten dekt", met een "doorglipperpercentage onder 20 PPM" en "meer dan 85% minder handmatige review", volledig traceerbaar (DBG jaarverslag 2025) single source, primair. Dat is het enige elektronicafabriek-cijfer dat ik vond dat dekking, doorglippers en reductie samen geeft **Een product van derden dat aan tier-1-EMS wordt verkocht, moet de 95% dekking, 20 PPM doorglippers en 85% reductie van DBG evenaren**; onder tier 1 is de benchmark de 50–85% reductie die leveranciers adverteren.

Anomaliedetectie die alleen op goede onderdelen is getraind, is snel (EfficientAD-S kost 2.2 ms per beeld (EfficientAD, WACV 2024)) maar zwak op "logische" anomalieën zoals ontbrekende of verkeerde onderdelen, en de beste methoden halen gemiddeld onder 60% AU-PRO op de moeilijkere MVTec AD 2 (MVTec AD 2) single source. Ontbrekende en verkeerde onderdelen kunnen beter tegen CAD en BOM worden gecontroleerd.

Het plafond tegen elke prijs: lagen machines, zeer weinig mensen, maar niet nul

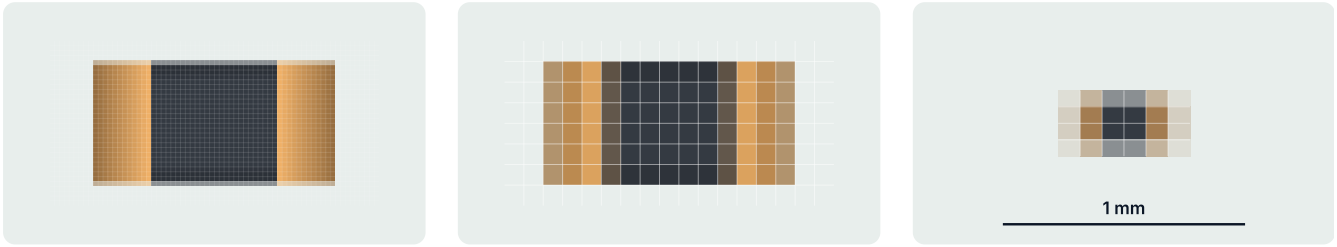
Ik vond geen elektronicafabriek die handmatige review volledig verwijderde en nul doorglippers bereikte. De praktijk voor IPC-A-610 Klasse 3-producten (hoge betrouwbaarheid) laat nog steeds getrainde inspecteurs de grensgevallen beslissen die automatische drempels niet kunnen classificeren, en één PCB-assemblagebedrijf noemt "geautomatiseerde inspectie kan menselijke aftekening vervangen" een "veelvoorkomend misverstand" (PCBCart) single source. De menselijke nulmeting moet ook worden gecorrigeerd: verwijzend naar een Sandia-studie uit 2012 vangt een enkele visuele inspecteur op zijn best ongeveer **80%** van de defecten, twee in serie ongeveer 96% (Instrumental) single source, geciteerd door een leverancier. De verdedigbare claim is "AI plus een kleine menselijke reviewlus verslaat volledig handmatige inspectie", niet "AI vervangt inspectie".

Zonder budgetlimiet is het plafond 3D SPI, 3D-AOI, röntgeninspectie, AI-review en AI-uiterlijkstations op elkaar gestapeld, met mensen alleen nog voor grensgevallen, nieuwe producten en ombouw, nieuwe defecttypen en aftekening door de klant. Vlaggenschipfabrieken melden 85–92% minder handmatige inspectie (I-Connect007, 2021, Wi-Advance) vendor claims; mijn schatting van het budgetloze plafond is reviewarbeid op 8–15% van vandaag estimate. Eindassemblage blijft arbeidsintensief: het automatiseren van de iPhone 16-knoppen werd gestopt vanwege hoge defectpercentages (9to5Mac) single source.

Een algemene hogeresolutiecamera is geen AOI

Productie-AOI's nemen beeld per tegel op met 8–20 µm per pixel, met 8–12 MP-camera's, programmeerbare meerhoekige meerkleurige belichting en gestructureerd-licht-3D (Koh Young, Saki) verified. De industrie merkte lang geleden op dat een 01005-onderdeel bij 25 µm per pixel slechts ongeveer 8×16 pixels beslaat, "mogelijk niet genoeg" voor betrouwbare beoordeling (Circuits Assembly, 2007). Met dezelfde rekensom estimate haalt een 4K-camera over een paneel van 300 mm ongeveer **78 µm per pixel**: een imperiaal 0402-onderdeel is ongeveer 13×6 pixels, genoeg voor aanwezigheid en grove offset maar niet voor de vorm van de verbinding, en een 01005 is ongeveer 5×3 pixels, in feite onzichtbaar. Het bereiken van 15 µm per pixel over een heel paneel zou ongeveer 330 megapixels vergen, en daarom bewegen AOI's een camera over de plaat. Wanneer iemand "0402" zegt, controleer dan welke hij bedoelt: metrisch 0402 is imperiaal 01005, ver buiten elke groothoekcamera.

FIG 04 **Een universele camera is geen AOI** ***Een 4K-camera over een paneel van 300 mm geeft ongeveer 78 µm per pixel: een 0402-onderdeel krimpt tot ≈13×6 pixels en een 01005 verdwijnt.***



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

4K-camera over een paneel van 300 mm: ≈78 µm/px
0402 krimpt tot ≈13×6 px: genoeg voor aanwezigheid en grove offset

zelfde camera op een 01005
≈5×3 px: effectief onzichtbaar

Bron: Circuits Assembly; berekening van de auteur voor een paneel van 300 mm en een 4K-sensor. Pixelrasters zijn illustratief.

Op lijnen die al AOI's hebben, is de beste plaats voor AI bovenop de beelden van de AOI. Alleen lijnen zonder AOI, of end-of-line-stations, rechtvaardigen een afgesloten camerastation met gecontroleerde belichting; een open hoge-resolutiecamera boven de lijn is het meest blootgesteld aan omgevingslicht en schaduwen (CSDN).

Assemblage en handling: geleerde perceptie plus geprogrammeerde beweging is in productie

Het IFR-rapport 2026 zegt dat "machinevisie voor inspectie en robotgeleiding volwassen en industrieklasse is", terwijl "algemene fysieke AI en leren uit simulatie en ervaring nog niet industrieklasse zijn" (IFR World Robotics 2026 managementsamenvatting) single source, primair . Elektronica is de grootste robotafnemer (142,235 eenheden wereldwijd geïnstalleerd in 2025, 59% van alle installaties in China (IFR, IFR China-uitgave)), toch heeft China slechts 166 robots per 10,000 productiearbeiders tegen 307 in de VS (IFR, april 2026). Het patroon in productie is geleerde perceptie plus deterministische beweging; end-to-end geleerde beweging is alleen op gestructureerde stations nabij productiekwaliteit (hoofdstukken 3 en 6).

Wat elke regio kan gebruiken

Vasteland-China vertrouwt op binnenlandse modellen en lokale gewichten; in Taiwan beslist de identiteit van de klant; in de VS doen invoerrechten en certificering dat.

Element	Vasteland-China	Taiwan (private fabrieken)	Zuidoost-Azië	Verenigde Staten
Claude (elk kanaal)	● Niet beschikbaar	● Voorwaardelijk: de leverancier mag niet in meerderheid in eigendom zijn van of worden gecontroleerd door personen of entiteiten in vasteland-China, Hongkong of Macau	● Idem	● Idem
GPT- en Gemini-cloud	● Niet beschikbaar	● Beschikbaar	● Beschikbaar (lijsten niet land voor land gecontroleerd)	● Beschikbaar

Element	Vasteland-China	Taiwan (private fabrieken)	Zuidoost-Azië	Verenigde Staten
Chinese cloudmodellen (Qwen, DeepSeek, Doubao, GLM, Kimi)	● Beschikbaar (gemelde modellen; het aanbieden van een dienst vereist registratie)	● Legaal, maar hosting op het vasteland doorstaat zelden de beveiligingsreview van de klant; verboden in de publieke sector, staatsbedrijven en kritieke infrastructuur	● Legaal (internationale nodes)	● Legaal, strikte review
Lokale open gewichten van Chinese oorsprong (Qwen, GLM, DeepSeek)	● Eerste keuze; private in-plant-implementatie vereist geen model- of algoritmemelding (lezing door advocatenkantoor)	● Beschikbaar (publieke sector beperkt)	● Beschikbaar	● Voorwaardelijk: oorsprongsreview
Lokale open gewichten van niet-Chinese oorsprong (bv. Gemma 4)	● Beschikbaar	● Beschikbaar	● Beschikbaar	● Beschikbaar
Klassieke niet-generatieve visiemodellen	● Beschikbaar (buiten de regels voor generatieve AI)	● Beschikbaar	● Beschikbaar	● Beschikbaar
Jetson Thor/T5000/T4000, DGX Spark	● Beschikbaar: NVIDIA classificeert ze als 5A992.c, geen Amerikaanse licentie nodig voor een gewone civiele fabriek; Spark heeft een China-SKU; kanaalprijzen ongeveer 40–60% boven de Amerikaanse adviesprijs	● Beschikbaar	● Beschikbaar	● Beschikbaar
RTX 5090, RTX PRO 6000, DGX Station	● Niet beschikbaar: de eerste twee zijn 4A090.a, DGX Station 5A992.z; legale alternatieven zijn 4090D, 5090 D V2, RTX PRO 5000 72GB, RTX 6000D, L20 (4A994.l)	● Beschikbaar	● Beschikbaar, maar kopers met hoofdkantoor in of uiteindelijk eigendom van China tellen als Chinees; Maleisië vereist vergunningen voor export en doorvoer	● Beschikbaar
Mac Studio M5 Ultra 512 GB	● Beschikbaar (Apple China-winkel)	● Beschikbaar	● Vermoedelijk beschikbaar	● Beschikbaar
Huawei Ascend en andere Chinese accelerators	● Beschikbaar	● Hoog risico	● Hoog risico	● In de praktijk niet bruikbaar
Vasteland-gemaakte vaste armen, AOI's, industriële camera's	● Beschikbaar	● Invoerbaar; sinds 1 december 2025 kunnen de autoriteiten rechtstreeks optreden tegen vasteland-merken, en kunnen agentschappen, staatsbedrijven, bepaalde stichtingen en aanbieders van kritieke infrastructuur worden beperkt of uitgesloten van het gebruik ervan	● Beschikbaar	● Legaal, maar gecombineerd invoerrecht is ongeveer 40% op armen, 37.5% op instrumenten en camera's van het AOI-type, 25% op edge-computers

Element	Vasteland-China	Taiwan (private fabrieken)	Zuidoost-Azië	Verenigde Staten
Buitenlandse AMR's en mobiele manipulatoren	● Beschikbaar	● Beschikbaar	● Beschikbaar	● Nieuwe modellen vereisen voorwaardelijke goedkeuring van het Amerikaanse ministerie van Oorlog
Hikvision- en Dahua-camera's	● Beschikbaar	● Privégebruik toegestaan, publieke sector verboden	● Beschikbaar	● Nieuwe FCC-autorisaties geblokkeerd; verboden voor federale aannemers
Robotveiligheid en certificering	● GB 11291.1-2011 en GB 11291.2-2013 verplicht, sinds april 2026 in herziening; CR-certificering vrijwillig; GB/T 39404-2026 dekt cybersecurity van de controller	● Standaard afrastering; hekwerkloze samenwerking moet voldoen aan CNS 14490 of ISO 10218 met een schriftelijke beoordeling	● Per klant en exportbestemming, schriftelijk bevestigd vóór verzending	● Elektrische apparatuur moet "goedgekeurd" zijn (in de praktijk een NRTL-keurmerk of veldevaluatie); UL 1740 vereist sinds 11 augustus 2025 dat robotcellen voldoen aan ISO 10218-2
Fabs en fabrieken met gecontroleerde producten	—	● De inkoopcontracten van TSMC vereisen SEMI E187; vasteland-apparatuur wordt uit 2 nm-fabs verwijderd	—	● ITAR-technische gegevens omvatten "foto's"; ontzeggingsbeleid voor China; door het vasteland gecontroleerde leveranciers kunnen defensie-elektronicafabrieken in de praktijk niet bedienen

Bronnen bij de tabel: (Anthropic ondersteunde regio's, OpenAI ondersteunde landen, Gemini-beschikbaarheid, CAC tussentijdse maatregelen voor generatieve AI, NVIDIA exportclassificatiegegevens, eCFR 15 CFR 774, NVIDIA China DGX Spark, Apple China, Covington over nieuwe BIS-regels, Greenberg Traurig, juni 2026, Fulcrum, Tom's Hardware over Ascend-richtlijnen, Taiwan BOFT-invoerlijst, Geschiedenis Cyber Security Management Act Taiwan, Taiwan-review-regels voor producten die de nationale cybersecurity in gevaar brengen, Taipei Times, USITC HTS hoofdstuk 99, Latham & Watkins, IPVM, China nationaal normenplatform, GB 11291.2, Intertek UL 1740-update, eCFR 22 CFR 126.1).

Vasteland-China: een volledige stack, maar noch gesloten frontier-modellen noch de sterkste losse machines kunnen binnenkomen. Gemelde binnenlandse cloudmodellen, lokale open gewichten (Qwen onder Apache-2.0, GLM en DeepSeek onder MIT), klassieke visiemodellen en binnenlandse AOI's, camera's en cobots zijn allemaal bruikbaar. De eigen exportclassificaties van NVIDIA vermelden de Jetson AGX Thor-kit, de T5000/T4000-modules en DGX Spark als 5A992.c, zodat een gewone EMS- of PCBA-fabriek geen Amerikaanse licentie nodig heeft, terwijl de RTX 5090 en RTX PRO 6000 (4A090.a) en DGX Station (5A992.z, TPP 58,484) niet op het vasteland kunnen worden gekocht (NVIDIA exportclassificaties, eCFR) [primair; de juridische lezing is van mij en is geen juridisch advies]. RS China vermeldt de Thor-kit op RMB 59,104 inclusief belasting, ongeveer 1.4 keer de Amerikaanse adviesprijs (RS China).

Vier advocatenkantoren zijn het erover eens dat een private in-plant-implementatie voor eigen gebruik van de fabriek geen model- of algoritmemelding vereist, terwijl een gehoste multi-tenant-dienst in een grijze zone zit (Covington, FPF, JunHe) verified: consistente lezingen, geen officiële uitspraak. GB 11291.1-2011 en GB 11291.2-2013 blijven de verplichte robotveiligheidsbasis en zijn sinds 28 april 2026 in herziening; GB/T 39404-2026 over controllercybersecurity trad op 1 november 2026 in werking (herzieningsplan 20262542, GB/T 39404-2026) single source, primair. Tot de nieuwe GB 11291 er is (2028–2029 estimate), ontwerp naar ISO 10218:2025 en bewaar de GB 11291-conformiteitsdossiers.

Taiwan: private fabrieken kunnen vrijwel alles gebruiken; de identiteit van de klant beslist. Frontier-API's, open gewichten en vasteland-apparatuur zijn allemaal legaal (data.gov.tw). Maar sinds de gewijzigde Cyber Security Management Act op **1 december 2025** in werking trad, kunnen de autoriteiten rechtstreeks optreden tegen producten die "vasteland-merken" zijn en deze beperken bij overheidsinstanties, staatsbedrijven en aanbieders van kritieke infrastructuur (Laws & Regulations Database, reviewregels) verified; de lijst van aanbieders van kritieke infrastructuur is vertrouwelijk; en TSMC heeft SEMI E187 in zijn inkoopcontracten opgenomen (CNA) en zou naar verluidt vaste-

land-apparatuur uit zijn 2 nm-fabs verwijderen (TrendForce) single source. Private EMS-, assemblage- en kleine PCBA-fabrieken staan open; fabs, parknutsbedrijven en overheidsgerelateerde klanten zijn in de praktijk gesloten. Een door het vasteland gecontroleerde leverancier moet naar E187 ontwerpen, een software bill of materials leveren, offline updates aanbieden en nooit op afstand verbinding maken vanuit het vasteland.

Zuidoost-Azië: de minste beperkingen; de problemen zijn trage terugverdientijd en Chinese moeders. Alles is bruikbaar, maar de Amerikaanse richtlijn van 31 mei 2026 vereist een licentie voor de export van geavanceerde rekenitems naar entiteiten met hoofdkantoor in, of uiteindelijk eigendom van, China, zelfs wanneer de ontvanger zich in een derde land bevindt (Greenberg Traurig) single source, en Maleisië heeft zijn eigen vergunningen voor geavanceerde chips (Fulcrum). De vraag verschuift hierheen: TrendForce verwachtte dat Thailand, Vietnam en Maleisië in 2026 bijna 50% van de server-ODM-SMT-capaciteit zouden benaderen (TrendForce) single source, een prognose uit 2023.

Verenigde Staten: de volledige stack, tegen de prijs van invoerrechten, certificering en gecontroleerde gegevens. Onder hoofdstuk 99 van het Amerikaanse tariefschema betaalt een op het vasteland gemaakte robotarm ongeveer 40% gecombineerd invoerrecht, instrumenten en camera's van het AOI-type ongeveer 37.5% en edge-computers 25%; gemaakt in Maleisië of Indonesië ongeveer 10–12.5%, in Vietnam, Thailand of de Filipijnen 12.5–15%, in Taiwan 10%, met computers op 0% buiten het vasteland (USITC HTS hoofdstuk 99, Great Lakes Customs Law) primair schema + estimate. De kennisgeving van USTR van 10 november verlengde 178 oude Section 301-uitsluitingen tot 9 januari 2027, waarvan geen enkele robots, inspectie-instrumenten, camera's of computers dekt (Federal Register 2026-22925) verified.

OSHA vereist dat elektrische apparatuur op de werkplek "goedgekeurd" is (in de praktijk een NRTL-keurmerk, of een veldevaluatie van ongeveer $3,000 tot $20,000+ per cel), en sinds 11 augustus 2025 brengt UL 1740 hele robotcellen onder ISO 10218-2 (eCFR 1910.399, Intertek, Industrial Monitor Direct). Defensie-elektronica is in de praktijk gesloten voor door het vasteland gecontroleerde leveranciers: ITAR behandelt productiefoto's als technische gegevens en past een ontzeggingsbeleid toe op China (eCFR 22 CFR 120.33, eCFR 22 CFR 126.1). Een cel voor de VS moet een vaste arm, onbeperkte camera's en lokale of VS-cloudmodellen gebruiken, met eindassemblage in Zuidoost-Azië waar dat de oorsprong werkelijk verandert.

Bij elk budget verschilt het plafond per regio maar bereikt het nergens nul mensen estimate. Het vasteland kan binnenlandse 3D SPI, 3D-AOI, röntgeninspectie, AI-review en een intern AI-team stapelen om reviewarbeid terug te brengen tot 8–15%, maar zonder gesloten frontier-modellen of DGX Station moet fabrieksbrede LLM-Q&A draaien op lokale open gewichten en binnenlandse rekenkracht (een volledig model voor veel gebruikers vereist ten minste twee acht-kaart Ascend-servers). Taiwan en Zuidoost-Azië kunnen technisch het beste van alles gebruiken, inclusief frontier-API's en DGX Station, beperkt door klantidentiteit, vasteland-merken en licenties voor Chinese moeders. De VS hebben geen technische limieten, tegen de prijs van invoerrechten, NRTL-certificering en personeelsscheiding voor gecontroleerde gegevens. Overal blijven mensen voor grensgevallen, nieuwe producten en ombouw, nieuwe defecttypen en aftekening door de klant.

Een product dat over regio's wordt verkocht, heeft dus twee modelstacks nodig: een vasteland-stack op gemelde binnenlandse modellen of lokale open gewichten, en een overzeese stack op frontier-API's plus een lokaal model van niet-Chinese oorsprong. Het beleid van Anthropic, van kracht op 12 november 2026, oordeelt op basis van eigendom en controle, zodat een door het vasteland gecontroleerde leverancier niet op Claude kan bouwen, zelfs niet via een overzeese dochter, een cloudplatform of een reseller (Anthropic) single source, primair.

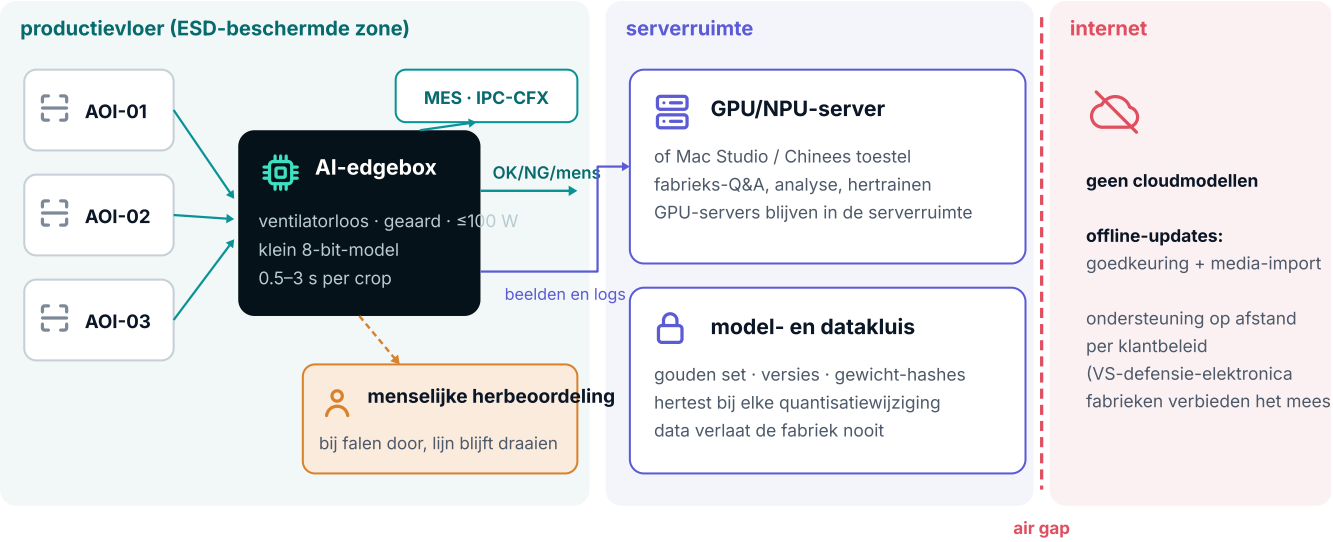
On-premises

02 AI binnen de fabriek draaien: cleanrooms, luchtspleten, hardware en kwantisering

In een luchtgescheiden fabriek is de beperking proces, niet rekenkracht

Volledig offline visuele inspectie is routine; de beperkingen zijn procedureel. De MIIT-richtlijn van 2024 over de beveiliging van industriële besturingssystemen vraagt om zonering, firewalls of datadiodes, gecontroleerd tijdelijk onderhoud op afstand, gescande media en geteste patches (richtlijntekst) single source ; halfgeleiderklanten voegen SEMI E187 en E188 toe, die TSMC in 2023 in zijn inkoopcontracten opnam (SEMI, TXOne) verified . Resultaten gaan terug naar MES via IPC-CFX en plaatidentiteit reist via IPC-HERMES-9852 (Hermes-standaard). De veelvoorkomende valkuil is licenties die een netwerk veronderstellen: de offlinemodus van Roboflow cacht gewichten op een lease van 30 dagen (Roboflow). Een offline product heeft offline activering nodig, ondertekende updates op gescande media, een software bill of materials, USB-controle, lokale auditlogs en een CFX-adapter.

FIG 05 **Een air-gapped uitrol** ***Een kleine box aan de lijn beslist; de serverruimte beantwoordt vragen; bij falen vallen ze terug op mensen en data verlaat de fabriek nooit.***



Bron: Synthese van de auteur (hoofdstuk 2).

Gesloten frontier-modellen komen in de praktijk niet in de serverruimte van een fabriek. Gemini wordt alleen luchtgescheiden aangeboden via Google Distributed Cloud voor overheden en grote ondernemingen (Google Cloud); Foundry Local van Microsoft opent grote modellen alleen voor “gekwalificeerde klanten” en noemt geen GPT-model (Microsoft); Claude wordt alleen aangeboden in cloud-overheidsregio’s, zonder on-premises-implementatie (Claude voor overheden); geen van de drie wordt in vasteland-China aangeboden. Fabrieksapparatuur moet rond open gewichten worden ontworpen; frontier-modellen kunnen hoogstens beelden labelen of uitleggen op het hoofdkantoor, offline, wanneer de klant geanonimiseerde beelden naar buiten laat.

Serverruimte-hardware moet als kantoorwaardig worden behandeld: de Mac Studio is geschikt voor 10–35 °C en DGX Spark voor 5–30 °C, met meldingen van thermische uitschakeling bij aanhoudende inferentie (Apple, NVIDIA-forum) verified . De lijn krijgt ventilatorloze industriële boxes, zoals een Jetson T4000 industriële PC geschikt voor

−25 tot 60 °C (AAEON) vendor claim. Omdat een uur stilstand een grote fabriek $36,000 tot $2.3 miljoen kost (Siemens True Cost of Downtime 2024) vendor-enquête, moet elke AI-stap op de lijn **fail open** zijn: bij elke fout valt hij terug op het bestaande AOI-plus-mensen-proces.

