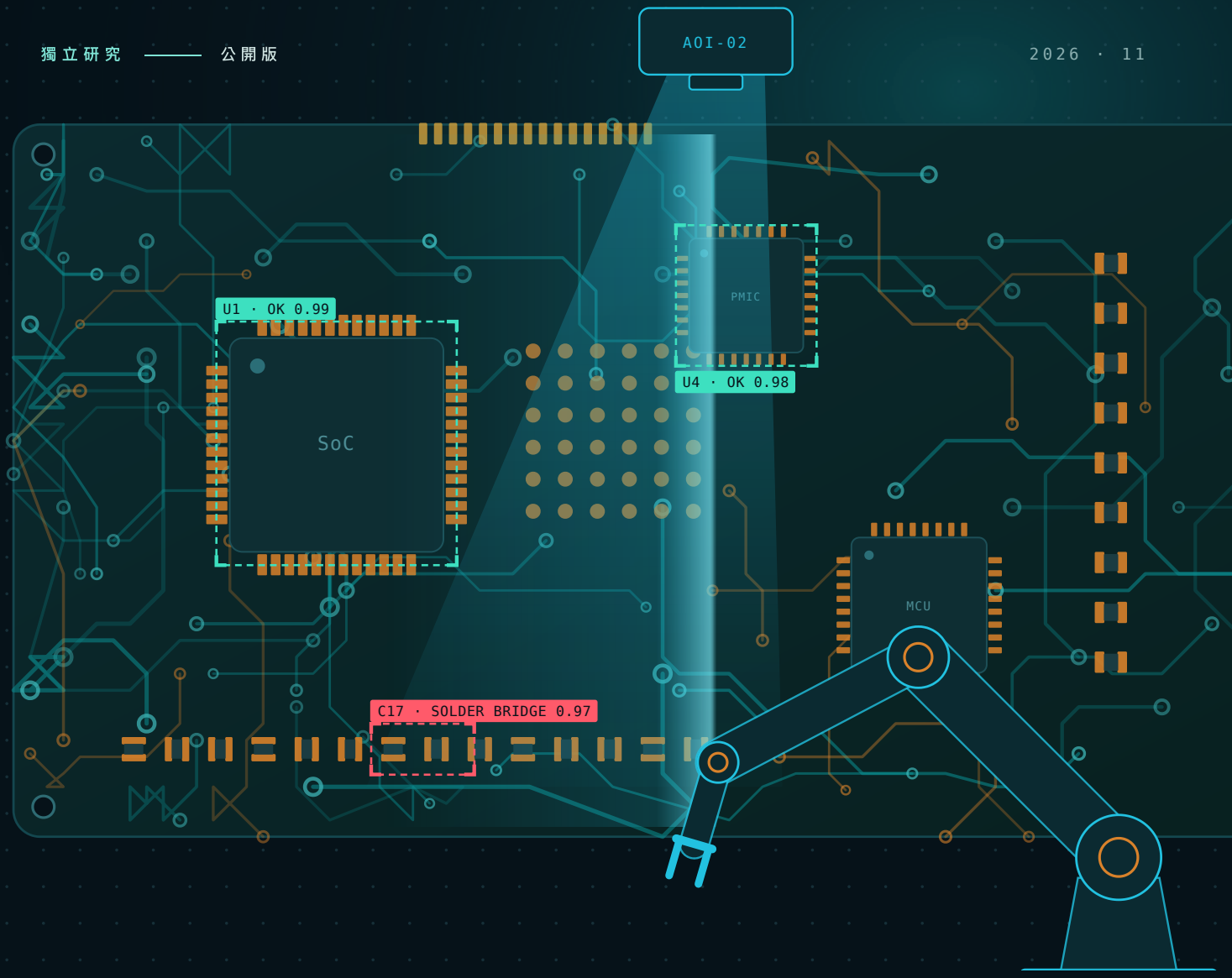


電子製造 · AI 檢測 × 機器人

電子工廠的 AI 視覺檢測與機械手臂

2026 年評估

主流配置 · 區域可行性 · 本地部署 · 老師—學生模型
投資報酬率 · 競爭 · 只有試點能回答的問題

關鍵發現 **AI 複判值得做，但需收窄範圍。具身機械手臂可以等。**

Tan Shuai
tanshuai.comi@tanshuai.com

資料截至 2026 年 11 月 11 日
中國大陸 · 台灣 · 東南亞 · 美國

目錄

目錄



◆ **執行摘要** 6

01 **今天的產線** 7
電子工廠今天跑什麼，以及技術上可行的範圍

02 **本地部署** 13
在工廠內跑 AI：潔淨室、不連網、硬體與量化

03 **老師與學生** 20
老師與學生：前沿模型可預標註檢測資料；機械手臂真正的老師仍是人

04 **成本與回本** 28
成本與回本：AI 複判數月回本，上下料以年計，揀出機械手臂永不回本

05 **市場與競爭** 34
市場、競爭與買方如何付款

06 **具身智慧** 41
具身智慧：先改變機器人怎麼被教，之後才改變它們怎麼動

07 **待解問題** 46
公開紀錄無法回答的問題，以及試點如何釐清

◆ **關於作者** 48
Tan Shuai

文中的可信度標籤

- **verified** 兩個或以上獨立來源，或兩份獨立的一手文件，意見一致
- **single source** 僅單一來源，包括一手文件
- **vendor claim** 由產品或服務的銷售方發布
- **estimate** 作者根據所引資料所做的計算或推論

如何閱讀本報告

關鍵數字都帶有可信度標籤。**verified** 表示兩個或以上獨立來源，或兩份獨立的一手文件，意見一致；**single source** 表示僅單一來源（包括一手文件）；**vendor claim** 表示由產品或服務的銷售方發布，即便被廣泛轉述；**estimate** 表示我根據所引資料所做的計算或推論，其輸入均有出處。「一手」描述的是來源類型（年報、招股書、法規、標準、官方統計）；它不改變可信度等級。來源衝突時，數字並列呈現、不取平均。匯率換算使用中國人民銀行 2026 年 11 月 9 日中間價，1 美元 = 人民幣 6.7188（Xinhua），台灣則用 SMBC 同日定價，1 美元 = 新台幣 32.12（SMBC）。中國大陸的人力與補貼資料使用電子製造最集中的珠江三角洲（深圳官方工資資料），並以長江三角洲作對照。業界訪談所得的說法，一律視為待查證的說法，絕不當作事實。公開紀錄無法回答的問題，都收錄在第 7 章。

圖 表

圖表

01	AI 今日在工廠現場適合做什麼	6
02	一條量產 SMT 線：人坐在 AOI 後面	7
03	約 97–99% 的 AOI 警報是誤報	8
04	通用相機不是 AOI	10
05	一個不連網的部署	13
06	潔淨與靜電分區決定算力能放哪裡	14
07	四類用途的速度門檻	15
08	本地算力：記憶體、價格與中國大陸可合法購置的範圍	16
09	一台大型工作站並不等於共用的助理	17
10	量化：8-bit 安全，4-bit 看配方，3-bit 及以下出局	18
11	老師候選：通用分數一代翻倍	21
12	「有沒有缺陷？」對通用模型仍然很難	22
13	兩條老師—學生流水線	23
14	金標準集要多大	24
15	端側策略延遲	25
16	機器人學生：只有工位結構化時才接近產線級	26
17	一名檢測或搬運工人一年的成本	29
18	價格階梯	30
19	各地區回本期	31
20	為什麼揀出不良板不回本	32
21	機器人、AOI 與邊緣電腦的美國關稅	33
22	市場規模相差十倍以上	35
23	自下而上：4 萬條線 → 1.6 萬可服務 → ≈人民幣 3.2 億一次性	36
24	改造窗口約五年	36
25	複判楔形市場三面被合圍	37
26	毛利在產品化硬體	38
27	買方如何付款	39
28	應收週轉天數	40
29	具身智慧堆疊	41
30	廠內機器人部署實際上揭露了什麼	42
31	每台機器人淨替代的人力	43
32	標準與法規時間線	44

執行摘要

關鍵發現

結論：對電子工廠而言，2026 年成熟、且廠方願意付錢的，是 AOI 影像上的 AI 複判與利基檢測，加上程式化的機器人上下料，而不是「大模型看板子 + 具身機器人揀板子」。

0.46—0.65 年
在美國、台灣與深圳，每線人民幣 5 萬元的 AI 複判回本期

≥95% · <20 PPM
頭部 EMS 自建 AI 複判的覆蓋率與漏失：賣給一線 EMS 的門檻

100—700 人民幣百萬
中國大陸可服務的第三方改裝市場，一次性，約五年窗口

0 分
2026 年 11 月旗艦已發表的工業缺陷基準分數：是老師，不是檢測員

1.1—1.5 年
在美國、台灣與深圳，兩班制協作臂上下料單元的回本期

0.003—0.14 FTE
揀出機械手臂在 SMT 線每班節省的人力：永不回本

檢測
ROI

把監督式深度學習應用在現有 AOI 已經產出的影像上，可將誤報或複判工時削減 60–90%（廠商說法）。DBG 這家大型 EMS 公司在其年報中披露，其自建系統覆蓋超過 95% 的報錯元件、漏失低於 20 PPM、人工複判減少逾 85%；這是第三方產品要賣進一線 EMS 必須跨過的門檻。以每線人民幣 5 萬元計，AI 複判在中國大陸、台灣與美國都能在一年內回本（深圳 0.65 年、台灣 0.53、美國 0.46），協作臂上下料單元 1.1–1.5 年，東南亞 2–6 年；而只為了揀出不良板加在 SMT 線上的機械手臂，在任何地方都永不回本，因為直通率 90% 以上時，揀出每班只佔 0.003–0.14 個人 estimate 。

模型
機器人

多模態大模型仍然當不了主檢測器。截至 2026 年 11 月 11 日，旗艦模型（Claude Fable 5.2、GPT-6.x、Gemini 4、Qwen3.8 等）沒有一個在任何工業缺陷或 PCB 基準上有公開分數，而在 ACM MM 2026 PCBA 賽題中，以 6,600 張真實產線影像訓練的前五名系統，在「有沒有缺陷」上僅得 77–82 分。前沿模型可以做預標註與分診。機械手臂方面，真正的老師仍是人類示範、真機強化學習與模擬：學生在有治具的結構化工位可達 90–99.4%，但在移動傳送帶、未見過的零件與新產品上僅 30–80%；一個前沿代理完成了 108 個準靜態真機任務中的 104 個，每個耗時 280 秒。

本地部署
部署

在潔淨、受控的工廠做不連網部署完全可行。普通 SMT 車間通常是靜電防護區，而非分級潔淨室：產線放一台無風扇的 Jetson 或國產 NPU 盒子，GPU 伺服器放機房。依 NVIDIA 自己的出口分類，Jetson Thor 與 DGX Spark 可賣給普通的中國大陸工廠、無需美國許可；DGX Station 則不行。精度預設 8-bit；廠商量化感知的 4-bit 有條件可用；3-bit 及以下出局。

市場
競爭

市場真實但有限。中國大陸約有 35,000–46,000 條 SMT 線，以及約 26,000–66,000 台未啟用 AI 的線上 AOI；第三方改裝的可服務市場約為一次性人民幣 1–7 億元，窗口約五年，而 AOI 原廠已開始把複判內建到新機 estimate 。競爭已經很激烈：複判楔形市場被 AOI 原廠、EMS 自建團隊與平台型方案商三面合圍，毛利落在產品化硬體上。

問題	判斷	關鍵數字（來源與可信度見各章）
① **主流電子廠今天跑什麼?**	一條量產 SMT 線是「3D SPI + 迴流焊前後 AOI + 抽樣 X 光 + ICT/FCT + 維修與目檢」，每台 AOI 後面都有人複判。AI 以三條路徑進入複判：AOI 原廠內建、頂級 EMS 自建、第三方加裝。組裝與搬運使用機械手臂、SCARA 與 AMR，具學習型感知與程式化運動；具身智慧仍在試點	節拍 ≈13–22 秒/板；一家 ODM 為 18 條線編列人民幣 3,060 萬元的 AOI 預算，每線約 3–5 台 AOI；約 97–99% 的 AOI 警報是誤報
② **技術能走多遠，錢能否買到零人力，各地區能用什麼?**	檢測以「專用檢測器 + AI 複判 + 少數人」達到工業級；就算錢無限也達不到零人力。中國大陸只用國產模型或本地開源權重；台灣的民間工廠幾乎什麼都能用，但晶圓廠與關鍵基礎設施實際上排除大陸品牌；東南亞限制最少；美國有全堆疊，外加關稅與認證	AI 複判削減誤報或複判工時 60–90% vendor claims；4K 相機看整板約 78 μm/像素，AOI 為 8–20 μm；大陸製手臂付約 40% 美國合併關稅，AOI 類儀器約 37.5%
③ **能否在潔淨室與不連網工廠本地部署，用什麼硬體，夠快夠準嗎?**	可以；限制在於流程，而非算力。產線放 ≤100 W 無風扇盒子，GPU 伺服器放機房；工位用 Thor，工廠問答用 Mac Studio 或國產一體機；中國大陸不用 DGX Station。模型大小與任務匹配時速度沒問題；保持 8-bit，有條件用廠商量化感知 4-bit，絕不用 3-bit	Thor 上 8B 視覺模型每張裁剪圖 ≈0.5 秒 estimate；Thor 開發套件中國含稅人民幣 59,104；DGX Station 12.3–14 萬美元，743B 模型供 32 名使用者合計 139 tokens/s
④ **最新旗艦能否當老師，跑幾十上百輪，蒸餾進本地學生，包括彈性製造與機械手臂?**	檢測：預標註與分診；僅對金標準集上「≥299 個缺陷、零漏檢」的類別自動採納老師標籤。機器人：老師仍是人類示範、真機強化學習與模擬，學生只在結構化工位接近量產級	11 月旗艦沒有任何工業缺陷分數；學生在訓練過的連接器上 97.8%，未見過的 29–67%，移動傳送帶 37–59%；前沿模型直接驅動關節 0–5.5%
⑤ **成本多少，各地區 ROI 如何?**	AI 複判在中國大陸、台灣與美國一年內回本；協作臂上下料 1.1–1.5 年；東南亞 2–6 年；SMT 線上的揀出機械手臂永不回本	每線人民幣 5 萬元複判盒：深圳 0.65 年、台灣 0.53、美國 0.46、越南 2.2 estimate；可接受的第三方加裝價格約每線一次性人民幣 1–3 萬元，或每年人民幣 0.5–1 萬元
⑥ **競爭者是否已夠強，現在開始還是等?**	需求尚未飽和，但供給擁擠；OEM AI 把改造窗口壓到約五年。現在以窄產品透過通路夥伴切入，不與平台型方案商正面交鋒；具身智慧等 2027–2028 年量產資料出來再押注	中國大陸約 2.6–6.6 萬台未啟用 AI 的線上 AOI，每年縮減 8–12% estimate；頭部 AI 視覺方案商 2025 年營收人民幣 10.86 億元、淨損人民幣 9.91 億元；最接近的上市可比公司 3C 營收 2026 年 1–5 月年減 79.9%

