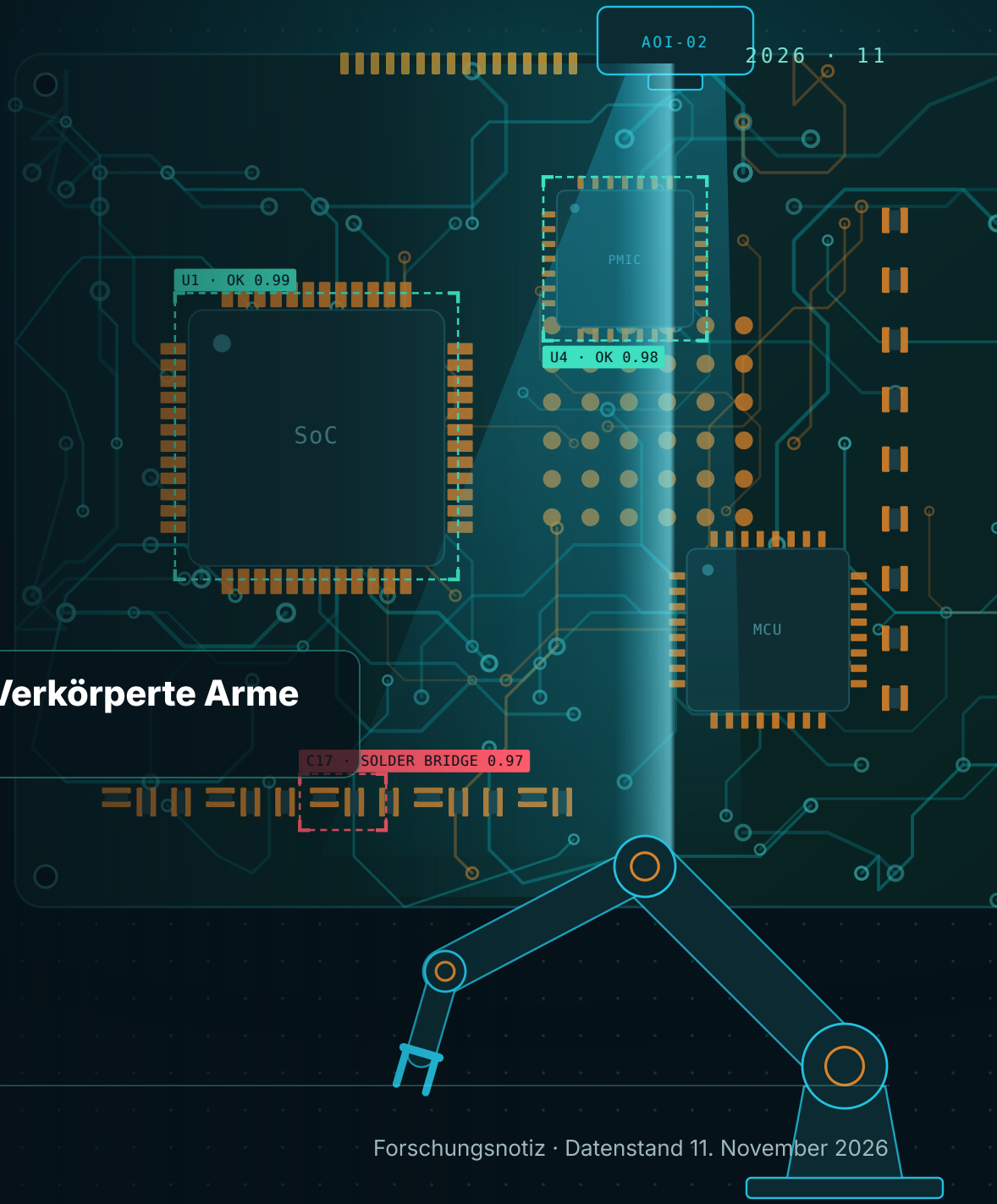
ELEKTRONIKFERTIGUNG · AI-INSPEKTION × ROBOTIK

AI an der SMT-Linie:

was sich 2026 amortisiert

AI-Sichtprüfung und Roboterarme in Elektronikfabriken

Festlandchina · Taiwan · Südostasien · Vereinigte Staaten



KERNERGEBNIS

AI-Nachprüfung lohnt sich, in engem Rahmen. Verkörperte Arme können warten.

Tan Shuai
tanshuai.comi@tanshuai.com

Forschungsnotiz · Datenstand 11. November 2026

Lohnend, aber eng gefasst, jetzt und mit Partnern

KERNERGEBNIS

Was 2026 ausgereift ist und wofür Fabriken zahlen, sind AI-Nachprüfung und Nischeninspektion auf AOI-Bildern plus programmierte Roboter-Maschinenbeschickung, nicht »ein großes Modell, das Platinen betrachtet, plus verkörperte Roboter, die sie greifen«.

0.46–0.65 yr

Amortisation der AI-Nachprüfung in den USA, Taiwan und Shenzhen, bei 50.000 RMB pro Linie

≥95% · <20 PPM

Die interne AI-Nachprüfung eines führenden EMS: die Messlatte für den Verkauf an Tier-1-EMS

100–700 RMB M

Adressierbarer Drittparteien-Nachrüstmarkt in Festlandchina, einmalig, ≈5-Jahres-Fenster

0 scores

Benchmark-Werte für Industriedefekte der Flaggschiffe vom November 2026

1.1–1.5 yr

Amortisation einer Cobot-Maschinenbeschickungszelle, zwei Schichten

0.003–0.14 FTE

Arbeit, die ein Ausschleus-Roboterarm pro Schicht einspart: amortisiert sich nie

Quelle: Bericht, Zusammenfassung und Kapitel 4–5; Amortisations- und Marktzahlen sind Schätzungen des Autors.

Sechs Fragen, sechs Antworten

① Was Fabriken betreiben

3D SPI, AOIs vor und nach dem Reflow, stichprobenartige Röntgenprüfung, ICT/FCT. Hinter jedem AOI prüft ein Mensch nach; Takt ≈13–22 s pro Platine.

② Wie weit die Technik reicht

Inspektion ist industrietauglich als Detektor + AI-Nachprüfung + einige Menschen; keine null Menschen zu jedem Preis. Festland: inländische Modelle; Taiwan: der Kunde entscheidet; USA: Zölle und Zertifizierung.

③ Vor Ort und luftspaltgetrennt

Praktikabel; die Grenzen sind prozessual. Eine lüfterlose Box ≤100 W an der Linie, Server im Serverraum, standardmäßig 8-Bit.

④ Lehrer und Schüler

Inspektion: Frontier-Modelle beschriften vor, Menschen entscheiden, ein Golden Set entscheidet über die Freigabe. Arme: Die Lehrer sind weiterhin menschliche Demos und Simulation.

⑤ Kosten und Amortisation

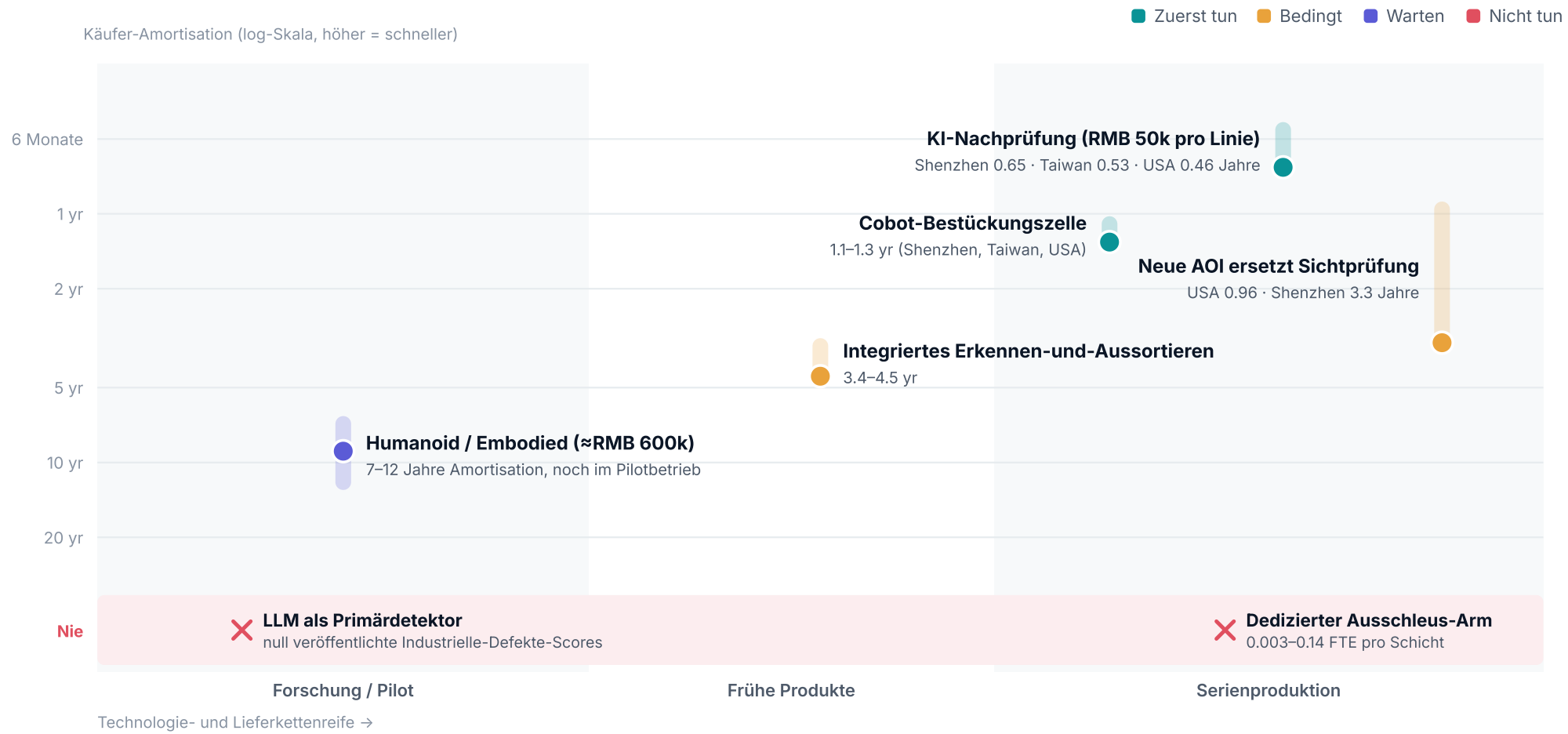
AI-Nachprüfung innerhalb eines Jahres in den USA, Taiwan und Shenzhen; Cobot-Beschickung 1.1–1.5 Jahre; Südostasien 2–6 Jahre; ein Ausschleusarm nie.

