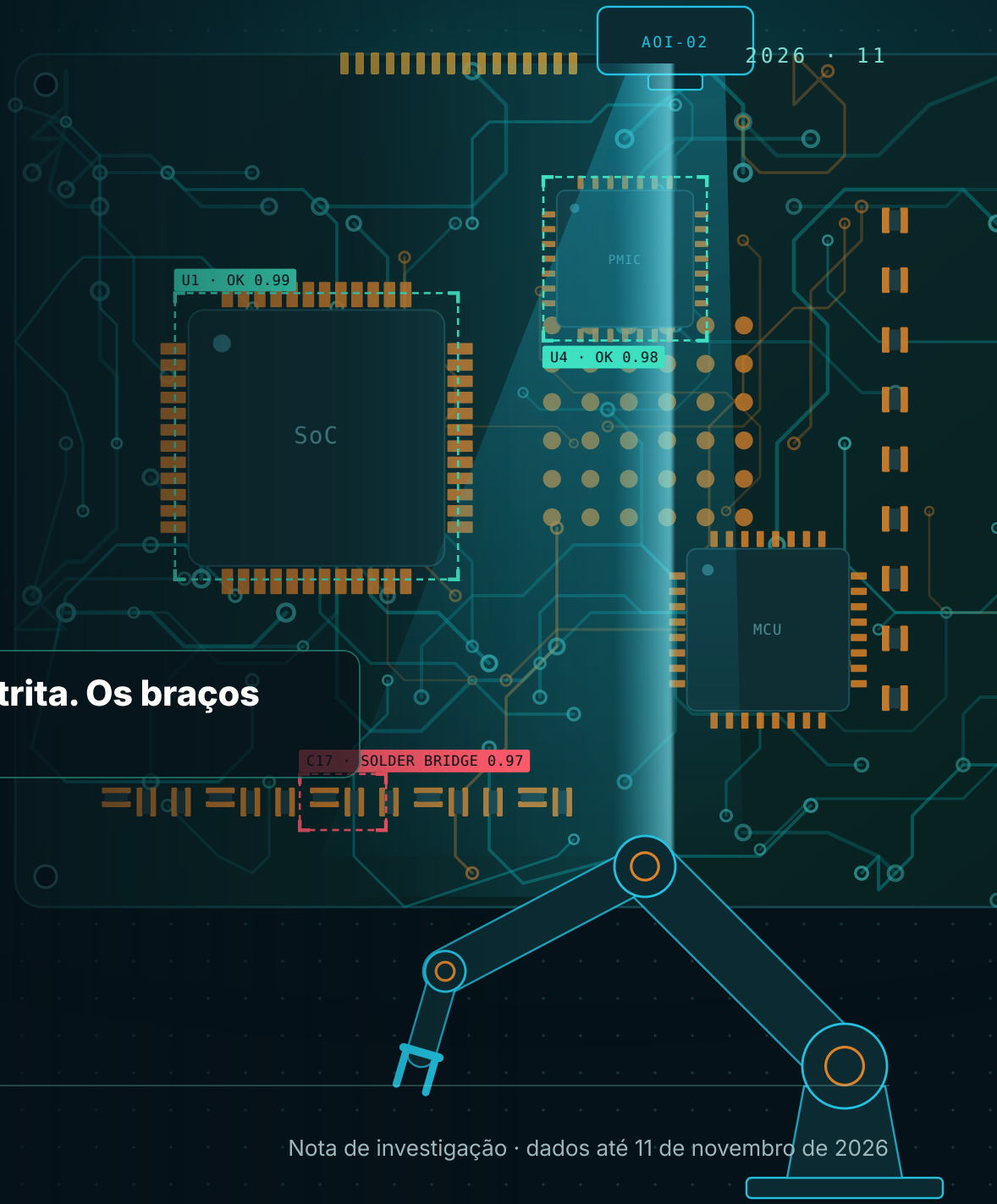
FABRICO DE ELETRÓNICA · INSPEÇÃO COM IA × ROBÓTICA

A IA na linha SMT:

o que se amortiza em 2026

Inspeção visual com IA e braços robóticos em fábricas de eletrónica

China continental · Taiwan · Sudeste Asiático · Estados Unidos



CONCLUSÃO PRINCIPAL

Vale a pena fazer revisão com IA, de forma restrita. Os braços incorporados podem esperar.

Tan Shuai
tanshuai.comi@tanshuai.com

Nota de investigação · dados até 11 de novembro de 2026

Vale a pena, mas de forma restrita, agora e com parceiros

CONCLUSÃO PRINCIPAL

O que está maduro em 2026, e pelo que as fábricas pagarão, é a revisão com IA e a inspeção de nicho sobre imagens de AOI, mais a alimentação programada de máquinas com robôs, não «um modelo grande a olhar para placas mais robôs incorporados a apanhá-las».

0.46–0.65 yr

Retorno da revisão com IA nos EUA, em Taiwan e em Shenzhen, a RMB 50 mil por linha

≥95% · <20 PPM

A revisão com IA interna de uma EMS de topo: a fasquia para vender a EMS de nível 1

100–700 RMB M

Mercado de retrofit de terceiros aproveitável na China continental, pagamento único, janela de ≈5 anos

0 scores

Pontuações em benchmarks de defeitos industriais dos modelos de topo de novembro de 2026

1.1–1.5 yr

Retorno de uma célula de alimentação de máquinas com cobot, dois turnos

0.003–0.14 FTE

Mão de obra que um braço robótico de retirada poupa por turno: nunca se amortiza

Fonte: sumário executivo do relatório e capítulos 4–5; os valores de retorno e de mercado são estimativas do autor.

Seis questões, seis respostas

① O que as fábricas executam

SPI 3D, AOIs antes e depois do reflow, raio-X por amostragem, ICT/FCT. Atrás de cada AOI há uma pessoa a rever; takt ≈13–22 s por placa.

② Até onde vai a tecnologia

A inspeção é de grau industrial como detetor + revisão com IA + algumas pessoas; não zero pessoas a qualquer preço. Continente: modelos nacionais; Taiwan: o cliente decide; EUA: direitos aduaneiros e certificação.

③ No local e isolado da rede

Prático; os limites são de procedimento. Uma caixa sem ventoinha de ≤100 W na linha, servidores na sala de servidores, 8 bits por defeito.

④ Professor e aluno

Inspeção: os modelos de fronteira pré-etiquetam, as pessoas arbitram, um conjunto dourado faz de portão. Braços: os professores continuam a ser demonstrações humanas e simulação.

