ELECTRONICS MANUFACTURING · AI INSPECTION × ROBOTICS

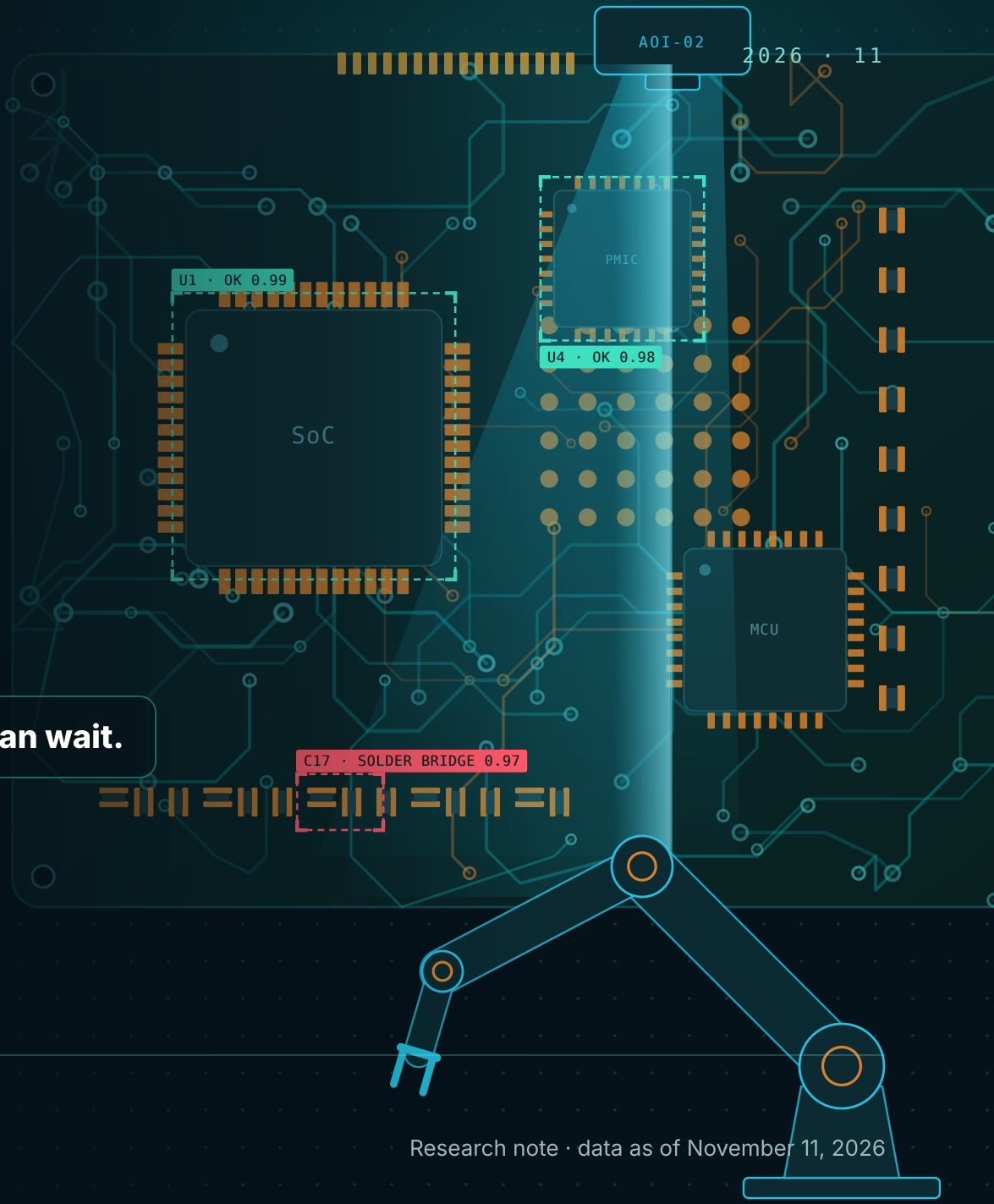
AI on the SMT line:

what pays back in 2026

AI visual inspection and robot arms in electronics factories

Mainland China · Taiwan · Southeast Asia · United States

KEY FINDING **AI review is worth doing, narrowly. Embodied arms can wait.**



Tan Shuai

tanshuai.com • i@tanshuai.com

Research note · data as of November 11, 2026



Worth doing, but narrowly, now and with partners

KEY FINDING

What is mature in 2026, and what plants will pay for, is AI review and niche inspection on AOI images plus programmed robot machine tending, not "a large model looking at boards plus embodied robots picking them".

0.46–0.65 yr

AI review payback in the US, Taiwan and Shenzhen, at RMB 50k per line

≥95% · <20 PPM

A top EMS's in-house AI review: the bar for selling to tier-1 EMS

100–700 RMB M

Serviceable third-party retrofit market in mainland China, one-off, ≈5-year window

0 scores

Industrial-defect benchmark scores for the November 2026 flagships

1.1–1.5 yr

Payback of a cobot machine-tending cell, two shifts

0.003–0.14 FTE

Labour a pick-out robot arm saves per shift: it never pays back

Source: report executive summary and Chapters 4–5; payback and market figures are the author's estimates.

Six questions, six answers

① What plants run

3D SPI, pre- and post-reflow AOIs, sampled X-ray, ICT/FCT. A person reviews behind every AOI; takt ≈13–22 s per board.

② How far the tech goes

Inspection is industrial-grade as detector + AI review + a few people; not zero people at any price. Mainland: domestic models; Taiwan: the customer decides; US: tariffs and certification.

③ On-premises and air-gapped

Practical; the limits are procedural. A ≤100 W fanless box at the line, servers in the server room, 8-bit by default.

④ Teacher and student

Inspection: frontier models pre-label, people adjudicate, a golden set gates. Arms: the teachers are still human demos and simulation.

⑤ Cost and payback

AI review within a year in the US, Taiwan and Shenzhen; cobot tending 1.1–1.5 years; Southeast Asia 2–6 years; a pick-out arm never.

⑥ Now or wait

Real demand, crowded supply; OEM AI leaves a ≈5-year retrofit window. Inspection now, narrowly; embodied AI after 2027–28 data.

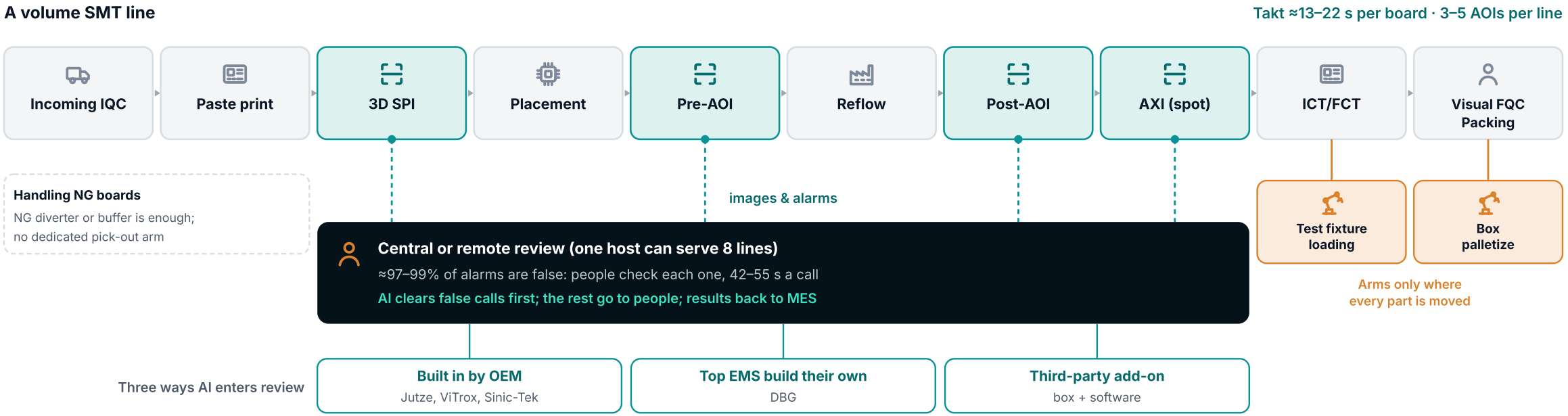
Source: report executive summary.

