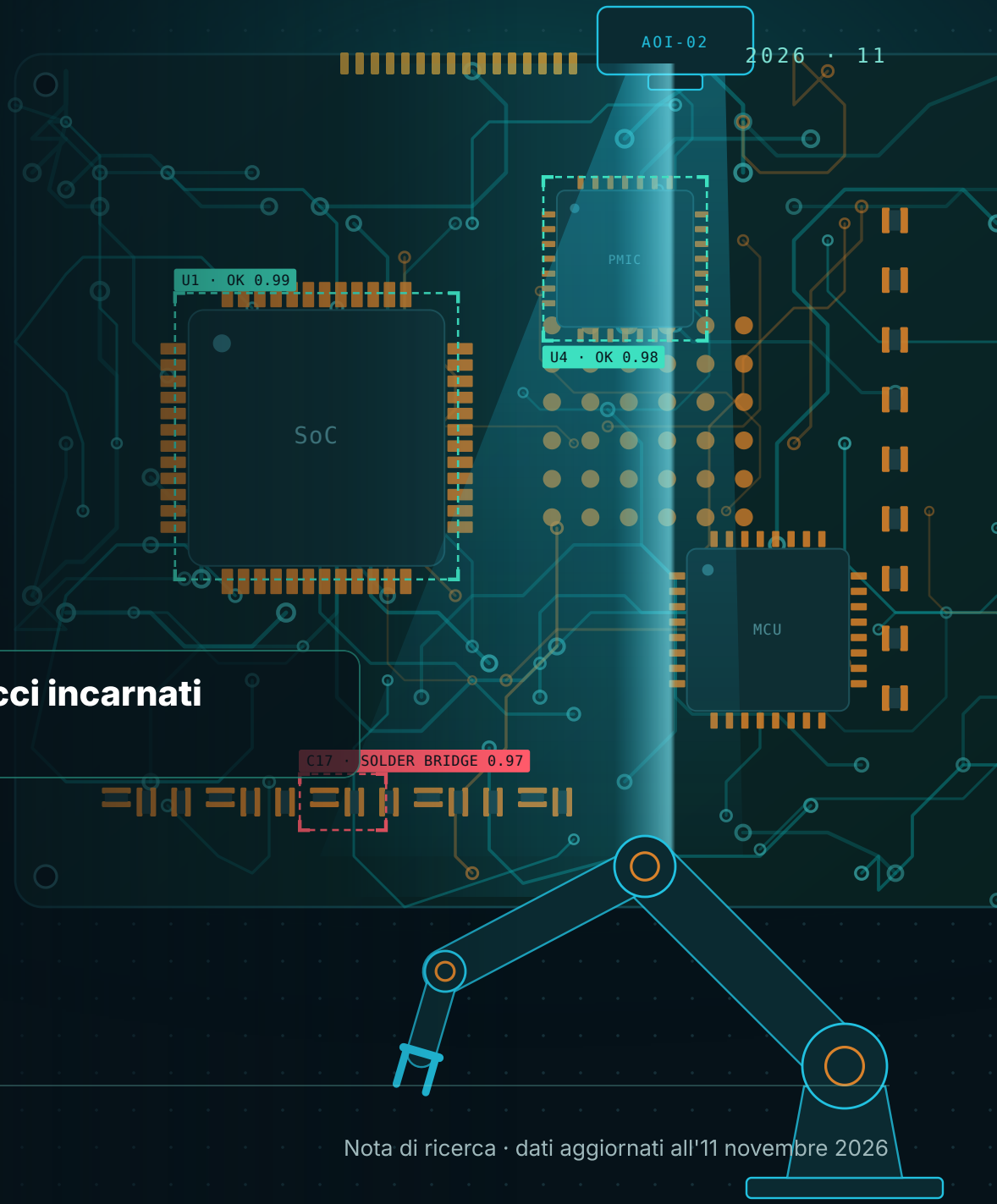


PRODUZIONE ELETTRONICA · ISPEZIONE AI × ROBOTICA

L'AI sulla linea SMT:

cosa ripaga nel 2026

Ispezione visiva AI e bracci robotici nelle fabbriche di elettronica

Cina continentale · Taiwan · Sud-est asiatico · Stati Uniti

RISULTATO CHIAVE

La revisione AI conviene, in modo mirato. I bracci incarnati possono attendere.

Tan Shuai
tanshuai.comi@tanshuai.com

Nota di ricerca · dati aggiornati all'11 novembre 2026

Conviene, ma in modo mirato, adesso e con partner

RISULTATO CHIAVE

Ciò che è maturo nel 2026, e per cui le fabbriche pagheranno, è la revisione AI e l'ispezione di nicchia su immagini AOI più l'alimentazione macchine robotica programmata, non "un grande modello che guarda le schede più robot incarnati che le prelevano".

0.46–0.65 yr

Rientro della revisione AI negli Stati Uniti, a Taiwan e a Shenzhen, a RMB 50k per linea

≥95% · <20 PPM

La revisione AI interna di un grande EMS: l'asticella per vendere agli EMS di primo livello

100–700 RMB M

Mercato di retrofit accessibile a terzi nella Cina continentale, una tantum, finestra di ≈5 anni

0 scores

Punteggi benchmark sui difetti industriali per i modelli di punta di novembre 2026

1.1–1.5 yr

Rientro di una cella cobot di alimentazione macchine, due turni

0.003–0.14 FTE

Manodopera risparmiata per turno da un braccio robotico di prelievo: non rientra mai

Fonte: sintesi del rapporto e capitoli 4–5; le cifre su rientro e mercato sono stime dell'autore.

Sei domande, sei risposte

① Cosa usano le fabbriche

SPI 3D, AOI pre- e post-reflow, raggi X campionati, ICT/FCT. Dietro ogni AOI c'è una persona che revisiona; takt ≈13–22 s per scheda.

② Quanto lontano arriva la tecnologia

L'ispezione è di livello industriale come detector + revisione AI + poche persone; non zero persone a qualsiasi prezzo. Cina continentale: modelli domestici; Taiwan: decide il cliente; Stati Uniti: dazi e certificazione.

③ On-premises e air-gapped

Pratico; i limiti sono procedurali. Un box senza ventola ≤100 W in linea, server nella sala server, 8 bit di default.

④ Insegnante e studente

Ispezione: i modelli frontier pre-etichettano, le persone arbitrano, un golden set fa da gate. Bracci: gli insegnanti sono ancora demo umane e simulazione.

⑤ Costi e rientro

Revisione AI entro un anno negli Stati Uniti, a Taiwan e a Shenzhen; alimentazione cobot 1.1–1.5 anni; Sud-est asiatico 2–6 anni; un braccio di prelievo mai.

⑥ Adesso o attendere

Domanda reale, offerta affollata; l'AI OEM lascia una finestra di retrofit di ≈5 anni. Ispezione adesso, in modo mirato; AI incarnata dopo i dati del 2027–28.

Fonte: sintesi del rapporto.