FIG 01 **AI 今日在工廠現場適合做什麼*AI 複判 AOI 誤報與協作臂上下料已成熟且回本快；專用的揀出機械手臂與作為主檢測器的大模型不值得做。***



來源： 回本來自作者模型（第 4 章）；成熟度軸為作者判斷。

要做什麼、不做什麼

	建議
要做	在 AOI 影像上做 AI 複判與缺陷分類，對象是配備舊款或多品牌 AOI、原廠不提供 AI 升級的中型 EMS；SMT AOI 覆蓋薄弱的檢測（通孔與手焊、三防漆、線纜、連接器與 FPC、整機外觀與螺絲）；每個零件都要搬運的協作臂上下料單元（測試治具、托盤、包裝），檢測作為加值項；在真實產線上建立金標準集與影子模式，作為一切聲明的依據
不做	通用的「AI + 機械手臂」平台；以多模態大語言模型當主檢測器；只為揀出 NG 板而加在 SMT 線上的機械手臂（NG 分流器或暫存機即可）；人形機器人與機器人即服務；控制環或安全功能中的前沿模型；工位堆一堆 GPU 記憶體跑「全尺寸」模型；直銷給晶圓廠、台灣關鍵基礎設施或美國國防電子廠
與誰合作	AOI 原廠與經銷商（通路與影像介面；經銷商占 Sinic-Tek 2025 年營收的 46.68%）；協作臂原廠（認證手臂與安全功能）；10–40 條線的 EMS 公司（滾動擴展與訂金）；泰國等地的本地整合商（滿足 BOI 本地連結規定）

01 今天的產線

電子工廠今天跑什麼，以及技術上可行的範圍

標準量產線：數台 AOI、抽樣 X 光、集中複判，每板 13–22 秒

上市 ODM 的一手申報文件顯示 2025–2026 年量產線的真實樣貌。Longcheer 的香港招股書（資料截至 2025 年 9 月 30 日）將其 SMT 品質檢驗列為「X 光檢測、錫膏檢測（SPI）、自動光學檢測（AOI）、PCBA 功能測試與目檢」；其製程依序為迴流焊前 AOI、迴流焊後 AOI、抽樣 X 光、NG 板送 SMT 維修站，再做最終目檢。其惠州、南昌與越南廠各有 34、24 與 8 條 SMT 線，以及 67、60 與 13 條最終組裝線（Longcheer H-share prospectus） single source, primary filing 。華勤 2023 年的筆電廠規劃 18 條 SMT 線，AOI 預算人民幣 3,060 萬元，每線約人民幣 170 萬元，以國產 3D AOI 價格計算，即每線三到五台 AOI，用於雙面板的迴流焊前後檢測（華勤 A-share prospectus） primary filing + estimate 。因此一條量產線通常有兩到四個複判點，而非一個。

這些機器早已高度自動化：華勤 稱其 SMT 線自動化率達 96%（華勤 H-share prospectus） single source 。剩下的人分布在複判、維修、目檢、測試與最終組裝，而沒有任何公司公布每線有多少複判員（第 7 章）。

三份獨立估計一致認同**每板 13–22 秒**的節拍（一份 2026 年河北五條線的環評文件（Hebei SMT project EIA）、華勤 的筆電線與 iFactory 案例），雙面板則每面 7–11 秒 estimate 。高混合度產線的難點不在產量，而在新產品導入與換型，而每一次換型對 AI 模型都是又一次冷啟動。

軟體方面，AOI 原廠自己的複判站早已集中化：Sinic-Tek 的維修站軟體可讓「一台主機控制八條 SMT 線的設備」（Sinic-Tek prospectus） single source, primary, 2022 ，而 Jabil 在 2026 年導入 ViTrox V-ONE 遠端 AOI 控制後，一名工程師可遠端管理 5–8 台 AOI，相較於本地的 1–2 台（AtlasPCB） vendor claim 。第三方 AI 複判產品正是透過 AOI 的影像與結果匯出（IPC-CFX、IPC-HERMES-9852 或原廠自有格式）插進這個中央站。AOI 本身也在快速加入 AI：Yamaha 稱其 AI 判定每次約一秒，30 張 NG 影像後即可協助檢測員，1,000 張後品質「非常高」（DirectIndustry/Yamaha） vendor claim 。

FIG 02 **一條量產 SMT 線：*人坐在 AOI 後面人力集中在自動光學檢測之後的複判座。那正是 AI 進入的地方。***



來源： 公開申報文件與廠商資料（華勤、光弘、思泰克等）；由作者彙整。

階段	主流產品與技術	代表性供應商	AI 進展到哪
印刷後	3D SPI（結構光、疊紋、部分雷射三角量測）	Koh Young、Sinic-Tek、TRI	OEM AI 錫膏判定已量產；Jutze 稱部分型號「全 AI」

階段	主流產品與技術	代表性供應商	AI 進展到哪
貼裝與迴流焊後	2D/3D AOI，多角度多色 LED 與結構光，每像素 8–20 µm，8–12 MP 相機	Koh Young、Omron、Saki、TRI、ViTrox、Mirtec、Jutze、Sinic-Tek 等	OEM AI 複判（Jutze 已量產、ViTrox V510Ai、Sinic-Tek 與 騰訊 合作）；頂級 EMS 自建（DBG）；第三方加裝
底部端子與隱藏焊點	抽樣或線上 X 光（AXI）	—	海康威視 稱其觀瀾模型對 X 光空洞的誤判率低於 1% vendor claim
複判與維修	中央複判站（一台主機、八條線）、遠端複判、維修站	AOI 原廠軟體、ViTrox V-ONE	AI 清除誤報，其餘交人工
目檢、IQC、FQC	多為人工；部分 AI 外觀機搭配機器人上下料	SmartMore 等	外觀機已量產；IQC/FQC 無公開人力資料
搬運與上下料	六軸機械手臂與協作臂、SCARA、AMR、2D/3D 視覺引導	FANUC、ABB、Yaskawa、KUKA、UR、Dobot、JAKA、Elite；Mech-Mind、Hikrobot	學習型感知加程式化運動已量產；VLA 與人形機器人仍在試點
資料與系統	MES、IPC-CFX、IPC-HERMES-9852	—	結果寫回 MES；舊設備支援程度不一

表格來源：上述申報文件與（Koh Young、Jutze 2025 annual report、Sinic-Tek 2025 annual report、ViTrox、Sohu on Hengtianyi、Xinhua on 海康威視）。

人坐在 AOI 後面，因為大多數警報是誤報

主流 SMT/PCBA 線的檢測鏈依序為：印刷後 3D SPI 檢查缺錫、錫量不足或過多、橋接與偏移；貼裝後（選配）與迴流焊後 AOI 檢查缺件、偏移、旋轉、反貼或立碑零件與錫量；有 BGA 或其他底部端子零件時加 X 光；然後是線上電路測試或飛針測試、功能測試與人工目檢。每台 AOI 後面都有一個複判座與維修站（Sinic-Tek prospectus、JLCPCB） verified 。迴流焊後 AOI 最常見、價值最高，但它只看形狀與面積；BGA 底下的焊點、空洞與潤濕需要 X 光或電性測試（JLCPCB）。

人力集中在複判，因為大多數警報是誤報。一家英國 EMS 在 917 個料號中記錄了**54,019 次誤報對 544 個真缺陷，約 99% 的警報**（arXiv 2505.03576） single source, preprint ；在 Siemens 的 Rastatt 廠，**AOI 判退的板子有 97.5% 其實是好的**（Siemens case study） vendor claim 。一條繁忙的產線每班約 448 次誤報，每次 42–55 秒，即每班 5.2–6.8 小時，約等於一名全職複判員（iFactory） vendor claim; hours are an estimate 。

FIG 03 **約 97–99% 的 AOI 警報是誤報 *今天由人逐一確認每一次警報；AI 複判先清除誤報，其餘交給人。***



來源： 西門子案例研究、iFactory（廠商說法）、光弘 2025 年報；複判工時由作者估計。

由「每班複判工時 = 板數 × 每板誤報數 × 每次秒數」推得，一名複判員約可涵蓋 0.7–1 條規則式 AOI 的量產線。一旦 AI 清除 60–90% 的誤報，單就判定時間而言一人可涵蓋 2–14 條線，但走動、NG 暫存處理與維修交接把本地涵蓋率限制在相鄰兩到四條線；只有集中或遠端複判才接近理論上限 estimate; seconds per call come from one source 。由於主流 AOI 每秒掃描 60–

80 cm²（Yunheng 2026 AOI guide），一塊 20×15 cm 的板面要 4–5 秒；AI 只需在一個節拍內處理少數標記區域，無需重新掃描整板。

已在量產中驗證的做法：在 AOI 自己的影像上做監督式分類

經過驗證的模式，是在工廠現有 AOI 的影像上做監督式深度學習分類。Fujitsu 在台灣 ADLINK 的 AOI 上加裝的模組，把誤報削減**逾 90%**（Fujitsu case）；OKI 自 2026 年 7 月起在其自有 EMS 業務用 AI 把 AOI 後的人工確認削減約 **80%**（Circuits Assembly）；Siemens 報告人工分析減少 60%、首年節省 160,287 歐元、八個月回本（Siemens case study）均為 vendor claims。這些案例有兩個共同點：幾乎沒人公布漏失率，且都保留人員處理剩餘警報。唯一給出漏失數字（0.8% 降到 0.3%，對比合約 0.5%）的案例，其內部其實自相矛盾（iFactory）。AI 削減的是複判工時，而非複判這個職務本身。

AI 複判已從第三方加裝走向 OEM 功能與自建系統，這改變了新進者的對標基準。Jutze 2025 年報將其「3D AOI 升級」列為量產，並稱其「用 AI 演算法在檢測端消除人工複判」（Jutze 2025 annual report）；Sinic-Tek 自 2019 年起開發「多模態 AI 輔助人工複判系統」，並與 騰訊 合作開發上下雙面 3D AOI，目標 99.5%+ 缺陷覆蓋與 98% 辨識準確率（Sinic-Tek 2025 annual report）primary filings 中的 vendor claims。大型 EMS 自建系統。DBG 2025 年報將「目檢 AI 判定」列為已完成專案，稱其「覆蓋超過 95% 的 AOI 報錯元件」、「漏失率低於 20 PPM」、「人工複判減少逾 85%」，且全程可追溯（DBG 2025 annual report）single source, primary。這是我找到的唯一同時給出覆蓋率、漏失與削減幅度的電子廠數字。**賣進一線 EMS 的第三方產品必須比照 DBG 的 95% 覆蓋率、20 PPM 漏失與 85% 削減**；在一線以下，基準是廠商宣傳的 50–85% 削減。

只以良品訓練的異常檢測速度很快（EfficientAD-S 每張影像 2.2 ms（EfficientAD, WACV 2024）），但對「邏輯型」異常（如缺件或錯件）很弱，而最佳方法在較難的 MVTec AD 2 上平均不到 60% AU-PRO（MVTec AD 2） single source。缺件與錯件最好對照 CAD 與 BOM 來查。

花再多的錢的天花板：層層機器、極少的人，但非零

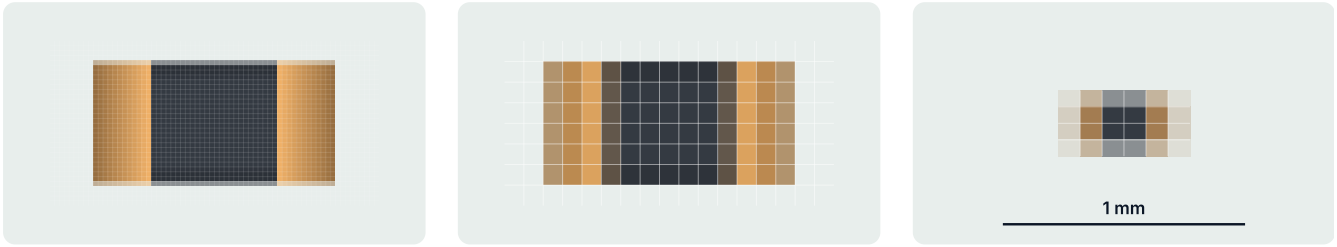
我沒找到任何電子廠完全移除人工複判且做到零漏失。IPC-A-610 Class 3（高可靠度）產品的實務仍是讓受訓檢測員裁定自動門檻無法分類的邊界案例，而一家 PCB 組裝廠稱「自動檢測可取代人工簽核」是「常見誤解」（PCBCart） single source。人類基準也需要校正：援引 2012 年 Sandia 研究，單一目檢員最多約能抓出 **80%** 的缺陷，兩人串聯約 96%（Instrumental） single source, cited by a vendor。站得住的說法是「AI 加一小圈人工複判，勝過全人工檢測」，而非「AI 取代檢測」。

預算無上限時，天花板是 3D SPI、3D AOI、X 光、AI 複判與 AI 外觀站層層堆疊，人員只留給邊界案例、新產品與換型、新缺陷類型與客戶簽核。旗艦廠報告人工檢測減少 85–92%（I-Connect007, 2021、WiAdvance） vendor claims；我對無預算天花板的估計是複判人力降到今天的 8–15% estimate。最終組裝仍屬勞力密集：iPhone 16 按鍵的自動化因缺陷率過高而放棄（9to5Mac） single source。

一般高解析度相機不是 AOI

量產 AOI 逐格成像，每像素 8–20 µm，配 8–12 MP 相機、可程式化多角度多色光源與結構光 3D（Koh Young、Saki） verified。業界很早就指出，25 µm/像素下一個 01005 零件只佔約 8×16 像素，「可能不足以」可靠判定（Circuits Assembly, 2007）。按同樣的算法 estimate，4K 相機看 300 mm 整板約 **78 µm/像素**：一個英制 0402 零件約 13×6 像素，夠判斷有無與明顯偏移、但不夠看焊點形狀；01005 約 5×3 像素，形同看不見。要對整板達到 15 µm/像素需約 3.3 億像素，這就是 AOI 要移動相機逐格拍的原因。當有人說「0402」時，先確認他指的是哪一種：公制 0402 就是英制 01005，遠超出任何廣角相機的能力。