Where AI fits on the factory floor today



Source: Payback from the author's model (Chapter 4); maturity axis is the author's judgement.

The people sit behind the AOI. That is where AI enters



Source: Public filings and vendor material (Huaqin, DBG, Sinic-Tek and others); compiled by the author.

About 97–99% of AOI alarms are false calls

For every 100 AOI alarms



Today: people check every call

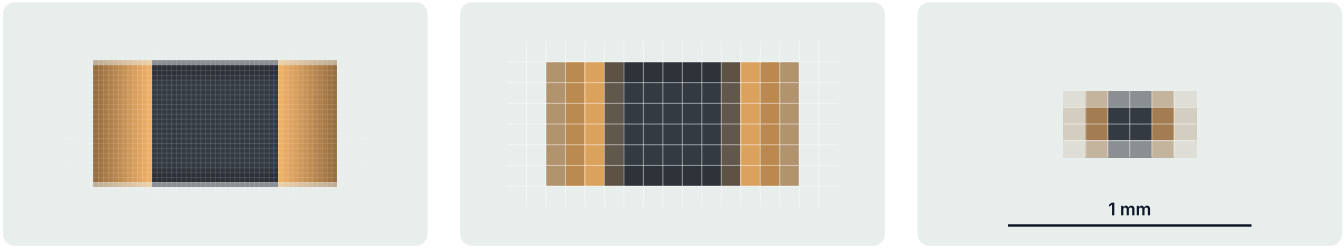
42–55 s each
≈448 false calls per shift on one busy line
→ 5.2–6.8 h per shift ≈ 1 full-time reviewer
case staffing: 2 per shift

AI review: clear false calls, pass the rest to people

False calls or review hours cut 60–90% (vendor claims)
Top-EMS in-house benchmark: covers ≥95% of flagged parts, escapes <20 PPM, human review down ≥85%
fail-open to people, line keeps running

Source: Siemens case study, iFactory (vendor claim), DBG 2025 annual report; review hours estimated by the author.

A general-purpose camera is not an AOI



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

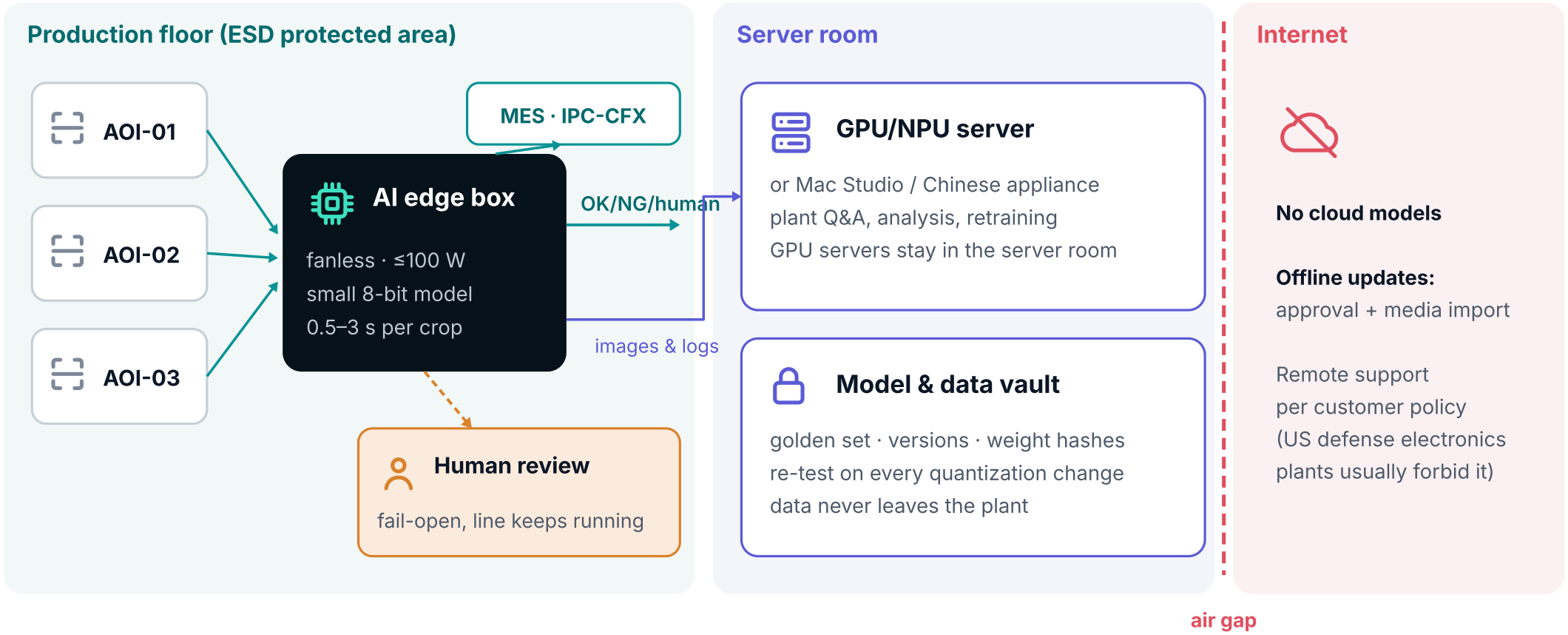
4K camera over a 300 mm panel: ≈78 µm/px
0402 shrinks to ≈13×6 px: enough for presence and gross offset only

Same camera on an 01005
≈5×3 px: effectively invisible

- **AOI:** 8–20 µm per pixel, multi-angle multi-colour lighting and structured light
- **4K over a 300 mm panel:** ≈78 µm per pixel; an 0402 is ≈13×6 pixels, enough for presence, not for joints
- **01005:** ≈5×3 pixels, effectively invisible
- **So:** the AI works on the AOI's own images; it does not replace the optics

Source: Circuits Assembly; author's calculation for a 300 mm panel and a 4K sensor. Pixel grids are illustrative.

An air-gapped deployment: the box decides, the server room answers

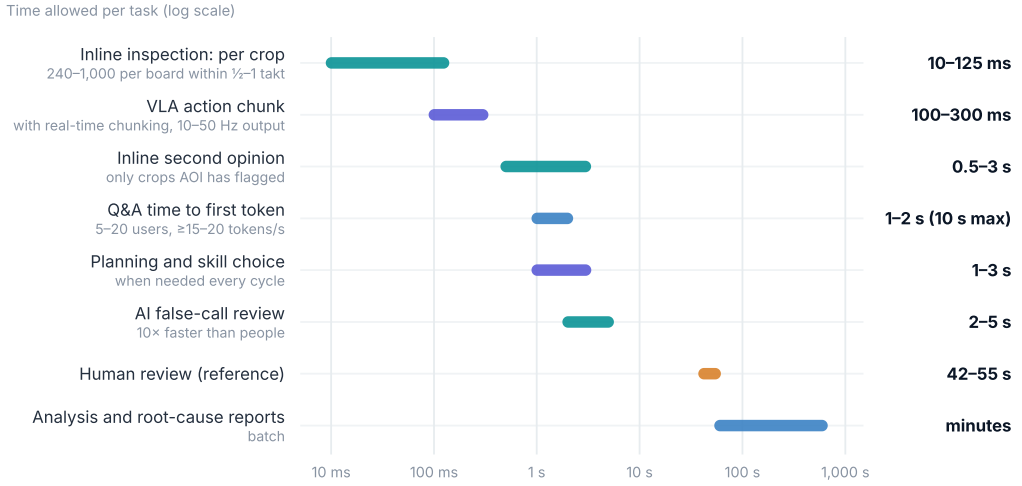


Source: Author's synthesis (Chapter 2).