Clean- en ESD-zones: een gewone SMT-hal is een ESD-beschermde zone, geen geclassificeerde cleanroom

Wat beperkt waar rekenkracht kan staan, is statische elektriciteit en luchtstroom, niet de reinheidsklasse. De meeste SMT-assemblagegebieden zijn “niet formeel deeltjesgecontroleerd”, en wereldklasse-fabrieken “kunnen ISO 8 bereiken of benaderen” (Circuitnet-expertpanel) single source. Medische PCBA en cameramodules zitten doorgaans in ISO 7, bare die en optische uitlijning in lokale ISO 5–6 (PCBSync) single source; cleanroombouwers noemen ISO 7–8 voor SMT in het algemeen (Angstrom Technology) vendor claim.

De universele eis is ESD-beheersing: ANSI/ESD S20.20-2021 vereist geïsoleerde geleiders onder 35 V en houdt sterkveld-isolatoren ten minste 30 cm van gevoelige apparaten (ANSI, ESDA, Desco) verified. Fabs plaatsen rekenkracht in de sub-fab of het datacenter, in vloeistofgekoelde racks die voldoen aan SEMI S2/S8/F47 (Prodrive) vendor claim.

Of apparatuur een schone ruimte mag binnengaan, wordt getest op deeltjesemissie onder ISO 14644-14; robots hebben dergelijke certificaten (Fraunhofer IPA), maar ik vond geen industriële PC, edge-AI-box of GPU-server met zo’n certificaat. CE, FCC en IEC 60068 tonen dat een box zijn omgeving verdraagt, niet dat hij een ESD-beschermde zone mag binnengaan; de ESD-coördinator van de klant controleert dat. GPU-servers gaan nooit de lijn op: een DGX Station trok 1,292 W bij de GPU in tests (StorageReview).

FIG 06 **Clean- en ESD-zones bepalen waar rekenkracht mag staan** ***Een gewone SMT-hal is een ESD-beschermde zone, geen cleanroom. GPU- en NPU-servers horen in de serverruimte.***



Bron: ANSI/ESD S20.20, IEC 61340-5-1, ISO 14644, SEMI-normen en leveranciersmateriaal; samengesteld door de auteur.

Zone	Typische omgeving	Rekenkracht aan de lijn	GPU/NPU-servers
Gewone SMT/PCBA-hal (consument, industrieel)	ESD-beschermde zone onder S20.20 of IEC 61340-5-1, meestal geen geclassificeerde cleanroom (≈ISO 8-equivalent); 40–60% RV	Ventilatorloze box in een geaarde metalen behuizing (Jetson T4000/Thor, Chinese NPU-box) naast de AOI of reviewplek; plastic monitors en behuizingen uit de zone of ≥30 cm van platen	Serverruimte of klimaatgecontroleerd IT-rek

Zone	Typische omgeving	Rekenkracht aan de lijn	GPU/NPU-servers
Medisch, cameramodules, EV-elektronica (ISO 7–8-cleanroom)	Cleanroom plus ESD-bescherming	Ventilatorloos, IP65/66-afgedicht, glad en makkelijk af te vegen; vraag ISO 14644-14-gegevens of een IPA-certificaat	Buiten de cleanroom, ingenetwerkt
Bare die, wire bond, optische stations (lokale ISO 5–6)	Mini-omgeving	Alleen camera's en verlichting binnen; rekenbox buiten	Serverruimte
Fab-cleanroom (ISO 5 en schoner)	Fab-specifieke ESD- en elektrostatische-aantrekkingsplannen, apparatuur naar SEMI E78; cybersecurity naar SEMI E187/E188	Alleen tool-geïntegreerde, fab-gekwalificeerde controllers; geen add-on-PC's met ventilatoren	Vloeistofgekoelde sub-fab-racks of het fab-datacenter

De tabel is mijn synthese van de bovenstaande bronnen estimate. Voor een product betekent dat een ventilatorloze lijnzijdige box onder ongeveer 100 W, een metalen behuizing verbonden met beschermende aarde, 12–24 V DC-ingang, werking bij breed temperatuurbereik bij echte kastluchtstroom, en een ISO 14644-14-deeltjesemissierapport klaar voor klanten op ISO 7 en schoner. Elk ontwerp waarbij "één GPU-server vier lijnen bedient", plaatst de server in de serverruimte en laat aan de lijn slechts een kleine box of een camera.

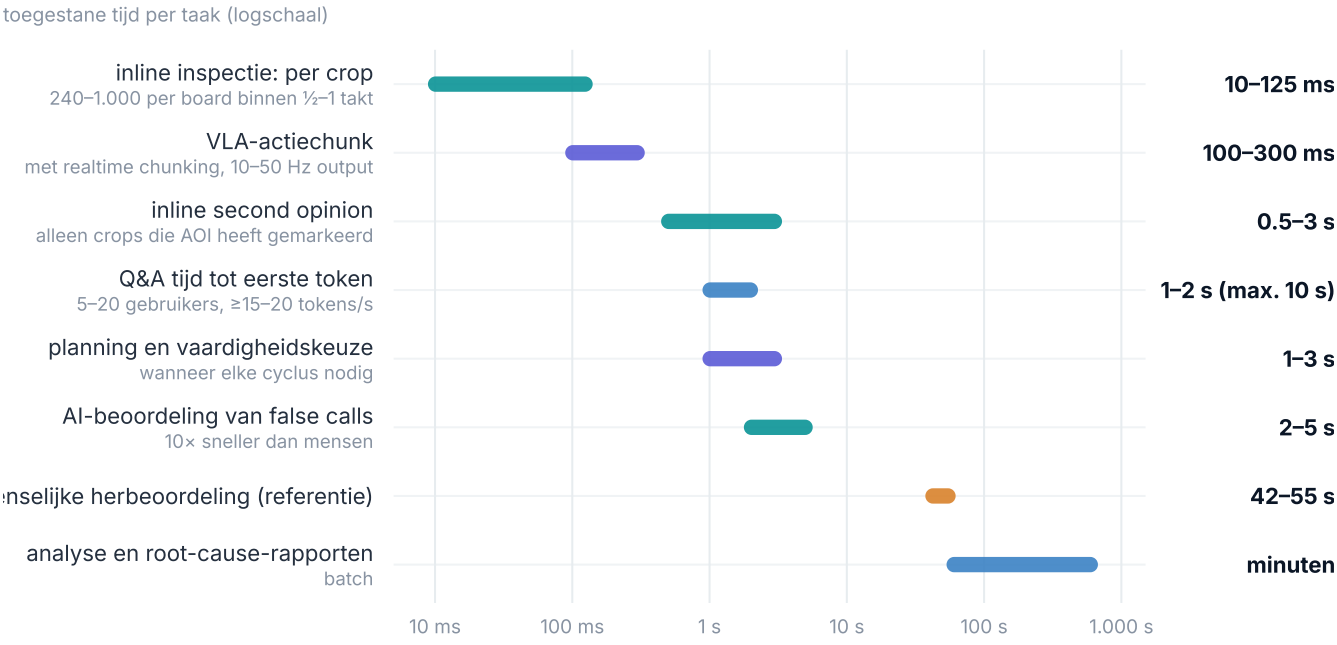
Snelheidslimieten per gebruik: nauwkeurigheid is moeilijker dan snelheid

Fabrieken en leveranciers publiceren vrijwel nooit AI-latentie- of beschikbaarheidsdoelen. De onderstaande drempels zijn afgeleid van takt, menselijke reviewtijd, robotbesturingsfrequenties en interactieonderzoek estimate:

Gebruik	Snelheidslimiet	Nauwkeurigheidslimiet	Beschikbaarheid
Inline-inspectie van elk onderdeel	Binnen ½–1 takt: 240–1,000 uitsnedes in een takt van 10–30 s, 8–100 uitsnedes per seconde	Doorglippers niet slechter dan de huidige "AOI plus mensen" (contractvoorbeeld <0.5%)	Fail open, stop de lijn nooit; doel ≥99.9% (geen publieke SLA)
Inline tweede mening op alleen AOI-gemarkeerde uitsnedes	Markeringen (meestal ≤5–20 per plaat) binnen ½ takt: 0.5–3 s per uitsnede	Idem; getraind op de eigen beelden van de lijn	Valt terug op menselijke review
AI-review van valse alarmen	2–5 s per oproep, meer dan tien keer sneller dan de 42–55 s van een persoon	Doorglippers niet slechter dan een menselijke reviewer	Degradatie voegt mensen toe, stopt de lijn nooit
Robotbesturing (VLA)	100–300 ms per actie-chunk met realtime-chunking, 10–50 Hz-output	≥99% taakslagingskans om geprogrammeerde beweging te verslaan	Veiligheidssysteem onafhankelijk van het model
Robotplanning en vaardigheidskeuze	1–3 s wanneer elke cyclus nodig; ≤1 min alleen bij uitzonderingen of ombouw	Output moet verifieerbare code of een vaardigheids-ID zijn	Valt terug op het vaste programma
Q&A (SOP's, MES-query's)	Eerste token idealiter ≤1–2 s, hooguit 10 s; ≥15–20 tokens/s per gebruiker; 5–20 gelijktijdige gebruikers	Nauwkeurige citaten; trouw over lange context	Kantoor-IT-waardig (≈99–99.5%)
Analyse en root-cause-rapporten	Minuten zijn prima (batch)	Hoogste; trouw over lange context	Kantoor-IT-waardig

Physical Intelligence mat 97 ms inferentie voor een VLA van 3B-klasse (ongeveer 108–139 ms met netwerk en voorbewerking), stelde vast dat realtime-chunking de doorvoer ongewijzigd houdt tot 200 ms extra vertraging, en voltooide fijne taken zoals het aanstrijken van een lucifer of het insteken van een netwerkkabel bij meer dan 300 ms (Physical Intelligence) vendor-onderzoek ; interactieonderzoek legt de grens voor het vasthouden van een gedachtegang bij één seconde en voor aandacht bij tien (NN/g). Nauwkeurigheid is het lastige deel: open Qwen3-VL-8B haalt gemiddeld slechts 56.8 zero-shot over vijf industriële anomaliedatasets (IndusAgent) single source . Geen enkele hardware redt een vision-language-model dat niet op de eigen beelden van de lijn is afgestemd.

FIG 07 **Snelheidsgrenzen per toepassing***Inline-inspectie wordt gemeten in milliseconden, Q&A in seconden, analyse in minuten. Mensen doen 42–55 s over een melding.*



Bron: Synthese van de auteur (hoofdstukken 1–2).

Hardware: Thor voor het station, een Mac Studio of binnenlands apparaat voor fabrieks-Q&A, en geen DGX Station op het vasteland

Apparaat	Geheugen en bandbreedte	Prijs (november 2026)	Waar het goed voor is	Vasteland-China
Jetson AGX Thor devkit / T5000 / T4000	128 GB LPDDR5X, 273 GB/s; 40–130 W	$3,499 bij lancering → $5,499 (juli 2026); T5000 $4,999, T4000 $2,999; RS China devkit RMB 59,104 incl. btw	2–8B vision-modellen die uitsnedes hercontroleren; realtime VLA's van 3B-klasse	● Ja (5A992.c, geen licentie voor civiele fabrieken)
DGX Spark	128 GB, 273 GB/s	$3,999 → $4,699 → $6,950; hetzelfde-chip ASUS GX10 China-model RMB 32,999 op JD (okt 2025)	Multi-user 8–35B MoE	● Ja (5A992.c, China-model, geautoriseerde kanalen)
RTX PRO 6000	96 GB GDDR7	≈$8,565 → $13,250 (juni 2026) → ≈$16,005 gemiddelde eBay-notering in november	Multi-user 27–35B FP8; 120B-klasse MoE op 4-bit	● Nee (4A090.a)
Mac Studio M5 Ultra 512 GB	512 GB unified, 1.2 TB/s; 480 W max continu	Vanaf $22,299 in de VS; ≈RMB 173–186k geschat in China	DeepSeek 671B op 4-bit voor 1–3 gebruikers	● Ja

Apparaat	Geheugen en bandbreedte	Prijs (november 2026)	Waar het goed voor is	Vasteland-China
DGX Station GB300	748 GB (252 GB HBM3e op 7.1 TB/s + 496 GB LPDDR5X op 396 GB/s); 1,292 W gemeten bij de GPU	$123–140k straatprijs in november (≈RMB 830–940k vóór invoerrecht); $95–117k in augustus	Hoge concurrency op 120–235B MoE dat in HBM past; slechts enkele gebruikers op 433–743B-modellen	Nee (5A992.z, TPP 58,484)
Huawei Atlas 300I Duo	96 GB LPDDR4X, 408 GB/s	RMB 13,300 van een JD-handelaar	Batchreview in laag volume; weinig ondersteunde modellen	Ja (binnenlands)
Huawei Atlas 800I A2 (8 × Ascend 910B 64 GB)	8 × 64 GB HBM	Niet verkregen; volledige DeepSeek vereist er twee op W8A8, vier op BF16	Fabrieksbrede multi-user Q&A en analyse	Ja (binnenlands)
RTX 5090 D V2 (China-model)	24 GB	RMB 16,499 bij lancering (aug 2025), sindsdien gestegen met geheugenprijzen	8B vision-modellen, kleine VLA's	Ja

Bronnen: (NVIDIA China Thor, NVIDIA Thor-lancering, CNX over prijsstijgingen, NVIDIA exportclassificatiegegevens, RS China, Implicator, IThome over ASUS GX10, Tom's Hardware RTX PRO 6000, GPU Poet, Apple-specificaties, TechRadar, IThome over M5 Ultra China-prijzen, Wccftech, StorageReview, CyberPowerPC, monitors.com, Tom's Hardware DGX Station, vLLM-Ascend-ondersteuningsmatrix, China Telecom Cloud, Huawei Atlas 800I A2, TweakTown). Vasteland-prijzen voor de 512 GB Mac Studio zijn geëxtrapoleerd uit Amerikaanse configuratiepremies en zijn laagbetrouwbaar estimate.

FIG 08 **On-prem rekenkracht: geheugen, prijs en wat legaal te koop is op het Chinese vasteland** ***Exportclassificaties bepalen de shortlist eerder dan benchmarks.***



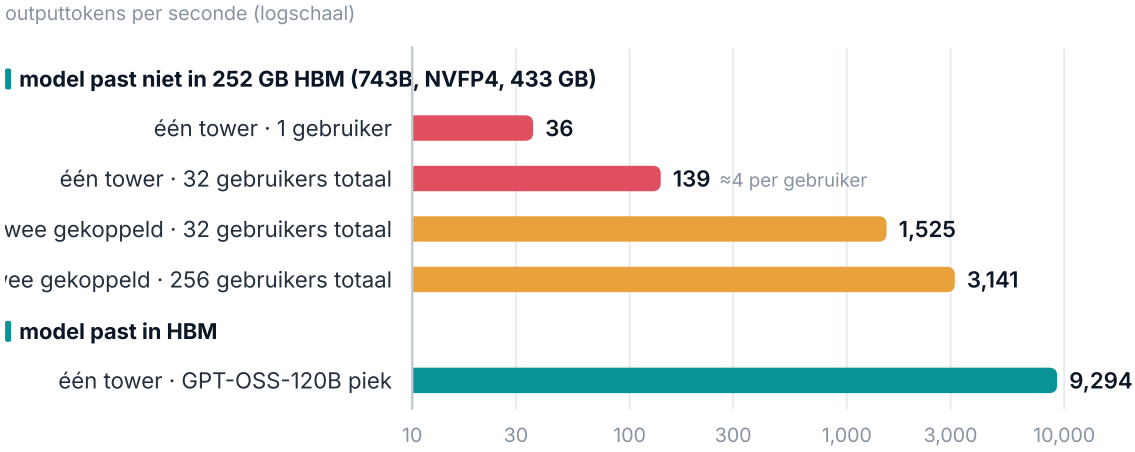
Bron: NVIDIA-exportclassificaties, Apple, RS China, JD.com en reviewprijzen (november 2026).

Jetson Thor maakt station-niveau-implementatie werkbaar. De TensorRT Edge-LLM-tests van NVIDIA tonen Qwen3-VL-8B (NVFP4) dat 2,048 tokens prefilt in 104.5 ms en decodeert met 44.4 tokens/s op Thor (TensorRT Edge-LLM) vendor claim; voor een AOI-uitsnede en een antwoord van 20 tokens is dat ongeveer 0.5 s met het 8B-model en 0.2 s met het 2B-model, of 8.5–21 uitsnedes per seconde gebatcht estimate. Uitvoerlengte, niet het aantal parameters, bepaalt de latentie, dus constrained decoding naar een klassetoken of korte JSON hoort bij het ontwerp. GR00T N1.7 (3B) draait op 25.1 Hz op Thor (Jetson AI Lab) vendor claim. Een Orin NX doet ongeveer 1.8 s per uitsnede en een Atlas 300I Duo heeft een p95 boven 23 s, alleen geschikt voor batchwerk (CSDN) single source. Op het vasteland moet je NVIDIA-producten op 40–60% boven de Amerikaanse adviesprijs begroten estimate.

Een 512 GB Mac Studio maakt een volledig model op één machine mogelijk, voor een paar gebruikers. De M5 Ultra (augustus 2026) biedt tot 512 GB op 1.2 TB/s vanaf ongeveer $22,299 (Apple, TechRadar); DeepSeek 671B op 4-bit (≈380–404 GB) decodeert met 17–20 tokens/s op een M3 Ultra (Pinggy) verified en naar schatting 25–29 op de M5 Ultra. Lange prompts zijn het zwakke punt: een retrievalprompt van 10,000 tokens duurt ruwweg 17–70 s estimate, boven de interactielimiet van tien seconden.

DGX Station is de sterkste losse machine, maar niet beschikbaar op het vasteland en vijf tot zes keer de prijs van een Mac. De geleverde GB300 heeft **748 GB**, waarvan slechts 252 GB snel HBM is (Wccftech, Tom's Hardware, aug 2026) verified. Onafhankelijke tests tonen ongeveer 9,300 output-tokens/s op GPT-OSS-120B, dat in HBM past, maar een 743B-model dat naar LPDDR5X lekt, gaf 32 gebruikers slechts 139 tokens/s in totaal, ongeveer 4 elk; twee gekoppelde eenheden haalden 1,525 (StorageReview, aug 2026, StorageReview, okt 2026) single source. Straatprijzen stegen tot $123–140k in november (CyberPowerPC, monitors.com). Op het vasteland zijn de realistische opties voor dezelfde taak de Mac Studio, CPU+GPU-servers in KTransformers-stijl (KTransformers) vendor-test of binnenlandse Ascend-servers.

FIG 09 **Eén groot workstation maakt geen gedeelde assistent** ***Wanneer het model niet in hoogbandbreedtegeheugen past, krijgen 32 gelijktijdige gebruikers elk ongeveer 4 tokens per seconde.***



Bron: StorageReview (aug en okt 2026); ServeTheHome. Onafhankelijke tests uit één bron.

Geheugentekorten bewogen dezelfde onderdelen 20–100% in een jaar (DGX Spark ging van $3,999 naar $6,950 (TechRadar)), dus apparatuuroffertes moeten worden geïndexeerd in plaats van vastgezet.

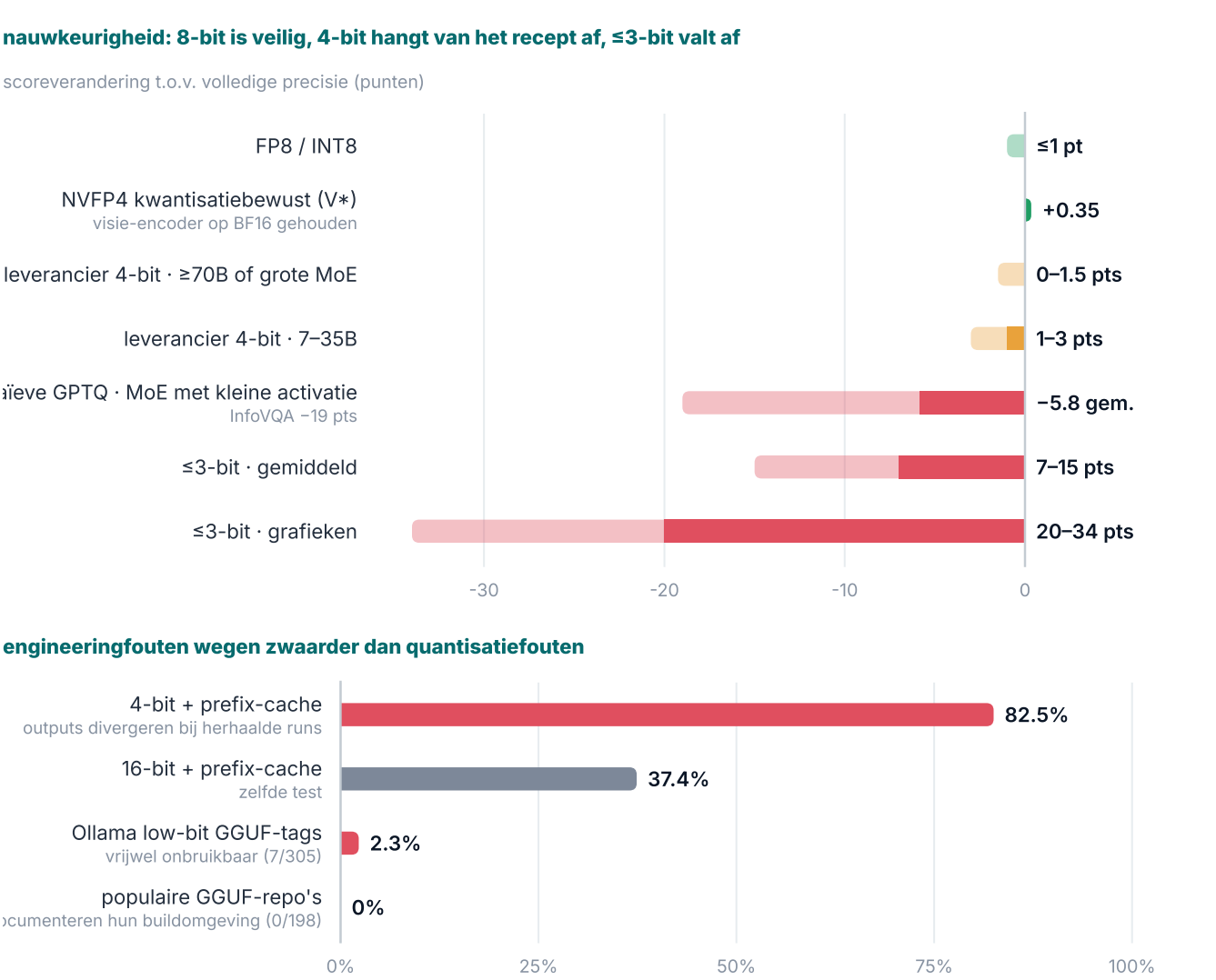
Binnenlandse accelerators draaien volledige modellen; softwareonderhoud is het knelpunt. Ascend 910B heeft 64 GB HBM, maar vLLM-Ascend markeert DeepSeek V3/R1 als "niet langer onderhouden" en ondersteunt slechts een paar modellen op kleinere kaarten (vLLM-Ascend) single source, primair, en elk nieuw model kost één tot twee maanden om aan te passen (Caijing/Anue); lever een bevroren, gevalideerde modelversie. Het serveren van volledige DeepSeek 671B aan 20–100 gebruikers vereist ten minste twee acht-kaart Atlas 800I A2-servers (China Telecom Cloud) verified, toch bedroegen de openbare aanbestedingen voor LLM-apparaten in China in januari–september 2025 slechts RMB 27.5 miljoen (C114) single source: fabrieken betalen voor integratie en dienstverlening, niet voor een box met rekenkracht.

Kwantisering: 8-bit is veilig, 4-bit hangt af van het recept, 3-bit en lager is uitgesloten

8-bit (FP8 of INT8) is in wezen verliesvrij voor vision-language-modellen (Red Hat, Hugging Face) verified. Op 4-bit hangt het verlies af van grootte en recept: de 72B AWQ-build van Qwen verliest minder dan een punt op MMMU terwijl de 7B 2.8 verliest (Qwen AWQ) verified, en detailzware taken zakken verder (naïef 4-bit GPTQ kost Qwen3-VL-30B-A3B 19 punten op InfoVQA (VEQ)). Op 3-bit en lager is het gemiddelde verlies 7–15 punten en verliezen grafie-

ken 20–34 (VLMQ) verified; boven 64k tokens verliest 4-bit tot 59%, erger op niet-Engelse invoer (arXiv 2505.20276). Houd de vision-encoder op FP16/BF16.

FIG 10 **Quantisatie: 8-bit is veilig, 4-bit hangt van het recept af, 3-bit en lager valt af** ***Engineeringfouten in de serving-stack kosten meer nauwkeurigheid dan de quantisatie zelf.***



Bron: NVIDIA Nemotron-modelkaart, VLMQ, VEQ, arXiv 2609.05881, 2610.19498, 2609.04748 en anderen.