FIG 04 ***通用相機不是 AOI：4K 相機看 300 mm 整板時每像素約 78 微米：0402 元件縮到 ≈13×6 像素，01005 直接消失。***



AOI：8–20 微米/像素
0402 約 100×50 像素

4K 相機看 300 mm 整板：約 78 微米/像素
0402 縮到只剩約 13×6 像素：僅夠判斷有無與明顯偏移

同一台相機看 01005
約 5×3 像素：等於看不見

來源： Circuits Assembly；作者針對 300 mm 整板與 4K 感測器的計算。像素網格為示意。

在已有 AOI 的產線上，AI 最好的落點是疊在 AOI 影像之上。只有沒有 AOI 的產線，或線末站，才值得架設有受控光源的封閉相機站；架在產線上的開放式高解析度相機最容易受環境光與陰影干擾（CSDN）。

組裝與搬運：學習型感知加程式化運動已在量產

IFR 2026 年報告稱「用於檢測與機器人引導的機器視覺已成熟、達工業級」，而「通用實體 AI 與從模擬和經驗中學習尚未達工業級」（IFR World Robotics 2026 executive summary） single source, primary 。電子業是最大的機器人買家（2025 年全球裝機 142,235 台，占中國裝機量的 59%（IFR、IFR China release)），但中國每萬名製造業工人僅 166 台機器人，美國為 307 台（IFR, April 2026）。量產中的模式是學習型感知加確定性運動；端到端學習運動只在結構化工位接近量產級（第 3 與第 6 章）。

各地區能用什麼

中國大陸依賴國產模型與本地權重；台灣由客戶身分決定；美國由關稅與認證決定。

項目	中國大陸	台灣（民間工廠）	東南亞	美國
Claude（任何管道）	● 不可用	● 有條件：供應商不得由中國大陸、香港或澳門的個人或實體多數持股或控制	● 同左	● 同左
GPT 與 Gemini 雲端	● 不可用	● 可用	● 可用（未逐國查核清單）	● 可用
中國雲端模型（Qwen、DeepSeek、Doubao、GLM、Kimi）	● 可用（已備案模型；提供服務需登記）	● 合法，但大陸託管很少通過客戶資安審查；公部門、國營企業與關鍵基礎設施禁用	● 合法（國際節點）	● 合法，審查嚴格
中國來源的本地開源權重（Qwen、GLM、DeepSeek）	● 首選；私有廠內部署無需模型或演算法備案（律師解讀）	● 可用（公部門受限）	● 可用	● 有條件：來源審查
非中國來源的本地開源權重（如 Gemma 4）	● 可用	● 可用	● 可用	● 可用
傳統非生成式視覺模型	● 可用（在生成式 AI 規則之外）	● 可用	● 可用	● 可用

項目	中國大陸	台灣（民間工廠）	東南亞	美國
Jetson Thor/T5000/T4000、DGX Spark	● 可用：NVIDIA 將其列為 5A992.c，普通民用工廠無需美國許可；Spark 有中國版；通路價約高於美國定價 40–60%	● 可用	● 可用	● 可用
RTX 5090、RTX PRO 6000、DGX Station	● 不可用：前兩者為 4A090.a，DGX Station 為 5A992.z；合法替代為 4090D、5090 D V2、RTX PRO 5000 72GB、RTX 6000D、L20（4A994.l）	● 可用	● 可用，但總部在中國或最終由中國所有的買家視同中國；馬來西亞出口與轉運需許可	● 可用
Mac Studio M5 Ultra 512 GB	● 可用（蘋果 中國官網）	● 可用	● 推定可用	● 可用
Huawei Ascend 與其他中國加速器	● 可用	● 高風險	● 高風險	● 實際上不可用
大陸製固定機械手臂、AOI、工業相機	● 可用	● 可進口；自 2025 年 12 月 1 日起主管機關可逕行處理大陸品牌，機關、國營企業、特定基金會與關鍵基礎設施供應者可被限制或禁止使用	● 可用	● 合法，但合併關稅手臂約 40%、AOI 類儀器與相機 37.5%、邊緣電腦 25%
外國製 AMR 與移動機械臂	● 可用	● 可用	● 可用	● 新型號需美國戰爭部有條件核准
海康威視 與 Dahua 相機	● 可用	● 民間可用、公部門禁用	● 可用	● 新 FCC 授權受阻；聯邦承包商禁用
機器人安全與認證	● GB 11291.1-2011 與 GB 11291.2-2013 強制，自 2026 年 4 月起修訂中；CR 認證自願；GB/T 39404-2026 涵蓋控制器資安	● 預設圍欄；無圍欄協作須符合 CNS 14490 或 ISO 10218 並附書面評估	● 依客戶與出口目的地，出貨前書面確認	● 電氣設備須「核准」（實務上為 NRTL 標誌或現場評估）；UL 1740 自 2025 年 8 月 11 日起要求機器人單元符合 ISO 10218-2
晶圓廠與受控產品廠	—	● TSMC 採購合約要求 SEMI E187；2 nm 晶圓廠正移除大陸設備	—	● ITAR 技術資料含「照片」；對中國採否決政策；大陸控制的供應商實際上無法服務國防電子廠

表格來源：（Anthropic supported regions、OpenAI supported countries、Gemini availability、CAC interim measures on generative AI、NVIDIA export classification data、eCFR 15 CFR 774、NVIDIA China DGX Spark、蘋果 China、Covington on new BIS rules、Greenberg Traurig, June 2026、Fulcrum、Tom's Hardware on Ascend guidance、Taiwan BOFT import list、Taiwan Cyber Security Management Act history、Taiwan review rules for products endangering national cybersecurity、Taipei Times、USITC HTS chapter 99、Latham & Watkins、IPVM、China national standards platform, GB 11291.2、Intertek UL 1740 update、eCFR 22 CFR 126.1）。

中國大陸：全堆疊齊備，但閉源前沿模型與最強的單機都進不來。 已備案的國產雲端模型、本地開源權重（Qwen 採 Apache-2.0、GLM 與 DeepSeek 採 MIT）、傳統視覺模型，以及國產 AOI、相機與協作臂都可用。NVIDIA 自己的出口分類將 Jetson AGX Thor 套件、T5000/T4000 模組與 DGX Spark 列為 5A992.c，因此普通 EMS 或 PCBA 工廠無需美國許可；而 RTX 5090 與 RTX PRO 6000（4A090.a）及 DGX Station（5A992.z，TPP 58,484）在大陸買不到（NVIDIA export classifications、eCFR） primary; the legal reading is mine and not legal advice 。RS China 的 Thor 套件含稅標價人民幣 59,104，約為美國定價的 1.4 倍（RS China）。

四家律師事務所一致認為，供工廠自用的私有廠內部署無需模型或演算法備案，而託管式多租戶服務處於灰色地帶（Covington、FPF、JunHe） verified: consistent readings, no official ruling 。GB 11291.1-2011 與 GB 11291.2-2013 仍是機器人安全的強制基線，自 2026 年 4 月 28 日起修訂中；控制器資安的 GB/T 39404-2026 於 2026 年 11 月 1 日施行（revision plan 20262542、GB/T 39404-2026） single source, primary 。在新版 GB 11291 出爐前（2028–2029 estimate），依 ISO 10218:2025 設計並保留 GB 11291 合規文件。

台灣：民間工廠幾乎什麼都能用；由客戶身分決定。 前沿 API、開源權重與大陸設備都合法（data.gov.tw）。但自修正後《資通安全管理法》於 **2025 年 12 月 1 日**施行起，主管機關可逕行處理「大陸品牌」產品，並在政府機關、國營企業與關鍵基礎設施供應者限制之（Laws & Regulations Database、review rules） verified ；關鍵基礎設施供應者清單屬機密；

TSMC 已把 SEMI E187 寫入採購合約（CNA），並傳聞正從其 2 nm 晶圓廠移除大陸設備（TrendForce） single source。民間 EMS、組裝與小型 PCBA 廠是開放的；晶圓廠、園區公用事業與政府相關客戶實際上是關閉的。大陸控制的供應商應按 E187 設計、提供軟體物料清單、支援離線更新，且絕不從大陸遠端連線。

東南亞：限制最少；問題在回本慢與中國母公司。 什麼都能用，但美國 2026 年 5 月 31 日的指引要求，向總部在中國或最終由中國所有的實體出口先進運算品項須取得許可，即便收貨人在第三國亦然（Greenberg Traurig） single source，而馬來西亞對先進晶片自有其許可規定（Fulcrum）。需求正往這裡移：TrendForce 預期泰國、越南與馬來西亞在 2026 年將接近伺服器 ODM SMT 產能的 50%（TrendForce） single source, a 2023 forecast。

美國：全堆疊齊備，代價是關稅、認證與受控資料。 依美國關稅表第 99 章，大陸製機械手臂合併關稅約 40%，AOI 類儀器與相機約 37.5%，邊緣電腦 25%；馬來西亞或印尼製約 10–12.5%，越南、泰國或菲律賓製 12.5–15%，台灣製 10%，電腦在大陸以外為 0%（USITC HTS chapter 99、Great Lakes Customs Law） primary schedule + estimate。USTR 11 月 10 日的公告把 178 項舊 301 排除延長到 2027 年 1 月 9 日，其中沒有一項涵蓋機器人、檢測儀器、相機或電腦（Federal Register 2026-22925） verified。

OSHA 要求工作場所電氣設備須「核准」（實務上為 NRTL 標誌，或每個單元約 3,000 至 20,000+ 美元的現場評估），且自 2025 年 8 月 11 日起 UL 1740 把整個機器人單元納入 ISO 10218-2（eCFR 1910.399、Intertek、Industrial Monitor Direct）。國防電子對大陸控制的供應商實際上是關閉的：ITAR 把生產照片視為技術資料，並對中國採否決政策（eCFR 22 CFR 120.33、eCFR 22 CFR 126.1）。銷美的單元應使用固定手臂、不受限相機與本地或美國雲端模型，並在東南亞做最終組裝（若那能真正改變原產地）。

任何預算下，天花板因地區而異，但哪裡都達不到零人力 estimate。中國大陸可堆疊國產 3D SPI、3D AOI、X 光、AI 複判與自建 AI 團隊，把複判人力降到 8–15%，但沒有閉源前沿模型或 DGX Station，全廠大語言模型問答得靠本地開源權重與國產算力（供多人使用的全尺寸模型至少需兩台八卡 Ascend 伺服器）。台灣與東南亞技術上可用最好的所有工具，包括前沿 API 與 DGX Station，受客戶身分、大陸品牌與中國母公司許可所限。美國沒有技術限制，代價是關稅、NRTL 認證與受控資料的人員隔離。任何地方都保留人員處理邊界案例、新產品與換型、新缺陷類型與客戶簽核。

因此，跨區域銷售的產品需要兩套模型堆疊：一套大陸堆疊（已備案國產模型或本地開源權重），一套海外堆疊（前沿 API 加一個非中國來源的本地模型）。Anthropic 自 2026 年 11 月 12 日起生效的政策以所有權與控制權判斷，因此大陸控制的供應商即使透過海外子公司、雲端平台或經銷商，也無法建立在 Claude 之上（Anthropic） single source, primary。

02 本地部署
在工廠內跑 AI：潔淨室、不連網、硬體與量化

在不連網工廠，限制在於流程，而非算力

完全離線的視覺檢測是常態；限制在於流程。中國 2024 年工信部工業控制系統安全指引要求分區、防火牆或資料二極體、受稽核的限時遠端維運、掃描過的介質與測試過的修補檔（guide text） single source；半導體客戶再加上 SEMI E187 與 E188，TSMC 在 2023 年已將其寫入採購合約（SEMI、TXOne） verified。結果經 IPC-CFX 回寫 MES，板身識別經 IPC-HERMES-9852 傳遞（Hermes standard）。常見的陷阱是假設有網路的授權機制：Roboflow 的離線模式以 30 天租約快取權重（Roboflow）。離線產品需要離線啟用、經掃描介質的簽章更新、軟體物料清單、USB 控管、本地稽核日誌與 CFX 轉接器。

FIG 05 ***一個不連網的部署線邊小盒負責判定；機房負責回答問題；故障時退回人工，資料從不出廠。***



來源： 作者綜合整理（第 2 章）。

實際上，閉源前沿模型進不了工廠機房。Gemini 僅透過 Google Distributed Cloud 提供給政府與大型企業的不連網方案（Google Cloud）；Microsoft 的 Foundry Local 只對「合格客戶」開放大型模型，且未點名任何 GPT 模型（Microsoft）；Claude 只在雲端政府區域提供，無本地部署（Claude for government）；三者皆不在中國大陸提供。廠內設備必須圍繞開源權重設計；前沿模型最多只能在總部離線為影像打標籤或解釋，且須客戶書面同意匿名影像外流。

機房硬體應視為辦公室級：Mac Studio 額定 10–35 °C，DGX Spark 額定 5–30 °C，並有持續推論下熱關機的報告（蘋果、NVIDIA forum） verified。產線放無風扇工業盒，如額定 −25 至 60 °C 的 Jetson T4000 工業電腦（AAEON） vendor claim。由於大型工廠每小時停機損失 36,000 至 230 萬美元（Siemens True Cost of Downtime 2024） vendor survey，產線上每個 AI 步驟都必須**失效即放行**：任何故障都退回現有的「AOI 加人」流程。

潔淨與靜電分區：普通 SMT 車間是靜電防護區，不是分級潔淨室

限制算力能放哪裡的是靜電與氣流，而非潔淨等級。多數 SMT 組裝區「未正式做微粒控制」，世界級工廠「可能達到或接近 ISO 8」（Circuitnet expert panel） single source。醫療 PCBA 與鏡頭模組通常落在 ISO 7，裸晶片與光學對位在局部 ISO 5–6（PCBSync） single source；潔淨室建商對 SMT 一般報 ISO 7–8（Angstrom Technology） vendor claim。

普遍的共同要求是靜電防護：ANSI/ESD S20.20-2021 要求孤立導體低於 35 V，強場絕緣體與敏感元件至少保持 30 cm（ANSI、ESDA、Desco） verified 。晶圓廠把算力放在次潔淨層或資料中心，用符合 SEMI S2/S8/F47 的液冷機櫃（Prodrive） vendor claim 。