⑥ Jetzt oder warten

Reale Nachfrage, überfülltes Angebot; OEM-AI lässt ein ≈5-Jahres-Nachrüstfenster. Inspektion jetzt, eng gefasst; verkörperte AI nach Daten von 2027–28.

Quelle: Zusammenfassung des Berichts.

Wo KI heute in der Fabrik passt



Quelle: Amortisation aus dem Modell des Autors (Kapitel 4); die Reifeachse ist eine Einschätzung des Autors.

Die Menschen sitzen hinter dem AOI. Genau dort tritt AI ein

Eine SMT-Serienlinie

Takt ≈13–22 s pro Platine · 3–5 AOIs pro Linie



Quelle: Öffentliche Berichte und Herstellerunterlagen (Huaqin, DBG, Sinic-Tek u. a.); vom Autor zusammengestellt.

Etwa 97–99% der AOI-Alarme sind Fehlalarme

Pro 100 AOI-Alarme



Heute: Menschen prüfen jeden Alarm

42–55 s je

≈448 Fehlalarme pro Schicht auf einer stark ausgelasteten Linie

→ 5.2–6.8 h pro Schicht ≈ 1 Vollzeit-Prüfer

Fallbesetzung: 2 pro Schicht

KI-Nachprüfung: Fehlalarme abräumen, Rest an Menschen

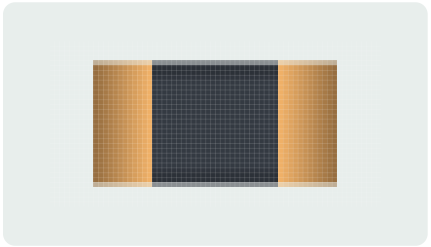
Fehlalarme oder Prüfstunden um 60–90% reduziert (Herstellerangaben)

Top-EMS-Eigenbau-Benchmark: deckt ≥95% der markierten Bauteile ab, Ausfälle <20 PPM, menschliche Nachprüfung ≥85% reduziert

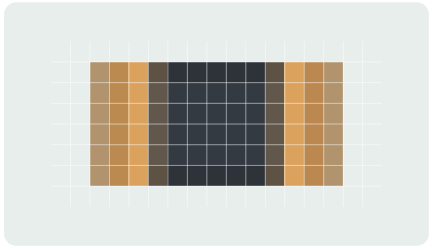
fail-open zu Menschen, Linie läuft weiter

Quelle: Siemens-Fallstudie, iFactory (Herstellerangabe), DBG-Geschäftsbericht 2025; Prüfstunden vom Autor geschätzt.

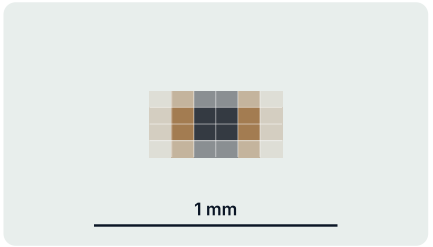
Eine Allzweckkamera ist keine AOI



AOI: 8–20 µm pro Pixel
0402 ≈ 100×50 px



4K-Kamera über ein 300-mm-Panel: ≈78 µm/px
0402 schrumpft auf ≈13×6 px: genug für Anwesenheit und grobe Verschiebung

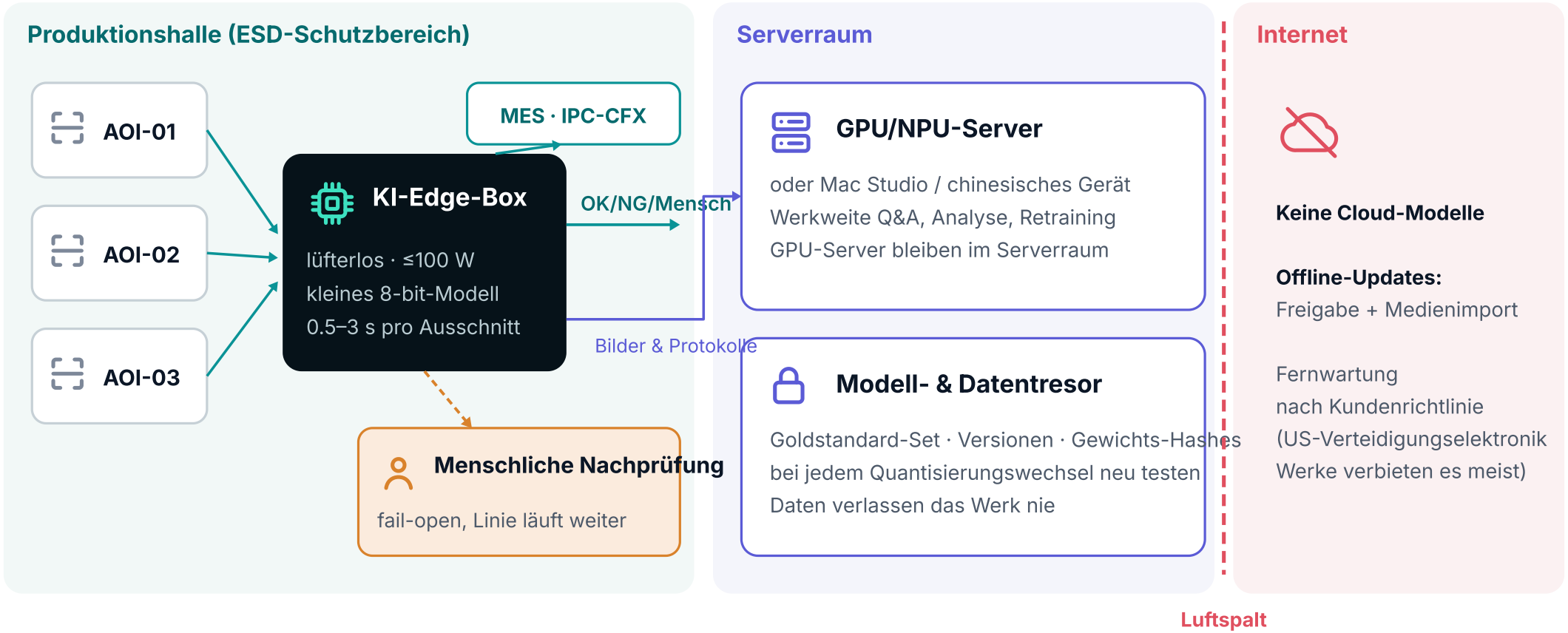


Dieselbe Kamera auf einem 01005
≈5×3 px: praktisch unsichtbar

- **AOI:** 8–20 µm pro Pixel, mehrwinklige mehrfarbige Beleuchtung und strukturiertes Licht
- **4K über eine 300-mm-Platine:** ≈78 µm pro Pixel; ein 0402 ist ≈13×6 Pixel, genug für die Anwesenheit, nicht für Lötstellen
- **01005:** ≈5×3 Pixel, praktisch unsichtbar
- **Also:** die AI arbeitet auf den eigenen Bildern des AOI; sie ersetzt nicht die Optik

Quelle: Circuits Assembly; Berechnung des Autors für ein 300-mm-Panel und einen 4K-Sensor. Die Pixelraster sind illustrativ.