⑤ Custo e retorno

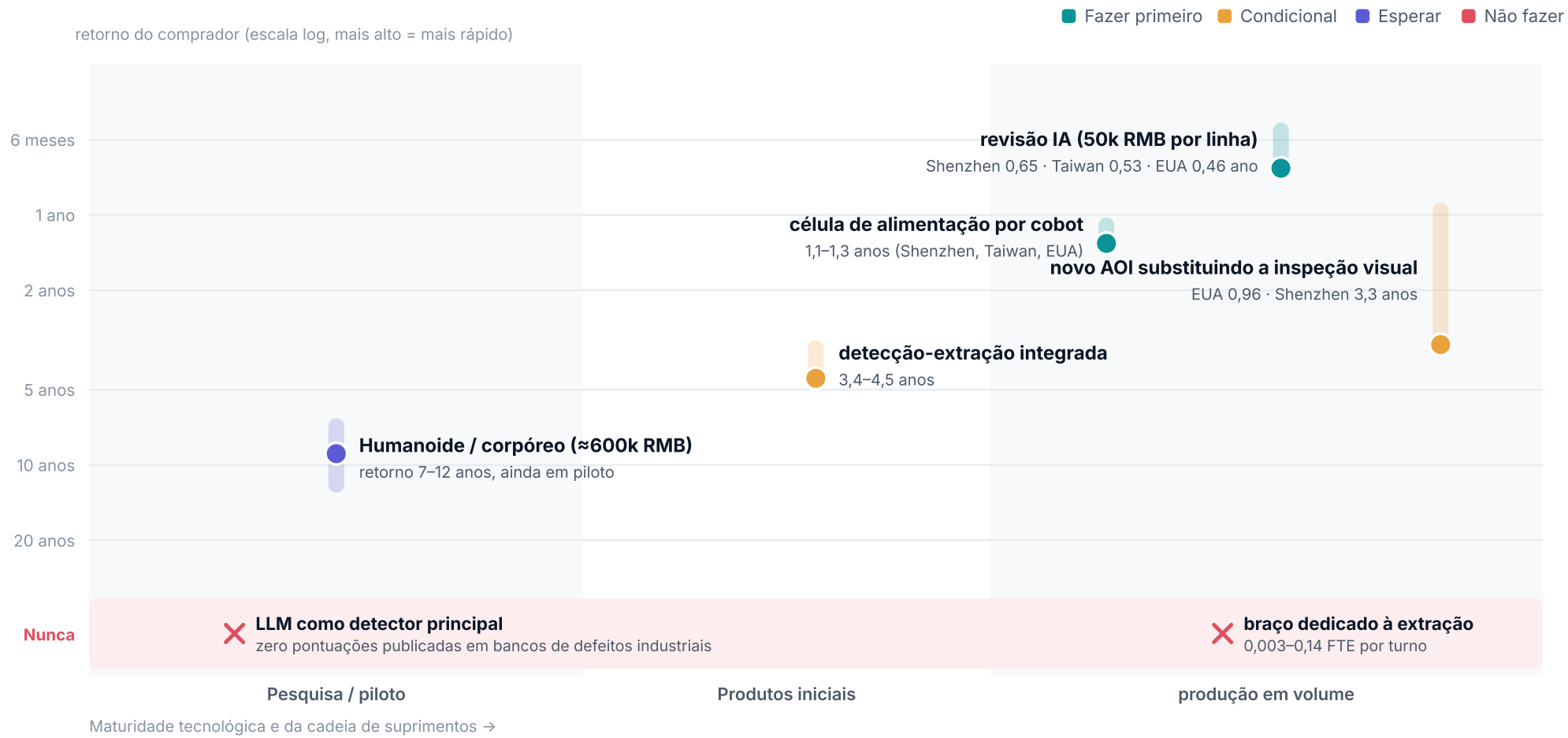
Revisão com IA em menos de um ano nos EUA, em Taiwan e em Shenzhen; alimentação com cobot 1.1–1.5 anos; Sudeste Asiático 2–6 anos; um braço de retirada, nunca.

⑥ Agora ou esperar

Procura real, oferta saturada; a IA dos OEM deixa uma janela de retrofit de ≈5 anos. Inspeção agora, de forma restrita; IA incorporada depois dos dados de 2027–28.

Fonte: sumário executivo do relatório.

Onde a IA se encaixa hoje no chão de fábrica

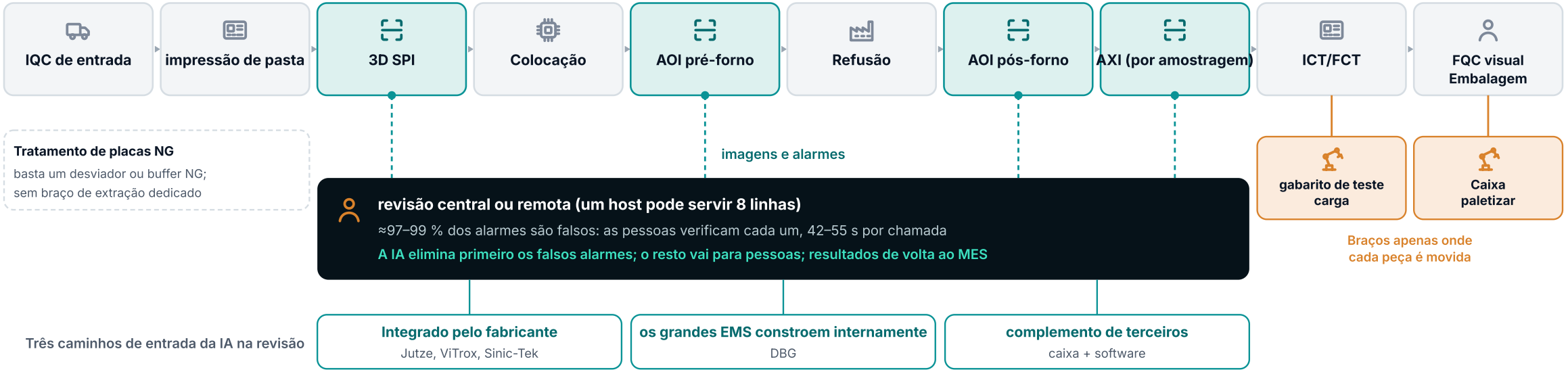


Fonte: Retorno do modelo do autor (capítulo 4); o eixo de maturidade é um julgamento do autor.

As pessoas sentam-se atrás do AOI. É aí que a IA entra

Uma linha SMT em volume

takt ≈13–22 s por placa · 3–5 AOI por linha



Fonte: Documentos públicos e material de fornecedores (Huaqin, DBG, Sinic-Tek e outros); compilado pelo autor.

Cerca de 97–99 % dos alarmes AOI são falsos

Para cada 100 alarmes AOI



Hoje: as pessoas verificam cada alarme

42–55 s cada

≈448 falsos alarmes por turno numa linha ocupada

→ 5,2–6,8 h por turno ≈ 1 revisor em tempo integral

lotação do caso: 2 por turno

revisão IA: eliminar falsos alarmes, passar o resto às pessoas

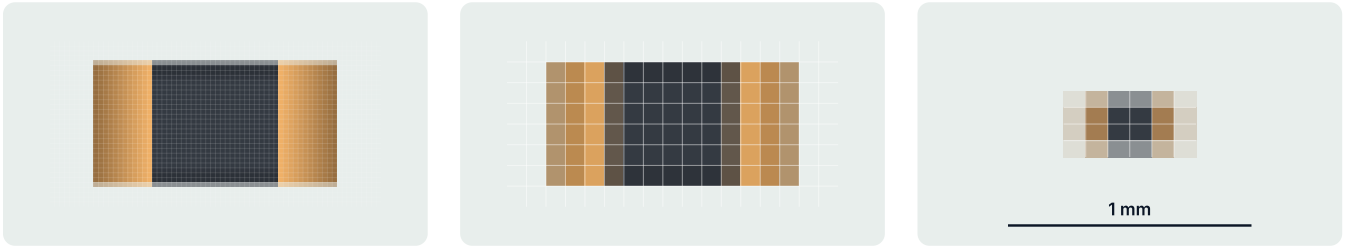
falsos alarmes ou horas de revisão reduzidos 60–90 % (alegações do fornecedor)

referência interna dos grandes EMS: cobre ≥95 % dos componentes sinalizados, escapes <20 PPM, revisão humana reduzida ≥85 %

abertura por falha para pessoas, a linha continua

Fonte: Estudo de caso Siemens, iFactory (alegação do fornecedor), relatório anual DBG 2025; horas de revisão estimadas pelo autor.

Uma câmera de uso geral não é um AOI



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

Câmera 4K sobre um painel de 300 mm: ≈78 µm/px
O 0402 encolhe para ≈13×6 px: só dá para presença e desvio grosseiro

Mesma câmera sobre um 01005
≈5×3 px: efetivamente invisível

- **AOI:** 8–20 µm por píxel, iluminação multiângulo multicolor e luz estruturada
- **4K sobre um painel de 300 mm:** ≈78 µm por píxel; um 0402 são ≈13×6 píxeis, suficientes para a presença, não para as soldaduras
- **01005:** ≈5×3 píxeis, efetivamente invisível
- **Portanto:** a IA trabalha sobre as próprias imagens do AOI; não substitui a ótica

Fonte: Circuits Assembly; cálculo do autor para um painel de 300 mm e um sensor 4K. As grades de pixels são ilustrativas.

Uma implantação isolada da rede: a caixa decide, a sala de servidores responde



Fonte: Síntese do autor (capítulo 2).