Twee waarschuwingen wegen zwaarder. Ten eerste heeft **per 11 november 2026 geen studie of modelkaart kwantisering op PCB/SMT-defectbeelden gemeten**, dus valideer opnieuw op de golden set. Ten tweede zijn fouten in de serving-stack vaak groter dan kwantiseringsfouten: hetzelfde model scoorde ongeveer 30 punten lager onder Ollama dan onder de referentie-implementatie (SiliconBench); 7 van Ollama's 305 officiële codemodeltags waren in wezen kapot hoewel de bestanden intact waren (arXiv 2609.05881); geen van 198 populaire GGUF-repository's noteerde zijn build (arXiv 2610.19498); en bij hergebruik van prefix-cache divergeerde 82.5% van de episodes op 4-bit tegen 37.4% op 16-bit (arXiv 2609.04748) single sources. Quantization-aware 4-bit van de leverancier is bijna verliesvrij op detailtaken (NVIDIA Nemotron 3.5) vendor claim, maar 4-bit VLA's verliezen 1.5–4.3 punten onder verstoring en draaien langzamer op Thor (LoopQ) single source. Beheer een gekwantiseerd model als een nieuwe meter: pin gewicht-hashes, noteer de build, reset prefix-caching en valideer opnieuw na elke wijziging van precisie, engine of template.

Precisie	≥70B dense of ≥100B MoE	7–35B	≤4B of detailzware taken	Oordeel
FP8 / INT8	Veilig (≤1 punt)	Veilig	Veilig; de standaard voor inspectie	Gebruiken
4-bit AWQ/GPTQ (leveranciersrecept, vision-encoder FP16)	0–1.5 punten	1–3 punten; naïef GPTQ op small-active MoE −5.8 gemiddeld, −19 op InfoVQA	2–3 punten; valideer detailtaken eerst	Voorwaardelijk
NVFP4 (leveranciersrecept)	≈0.3–1 punt onder FP8 (leveranciersdata)	Quantization-aware builds bijna verliesvrij op V*, OCRBench	Geen publieke data	Voorwaardelijk (geef de voorkeur aan quantization-aware builds)
GGUF Q4_K_M	Ongetest	Hoge variantie op domeintaken	Niet voor kleine tekst of defecten	Niet voor detailbeoordeling
GGUF Q5_K_M en hoger	Vermoedelijk veilig	Bijna verliesvrij in twee tests	De ondergrens voor detail op llama.cpp	Gebruiken
≤3-bit (inclusief 1T-modellen geperst tot 1.8–2 bits op één machine)	Niet voor beslissingen	Nee	Nee	Nee

Welke configuratie voor welke taak

Configuratie	Inline-inspectie	AI-review van valse alarmen	Robotplanning en -besturing	Q&A en analyse
Toegewijde detectors of anomalie-modellen (YOLO, RF-DETR, PatchCore-klasse), INT8/FP16, elke NPU- of GPU-edgebox	● Ja: de hoofdbeoordelaar, 2–12 ms per uitsnede	● Ja (classificator afgestemd op de lijn)	● Alleen perceptie	—
2–4B vision-modellen, NVFP4/AWQ/FP8 met FP16 vision-encoder, op Thor/T4000 of DGX Spark	● Voorwaardelijk: hercontroleer AOI-gemarkeerde uitsnedes op 0.2–0.3 s elk	● Voorwaardelijk, onder 1 s na afstemming	● VLA-backbone (2–3B VLA op ≈40–130 ms per chunk)	● Zwak
7–9B vision-modellen (Qwen3-VL-8B, MiniCPM-V 4.5, GLM-4.6V-Flash), FP8 of 4-bit van de leverancier, op Thor, Spark of een kaart van 24–32 GB	● Voorwaardelijk: ≈0.5 s per gemarkeerde uitsnede	● Ja: 0.5–1 s per oproep, beste waarde	● Planning en vaardigheidskeuze (1–3 s), niet in de regelkring	● Basis-SOP-Q&A voor één of enkele gebruikers
26–35B MoE (Gemma-4-26B-A4B, Qwen3.6-35B-A3B), NVFP4/FP8	● Nee (te zwaar)	● Ja, als uitlegger of voor moeilijke gevallen	● Planning	● Ja (8 gebruikers op een Spark, ≈0.1 s eerste token)
106–241B MoE op 4-bit op een Mac Studio, RTX PRO 6000 of Spark	● Nee	● Nee	● Offline planning	● Ja, voor één of enkele gebruikers

Configuratie	Inline-inspectie	AI-review van valse alarmen	Robotplanning en -besturing	Q&A en analyse
DeepSeek 671B of Kimi K2.x 1T op 4-bit op een 512 GB Mac, KTransformers-server of DGX Station	Nee	Nee	Alleen offline programmageneratie	Voorwaardelijk: 1–3 gebruikers en korte prompts; lange prompts duren 16–170 s
Alles op ≤3-bit; GGUF Q4_K_M voor detailwerk met kleine modellen; Ollama voor veel gebruikers	Nee	Nee	Nee	Nee

Bronnen: de TensorRT Edge-LLM-, Jetson AI Lab-, Apple ML Research- en KTransformers-data hierboven, en (DGX Spark-tests, RTX PRO 6000-tests, Hardware Corner over Atlas, Ultralytics YOLO26, RF-DETR) **gemeenschaps- en vendor-tests, meestal single source** .

Dus, op de twee gebruikelijke zorgen: **het hoeft niet langzaam te zijn** als het model bij de taak past (kleine en middelgrote vision-modellen op Thor of Spark halen elk inline- en reviewbudget; 671B–1T-modellen blijven van de lijn), en **kwantisering hoeft het niet verkeerd te maken** met FP8/INT8 standaard, of 4-bit van de leverancier boven 7B gevalideerd op de golden set. De grotere risico's zijn de ruwweg 57-punts zero-shot-industrienauwkeurigheid van algemene modellen en fouten in de serving-stack. Geef inspectieverzoeken vaste batchgroottes en een eigen wachtrij (temperatuur 0 is niet bit-reproduceerbaar onder batching (vLLM-documentatie)), houd een N+1 hot standby, en maak elke stap op de lijn fail open.

Leraar en leerling

03 Leraar en leerling: frontier-modellen kunnen inspectiedata pre-labelen; voor robotarmen zijn de echte leraren nog altijd mensen

Leraarkandidaten november 2026: generieke visie verdubbelde, scores voor industriedefecten ontbreken

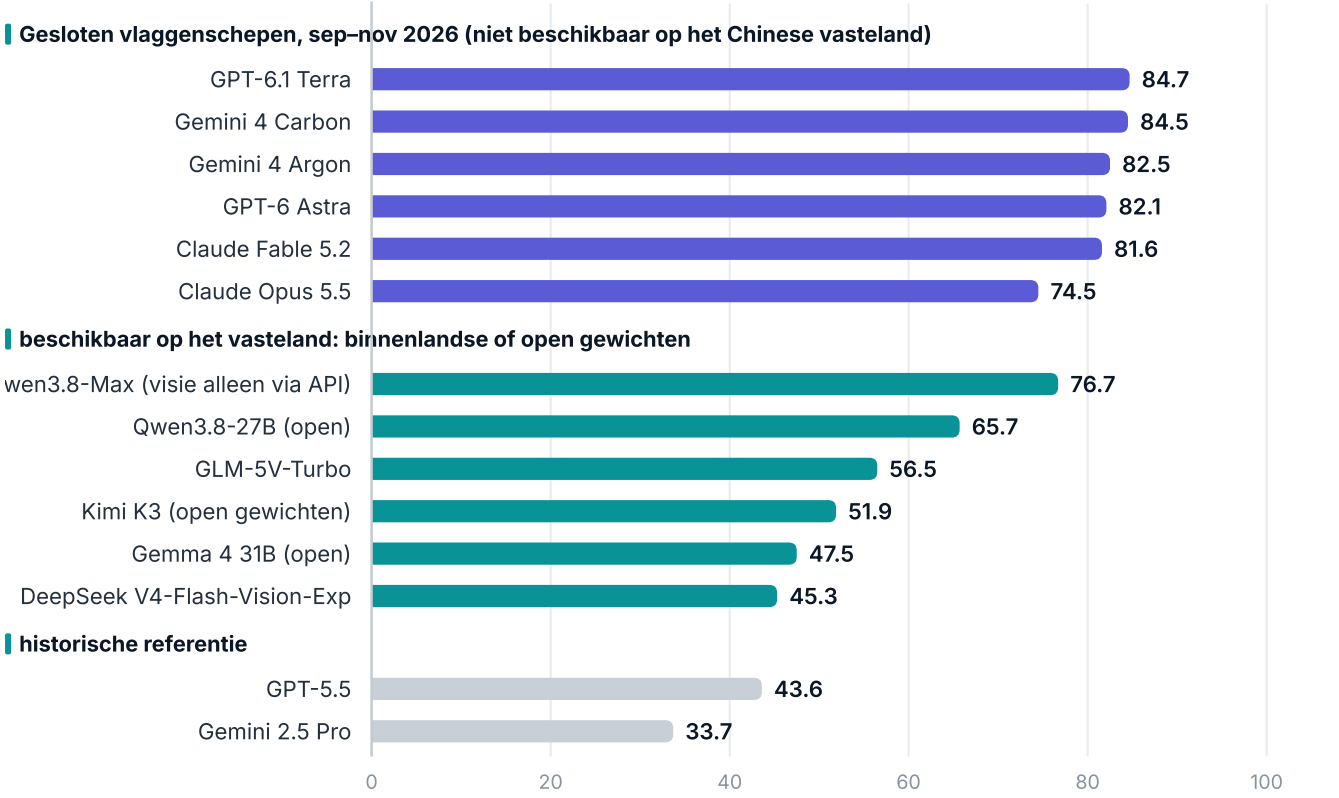
Elk gesloten vlaggenschip dat in november 2026 als leraar zou kunnen fungeren, accepteert beelden: Claude Fable 5.2 (algemeen beschikbaar 9 november, $8/$40 per miljoen tokens) en Opus 5.5 ($4/$20) (Anthropic, Claude-modellenoverzicht) verified; GPT-6 Astra ($10/$50) en GPT-6.1 Terra ($1/$5) (OpenAI-prijzen); en Gemini 4 Argon en Carbon, in fasen geopend (Google Carbon). Bruikbaar in vasteland-China zijn Qwen3.8-Max (vision via API) en open modellen zoals Qwen3.8-27B, Kimi K3, GLM-5V-Turbo, DeepSeek V4-Flash-Vision-Exp en Gemma 4 (Roboflow-blog, Hugging Face).

De kernkloof moet helder worden benoemd: geen van deze modellen heeft een gepubliceerde score op MMAD, MVTec AD 2, VisA, Real-IAD, UniPCB of welke PCB/SMT-dataset dan ook. MMAD heeft nog steeds geen leaderboard (MMAD GitHub), en de twee nieuwste artikelen over industriële anomalieën (augustus–september 2026) gebruiken nog steeds Gemini 2.5 Pro en Claude Sonnet 4 als hun nieuwste algemene baselines (InspectorGPT, Anomaly-LR). Een full-text-zoekopdracht op arXiv op 11 november vond nul hits voor “Fable 5.2” of “Gemini 4 Carbon” in deze context (arXiv API) single source, afwezigheid van bewijs. De dichtstbijzijnde vervangers:

- Het nieuwste frontier-model met een MMAD-score is GPT-5.5: 69.95% op “defect of niet” en 76.20% gemiddeld over zeven taken op een subset van 500 beelden, tegen 95.24% en 86.65% voor menselijke experts (AnomalyClaw) single source; GPT-4o scoorde 68.63% en 74.92% in 2024 (MMAD, ICLR 2025), ongeveer 1.3 punten vooruitgang in twee jaar.
- Op LDU-Bench, echte SEM-reviewbeelden van IC-lithografie en de enige elektronica-benchmark met 2026-modellen, is goed-of-slecht-triage redelijk (macro F1 0.932 voor GPT-5.4, 0.884 voor GLM-5V-Turbo, 0.876 voor Claude Opus 4.6, 0.504 voor Gemma 4 31B), maar het classificeren van 11 defectmorfologieën haalt slechts 0.20–0.41 (LDU-Bench) single source.
- Op de onafhankelijke vision-evals van Roboflow (bijgewerkt 10 november 2026) is objectdetectie-mAP@50 84.5 voor Gemini 4 Carbon, 84.7 voor GPT-6.1 Terra en 81.6 voor Fable 5.2, tegen 33.7, 43.6 en 38.6 voor Gemini 2.5 Pro, GPT-5.5 en Claude Opus 4.8 (Roboflow Vision Evals) single source, onafhankelijke beoordelaar, private datasets.

FIG 11 **Leraarkandidaten: generieke scores verdubbelden in een generatie** ***Maar geen van de vlaggenschepen van 2026 heeft een gepubliceerde score op een industriële defectbenchmark.***

Roboflow generieke detectie mAP@50



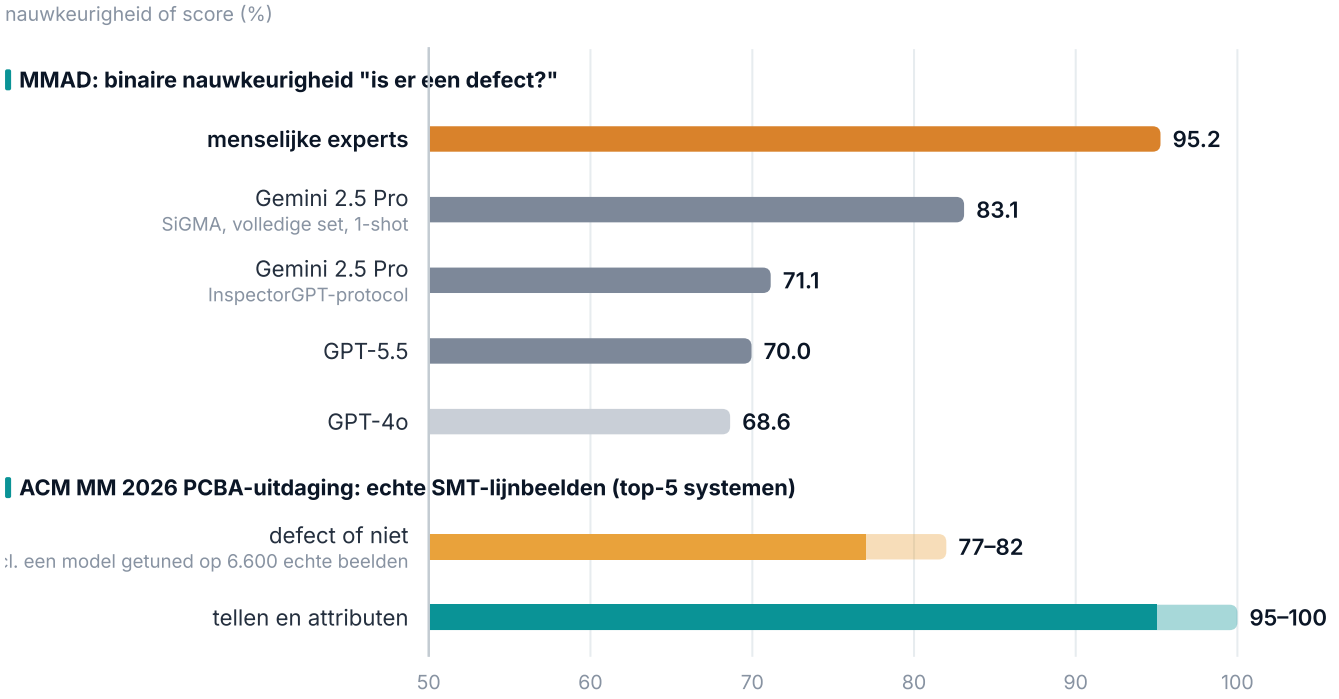
industriële defectbenchmarks (MMAD, MVTec AD 2, VisA, Real-IAD, UniPCB): nul gepubliceerde scores voor deze vlaggensche
generieke detectiescores verdubbelden in een generatie, maar dat is geen defectoordeel; test op je eigen gouden set

Bron: Roboflow Vision Evals; releasenotes van modellen; MMAD, MVTec AD 2 en andere benchmarkpagina's (per 11 nov 2026).

Leraarkandidaat (nov 2026)	Prijs per M tokens (in/uit)	Roboflow-detectie mAP@50 / tellen	Benchmarks voor industriedefecten	Bruikbaar in vasteland-China
Claude Fable 5.2	$8 / $40	81.6 / 86.5	Geen	Nee
Claude Opus 5.5	$4 / $20	74.5 / 80.6	Geen	Nee
GPT-6 Astra	$10 / $50	82.1 / 80.2	Geen	Nee
GPT-6.1 Terra	$1 / $5	84.7 / 80.2	Geen	Nee
Gemini 4 Carbon	$2 / $10 (promotie)	84.5 / 90.1	Geen	Nee
Qwen3.8-Max (vision alleen via API)	$2 / $6	76.7 / 81.1	Geen	Ja
Qwen3.8-27B (open)	Lokaal	65.7 / 64.9	Geen	Ja
GLM-5V-Turbo	$1.20 / $4.00	56.5 / 48.6	LDU-Bench-triage 0.884	Ja
Gemma 4 31B (open)	Lokaal	47.5 / 51.4	LDU-Bench-triage 0.504	Ja (lokale gewichten)
Referentie: GPT-5.5 / GPT-4o	—	43.6 / 68.0 (GPT-5.5)	MMAD binair 69.95% / 68.63%	Nee
Menselijke experts	—	—	MMAD binair 95.24%	—

Nieuwere artikelen testen nog de vorige generatie, maar wijzen dezelfde kant op. Op de volledige MMAD-set scoren menselijke experts 95.2% op defect-of-niet en Gemini 2.5 Pro 83.1% (SiGMA), maar hetzelfde model scoorde in een ander artikel slechts 71.14% (InspectorGPT): het testprotocol beweegt de score evenveel als een modelgeneratie. In de CVPR 2026 VAND 4.0-uitdaging piekten kant-en-klare vision-language-modellen op 61.88 AP tegen 90.84 voor gesuperviseerde methoden, en de organisatoren concludeerden dat ze “momenteel geen fijn afgestemde classificatoren kunnen vervangen” (VAND 4.0) single source. In de ACM MM 2026 PCBA-uitdaging scoorden de top vijf 76.92–82.47 op defect-of-niet terwijl componenttype en pincount 95–99.5 haalden (arXiv 2609.21276) single source: herkennen is veel makkelijker dan beoordelen.

FIG 12 **"Is er een defect?" blijft moeilijk voor algemene modellen** ***Het beste algemene model loopt meer dan 12 punten achter op menselijke experts op MMAD; getunede systemen doen het beter op echte lijnbeelden.***



Bron: MMAD; SiGMA; InspectorGPT; ACM MM 2026 PCBA-uitdaging.

Drie conclusies volgen. Zien en tellen verdubbelden in een generatie, maar het beoordelen van afwijkingen beweegt langzaam, hangt sterk af van het protocol en kent systematische vertekeningen; hoeveel de nieuwe modellen winnen, kan alleen op de eigen beelden van een fabriek worden gemeten. Binnen hun trainingsdistributie verslaan fijn afgestemde 7B-specialisten algemene modellen nog altijd (Anomaly-LR, UniPCB). En echte lijndata doen scores instorten: op geassembleerde elektronische panelen viel de AUROC op beeldniveau van 97–98 op MVTec AD naar 31.5–78.6 (arXiv 2606.01992) single source. Voor vasteland-bedrijven ligt de sterkste beschikbare leraar, Qwen3.8-Max, ongeveer 8 punten onder de wereldwijde frontier.

Tientallen of honderden rondes draaien: geld is niet het probleem, labelkwaliteit wel

Geld is niet het probleem. Bij de adviesprijzen van november 2026 kosten 1,000 beelden van één megapixel ongeveer $3 (GPT-6.1 Terra) tot $32 (GPT-6 Astra) per ronde, en honderd rondes over 100,000 beelden ongeveer $32,000–135,000 vóór batchkortingen estimate (Claude vision-documentatie, OpenAI vision-gids, Roboflow Vision Evals). Voor eenvoudige labels kost een persoon ongeveer evenveel als Terra; het voordeel van de leraar is snelheid en schaal, niet prijs. De NDA’s van klanten houden productbeelden doorgaans binnen de fabriek, dus de golden set wordt ter plaatse gelabeld.

Het echte obstakel is kwaliteit, en meer rondes repareren systematische fouten niet. Roboflow draait elke taak drie keer en scores bewegen slechts ±0.1–1.4 punten (Roboflow), dus een model herhaalt meestal zijn fouten en meerderheidsstemming kan het niet redden. Rondes helpen alleen wanneer elke ronde informatie toevoegt: één of twee goede referentiebeelden voegen ongeveer 3–4 punten toe, meer voegt niets toe (AD-Copilot); een model tegen een goede referentie laten argumenteren over meerdere beurten voegt 6.23 punten cross-domein-AUROC toe voor GPT-5.5 en 2.95 op MMAD, met een piek na ongeveer drie oproepen (AnomalyClaw) single sources. “Een leerling getraind op leraarlabels benadert menselijke labels” is alleen op natuurlijke beelden aangetoond: een ResNet-101 getraind op Qwen3-VL-labels van COCO en Objects365 bereikte 95.7–99.4% van het menselijke-labelresultaat, en de auteurs zeggen dat domeindata ongetest is (arXiv 2602.20972) single source.

FIG 13 **Twee leraar–leerling-pipelines** ***Voor inspectie pre-labelt een frontiermodel en beslissen mensen; voor robotarmen zijn de echte leraren nog steeds menselijke demonstraties.***



Bron: Synthese van de auteur (hoofdstuk 3); kosten en steekproefgroottes zijn schattingen.

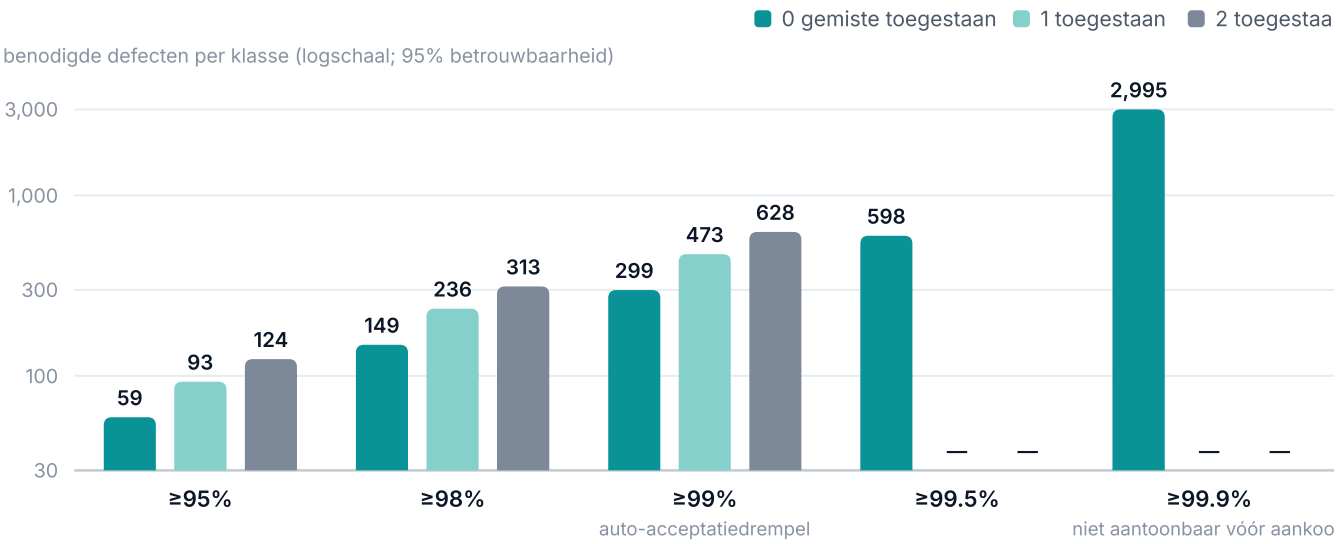
Een werkbaar protocol estimate:

1. **Verzamel de golden set uit het eigen AOI-archief van de fabriek en bewaarde NG-platen**, gestratificeerd in 12 echte defectklassen (ontbrekend, offset, tombstone, polariteit, verkeerd onderdeel, brug/kortsluiting, onvoldoende soldeer, overmatig soldeer, soldeerbal, losgelaten lead/open, beschadigd onderdeel, vreemd materiaal) en vier oorzaken van valse oproepen (schaduw, flux, zeefdruk, tolerantie), plus een “driftslice” van ten minste 300 defecten van een tweede lijn of andere belichting, plaatkleur of artikelnummers.
2. **Label blind met twee of drie IPC-A-610-getrainde inspecteurs**, met een senior engineer die beslist en röntgeninspectie, elektrische test of dwarsdoorsneden die dubbelzinnige monsters beslechten. Accepteer een klassedefinitie alleen als de overeenstemming tussen inspecteurs $\kappa \geq 0.75$ bereikt (≥ 0.90 voor kritieke klassen); de menselijke overeenstemming over 11 elektronische defectmorfologieën in LDU-Bench was slechts 0.727, dus sommige klassen moeten worden samengevoegd of opnieuw gedefinieerd. Attribuutovereenstemming in AIAG-stijl (effectiviteit ≥90%, misser ≤2%, vals alarm ≤5%, $\kappa > 0.75$; >0.90 voor release) is een verstandig minimum (Fabrico, iFactory).
3. **Bepaal de omvang met exacte binominale (Clopper–Pearson) eenzijdige 95%-grenzen** (binominale betrouwbaarheidsintervallen, rule of three:

Te bewijzen recall	0 missers toegestaan	1 toegestaan	2 toegestaan
≥95%	59 defecten	93	124

Te bewijzen recall	0 missers toegestaan	1 toegestaan	2 toegestaan
≥98%	149	236	313
≥99%	299	473	628
≥99.5%	598	—	—
≥99.9%	2,995	—	—

FIG 14 **Hoe groot de gouden set moet zijn*99% recall bewijzen met nul gemist vereist 299 defecten per klasse; 99.9% kan niet vóór aankoop worden bewezen.***



Bron: Exacte binomiale betrouwbaarheidsintervallen (95%).