設備能否進入潔淨區，依 ISO 14644-14 測試微粒逸散；機器人持有此類證書（Fraunhofer IPA），但我沒找到任何工業電腦、邊緣 AI 盒或 GPU 伺服器有此證書。CE、FCC 與 IEC 60068 只證明盒子能耐受其環境，不證明它能進入靜電防護區；這由客戶的靜電防護協調員把關。GPU 伺服器從不上產線：DGX Station 測試中 GPU 耗電 1,292 W（StorageReview）。

FIG 06 ***潔淨與靜電分區決定算力能放哪裡 普通 SMT 車間是靜電防護區，不是潔淨室。GPU 與 NPU 伺服器屬於機房。***



來源： ANSI/ESD S20.20、IEC 61340-5-1、ISO 14644、SEMI 標準與廠商資料；由作者彙整。

區域	典型環境	產線算力	GPU/NPU 伺服器
普通 SMT/PCBA 車間（消費、工業）	S20.20 或 IEC 61340-5-1 下的靜電防護區，通常非分級潔淨室（≈ISO 8 等效）；相對濕度 40–60%	接地金屬外殼的無風扇盒（Jetson T4000/Thor、國產 NPU 盒），靠近 AOI 或複判座；塑膠螢幕與外殼移出該區或與板子保持 ≥30 cm	機房或空調 IT 機櫃
醫療、鏡頭模組、電動車電子（ISO 7–8 潔淨室）	潔淨室加靜電防護	無風扇、IP65/66 密封、表面平滑易擦拭；索取 ISO 14644-14 資料或 IPA 證書	潔淨室外，網路接入
裸晶片、打線、光學工位（局部 ISO 5–6）	微環境	只有相機與光源在內；算力盒在外	機房
晶圓廠潔淨室（ISO 5 及更潔淨）	晶圓廠專用靜電與靜電吸附計畫，設備依 SEMI E78；資安依 SEMI E187/E188	僅工具整合、經晶圓廠認證的控制器；不加裝帶風扇的 PC	次潔淨層液冷機櫃或晶圓廠資料中心

本表是我對上述來源的綜合整理 estimate 。對產品而言，這意味著產線側無風扇盒約 100 W 以下、金屬外殼接保護地、12–24 V DC 輸入、在真實機櫃氣流下寬溫運作，並備妥 ISO 14644-14 微粒逸散報告，供 ISO 7 及更潔淨的客戶索取。任何「一台 GPU 伺服器服務四條線」的設計，都是把伺服器放機房，產線只留小盒或相機。

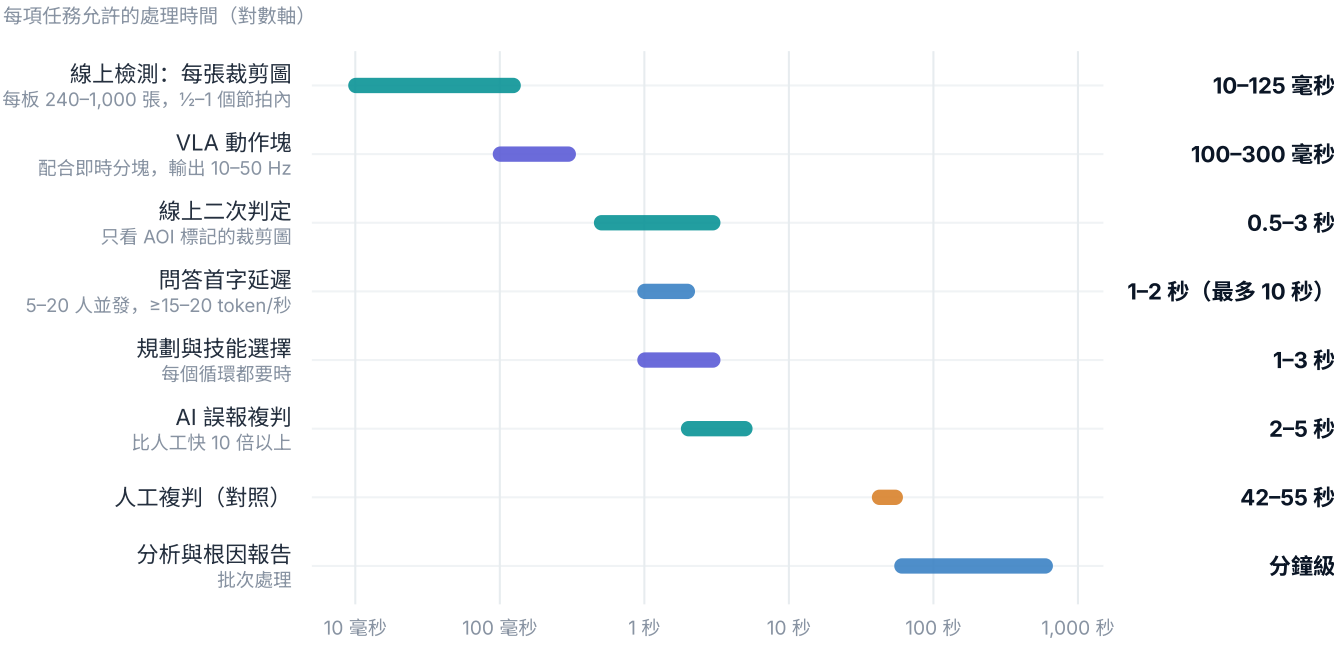
依用途的速度門檻：準確比速度更難

工廠與廠商幾乎從不公布 AI 延遲或可用率目標。以下門檻由節拍、人工複判時間、機器人控制速率與互動研究推導 estimate ：

用途	速度門檻	準確門檻	可用率
每個零件的線上檢測	½–1 個節拍內：10–30 秒節拍內 240–1,000 張裁剪圖，每秒 8–100 張	漏失不劣於今天的「AOI 加人」（合約例 <0.5%）	失效即放行，不停線；目標 ≥99.9%（無公開 SLA）
僅對 AOI 標記裁剪圖的線上二次判定	½ 節拍內處理標記（通常每板 ≤5–20 個）：每張 0.5–3 秒	同左；以產線自身影像訓練	退回人工複判
誤報的 AI 複判	每次 2–5 秒，比人工 42–55 秒快十倍以上	漏失不劣於人工複判員	降級時加人，永不停線
機器人控制（VLA）	每個動作塊 100–300 ms，配合即時分塊，輸出 10–50 Hz	≥99% 任務成功率以勝過程式化運動	安全系統獨立於模型
機器人規劃與技能選擇	每個循環需要時 1–3 秒；僅例外或換型時 ≤1 分鐘	輸出須為可驗證的程式碼或技能 ID	退回固定程式
問答（SOP、MES 查詢）	首字理想 ≤1–2 秒，最多 10 秒；每名使用者 ≥15–20 tokens/s；5–20 人並發	引用準確；長上下文保真	辦公室 IT 級（≈99–99.5%）
分析與根因報告	分鐘級即可（批次）	最高；長上下文保真	辦公室 IT 級

Physical Intelligence 量得 3B 級 VLA 推論 97 ms（含網路與前處理約 108–139 ms），發現即時分塊在最多 200 ms 額外延遲下吞吐不變，並在逾 300 ms 下完成點火柴、插網路線等精細任務（Physical Intelligence） vendor research；互動研究把維持思路的界限定在一秒、注意力在十秒（NN/g）。準確才是難點：開源 Qwen3-VL-8B 在五個工業異常資料集上的零樣本平均僅 56.8（IndusAgent） single source。再好的硬體也救不了沒在產線自身影像上調校過的視覺語言模型。

FIG 07 **四類用途的速度門檻*線上檢測以毫秒計，問答以秒計，分析以分鐘計。人每次複判要 42–55 秒。***



來源： 作者綜合整理（第 1–2 章）。

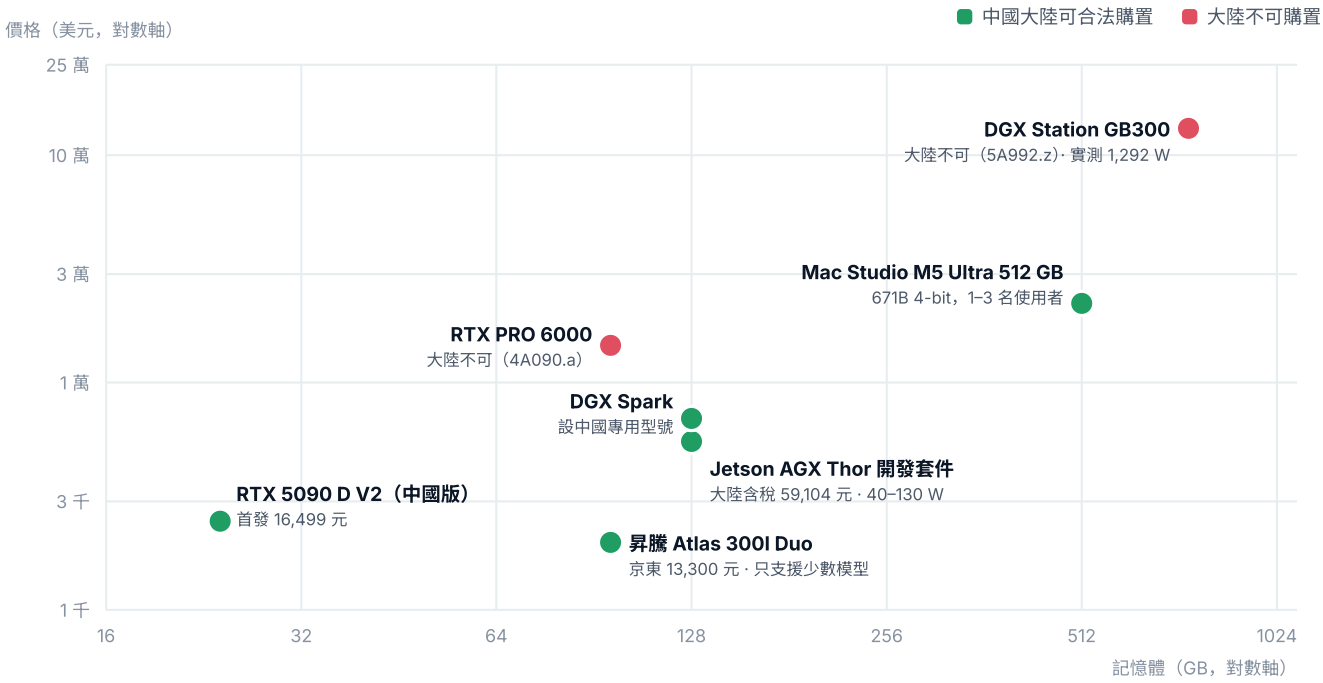
硬體：工位用 Thor，工廠問答用 Mac Studio 或國產一體機，中國大陸不用 DGX Station

設備	記憶體與頻寬	價格（2026 年 11 月）	適合用途	中國大陸
Jetson AGX Thor 開發套件 / T5000 / T4000	128 GB LPDDR5X，273 GB/s；40–130 W	上市 $3,499 → $5,499（2026 年 7 月）；T5000 $4,999，T4000 $2,999；RS China 開發套件含稅人民幣 59,104	2–8B 視覺模型複核裁剪圖；即時 3B 級 VLA	● 可（5A992.c，民用工廠免許可）

設備	記憶體與頻寬	價格（2026 年 11 月）	適合用途	中國大陸
DGX Spark	128 GB，273 GB/s	$3,999 → $4,699 → $6,950；同晶片 ASUS GX10 中國版京東人民幣 32,999（2025 年 10 月）	多人 8–35B MoE	● 可（5A992.c，中國版，授權通路）
RTX PRO 6000	96 GB GDDR7	≈$8,565 → $13,250（2026 年 6 月）→ 11 月 eBay 均價 ≈$16,005	多人 27–35B FP8；4-bit 120B 級 MoE	● 否（4A090.a）
Mac Studio M5 Ultra 512 GB	512 GB 統一記憶體，1.2 TB/s；最大持續 480 W	美國 $22,299 起；中國估約人民幣 173–186k	4-bit DeepSeek 671B 供 1–3 名使用者	● 可
DGX Station GB300	748 GB（252 GB HBM3e @7.1 TB/s + 496 GB LPDDR5X @396 GB/s）；GPU 實測 1,292 W	11 月市價 $123–140k（關稅前 ≈人民幣 830–940k）；8 月 $95–117k	裝得進 HBM 的 120–235B MoE 高並發；433–743B 模型僅少數使用者	● 否（5A992.z，TPP 58,484）
Huawei Atlas 300I Duo	96 GB LPDDR4X，408 GB/s	京東商家人民幣 13,300	低量批次複判；支援模型少	● 可（國產）
Huawei Atlas 800I A2（8 × Ascend 910B 64 GB）	8 × 64 GB HBM	未取得；完整 DeepSeek 需兩台（W8A8）、四台（BF16）	全廠多人問答與分析	● 可（國產）
RTX 5090 D V2（中國版）	24 GB	上市人民幣 16,499（2025 年 8 月），隨記憶體價格上漲	8B 視覺模型、小型 VLA	● 可

來源：（NVIDIA China Thor、NVIDIA Thor launch、CNX on price rises、NVIDIA export classification data、RS China、Implicator、IThome on ASUS GX10、Tom's Hardware RTX PRO 6000、GPU Poet、蘋果 specifications、TechRadar、IThome on M5 Ultra China pricing、Wccftech、StorageReview、CyberPowerPC、monitors.com、Tom's Hardware DGX Station、vLLM-Ascend support matrix、China Telecom Cloud、Huawei Atlas 800I A2、TweakTown）。512 GB Mac Studio 的中國價格由美國配置溢價外推，屬低可信度 estimate 。