Dove l'AI si inserisce oggi in fabbrica

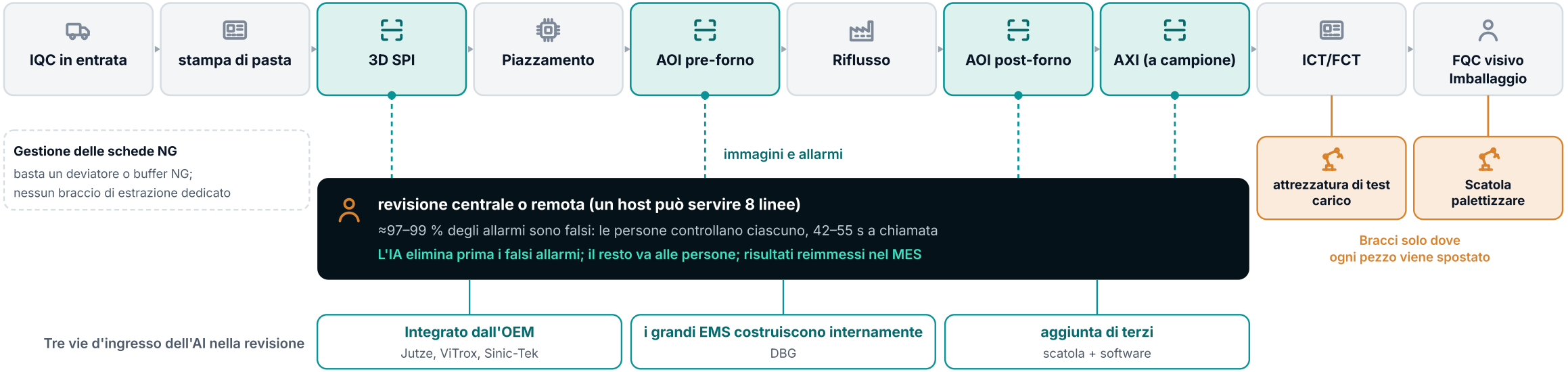


Fonte: Ritorno dal modello dell'autore (capitolo 4); l'asse di maturità è un giudizio dell'autore.

Le persone siedono dietro l'AOI. È lì che entra l'AI

Una linea SMT in volume

takt ≈13–22 s per scheda · 3–5 AOI per linea



Fonte: Documenti pubblici e materiale dei fornitori (Huaqin, DBG, Sinic-Tek e altri); compilato dall'autore.

Circa il 97–99 % degli allarmi AOI sono falsi

Per ogni 100 allarmi AOI



Oggi: le persone controllano ogni allarme

42–55 s ciascuno

≈448 falsi allarmi per turno su una linea affollata

→ 5,2–6,8 h per turno ≈ 1 revisore a tempo pieno

dotazione del caso: 2 per turno

revisione AI: eliminare i falsi allarmi, passare il resto alle persone

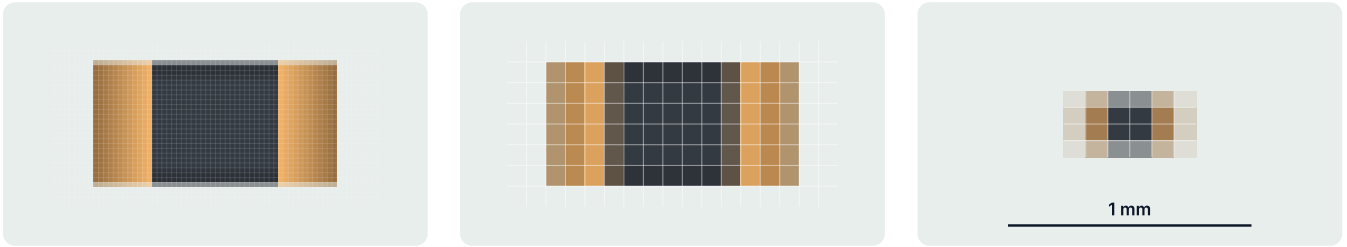
falsi allarmi o ore di revisione ridotti del 60–90 % (affermazioni del fornitore)

benchmark interno dei grandi EMS: copre ≥95 % dei componenti segnalati, fughe <20 PPM, revisione umana ridotta ≥85 %

fail-open verso le persone, la linea prosegue

Fonte: Caso di studio Siemens, iFactory (affermazione del fornitore), rapporto annuale DBG 2025; ore di revisione stimate dall'autore.

Una fotocamera generica non è un AOI



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

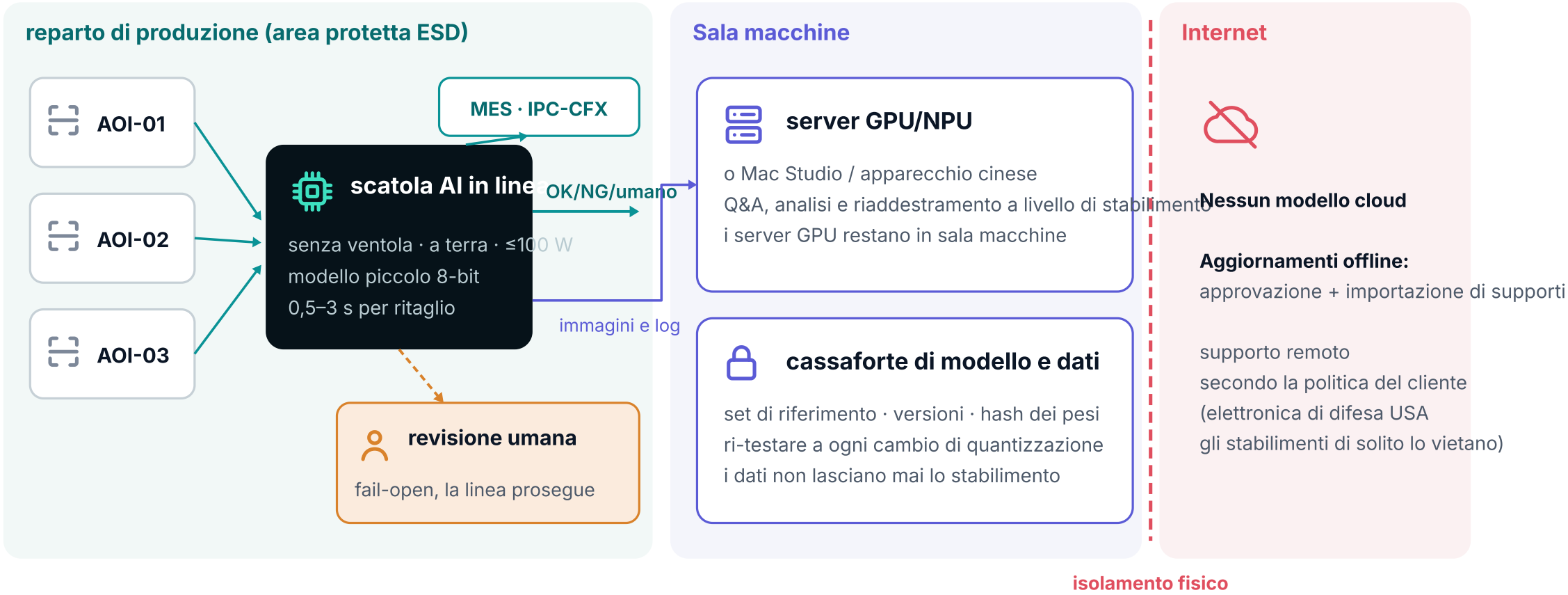
Fotocamera 4K su un pannello da 300 mm: ≈78 µm/px
Il 0402 si riduce a ≈13×6 px: appena sufficiente per presenza e scostamento grossolano

Stessa fotocamera su un 01005
≈5×3 px: di fatto invisibile

- **AOI:** 8–20 µm per pixel, illuminazione multiangolo multicolore e luce strutturata
- **4K su un pannello da 300 mm:** ≈78 µm per pixel; un 0402 è ≈13×6 pixel, sufficiente per la presenza, non per i giunti
- **01005:** ≈5×3 pixel, di fatto invisibile
- **Quindi:** l'AI lavora sulle immagini stesse dell'AOI; non sostituisce l'ottica

Fonte: Circuits Assembly; calcolo dell'autore per un pannello da 300 mm e un sensore 4K. Le griglie di pixel sono illustrative.

Una distribuzione air-gapped: il box decide, la sala server risponde



Fonte: Sintesi dell'autore (capitolo 2).