Clean and ESD zones decide where compute sits; speed limits depend on the use

Cleanliness →

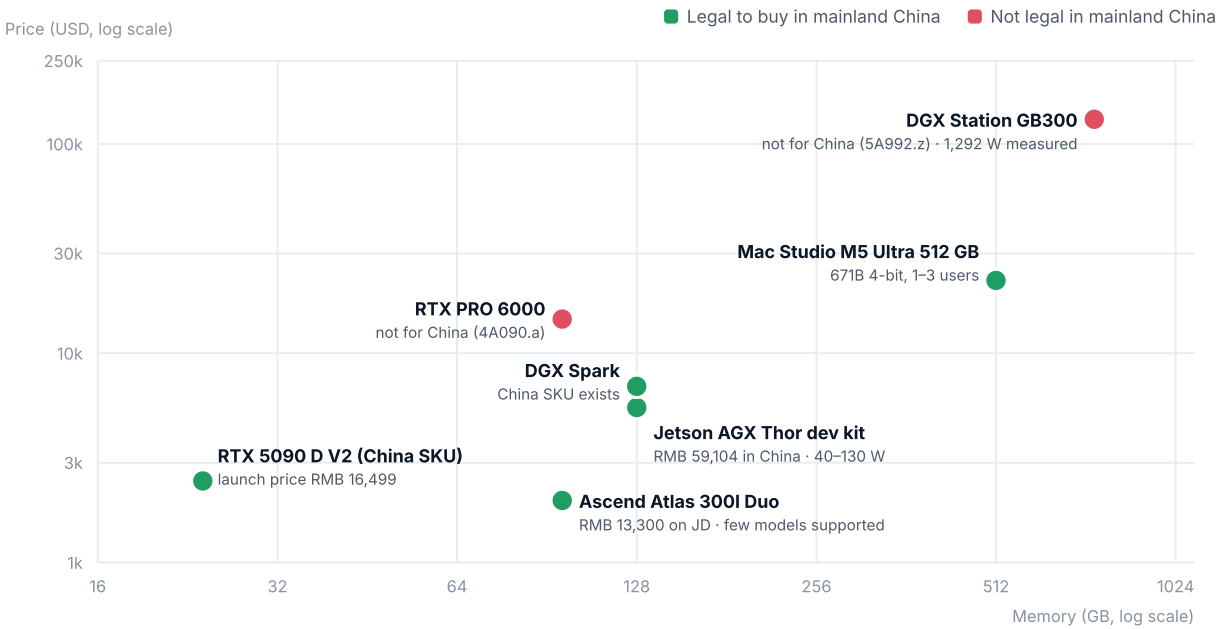
Ordinary SMT/PCBA hall	Clean room ISO 7–8	Mini-environment ISO 5–6	Fab, ISO 5 or cleaner
ESD protected area (≈ISO 8) RH 40–60%	medical, camera modules, EV clean room + ESD	bare die, wire bond, optics mini-environment	SEMI E78 ESD · E187/E188 cyber fab-specific plans
Line-side compute Fanless, grounded metal box next to the AOI or review seat	**Line-side compute** Fanless, sealed IP65/66, easy to wipe; ask for ISO 14644-14 data	**Line-side compute** Cameras and lights only; compute box outside	**Line-side compute** Only tool-integrated, fab-qualified controllers
GPU/NPU server Server room or air-conditioned IT rack	**GPU/NPU server** outside the clean room, networked in	**GPU/NPU server** Server room	**GPU/NPU server** Sub-fab liquid-cooled racks or data center



Key point An ordinary SMT hall is an ESD protected area, not a graded clean room: a fanless box in a grounded metal case sits at the AOI or review seat, and GPU servers never go on the line.

Source: ANSI/ESD S20.20, IEC 61340-5-1, ISO 14644, SEMI standards and vendor material; compiled by the author. Author's synthesis (Chapters 1–2).

Export rules pick the hardware short list before benchmarks do

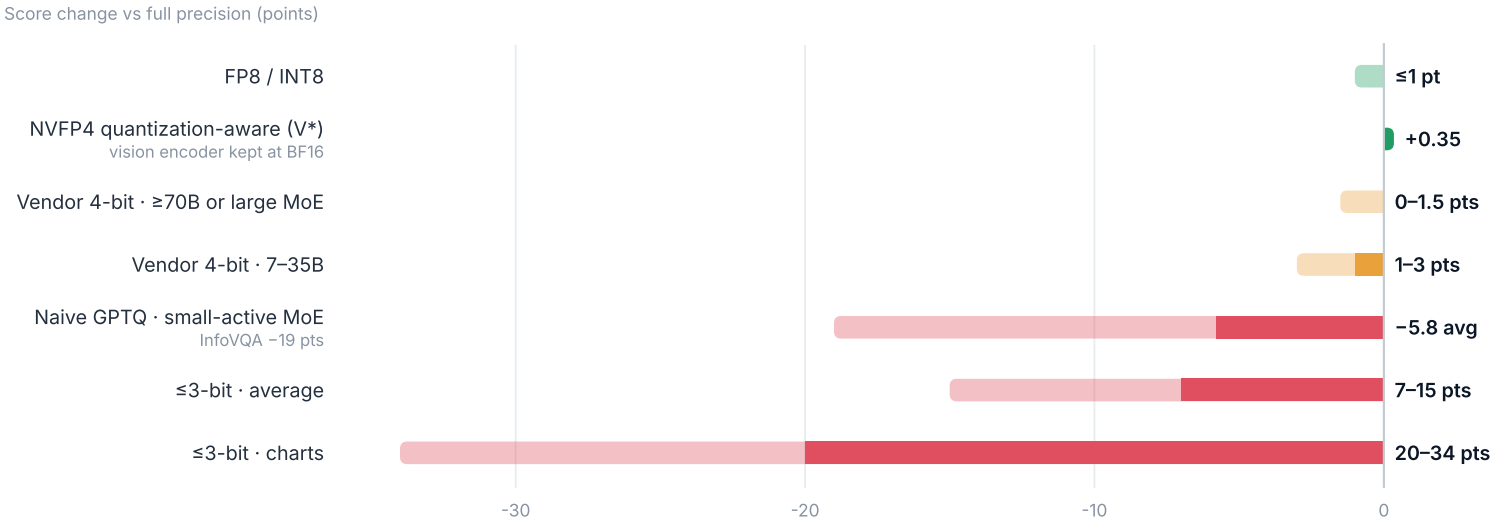


- **Jetson Thor:** 128 GB, 40–130 W; 2–8B vision models on AOI crops, 3B-class VLAs in real time; RMB 59,104 in China
- **Mac Studio 512 GB:** DeepSeek 671B at 4-bit for 1–3 users
- **DGX Station:** 5A992.z, not for the mainland; a 743B model gives 32 users ≈4 tokens/s each
- **Domestic accelerators:** two eight-card Ascend servers for full-size multi-user Q&A

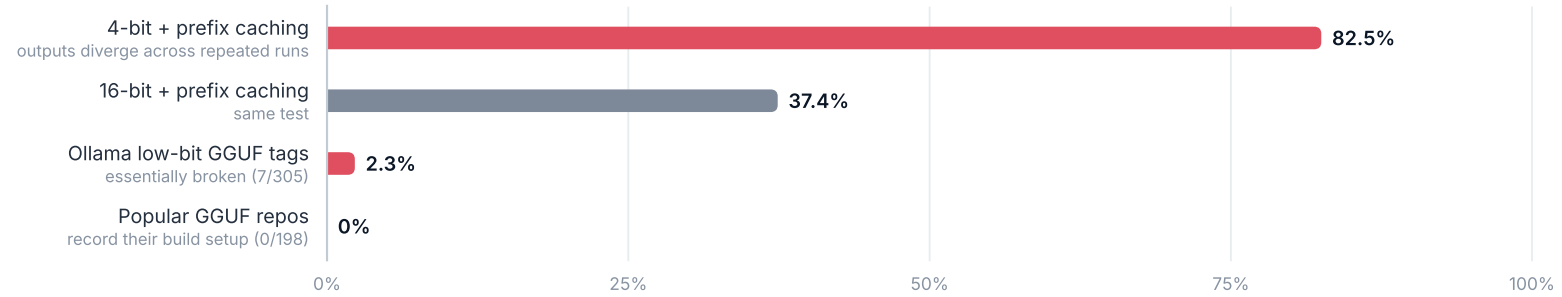
Source: NVIDIA export classifications, Apple, RS China, JD.com and review prices (November 2026). StorageReview (Aug and Oct 2026); ServeTheHome. Single-source independent tests.