FIG 08 **本地算力：記憶體、價格與中國大陸可合法購置的範圍*出口管制分類在基準測試之前就先決定了候選清單。***



來源： NVIDIA 出口分類、蘋果、RS China、京東與評測價格（2026 年 11 月）。

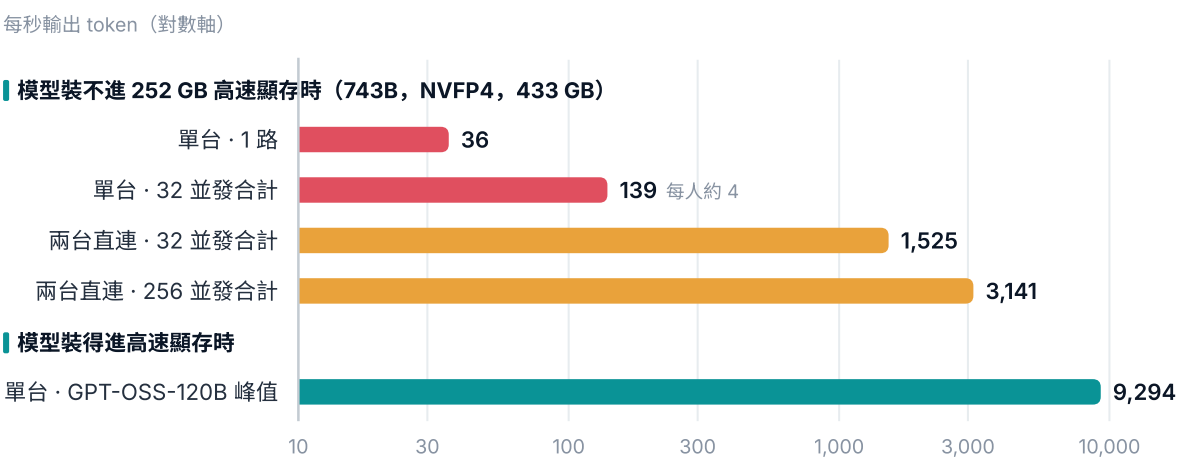
Jetson Thor 讓工位級部署可行。 NVIDIA 的 TensorRT Edge-LLM 測試顯示，Qwen3-VL-8B（NVFP4）在 Thor 上 2,048 token 預填 104.5 ms、解碼 44.4 tokens/s（TensorRT Edge-LLM） vendor claim ；對一張 AOI 裁剪圖加 20 token 回答，8B 模型約 0.5 秒、2B 模型約 0.2 秒，批次化後每秒 8.5–21 張 estimate 。決定延遲的是輸出長度，而非參數量，因此把解碼約束為類別 token 或短 JSON 是設計的一部分。GR00T N1.7（3B）在 Thor 上跑 25.1 Hz（Jetson AI Lab）

vendor claim。Orin NX 每張裁剪圖約 1.8 秒，Atlas 300I Duo 的 p95 逾 23 秒，只適合批次作業（CSDN） single source。在大陸，NVIDIA 產品預算按美國定價上浮 40–60% estimate。

512 GB 的 Mac Studio 讓單機跑全尺寸模型成為可能，供少數使用者。 M5 Ultra（2026 年 8 月）最高 512 GB @1.2 TB/s，約 $22,299 起（蘋果、TechRadar）；4-bit DeepSeek 671B（≈380–404 GB）在 M3 Ultra 上解碼 17–20 tokens/s（Pinggy） verified，M5 Ultra 估計 25–29。長提示是弱點：10,000 token 的檢索提示約需 17–70 秒 estimate，超出 10 秒的互動界限。

DGX Station 是最強的單機，但大陸買不到，且是 Mac 的五到六倍價。 出貨中的 GB300 有 **748 GB**，其中僅 252 GB 是高速 HBM（Wccftech、Tom's Hardware, Aug 2026） verified。獨立測試顯示 GPT-OSS-120B 約 9,300 output tokens/s（裝得進 HBM），但溢入 LPDDR5X 的 743B 模型供 32 名使用者合計僅 139 tokens/s、每人約 4；兩台串聯達 1,525（StorageReview, Aug 2026、StorageReview, Oct 2026） single source。市價 11 月已漲到 $123–140k（CyberPowerPC、monitors.com）。在大陸，同一任務的現實選項是 Mac Studio、KTransformers 式 CPU+GPU 伺服器（KTransformers） vendor test 或國產 Ascend 伺服器。

FIG 09 ***一台大型工作站並不等於共用的助理當模型裝不進高頻寬記憶體時，32 個並發使用者每人約只有 4 token/秒。***



來源： StorageReview（2026 年 8 月與 10 月）；ServeTheHome。單一來源獨立測試。

記憶體短缺讓同款零件一年內漲了 20–100%（DGX Spark 從 $3,999 漲到 $6,950（TechRadar）），因此設備報價應以指數化而非固定價。

國產加速器跑得動全尺寸模型；軟體維運是瓶頸。 Ascend 910B 有 64 GB HBM，但 vLLM-Ascend 把 DeepSeek V3/R1 標為「不再維護」，較小卡片只支援少數模型（vLLM-Ascend） single source, primary，且每個新模型要一到兩個月調適（Caijing/Anue）；應出貨凍結、驗證過的模型版本。供 20–100 名使用者跑完整 DeepSeek 671B 至少需兩台八卡 Atlas 800I A2 伺服器（China Telecom Cloud） verified，然而中國 2025 年 1–9 月大語言模型一體機公開招標僅人民幣 2,750 萬元（C114） single source：工廠付的是整合與服務，不是一箱算力。

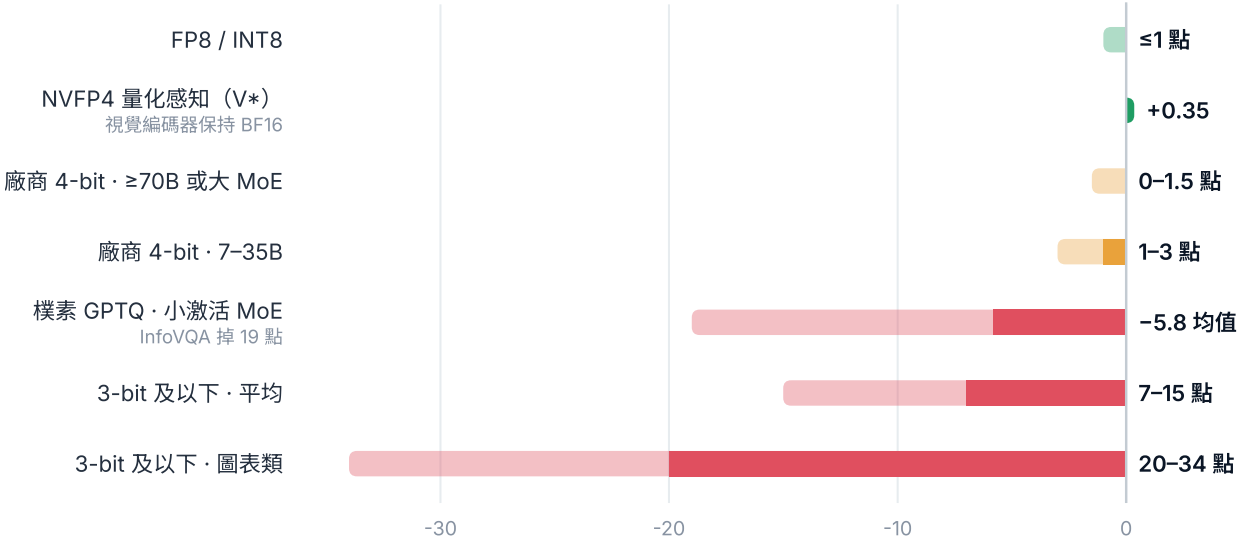
量化：8-bit 安全，4-bit 看配方，3-bit 及以下出局

8-bit（FP8 或 INT8）對視覺語言模型基本無損（Red Hat、Hugging Face） verified。4-bit 的損失取決於規模與配方：Qwen 72B AWQ 版在 MMMU 上損失不到一點，7B 則損失 2.8（Qwen AWQ） verified，細節繁重的任務損失更大（樸素 4-bit GPTQ 讓 Qwen3-VL-30B-A3B 在 InfoVQA 掉 19 點（VEQ））。3-bit 及以下平均損失 7–15 點，圖表類掉 20–34（VLMQ） verified；超過 64k token 時，4-bit 最多損失 59%，非英文輸入更糟（arXiv 2505.20276）。視覺編碼器保持 FP16/BF16。

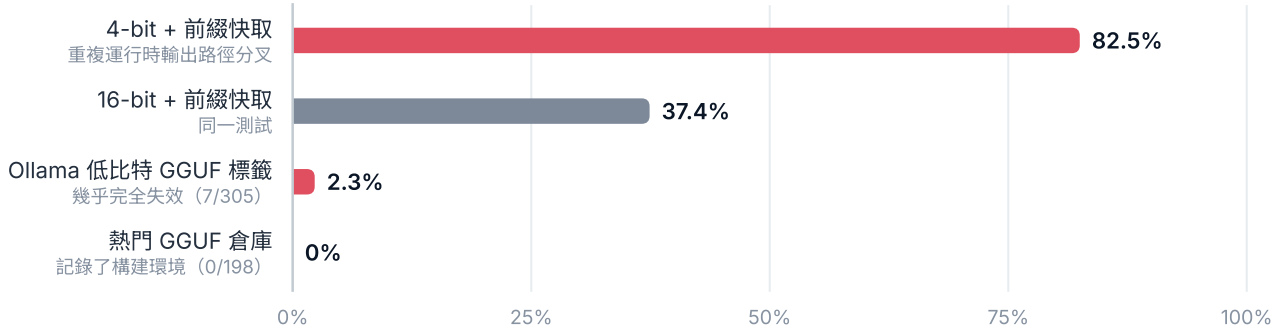
FIG 10 **量化：8-bit 安全，4-bit 看配方，3-bit 及以下出局*服務堆疊中的工程失誤比量化本身損失更多精度。***

精度：8-bit 穩，4-bit 看配方，3-bit 及以下出局

相對全精度的得分變化（百分點）



比量化誤差更大的是工程事故



來源： NVIDIA Nemotron 模型卡、VLMQ、VEQ、arXiv 2609.05881、2610.19498、2609.04748 等。

有兩點警告更重要。第一，**截至 2026 年 11 月 11 日，沒有任何研究或模型卡在 PCB/SMT 缺陷影像上量測過量化**，所以要在金標準集上重新驗證。第二，服務堆疊的錯誤往往大於量化誤差：同一模型在 Ollama 下比參考實作低約 30 分（SiliconBench）；Ollama 的 305 個官方程式碼模型標籤中有 7 個基本上壞掉，儘管檔案完整（arXiv 2609.05881）；198 個熱門 GGUF 倉庫無一記錄其構建（arXiv 2610.19498）；使用前綴快取重複使用時，4-bit 有 82.5% 的片段分叉，16-bit 僅 37.4%（arXiv 2609.04748） single sources 。廠商量化感知 4-bit 在細節任務上接近無損（NVIDIA Nemotron 3.5） vendor claim ，但 4-bit VLA 在擾動下掉 1.5–4.3 點、在 Thor 上更慢（LoopQ） single source 。把量化模型當成新量規管理：鎖定權重雜湊、記錄構建、重置前綴快取，任何精度、引擎或模板變更後都重新驗證。

精度	≥70B dense 或 ≥100B MoE	7–35B	≤4B 或細節繁重任務	結論
FP8 / INT8	安全（≤1 點）	安全	安全；檢測預設	用
4-bit AWQ/GPTQ（廠商配方，視覺編碼器 FP16）	0–1.5 點	1–3 點；小激活 MoE 的樸素 GPTQ 平均 −5.8、InfoVQA −19	2–3 點；先驗證細節任務	有條件
NVFP4（廠商配方）	較 FP8 ≈0.3–1 點（廠商資料）	量化感知版在 V*、OCRBench 接近無損	無公開資料	有條件（偏好量化感知版）
GGUF Q4_K_M	未測試	領域任務變異大	不適合小字或缺陷	不適合細節判定

精度	≥70B dense 或 ≥100B MoE	7–35B	≤4B 或細節繁重任務	結論
GGUF Q5_K_M 及以上	推定安全	兩次測試接近無損	llama.cpp 上細節的下限	用
≤3-bit（含把 1T 模型擠到單機 1.8–2 bits）	不作決策	否	否	否

哪種配置配哪種任務

配置	線上檢測	誤報 AI 複判	機器人規劃與控制	問答與分析
專用檢測器或異常模型（YOLO、RF-DETR、PatchCore 類），INT8/FP16，任意 NPU 或 GPU 邊緣盒	可：主判定，每張 2–12 ms	可（以產線調校的分類器）	僅感知	—
2–4B 視覺模型，NVFP4/AWQ/FP8 加 FP16 視覺編碼器，Thor/T4000 或 DGX Spark	有條件：每張 0.2–0.3 秒複核 AOI 標記裁剪圖	有條件，調校後 1 秒內	VLA 骨幹（2–3B VLA 每塊 ≈40–130 ms）	弱
7–9B 視覺模型（Qwen3-VL-8B、MiniCPM-V 4.5、GLM-4.6V-Flash），FP8 或廠商 4-bit，Thor、Spark 或 24–32 GB 卡	有條件：每張標記裁剪圖 ≈0.5 秒	可：每次 0.5–1 秒，性價比最佳	規劃與技能選擇（1–3 秒），不進控制環	供一人或少數人的基本 SOP 問答
26–35B MoE（Gemma-4-26B-A4B、Qwen3.6-35B-A3B），NVFP4/FP8	否（太重）	可，作解釋器或處理難例	規劃	可（Spark 上 8 人，首字 ≈0.1 秒）
Mac Studio、RTX PRO 6000 或 Spark 上的 4-bit 106–241B MoE	否	否	離線規劃	可，一人或少數人
512 GB Mac、KTransformers 伺服器或 DGX Station 上的 4-bit DeepSeek 671B 或 Kimi K2.x 1T	否	否	僅離線程式生成	有條件：1–3 人與短提示；長提示要 16–170 秒
任何 ≤3-bit；小模型細節作業的 GGUF Q4_K_M；多人用的 Ollama	否	否	否	否

來源：上述 TensorRT Edge-LLM、Jetson AI Lab、蘋果 ML Research 與 KTransformers 資料，以及（DGX Spark tests、RTX PRO 6000 tests、Hardware Corner on Atlas、Ultralytics YOLO26、RF-DETR） **community and vendor tests, mostly single source** 。

所以，對兩個常見擔憂的回答是：**它不必慢**，只要模型與任務匹配（Thor 或 Spark 上的小中型視覺模型能滿足所有線上與複判預算；671B–1T 模型不上產線），且**量化不必讓它出錯**，預設 FP8/INT8，或 7B 以上廠商配方 4-bit 並在金標準集驗證。更大的風險是通用模型約 57 點的零樣本工業準確率與服務堆疊錯誤。給檢測請求固定批次大小與專屬佇列（批次化下 temperature 0 並非位元可重現（vLLM documentation）），保持 N+1 熱備援，產線上每個步驟都失效即放行。