Zestig defecten per klasse zijn genoeg voor een pass/fail-scherm; 300 zijn nodig om over automatische acceptatie te beslissen. De volledige set, ongeveer 3,600 defecten plus 3,500 valse-oproep-uitsnedes, kost ongeveer 37–57 arbeidsuren om te labelen en te beoordelen, en het draaien van een leraarkandidaat erover kost $23–196; de echte kost is twee tot zes weken verzamelen, bepaald door de zeldzaamste klasse estimate. Stuur uitsnedes van losse componenten van de AOI, geen hele platen, met het doelwit op ten minste 42–48 pixels breed (arXiv 2610.09313).

1. **Accepteer leraarlabels alleen automatisch voor een klasse met nul missers op ≥299 golden defects**, model-waarheid-overeenstemming niet slechter dan mens-mens-overeenstemming, en ten minste 50% van de valse oproepen opgeruimd. Kritieke klassen zoals kortsluitingen, ontbrekende onderdelen, polariteit en verkeerde onderdelen vereisen 99.5–99.9% (598–2,995 defecten); anders gaat alles wat de AI niet met hoge betrouwbaarheid kan opruimen naar een persoon. Een doorglipperpercentage onder 0.1% van de defecten kan niet vóór aankoop worden bewezen (2,995 defecten met nul missers; 0.01% zou 29,956 vergen), dus de lijn houdt menselijke steekproeven van AI-opgeruimde oproepen, downstream-ICT/FCT als vangnet en SPC op doorglippers. Geverifieerde labels worden vervolgens gedestilleerd in een conventionele detector of classificator en opnieuw gevalideerd op elk nieuw product.

Ook voorwaarden en dataregels bepalen de leraar. Anthropic verbiedt het gebruik van zijn diensten om concurrerende modellen te trainen (Anthropic commerciële voorwaarden); OpenAI staat toe dat outputs in-house classificatoren trainen die niet aan derden worden verkocht (OpenAI zakelijke voorwaarden); DeepSeek staat destillatie expliciet toe (DeepSeek-voorwaarden) [primaire voorwaarden]. Geen enkele definieert een “concurrerend model”, dus commercieel gebruik vereist juridische toetsing, en de dataregels van klanten houden beelden doorgaans ter plaatse. De veiligste standaard is een lokaal draaiende open-gewicht-leraar, met cloud-frontier-modellen alleen wanneer de klant dat schriftelijk toestaat.

Inspectieleerlingen: klein en snel; de knelpunten zijn optiek, data en validatie

De leerling is niet het knelpunt: YOLO26n doet 3.80 ms per beeld in INT8 op een Orin Nano Super van $249–399 (Ultralytics) vendor-test, NVIDIA meldt 98.51% op drie PCB-defectklassen na het aanpassen van een foundation-model op ongeveer 600 beelden (NVIDIA) vendor claim, en synthetische defecten voegen slechts 0.9–2.5 mAP-punten toe (arXiv 2605.04635). De beperkingen zijn optiek, lijndata, validatie en licenties: Ultralytics YOLO is AGPL-3.0 of enterprise-gelicentieerd (Ultralytics), dus gesloten producten hebben een licentie nodig of een Apache-2.0-detector zoals RF-DETR.

Robotleerlingen: gestructureerde stations zijn nabij productiekwaliteit; flexibel en dynamisch werk niet

In november 2026 zijn er drie soorten robotleraar, en slechts één kan een lokale leerling worden: open, fijn afstelbare VLA's (π0/π0.5, GR00T N1.7, X-VLA, RDT2, SmolVLA). Gesloten robotmodellen (π0.6/π0.7 via een bèta-API, Gemini Robotics 2 voor vertrouwde testers) kunnen niet ter plaatse worden ingezet (The AI Insider, DeepMind). Frontier-algemene modellen zijn langzaam en slecht in fijne manipulatie: in de test van Artificial Analysis scoorden ze 46–69 met 28–45% grijpkans en ongeveer 30 minuten per taak, "veel te traag en te duur voor productiegebruik" (AA-Robotics); GPT-6 Astra haalde een precisieslot slechts 2 keer in 20 (RoboCurve); en het direct aansturen van gewrichten geeft 0–5.5% op LIBERO op 0.2–0.4 Hz tegen ongeveer 83 Hz die nodig is (Anthropic) verified over drie bronnen. De 64.9% van Fable 5.2 op de gesimuleerde RoboDojo-benchmark is een interne run (Anthropic) vendor claim.

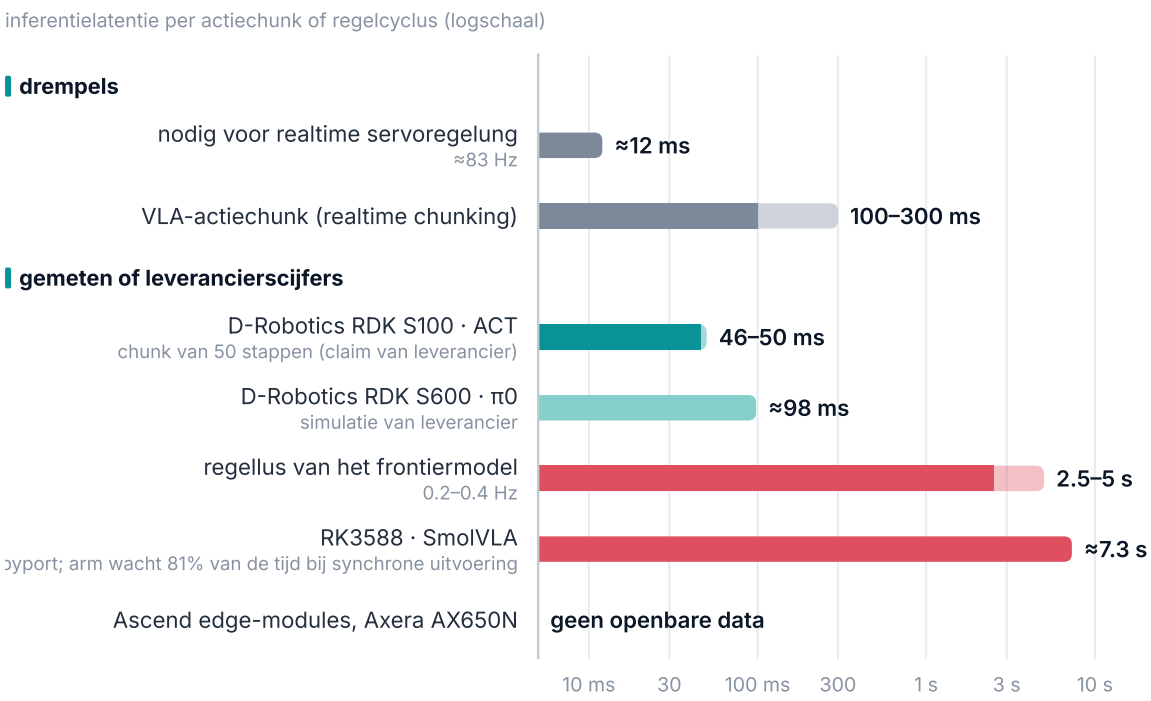
Studies van derden sinds oktober zijn het eens: ongeveer 33–44% op LIBERO-10 voor de drie leidende agents (arXiv 2610.30051); agents die ten onrechte succes claimen op 24–39% van de subtaken (arXiv 2611.08209); en 104 van 108 quasistatische taken op echte robots voor Opus 5.5, maar op 280 seconden elk en zonder insertie of assemblage (AMP1) single sources. Frontier-modellen horen offline: programma's schrijven, taken orkestreren en succescriteria opstellen.

Leerlingkwaliteit heeft een duidelijke rand: goed binnen de trainingsdistributie, instortend daarbuiten. π0.7 "overschrijdt vaak 90%" op geziene taken en haalt 60–80% op ongeziene (π0.7-artikel) vendor-onderzoek; ContactDP haalt 97.8% op getrainde connectoren maar 28.6–66.7% op ongeziene (ContactDP); lichte camera- of posewijzigingen laten 95% onder 30% zakken in LIBERO-Plus (LIBERO-Plus, CVPR 2026); geleerd grijpen van bewegende transportbanden piekt op 58.89% (ForeTime-VLA) single sources, een taak die klassieke visietracking al jaren in productie doet.

Het nabij-productiebewijs komt allemaal van gestructureerde stations, menselijke demonstraties en correctie op echte robots. Neuromeka hield een state machine en vooraf aangeleerde beweging en leerde alleen grijpen, kabelinsertie en solderen, elk uit minder dan 20 minuten data: 322 van 324 in productie, 12.8% langzamer dan een persoon (Neuromeka) single source, vendor. Fraunhofer haalde 92.8% over vier connectoren uit 68 minuten demonstraties (Fraunhofer); HIL-SERL van Berkeley slaagde 100 van de 100 keer op RAM-, SSD- en USB-insertie na 1–2.5 uur RL op echte robots (HIL-SERL); AgiBot draaide 18–20 seconden per stap binnen een takt van 20 seconden op een tablettestlijn (IThome) vendor claim.

Destillatie koopt snelheid, geen generalisatie: π0.5 draait op 0.70 Hz op Jetson Orin raw en 10.30 Hz na asynchrone scheduling (EagleVLA); laagdestillatie brengt het van ongeveer 3 Hz naar ongeveer 10 Hz (Shallow-π) verified. Een leerling erft alleen wat de leraar dekte; elk nieuw product vereist nog steeds taakdata.

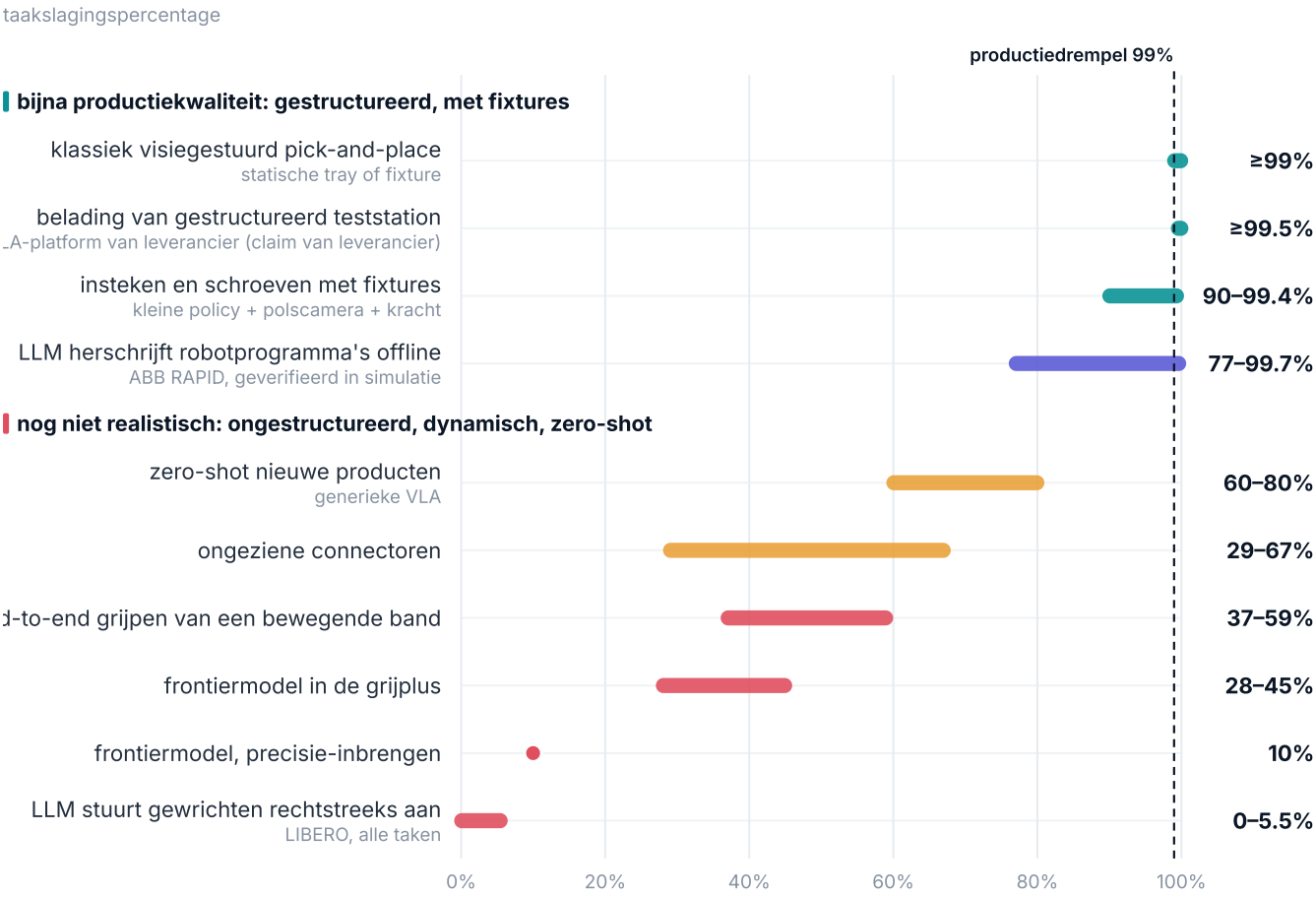
FIG 15 **Edge-policylatentie***Een kleine ACT-policy draait in ongeveer 50 ms op een edge-board; een VLA geport naar een hobbychip doet er 7.3 s over.*



Bron: D-Robotics-documentatie en -forum (leverancierscijfers); community-ports op GitHub; onderzoek van Anthropic.

Op Chinese edge-chips is het enige gemeten geleerde beleid een ACT-beleid op de RDK S100 van D-Robotics op 46–50 ms per chunk van 50 stappen, zonder gepubliceerd slagingspercentage (D-Robotics RDK S-documentatie) vendor claim; een hobbyport van SmolVLA op een RK3588 doet ongeveer 7.3 s per chunk (GitHub). Kwantisering koopt snelheid niet gratis: geen 4-bit π0.5-build was sneller dan 16-bit bij batchgrootte 1 (LoopQ). Demonstraties zijn duur (Reuters meldde ongeveer 300 pogingen per bruikbare beweging voor een beginnende teleoperator (Notebookcheck)), en validatie kost meer: het bewijzen van een faalkans ≤0.1% bij 95% betrouwbaarheid vergt ongeveer 3,000 opeenvolgende successen, 17–22 roboturen estimate. Geen geleerd manipulatiebeleid heeft een veiligheidscertificering.

FIG 16 **Robotleerlingen: bijna productiekwaliteit alleen als het station gestructureerd is** ***Stations met fixtures halen 90–99%+; ongestructureerde, dynamische en zero-shot taken zijn nog niet realistisch.***



Bron: Papers en leveranciersmateriaal uit hoofdstuk 3; sommige cijfers zijn claims van leveranciers.

Elektronicataak	Werkbare leerling	Verwacht slagingspercentage	Snelheid vs een persoon	Kost per ombouw	Oordeel 2026–2028
High-mix pick-and-place, kitting, machinebediening (statische tray of opstelling)	Geleerde perceptie plus geprogrammeerde of geplande beweging	≥99% met klassieke visiegeleiding	Idem aan conventionele stations	Uren	Nu haalbaar
Gefixeerde connector-, FFC- en kabelinsertie, schroeven, solderen	State machine + klein BC/diffusion/ACT-beleid + polscamera + kracht/koppel (idealiter tast) + menselijke correctie	90–99.4%	10–15% langzamer	10–300 demo's (≈15–70 min) + 1,000–3,000 validatieruns (6–22 roboturen)	Haalbaar, R&D-zwaar maanden voor het eerste station
Gestructureerd teststation-bedienen en OK/NG-sorteren	VLA-platform van leverancier + digitale tweeling + RL op echte robots	≥99.5% vendor claim, tegenstrijdig	18–20 s per stap, binnen een takt van 20 s	≤4 h hertraining + 15 min kalibratie	Haalbaar op een leveranciersplatform
Ombouwprogrammering	LLM schrijft of bewerkt robotprogramma's offline, geverifieerd in simulatie	77–99.7% op ABB RAPID-herschrijvingen	Geen effect op takt	Minuten tot uren per programma	Haalbaar als engineeringtool

Elektronicataak	Werkbare leerling	Verwacht slagingspercentage	Snelheid vs een persoon	Kost per ombouw	Oordeel 2026–2028
End-to-end geleerd grijpen van bewegende transportbanden	VLA-leerling	37–59%; ≤37% op snelle banden	—	≥458 demo's per opstelling	**Niet vóór 2028; gebruik klassieke tracking plus geleerde perceptie**
Nieuwe producten zonder nieuwe data (zero-shot)	Generieke VLA	60–80%; ongeziene connectoren 29–67%	—	—	**Niet realistisch**
Frontier-LLM in de lus die de arm aanstuurt	GPT-6-, Claude-, Gemini-agents	Grijpen 28–45%; precisie-insertie 10%	≈30 min of 280 s per taak vs ≈20 s	—	**Niet realistisch**
Vrij geleide kabelbomen en meerstaps flexibele assemblage	—	Geen productiebewijs in 2026	—	—	**Niet vóór 2028**

Bronnen: de bovenstaande studies en (Universiteit van Passau over RAPID-herschrijving, GraspVLA, Komax) **meestal single source en vendordata** . In 2026–2028 is de waarde van een groot model voor robots een vooraf getrainde prior en een offline programmeur, niet iets dat een leerling "lesgeeft". Elke nabij-productiecase die ik vond (Neuromeka, AgiBot bij Longcheer, Fraunhofer, ContactDP) vertrouwt op menselijke demonstraties, kracht- of tastsensoren, simulatie of digitale tweelingen en gesuperviseerde RL. Geen enkele vertrouwt op labels van een frontier-model.

Kosten en terugverdientijd

04 Kosten en terugverdientijd: AI-review verdient zich in maanden terug, machinebediening in jaren, uithaalarmen nooit

Loonkosten: de VS is ongeveer 5.5 keer Shenzhen, Taiwan ongeveer 1.5 keer, het grootste deel van Zuidoost-Azië onder de helft

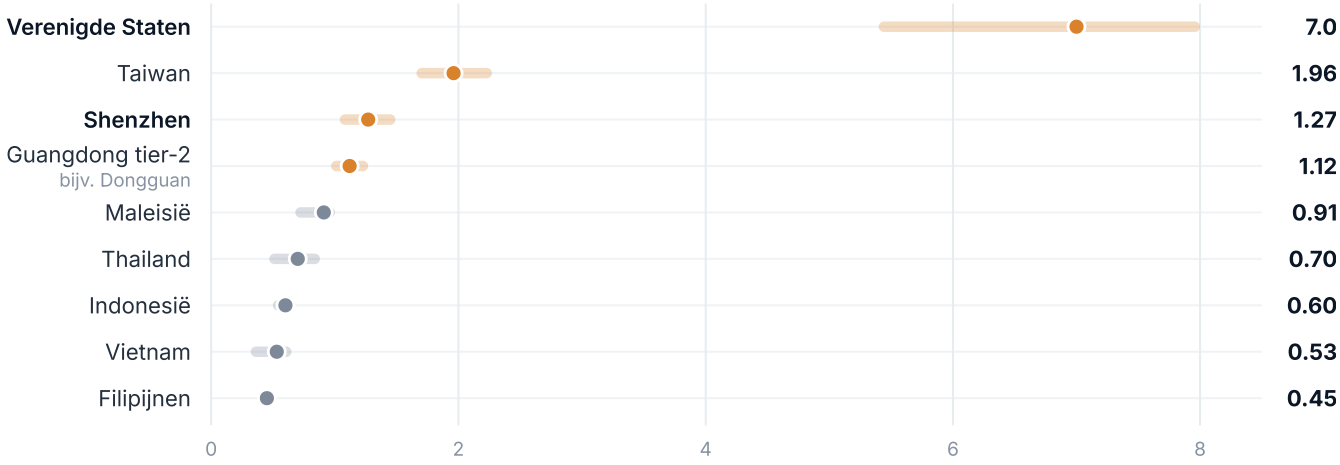
Het rendement op het vervangen van arbeid begint bij wat die arbeid kost. De officiële loonrichtlijn van Shenzhen zet de mediaan van 2024 voor een kwaliteitsinspecteur in de productie op **RMB 74,074** per jaar; lonen zijn 86.92% van de totale arbeidskosten in de elektronicaproductie, dus de volledig belaste kost is ongeveer RMB 85,000. Het JETRO-onderzoek van 2025 zet de jaarlijkse kost van een Chinese fabrieksarbeider op $12,667, ongeveer RMB 85,100, binnen 0.2% (loonrichtlijn Shenzhen, JETRO 2025) verified .

Regio	Volledig belaste jaarkost van één inspecteur of handler (laag / basis / hoog)	Grondslag
Vasteland-China, Shenzhen	RMB 70k / **85k** / 100k ($10.4k / **12.7k** / 14.9k)	Officiële mediaan × 1.15; komt overeen met JETRO verified
Guangdong tier-2-loongebieden (Dongguan enz.)	RMB 65k / **75k** / 85k	Guangdong-mediaan 2023 × 1.15 plus twee jaar loonsverhogingen estimate
Jangtsekiangdelta (Suzhou, Kunshan, Wuxi)	Als Shenzhen	Geen grondslag voor goedkopere inspectiearbeid; het hoogste minimumloon van Jiangsu ligt ongeveer 5.6% boven dat van Shenzhen estimate
Taiwan	$16.6k / **19.6k** / 22.7k (NT$534k / **630k** / 729k)	Beroepsloononderzoek van het ministerie van Arbeid plus ≈1.18 wettelijke lasten; JETRO als hoge kant estimate
Maleisië	$6.8k / **9.1k** / 10.0k	JETRO en minimumlonen estimate
Thailand	$4.7k / **7.0k** / 8.8k	Idem
Indonesië	$5.0k / **6.0k** / 6.3k	Idem
Vietnam	$3.2k / **5.3k** / 6.5k	Idem
Filipijnen	$4.0k / **4.5k** / 4.9k	Idem
Verenigde Staten	$54k / **70k** / 80k	BLS mei 2025 mediaan inspecteur $48,570 ÷ lonen op 65.5% van de vergoeding estimate

Bronnen: (JETRO 2025, loonrichtlijn Guangdong, Wuxi HR-bureau, Taiwan beroepsloononderzoek, PTS News, BLS OEWS, BLS ECEC).

FIG 17 **Wat één inspectie- of handlingmedewerker per jaar kost *Volledig belast zijn de VS ongeveer 5.5× Shenzhen en Taiwan ongeveer 1.5×.***

volledige jaarkosten per QC- of handlingmedewerker ($10k; punt = basis, balk = laag–hoog)



Bron: Officiële loonrichtlijnen van Shenzhen, Ministerie van Arbeid van Taiwan, JETRO, BLS; schattingen van de auteur.

Drie correcties leggen de grondslag. Ten eerste verhoogde Shenzhen in juli 2026 zijn pensioenbijdragevloer van RMB 4,775 naar 5,521 per maand (Shenzhen Bendibao) single source, dus een inspecteur in Shenzhen kost in 2026 waarschijnlijk RMB 80–90k. Ten tweede is een post geen persoon: twee ploegen van 12 uur vereisen ongeveer 2.0 mensen (zes dagen per week) tot 2.4 (het hele jaar) in de Parelrivierdelta, ongeveer 2.7–3.1 in Zuidoost-Azië onder overwerkplafonds, en 4.2–4.5 voor 24×7 in de VS en Taiwan estimate. AI-review laat één reviewer doorgaans twee tot vier lijnen dekken, dus de realistische besparing is **0.5–1.5 voltijdse mensen per lijn, niet één persoon per lijn per ploeg**. Ten derde jagen kopers in Taiwan, Vietnam en Maleisië niet alleen op kosten: in het JETRO-onderzoek zei 61.1% van de bedrijven in Taiwan, 48.2% in Vietnam en 44.9% in Maleisië dat werving de afgelopen twee jaar moeilijker was geworden, tegen 18.6% in vasteland-China (JETRO 2025) single source.