Le zone pulite ed ESD decidono dove risiede il calcolo; i limiti di velocità dipendono dall'uso

Pulizia →

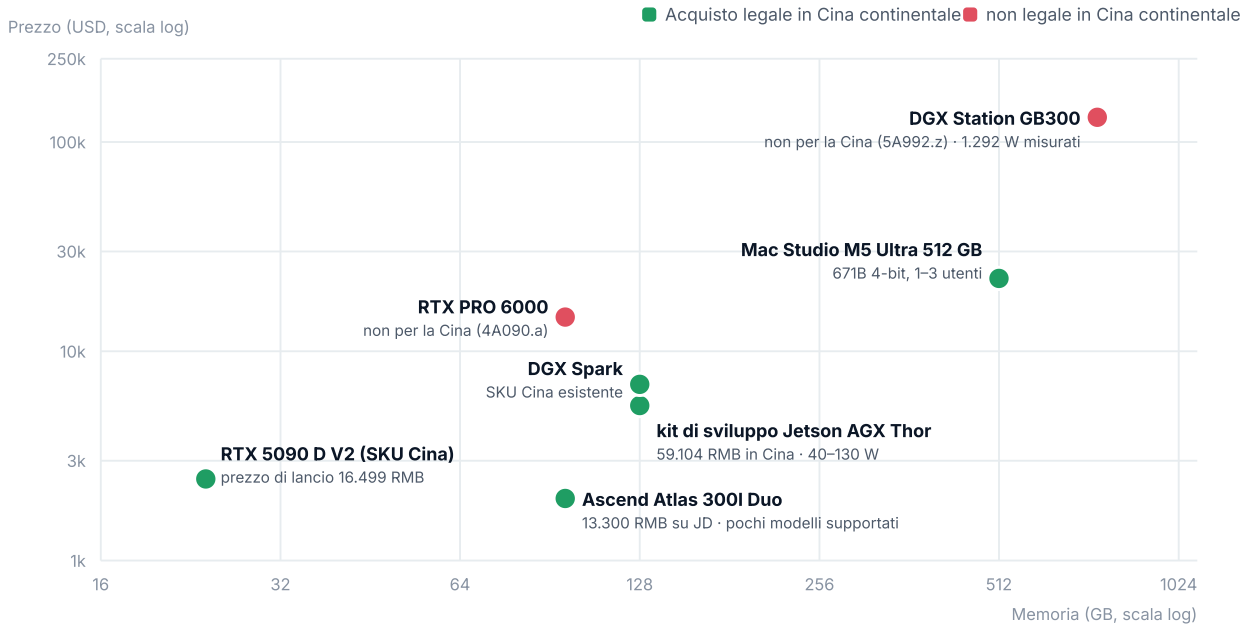
reparto SMT/PCBA ordinario	camera pulita ISO 7–8	mini-ambiente ISO 5–6	fab, ISO 5 o più pulito
area protetta ESD (≈ISO 8) UR 40–60 %	medicale, moduli fotocamera, VE camera pulita + ESD	die nudo, wire bond, ottica mini-ambiente	SEMI E78 ESD · E187/E188 cyber piani specifici per fab
calcolo in linea scatola senza ventola con involucro metallico a terra, vicino all'AOI o alla postazione di revisione	**calcolo in linea** Senza ventola, sigillato IP65/66, facile da pulire; richiedere i dati ISO 14644-14	**calcolo in linea** Solo fotocamere e luci; la scatola di calcolo fuori	**calcolo in linea** solo controller integrati nello strumento e qualificati per fab
server GPU/NPU sala macchine o rack IT climatizzato	**server GPU/NPU** fuori dalla camera pulita, rete verso l'interno	**server GPU/NPU** Sala macchine	**server GPU/NPU** rack raffreddati a liquido in sottofabbrica o data center



Punto chiave Una normale sala SMT è un'area protetta ESD, non una camera bianca classificata: un box senza ventola in un involucro metallico collegato a terra sta accanto all'AOI o alla postazione di revisione, e i server GPU non vanno mai in linea.

Fonte: ANSI/ESD S20.20, IEC 61340-5-1, ISO 14644, norme SEMI e materiale dei fornitori; compilato dall'autore. Sintesi dell'autore (capitoli 1–2).

Le regole sull'export selezionano la short list hardware prima dei benchmark



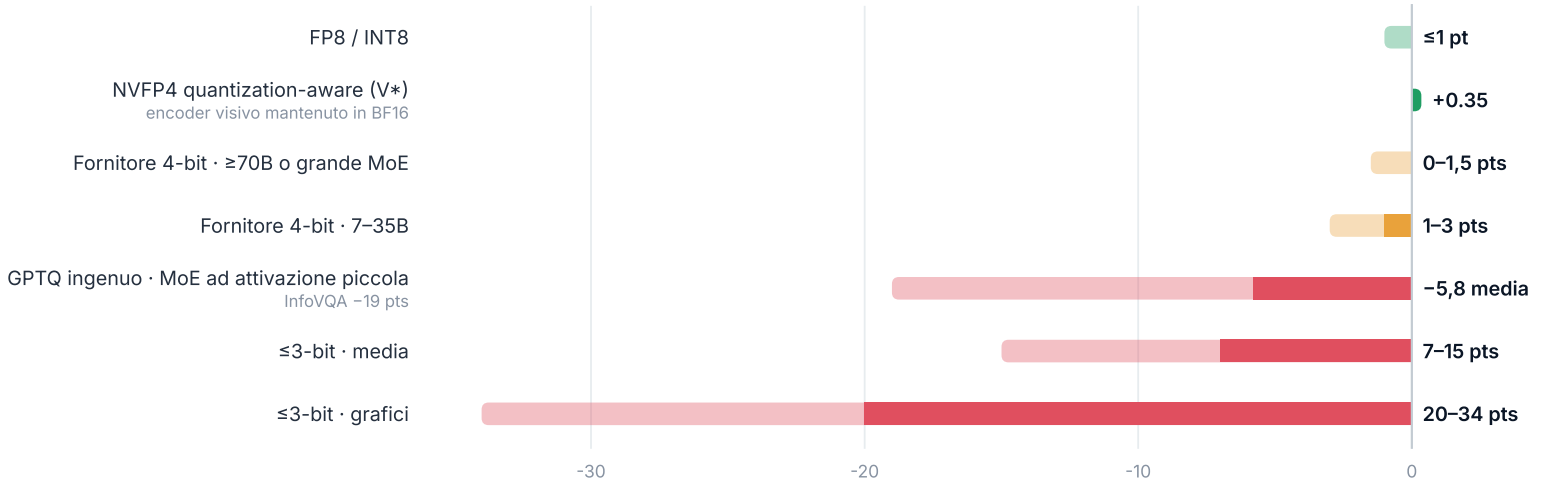
- **Jetson Thor:** 128 GB, 40–130 W; modelli di visione 2–8B su ritagli AOI, VLA di classe 3B in tempo reale; RMB 59,104 in Cina
- **Mac Studio 512 GB:** DeepSeek 671B a 4 bit per 1–3 utenti
- **DGX Station:** 5A992.z, non per la Cina continentale; un modello 743B dà a 32 utenti ≈4 token/s ciascuno
- **Acceleratori domestici:** due server Ascend a otto schede per Q&A multiutente a grandezza piena

Fonte: Classificazioni all'esportazione NVIDIA, Apple, RS China, JD.com e prezzi di recensione (novembre 2026). StorageReview (ago. e ott. 2026); ServeTheHome. Test indipendenti da singola fonte.