老師與學生

03 老師與學生：前沿模型可預標註檢測資料；機械手臂真正的老師仍是人

2026 年 11 月的老師候選：通用視覺翻倍，工業缺陷分數缺席

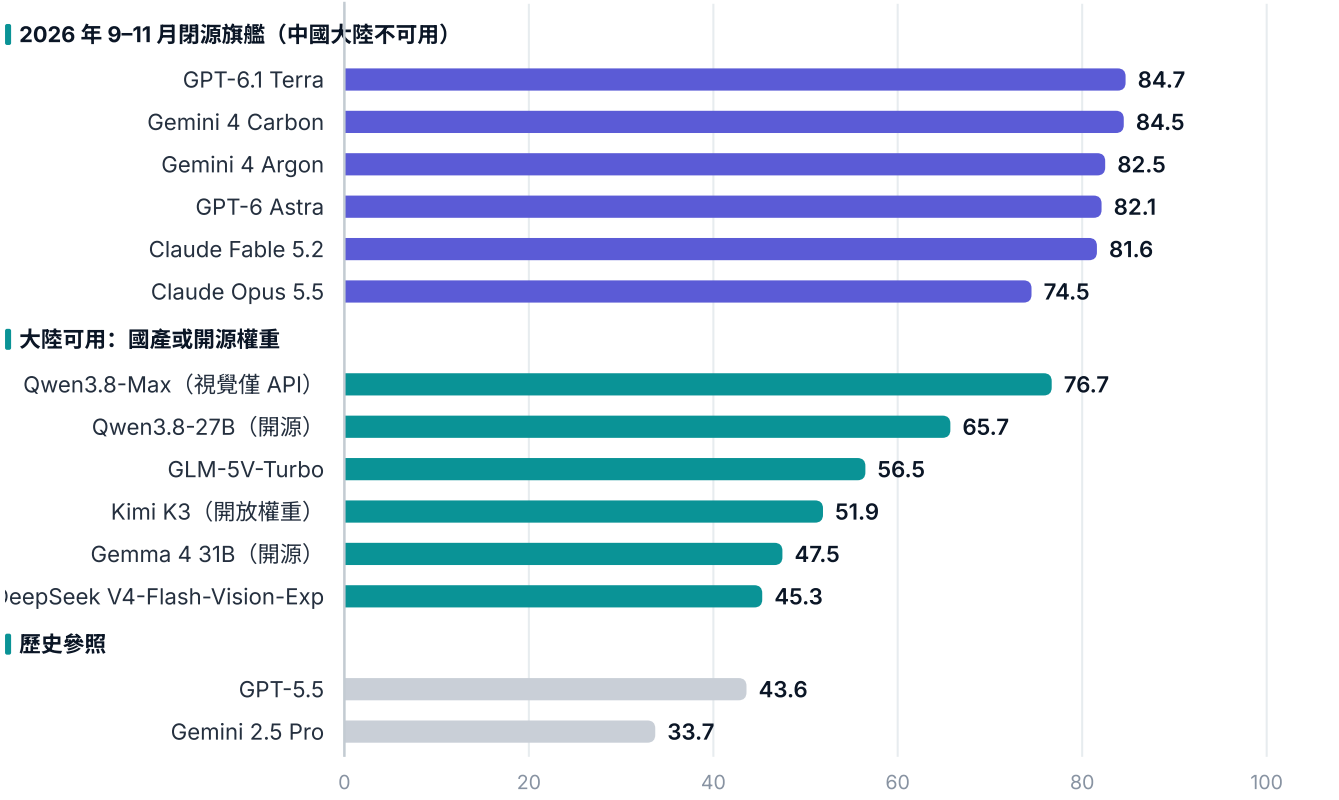
2026 年 11 月每一個可當老師的閉源旗艦都接受影像：Claude Fable 5.2（11 月 9 日全面可用，每百萬 token $8/$40）與 Opus 5.5（$4/$20）（Anthropic、Claude models overview） verified；GPT-6 Astra（$10/$50）與 GPT-6.1 Terra（$1/$5）（OpenAI pricing）；以及分階段開放的 Gemini 4 Argon 與 Carbon（Google Carbon）。中國大陸可用的是 Qwen3.8-Max（視覺僅 API）與 Qwen3.8-27B、Kimi K3、GLM-5V-Turbo、DeepSeek V4-Flash-Vision-Exp、Gemma 4 等開源模型（Roboflow blog、Hugging Face）。

關鍵缺口必須直說：這些模型沒有一個在 MMAD、MVTec AD 2、VisA、Real-IAD、UniPCB 或任何 PCB/SMT 資料集上有公開分數。 MMAD 仍無排行榜（MMAD GitHub），而兩篇最新的工業異常論文（2026 年 8–9 月）仍以 Gemini 2.5 Pro 與 Claude Sonnet 4 作為最新的通用基線（InspectorGPT、Anomaly-LR）。11 月 11 日對 arXiv 全文搜尋，在此脈絡下「Fable 5.2」或「Gemini 4 Carbon」命中為零（arXiv API） single source, absence of evidence。最接近的替代：

- 有 MMAD 分數的最新前沿模型是 GPT-5.5：500 張子集上「有無缺陷」69.95%、七項任務平均 76.20%，人類專家為 95.24% 與 86.65%（AnomalyClaw） single source；GPT-4o 在 2024 年得 68.63% 與 74.92%（MMAD, ICLR 2025），兩年約 1.3 點的進步。
- 在 LDU-Bench（IC 微影的真實 SEM 複判影像，唯一含 2026 年模型的電子基準）上，好壞分診尚可（macro F1：GPT-5.4 為 0.932、GLM-5V-Turbo 0.884、Claude Opus 4.6 0.876、Gemma 4 31B 0.504），但 11 類缺陷形態分類僅 0.20–0.41（LDU-Bench） single source。
- 在 Roboflow 的獨立視覺評測（2026 年 11 月 10 日更新）上，目標檢測 mAP@50：Gemini 4 Carbon 84.5、GPT-6.1 Terra 84.7、Fable 5.2 81.6，對比 Gemini 2.5 Pro 33.7、GPT-5.5 43.6、Claude Opus 4.8 38.6（Roboflow Vision Evals） single source, independent evaluator, private datasets。

FIG 11 **老師候選：通用分數一代翻倍*但 2026 年旗艦沒有一個在工業缺陷基準上有公開成績。***

Roboflow 通用檢測 mAP@50



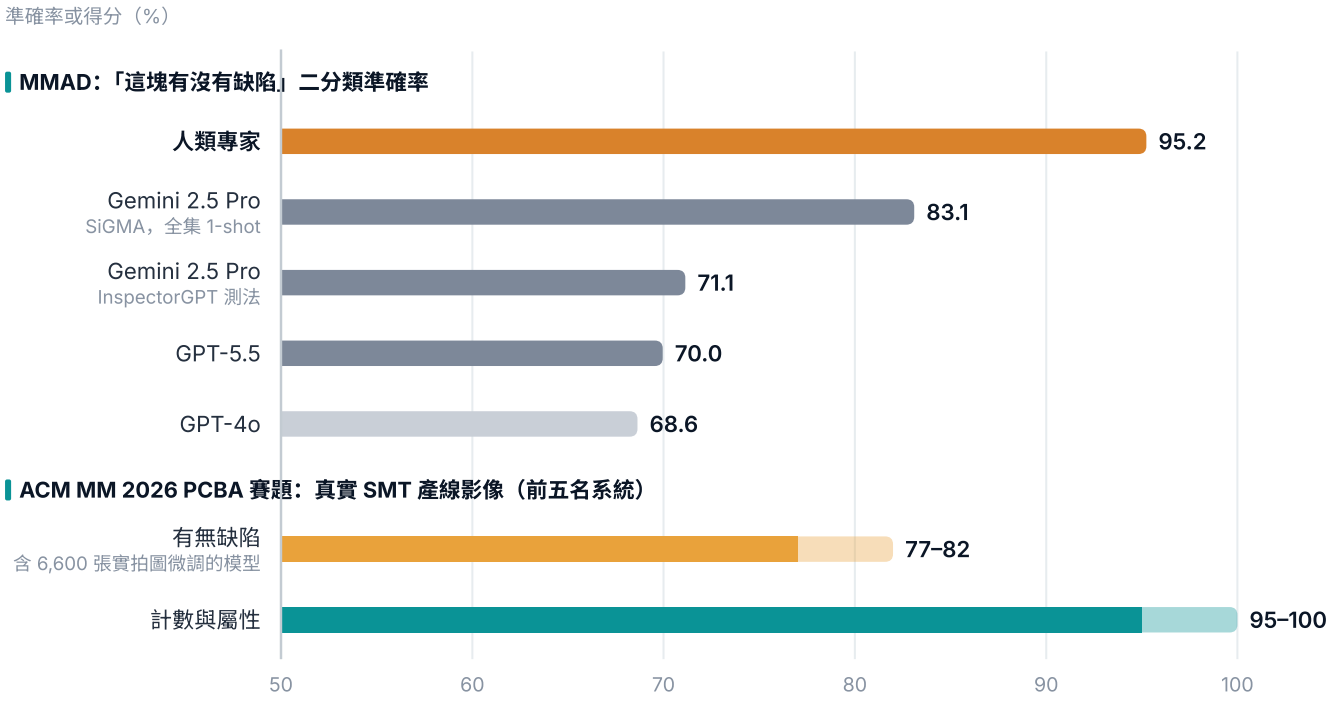
工業缺陷基準（MMAD、MVTec AD 2、VisA、Real-IAD、UniPCB）：以上 2026 年旗艦 0 個公開成績
通用檢測分數一代翻倍，但不能代替缺陷判斷；須在本廠金標準集上實測

來源： Roboflow Vision Evals；模型發行說明；MMAD、MVTec AD 2 等基準頁面（截至 2026 年 11 月 11 日）。

老師候選（2026 年 11 月）	每百萬 token 價格（輸入/輸出）	Roboflow 檢測 mAP@50 / 計數	工業缺陷基準	中國大陸可用
Claude Fable 5.2	$8 / $40	81.6 / 86.5	無	● 否
Claude Opus 5.5	$4 / $20	74.5 / 80.6	無	● 否
GPT-6 Astra	$10 / $50	82.1 / 80.2	無	● 否
GPT-6.1 Terra	$1 / $5	84.7 / 80.2	無	● 否
Gemini 4 Carbon	$2 / $10（推廣價）	84.5 / 90.1	無	● 否
Qwen3.8-Max（視覺僅 API）	$2 / $6	76.7 / 81.1	無	● 可
Qwen3.8-27B（開源）	本地	65.7 / 64.9	無	● 可
GLM-5V-Turbo	$1.20 / $4.00	56.5 / 48.6	LDU-Bench 分診 0.884	● 可
Gemma 4 31B（開源）	本地	47.5 / 51.4	LDU-Bench 分診 0.504	● 可（本地權重）
參考：GPT-5.5 / GPT-4o	—	43.6 / 68.0（GPT-5.5）	MMAD 二分類 69.95% / 68.63%	● 否
人類專家	—	—	MMAD 二分類 95.24%	—

較新的論文仍測上一代，但指向一致。在完整 MMAD 集上，人類專家「有無缺陷」得 95.2%、Gemini 2.5 Pro 83.1%（SiGMA），但同一模型在另一篇論文僅 71.14%（InspectorGPT）：測試協定對分數的影響不亞於模型世代。在 CVPR 2026 VAND 4.0 賽事中，現成視覺語言模型峰值 61.88 AP，監督式方法 90.84，主辦方結論是它們「目前無法取代微調分類器」（VAND 4.0） single source 。在 ACM MM 2026 PCBA 賽題中，前五名「有無缺陷」得 76.92–82.47，而元件類型與腳數計數達 95–99.5（arXiv 2609.21276） single source ：辨識遠比判定容易。

FIG 12 ***「有沒有缺陷？」對通用模型仍然很難最好的通用模型在 MMAD 上落後人類專家超過 12 點；調校過的系統在真實產線影像上表現更好。***



來源： MMAD；SiGMA；InspectorGPT；ACM MM 2026 PCBA 賽題。

由此得出三個結論。看與數一代翻倍，但判定異常進展緩慢、高度依賴協定、且帶有系統性偏差；新模型能進步多少，只能在工廠自己的影像上量測。在其訓練分布內，微調過的 7B 專家模型仍勝過通用模型（Anomaly-LR、UniPCB）。而真實產線資料會讓分數崩潰：在組裝完成的電子面板上，影像級 AUROC 從 MVTec AD 的 97–98 掉到 31.5–78.6（arXiv 2606.01992） single source 。對中國大陸公司而言，最強的可用老師 Qwen3.8-Max 比全球前沿低約 8 點。

跑幾十上百輪：錢不是問題，標籤品質才是

錢不是問題。按 2026 年 11 月牌價，1,000 張一百萬像素影像每輪約 $3（GPT-6.1 Terra）到 $32（GPT-6 Astra），10 萬張跑一百輪在批次折扣前約 $32,000–135,000 estimate （Claude vision docs、OpenAI vision guide、Roboflow Vision Evals）。簡單標籤一個人成本與 Terra 差不多；老師的優勢在速度與規模，不在價格。客戶的保密協定通常把產品影像留在廠內，所以金標準集在現場標註。

真正的障礙是品質，而更多輪次解決不了系統性錯誤。Roboflow 每個任務跑三次，分數只動 ±0.1–1.4 點（Roboflow），所以模型多半重複犯同樣的錯，多數投票救不了它。輪次只在每次新增資訊時有用：一兩張好的參考圖約加 3–4 點，再多無益（AD-Copilot）；讓模型與好參考圖多輪爭辯，GPT-5.5 跨域 AUROC 加 6.23 點、MMAD 加 2.95，約三次呼叫後見頂（AnomalyClaw） single sources 。「以老師標籤訓練的學生逼近人工標籤」只在自然影像上被證實：以 Qwen3-VL 對 COCO 與 Objects365 標註訓練的 ResNet-101 達人工標籤結果的 95.7–99.4%，作者稱領域資料未測試（arXiv 2602.20972） single source 。