Eine luftspaltgetrennte Bereitstellung: die Box entscheidet, der Serverraum antwortet

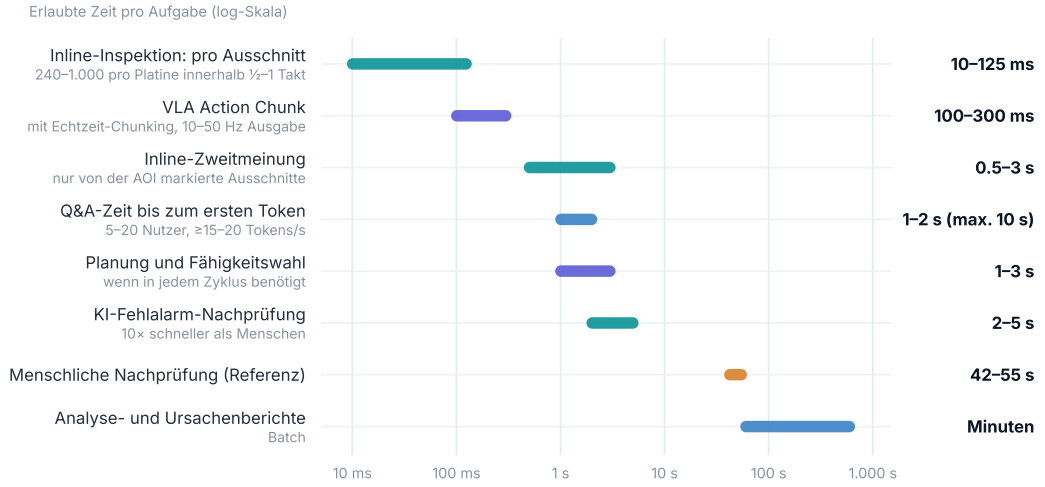


Quelle: Zusammenfassung des Autors (Kapitel 2).

Reinraum- und ESD-Zonen entscheiden, wo Rechenleistung steht; Geschwindigkeitsgrenzen hängen vom Einsatz ab

Sauberkeit →

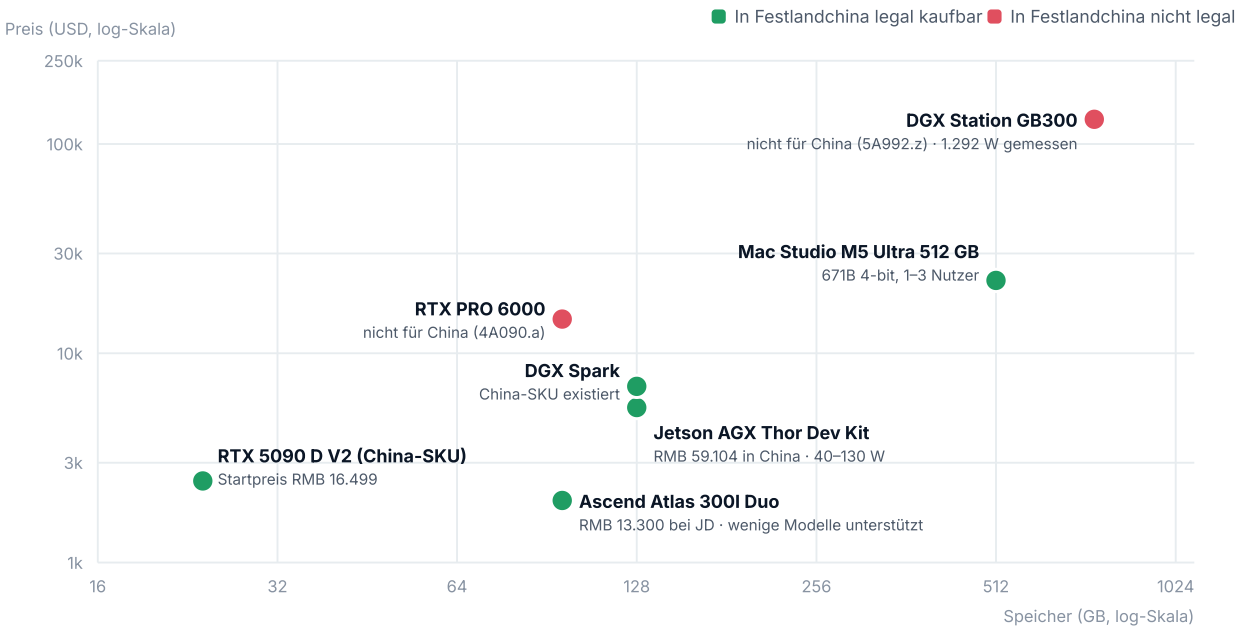
Normale SMT/PCBA-Halle	Reinraum ISO 7–8	Mini-Umgebung ISO 5–6	Fab, ISO 5 oder sauberer
ESD-Schutzbereich (≈ISO 8) **Luftfeuchte 40–60%**	Medizin, Kameramodule, EV **Reinraum + ESD**	Bare Die, Drahtbonden, Optik **Mini-Umgebung**	SEMI E78 ESD · E187/E188 Cyber **fabspezifische Pläne**
Liniennahe Rechenleistung Lüfterlose, geerdete Metallbox neben der AOI oder dem Prüfplatz	**Liniennahe Rechenleistung** Lüfterlos, IP65/66-dicht, leicht zu reinigen; ISO 14644-14-Daten anfordern	**Liniennahe Rechenleistung** Nur Kameras und Licht; Rechenbox außerhalb	**Liniennahe Rechenleistung** Nur toolintegrierte, fabqualifizierte Controller
GPU/NPU-Server Serverraum oder klimatisiertes IT-Rack	**GPU/NPU-Server** außerhalb des Reinraums, vernetzt hinein	**GPU/NPU-Server** Serverraum	**GPU/NPU-Server** Flüssiggekühlte Racks im Sub-Fab oder Rechenzentrum



Kernpunkt Eine gewöhnliche SMT-Halle ist ein ESD-Schutzbereich, kein klassifizierter Reinraum: eine lüfterlose Box in einem geerdeten Metallgehäuse steht am AOI- oder Prüfplatz, und GPU-Server kommen nie an die Linie.

Quelle: ANSI/ESD S20.20, IEC 61340-5-1, ISO 14644, SEMI-Normen und Herstellerunterlagen; vom Autor zusammengestellt. Zusammenfassung des Autors (Kapitel 1–2).

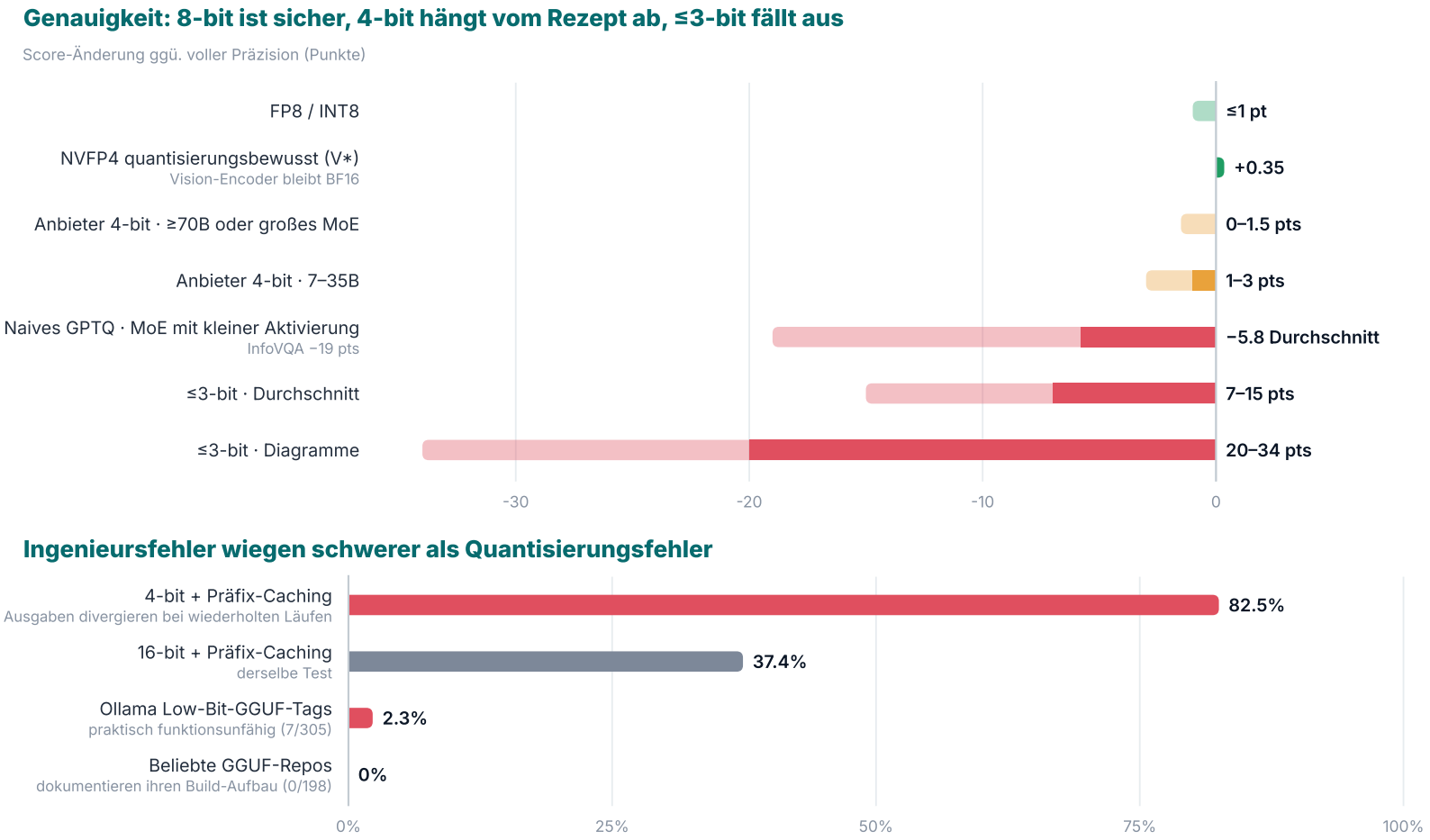
Exportregeln bestimmen die Hardware-Shortlist, bevor Benchmarks es tun