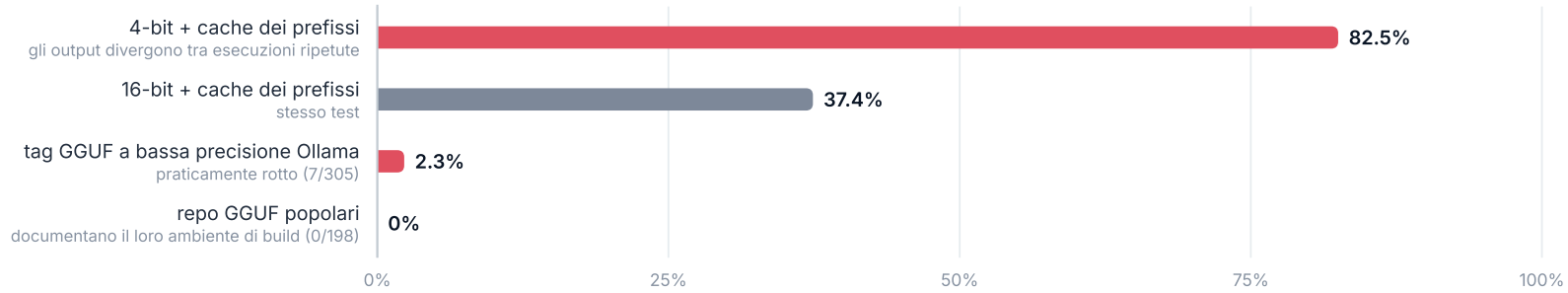
Quantizzazione: 8-bit è sicuro, 4-bit dipende dalla ricetta, 3-bit e meno è escluso

Precisione: 8-bit sicuro, 4-bit secondo la ricetta, ≤3-bit escluso

variazione del punteggio vs piena precisione (punti)

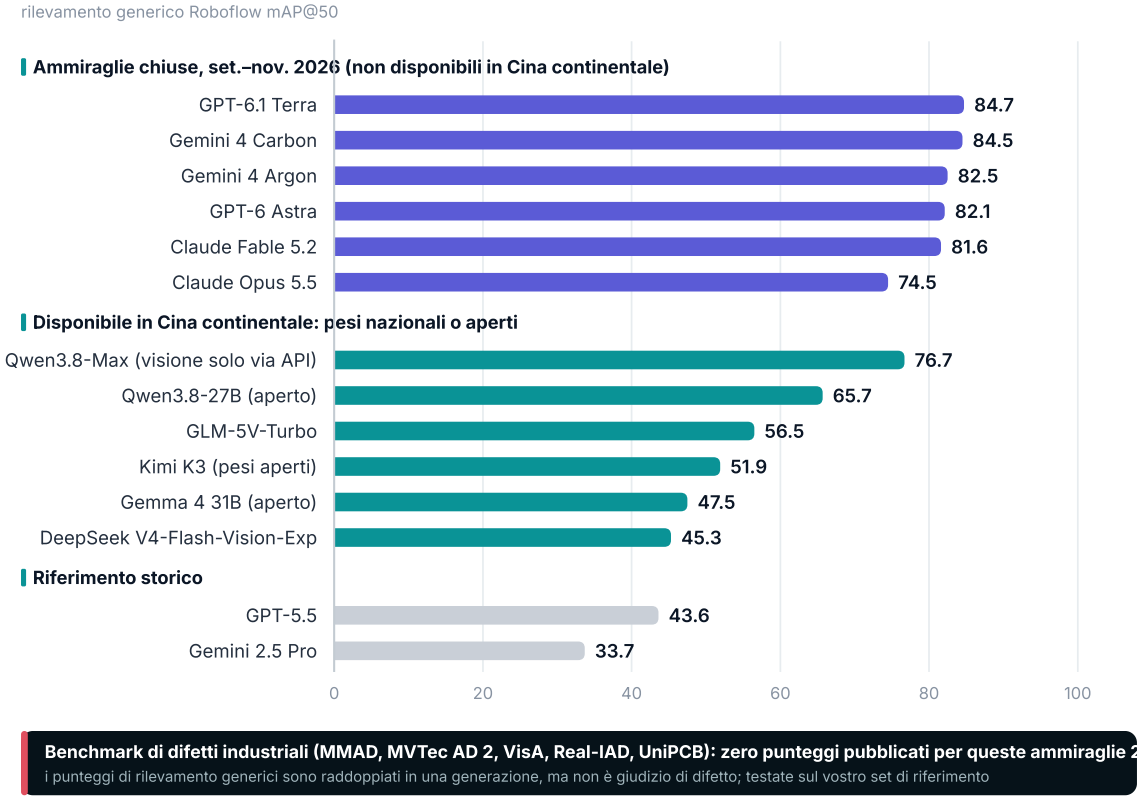


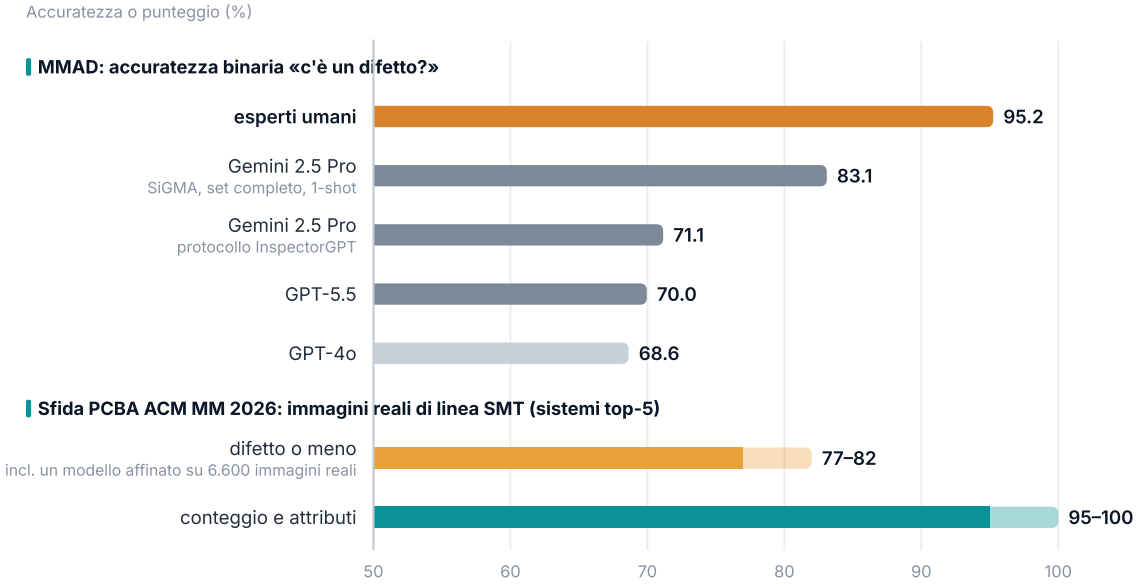
i guasti ingegneristici pesano più dell'errore di quantizzazione



Fonte: Scheda modello NVIDIA Nemotron, VLMQ, VEQ, arXiv 2609.05881, 2610.19498, 2609.04748 e altri.

La visione generica è raddoppiata in una generazione; mancano i punteggi sui difetti industriali





Benchmark di difetti industriali (MMAD, MVTec AD 2, VisA, Real-IAD, UniPCB): zero punteggi pubblicati per queste ammiraglie 2

i punteggi di rilevamento generici sono raddoppiati in una generazione, ma non è giudizio di difetto; testate sul vostro set di riferimento

Fonte: Roboflow Vision Evals; note di rilascio dei modelli; pagine di MMAD, MVTec AD 2 e altri benchmark (all'11 nov. 2026). MMAD; SiGMA; InspectorGPT; sfida PCBA ACM MM 2026.