Quantization: 8-bit is safe, 4-bit depends on the recipe, 3-bit and below is out

Accuracy: 8-bit is safe, 4-bit depends on the recipe, ≤3-bit is out



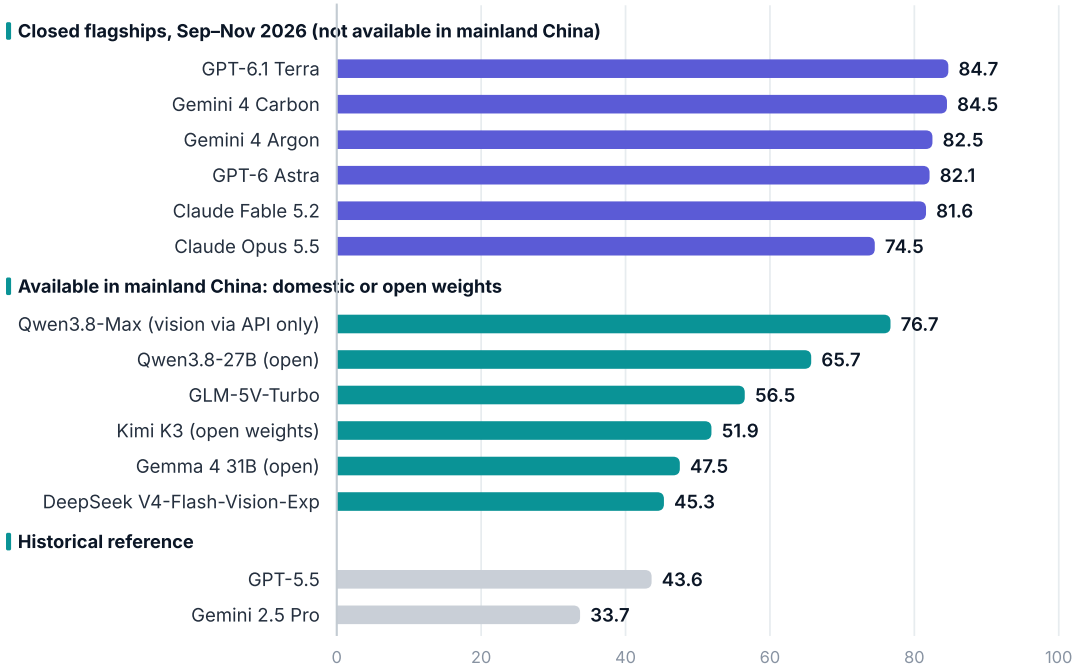
Engineering failures outweigh quantization error



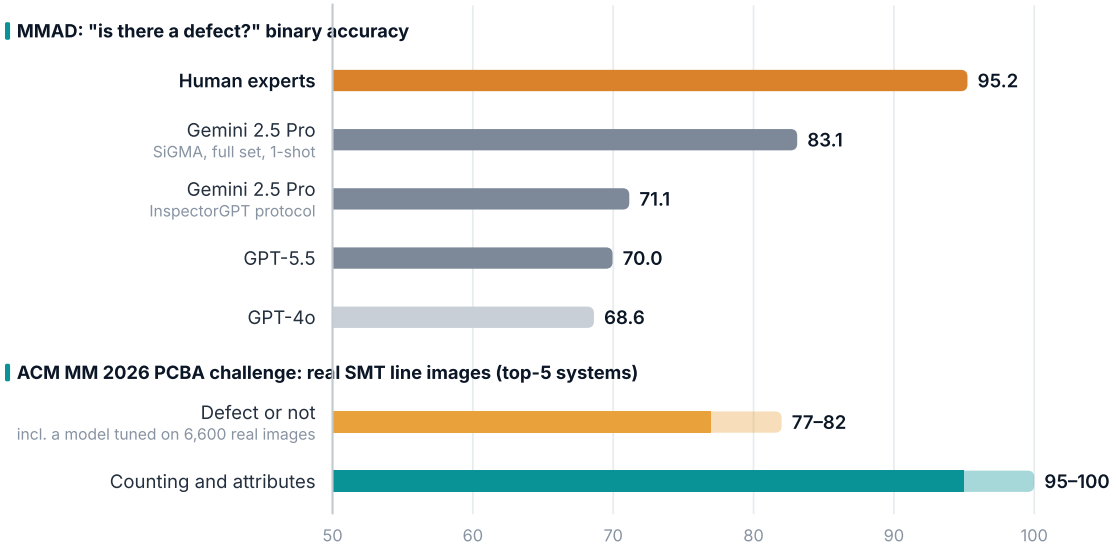
Source: NVIDIA Nemotron model card, VLMQ, VEQ, arXiv 2609.05881, 2610.19498, 2609.04748 and others.

Generic vision doubled in a generation; industrial-defect scores are missing

Roboflow generic detection mAP@50



Accuracy or score (%)



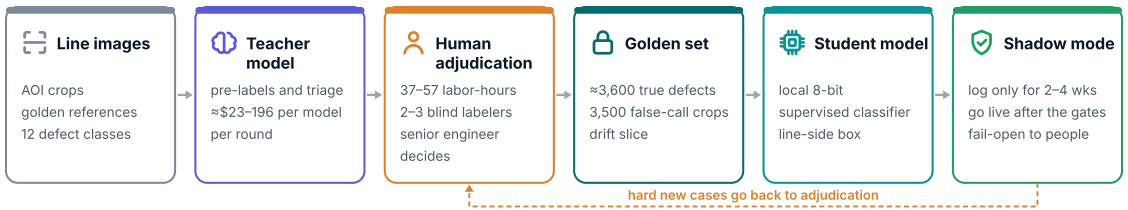
Industrial defect benchmarks (MMAD, MVTec AD 2, VisA, Real-IAD, UniPCB): zero published scores for these 2026 flagships

Generic detection scores doubled in a generation, but that is not defect judgement; test on your own golden set

Source: Roboflow Vision Evals; model release notes; MMAD, MVTec AD 2 and other benchmark pages (as of 11 Nov 2026). MMAD; SiGMA; InspectorGPT; ACM MM 2026 PCBA challenge.

Frontier models pre-label, people adjudicate, a golden set gates

Inspection: the frontier model pre-labels, people adjudicate, a golden set gates



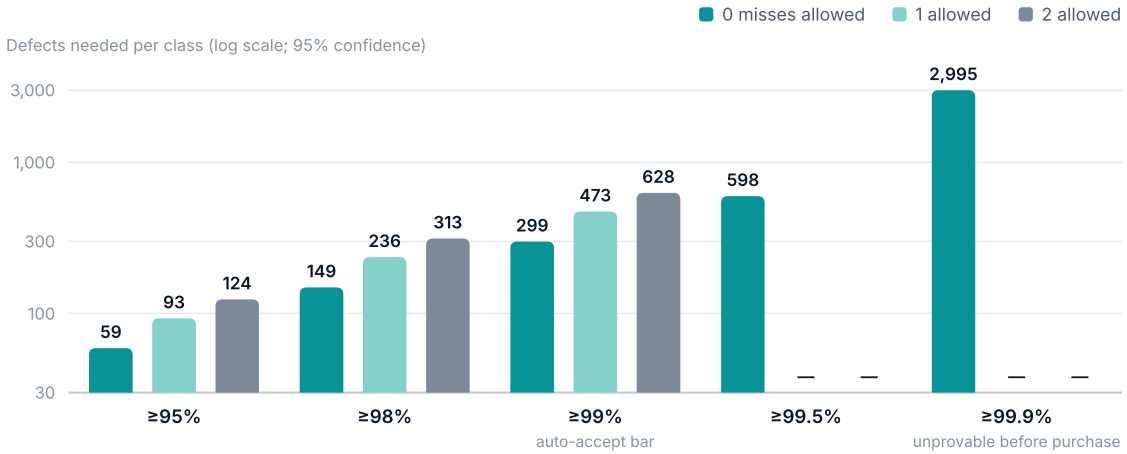
Bar for accepting teacher labels automatically