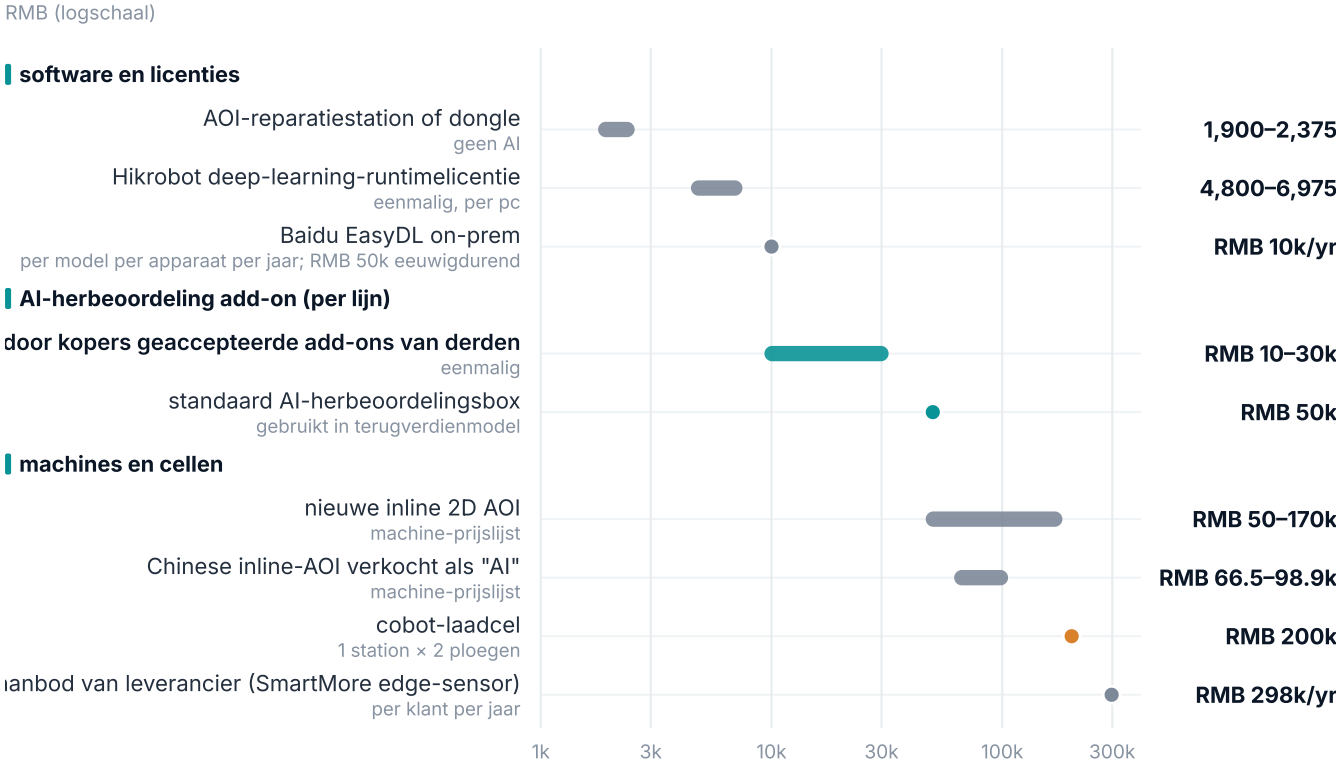
Apparatuurprijzen: vier grondslagen die niet mogen worden gemiddeld

AOI-prijzen verschillen vooral omdat ze verschillende dingen meten. Online-noteringen (november 2026, nieuw, gebruikt en verhuurd gemengd) lopen ongeveer RMB 17–80k voor offline of desktop-2D, RMB 50–170k voor inline-2D en RMB 80–475k voor inline-3D, tot RMB 700k voor Koreaanse 3D (1688 inline-AOI, 1688 offline-AOI, 1688 3D-AOI) single source, adviesprijzen. Een aanbesteding van november 2025 begrensde één 3D-AOI plus één 3D SPI op RMB 800k (aanbestedingsbericht). Het prospectus van Sinic-Tek impliceert RMB 428–512k per 3D-AOI in 2019–2022, ongeveer RMB 340–470k volgens het rapport van 2025 (Sinic-Tek prospectus, Sinic-Tek jaarverslag 2025) primair + estimate. Een overzeese leverancier schat $3,000–20,000 voor desktop-2D, $30–70k inline-2D, $50–90k instap-3D en $100–200k+ high-end-3D, met dezelfde inlinemachine op $30–50k in China (Averroes) vendor-schatting van een verkoper van concurrerende AI-software.

AI-review zelf heeft geen openbare RMB-prijs; leveranciers quoteren per geval na een gratis evaluatie ter plaatse (Sohu over Hengtianyi), en zoekopdrachten in e-commerce-noteringen, aanbestedingen, leveranciersartikelen en Taiwanese media leverden geen openbare quote op (1688, 51SMT). De ankers: ongeveer €27k per lijn geïmpliceerd door de Siemens-case estimate; $21,924 adviesprijs voor een remote-review-netwerkmodule van een Amerikaanse distributeur (Manncorp) vendor-adviesprijs; ongeveer RMB 298k per klant per jaar voor de edge-AI-sensoren van SmartMore in 2025 (36Kr) single source. Een binnenlandse inline-2D-AOI verkocht als "AI" staat op RMB 66.5k (1688 inline-AOI), en een NG-buffer voor handmatige herinspectie op ongeveer RMB 50k (1688 AOI-review). Het plafond voor een derden-review-add-on is dus ongeveer RMB 10–30k per lijn, of een vergelijkbare jaarlijkse vergoeding; daarboven vervangt de koper liever de machine door een binnenlandse AOI met ingebouwde AI estimate.

Anker	Prijs (RMB)	Wat het koopt	Bron
AOI-reparatiestation of softwaredongle (oudere TRI, Zhenhuaxing)	≈1,900–2,375	Licentiesleutel, geen AI	1688 [adviesprijs]
Algemene deep-learning-vision-runtime (Hikrobot VisionMaster-dongle)	≈4,800–6,975 eenmalig per PC	Het platform; de klant bouwt en onderhoudt modellen	1688 [adviesprijs]
Baidu EasyDL on-premises-implementatie	10,000 per model per apparaat per jaar, of 50,000 eeuwigdurend	Verpakte inferentiedienst	Baidu AI-documentatie [prijspagina 2022]
Binnenlandse inline-AOI verkocht als "AI"	66,500–98,900 per machine	Machinevervanging	1688 [adviesprijs]
Edge-AI-sensoren van SmartMore	≈298,000 per klant per jaar	Leveranciersoplossing over meerdere machines	36Kr single source

FIG 18 **De prijsladder** ***Van een paar duizend RMB softwarelicentie tot RMB 200–300k voor een robotcel of een AI-aanbod van een leverancier.***



Bron: 1688-listings, Baidu AI-documentatie (prijslijst 2022), 36Kr, een branche-interview (niet geverifieerd).

Een jaarlijkse vergoeding in Baidu-stijl, met één inferentieserver voor één tot vier lijnen en elk één tot vier modellen, impliceert een abonnement van RMB 2.5–40k per lijn per jaar, realistisch ongeveer RMB 5–10k; op 20% van een eeuwigdurende licentie breken abonnement en aankoop in jaar vijf gelijk, een structuur die het kopiëren waard is estimate. Kopers zijn gewend aan bijna nul softwarekosten: de volledige "overige" omzet van Sinic-Tek (reserveonderdelen, dienstverlening, verhuur, accessoires) was 7.65% van de verkoop in 2025, ongeveer RMB 2,300–3,100 per geïnstalleerde machine per jaar (Sinic-Tek jaarverslag 2025) primair + estimate. Een AI-abonnement van RMB 5–10k per lijn per jaar is twee tot vijf keer wat kopers nu aan AOI-software uitgeven, dus het moet worden gerecht-

vaardigd door de review-uren die het vrijmaakt (ongeveer RMB 75–90k per voltijds persoon per jaar), niet door vergelijking met bestaande softwarekosten.

Voor robotarmen daalde de gemiddelde cobotprijs in China tot ongeveer RMB 58,600 in 2025 (GGII) estimate, en het zesassige gemiddelde van Dobot daalde dat jaar met 19.1% toen goedkopere modellen aandeel wonnen (Huxiu) verified, ook al stegen de adviesprijzen van dezelfde modellen in 2026 met 5–15% (Sina Finance). In de VS staat een FANUC CRX-10iA op $51,434 (Vention) en zetten compilaties een UR7e op $30–48k (cobotcost) estimate. De arm is slechts 40–50% van een Amerikaanse cel (Standard Bots); in China lopen camera's, grippers, veiligheid en engineering makkelijk op tot twee tot vier keer een arm van RMB 40k, dus begroot absolute engineeringkosten in plaats van een veelvoud estimate.

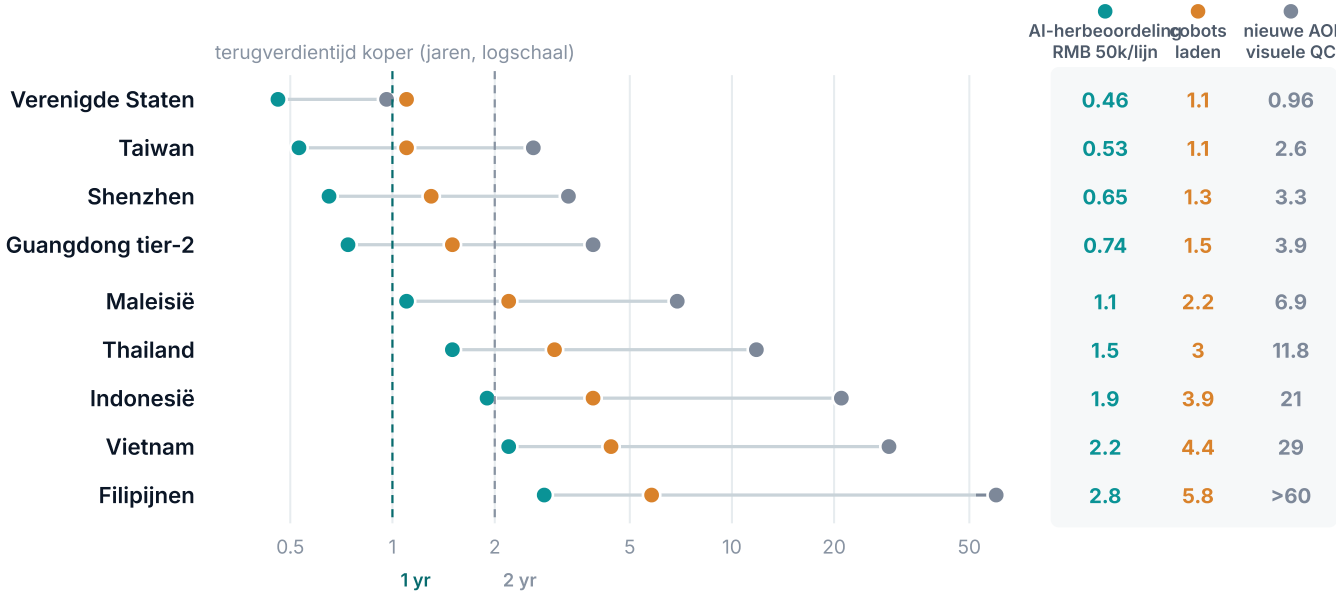
Terugverdientijd: review in maanden, bediening in jaren, uithalen nooit

Eenvoudige terugverdientijd = investering ÷ (bespaarde mensen × belaste loonkost − jaarlijkse lopende kost), voor een tweepleegenlijn zonder subsidies of invoerrechten. Taiwanese apparatuur is geprijsd op 1.3 keer China, Vietnam, Maleisië en Thailand op 1.15 keer, Indonesië en de Filipijnen op 1.2 keer, en de VS op overzeese prijzen estimate.

Scenario	Investering (China / VS)	Bespaarde mensen	Jaarlijkse lopende kost
(a) AI-reviewbox of AI-camerabox	RMB 50k / $30k per lijn	1.0 per lijn	15%
(b) Nieuwe inline-2D-AOI die handmatige visuele controles vervangt (0.5 reviewer per ploeg blijft)	RMB 200k / $60k	1.0	12%
(c) Cobotlaadcel (één station × twee ploegen)	RMB 200k / $140k	2.0	8%
(d) Geïntegreerde detecteer-en-pak-cel (inspectie 1.0 + pakken 0.07)	RMB 300k / $200k	1.07	8%
(e) Een robotarm alleen toegevoegd om NG-platen eruit te halen	RMB 150k / $100k	0.07	8%

Regio	(a) AI-review	(b) Nieuwe AOI	(c) Cobotcel	(d) Detecteer-en-pak	(e) Uithaalarm
Shenzhen	0.65 jr (0.54–0.80)	3.3 (2.6–4.4)	**1.3** (1.1–1.6)	4.5 (3.6–5.9)	Nooit
Guangdong tier-2	0.74	3.9	1.5	5.3	Nooit
Taiwan	0.53	2.6	1.1	3.6	Nooit
Maleisië	1.1	6.9	2.2	9.2	Nooit
Thailand	1.5	11.8	3.0	15	Nooit
Indonesië	1.9	21	3.9	25	Nooit
Vietnam	2.2 (1.6–4.5)	29	4.4	34	Nooit
Filipijnen	2.8	>60 (lopende kost ≈ besparing)	5.8	≈100	Nooit
Verenigde Staten	0.46	**0.96**	**1.1**	3.4	Nooit

FIG 19 **Terugverdientijd per regio** ***AI-herbeoordeling betaalt zich terug in maanden in de VS, Taiwan en Shenzhen en in 1–3 jaar in Zuidoost-Azië; cobot-belading in jaren.***



een robotarm enkel om NG-boards uit te nemen betaalt zich nergens terug
bij first-pass yield boven 90% kost uitnemen slechts 0.003–0.14 FTE per ploeg

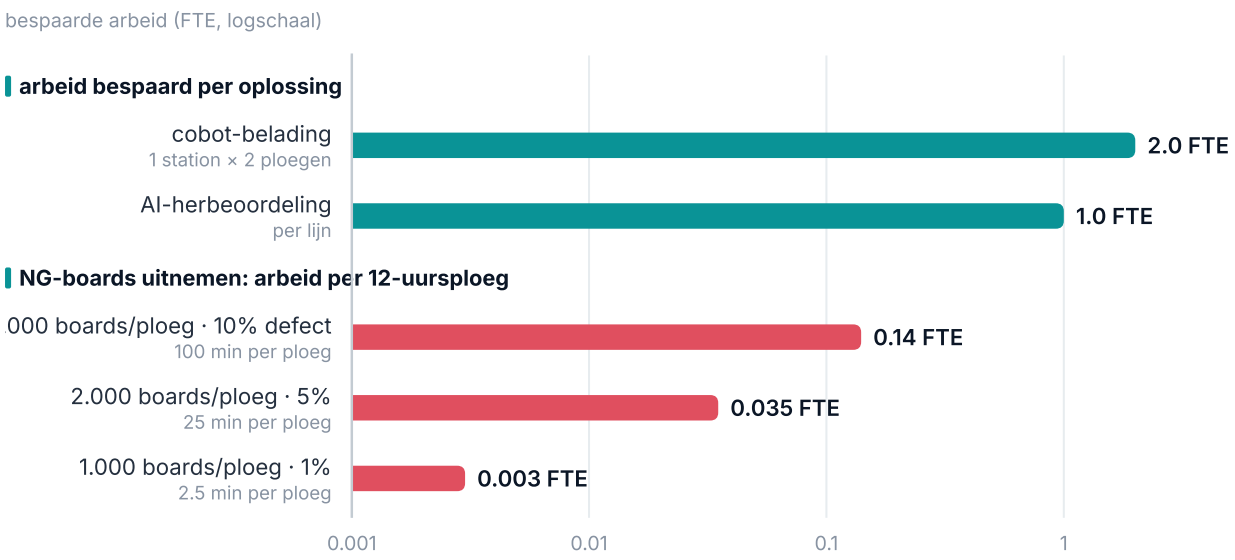
Bron: Model van de auteur: RMB 50k AI-herbeoordelingsbox per lijn; cobot-beladingscel van RMB 200k of $140k, twee ploegen.

Alle waarden zijn schattingen uit de bovenstaande invoer. Gevoeligheden: halveer de bespaarde mensen en elke terugverdientijd verdubbelt ruwweg; draai één ploeg en (b)–(d) verdubbelt; 100% eerstejaarsafschrijving in de VS verkort de terugverdientijd met ongeveer 20%, terwijl invoerrechten op Chinese, Japanse of Koreaanse apparatuur hem met ongeveer 10–40% verlengen; een lokale subsidie van 15% in China verkort hem met ongeveer 15%. Ik vond geen onafhankelijk onderzoek naar de terugverdientijddrempels van kopers; leveranciers zeggen dat AOI "doorgaans in 9–18 maanden" terugverdient (Averroes) en de Siemens-case claimt acht maanden vendor claims. Ik gebruik één en twee jaar als screeningslijnen. Het model laat kwaliteitswinst (garantie, retouren) en verloop- en trainingskosten buiten beschouwing, dus het is conservatief.

Het contra-intuïtieve deel van "detecteren en pakken" is dat de arbeid in review zit, niet in het van de lijn halen van de paar echte NG-platen. Bij 15 seconden handling per NG-plaat estimate:

Platen per ploeg	Defectpercentage	Paktijd per ploeg van 12 uur	Mensen
1,000	1%	2.5 min	0.003
2,000	5%	25 min	0.035
4,000	10%	100 min	0.14

FIG 20 **Waarom slechte boards uitnemen zich niet terugbetaalt** ***Boven 90% first-pass yield kost uitnemen slechts 0.003–0.14 van een persoon per ploeg.***



Bron: Schattingen van de auteur.

Het basisscenario (2,000 platen per ploeg, 5% defect, twee ploegen) is 0.07 van een persoon, ongeveer RMB 6,000 per jaar waard in Shenzhen, minder dan de ruwweg RMB 12,000 jaarlijkse lopende kost van de arm, dus het verdient zich nergens terug; zelfs het slechtste geval (4,000 platen per ploeg, 10% defect) duurt ongeveer 13 jaar in Shenzhen en 9 in de VS estimate. Als de zorg het stoppen van de lijn om te pakken is, lost een standaard NG-afleider of bufferband het op zonder arm. Mensen in de branche zeggen “boven 90% first-pass-opbrengst, stop de lijn en haal de plaat er met de hand uit”, en **dat klopt voor pakken, maar niet voor inspectie**. Het productontwerp op een SMT-lijn is daarom “inspecteer plus markeer of leid af”, met robotarmen voor plaatsen waar elk onderdeel wordt gehanteerd (testopstellingbelading, tray- en magazijnhandling, verpakking), waar inspectie een bijna gratis add-on wordt en scenario (c) geldt.

Subsidies, belasting en invoerrechten bewegen de terugverdientijd met 15–25%, maar niet de rangorde

Vasteland-China: steun verlaagt vooral de financieringskost; losse machines komen zelden rechtstreeks in aanmerking. Een nationale leningssubsidie betaalt 1.5 procentpunt rente gedurende maximaal twee jaar in 14 prioriteitsketens (Yicai) verified, ongeveer RMB 9,000 op een machine van RMB 300k; de upgradepremie van Guangdong van maximaal 20% vereist een geregistreerd project met ten minste RMB 7 miljoen nieuwe apparatuur en sluit Shenzhen uit (Guangdong-regels), en Dongguan betaalde in juli 2026 daadwerkelijk 4.42% (Dongguan-toewijzing) single source, primair. Een AI-inspectieapparaat wordt alleen goedgekeurd als apparatuur “met ondersteunende software” binnen het geregistreerde project van de koper, dus het basisscenario gaat uit van geen subsidie.

Taiwan: incentives gaan naar de koper. Onder artikel 10-1 van de Wet op industriële innovatie, in mei 2025 gewijzigd om AI op te nemen, kan een koper 5% van de uitgaven in 2025–2029 aan nieuwe slimme machines of AI (inclusief CNN/DNN-gebaseerde inspectie) verrekenen met de inkomstenbelasting van dat jaar, of 3% per jaar gedurende drie jaar (Wet, verrekenregels) verified; subsidies aan de leverancierszijde zijn gesloten voor vasteland-kapitaal (MOEA) single source.

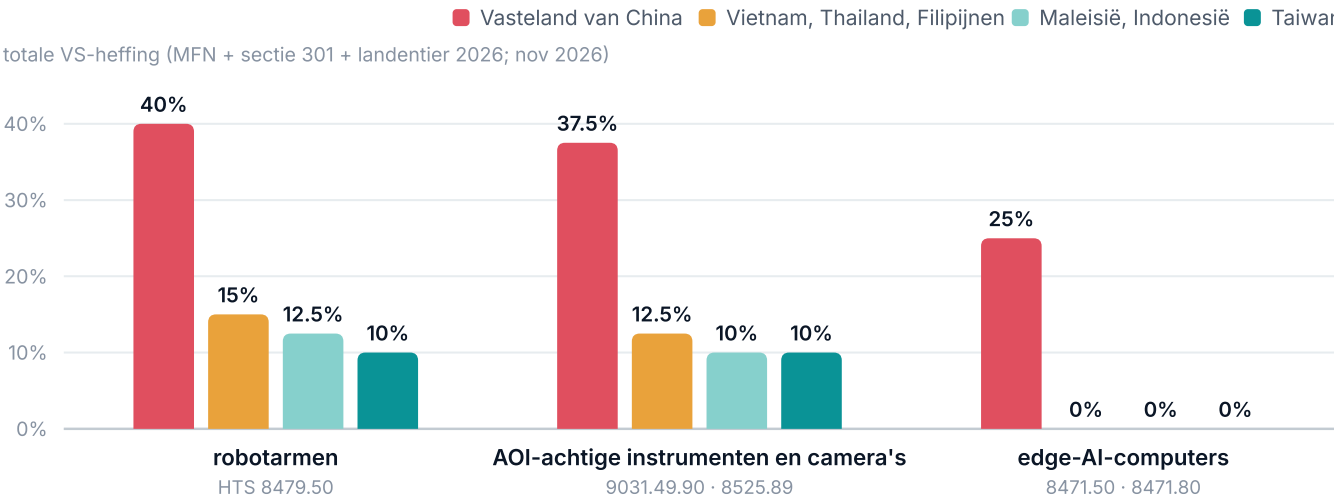
Zuidoost-Azië: alleen Maleisië en Thailand belonen de aankoop van automatisering. De 200%-automatiseringskapitaalaftrek van Maleisië over de eerste RM10 miljoen is voor een winstgevende koper ongeveer 24% van de capex waard (MIDA, PwC Maleisië) verified; de BOI van Thailand geeft een driejarige belastingvrijstelling tot 50–100% van een upgrade-investering, het hogere plafond wanneer Thaise automatiseringsleveranciers betrokken zijn (BOI) single source, primair, dus het betreden van Thailand werkt het best met een lokale integrator. Vietnam, Indonesië en

de Filipijnen belonen het bouwen van fabrieken in plaats van het kopen van automatisering (Vietnam Law Magazine, PwC Filipijnen).

Verenigde Staten: afschrijving verkort de terugverdientijd; invoerrechten verlengen hem. Activa gekocht na 19 januari 2025 krijgen permanente 100% eerstejaarsafschrijving, ongeveer een 21% federaal belastingschild voor een winstgevende koper (IRS Notice 2026-11) single source, primair ; het Section 232-onderzoek naar robots en industriële machines eindigde op 28 oktober 2026 met onderhandelingen in plaats van nieuwe invoerrechten, te herbeoordelen binnen 180 dagen (White House, Mayer Brown) verified . Eindassemblage in Zuidoost-Azië verlaagt de toeslag op armen en instrumenten van het AOI-type van ongeveer 37.5–40% naar 10–15% (hoofdstuk 1).

Regio	Belangrijkste instrumenten	Effect op terugverdientijd koper estimate
Vasteland-China	1.5-punts leningsubsidie voor maximaal twee jaar; provinciale en stedelijke upgradepremies (Guangdong vereist ≥RMB 7M geregistreerde projecten; Dongguan betaalde daadwerkelijk 4.42%); maatregelen Shenzhen 2026–2029 (tarieven niet openbaar)	Ongeveer 0–5% voor een losse machine; 4–15% binnen een groot project
Taiwan	Wet §10-1-verrekening, 5% of 3% × 3 jaar	Ongeveer 5% korter, als de koper genoeg belasting betaalt
Maleisië	200%-automatiseringskapitaalaftrek (eerste RM10M)	Ongeveer 15–25% korter voor winstgevende kopers
Thailand	BOI driejarige belastingvrijstelling tot 50–100% van de investering	Ongeveer 15–25% korter voor winstgevende kopers buiten een belastingvakantie
Vietnam, Indonesië, Filipijnen	Inkomstenbelastingincentives voor het bouwen van fabrieken; extra 20% apparaatafschrijving in de Filipijnen	Vrijwel niets voor losse automatiseringsaankopen
Verenigde Staten	100% eerstejaarsafschrijving; ≈37.5–40% gecombineerd invoerrecht op vasteland-apparatuur	Afschrijving ≈20% korter; invoerrechten ≈10–40% langer

FIG 21 **VS-heffing op robots, AOI en edge-computers** ***Armen uit het Chinese vasteland betalen ongeveer 40%; verzonden vanuit Zuidoost-Azië of Taiwan, 10–15%.***



Bron: USITC Harmonized Tariff Schedule hoofdstuk 99; Federal Register; USTR (november 2026).