I modelli frontier pre-etichettano, le persone arbitrano, un golden set fa da gate

Ispezione: il modello di frontiera pre-etichetta, le persone arbitrano, un set di riferimento valida

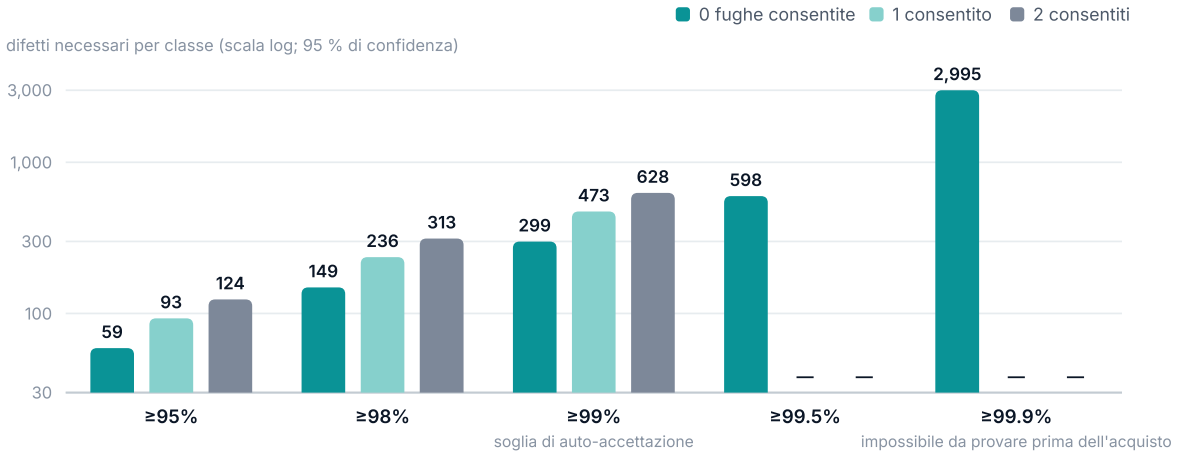


soglia per accettare automaticamente le etichette del maestro

Per classe: ≥299 difetti, zero fughe (recall ≥99 % al 95 % di confidenza), accordo non peggiore di umano–umano; altrimenti solo pre-etichettatura

Bracci: i maestri restano le demo umane, il RL su robot reale e la simulazione modelli di frontiera aiutano solo offline: programmazione e cambi serie



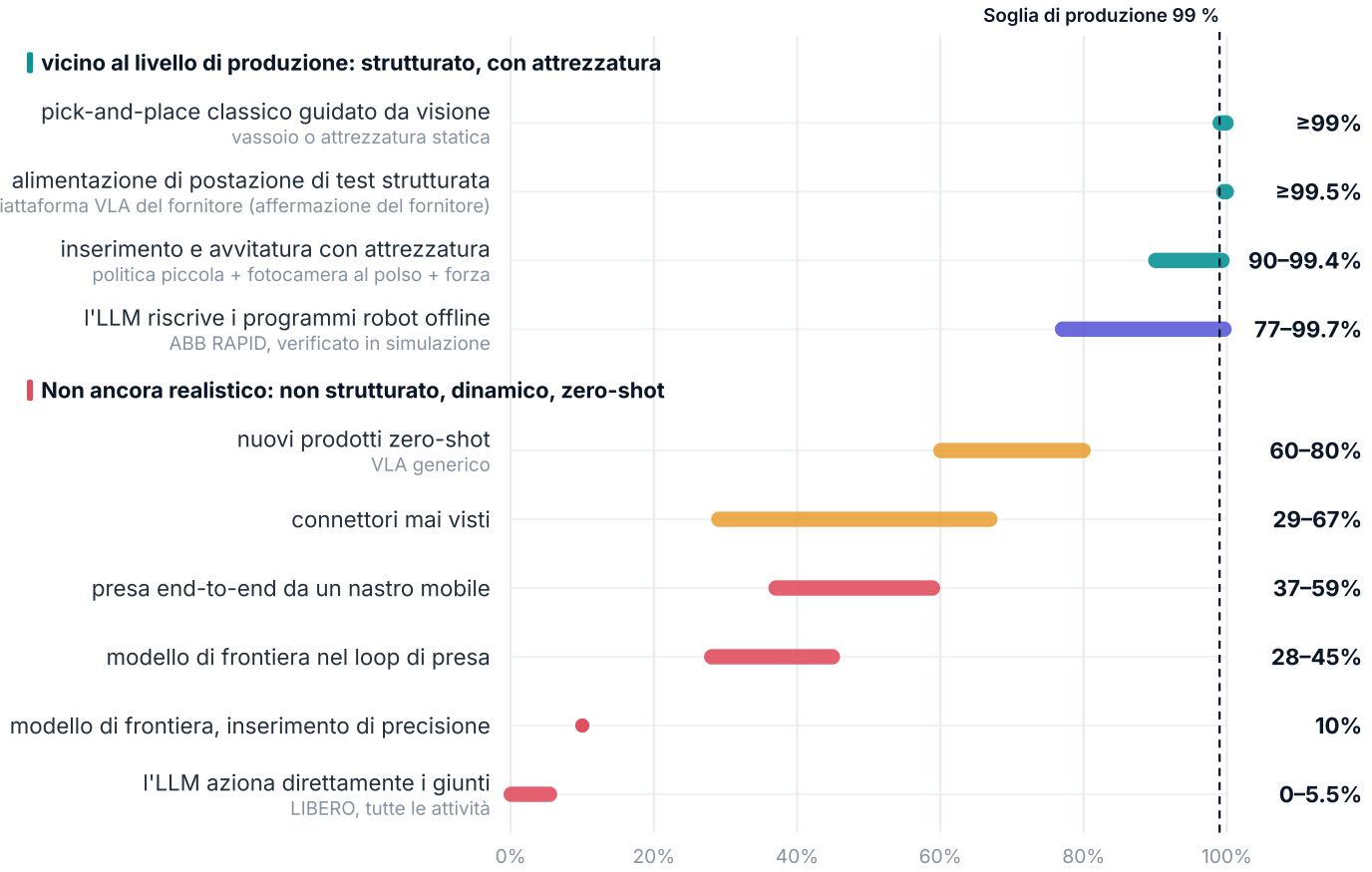


Punto chiave Accetta automaticamente le etichette dell'insegnante solo per una classe con zero errori su ≥299 difetti golden (99% di recall al 95% di confidenza); le classi critiche come cortocircuiti, parti mancanti e polarità richiedono di più.

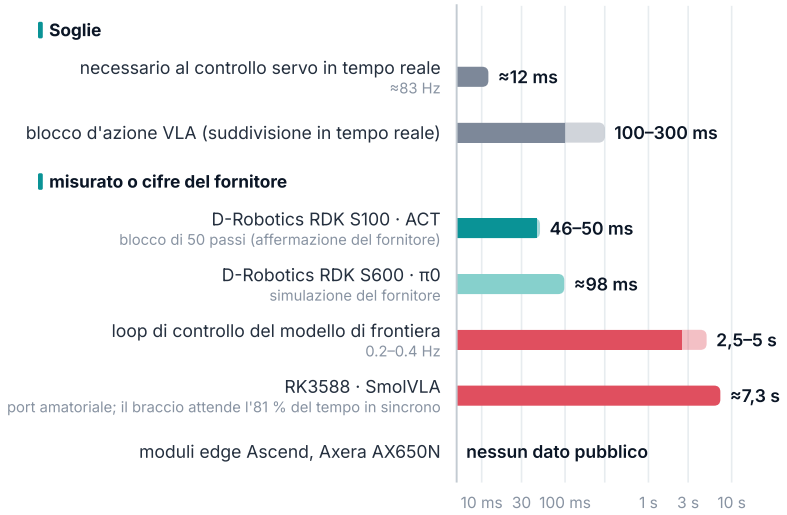
Fonte: Sintesi dell'autore (capitolo 3); costi e dimensioni del campione sono stime. Intervalli di confidenza binomiali esatti (95 %).