Per class: ≥299 defects, zero misses (recall ≥99% at 95% conf.), agreement no worse than human–human; else pre-label only

Arms: the teachers are still human demos, real-robot RL and simulation

Frontier models help only offline: programming and changeovers



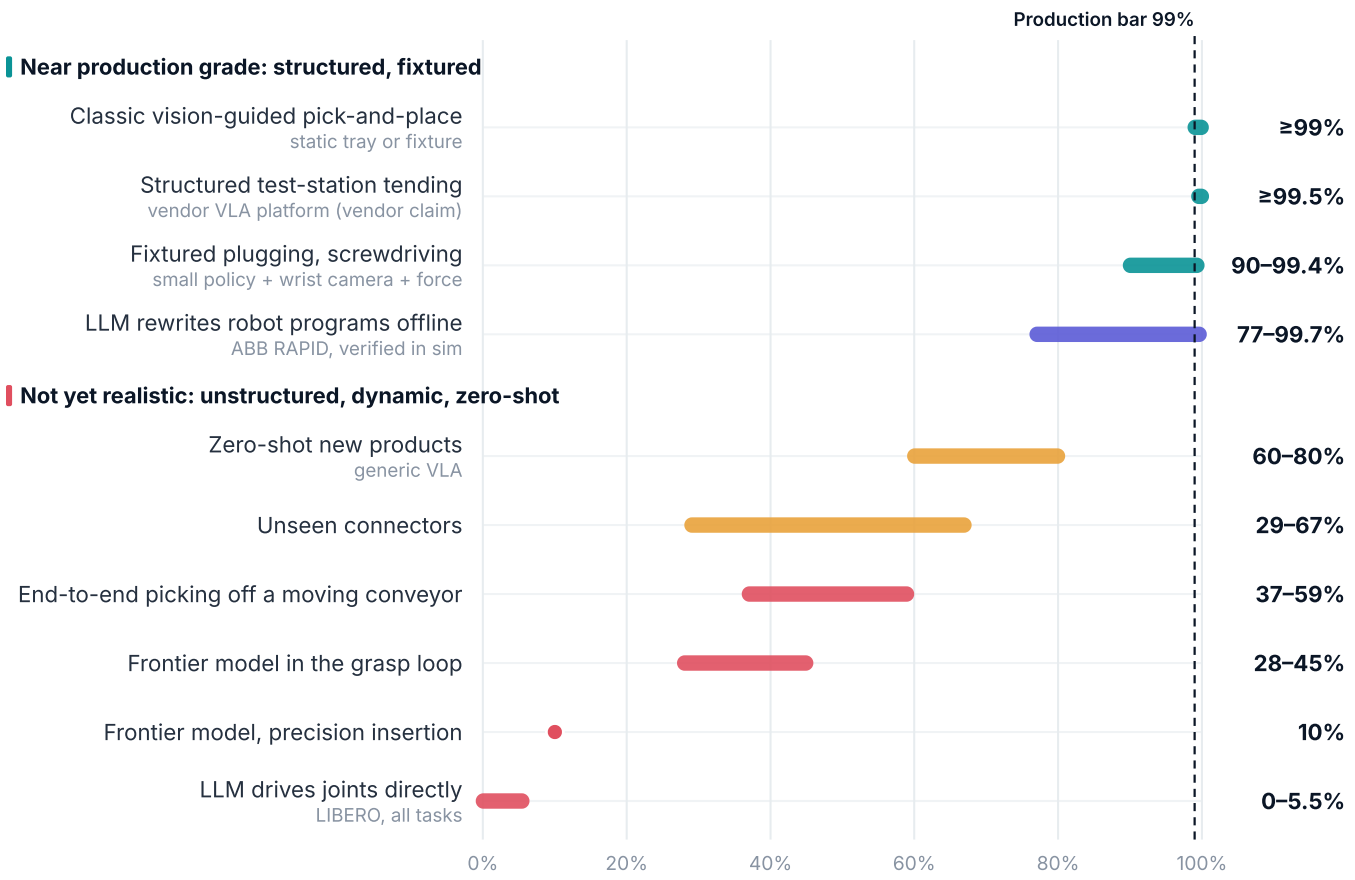


Key point Auto-accept the teacher's labels only for a class with zero misses on ≥299 golden defects (99% recall at 95% confidence); critical classes such as shorts, missing parts and polarity need more.

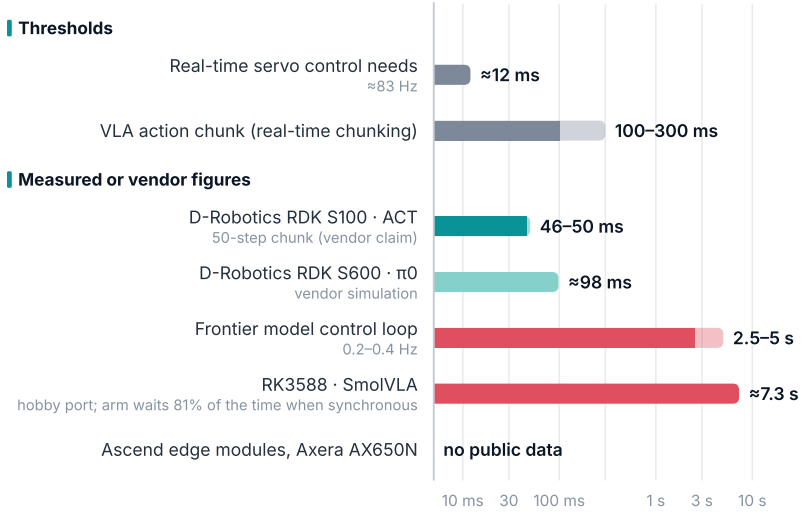
Source: Author's synthesis (Chapter 3); costs and sample sizes are estimates. Exact binomial confidence intervals (95%).

Robot students: near production grade only when the station is structured

Task success rate

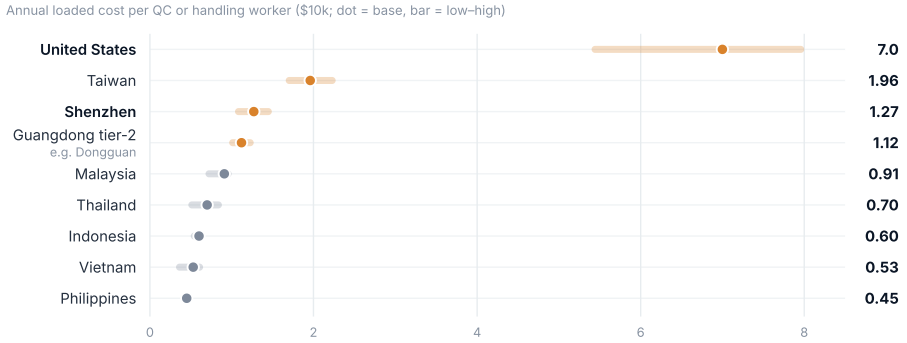


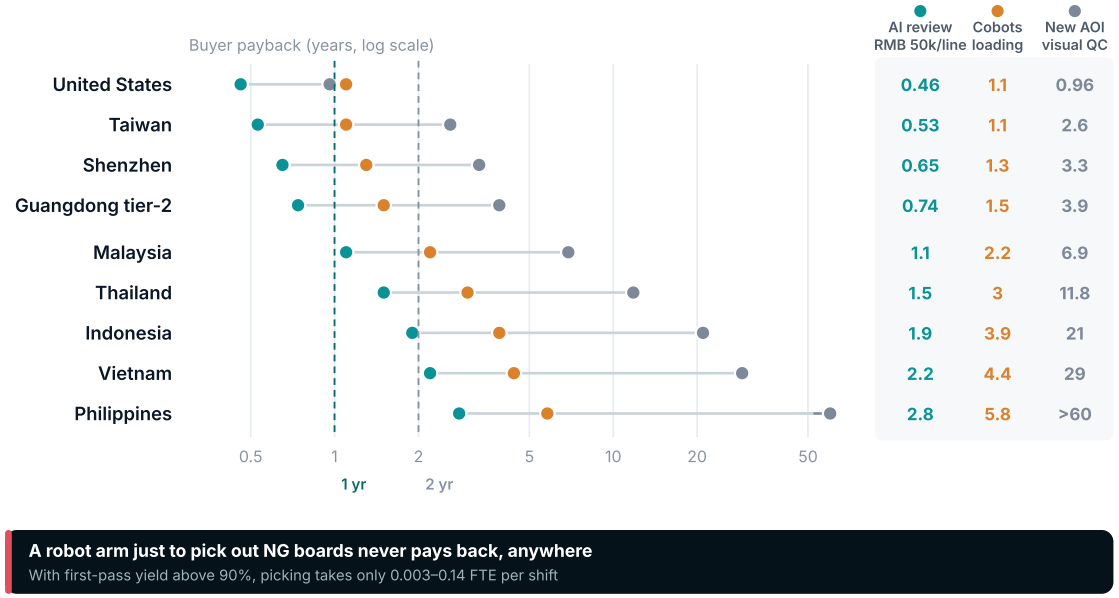
Inference latency per action chunk or control cycle (log scale)