As zonas limpas e de ESD decidem onde fica o poder de computação; os limites de velocidade dependem do uso

Limpeza →

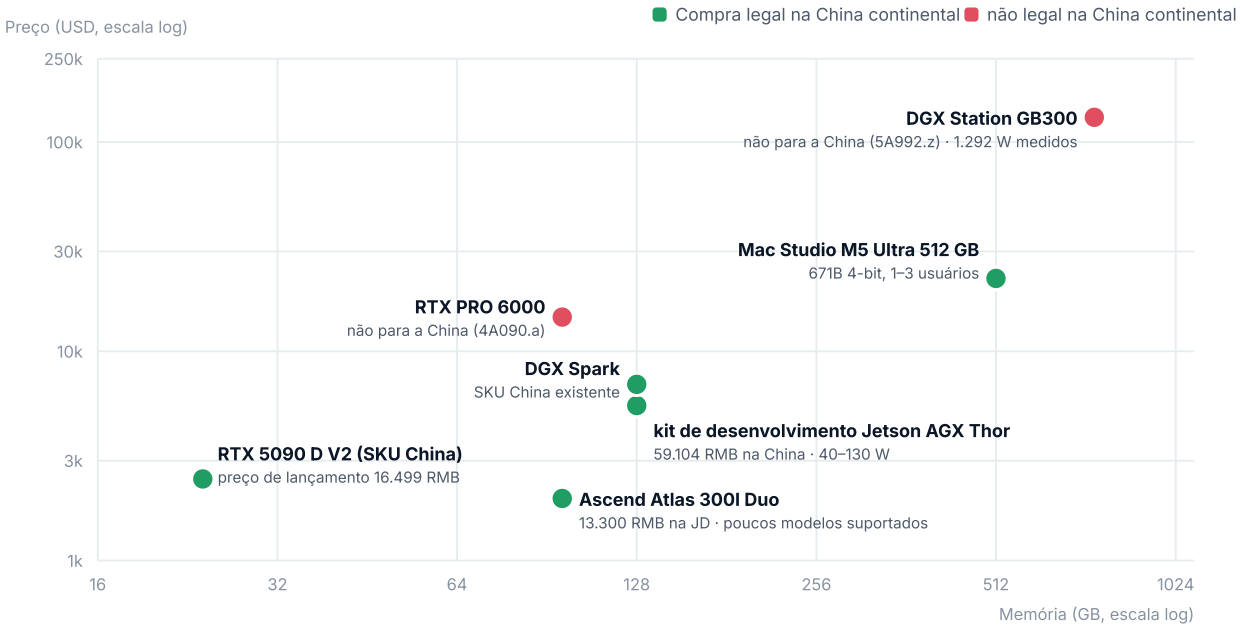
nave SMT/PCBA comum	sala limpa ISO 7–8	miniambiente ISO 5–6	fab, ISO 5 ou mais limpo
área protegida ESD (≈ISO 8) UR 40–60 %	médico, módulos de câmera, VE sala limpa + ESD	die nu, wire bond, óptica miniambiente	SEMI E78 ESD · E187/E188 ciber planos específicos de fab
computação na linha caixa sem ventoinha com invólucro metálico aterrado, perto do AOI ou do posto de revisão	**computação na linha** Sem ventoinha, vedado IP65/66, fácil de limpar; pedir dados ISO 14644-14	**computação na linha** Só câmeras e luzes; a caixa de cálculo fora	**computação na linha** apenas controladores integrados à ferramenta e qualificados para fab
servidor GPU/NPU sala de máquinas ou rack de TI climatizado	**servidor GPU/NPU** fora da sala limpa, rede para dentro	**servidor GPU/NPU** Sala de máquinas	**servidor GPU/NPU** racks refrigerados a líquido na subfab ou data center



Ponto-chave Uma sala SMT comum é uma área protegida contra ESD, não uma sala limpa classificada: uma caixa sem ventoinha numa caixa metálica ligada à terra fica junto ao AOI ou ao posto de revisão, e os servidores GPU nunca vão para a linha.

Fonte: ANSI/ESD S20.20, IEC 61340-5-1, ISO 14644, normas SEMI e material de fornecedores; compilado pelo autor. Síntese do autor (capítulos 1–2).

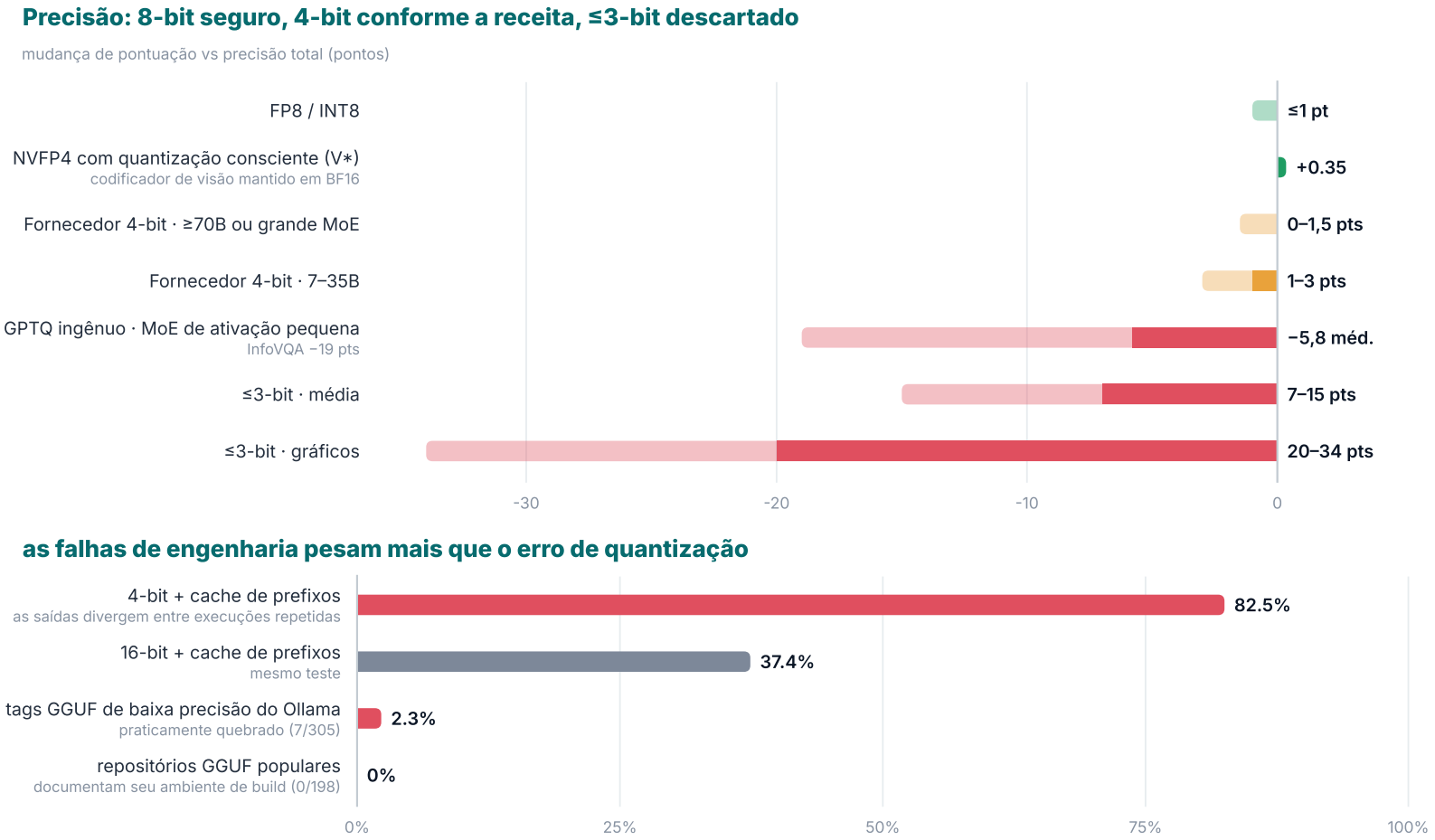
As regras de exportação escolhem a lista curta de hardware antes dos benchmarks