Per regio: in vasteland-China verdient een AI-reviewbox zich in maanden terug en een cobotcel in 1.3–1.5 jaar, terwijl een nieuwe AOI die alleen wordt gekocht om arbeid te besparen 3–4 jaar kost en geïntegreerd detecteer-en-pak

4.5–5.3, beide boven een tweejaren-scherm. Taiwan doet het iets beter dan Shenzhen (0.53 jaar voor review, 1.1 voor een cobotcel), en werving is daar moeilijker. In Maleisië en Thailand kost een AI-box 1–1.5 jaar en een cobotcel 2–3, ongeveer 15–25% sneller met lokale belastingincentives voor winstgevende kopers. In Indonesië, Vietnam en de Filipijnen verdient alleen een zeer goedkope AI-box zich binnen 2–3 jaar terug; de vraag daar komt meer voort uit kwaliteitseisen van klanten en personeelstekorten. In de VS verdient elke optie behalve een pure uithaalarm zich in 0.2–3.4 jaar terug, met invoerrechten, de Section 232-herbeoordeling en NRTL-certificering als belangrijkste variabelen.

05 Markt en concurrentie

Markt, concurrentie en hoe kopers betalen

Marktomvang: gepubliceerde cijfers verschillen met veelvouden; alleen een bottom-up-telling is bruikbaar

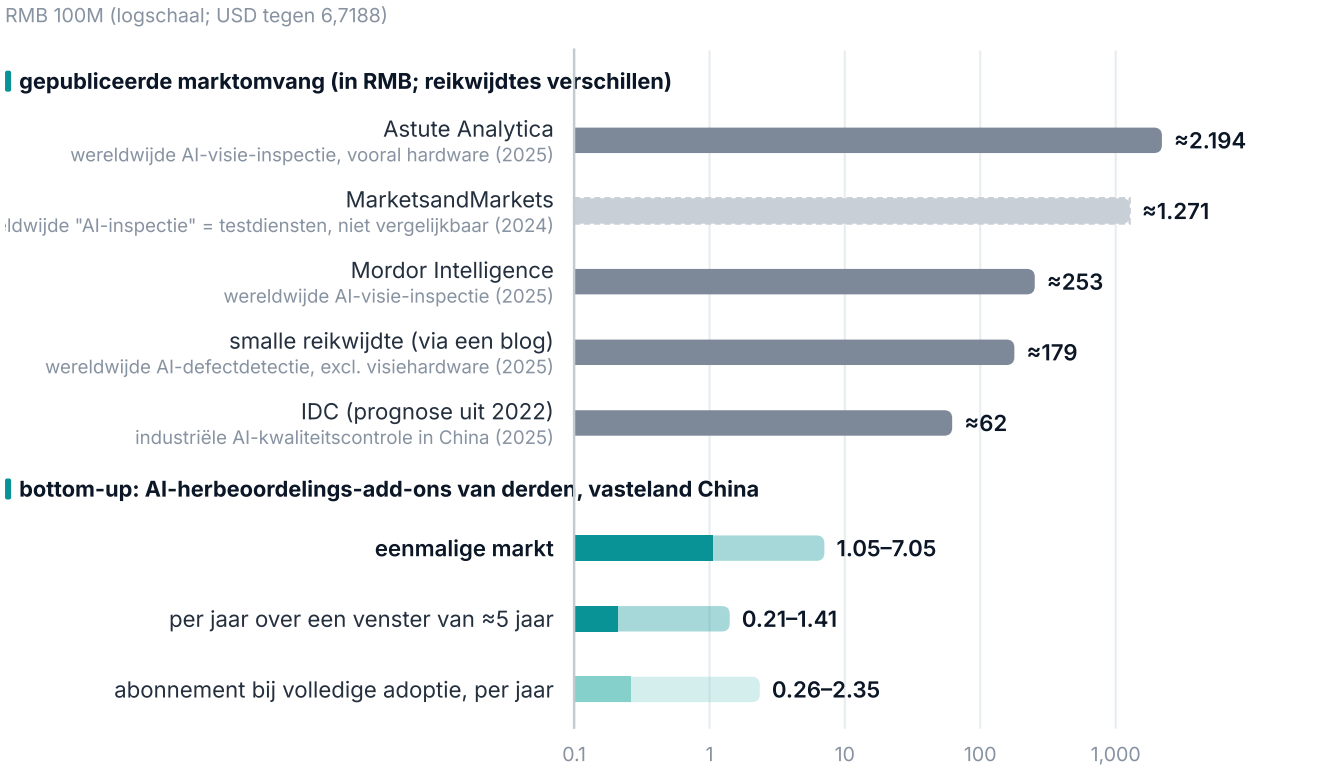
Elk marktcijfer moet met zijn reikwijdte worden gelezen. De machinevisiemarkt van China was RMB 18.15 miljard in 2024 (Zhiyan) verified, met 3C-elektronica als grootste afnemer; wereldwijde schattingen voor machinevisie verschillen ongeveer drievoudig en voor AOI ongeveer 6.5-voudig (MarketsandMarkets, Vision Systems Design/Interact Analysis, Zhiyan). Bottom-up uit de omzetten van Koh Young, TRI, Jutze, Sinic-Tek en ViTrox ligt de wereldwijde SMT-inspectieapparatuur in de orde van $1–2 miljard estimate.

Er is geen zelfstandige markt voor "AI-visuele-inspectiesoftware". De gepubliceerde cijfers verschillen meer dan tienvoudig en omvatten allemaal hardware of diensten:

Reikwijdte	Uitgever	Cijfer
Wereldwijde "AI-detectie van industriële defecten" (smal, exclusief vision-hardware)	Naamloos bedrijf, via een iFactory-blog	$2.66 miljard in 2025, 8.6% CAGR tot 2035, elektronica 34%
Wereldwijde "AI-visuele inspectie"	Mordor Intelligence	$3.76 miljard in 2025, $12.8 miljard in 2031, 22.67% CAGR
Wereldwijde "AI-visuele inspectie" (voornamelijk hardware)	Astute Analytica	≈$32.66 miljard in 2025
Wereldwijde "AI-inspectie" (test- en certificeringsdiensten; niet vergelijkbaar)	MarketsandMarkets	$18.92 miljard in 2024
China "industriële AI-kwaliteitsinspectie"	IDC, prognose uit 2022	$958 miljoen in 2025 (≈RMB 6.2 miljard), 91.5% in communicatie, elektronica en twee andere sectoren

Bronnen: (iFactory, GII/Mordor, GII/Astute, MarketsandMarkets, IDC) elk single source, meestal samenvattingen van betaalde rapporten.

FIG 22 **Marktomvang verschilt meer dan tienvoudig*Alleen de smalle reikwijdte en een bottom-up telling zijn bruikbaar voor beslissingen.***



Bron: Mordor Intelligence, Astute Analytica, MarketsandMarkets, IDC (2022), iFactory; bottom-up schatting van de auteur.

Bottom-up is er slechts één openbare telling van SMT-lijnen: New Venture Research schatte ongeveer 73,231 lijnen wereldwijd in 2020, 47% daarvan in China (19% in 2000), ongeveer 70% high-end en 30% low-end (EMSNow, NVR SMT2020-samenvatting) single source, 2020 ; NVR behoort nu tot IPC en nieuwere data zijn betaald (NVR). Groei van 3–5% per jaar met een China-aandeel van 40–47% geeft ongeveer 87,000–98,000 lijnen wereldwijd in 2026 en 35,000–46,000 in vasteland-China (basis 40,000) estimate .

Invoer	Laag	Basis	Hoog	Grondslag
SMT-lijnen vasteland-China (2026)	35,000	40,000	46,000	NVR 2020-basis × groei × aandeel
Aandeel zonder AI-review	75%	80%	85%	Retrofitvenster, hieronder
Aandeel dat een derde kan bereiken (niet in-house, niet aan OEM-AI gebonden, kan betalen)	40%	50%	60%	Aanname: ≈30% low-end-lijnen plus enkele high-end-lijnen met gemengde merken, exclusief in-house-bouwsels in DBG-stijl
Adresseerbare lijnen	**10,500**	**16,000**	**23,500**	Product
Eenmalige prijs per lijn	RMB 10k	RMB 20k	RMB 30k	Prijsladder en plafond (hoofdstuk 4)
Eenmalige adresseerbare markt (geïnstalleerde basis)	**RMB 105M**	**RMB 320M**	**RMB 705M**	Product
Per jaar over een venster van ≈5 jaar	RMB 21M	RMB 64M	RMB 141M	Venster krimpt 8–12% per jaar
Abonnement (RMB 2.5k / 5k / 10k per lijn per jaar, volledige adoptie)	RMB 26M per jaar	RMB 80M per jaar	RMB 235M per jaar	Anker van jaarlijkse vergoeding in Baidu-stijl

FIG 23 **Bottom-up: 40.000 lijnen → 16.000 serviceerbaar → ≈RMB 320M eenmalig*De niche van AI-herbeoordeling door derden in het Chinese vasteland is reëel, maar klein.***



een nieuwkomer die in 3 jaar 2–5% van de serviceerbare lijnen wint: ≈320–800 lijnen, ≈RMB 2–5M per jaar
als abonnement (RMB 5k per lijn per jaar): RMB 1.6–4M recurring, slechts 0.3–2.3% van IDC's prognose voor industriële AI-QC in China

Bron: NVR (basis 2020); schattingen van de auteur.

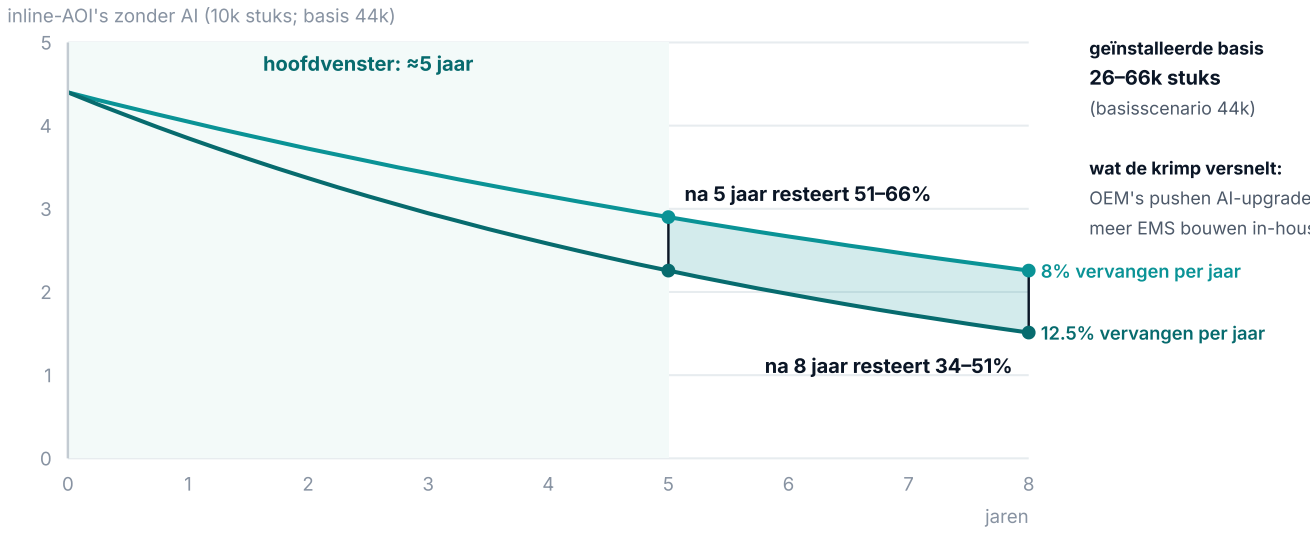
Dit zijn allemaal schattingen; noch het aandeel noch de prijs is openbaar gemeten. De niche is slechts 0.3–2.3% per jaar waard van de IDC-prognose voor alle industriële AI-kwaliteitsinspectie van China in 2025 (ongeveer RMB 6.2 miljard). Een nieuwkomer die in drie jaar 2–5% van de basis-adresseerbare lijnen wint, zou ongeveer 320–800 lijnen hebben: RMB 6.4–16 miljoen in totaal bij RMB 20,000 per lijn, ongeveer RMB 2–5 miljoen per jaar, of RMB 1.6–4 miljoen per jaar aan terugkerende omzet bij RMB 5,000 per lijn estimate. Dat is genoeg voor een businessunit binnen een bedrijf, niet voor een zelfstandig bedrijf op ventureschaal. Elke 1,000 geïnstalleerde lijnen in Taiwan, Zuidoost-Azië of de VS is ruwweg RMB 3–15 miljoen eenmalig waard tegen vasteland-prijzen; prijzen zijn hoger in Taiwan en de VS, maar kopers daar gebruiken vaker OEM-3D-AOI's met AI-opties estimate.

Het retrofitvenster is ongeveer vijf jaar

Geen AOI-maker publiceert de levensduur van machines; EMS-bedrijven schrijven apparatuur af over 3–10 jaar (DBG jaarverslag 2025), en oudere AOI's verliezen software-updates en reserveonderdelen na end-of-life (I-Connect007) single source. Ik gebruik 8–12 jaar (basis 10).

Bij 1.0–1.65 inline-AOI's per lijn heeft het vasteland ongeveer 35,000–78,000 inline-AOI's in dienst (basis ongeveer 55,000), consistent met bekende leveringen (Circuits Assembly, Sinic-Tek jaarverslag 2025). Geen bron meet hoeveel er AI-review draaien; aangenomen dat binnenlandse machines dit sinds ongeveer 2023 aanbieden en 40–60% van de recente leveringen het gebruikt, **draaien ongeveer 26,000–66,000 inline-AOI's in vasteland-China (basis ongeveer 44,000) geen AI** estimate. Als elke nieuwe machine nu met AI wordt geleverd, krimpt die pool 8–12.5% per jaar: 51–66% blijft na vijf jaar en 34–51% na acht, sneller als AOI-makers AI-upgrades naar nieuwere machines pushen. **Het grootste deel van de kans valt in de komende vijf jaar of zo**, met middelgrote EMS die oudere binnenlandse 2D-AOI's, gemengde merken en geen AI-upgradepad in de kern draaien.

FIG 24 **Het retrofitvenster is ongeveer vijf jaar** ***Inline-AOI's zonder AI krimpen 8–12.5% per jaar naarmate ze worden vervangen of geüpgraded.***

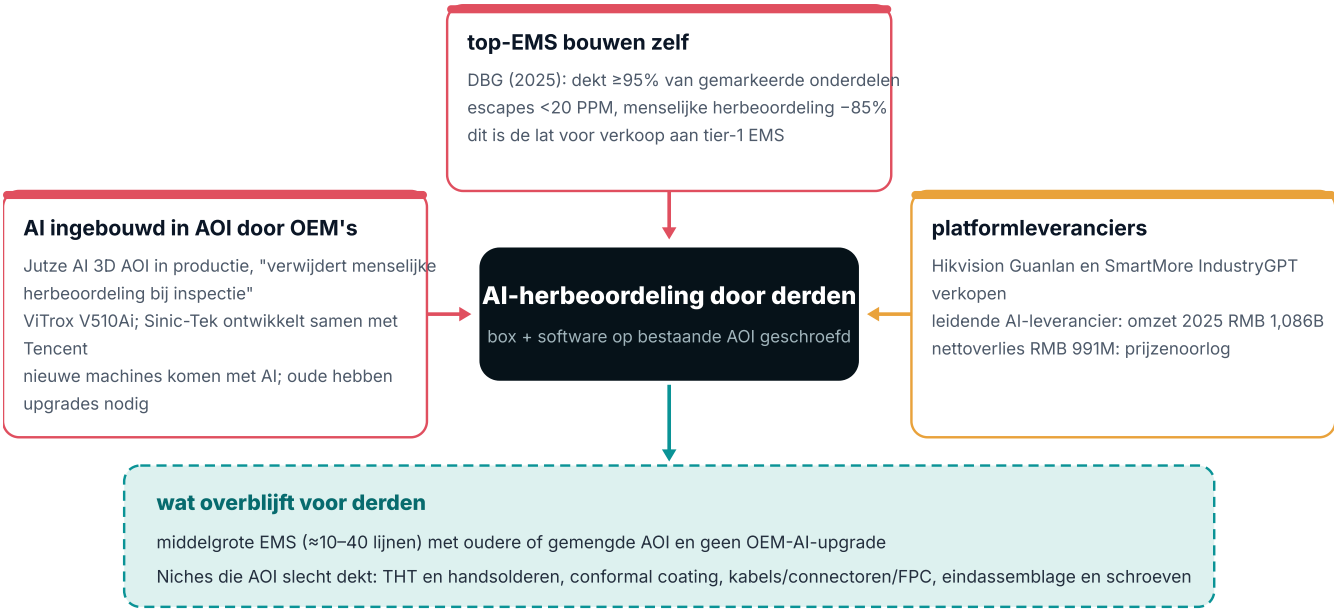


Bron: Schattingen van de auteur; geïnstalleerde basis en aandelen niet onafhankelijk geverifieerd.

De concurrenten zijn al sterk: de reviewwig is omsingeld

De AI-reviewwig vernauwt vanuit drie richtingen. AOI- en visieleiders bouwen AI in nieuwe platforms: Jutze heeft "AI-algoritmen volledig overgenomen" (Jutze investor relations, juni 2026), de industriële AI-omzet van OPT groeide in 2025 met 96.77% (OPT jaarverslag 2025) single source, primair, en ViTrox claimt tot 95% nauwkeurigheid voor autonome review (ViTrox) vendor claim. Grote fabrikanten bouwen zelf (iThome). En door durfkapitaal gefinancierde AI-bedrijven verkopen op schaal: de edge-sensorklanten van SmartMore groeiden in twee jaar van 65 naar 251 (36Kr) single source, Hikvision verkoopt zijn Guanlan-model voor inspectie (Xinhua), en "oog, brein en hand"-integratie wordt al aangeboden door SmartMore, Hikrobot en Mech-Mind (36Kr, Securities Star).

FIG 25 **De herbeoordelingswig is aan drie zijden omsingeld** ***Ruimte voor derden zit in middelgrote EMS en in inspectieniches die AOI slecht dekt.***



Bron: Jaarverslagen van Jutze, Sinic-Tek en DBG; publiek materiaal van Hikvision en SmartMore; samengesteld door de auteur.

De ruimte die overblijft voor een nieuwkomer is smal: middelgrote EMS met oudere 2D/3D-AOI's en geen AI-upgrade van de maker; lijnen die meerdere AOI-merken mengen (één retrofit-leverancier ondersteunt al Omron, VI, Koh Young en Mirtec (Sohu over Hengtianyi)); niches die SMT-AOI's slecht dekken, zoals through-hole en handgesoldeerde verbindingen, conformal coating, kabels, connectoren en FPC's, en einduiterlijk en schroeven; en kleine fabrieken die een AOI van RMB 230–280k niet kunnen betalen.

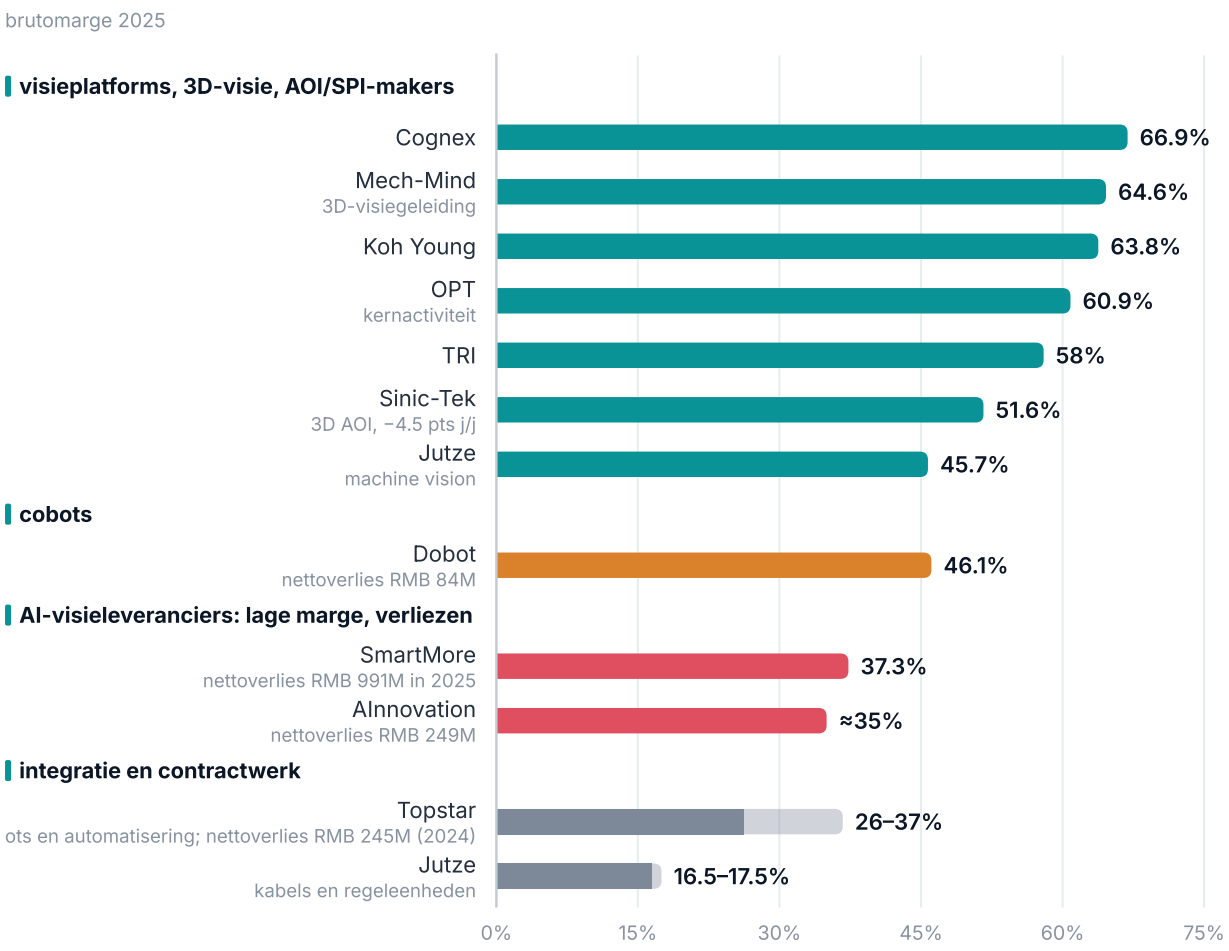
Het dichtstbijzijnde beursgenoteerde vergelijkingsbedrijf geeft een waarschuwing. Weiyi Zhizao, dat AI-inspectieapparatuur en robots verkoopt, zag de 3C-omzet in januari–mei 2026 met 79.9% op jaarbasis dalen, erkende "toenemende concurrentie en prijsdruk", stelde zijn Hongkong-notering op 25 oktober 2026 uit en maakte 15 vóór voltooiing beëindigde projecten bekend, waarvan 12 embodied-robotprojecten (Weiyi Zhizao prospectus, uitstelbericht) single source, primair. AInnovation eindigde 2025 met 29.8% van zijn vorderingen voorzien (AInnovation jaarverslag 2025). De 3C-inspectiewig is overvol, en de prijszettingskracht per lijn is zwak.

De marge zit bij geproductiseerde hardware

Positie in de keten	Representatieve bedrijven en brutomarge 2025	Winst en incasso
Visieplatforms	Cognex 66.9%	Winstgevend; stoot ≈$22M omzet met lage marge af
AOI/SPI-apparatuur	Koh Young 63.8%; TRI 58%; OPT-kernactiviteit 60.85%; Sinic-Tek 3D-AOI 51.62% (−4.53 punten op jaarbasis); Jutze machinevisie 45.74%	Winstgevend; vorderingsdagen 139 (Jutze) tot 247–274 (OPT)
3D-visiegeleiding	Mech-Mind 64.6%	—
Leveranciers van AI-visieoplossingen	SmartMore 37.3%; AInnovation ≈35%	Verlieslatend: nettoverlies SmartMore RMB 991M in 2025, AInnovation RMB 249M
Cobots	Dobot 46.1%	Nettoverlies Dobot RMB 84M; UR-omzet van $326M (2022) naar ≈$232M (2025, prognose)
Systeemintegratie	Topstar robotica en automatisering 26.2–36.7%	Nettoverlies RMB 245M (2024); vorderingsdagen 180–240
Contracthardware	Jutze kabels en besturingseenheden 16.5–17.5%	—

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

FIG 26 **De marge zit in geproductiseerde hardware** ***Visieplatforms en apparatuurfabrikanten verdienen 46–67% brutomarge; AI-leveranciers 35–37% en lijden verlies.***



Bron: Jaarverslagen of prospectussen van 2025.