Source: Papers and vendor material cited in Chapter 3; some figures are vendor claims. D-Robotics documentation and forum (vendor figures); GitHub community ports; Anthropic research.

AI review pays back in months; machine tending in about a year



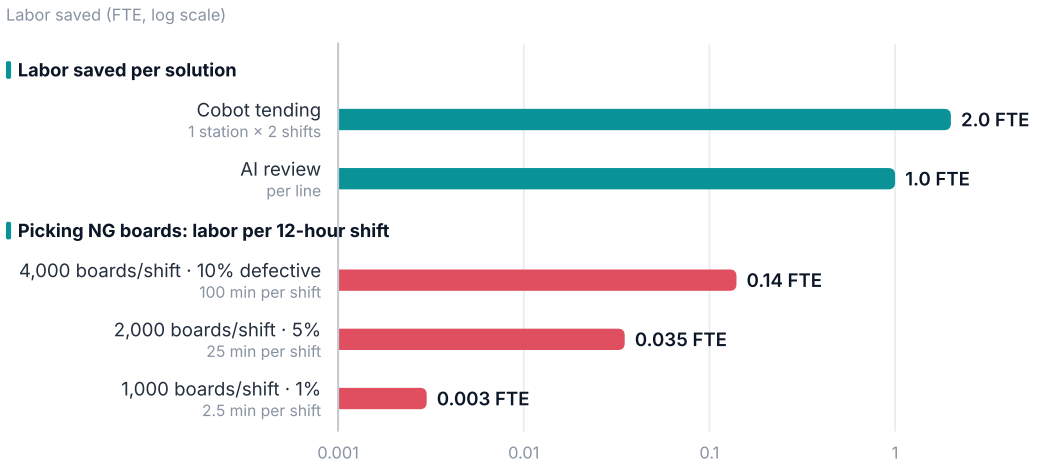


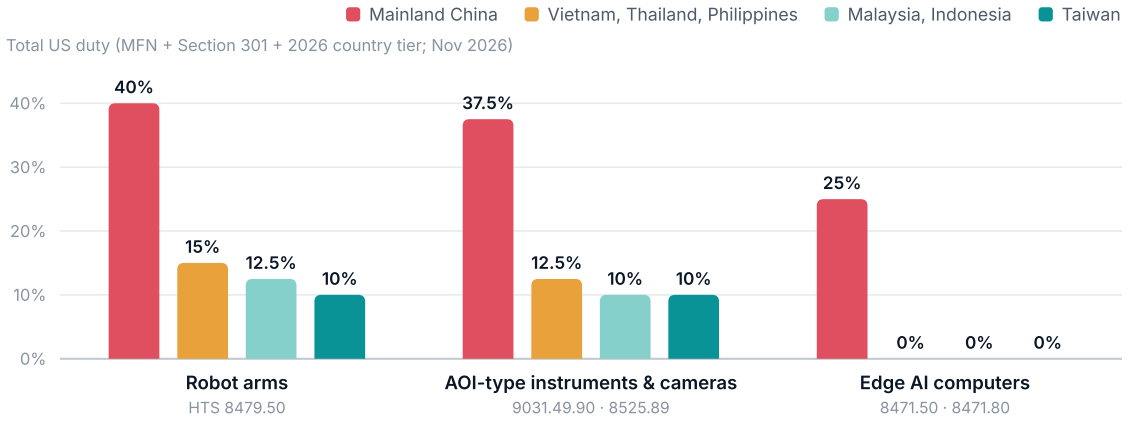
A robot arm just to pick out NG boards never pays back, anywhere
With first-pass yield above 90%, picking takes only 0.003–0.14 FTE per shift

Key point At RMB 50,000 per line, AI review pays back in 0.46 years in the US, 0.53 in Taiwan, 0.65 in Shenzhen and 2.2 in Vietnam; a two-shift cobot tending cell in 1.1–1.5 years [estimates].

Source: Shenzhen official wage guidance, Taiwan Ministry of Labor, JETRO, BLS; author's estimates. Author's model: RMB 50k AI review box per line; cobot tending cell at RMB 200k or $140k, two shifts.

A pick-out arm never pays back; where an arm is made changes its US duty





Key point Above 90% first-pass yield, picking out NG boards takes 0.003–0.14 of a person per shift, so an NG diverter or buffer does the job; a mainland-made arm also pays about 40% combined US duty.

Source: Author's estimates. USITC Harmonized Tariff Schedule chapter 99; Federal Register; USTR (November 2026).