- **Jetson Thor:** 128 GB, 40–130 W; modelos de visão 2–8B sobre recortes de AOI, VLAs de classe 3B em tempo real; RMB 59,104 na China
- **Mac Studio 512 GB:** DeepSeek 671B a 4 bits para 1–3 utilizadores
- **DGX Station:** 5A992.z, não disponível para o continente; um modelo 743B dá a 32 utilizadores ≈4 tokens/s cada
- **Aceleradores nacionais:** dois servidores Ascend de oito placas para Q&A multiusuário em tamanho real

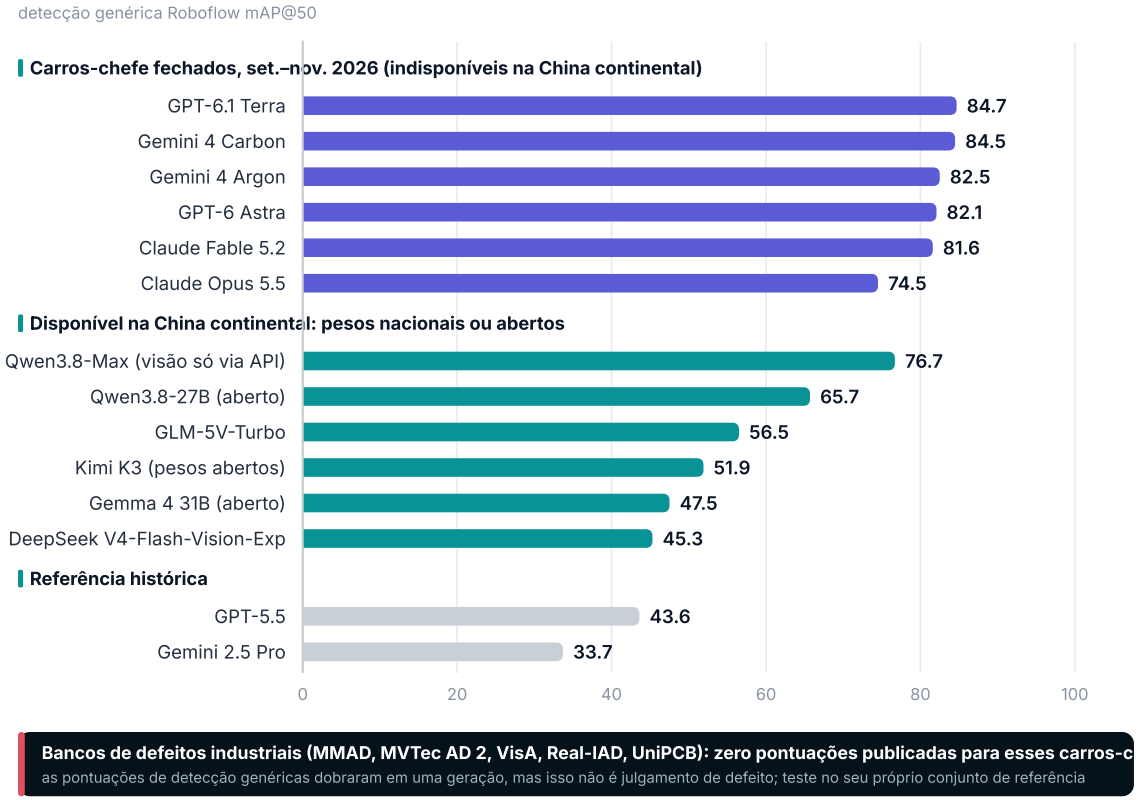
Fonte: Classificações de exportação da NVIDIA, Apple, RS China, JD.com e preços de revisão (novembro de 2026). StorageReview (ago. e out. 2026); ServeTheHome. Testes independentes de fonte única.

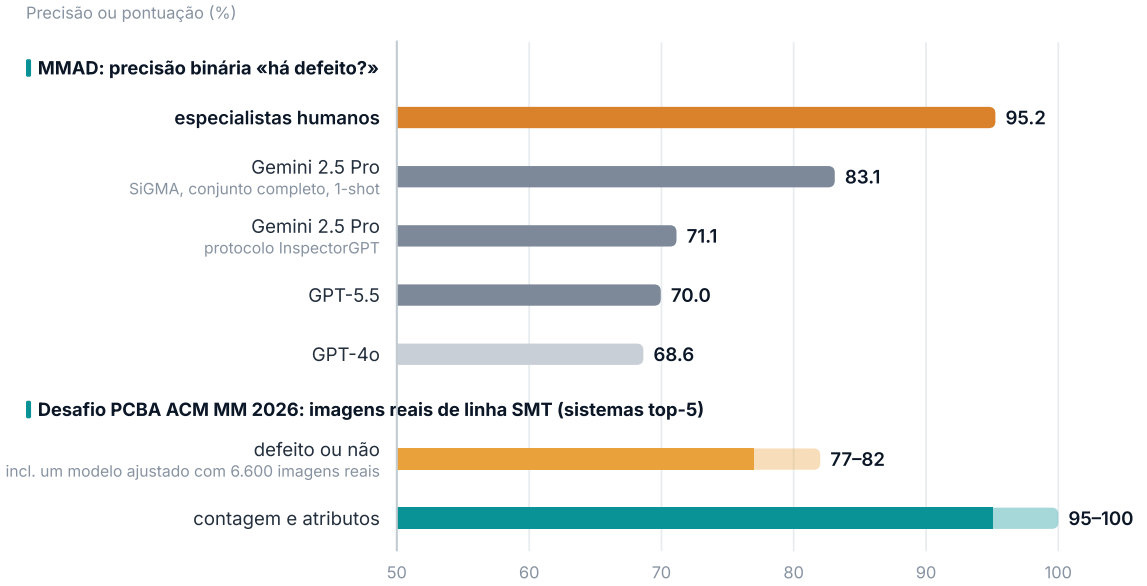
Quantização: 8-bit é seguro, 4-bit depende da receita, 3-bit e menos é descartado



Fonte: Ficha do modelo NVIDIA Nemotron, VLMQ, VEQ, arXiv 2609.05881, 2610.19498, 2609.04748 e outros.

A visão genérica duplicou numa geração; faltam as pontuações de defeitos industriais