- **Jetson Thor:** 128 GB, 40–130 W; 2–8B-Vision-Modelle auf AOI-Ausschnitten, VLAs der 3B-Klasse in Echtzeit; 59.104 RMB in China
- **Mac Studio 512 GB:** DeepSeek 671B in 4-Bit für 1–3 Nutzer
- **DGX Station:** 5A992.z, nicht für das Festland; ein 743B-Modell liefert 32 Nutzern ≈4 Tokens/s je Nutzer
- **Inländische Beschleuniger:** zwei Ascend-Server mit acht Karten für vollwertige Multi-User-Q&A

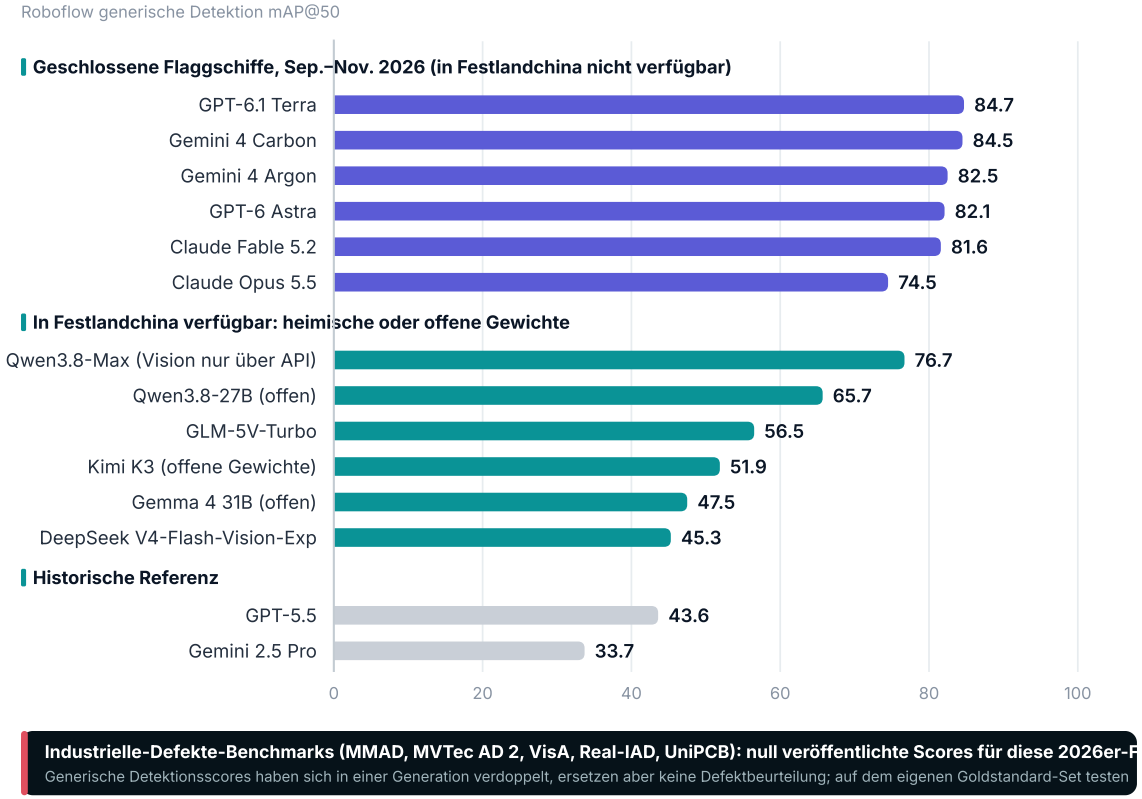
Quelle: NVIDIA-Exportklassifizierungen, Apple, RS China, JD.com und Testpreise (November 2026). StorageReview (Aug. und Okt. 2026); ServeTheHome. Unabhängige Tests aus Einzelquellen.

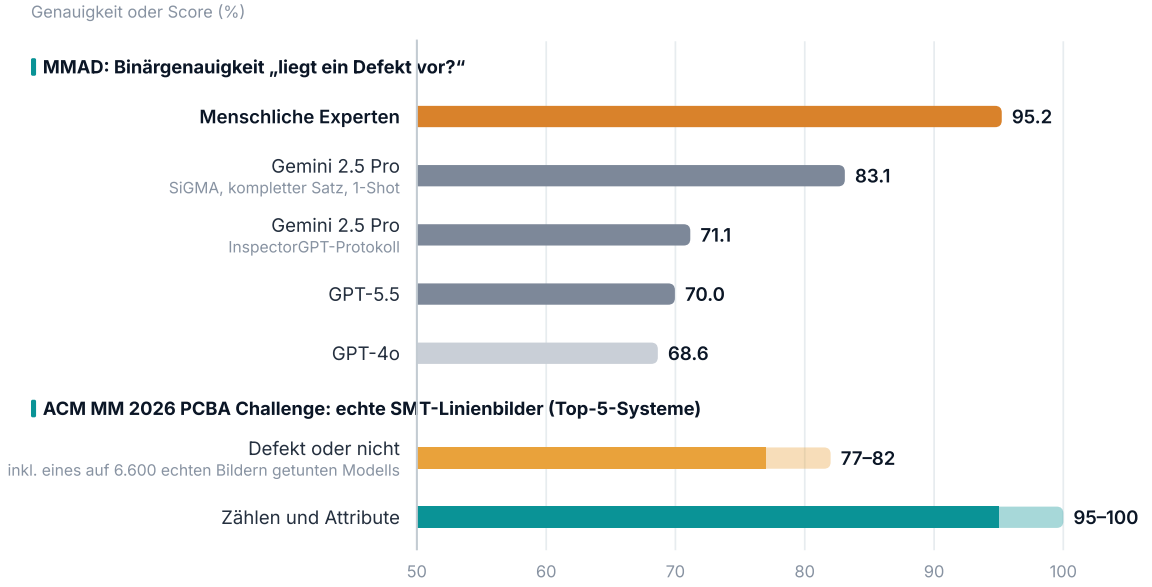
Quantisierung: 8-Bit ist sicher, 4-Bit hängt vom Rezept ab, 3-Bit und darunter fällt aus



Quelle: NVIDIA-Nemotron-Modellkarte, VLMQ, VEQ, arXiv 2609.05881, 2610.19498, 2609.04748 u. a.

Generische Bilderkennung verdoppelte sich in einer Generation; Werte für Industriedefekte fehlen





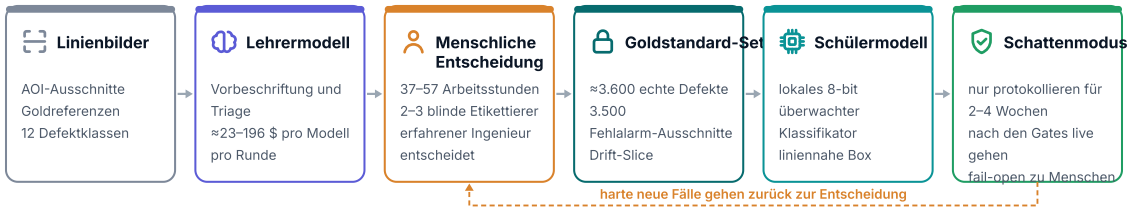
Industrielle-Defekte-Benchmarks (MMAD, MVTec AD 2, VisA, Real-IAD, UniPCB): null veröffentlichte Scores für diese 2026er-F

Generische Detektionsscores haben sich in einer Generation verdoppelt, ersetzen aber keine Defektbeurteilung; auf dem eigenen Goldstandard-Set testen

Quelle: Roboflow Vision Evals; Modell-Release-Notes; MMAD, MVTec AD 2 und weitere Benchmark-Seiten (Stand 11. Nov. 2026). MMAD; SiGMA; InspectorGPT; ACM MM 2026 PCBA Challenge.