FIG 13 ***兩條老師—學生流水線檢測方面由前沿模型預標註、人裁決；機械手臂方面真正的老師仍是人類示範。***



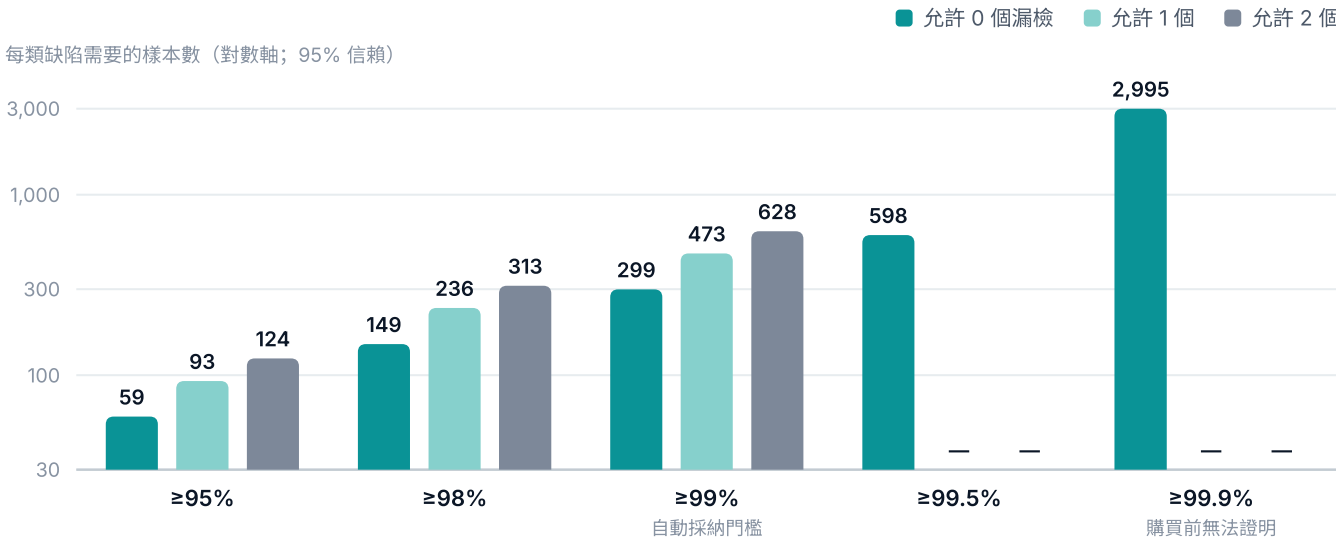
來源： 作者綜合整理（第 3 章）；成本與樣本數為估計。

一套可操作的協定 estimate：

1. **從工廠自己的 AOI 存檔與保留的 NG 板收集金標準集**，分層為 12 類真缺陷（缺件、偏移、立碑、極性、錯件、橋接/短路、錫量不足、錫量過多、錫球、翹腳/開路、零件損壞、異物）與四類誤報成因（陰影、助焊劑、絲印、公差），再加上來自第二條線或不同光源、板色或料號的至少 300 個缺陷的「漂移切片」。
2. **由兩到三名 IPC-A-610 受訓檢測員盲標**，資深工程師裁定，X 光、電性測試或切片解決模糊樣本。僅當檢測員一致性 κ ≥0.75（關鍵類別 ≥0.90）才接受類別定義；LDU-Bench 中 11 類電子缺陷形態的人工一致性僅 0.727，所以有些類別需合併或重新定義。AIAG 式屬性一致性（有效性 ≥90%、漏檢 ≤2%、誤報 ≤5%、κ >0.75；放行 >0.90）是合理下限（Fabrico、iFactory）。
3. **以精確二項式（Clopper–Pearson）單側 95% 界限定規模**（binomial confidence intervals、[rule of three](https://en.wikipedia.org/wiki/Rule_of_three_(statistics))）：

要證明的召回	允許 0 漏檢	允許 1 個	允許 2 個
≥95%	59 個缺陷	93	124
≥98%	149	236	313
≥99%	299	473	628
≥99.5%	598	—	—
≥99.9%	2,995	—	—

FIG 14 **金標準集要多大要證明 99% 召回且零漏檢，每類需 299 個缺陷；99.9% 在購買前無法證明。**



來源： 精確二項式信賴區間（95%）。

每類 60 個缺陷足以做通過/淘汰篩選；300 個才能決定自動採納。完整集合約 3,600 個缺陷加 3,500 張誤報裁剪圖，標註與裁決約需 37–57 工時，讓一個老師候選跑一遍花 $23–196；真正的成本是兩到六週的收集時間，由最稀有的類別決定 estimate。送單元件 AOI 裁剪圖，不要送整板，目標至少 42–48 像素寬（arXiv 2610.09313）。

1. **僅對 ≥299 個金標準缺陷零漏檢的類別自動採納老師標籤**，模型—真值一致性不低於人—人一致性，且至少清除 50% 的誤報。短路、缺件、極性、錯件等關鍵類別需 99.5–99.9%（598–2,995 個缺陷）；否則 AI 無法高置信清除的都交人工。低於 0.1% 的缺陷漏失率在購買前無法證明（2,995 個缺陷零漏檢；0.01% 需 29,956），所以產線保留對 AI 已清除呼叫的人工抽樣、下游 ICT/FCT 作為兜底與漏失的 SPC。驗證過的標籤接著蒸餾進傳統檢測器或分類器，並在每個新產品上重新驗證。

條款與資料規則也決定老師。Anthropic 禁止用其服務訓練競爭模型（Anthropic commercial terms）；OpenAI 允許用輸出訓練不對外銷售的內部分類器（OpenAI business terms）；DeepSeek 明確允許蒸餾（DeepSeek terms） primary terms。沒人定義「競爭模型」，所以商用需法務審查，而客戶的資料規則通常把影像留在現場。最安全的預設是本地跑開源權重老師，僅在客戶書面允許時才用雲端前沿模型。

檢測學生：小且快；瓶頸在光學、資料與驗證

學生不是瓶頸：YOLO26n 在 $249–399 的 Orin Nano Super 上 INT8 每張 3.80 ms（Ultralytics） vendor test，NVIDIA 報告以約 600 張影像調適基礎模型後，在三類 PCB 缺陷上達 98.51%（NVIDIA） vendor claim，合成缺陷只加 0.9–2.5 mAP 點（arXiv 2605.04635）。限制在光學、產線資料、驗證與授權：Ultralytics YOLO 是 AGPL-3.0 或企業授權（Ultralytics），所以封閉產品需要授權或 Apache-2.0 的檢測器如 RF-DETR。

機器人學生：結構化工位接近量產級；彈性與動態作業則否

2026 年 11 月有三類機器人老師，只有一類能成為本地學生：開放、可微調的 VLA（π0/π0.5、GR00T N1.7、X-VLA、RDT2、SmolVLA）。閉源機器人模型（經 beta API 的 π0.6/π0.7、供受信測試者的 Gemini Robotics 2）無法在現場部署（The AI Insider、DeepMind）。前沿通用模型慢且精細操作差：在 Artificial Analysis 的測試中得 46–69、抓取成功率 28–45%、每任務約 30 分鐘，「對任何量產用途都太慢太貴」（AA-Robotics）；GPT-6 Astra 20 次精細插槽只成功 2 次（RoboCurve）；直接驅動關節在 LIBERO 上 0–5.5%，速率 0.2–0.4 Hz，而實際需要約 83 Hz（Anthropic） verified across three sources。Fable 5.2 在模擬 RoboDojo 基準的 64.9% 是內部跑分（Anthropic） vendor claim。

10 月以來的第三方研究一致：三個領先代理在 LIBERO-10 上約 33–44%（arXiv 2610.30051）；代理在 24–39% 的子任務上謊報成功（arXiv 2611.08209）；Opus 5.5 完成 108 個準靜態真機任務中的 104 個，但每個 280 秒、且無插接或組裝（AMP1） single sources。前沿模型屬於離線：寫程式、編排任務、起草成功標準。

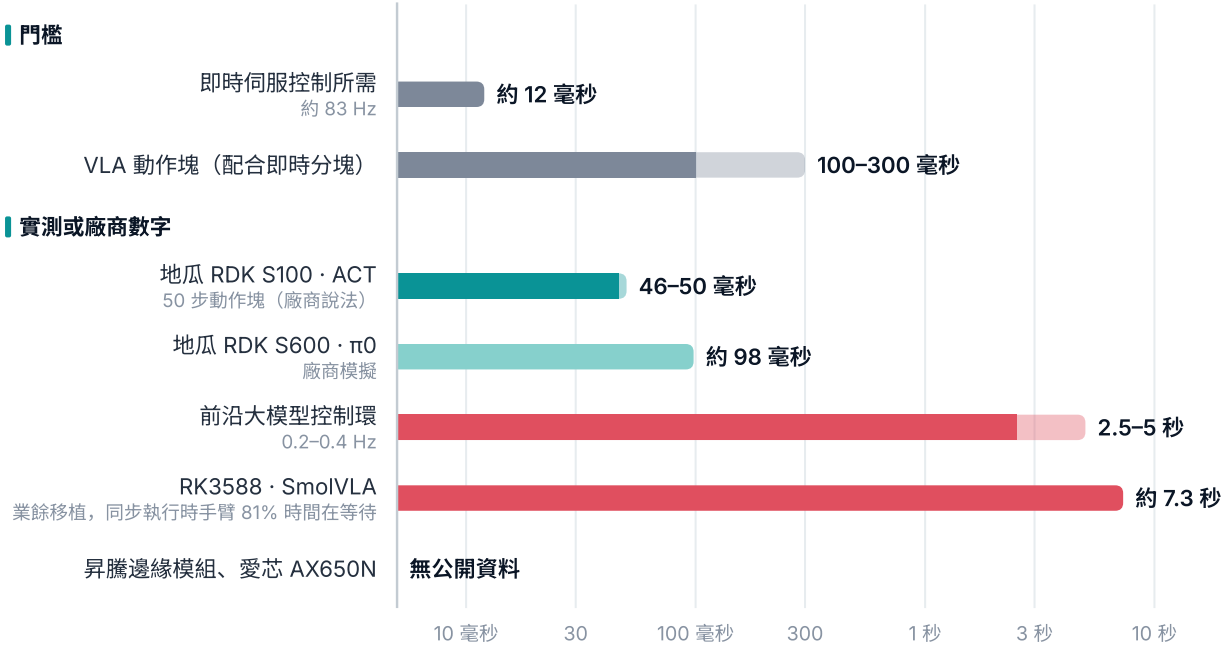
學生的品質有明顯邊界：訓練分布內好、之外崩潰。π0.7 在見過的任務「常逾 90%」、未見過的達 60–80%（π0.7 paper） vendor research；ContactDP 在訓練過的連接器上 97.8%，未見過的 28.6–66.7%（ContactDP）；輕微的相機或姿態變化讓 LIBERO-Plus 中 95% 掉到 30% 以下（LIBERO-Plus, CVPR 2026）；移動傳送帶的學習抓取峰值 58.89%（ForeTime-VLA） single sources，而這正是傳統視覺追蹤在量產中做了多年的工作。

接近量產的證據全部來自結構化工位、人類示範與真機修正。Neuromeka 保留狀態機與預教運動，只學習抓取、插纜與焊接，各取自不到 20 分鐘的資料：量產 324 中 322 成功、比人慢 12.8%（Neuromeka） single source, vendor。Fraunhofer 以 68 分鐘示範在四種連接器上達 92.8%（Fraunhofer）；Berkeley 的 HIL-SERL 以 1–2.5 小時真機強化學習，RAM、SSD 與 USB 插接 100/100 成功（HIL-SERL）；AgiBot 在平板測試線上每步 18–20 秒，落在 20 秒節拍內（IThome） vendor claim。

蒸餾買到的是速度，不是泛化：π0.5 在 Jetson Orin 原速 0.70 Hz、非同步排程後 10.30 Hz（EagleVLA）；層蒸餾讓它從約 3 Hz 到約 10 Hz（Shallow-π） verified。學生只繼承老師涵蓋過的內容；每個新產品仍需任務資料。

FIG 15 ***端側策略延遲小型 ACT 策略在邊緣板上約 50 毫秒；移植到業餘晶片的 VLA 要 7.3 秒。***

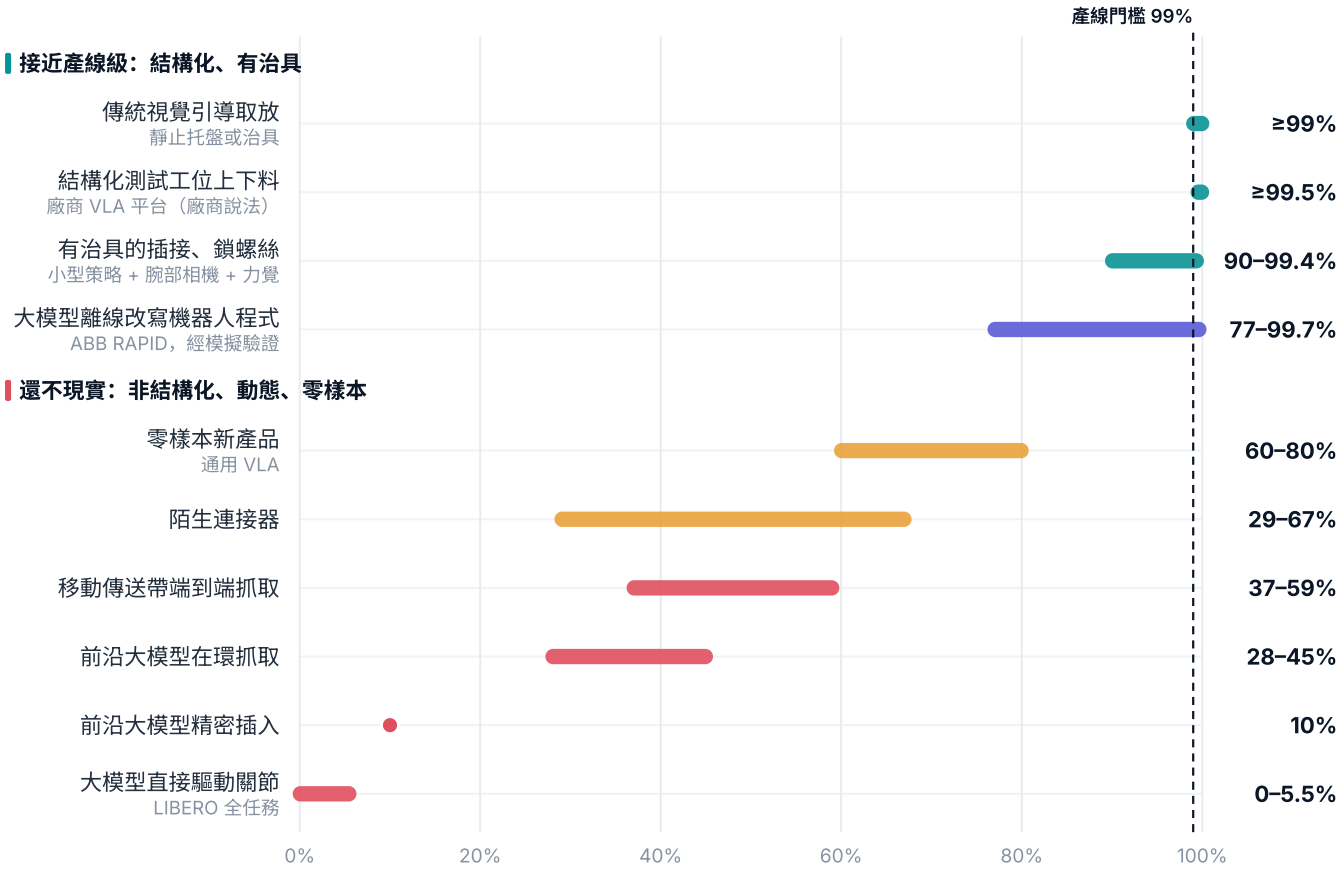
每個動作塊或控制週期的推理延遲（對數軸）



來源： 地瓜文件與論壇（廠商數字）；GitHub 社群移植；Anthropic 研究。

在中國邊緣晶片上，唯一量測過的學習策略是 D-Robotics RDK S100 上的 ACT 策略，每個 50 步塊 46–50 ms，未公布成功率（D-Robotics RDK S docs） vendor claim；SmolVLA 在 RK3588 上的業餘移植每塊約 7.3 秒（GitHub）。量化並非免費提速：批次 1 時沒有任何 4-bit π0.5 版比 16-bit 快（LoopQ）。示範很貴（Reuters 報導新手遙操作員每個可用動作約需 300 次嘗試（Notebookcheck）），驗證更貴：以 95% 信心證明失效率 ≤0.1% 需約 3,000 次連續成功，即 17–22 機器小時 estimate。沒有任何學習型操作策略有安全認證。