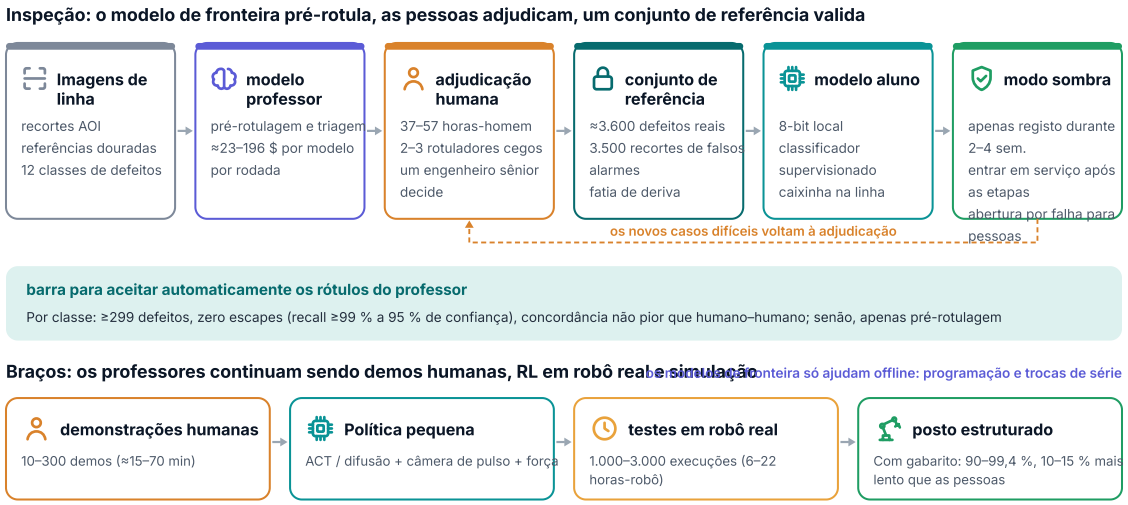


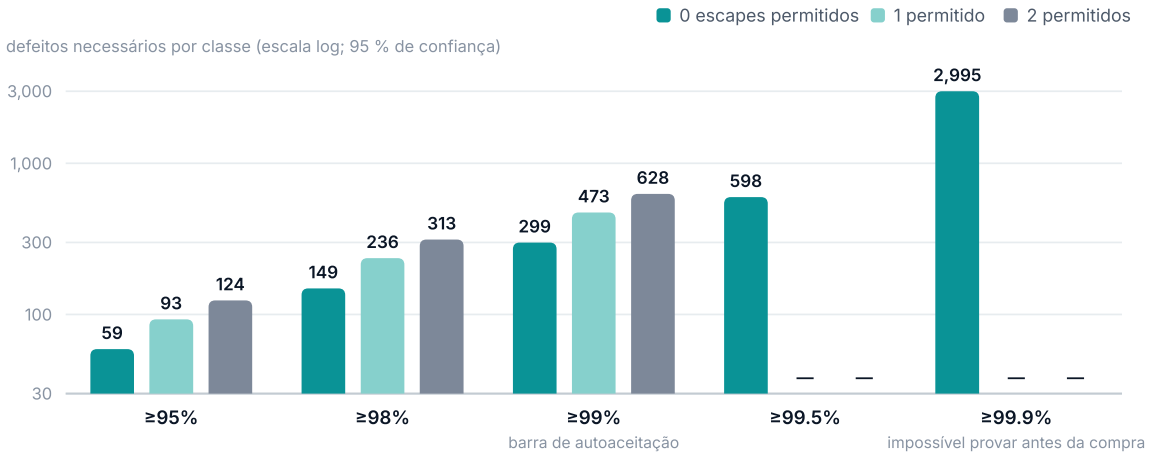


Bancos de defeitos industriais (MMAD, MVTec AD 2, VisA, Real-IAD, UniPCB): zero pontuações publicadas para esses carros-c
as pontuações de detecção genéricas dobraram em uma geração, mas isso não é julgamento de defeito; teste no seu próprio conjunto de referência

Fonte: Roboflow Vision Evals; notas de versão de modelos; páginas de MMAD, MVTec AD 2 e outros bancos (em 11 nov. 2026). MMAD; SiGMA; InspectorGPT; desafio PCBA ACM MM 2026.

Os modelos de fronteira pré-etiquetam, as pessoas arbitram, um conjunto dourado faz de portão



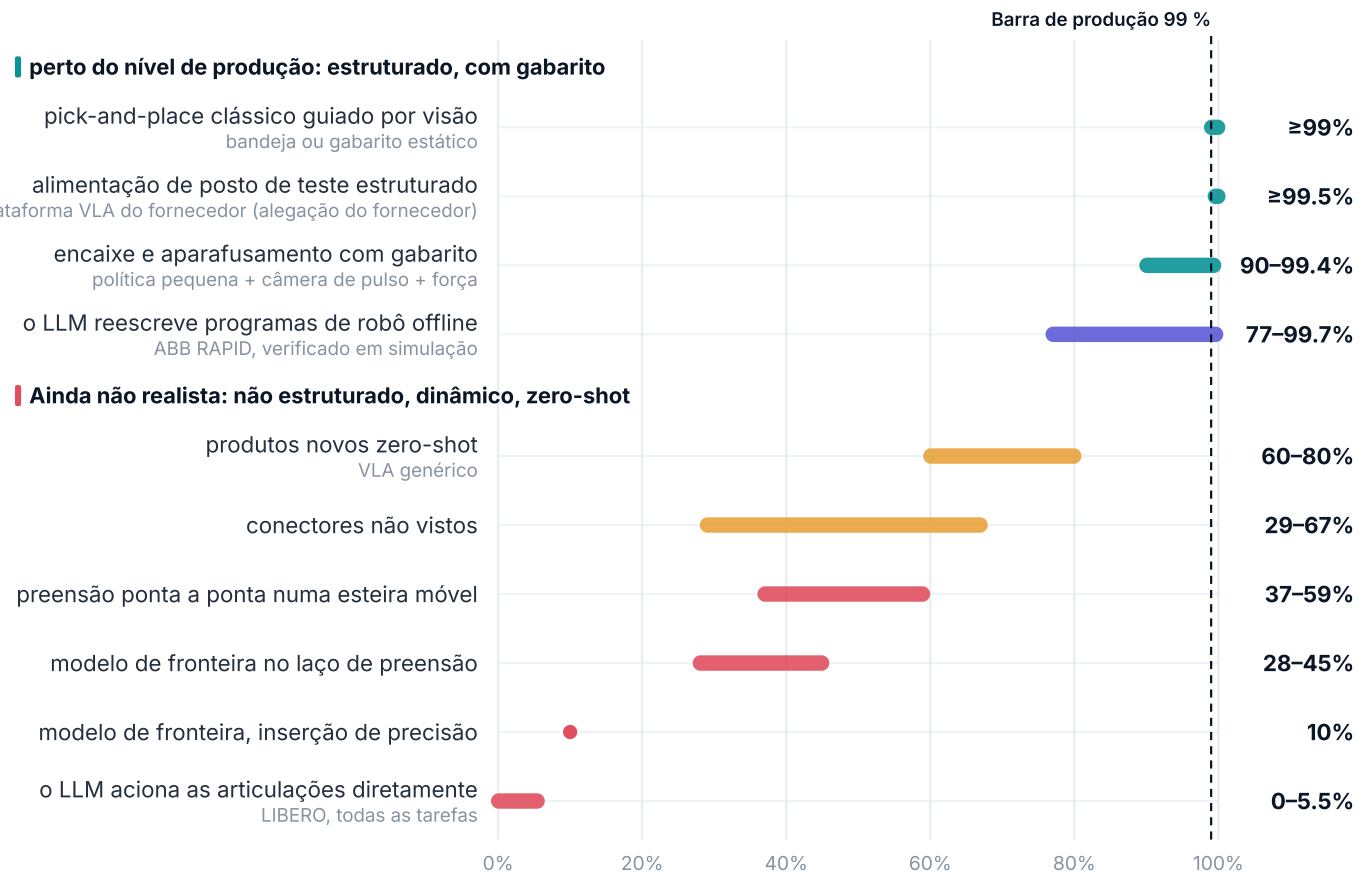


Ponto-chave Aceitar automaticamente as etiquetas do professor apenas para uma classe com zero falhas sobre ≥299 defeitos dourados (99% de sensibilidade com 95% de confiança); classes críticas como curto-circuitos, peças em falta e polaridade precisam de mais.

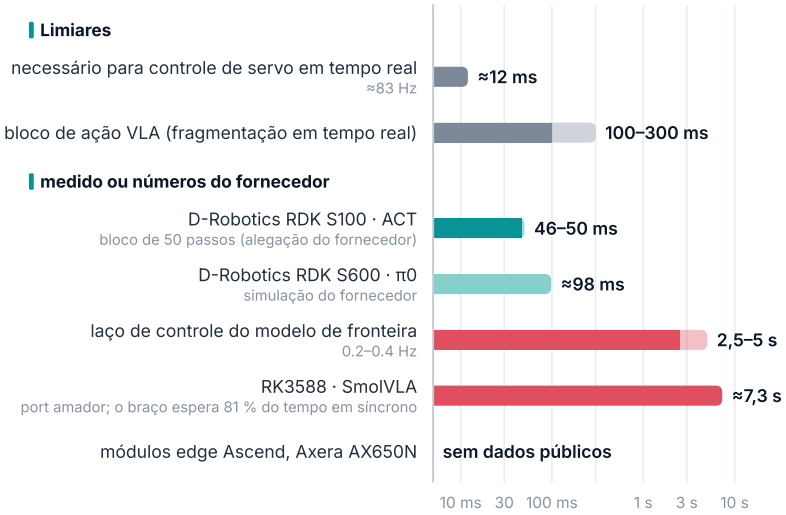
Fonte: Síntese do autor (capítulo 3); custos e tamanhos de amostra são estimativas. Intervalos de confiança binomiais exatos (95 %).