Real but small: 16,000 serviceable lines, ≈RMB 320M one-off, a five-year window

China SMT lines
40,000 lines
35k–46k

× 80% × 50%
no AI yet: 75–85%
reachable: 40–60%

Serviceable
16,000 lines
10.5k–23.5k

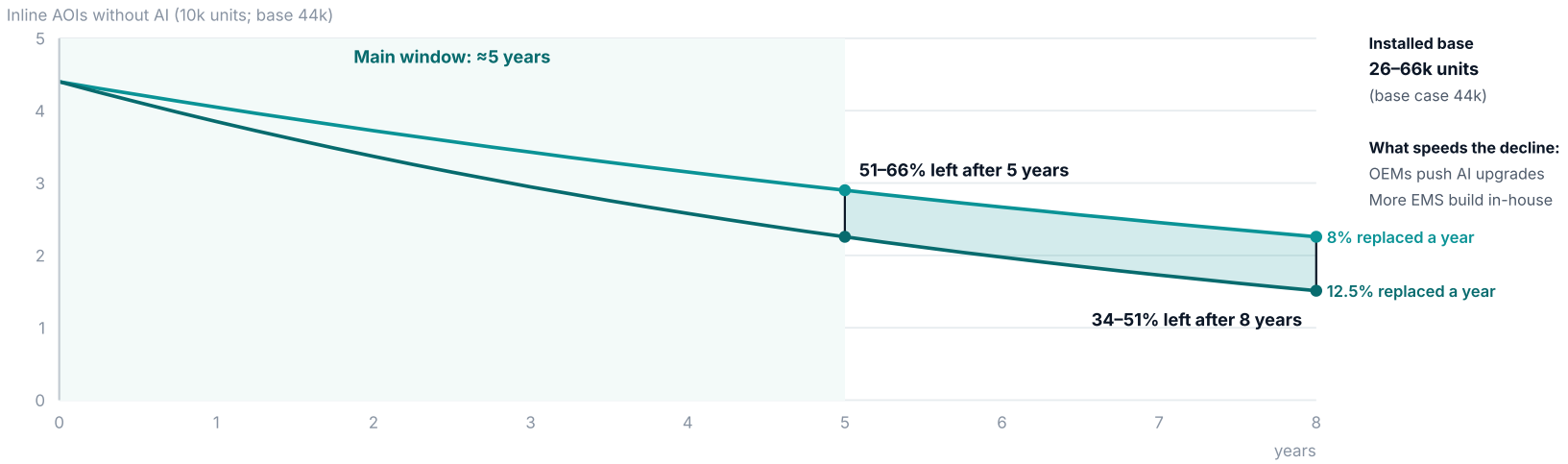
× RMB 20k
RMB 10–30k one-off

One-off market
RMB 320M
RMB 105–705M

÷ ≈5 years
shrinks 8–12%/yr

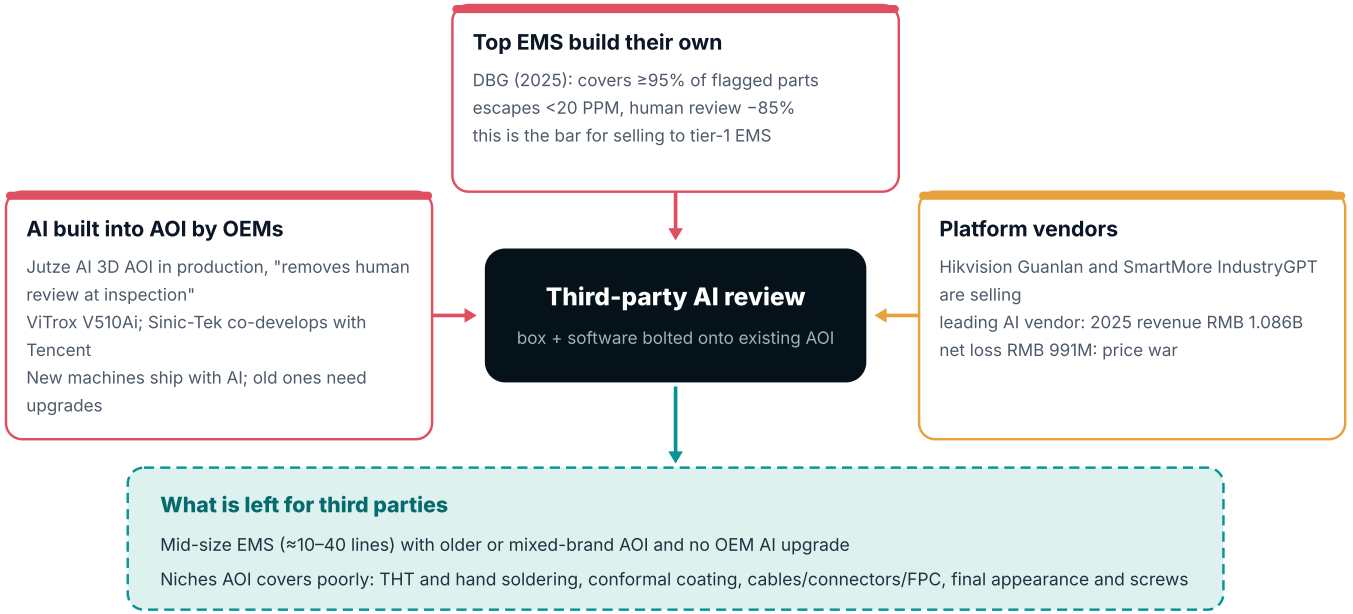
5-year window
RMB 64M/yr
RMB 21–141M

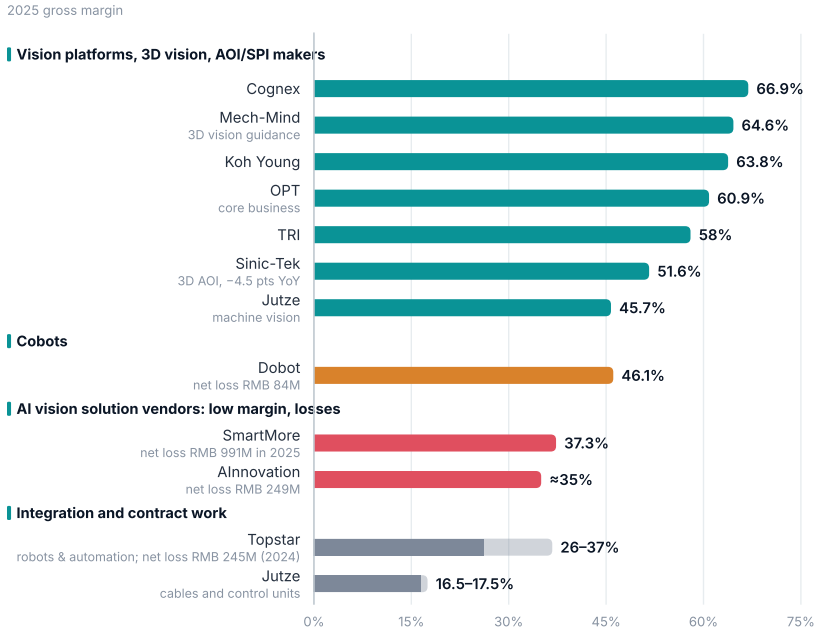
A newcomer winning 2–5% of serviceable lines in 3 years: ≈320–800 lines, ≈RMB 2–5M a year
As a subscription (RMB 5k per line a year): RMB 1.6–4M recurring, only 0.3–2.3% of IDC's China industrial AI QC forecast



Source: NVR (2020 base); author's estimates. Author's estimates; installed base and shares not independently verified.

The review wedge is encircled; the margin sits in productized hardware

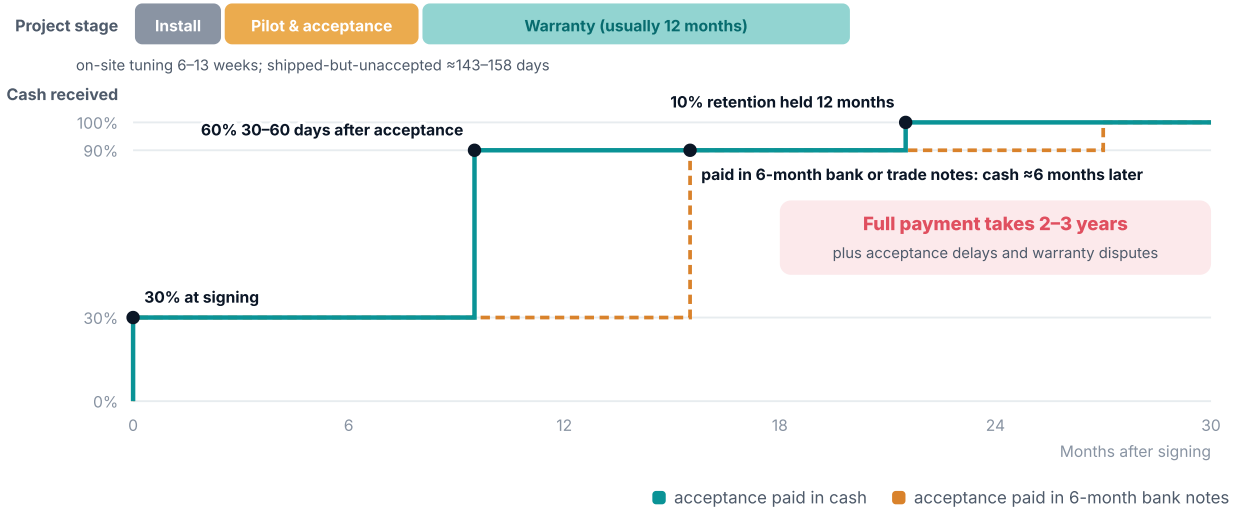


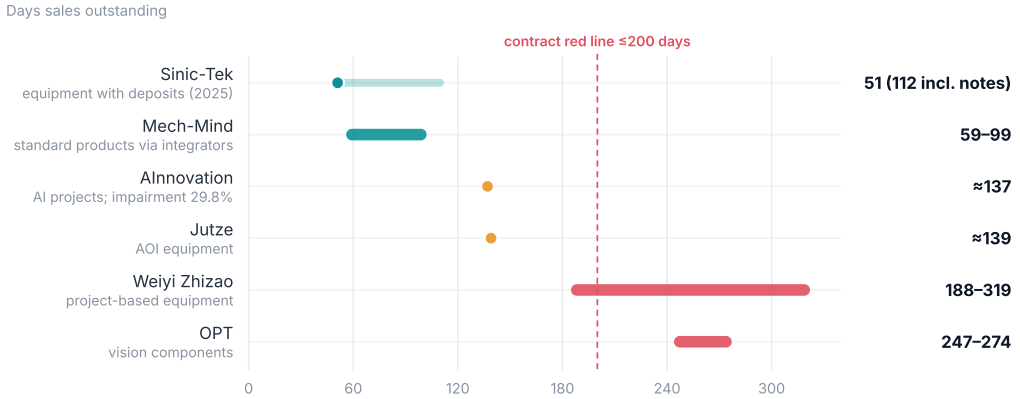


Evidence The leading AI vision vendor had 2025 revenue of RMB 1.086 billion and a net loss of RMB 991 million; the closest listed comparable's 3C revenue fell 79.9% year on year in January–May 2026.