Een nieuwkomer behoudt zijn marge alleen door inspectie om te zetten in een product, apparatuur of software; project-voor-project-afstemming duwt marges naar ongeveer 35% met aanhoudende verliezen. De mechanismen werken ook tegen kleine bedrijven: omzet wordt pas erkend na acceptatie ter plaatse, vooruitbetalingen zijn klein (Jutze, OPT), pilots zijn meestal gratis (A3/Mindtrace), en merkeigenaren specificeren vaak de inspectieapparatuur op de lijnen van hun contractfabrikanten.

Hoe kopers betalen: 30% bij ondertekening, 60% na acceptatie, 10% een jaar vastgehouden

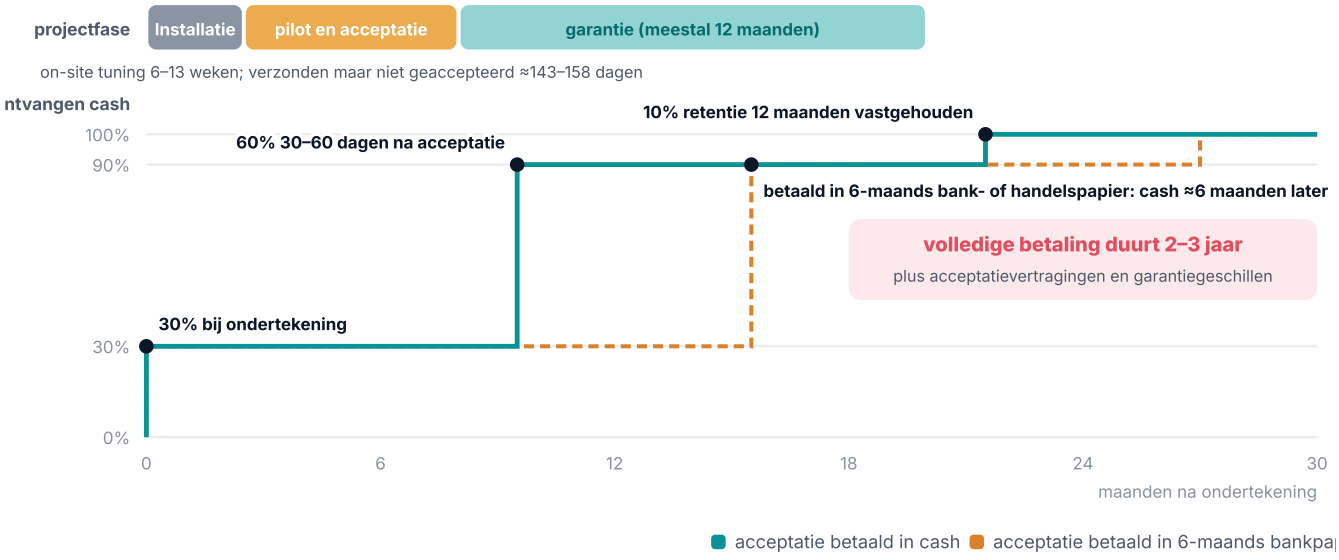
De betalingsvoorwaarden die vasteland-elektronicafabrieken voor inspectieapparatuur gebruiken, staan uitgeschreven in 13 met naam genoemde kopercontracten die zijn opgenomen in het antwoord op de noteringsvragen van Sinic-Tek (Sinic-Tek antwoord op vragen 8-1-1) single source: één primaire deponering, meerdere onafhankelijke kopers, 2020–2022 :

Koper (jaar)	Ondertekening	Levering	Acceptatie	Retentie of saldo
Een Hikvision-eenheid (2021)	30% binnen 15 dagen	—	65% binnen 30 dagen na definitieve acceptatie	5% garantiertentie betaald 12 maanden na acceptatie; deel betaald in bankacceptatienota's van 6+ maanden
Nanchang Huaqin (2020–2021)	30% binnen 10 werkdagen	—	60% op 60 dagen na acceptatie	10% na een garantie van één jaar

Koper (jaar)	Ondertekening	Levering	Acceptatie	Retentie of saldo
Guangdong Yinuo (2021, 3D-AOI)	30% binnen 10 werkdagen	—	60% binnen 30 dagen	10% na drie maanden storingsvrij draaien
Xiamen Hongxin, Jingmen Hongyi (2022)	30% binnen 10 werkdagen	—	60% binnen 30 werkdagen na definitieve acceptatie, banknota's van 6 maanden	10% na een garantie van één jaar
AKM Panyu, Suzhou (2021–2022)	30% binnen 7–10 werkdagen	—	60% een maand na acceptatie en facturering, banknota's van 180 dagen	10% na 360 dagen
BYD Precision, Shenzhen, Huizhou (2020)	—	30% bij aankomst	70% na acceptatie en facturering	Alles in commerciële nota's van 6 maanden
Shenzhen H&T (2020)	—	—	90% na acceptatie en facturering	10% na zes maanden storingsvrij gebruik

Andere primaire deponeringen tonen dezelfde structuur: de grote projecten van Mech-Mind betalen 30% bij ondertekening, 30% vóór levering, 30% na acceptatie en 10% garantiertentie over 12–15 maanden (Mech-Mind prospectus); GKG zegt dat retentie in de apparatuurindustrie doorgaans 5–10% van het contract is, een jaar na acceptatie betaald (GKG antwoord op vragen) single sources, primair. De norm is dus ongeveer 30% bij ondertekening (20–30%, nul bij sommige fabrieken zoals BYD), 60–65% na acceptatie (30–60 dagen later betaald, vaak in bank- of commerciële nota's van zes maanden, wat bruikbare cash nog een half jaar vertraagt), en 5–10% retentie gedurende 3–12 maanden (12 meest gebruikelijk) verified over meerdere contracten.

FIG 27 **Hoe kopers betalen *30% bij ondertekening, 60% na acceptatie, 10% een jaar vastgehouden: volledige betaling duurt 2–3 jaar.***



Bron: Antwoord op vragen van Sinic-Tek, Mech-Mind, GKG, JT en andere stukken; schema op gebruikelijke voorwaarden, timing geschat.

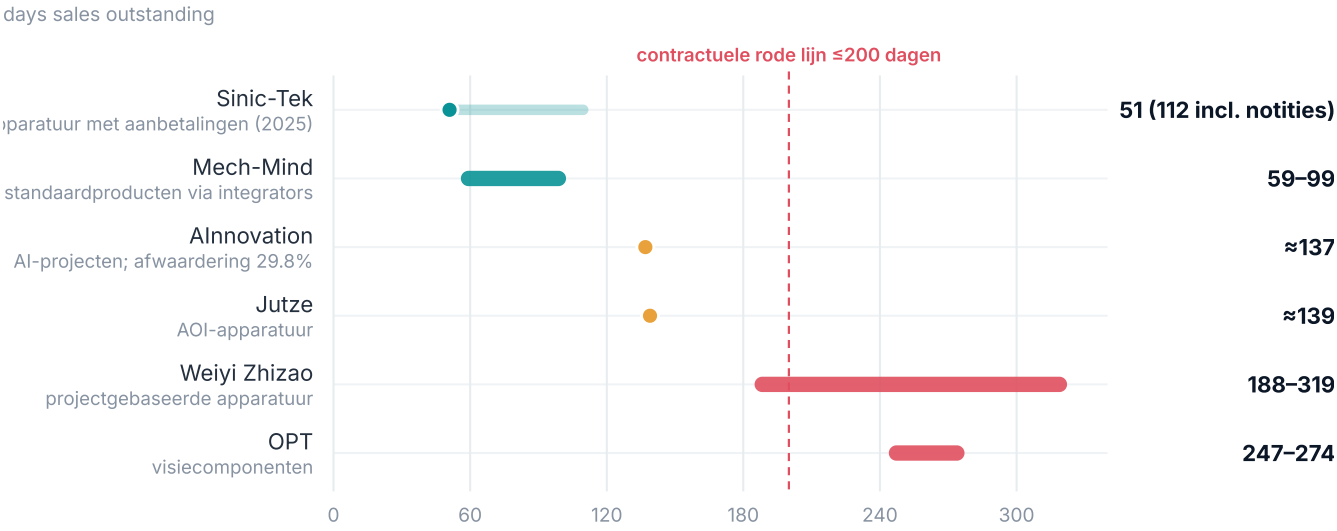
Acceptatie en proeven verlopen traag en worden door de verkoper betaald. De verzonden-maar-niet-geaccepteerde goederen van Sinic-Tek deden er ongeveer 143–158 dagen over om te rouleren, omdat grote EMS- en merkklanten hele-lijn-integratietests en een verlengde proefperiode nodig hebben (Sinic-Tek antwoord op vragen 8-1-2); de AI-projecten van Weiyi duren 4–8 maanden van evaluatie tot implementatie (Weiyi Zhizao prospectus) single sources, primair. Er zijn geen conversiedata voor AI-reviewpilots; Sinic-Tek converteerde slechts ongeveer 18–19% van zijn

demomachines rechtstreeks in verkoop. Ik reken op één op drie tot één op vijf voor gratis pilots en ongeveer 80% voor betaalde pilots na shortlisting estimate.

Bedrijf	Bedrijfsmodel	Dagen uitstaande verkoop
Sinic-Tek (2025)	Apparatuur met aanbetalingen	≈51 (≈112 inclusief nota's)
Mech-Mind	Standaardproducten via integratoren	59–99
Jutze	AOI-apparatuur	≈139
AInnovation (2025)	AI-projecten	≈137 einde jaar, met 29.8% bijzondere waardevermindering
Weiyi Zhizao	Projectgebaseerde apparatuur	188–319
OPT	Visiecomponenten	247–274

Bronnen: de jaarverslagen en prospectussen hierboven; vorderingen (inclusief contractactiva) ÷ omzet × 365 primair + estimate.

FIG 28 **Days sales outstanding** ***Standaardproducten 51–99 dagen; projectgebaseerde apparatuur 188–319 dagen.***



Bron: Bedrijfsstukken; vorderingen (incl. contractactiva) ÷ omzet × 365.

AI-reviewlicenties die via AOI-makers of integratoren worden verkocht, zouden moeten incasseren als de 60–100 dagen van Mech-Mind; directe projectverkoop als de 190–320 dagen van Weiyi. Samengevoegd als een cash-tijdlijn: gratis pilot één tot drie maanden, kwalificatie bij een grote ODM tot ongeveer twaalf maanden, bouw, levering en acceptatie drie tot zeven maanden; 30% komt met de bestelling, 60% ongeveer vier tot negen maanden erna en de laatste 10% ongeveer 16–21 maanden erna. **Van de eerste pilot bij een grote EMS tot volledige betaling kan twee tot drie jaar duren**; middelgrote EMS zonder formele aanbestedingen kunnen de kwalificatie tot weken verkorten estimate. De acceptatiecriteria in de openbaargemaakte contracten dekken alleen hele-lijn-integratie, doorvoer (één Taiwanees project vereiste 100 eenheden per uur) en maanden storingsvrij draaien; geen enkel specificeert doorglipper- of vals-alarmpercentages (Anda antwoord op vragen) single source, primair. Schriftelijke acceptatiecriteria voor AI-review moeten vóór de pilot worden overeengekomen.

Vier faalpatronen

Gedocumenteerde mislukkingen vallen in vier patronen: verzenden voordat het product betrouwbaar was (Rethink (The Robot Report)); een softwarelaag op de hardware van een leider die het ecosysteem van de leider later verving (READY Robotics (The Robot Report)); geld opmaken aan de onderkant van een cyclus (Dextrous Robotics (Ro-

bots & Startups)); en modelbedrijven die eindigen in acqui-hires (Covariant (GeekWire)). De lessen: productiseer smal in plaats van project voor project af te stemmen; houd werkkapitaal aan voor acceptatiegebaseerde omzet en lange incasso's; lever lokale service voor fabrieken die 24 uur draaien; en schaal niet op voordat het product is bewezen.

Timing: inspectie wint niets bij wachten; armhardware wordt goedkoper

De argumenten voor wachten zijn zwak. Hardware werd in 2026 niet goedkoper: robotmakers verhoogden prijzen met 5–15%, sommige producten 25–50% (Sina Finance) single source, en NVIDIA-kits stegen voor de derde keer (36Kr). Productiewaardige VLA's komen niet binnen het venster (IFR, Instrumental). Intussen verwijdert elke AOI-vervanging een retrofitdoelwit en legt elke OEM-AI-implementatie de data en het proces van een lijn vast; modelcapaciteit commoditiseert, en wat schaars is, is toegang tot fabrieken en gelabelde fabrieksdata. Lokale ondersteuning heeft ook vensters: de regels van Shenzhen 2026 voor "AI-pionierstad", met maximaal RMB 2 miljoen voor AI-toepassingsdemonstratieprojecten en modelvouchers die 30% van aankopen dekken, lopen tot **30 juni 2027** (Yicai, mei 2026) voorwaarden single source.

Wat nog snel goedkoper wordt, zijn humanoïden en behendige handen (handen daalden in twee jaar van ongeveer RMB 50,000 naar onder 10,000 (Sina Finance)), en instap-cobots naarmate de productmix naar beneden beweegt; dat pleit voor het uitstellen van het embodied-deel, niet van het inspectiedeel. **Algemeen: nu inspectie, als smal product; nu robotcellen, via partners en integratie; belichaamde AI later, met vandaag alleen een R&D-station.**

Belichaamde AI

06 Belichaamde AI: eerst verandert het hoe robots worden aangeleerd, later hoe ze bewegen

Wat verandert is perceptie en programmering; beweging blijft deterministisch

Elke robotarmleider market nu “fysieke AI”, maar wat is geleverd valt in drie lagen: geleerde perceptie zonder aanleren (AI Masterless Vision van FANUC, leverbaar vanaf januari 2027 (FANUC); neurale grijpselectie van KUKA (KUKA)); grote modellen die programma’s schrijven en taken plannen (iiQWorks.Copilot van KUKA, dat waarschuwt dat AI-gegenereerde code in een digitale tweeling moet worden geverifieerd (KUKA); Yaskawa met Gemini Robotics-ER, waar “generatieve AI beslist wat te doen” en de robot beslist “hoe te bewegen” (Yaskawa)); en reken- en dataplatforms (UR’s AI Accelerator met een Jetson Orin (Universal Robots)) vendor claims. Geen leider publiceert productieslagingspercentages voor zijn AI-functies, en de Japanse leiders werken samen met Google, SoftBank, Microsoft en anderen in plaats van zelf frontier-VLA’s te bouwen (FANUC-nieuwsindex): de geleerde beleidslaag komt van modelbedrijven, de gecertificeerde armen en controllers van de armfabrikanten.

FIG 29 **De embodied-AI-stack** ***Grote modellen plannen offline, kleine policies draaien in de cel, beweging en veiligheid blijven deterministisch.***



Bron: Synthese van de auteur (hoofdstukken 3 en 6); ISO 10218:2025; EU-Machinerichtlijn.

Hoe ver van ongestructureerd, flexibel werk in kleine series: gestructureerde stations in 1–3 jaar, flexibele assemblage 5+

Echte fabriekscijfers beginnen te verschijnen, maar ze zijn door leveranciers gerapporteerd, smal en afkomstig van gestructureerde stations. De acht humanoïden van AgiBot op de tablet-eindtestlijn van Longcheer logden 64,828 taken op 99.99% over meer dan 64 uur (AgiBot, Yicai Global) vendor en klant, maar dit is bedienwerk dat een conventionele arm met visie kan doen. UBTech zegt dat zijn Walker S2 op 30–50% van menselijke efficiëntie werkt (Irish Ti-

mes/FT) single source. De 11 maanden van Figure 02 bij BMW Spartanburg leverden volumecijfers op (1,250+ uur, 90,000+ onderdelen), terwijl de cyclus-, succes- en interventiecijfers doelen zijn, geen resultaten (Figure, BMW Group) verified, alleen tellers. Nucleus zegt dat ongeveer 40% van zijn fabrieksvideo wordt geteleopereerd (Interesting Engineering) vendor claim, en een Reuters-onderzoek noemde ongeveer 20% van de output van een arbeider voor sommige Chinese humanoïden (Notebookcheck) single source. Geen enkele implementatie publiceert slagingspercentage, cyclustijd en interventiepercentage samen, en geen enkele is gecontroleerd, terwijl juist die drie zijn wat een terugverdientijdberekening nodig heeft.

Implementatie	Slagingspercentage	Takt	Interventies of teleoperatie	Uptime of faalinterval	Volumerun
Figure 02 @ BMW Spartanburg	Alleen doel (>99% per ploeg)	Alleen doel (84 s)	Alleen doel (0 per ploeg)	Niet bekendgemaakt (onderarmtop faalpunt)	1,250 h
Agility Digit @ GXO	Niet bekendgemaakt	Niet bekendgemaakt	Niet bekendgemaakt	Niet bekendgemaakt	100,000 totes
Apptronik Apollo @ Mercedes-Benz	Niet bekendgemaakt	Niet bekendgemaakt	Niet bekendgemaakt	Niet bekendgemaakt	Niet bekendgemaakt
Agile ONE @ Audi	Niet bekendgemaakt	Niet bekendgemaakt	Niet bekendgemaakt	Niet bekendgemaakt	Niet bekendgemaakt
Nucleus (naamloze fabriek)	Niet bekendgemaakt	Niet bekendgemaakt	≈40% geteleopereerd (methode niet bekendgemaakt)	Niet bekendgemaakt	≈2 h video
AgiBot G2 @ Longcheer	99.99% van 64,828 taken	18–20 s per stap	Niet bekendgemaakt	Eén communicatiefout, ≈2 min herstel	64+ h
Neuromeka hybride cel	322/324	12.8% langzamer dan een persoon	Niet bekendgemaakt	Niet bekendgemaakt	324 cycli
Physical Intelligence π0.6 bij een klantlocatie	—	—	96.4% autonoom over een ploeg	Niet bekendgemaakt	Eén ploeg

Bronnen: dit hoofdstuk en hoofdstuk 3 meestal vendor claims.

FIG 30 **Wat in-plant robotuitrollen echt openbaar maken** ***De meeste maken alleen volume bekend; geen enkele maakt succespercentage, takt en interventiefrequentie samen bekend.***

	slagingspercentage	takt	interventies / teleoperatie	uptime / storingen	volumeproductie
Figure 02 @ BMW	alleen doel >99%/ploeg	alleen doel 84 s	alleen doel 0/ploeg	n/b	1.250 h
Agility Digit @ GXO	n/b	n/b	n/b	n/b	100.000 bakken
Apptronik Apollo @ Mercedes	n/b	n/b	n/b	n/b	n/b
Agile ONE @ Audi	n/b	n/b	n/b	n/b	n/b
us (niet nader genoemde fabriek)	n/b	n/b	≈40% telebediend*	n/b	≈2 h video
AgiBot G2 @ Longcheer	99.99%	18–20 s per stap	n/b	één storing, 2 min	64+ h
hybride cel van Neuromeka	322/324	12.8% langzamer	n/b	n/b	324 cycli
π0.6 @ klantlocatie	—	—	96.4% autonoom	n/b	1 ploeg

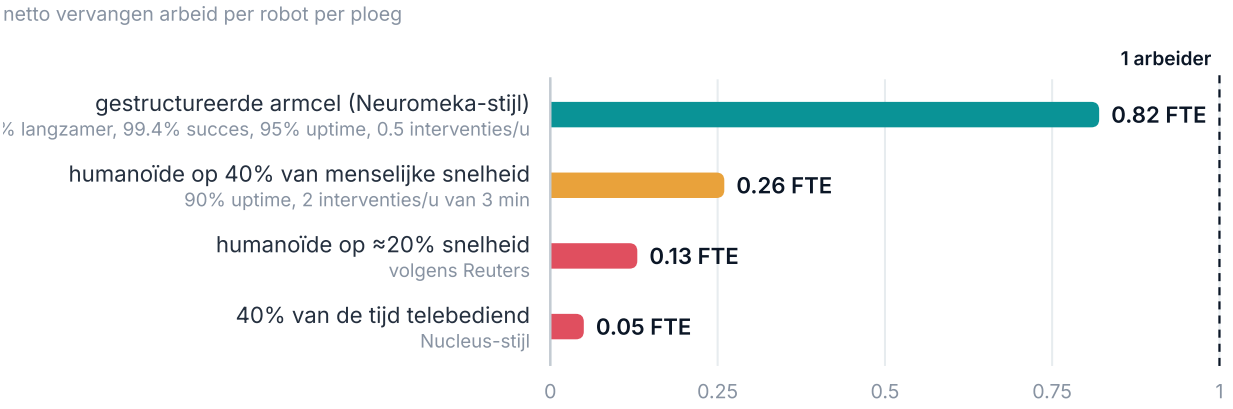
* methode niet bekendgemaakt

bekendgemaakt (meestal claims van leveranciers) alleen doelen n/b

Bron: Publiek materiaal van BMW, Figure, Agility, Mercedes-Benz, Audi, AgiBot, Neuromeka, Physical Intelligence en anderen.

Teleoperatie telt zwaarder voor de economie dan het slagingspercentage. Netto vervangen mensen per robot per ploeg ≈ (menselijke cyclus ÷ robotcyclus) × uptime × succes, minus interventietijd en teleoperatie estimate, mijn formule. Een gestructureerd station in Neuromeka-stijl (12.8% langzamer, 99.4% succes, 95% uptime, 0.5 interventies van 2 minuten per uur) vervangt ongeveer 0.82 mensen; een humanoïde op 40% van menselijke snelheid, 90% uptime en twee interventies van 3 minuten per uur ongeveer 0.26; op de 20% van Reuters, ongeveer 0.13; en als, zoals Nucleus, één operator hem 40% van de tijd teleopereert, ongeveer 0.05, vrijwel niets. Bij ongeveer RMB 600k per humanoïde en RMB 85k per persoon per jaar op het vasteland verdienen de laatste drie zich over decennia terug; alleen het gestructureerde armstation heeft een realistische terugverdientijd (hoofdstuk 4) estimate.

FIG 31 **Netto vervangen arbeid per robot** ***Een gestructureerde armcel vervangt ongeveer 0.82 van een werknemer; een humanoïde die 40% van de tijd telebediend wordt, ongeveer 0.05.***



Bron: Formulegebaseerde schattingen van de auteur.

Buiten gestructureerde stations scoren algemene modellen nog 40–90% op nieuwe taken: Gemini Robotics 2 haalt 74.2% op lab-pick-and-place, 78.9% op gereedschapskitting en 44% op het knopen van een zak (eesel/DeepMind); op RoboChallenge Table30, 30 echte-robottaken, scoort π0.5 43.7% en het beste open model 50.3% (RoboChallenge, QbitAI).

Externe prognoses verschillen, maar geen enkele ondersteunt wijdverbreid gebruik binnen twee jaar. De IFR zegt dat algemene fysieke AI “nog niet industrieklasse” is en dat de meeste humanoïden “nog prototypes of pilots” zijn

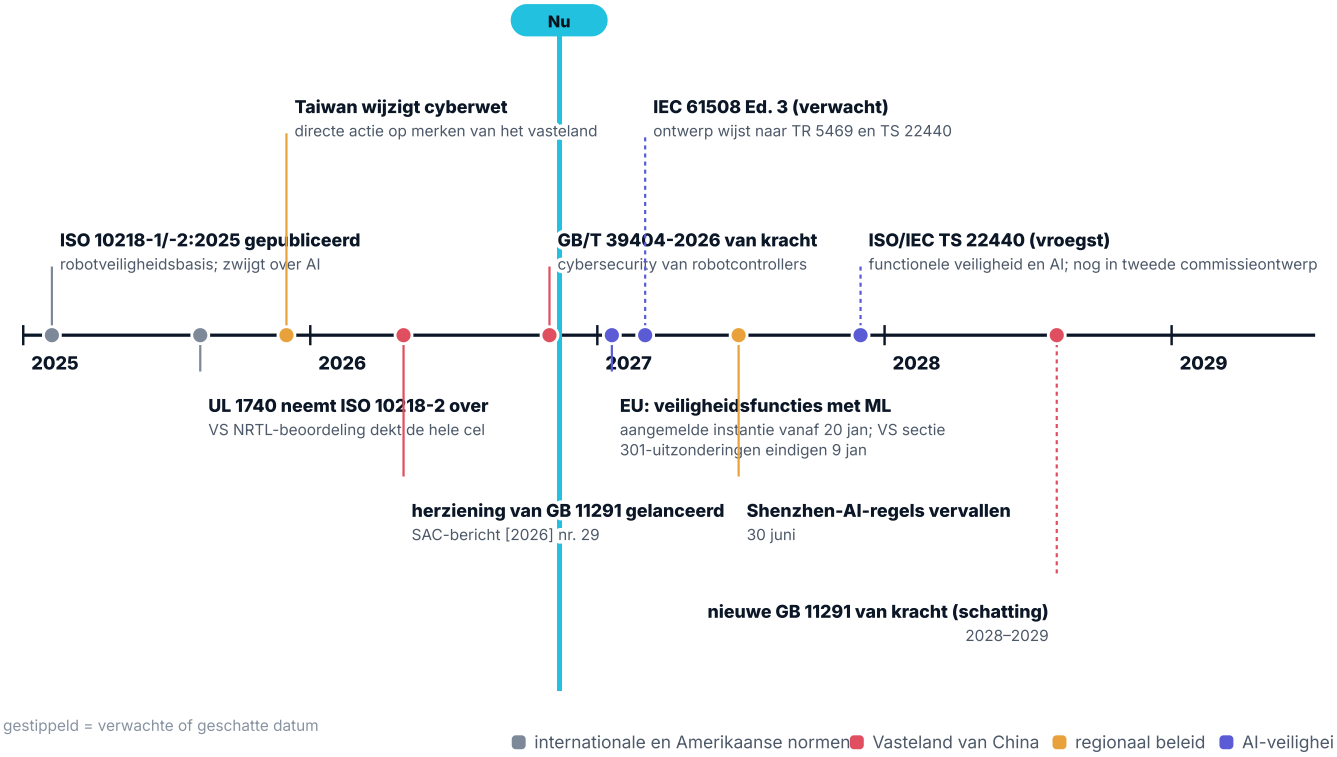
(IFR); Interact Analysis schat dat slechts ongeveer 10% van de in 2025 geproduceerde humanoïden in echt gebruik is gegaan, met het commerciële keerpunt “na 2032” (automation.com); UBS acht een humanoïde-“EV-moment” vóór 2030 onwaarschijnlijk (Yicai Global); Chinese insiders schatten nog twee tot drie jaar vóór fabrieksschaalgebruik (Gasgoo) single sources. Mijn oordeel estimate: in goed gefinancierde elektronica- en autofabrieken met de leverancier ter plaatse zullen **gestructureerde stationtaken** (bedienen, OK/NG-sorteren, gefixeerde insertie) in ongeveer 1–3 jaar ≥99% bereiken op bijna menselijke snelheid; **kitting van gemengde onderdelen** met algemene VLA’s in ongeveer 3–5; **vrij geleide kabelbomen en meerstaps flexibele assemblage** in meer dan 5. Humanoïde-terugverdientijd werkt vandaag niet: een Walker S2 op ongeveer RMB 600k (10jqka) single source, op 30–50% van menselijke efficiëntie over twee ploegen, staat gelijk aan ongeveer 0.6–1.0 arbeiders, een eenvoudige terugverdientijd van 7–12 jaar tegen vasteland-loonkosten vóór integratie, onderhoud en stilstand estimate.