Alunos robô: perto do nível de produção só quando o posto é estruturado

taxa de sucesso da tarefa



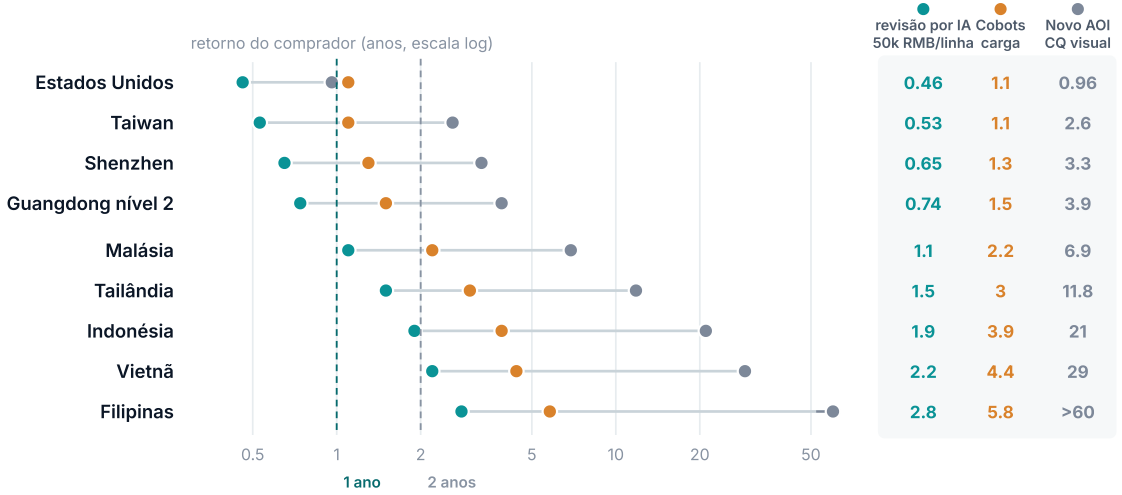
latência de inferência por bloco de ação ou ciclo de controle (escala log)



Fonte: Artigos e material de fornecedores citados no capítulo 3; alguns números são alegações do fornecedor. Documentação e fórum da D-Robotics (números do fornecedor); ports comunitários do GitHub; pesquisa da Anthropic.

A revisão com IA amortiza-se em meses; a alimentação de máquinas em cerca de um ano





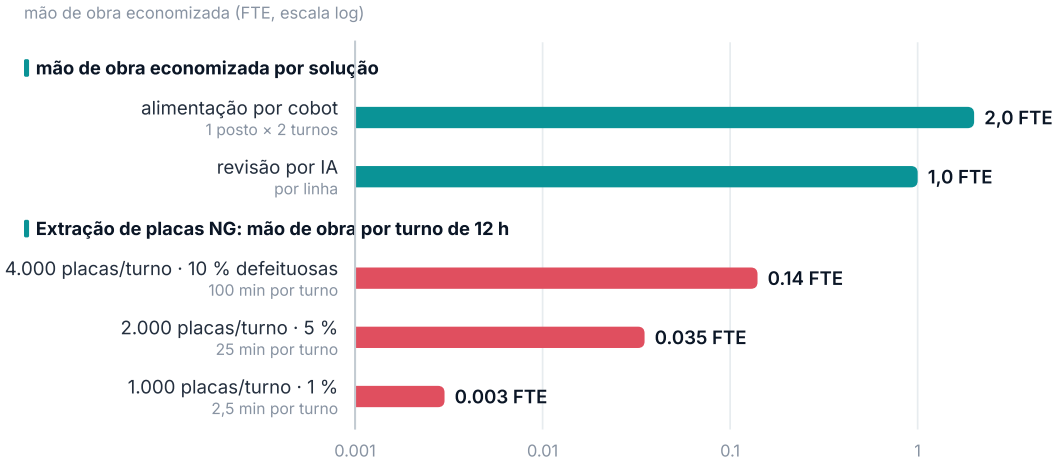
	revisão por IA 50k RMB/linha	Cobots carga	Novo AOI CQ visual
Estados Unidos	0.46	1.1	0.96
Taiwan	0.53	1.1	2.6
Shenzhen	0.65	1.3	3.3
Guangdong nível 2	0.74	1.5	3.9
Malásia	1.1	2.2	6.9
Tailândia	1.5	3	11.8
Indonésia	1.9	3.9	21
Vietnã	2.2	4.4	29
Filipinas	2.8	5.8	>60

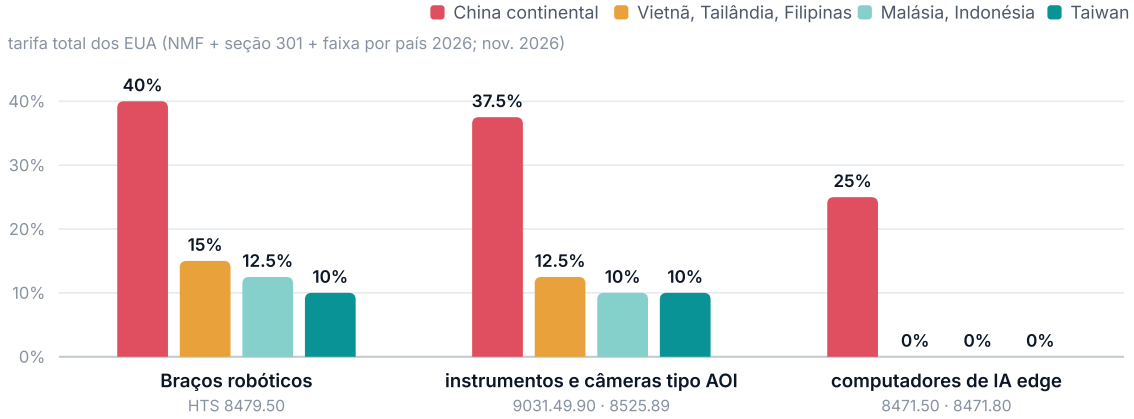
Um braço robótico só para extrair placas NG nunca se paga, em lugar nenhum
Com rendimento de primeira passagem acima de 90 %, a extração ocupa só 0,003–0,14 FTE por turno

Ponto-chave A RMB 50,000 por linha, a revisão com IA amortiza-se em 0.46 anos nos EUA, 0.53 em Taiwan, 0.65 em Shenzhen e 2.2 no Vietname; uma célula de alimentação com cobot a dois turnos em 1.1–1.5 anos [estimates].

Fonte: Orientação salarial oficial de Shenzhen, Ministério do Trabalho de Taiwan, JETRO, BLS; estimativas do autor. Modelo do autor: caixa de revisão IA de 50k RMB por linha; célula de alimentação por cobot a 200k RMB ou 140k $, dois turnos.

Um braço de retirada nunca se amortiza; o local de fabrico de um braço altera o seu direito aduaneiro nos EUA





Ponto-chave Com um rendimento à primeira passagem acima de 90%, retirar placas NG ocupa 0.003–0.14 de uma pessoa por turno, pelo que um desviador ou buffer de NG faz o trabalho; além disso, um braço fabricado no continente paga cerca de 40% de direito aduaneiro combinado nos EUA.

Fonte: Estimativas do autor. Capítulo 99 do Tarifário Harmonizado USITC; Federal Register; USTR (novembro de 2026).