Studenti robot: vicini al livello di produzione solo se la postazione è strutturata

tasso di successo dell'attività



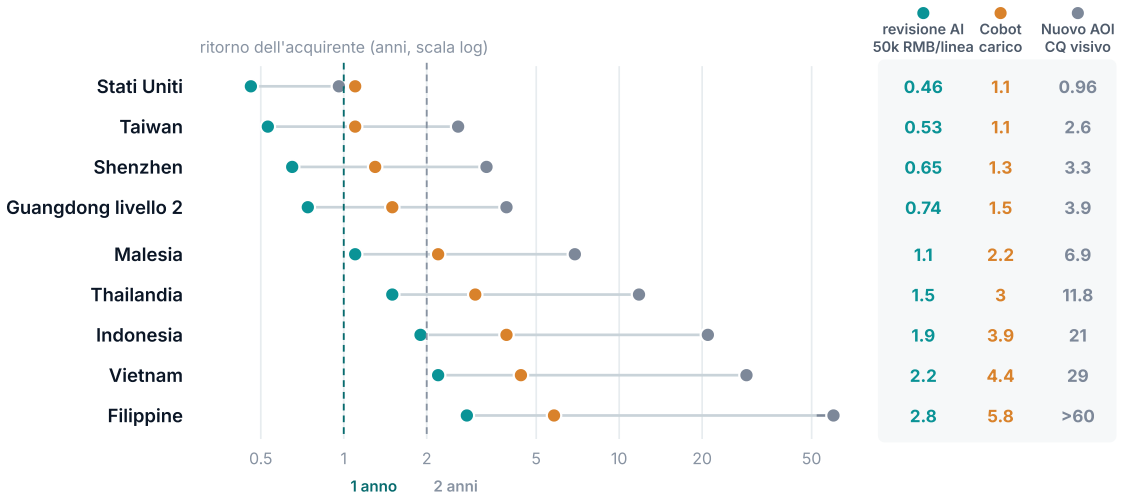
latenza di inferenza per blocco d'azione o ciclo di controllo (scala log)



Fonte: Articoli e materiale dei fornitori citati nel capitolo 3; alcune cifre sono affermazioni del fornitore. Documentazione e forum D-Robotics (cifre del fornitore); port comunitari GitHub; ricerca Anthropic.

La revisione AI rientra in mesi; l'alimentazione macchine in circa un anno



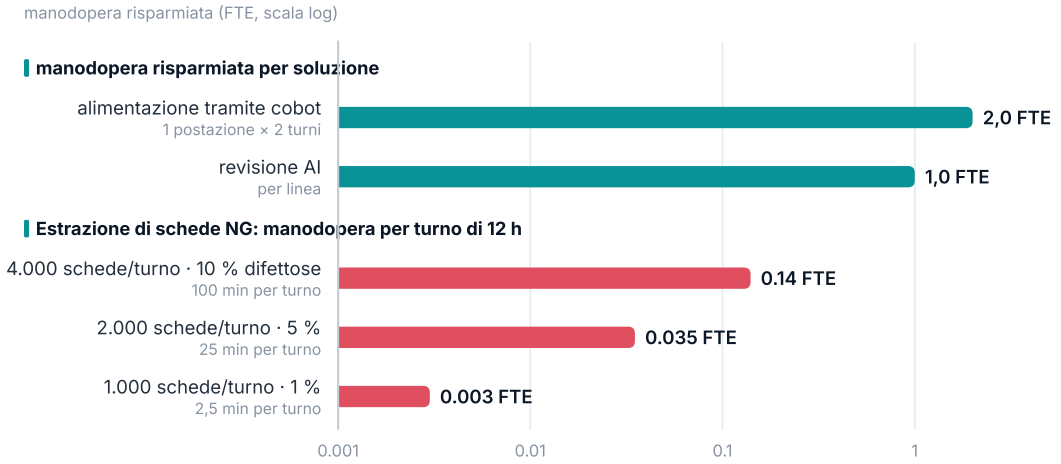


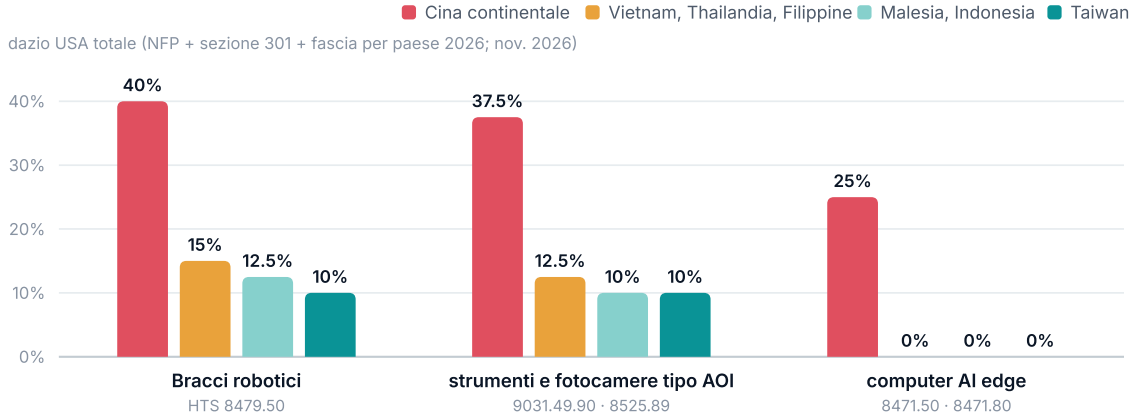
Un braccio robotico solo per estrarre le schede NG non ripaga mai, da nessuna parte
Con resa di primo passaggio oltre il 90 %, l'estrazione occupa solo 0,003–0,14 FTE per turno

Punto chiave A RMB 50,000 per linea, la revisione AI rientra in 0.46 anni negli Stati Uniti, 0.53 a Taiwan, 0.65 a Shenzhen e 2.2 in Vietnam; una cella cobot di alimentazione su due turni in 1.1–1.5 anni [estimates].

Fonte: Guida salariale ufficiale di Shenzhen, Ministero del Lavoro di Taiwan, JETRO, BLS; stime dell'autore. Modello dell'autore: scatola di revisione AI da 50k RMB per linea; cella di alimentazione cobot a 200k RMB o 140k $, due turni.

Un braccio di prelievo non rientra mai; dove un braccio viene prodotto cambia il suo dazio statunitense





Punto chiave Oltre il 90% di resa al primo passaggio, estrarre le schede NG richiede 0.003–0.14 di una persona per turno, quindi un deviatore o un buffer NG fa il lavoro; un braccio prodotto in Cina continentale paga inoltre circa il 40% di dazio statunitense combinato.

Fonte: Stime dell'autore. Capitolo 99 della Tariffa Armonizzata USITC; Federal Register; USTR (novembre 2026).