Frontier-Modelle beschriften vor, Menschen entscheiden, ein Golden Set entscheidet über die Freigabe

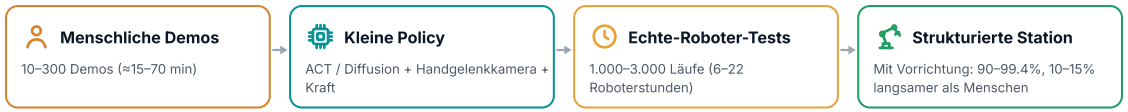
Inspektion: das Frontier-Modell beschriftet vor, Menschen entscheiden, ein Goldstandard-Set gated

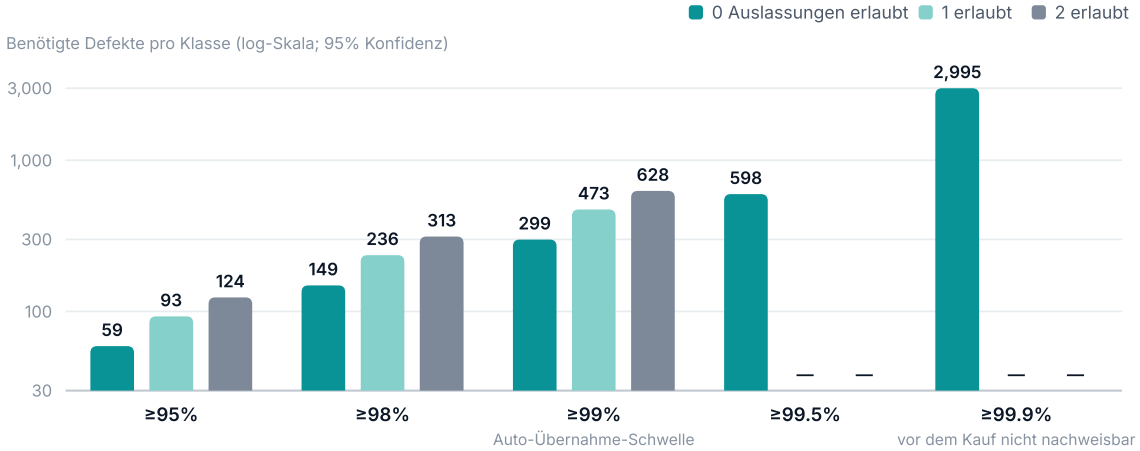


Hürde für die automatische Übernahme von Lehrerlabels

Pro Klasse: ≥299 Defekte, null Auslassungen (Recall ≥99% bei 95% Konfidenz), Übereinstimmung nicht schlechter als Mensch–Mensch; sonst nur Vorbeschriftung

Arme: Die Lehrer sind weiterhin menschliche Demos, echtes Roboter-RL und Simulation



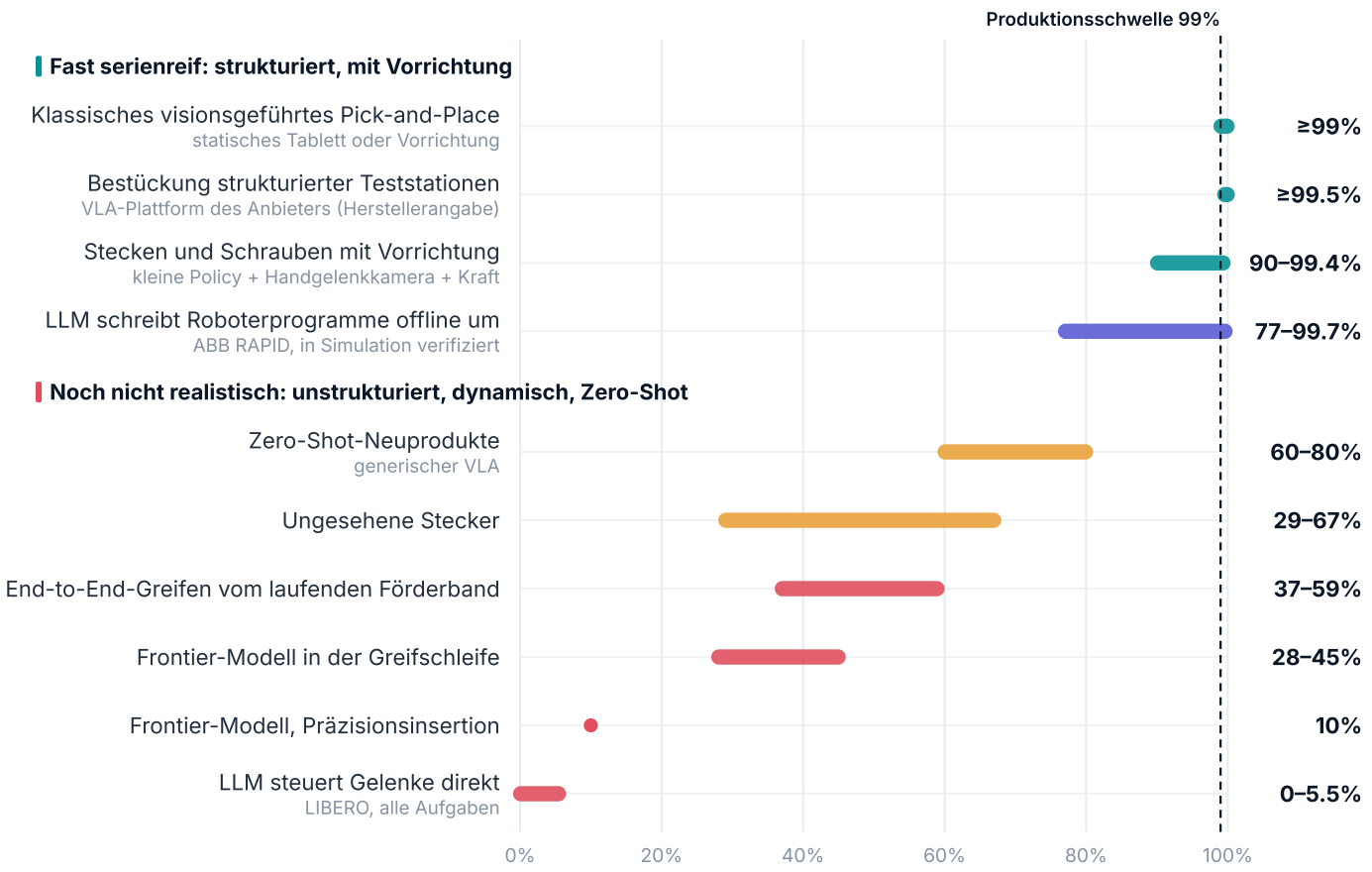


Kernpunkt Die Labels des Lehrers nur für eine Klasse mit null Übersehenen bei ≥299 Golden-Set-Defekten automatisch übernehmen (99% Recall bei 95% Konfidenz); kritische Klassen wie Kurzschlüsse, fehlende Teile und Polarität brauchen mehr.

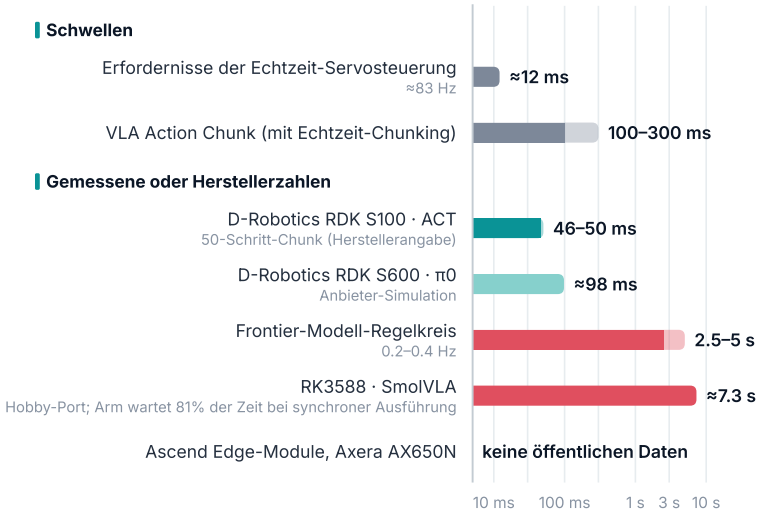
Quelle: Zusammenfassung des Autors (Kapitel 3); Kosten und Stichprobengrößen sind Schätzungen. Exakte binomiale Konfidenzintervalle (95%).