Real mas pequeno: 16,000 linhas aproveitáveis, ≈RMB 320M de pagamento único, uma janela de cinco anos

linhas SMT na China (2026)

40.000 linhas

35k–46k

× 80% × 50%

ainda sem IA: 75–85 %
alcançáveis: 40–60 %

Linhas atendíveis

16.000 linhas

10,5k–23,5k

× 20k RMB

10–30k RMB pontual

Mercado pontual

320 M RMB

105–705 M RMB

÷ ≈5 anos

encolhe 8–12 %/ano

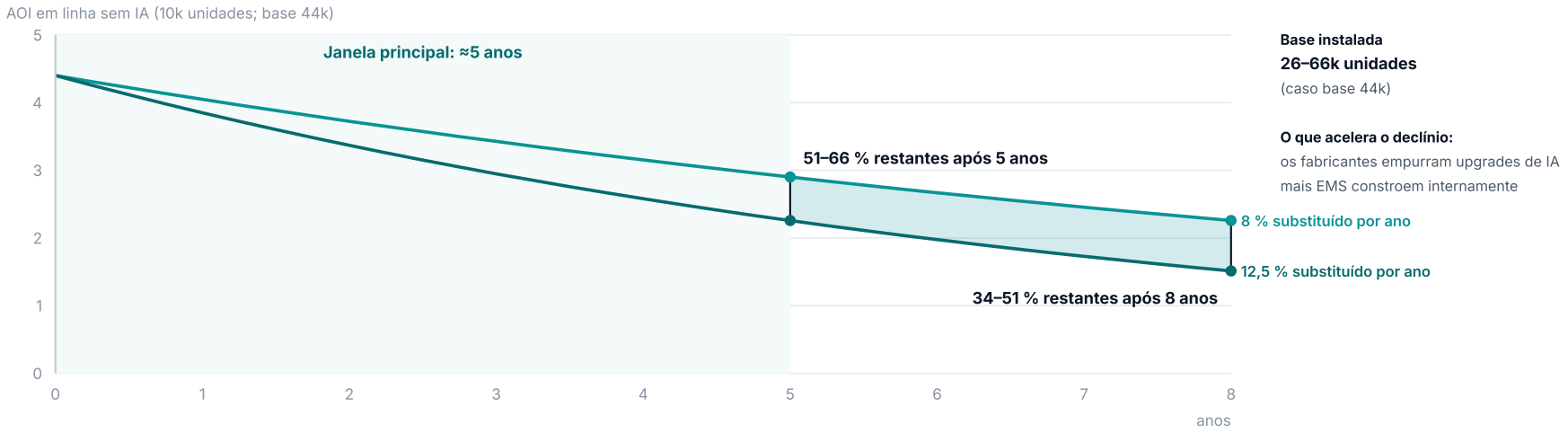
janela de 5 anos

RMB 64M/yr

21–141 M RMB

Um novo entrante ganhando 2–5 % das linhas atendíveis em 3 anos: ≈320–800 linhas, ≈2–5 M RMB por ano

Como assinatura (5k RMB por linha por ano): 1,6–4 M RMB recorrentes, só 0,3–2,3 % da previsão da IDC para o CQ industrial com IA na China



Fonte: NVR (base 2020); estimativas do autor. Estimativas do autor; base instalada e participações não verificadas de forma independente.

A cunha de revisão está cercada; a margem está no hardware transformado em produto

os grandes EMS constroem internamente

DBG (2025): cobre ≥95 % dos componentes sinalizados
escapes <20 PPM, revisão humana −85 %
essa é a barra para vender a EMS de primeira linha

IA integrada no AOI pelos fabricantes

Jutze AI 3D AOI em produção, «remove a revisão humana na inspeção»
ViTrox V510Ai; Sinic-Tek codesenvolve com Tencent
máquinas novas vêm com IA; as antigas precisam de upgrades

revisão IA de terceiros

caixa + software acoplado ao AOI existente

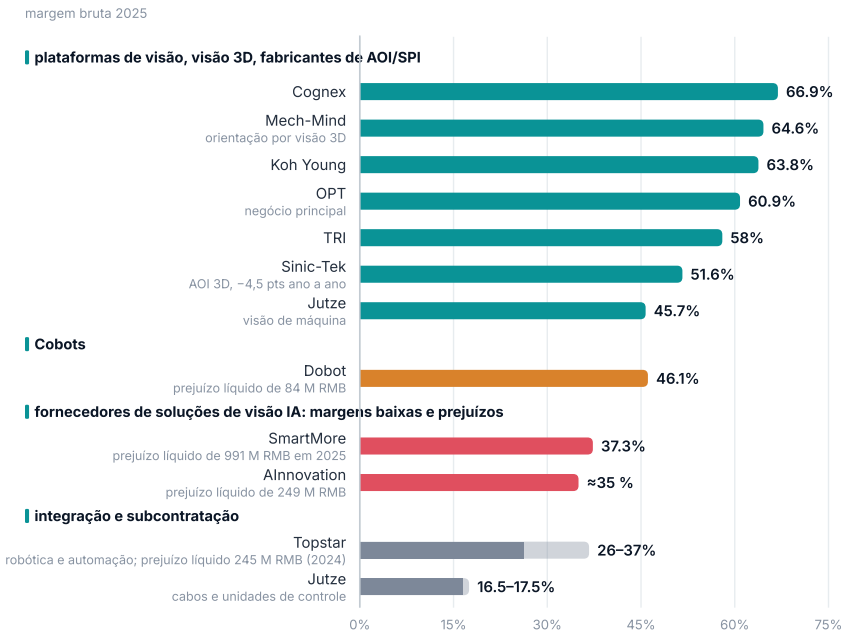
fornecedores de plataforma

Hikvision Guanlan e SmartMore IndustryGPT estão vendendo
fornecedor IA líder: receita 2025 de 1,086 bi RMB
prejuízo líquido de 991 M RMB: guerra de preços

O que resta para terceiros

EMS médios (≈10–40 linhas) com AOI antigo ou multimarca sem upgrade de IA do fabricante

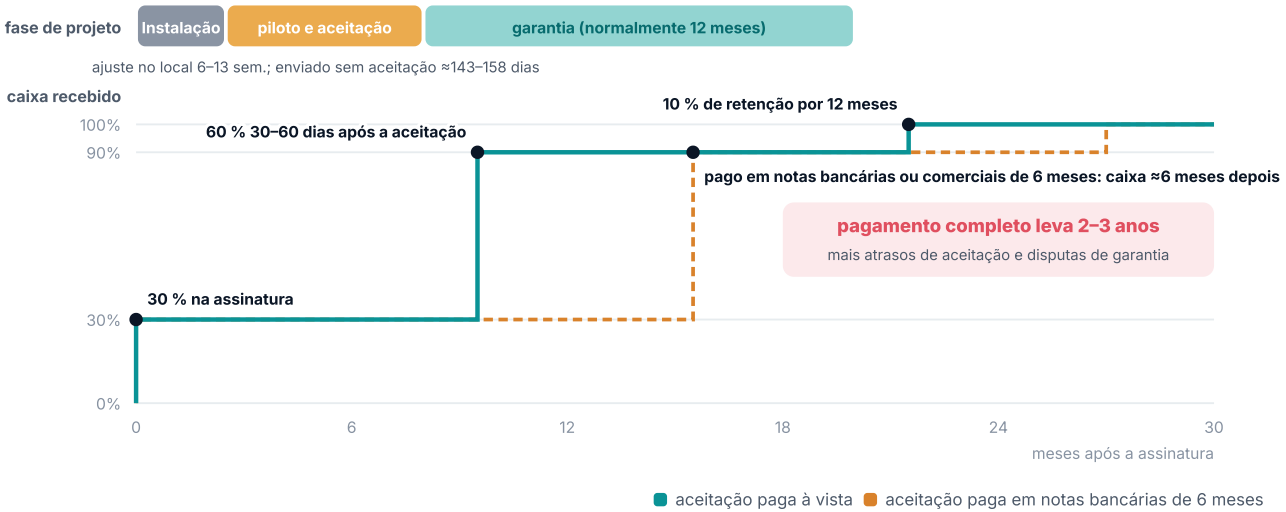
Nichos mal cobertos pelo AOI: THT e soldagem manual, verniz de proteção, cabos/conectores/FPC, aparência final e parafusos