Reale ma piccolo: 16,000 linee servibili, ≈RMB 320M una tantum, una finestra di cinque anni

linee SMT in Cina (2026)

40.000 righe

35k–46k

× 80% × 50%

ancora senza AI: 75–85 %
raggiungibili: 40–60 %

Linee servibili

16.000 righe

10,5k–23,5k

× 20k RMB

10–30k RMB una tantum

Mercato una tantum

320 M RMB

105–705 M RMB

÷ ≈5 anni

si riduce dell'8–12 %/anno

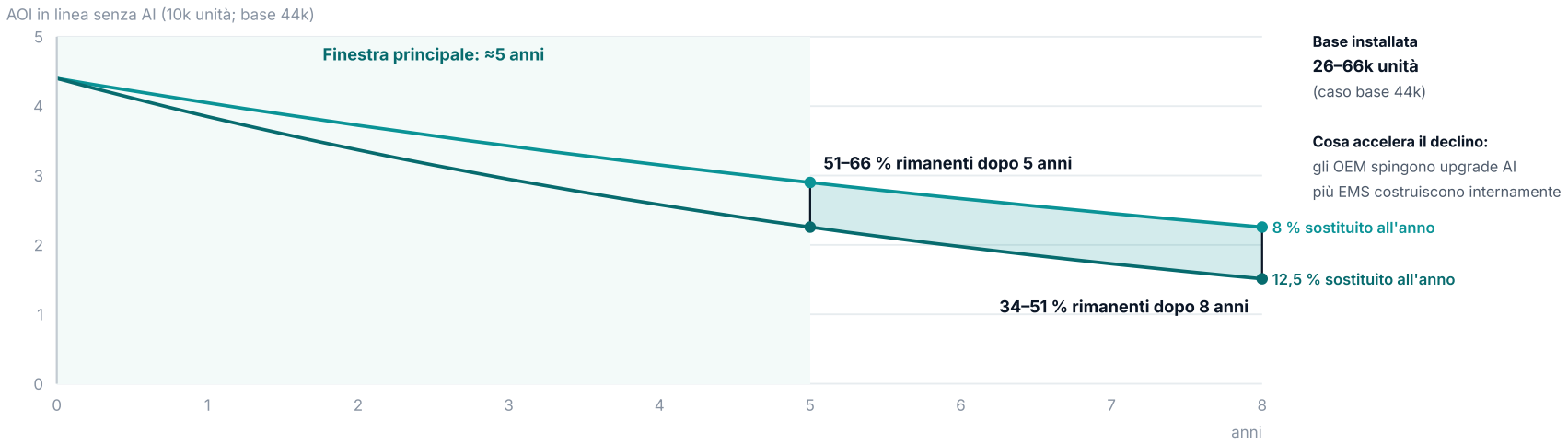
finestra di 5 anni

RMB 64M/yr

21–141 M RMB

Un nuovo entrante che conquista il 2–5 % delle linee servibili in 3 anni: ≈320–800 linee, ≈2–5 M RMB all'anno

Come abbonamento (5k RMB per linea all'anno): 1,6–4 M RMB ricorrenti, solo lo 0,3–2,3 % della previsione IDC del CQ industriale AI in Cina



Fonte: NVR (base 2020); stime dell'autore. Stime dell'autore; base installata e quote non verificate in modo indipendente.

Il cuneo della revisione è accerchiato; il margine sta nell'hardware prodotto in serie

i grandi EMS costruiscono internamente

DBG (2025): copre ≥95 % dei componenti segnalati
fughe <20 PPM, revisione umana −85 %
questa è la soglia per vendere agli EMS di prima fascia

AI integrata nell'AOI dagli OEM

Jutze AI 3D AOI in produzione, «elimina la revisione umana all'ispezione»
ViTrox V510Ai; Sinic-Tek co-sviluppa con Tencent
le macchine nuove includono l'AI; quelle esistenti richiedono upgrade

revisione AI di terzi

scatola + software agganciato all'AOI esistente

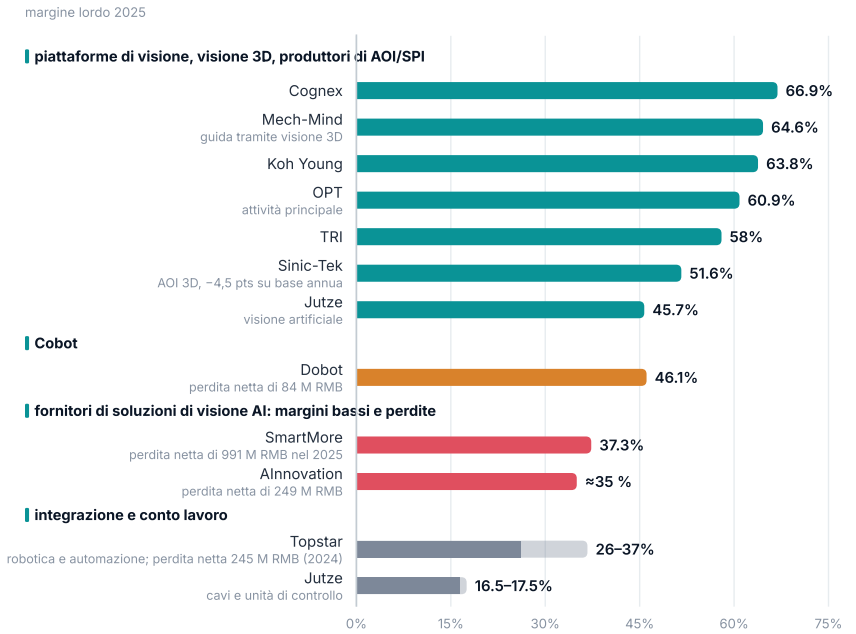
fornitori di piattaforma

Hikvision Guanlan e SmartMore IndustryGPT stanno vendendo
fornitore AI leader: ricavi 2025 di 1,086 mld RMB
perdita netta di 991 M RMB: guerra dei prezzi

Ciò che resta ai terzi

EMS medi (≈10–40 linee) con AOI vecchio o multimarca senza upgrade AI dell'OEM

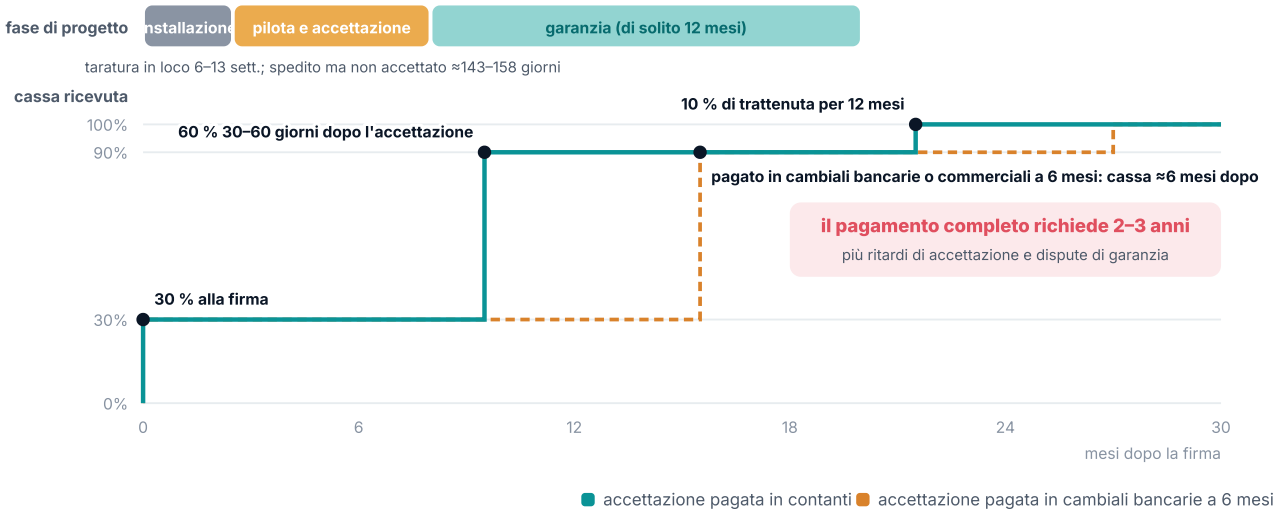
Nicchie mal coperte dall'AOI: THT e saldatura manuale, vernice protettiva, cavi/connettori/FPC, aspetto finale e viti

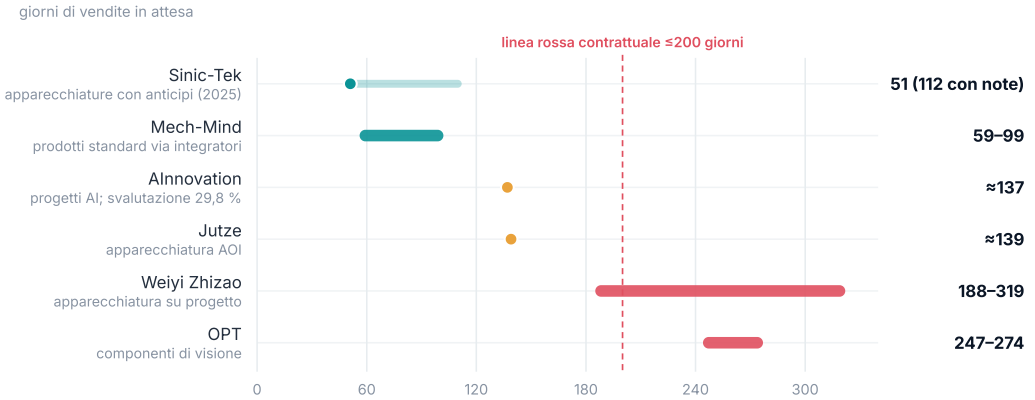


Evidenze Il principale fornitore di visione AI ha registrato nel 2025 ricavi per RMB 1.086 miliardi e una perdita netta di RMB 991 milioni; il ricavo 3C del comparabile quotato più vicino è calato del 79.9% anno su anno in gennaio–maggio 2026.