Roboter-Schüler: fast serienreif nur bei strukturierten Stationen

Aufgabenerfolgsrate



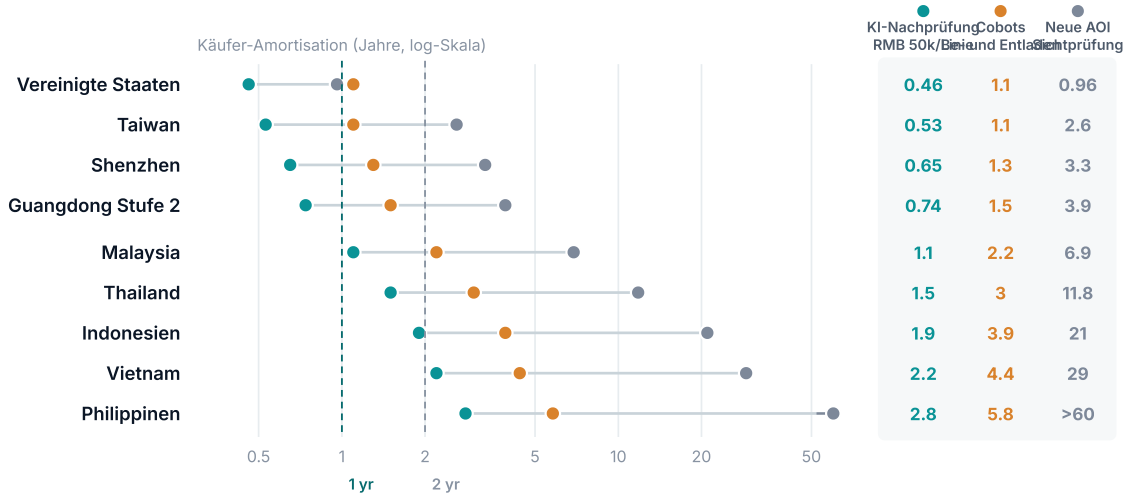
Inferenzlatenz pro Action Chunk oder Regelzyklus (log-Skala)



Quelle: In Kapitel 3 zitierte Paper und Herstellerunterlagen; einige Zahlen sind Herstellerangaben. D-Robotics-Dokumentation und -Forum (Herstellerangaben); GitHub-Community-Ports; Anthropic-Forschung.

AI-Nachprüfung amortisiert sich in Monaten; Maschinenbeschickung in etwa einem Jahr



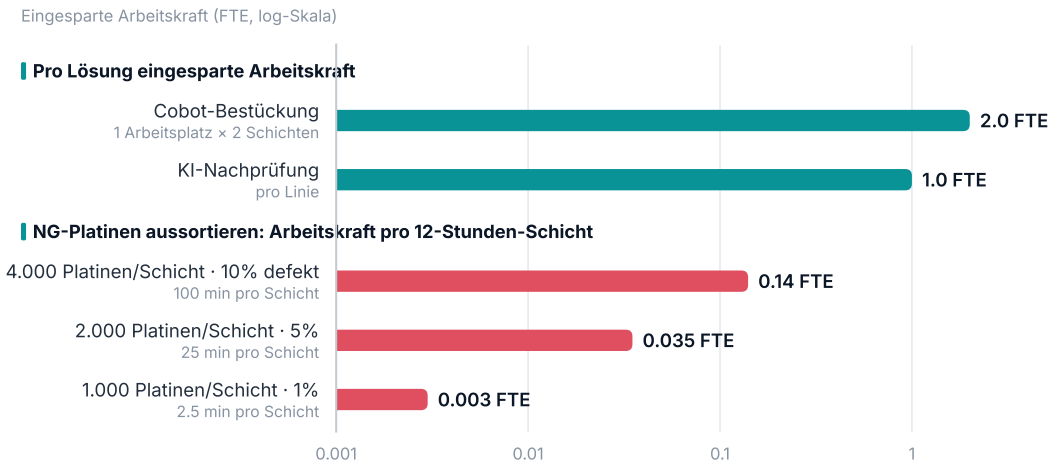


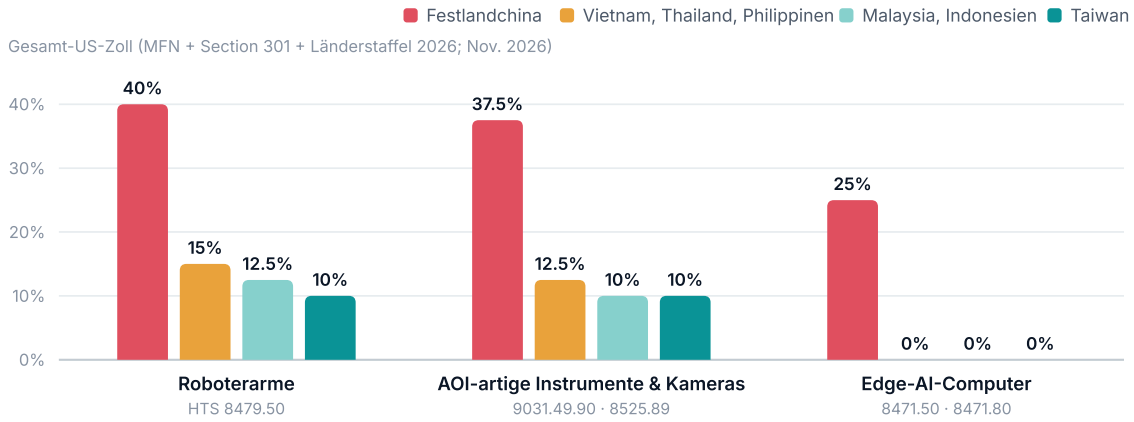
Ein Roboterarm nur zum Aussortieren von NG-Platinen amortisiert sich nirgendwo
Bei Erstausbeute über 90% kostet das Aussortieren nur 0.003–0.14 FTE pro Schicht

Kernpunkt Bei 50.000 RMB pro Linie amortisiert sich die AI-Nachprüfung in 0.46 Jahren in den USA, 0.53 in Taiwan, 0.65 in Shenzhen und 2.2 in Vietnam; eine Cobot-Beschickungszelle im Zweischichtbetrieb in 1.1–1.5 Jahren [estimates].

Quelle: Offizielle Lohnleitlinie Shenzhen, taiwanesisches Arbeitsministerium, JETRO, BLS; Schätzungen des Autors. Modell des Autors: KI-Nachprüfbox 50k RMB pro Linie; Cobot-Bestückungszelle für 200k RMB oder 140k $, zwei Schichten.

Ein Ausschleusarm amortisiert sich nie; wo ein Arm gefertigt wird, ändert seinen US-Zoll





Kernpunkt Bei über 90% Erstausbeute dauert das Ausschleusen von NG-Platinen 0.003–0.14 einer Person pro Schicht, sodass eine NG-Ausschleusung oder ein Puffer die Aufgabe übernimmt; ein auf dem Festland gefertigter Arm zahlt zudem etwa 40% kombinierten US-Zoll.

Quelle: Schätzungen des Autors. USITC Harmonized Tariff Schedule Kapitel 99; Federal Register; USTR (November 2026).

Real, aber klein: 16.000 adressierbare Linien, ≈320 Mio. RMB einmalig, ein Fünfjahresfenster

SMT-Linien in Festlandchina (2026)

40.000 Zeilen

35k–46k

× 80% × 50%

noch keine KI: 75–85%
erreichbar: 40–60%

Bedienbar

16.000 Zeilen

10.5k–23.5k

× RMB 20k

RMB 10–30k einmalig

Einmaliger Bestandsmarkt

RMB 320M

RMB 105–705 Mio.

÷ ≈5 Jahre

schrumpft 8–12%/Jahr

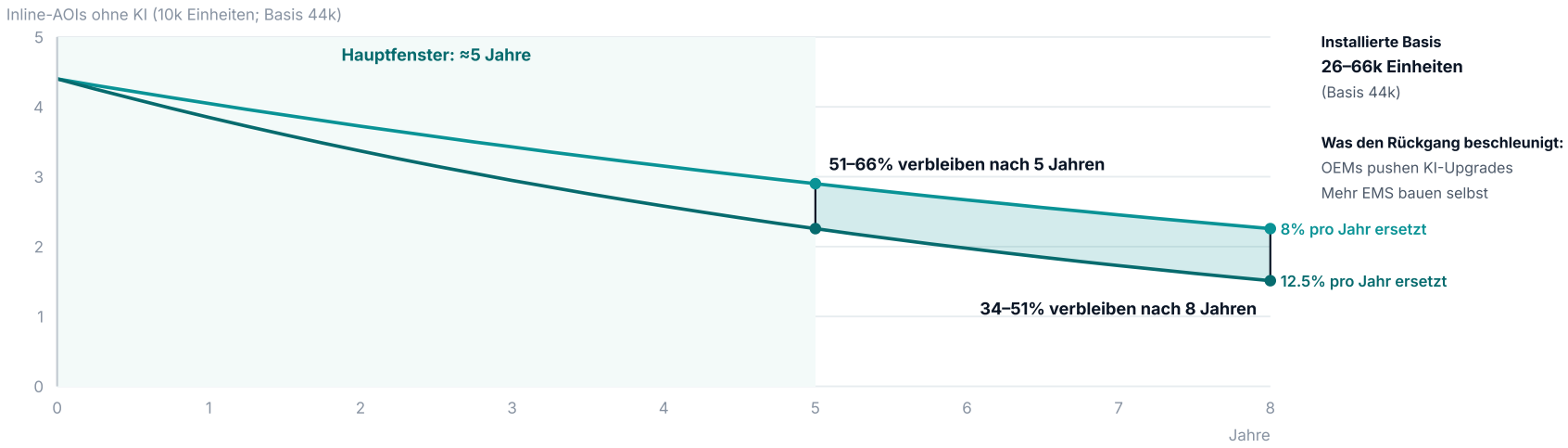
5-Jahres-Fenster

RMB 64M/yr

RMB 21–141 Mio.