Waar het vastzit

Schakel	Bewijs in 2026
Data	De industrie schat dat ten minste 10 miljoen uur hoogwaardige echte data nodig is en dat er begin 2026 pas ongeveer 500,000 was verzameld (Gasgoo); Reuters meldt ongeveer 500,000 uur tegen ongeveer 100 miljoen nodig (Notebookcheck); een beginnende teleoperator heeft ongeveer 300 pogingen per bruikbare beweging nodig; fabrieken delen procesdata niet; simulatiegetrainde beleidsregels zakken scherp op echte robots
Betrouwbaarheid	Gestructureerde stations 98–99.99% (vendor claims, falen ongedefinieerd); algemene modellen 40–90% op nieuwe taken; ongeziene connectoren 29–67%
Snelheid	Humanoïden op 30–50% van menselijke snelheid (UBTech), sommige op ongeveer 20% (Reuters); frontier-agents ongeveer 280 s per quasistatische echte-robottaak; LLM-regelkringen op 0.2–0.4 Hz tegen ongeveer 83 Hz nodig (Anthropic)
Kosten	Walker S2 gemiddeld ongeveer RMB 600k; terugverdientijd 7–12 jaar estimate
Veiligheidscertificering	ISO 10218-1/-2:2025 maakt functionele veiligheid explicieter en absorbeert samenwerkingseisen, maar zegt niets over AI (The Robot Report); ISO/IEC TS 22440 (functionele veiligheid en AI) was in november 2026 nog een tweede commissieontwerp (ISS, ISO Open Data); IEC 61508-3-richtlijnen ontraden AI boven SIL 1 (Analog Devices); onder de EU-machineverordening vereisen veiligheidsfuncties met machine learning vanaf 20 januari 2027 een beoordeling door een aangemelde instantie (EUR-Lex)
Handen en gereedschap	Behendige handen halen ongeveer 30 N (≈3 kg) per vinger tegen ongeveer 10 kg voor een mensenhand (Gasgoo); gereedschapswissels zijn nog handmatig
Integratie	Sensor-mislezingen, protocolmismatches en MES-koppelingen ter plaatse; interventiepercentages bijna nooit bekendgemaakt

De veiligheidsimplicatie is direct: de normen verbieden geleerde beleidsregels niet, noch erkennen ze deze. ISO 10218-1/-2:2025 zegt niets over AI; ISO/IEC TS 22440 (functionele veiligheid en AI) zit nog in commissieontwerp en is onwaarschijnlijk vóór eind 2027 of 2028 estimate; ISO/IEC TR 5469:2024 is alleen richtlijn en beveelt niet-AI-fallbacks, toezichthoudende beperkingen en redundante stemming aan; IEC 61508 editie 3 wordt begin 2027 verwacht (ISO Open Data, BSI, 61508 Association) verified. Geen gepubliceerde norm biedt nog een geaccepteerde route om machine-learningperceptie als veiligheidsfunctie te certificeren. Ook regionale regels bewegen: China herziet GB 11291 (van kracht rond 2028–2029) (herzieningsplan 20262543); UL 1740 heeft ISO 10218-2 sinds augustus 2025 in de Amerikaanse NRTL-review gebracht; de EU vereist vanaf 20 januari 2027 een beoordeling door een aangemelde instantie van veiligheidsfuncties met machine learning; en geen humanoïde is gecertificeerd voor hekwerkloos fabrieksgebruik zolang ISO 25785-1 nog in ontwerp is (i-SCOOP). De verdedigbare architectuur: machine learning neemt alleen procesbeslissingen (“welk onderdeel”, “is deze plaat goed”), terwijl een gecertificeerde deterministische veiligheidslaag (doorgaans PL d / Categorie 3) en een veiligheids-PLC snelheid, kracht en stops bepalen. Behandel elke claim van “camera-AI voor hekwerkloze veiligheid” als onbewezen totdat een PL d- of SIL 2-certificaat de machine-learningcomponent noemt.

FIG 32 **Tijdlijn van normen en regelgeving** ***Een certificeringstraject voor AI binnen veiligheidsfuncties is vóór 2027–2029 niet uitgekristalliseerd.***



Bron: ISO, IEC, UL, EUR-Lex, SAC en nationale regelgevingsdatabases.

Welke laag de kans is: modellen horen bij de reuzen, edge-boxes bij de chipmakers, kleine bedrijven bij integratie en sensorhardware

De modellaag heeft kapitaal en data nodig en hoort bij de reuzen (AgiBot trainde GO-2 op “tienduizenden uren” (iFeng Tech); de sterkste modellen zijn gesloten of alleen cloud). De edge-boxlaag hoort bij de chipmakers: Jetson Thor is de feitelijke standaard, en op het vasteland concurreren de RDK-boards van D-Robotics (D-Robotics) en Rockchip-coprocessors (CSDN), meestal zonder gepubliceerde VLA-resultaten.

Kleine bedrijven hebben twee openingen. De ene is sensorhardware die betrouwbaarheid rechtstreeks inkoopt: tast tilt het insertiesucces van ContactDP van 58.7% naar 97.8%, en kracht/koppel-sensoren en polscamera’s doen vergelijkbaar werk; datacollectiekits verkopen ook (10jqka), hoewel handen, sensoren en armen allemaal in prijs dalen. De andere is de integratielaag tussen armfabrikanten en modelbedrijven, die per station de taakdata, de veiligheidsenvelop en de MES-koppeling bezit, wat armfabrikanten vandaag niet als product verkopen. Concurrenten vallen in vier groepen: armleiders (FANUC, Yaskawa, KUKA, ABB, UR, JAKA, Dobot en andere), visie-plus-armbedrijven (Mech-Mind, Hikrobot, SmartMore), embodied-AI-bedrijven (AgiBot, UBTech, Unitree, Galbot, Figure, Physical Intelligence, Google DeepMind, NVIDIA) en toetreders uit de autochips.

Koperacceptatie: geen cijfer zonder noemer

Aangezien geen leverancier interventiepercentages, faalintervallen of langetermijn-succestellingen publiceert, horen deze in de acceptatievoorwaarden van het contract. Statistisch bewijzen 3,000 opeenvolgende productiecycli met nul fouten een faalkans van hoogstens 0.1% bij 95% betrouwbaarheid; 10 fouten in 3,000 bewijzen slechts ≤0.57%; het bewijzen van ≤1%, ≤0.5%, ≤0.1% en ≤0.01% met nul fouten vergt 299, 598, 2,995 en 29,956 cycli estimate, exact binomiaal. Naar die norm heeft de 322/324 van Neuromeka een tweezijdig 95%-interval van 97.8–99.9% en bewijst het geen ≥99%; de ruwweg 6 fouten van AgiBot in 64,828 taken impliceren 99.98–100%, maar alleen als “falen” wordt gedefinieerd zoals de koper het definieert estimate. Dit zijn de acceptatie-eisen die ik aan elke robot-

arm- of embodied-cel zou stellen; de drempels putten uit de eigen doelen van Figure bij BMW en uit de Neurome-ka- en AgiBot-cases, en het zijn oordelen, geen norm estimate :

Item	Eisen	Voorgestelde slaagdrempel
Definities	Ondertekende definities van cyclus, succes, falen (gedeeltelijk succes, her-grijpen, laten vallen, kwaliteitsdoorglipper), interventie (aanraking, teleoperatie op afstand, reset, noodstop) en hoe autonomie wordt berekend	Definities komen vóór elke metriek; geen percentage zonder formule
Interventies en teleoperatie	Per robotuur, naar type en duur; gemiddelde tijd tussen interventies; teleoperatieminuten als aandeel van de draaitijd; operators per robot	Gestructureerde stations ≤0.5 interventies per robotuur (≥2 h uit elkaar); nul teleoperatie tijdens acceptatie
Langetermijnsucces	≥3,000 opeenvolgende productiecycli over ≥3 ploegen, ≥2 artikelnummers en normale belichtings- en leveringsvariatie; fouten geclassificeerd; exacte ondergrens	Ondergrens ≥99.5% voor lijnstoppende inlinestations, ≥99.0% offline; kwaliteitsdoorglippers apart gerapporteerd
Cyclustijd	P50/P90/P99 per onderdeel inclusief herstel, tegen ≥30 getimede menselijke cycli op hetzelfde station	≤1.15 × menselijke P50, of binnen de lijntakt
Uptime en fouten	Draaiuren, downtime-log, fouten per component, onderdeelwisseltijd	≥95% uptime over ≥2 weken; MTTR ≤20 min met reserveonderdelen ter plaatse
Herstel	Gescripte foutinjectie: laten vallen, occlusie, belichtingswijziging, bandstop, MES- of netwerkuitval, stroomuitval, noodstop	≥90% autonoom herstel, binnen 2 min terug in productie
Ombouw	Uren, demonstraties, operatorvaardigheid en datacollectiekosten per nieuw artikelnummer of opstelling	Gestructureerde stations: ≤1 ploeg en ≤100 demo's per nieuw artikelnummer
Veiligheid	Arm en cel naar ISO 10218-1/-2:2025, celrisicobeoordeling, lokale regels (GB 11291 op het vasteland; CNS 14490 plus schriftelijke beoordeling in Taiwan; ANSI/A3 R15.06-2025 en UL 1740 in de VS)	Geen veiligheidsfunctie geïmplementeerd door machine learning; een gecertificeerde veiligheidslaag (doorgaans PL d / Cat 3) beperkt snelheid en kracht; geen hekwerkloze legged robots totdat ISO 25785-1 is gepubliceerd en beoordeeld
Rekenkracht en offline gedrag	Inferentiehardware, P99-latentie, gedrag zonder netwerk, modelversionering, hervalidatie na elke update	Takt-kritieke inferentie hangt niet af van de cloud; volledige her-acceptatie na elke model-, kwantiserings- of enginewijziging
Bewijs	Onbewerkte video en ruwe logs voor het hele acceptatievenster; waargenomen door het kwaliteitsteam van de klant of een derde partij	Geen highlightreels of alleen-leverancier-samenvattingen
Commerciële voorwaarden	Betalingsmijlpalen gekoppeld aan acceptatiemetrieken; beschikbaarheidsboetes wanneer per robotuur wordt gefactureerd; regels voor data-eigendom en toegang op afstand	Laatste betaling en retentie pas na de 3,000-cyclusacceptatie en ≥4 weken stabiel draaien
Economie	Netto vervangen mensen = snelheidsratio × uptime × succes − interventietijd − teleoperatieaandeel	Wijs elk voorstel af waarvan de netto-vervanging de geclaimde terugverdientijd niet kan dragen

Bronnen: (Figure, Neuromeka, Yicai Global, i-SCOOP); de drempels zijn mijn suggesties.

Open vragen

07 Wat de openbare documentatie niet kan beantwoorden, en hoe een pilot het beslecht

De vragen die openbare bronnen niet kunnen beantwoorden, vallen in vijf groepen: waarvoor AI-review werkelijk verkoopt en wat kopers willen betalen; hoe AOI-makers AI bundelen en oude machines upgraden; hoe oud de geïnstalleerde AOI's zijn en hoeveel er al AI draaien; hoeveel mensen en uren review werkelijk per lijn kost; en hoe pilots converteren en hoe kopers betalen. Daarbovenop komt hoe de nieuwste modellen, gekwantiseerde builds en robotarmen werkelijk presteren onder de omstandigheden van een bepaalde fabriek. Elk kan worden omgezet in een test met een duidelijke drempel. De onderstaande drempels zijn beslissingsregels die ik uit het bewijs in dit rapport heb afgeleid estimate ; herbereken ze zodra de eerste fabrieksdata binnenkomen.

Open vraag	Hoe te beslechten	Hoe een goed resultaat eruitziet
Prijs en bereidheid om te betalen voor AI-review	Een standaard offerteaanvraag (bijvoorbeeld vier lijnen met gemengde oudere 2D en nieuwere 3D-AOI's, twee ploegen, ongeveer 2,000 platen per ploeg, 12 defectklassen, alleen on-premises, MES- of IPC-CFX-integratie) verstuurd naar meerdere derden-leveranciers; een prijsgevoeligheidsonderzoek van vier vragen onder 20–30 EMS-fabrieken	De meeste offertes binnen RMB 40–120k eenmalig of 20–40k per jaar voor vier lijnen; ten minste de helft van de fabrieken accepteert RMB 10–30k per lijn met terugverdientijd binnen 12 maanden
AI-bundeling en upgrades van AOI-makers voor oudere machines	Korte gestructureerde interviews met ongeveer een dozijn AOI-makers en distributeurs	Of AI-review standaard, optioneel of afwezig is op 2025-26-modellen; of pre-2020-machines kunnen worden geüpgraded en tegen welke prijs
Leeftijd van geïnstalleerde AOI's en aandeel dat AI draait	Enquête ter plaatse bij ongeveer 30 EMS-fabrieken, gestratificeerd naar grootte en regio, met apparatuurlijsten beoordeeld onder NDA	Mediane AOI-leeftijd ≥5 jaar en ≥60% zonder AI betekent dat de retrofitpool reëel is
Reviewarbeid en -uren	Interviews en 2–4 weken AOI-alarm-, review-, reparatie- en ploegenlogs bij twee of meer fabrieken, met werksteekproeven	≥1 toegewijde reviewer per lijn per ploeg en ≥3 review-uren per lijn per ploeg, met ten minste 0.5 FTE per lijn die de fabriek wil herplaatsen
Prestatie van AI-review zelf	Golden set plus schaduwmodus op een echte lijn (hoofdstuk 3): dataovereenkomst in week 0, nulmeting en labelen in weken 1–2, schaduw draaien in weken 2–4, optioneel geassisteerd draaien in weken 4–8	Valse oproepen of review-uren ≥50–60% verlaagd; doorglippergrenzen per klasse niet slechter dan de menselijke nulmeting; nul missers op kritieke klassen; p95-latentie ≤½ takt; automatisch fail-open naar mensen
Leraar-labelkwaliteit van frontier-modellen	Draai twee of drie kandidaten op dezelfde golden set, elk drie promptformuleringen, één goed referentiebeeld, losse component-uitsnedes ≥42–48 px	Automatische acceptatie alleen voor klassen met nul missers op ≥299 defecten en overeenstemming op menselijk niveau; anders alleen pre-labelen en triage
Gekwantiseerde builds	Wijzig alleen de precisie; pin gewicht-hashes, template, voorbewerking, decoding, engine en batchgrootte; schakel prefix-caching uit of reset het; controleer dat de referentiebuild leveranciersresultaten reproduceert; scoor paren op de golden set en driftslice	Geen nieuwe missers op kritieke klassen; ≤1 nieuwe misser in 1,000 defecten; opruiming ≤2 punten omlaag; ≥99.5% identieke beslissingen over drie runs; anders terugvallen op FP8/INT8

Open vraag	Hoe te beslechten	Hoe een goed resultaat eruitziet
Lijnzijdige hardware, reinheid en ESD	≥2 eenheden van elke kandidaat-box stressgetest op tweemaal het verwachte alarmpercentage; 72-uurs soak bij echte kastluchtstroom; ESD-controle door de coördinator van de klant; ISO 14644-14-test voor klanten op ISO 7 en schoner	p95 ≤½ takt; geen crashes of throttling over 72 uur; ESD-aftekening; deeltjesrapport in de juiste klasse
Betalingsvoorwaarden, pilotregels en conversie	Betalingsclausules uit recente inspectieapparatuurcontracten bij de onderzochte fabrieken; interviews over pilotregels, acceptatiecriteria en kwalificatietijd	Ongeveer 30/60/10 met cash binnen 60 dagen na acceptatie; pilots ≤8–12 weken met een benoemde aftekeningseigenaar; ≥1 op 3 proeven converteert; een raamwerk dat ≥5–10 lijnen dekt in jaar één
Data, IT en aansprakelijkheid	Interviews en contracttoetsing met juridisch adviseur	Een lokale box mag verbinding maken en data benaderen; aansprakelijkheid beperkt tot softwaredefecten
Regionale compliance vóór vertrek naar het buitenland	Taiwan: schriftelijke status van de klant onder de Cyber Security Management Act, E187-zelfbeoordeling; VS: douane-uitspraak over hoe een cel wordt geclassificeerd, NRTL-offertes, ITAR-controle; Zuidoost-Azië: BOI-local-linkage en Maleisische aftrek-eligibiliteit, moeder van de klant; vasteland: CCC-reikwijdte voor gebundelde pc's en voedingen	De klant is geen gecontroleerde entiteit, en de terugverdientijd blijft ≤2 jaar met invoerrecht en certificering inbegrepen
Acceptatie van een robot- of embodied-cel	De 12-item acceptatietabel in hoofdstuk 6: ondertekende definities, ≥3,000 opeenvolgende productiecycli over ≥3 ploegen en ≥2 artikelnummers, gescripte foutinjectie, ≥30 getimede menselijke cycli, waargenomen	Ondergrens ≥99.5% inline (≥99.0% offline); cyclus ≤1.15 × mens; ≤0.5 interventies per uur en nul teleoperatie; ≥95% uptime; MTTR ≤20 min; ≥90% herstel
ACT- en VLA-beleidsregels op Chinese edge-chips	Vier weken: porten en micro-benchmark (≥1,000 inferenties, latentieverdeling, vermogen, soak van 60 minuten); gesloten-lus-simulatie met gemeten latentie geïnjecteerd; echte-robot-A/B tegen een GPU; duurtest van 8–24 uur (≥3,000 cycli)	ACT-chunks p95 ≤100 ms, VLA ≤300 ms; echte-robotsucces binnen 3 punten van de GPU en cyclus ≤1.2 × mens; anders het beleid op Jetson of een GPU houden
Cobot-, AMR- en integratieprijzen	Model-voor-model-offertes van meerdere cobotdistributeurs, en volledige-celoffertes van twee of drie integratoren voor testopstellingbelading (één station, twee ploegen)	Volledige cel ≤RMB 200k en terugverdientijd ≤1.5 jaar over twee ploegen in Shenzhen
SMT-lijntellingen in Taiwan, Zuidoost-Azië en de VS	IPC-data (voorheen New Venture Research) of de geïnstalleerde basis van lokale apparatuurdistributeurs	Genoeg adresseerbare lijnen per regio om lokale applicatie-engineering te rechtvaardigen

Van al deze bepalen er vier de meest gevoelige cijfers in dit rapport: de prijs die kopers willen betalen, de arbeid die AI-review werkelijk vrijmaakt, of AI-review werkelijk niet meer mist dan mensen, en hoe snel cash terugkomt. Ze zijn het waard om vóór elke significante toezegging te worden beslecht, door een fabriek die een aankoop overweegt of door een leverancier die een product overweegt.

Een pilotreeks die voor beide kanten werkt

Voor een fabriek, en voor een leverancier die een geloofwaardige referentie wil, is de volgorde hetzelfde:

1. **Weken 0–2: overeenstemming over data en definities.** Onderteken de dataovereenkomst, definieer de defectklassen en wat als doorglipper telt, en leg een nulmeting vast: valse oproepen per plaat, reviewseconden per oproep, downstream gevonden doorglippers.

2. **Weken 1–4: bouw de golden set** uit het AOI-archief en bewaarde NG-platen, blind gelabeld en beoordeeld (hoofdstuk 3).
3. **Weken 2–4 (of langer): schaduwmodus.** Het model logt zijn beslissingen terwijl mensen beslissen zoals voorheen; vergelijk per klasse.
4. **Weken 4–8: geassisteerd draaien op willekeurige ploegen,** waarbij het model alleen klassen opruimt die de poort hebben doorstaan en al het overige naar mensen gaat.
5. **Acceptatie, schriftelijk, vóór opschaling:** de resultaten per klasse, latentie, fail-open-gedrag, en wie modellen en labels bezit. Schaal lijn voor lijn op, en valideer opnieuw bij elk nieuw product, elke modelversie, kwantiserings- of enginewijziging.

Beperkingen van dit rapport

Dit is deskresearch. Ik heb voor deze editie geen primaire interviews met fabrieken afgenomen; veel lijndata komt van leveranciers en is niet onafhankelijk gecontroleerd. Verschillende overheidswebsites, aanbestedingsdatabases en wervingsplatforms waren niet toegankelijk, dus de subsidies voor de upgrade van Shenzhen 2026, de loonrichtlijn van de Jangtsekiangdelta, transactieprijzen voor AI-review, reviewers per lijn en pilotconversiepercentages zijn schattingen of analogieën; elk verschijnt hierboven als een open vraag. Wisselkoersen zijn de centrale koers van de People's Bank of China op 9 november 2026 ($1 = RMB 6.7188) en, voor Taiwan, de fixing van SMBC op dezelfde dag ($1 = NT$32.12); de Vietnamese dong gebruikt data van mei 2026. Alle drempels zijn beslissingsregels afgeleid uit het geciteerde bewijs, geen normen.

Over de auteur

Tan Shuai

Tan Shuai is een technoloog en oprichter met meer dan twintig jaar werk in chips, hardware en systeemsoftware, en 20 verleende patenten. Zijn werk voerde hem langs Silicon Valley, New York, Chicago, Taiwan, de Greater Bay Area (Shenzhen en Hongkong), Vietnam en Maleisië, met een langdurige focus op geavanceerde productie en toeleveringsketens. Vandaag werkt hij als AI-onderzoeker en -ontwikkelaar, gericht op hoe AI werkelijk op de fabrieksvloer wordt gebruikt.

Hij verwelkomt gesprekken met mensen die aan AI in de productie werken, of het nu gaat om het opbouwen van een team, het zoeken van een partner of het uitwisselen van aantekeningen: i@tanshuai.com · tanshuai.com

Opmerkingen en disclaimer

Dit rapport is onafhankelijk deskresearch op basis van openbare documenten die beschikbaar waren tot 11 november 2026. Het wordt uitsluitend ter informatie verstrekt en is geen investerings-, juridisch, fiscaal of technisch advies. Productnamen, prijzen, modelreleases, exportclassificaties, invoerrechten en regelgeving veranderen snel; verifieer alles waarop u vertrouwt tegen de primaire bron. Cijfers gemarkeerd met vendor claim of estimate moeten als zodanig worden gelezen. Bedrijfs- en productnamen behoren toe aan hun eigenaren en worden uitsluitend ter identificatie gebruikt; geen enkel bedrijf heeft dit rapport beoordeeld of gesponsord.

© 2026 Tan Shuai. U mag dit rapport ongewijzigd delen met bronvermelding.

ELEKTRONICAPRODUCTIE · AI-INSPECTIE × ROBOTICA

AI-visuele inspectie en robotarmen in elektronicafabrieken: beoordeling 2026

Elk kerncijfer draagt zijn bron en een betrouwbaarheidstag: verified, single source, vendor claim of estimate. Wat de openbare documentatie niet kan beantwoorden, is verzameld in hoofdstuk 7, met hoe een pilot elke vraag beslecht.

Tan Shuai
tanshuai.comi@tanshuai.com

OPENBARE EDITIE · GEGEVENS PER 11 NOVEMBER 2026