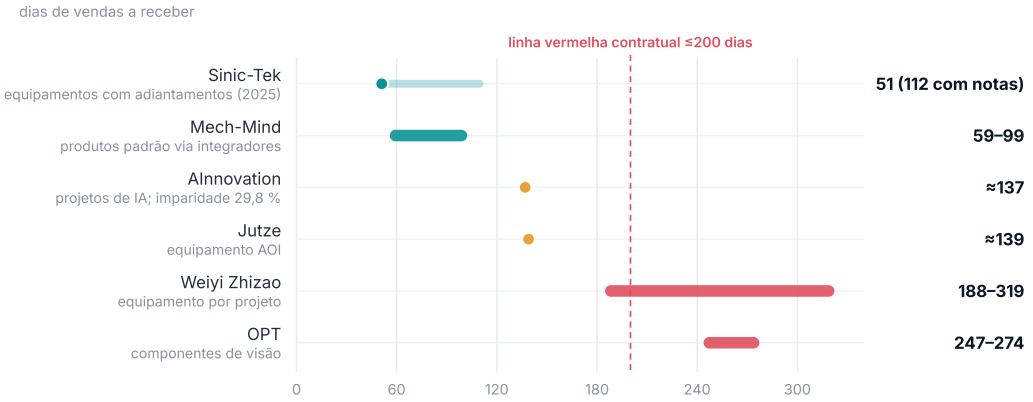


Evidência O fornecedor líder de visão com IA teve receitas em 2025 de RMB 1.086 mil milhões e um prejuízo líquido de RMB 991 milhões; a receita 3C do comparável cotado mais próximo caiu 79.9% em termos homólogos entre janeiro e maio de 2026.

Fonte: Relatórios anuais de Jutze, Sinic-Tek e DBG; material público de Hikvision e SmartMore; compilado pelo autor. Relatórios anuais ou prospectos de 2025.

Os compradores pagam 30/60/10; o pagamento integral demora dois a três anos





Ponto-chave 30% chegam com a encomenda, 60% cerca de quatro a nove meses depois e os últimos 10% cerca de 16–21 meses depois; desde um primeiro piloto numa EMS grande até ao pagamento integral podem passar dois a três anos [estimate].

Fonte: Resposta de consulta da Sinic-Tek, Mech-Mind, GKG, JT e outros documentos; esquema em condições usuais, prazo estimado. Documentos de empresa; contas a receber (incl. ativos contratuais) ÷ receita × 365.

Planear offline, agir com políticas pequenas e manter o movimento e a segurança deterministas

Camada	O que faz e até onde vai a evidência	Escala de tempo · quem a controla
Planejamento L4	**modelos de fronteira (offline)** escrever ou alterar programas de robô offline, trocas de série, explicar falhas reescreve programas ABB RAPID a 77–99,7 %, implantado após simulação laço de controle a só 0,2–0,4 Hz: fica fora do tempo real	**segundos a minutos** gigantes de modelos
Habilidades / polític L3	**ACT, políticas de difusão, VLAs pequenos** postos estruturados com gabarito 90–99,4 % chips edge: Jetson Thor, D-Robotics RDK, Ascend, Rockchip	**blocos de ação 100–300 ms** fabricantes de chips · fabricantes de robôs
controle de movime L2	**Controlador de robô** trajetórias, servo, controle de força, detecção de colisão código determinístico; as políticas aprendidas só definem metas	**milissegundos (tempo real exige >83 Hz)** fabricantes de robôs
Segurança L1	**PLC de segurança, scanners, limites de força e velocidade** certificado conforme ISO 10218-1/-2:2025 e regras locais Sem ML em funções de segurança: avaliação por organismo notificado da UE a partir de 20 jan. 2027	**tempo real estrito** robô certificado + integrador

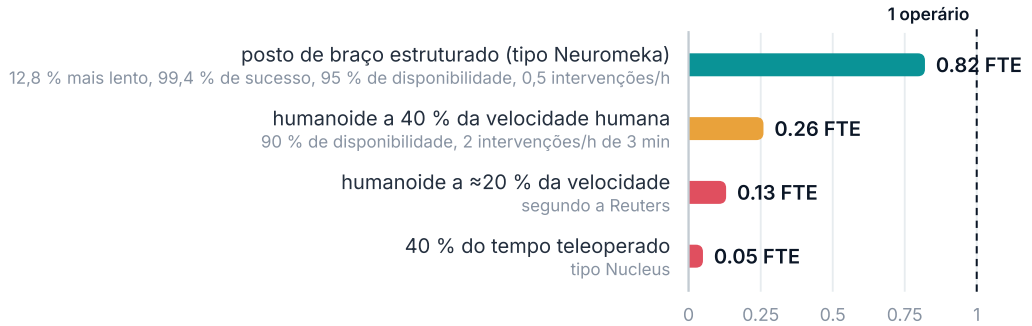
As pequenas empresas ganham na integração e no hardware de detecção: colocar percepção aprendida e políticas pequenas em células robóticas certificadas, um posto por vez

	taxa de sucesso	Takt	Intervenções / teleoperação	Disponibilidade / falhas	Volume
Figure 02 @ BMW	só objetivo >99 %/turno	só objetivo 84 s	só objetivo 0/turno	n/d	1.250 h
Agility Digit @ GXO	n/d	n/d	n/d	n/d	100.000 contentores
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 (fábrica não nomeada)	n/d	n/d	≈40 % teleoperado*	n/d	≈2 h de vídeo
AgiBot G2 @ Longcheer	99.99%	18–20 s por etapa	n/d	uma falha, 2 min	64+ h
célula híbrida Neuromeka	322/324	12,8 % mais lento	n/d	n/d	324 ciclos
π0.6 @ site do cliente	—	—	96,4 % autônomo	n/d	1 turno

* método não divulgado

Divulgado (sobretudo alegações do fornecedor) · só metas · n/d

mão de obra líquida substituída por robô e turno



Fonte: síntese do autor (capítulos 3 e 6) e estimativas; ISO 10218:2025; Regulamento de Máquinas da UE; material público da BMW, Figure, Agility, Mercedes-Benz e outros.

Construir de forma restrita, provar sobre as próprias imagens da fábrica e depois escalar

Construir	Revisão com IA e classificação de defeitos sobre imagens de AOI para EMS de média dimensão com AOIs antigos ou de marcas mistas; inspeção que o AOI cobre mal; células de cobot onde cada peça é manipulada
Não construir	Uma plataforma geral de IA mais braço; um LLM como inspetor principal; um braço de retirada numa linha SMT; humanoides e robôs como serviço; modelos de fronteira em malhas de controlo ou funções de segurança

Piloto

1. **Semanas 0–2:** acordo de dados, classes de defeitos, linha de base
2. **Semanas 1–4:** conjunto dourado, etiquetado às cegas e arbitrado
3. **Semanas 2–4+:** modo sombra, comparado por classe
4. **Semanas 4–8:** execução assistida apenas nas classes aprovadas
5. **Depois:** aceitação por escrito, escalar linha a linha e revalidar a cada alteração

Segurança Nada de aprendizagem automática dentro das funções de segurança: a UE exige avaliação por organismo notificado a partir de 20 de janeiro de 2027, e as vias de certificação de IA não ficarão definidas antes de 2027–2029.

Fonte: capítulos 6–7 do relatório e sumário executivo; ISO, IEC, UL, EUR-Lex, SAC e bases de dados regulatórias nacionais.

SOBRE O AUTOR

Tan Shuai

Tecnólogo e fundador, com mais de vinte anos entre chips, hardware e software de sistema, e 20 patentes concedidas. O seu trabalho levou-o por Silicon Valley, Nova Iorque, Chicago, Taiwan, a Grande Área da Baía (Shenzhen e Hong Kong), Vietname e Malásia, com um longo foco na indústria avançada e nas cadeias de abastecimento. Hoje trabalha como investigador e programador de IA, concentrado em como a IA é usada no chão de fábrica.

O relatório completo (todas as figuras, tabelas e fontes) e o artigo estão em tanshuai.com. Acolhe conversas com pessoas que trabalham em IA na indústria: i@tanshuai.com