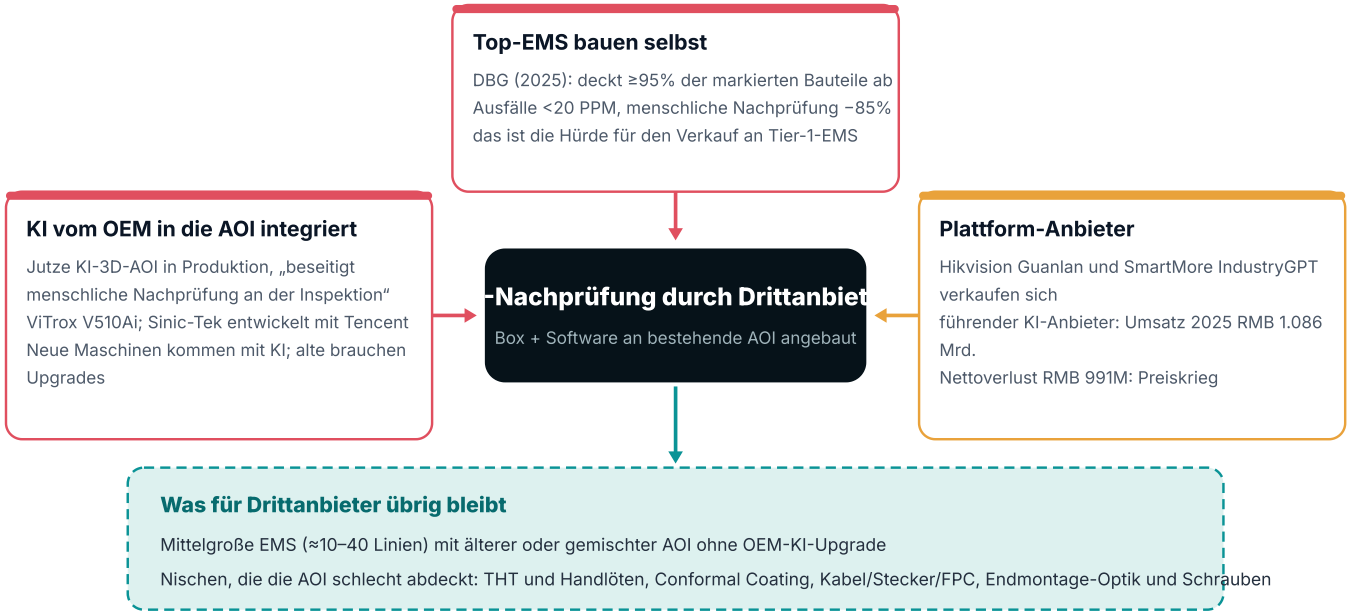
Ein Neueinsteiger gewinnt in 3 Jahren 2–5% der bedienbaren Linien: ≈320–800 Linien, ≈RMB 2–5 Mio. pro Jahr

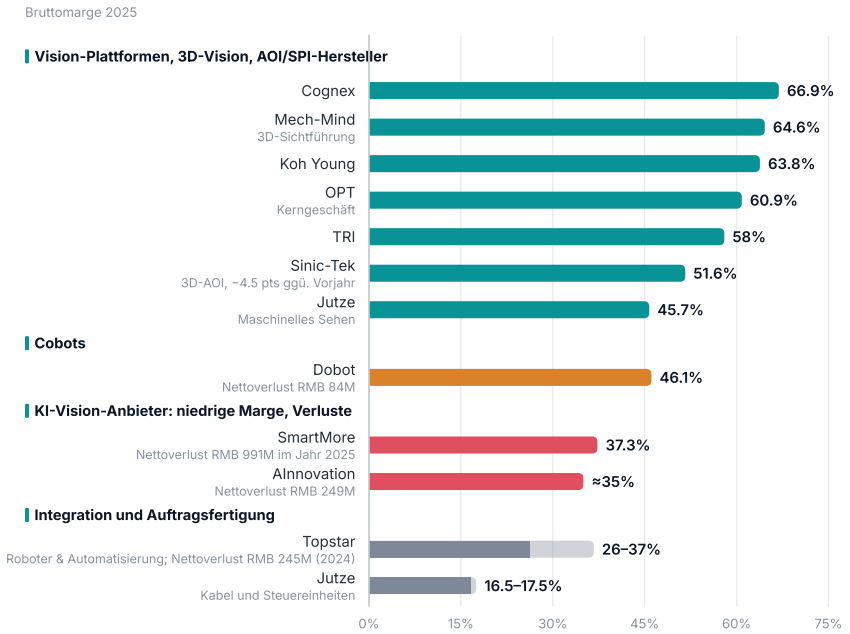
Als Abo (RMB 5k pro Linie und Jahr): RMB 1.6–4 Mio. wiederkehrend, nur 0.3–2.3% der IDC-Prognose für Chinas industrielle KI-QC



Quelle: NVR (Basis 2020); Schätzungen des Autors. Schätzungen des Autors; installierte Basis und Anteile nicht unabhängig verifiziert.

Der Review-Keil ist umzingelt; die Marge liegt in produktisierter Hardware

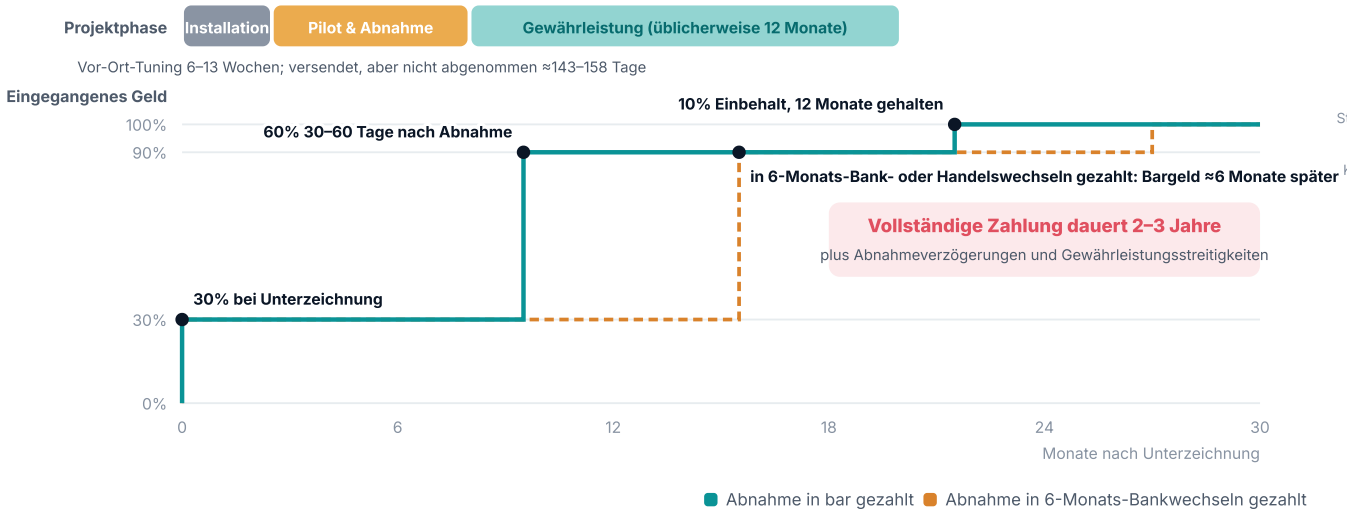


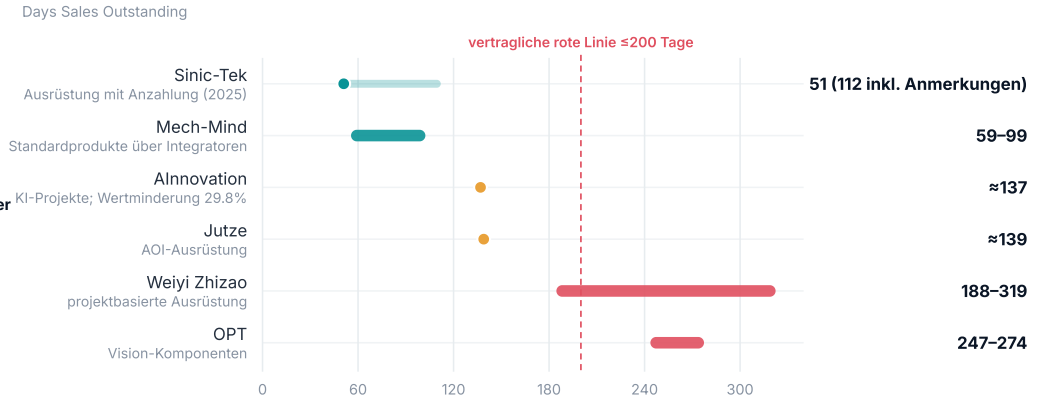


Belege Der führende AI-Vision-Anbieter erzielte 2025 einen Umsatz von 1.086 Milliarden RMB und einen Nettoverlust von 991 Millionen RMB; der 3C-Umsatz des nächsten börsennotierten Vergleichsunternehmens fiel im Januar–Mai 2026 um 79.9% im Jahresvergleich.