Source: Jutze, Sinic-Tek and DBG annual reports; Hikvision and SmartMore public material; compiled by the author. 2025 annual reports or prospectuses.

Buyers pay 30/60/10; full payment takes two to three years





Key point 30% arrives with the order, 60% about four to nine months later and the last 10% about 16–21 months later; from a first pilot at a large EMS to full payment can take two to three years [estimate].

Source: Sinic-Tek inquiry reply, Mech-Mind, GKG, JT and other filings; schematic on usual terms, timing estimated. Company filings; receivables (incl. contract assets) ÷ revenue × 365.

Plan offline, act with small policies, keep motion and safety deterministic

Layer	What it does, and how far the evidence goes	Time scale · who owns it
Planning L4	**Frontier models (offline)** write or change robot programs offline, changeovers, explain faults Rewrites ABB RAPID programs at 77–99.7%, deployed after simulation Control loop only 0.2–0.4 Hz: stays out of real time	**seconds to minutes** Model giants
Skills / policy L3	**ACT, diffusion policies, small VLAs** Fixtured, structured stations 90–99.4% Edge chips: Jetson Thor, D-Robotics RDK, Ascend, Rockchip	**action chunks 100–300 ms** Chip vendors · robot makers
Motion control L2	**Robot controller** trajectories, servo, force control, collision checks deterministic code; learned policies only set goals	**milliseconds (real time needs >83 Hz)** Robot makers
Safety L1	**Safety PLC, scanners, force and speed limits** certified to ISO 10218-1/-2:2025 and local rules No ML in safety functions: EU notified-body review from Jan 20, 2027	**hard real time** Certified robot + integrator

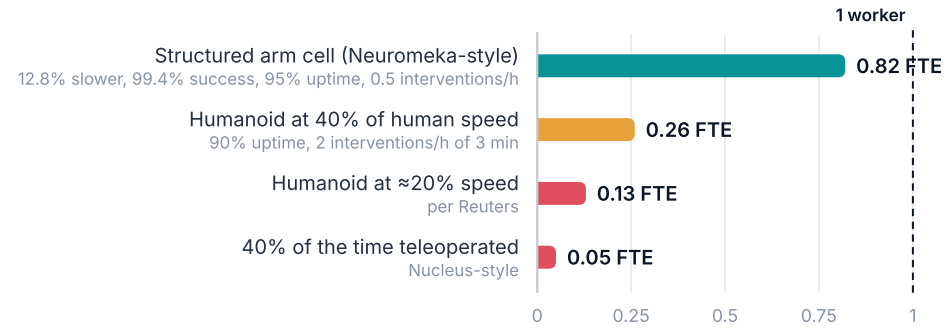
Small firms win in integration and sensing hardware: put learned perception and small policies into certified robot cells, one station at a time

	Success rate	Takt	Interventions / teleop	Uptime / faults	Volume run
Figure 02 @ BMW	target only >99%/shift	target only 84 s	target only 0/shift	n/d	1,250 h
Agility Digit @ GXO	n/d	n/d	n/d	n/d	100,000 totes
Apptronik Apollo @ Mercedes	n/d	n/d	n/d	n/d	n/d
Agile ONE @ Audi	n/d	n/d	n/d	n/d	n/d
Nucleus (unnamed plant)	n/d	n/d	≈40% teleoperated*	n/d	≈2 h of video
AgiBot G2 @ Longcheer	99.99%	18–20 s per step	n/d	one fault, 2 min	64+ h
Neuromeka hybrid cell	322/324	12.8% slower	n/d	n/d	324 cycles
π0.6 @ customer site	—	—	96.4% autonomous	n/d	1 shift

* method not disclosed

Disclosed (mostly vendor claims) · targets only · n/d

Net labor replaced per robot per shift



Source: author's synthesis (Chapters 3, 6) and estimates; ISO 10218:2025; EU Machinery Regulation; public material from BMW, Figure, Agility, Mercedes-Benz and others.

Build narrowly, prove it on the plant's own images, then scale

Build	AI review and defect classification on AOI images for mid-size EMS with older or mixed-brand AOIs; inspection the AOI covers poorly; cobot cells where every part is handled
Don't build	A general AI-plus-arm platform; an LLM as primary inspector; a pick-out arm on an SMT line; humanoids and robots-as-a-service; frontier models in control loops or safety functions

Pilot

1. **Weeks 0–2:** data agreement, defect classes, baseline
2. **Weeks 1–4:** golden set, labelled blind and adjudicated
3. **Weeks 2–4+:** shadow mode, compared per class
4. **Weeks 4–8:** assisted running on passed classes only
5. **Then:** written acceptance, scale line by line, revalidate on every change

Safety No machine learning inside safety functions: the EU requires notified-body assessment from January 20, 2027, and AI certification paths will not settle before 2027–2029.

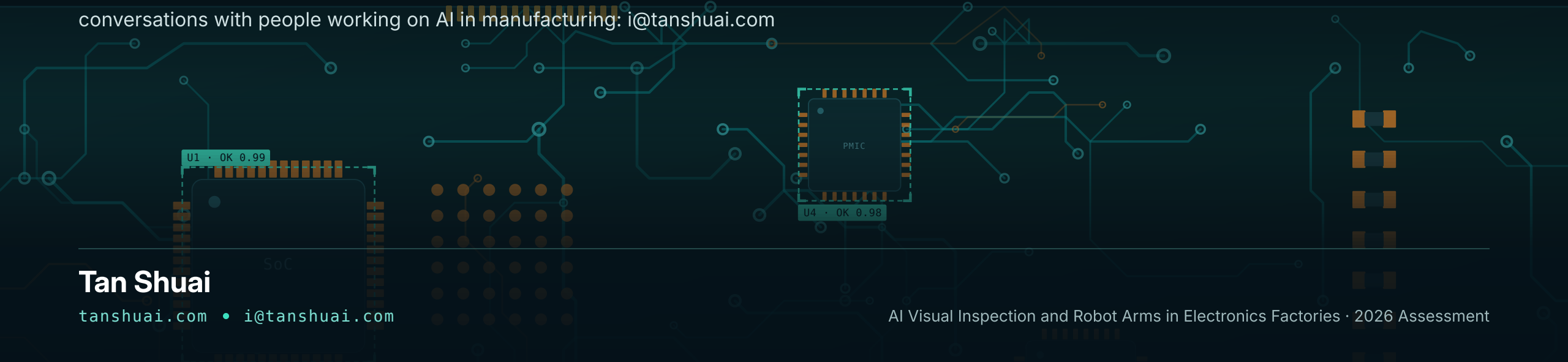
Source: report Chapters 6–7 and executive summary; ISO, IEC, UL, EUR-Lex, SAC and national regulation databases.

ABOUT THE AUTHOR

Tan Shuai

Technologist and founder, with more than twenty years across chips, hardware and system software and 20 granted patents. His work has taken him through Silicon Valley, New York, Chicago, Taiwan, the Greater Bay Area (Shenzhen and Hong Kong), Vietnam and Malaysia, with a long focus on advanced manufacturing and supply chains. Today he works as an AI researcher and developer, concentrating on how AI is used on the factory floor.

The full report (all figures, tables and sources) and the article are at tanshuai.com. He welcomes conversations with people working on AI in manufacturing: i@tanshuai.com



Tan Shuai
tanshuai.com • i@tanshuai.com