Fonte: Rapporti annuali di Jutze, Sinic-Tek e DBG; materiale pubblico di Hikvision e SmartMore; compilato dall'autore. Rapporti annuali o prospetti 2025.

Gli acquirenti pagano 30/60/10; il pagamento completo richiede due-tre anni





Punto chiave Il 30% arriva con l'ordine, il 60% circa quattro-nove mesi dopo e l'ultimo 10% circa 16–21 mesi dopo; dal primo progetto pilota presso un grande EMS al pagamento completo possono passare due-tre anni [estimate].

Fonte: Risposta a richiesta Sinic-Tek, Mech-Mind, GKG, JT e altri documenti; schema su condizioni usuali, tempistica stimata. Documenti aziendali; crediti (incl. attività contrattuali) ÷ ricavi × 365.

Pianifica offline, agisci con policy piccole, mantieni movimento e sicurezza deterministici

Livello	Cosa fa e fino a dove arriva l'evidenza	Scala temporale · chi la controlla
Pianificazione L4	**modelli di frontiera (offline)** scrivere o modificare programmi robot offline, cambi serie, spiegare i guasti riscrive i programmi ABB RAPID al 77–99,7 %, distribuito dopo simulazione loop di controllo a soli 0,2–0,4 Hz: resta fuori dal tempo reale	da secondi a minuti giganti dei modelli
Competenze / politi L3	**ACT, politiche di diffusione, piccoli VLA** postazioni strutturate con attrezzatura 90–99,4 % chip edge: Jetson Thor, D-Robotics RDK, Ascend, Rockchip	blocchi d'azione 100–300 ms produttori di chip · produttori di robot
controllo del movim L2	**Controller robot** traiettorie, servo, controllo di forza, rilevamento collisioni codice deterministico; le politiche apprese fissano solo gli obiettivi	millisecondi (il tempo reale richiede >83 Hz) produttori di robot
Sicurezza L1	**PLC di sicurezza, scanner, limiti di forza e velocità** certificato secondo ISO 10218-1/-2:2025 e regole locali Niente ML nelle funzioni di sicurezza: valutazione da organismo notificato UE dal 20 gen. 2027	tempo reale rigido robot certificato + integratore

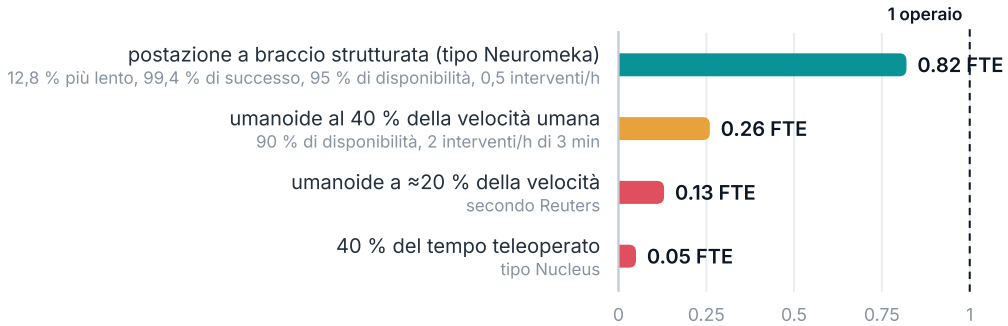
Le piccole aziende vincono nell'integrazione e nell'hardware di rilevamento: inserire percezione appresa e piccole politiche in celle robotiche certificate, una postazione alla volta

	tasso di successo	Takt	Interventi / teleoperazione	Disponibilità / guasti	Volume
Figure 02 @ BMW	solo obiettivo >99 %/turn	solo obiettivo 84 s	solo obiettivo 0/turno	n/d	1.250 h
Agility Digit @ GXO	n/d	n/d	n/d	n/d	100.000 contenitori
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 (stabilimento non nominato)	n/d	n/d	≈40 % teleoperato*	n/d	≈2 h di video
AgiBot G2 @ Longcheer	99.99%	18–20 s per fase	n/d	un guasto, 2 min	64+ h
cella ibrida Neuromeka	322/324	12,8 % più lento	n/d	n/d	324 cicli
π0.6 @ sede cliente	—	—	96,4 % autonomo	n/d	1 turno

* metodo non divulgato

Divulgato (soprattutto affermazioni del fornitore) · solo obiettivi · n/d

manodopera netta sostituita per robot e turno



Fonte: sintesi dell'autore (capitoli 3, 6) e stime; ISO 10218:2025; Regolamento macchine UE; materiale pubblico di BMW, Figure, Agility, Mercedes-Benz e altri.

Costruisci in modo mirato, dimostralo sulle immagini dello stabilimento, poi scala

Costruire	Revisione AI e classificazione dei difetti su immagini AOI per EMS di medie dimensioni con AOI più vecchi o di marche miste; ispezione che l'AOI copre male; celle cobot dove ogni parte viene manipolata
Non costruire	Una piattaforma generale AI-più-braccio; un LLM come ispettore primario; un braccio di prelievo su una linea SMT; umanoidi e robot-as-a-service; modelli frontier in loop di controllo o funzioni di sicurezza

Pilota

1. **Settimane 0–2:** accordo sui dati, classi di difetto, baseline
2. **Settimane 1–4:** golden set, etichettato in cieco e arbitrato
3. **Settimane 2–4+:** modalità ombra, confronto per classe
4. **Settimane 4–8:** esecuzione assistita solo sulle classi superate
5. **Poi:** accettazione scritta, scala linea per linea, rivalida a ogni modifica

Sicurezza Nessun machine learning dentro le funzioni di sicurezza: l'UE richiede la valutazione di un organismo notificato dal 20 gennaio 2027, e i percorsi di certificazione AI non si stabilizzeranno prima del 2027–2029.

Fonte: capitoli 6–7 del rapporto e sintesi; ISO, IEC, UL, EUR-Lex, SAC e banche dati regolatorie nazionali.

L'AUTORE

Tan Shuai

Tecnologo e fondatore, con oltre vent'anni di esperienza tra chip, hardware e software di sistema e 20 brevetti concessi. Il suo lavoro lo ha portato attraverso la Silicon Valley, New York, Chicago, Taiwan, la Greater Bay Area (Shenzhen e Hong Kong), Vietnam e Malaysia, con una lunga attenzione alla produzione avanzata e alle catene di fornitura. Oggi lavora come ricercatore e sviluppatore AI, concentrandosi su come l'AI viene utilizzata in fabbrica.

Il rapporto completo (tutte le figure, le tabelle e le fonti) e l'articolo sono su tanshuai.com. Accoglie volentieri conversazioni con chi lavora sull'AI nella produzione: i@tanshuai.com