Quelle: Geschäftsberichte von Jutze, Sinic-Tek und DBG; öffentliches Material von Hikvision und SmartMore; vom Autor zusammengestellt. Geschäftsberichte oder Prospekte 2025.

Käufer zahlen 30/60/10; die vollständige Zahlung dauert zwei bis drei Jahre





Kernpunkt 30% kommen mit der Bestellung, 60% etwa vier bis neun Monate später und die letzten 10% etwa 16–21 Monate später; vom ersten Pilotprojekt bei einem großen EMS bis zur vollständigen Zahlung können zwei bis drei Jahre vergehen [estimate].

Quelle: Antwort auf Anfrage bei Sinic-Tek, Mech-Mind, GKG, JT und weitere Unterlagen; schematisch zu üblichen Konditionen, Zeitangaben geschätzt. Unternehmensunterlagen; Forderungen (inkl. Vertragsvermögen) ÷ Umsatz × 365.

Offline planen, mit kleinen Policies handeln, Bewegung und Sicherheit deterministisch halten

Ebene	Was es tut und wie weit die Evidenz reicht	Zeitskala · wer sie besitzt
Planung L4	**Frontier-Modelle (offline)** Roboterprogramme offline schreiben oder ändern, Umrüstungen, Fehler erklären Schreibt ABB-RAPID-Programme zu 77–99.7% um, nach Simulation bereitgestellt Regelkreis nur 0.2–0.4 Hz: bleibt aus der Echtzeit	**Sekunden bis Minuten** Modell-Giganten
Fähigkeiten / Policy L3	**ACT, Diffusionspolicies, kleine VLAs** Strukturierte Stationen mit Vorrichtung 90–99.4% Edge-Chips: Jetson Thor, D-Robotics RDK, Ascend, Rockchip	**Action Chunks 100–300 ms** Chip-Anbieter · Roboterhersteller
Bewegungssteueru L2	**Robotersteuerung** Trajektorien, Servo, Kraftregelung, Kollisionsprüfung deterministischer Code; gelernte Policys setzen nur Ziele	**Millisekunden (Echtzeit braucht >83 Hz)** Roboterhersteller
Sicherheit L1	**Sicherheits-PLC, Scanner, Kraft- und Geschwindigkeitsbegrenzung** zertifiziert nach ISO 10218-1/-2:2025 und lokalen Vorschriften Kein ML in Sicherheitsfunktionen: EU-Bewertung durch benannte Stelle ab 20. Januar 2027	**harte Echtzeit** Zertifizierter Roboter + Integrator

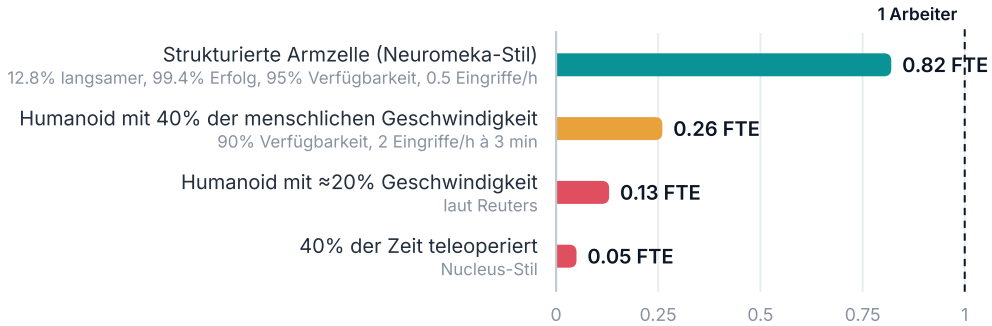
Kleine Firmen gewinnen bei Integration und Sensorik-Hardware: gelernte Wahrnehmung und kleine Policys in zertifizierte Roboterzellen bringen, Station für Station

	Erfolgsrate	Takt	Eingriffe / Teleoperation	Verfügbarkeit / Fehler	Serienlauf
Figure 02 @ BMW	nur Ziel >99%/Schicht	nur Ziel 84 s	nur Ziel 0/Schicht	k. A.	1.250 h
Agility Digit @ GXO	k. A.	k. A.	k. A.	k. A.	100.000 Behälter
Apptronik Apollo @ Mercedes	k. A.	k. A.	k. A.	k. A.	k. A.
Agile ONE @ Audi	k. A.	k. A.	k. A.	k. A.	k. A.
Nucleus (ungenanntes Werk)	k. A.	k. A.	≈40% teleoperiert*	k. A.	≈2 h Video
AgiBot G2 @ Longcheer	99.99%	18–20 s pro Schritt	k. A.	ein Fehler, 2 min	64+ h
Neuromeka Hybridzelle	322/324	12.8% langsamer	k. A.	k. A.	324 Zyklen
π0.6 @ Kundeneinsatzort	—	—	96.4% autonom	k. A.	1 Schicht

* Methode nicht offengelegt

Offengelegt (meist Herstellerangaben) · nur Ziele · k. A.

Netto ersetzte Arbeitskraft pro Roboter und Schicht



Quelle: Synthese des Autors (Kapitel 3, 6) und Schätzungen; ISO 10218:2025; EU-Maschinenverordnung; öffentliches Material von BMW, Figure, Agility, Mercedes-Benz und anderen.

Eng bauen, auf den eigenen Bildern der Fabrik beweisen, dann skalieren

Bauen	AI-Nachprüfung und Defektklassifikation auf AOI-Bildern für mittelgroße EMS mit älteren oder gemischten AOI-Marken; Inspektion, die das AOI schlecht abdeckt; Cobot-Zellen, wo jedes Teil gehandhabt wird
Nicht bauen	Eine allgemeine AI-plus-Arm-Plattform; ein LLM als primärer Prüfer; ein Ausschleusarm an einer SMT-Linie; Humanoide und Robots-as-a-Service; Frontier-Modelle in Regelkreisen oder Sicherheitsfunktionen

Pilotprojekt

1. **Wochen 0–2:** Datenvereinbarung, Defektklassen, Ausgangsbasis
2. **Wochen 1–4:** Golden Set, blind beschriftet und beurteilt
3. **Wochen 2–4+:** Schattenmodus, je Klasse verglichen
4. **Wochen 4–8:** unterstützter Betrieb nur bei bestandenen Klassen
5. **Dann:** schriftliche Abnahme, Linie für Linie skalieren, bei jeder Änderung erneut validieren

Sicherheit Kein maschinelles Lernen in Sicherheitsfunktionen: Die EU verlangt ab dem 20. Januar 2027 die Bewertung durch eine benannte Stelle, und AI-Zertifizierungswege werden sich vor 2027–2029 nicht festigen.

Quelle: Bericht, Kapitel 6–7 und Zusammenfassung; ISO, IEC, UL, EUR-Lex, SAC und nationale Regulierungsdatenbanken.

ÜBER DEN AUTOR

Tan Shuai

Technologe und Gründer, mit mehr als zwanzig Jahren Erfahrung in Chips, Hardware und Systemsoftware sowie 20 erteilten Patenten. Seine Arbeit führte ihn durch das Silicon Valley, New York, Chicago, Taiwan, die Greater Bay Area (Shenzhen und Hongkong), Vietnam und Malaysia, mit einem langjährigen Schwerpunkt auf fortschrittlicher Fertigung und Lieferketten. Heute arbeitet er als AI-Forscher und -Entwickler und konzentriert sich darauf, wie AI in der Fabrik eingesetzt wird.

Der vollständige Bericht (alle Abbildungen, Tabellen und Quellen) und der Artikel sind unter tanshuai.com verfügbar. Er freut sich über Gespräche mit Menschen, die in der Fertigung mit AI arbeiten: i@tanshuai.com