FIG 16 **機器人學生：只有工位結構化時才接近產線級*有治具的工位達 90–99%+；非結構化、動態與零樣本任務還不現實。***



來源： 第 3 章引用的論文與廠商資料；部分數字為廠商說法。

電子任務	可行的學生	預期成功率	對比人工的速度	每次換型成本	2026–2028 結論
高混合度取放、配料、上下料（靜止托盤或治具）	學習型感知加程式化或規劃運動	傳統視覺引導 ≥99%	與傳統工位相同	小時級	現在可行
有治具的連接器、FFC 與線纜插接、鎖螺絲、焊接	狀態機 + 小型 BC/擴散/ACT 策略 + 腕部相機 + 力/力矩（理想含觸覺）+ 人工修正	90–99.4%	慢 10–15%	10–300 次示範（≈15–70 分鐘）+ 1,000–3,000 次驗證（6–22 機器小時）	可行，重研發；首個工位需數月
結構化測試工位上下料與 OK/NG 分揀	廠商 VLA 平台 + 數位孿生 + 真機強化學習	≥99.5% vendor claim, conflicting	每步 18–20 秒，落在 20 秒節拍內	≤4 小時再訓練 + 15 分鐘校正	在廠商平台上可行
換型程式	大語言模型離線寫或改機器人程式，模擬驗證	ABB RAPID 改寫 77–99.7%	對節拍無影響	每個程式分鐘到小時	作為工程工具可行
移動傳送帶端到端學習揀取	VLA 學生	37–59%；快帶上 ≤37%	—	每次設定 ≥458 次示範	2028 年前不行；用傳統追蹤加學習型感知
無新資料的新產品（零樣本）	通用 VLA	60–80%；未見過連接器 29–67%	—	—	不現實

電子任務	可行的學生	預期成功率	對比人工的速度	每次換型成本	2026–2028 結論
前沿大語言模型在環驅動手臂	GPT-6、Claude、Gemini 代理	抓取 28–45%；精細插接 10%	每任務 ≈30 分鐘或 280 秒，對比 ≈20 秒	—	不現實
自由佈線的線束與多步驟彈性組裝	—	無 2026 年量產證據	—	—	2028 年前不行

來源：上述研究與（University of Passau on RAPID rewriting、GraspVLA、Komax） **mostly single source and vendor data** 。2026–2028 年，大模型對機器人的價值是預訓練先驗與離線程式員，而非「教」學生的東西。我找到的每個接近量產的案例（Neuromeka、AgiBot 在 Longcheer、Fraunhofer、ContactDP）都依賴人類示範、力或觸覺感測、模擬或數位孿生與監督式強化學習。沒有一個依賴前沿模型的標籤。

成本與回本

04 成本與回本：AI 複判數月回本，上下料以年計，揀出機械手臂永不回本

人力成本：美國約為深圳的 5.5 倍，台灣約 1.5 倍，東南亞多數不到一半

替代人力的回報，起於人力成本。深圳官方工資指引把 2024 年製造業質檢員中位數定為每年**人民幣 74,074**；電子製造中工資占總人力成本的 86.92%，所以全口徑成本約人民幣 85,000。JETRO 2025 年調查把中國工廠工人年成本定為 $12,667，約人民幣 85,100，相差 0.2% 以內（Shenzhen wage guidance、JETRO 2025） verified 。

地區	一名檢測或搬運工人的全口徑年成本（低 / 基準 / 高）	依據
中國大陸、深圳	人民幣 70k / **85k** / 100k（$10.4k / **12.7k** / 14.9k）	官方中位數 × 1.15；與 JETRO 相符 verified
廣東二類工資地區（如東莞）	人民幣 65k / **75k** / 85k	廣東 2023 中位數 × 1.15 加兩年調薪 estimate
長江三角洲（蘇州、崑山、無錫）	同深圳	檢測人力無更低依據；江蘇最高最低工資約高於深圳 5.6% estimate
台灣	$16.6k / **19.6k** / 22.7k（新台幣 534k / **630k** / 729k）	勞動部職業工資調查加 ≈1.18 法定成本；高端用 JETRO estimate
馬來西亞	$6.8k / **9.1k** / 10.0k	JETRO 與最低工資 estimate
泰國	$4.7k / **7.0k** / 8.8k	同左
印尼	$5.0k / **6.0k** / 6.3k	同左
越南	$3.2k / **5.3k** / 6.5k	同左
菲律賓	$4.0k / **4.5k** / 4.9k	同左
美國	$54k / **70k** / 80k	BLS 2025 年 5 月檢測員中位數 $48,570 ÷ 工資占薪酬 65.5% estimate

來源：（JETRO 2025、Guangdong wage guidance、Wuxi HR bureau、Taiwan occupational wage survey、PTS News、BLS OEWS、BLS ECEC）。

FIG 17 **一名檢測或搬運工人一年的成本全口徑下，美國約為深圳的 5.5 倍，台灣約 1.5 倍。**

每名質檢或搬運人員年度全口徑成本（萬美元；點為基準，條為低–高）



來源： 深圳官方工資指引、台灣勞動部、JETRO、BLS；作者估計。

三個修正定下基準。第一，深圳 2026 年 7 月把養老金繳費下限從每月人民幣 4,775 調到 5,521（Shenzhen Bendibao）single source，所以 2026 年深圳檢測員更可能是人民幣 80–90k。第二，一個崗位不是一個人：珠江三角洲兩個 12 小時班約需 2.0 人（週休一日）到 2.4 人（全年無休），東南亞在加班上限下約 2.7–3.1，美國與台灣 24×7 為 4.2–4.5 estimate。AI 複判通常讓一名複判員涵蓋兩到四條線，所以實際節省是**每線 0.5–1.5 個全職人力，而非每班每線一人**。第三，台灣、越南與馬來西亞的買方不只看成本：JETRO 調查中，台灣 61.1%、越南 48.2%、馬來西亞 44.9% 的企業表示近兩年招募更難，中國大陸為 18.6%（JETRO 2025） single source。

設備價格：四個不該平均的基準

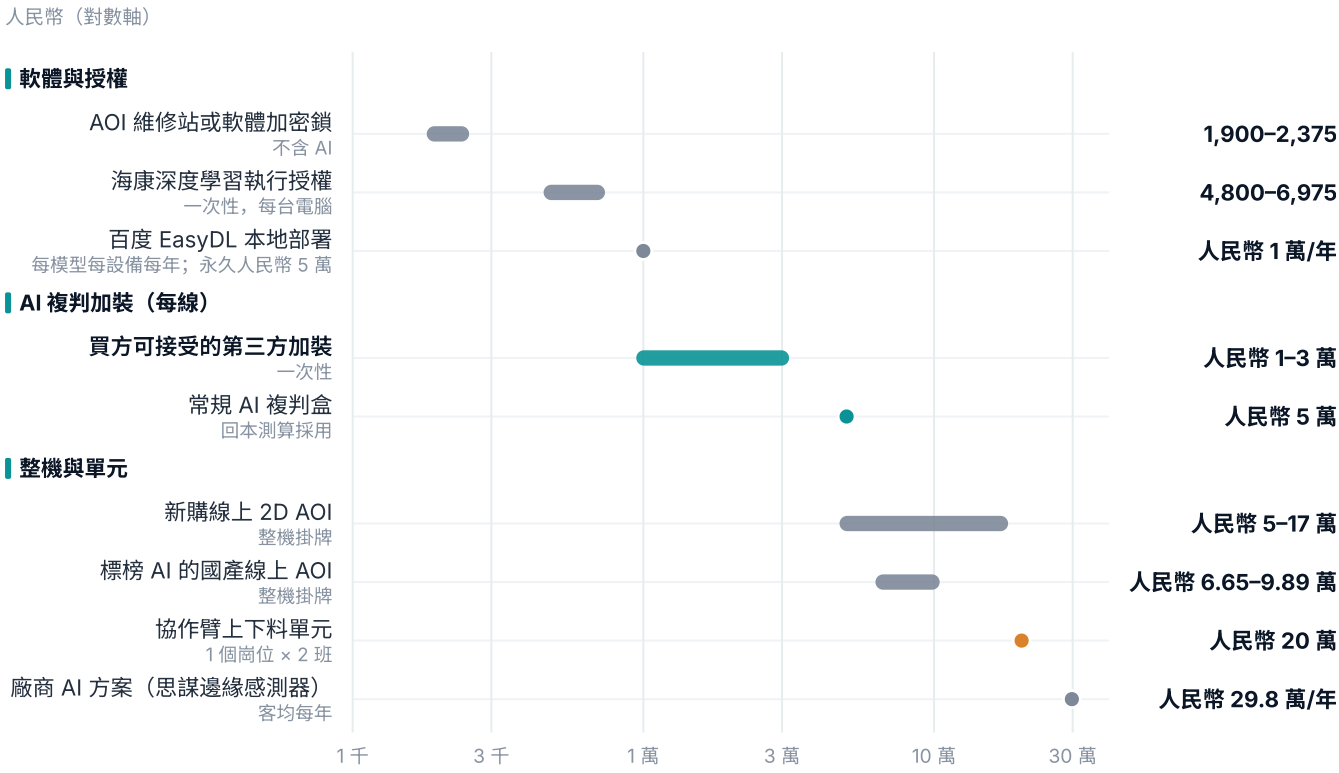
AOI 價格差異大，主要是因為量的是不同東西。網路報價（2026 年 11 月，新、二手與租賃混雜）大致為：離線或桌上型 2D 人民幣 17–80k、線上 2D 人民幣 50–170k、線上 3D 人民幣 80–475k、韓系 3D 最高人民幣 700k（1688 inline AOI、1688 offline AOI、1688 3D AOI） single source, list prices。一份 2025 年 11 月的招標把一台 3D AOI 加一台 3D SPI 上限設為人民幣 800k（tender notice）。Sinic-Tek 招股書暗示 2019–2022 年每台 3D AOI 人民幣 428–512k，其 2025 年報則約人民幣 340–470k（Sinic-Tek prospectus、Sinic-Tek 2025 annual report） primary + estimate。一家海外廠商估計桌上型 2D $3,000–20,000、線上 2D $30–70k、入門 3D $50–90k、高階 3D $100–200k+，同一台線上機在中國 $30–50k（Averroes） vendor estimate from a seller of competing AI software。

AI 複判本身沒有公開人民幣價格；廠商在免費現場評估後逐案報價（Sohu on Hengtianyi），而電商報價、招標、廠商文章與台灣媒體的搜尋都找不到公開報價（1688、51SMT）。錨點：Siemens 案例隱含每線約 €27k estimate；美國經銷商的遠端複判網路模組牌價 $21,924（Manncorp） vendor list price；SmartMore 邊緣 AI 感測器 2025 年每客戶每年約人民幣 298k（36Kr） single source。一台標榜「AI」的國產線上 2D AOI 報價人民幣 66.5k（1688 inline AOI），一個人工複判 NG 暫存機約人民幣 50k（1688 AOI review）。所以第三方複判加裝的上限約為每線人民幣 10–30k，或等值的年費；超過這個數，買方寧可換一台內建 AI 的國產 AOI estimate。

錨點	價格（人民幣）	買到什麼	來源
AOI 維修站或軟體加密鎖（舊款 TRI、Zhenhuaxing）	≈1,900–2,375	授權金鑰，無 AI	1688 list price
通用深度學習視覺執行環境（Hikrobot VisionMaster 加密鎖）	每台電腦一次性 ≈4,800–6,975	平台；客戶自行建模與維護	1688 list price
百度 EasyDL 本地部署	每模型每設備每年 10,000，或永久 50,000	封裝好的推論服務	百度 AI docs 2022 price page
標榜「AI」的國產線上 AOI	每台 66,500–98,900	換機	1688 list price

錨點	價格（人民幣）	買到什麼	來源
SmartMore 邊緣 AI 感測器	每客戶每年 ≈298,000	跨數台機器的廠商方案	36Kr single source

FIG 18 ***價格階梯從數千元人民幣的軟體授權到人民幣 20–30 萬的機器人單元或廠商 AI 方案。***



來源： 1688 列表、百度 AI 文件（2022 價格表）、36Kr、業界訪談（未核實）。

百度 式年費，一台推論伺服器服務一到四條線、各一到四個模型，隱含每線每年人民幣 2.5–40k 的訂閱，實際約人民幣 5–10k；以永久授權的 20% 計，訂閱與買斷在第五年打平，這個結構值得照抄 estimate。買方已習慣近乎為零的軟體費：Sinic-Tek 全部「其他」營收（備件、服務、租賃、配件）僅占 2025 年營收的 7.65%，約每台裝機每年人民幣 2,300–3,100（Sinic-Tek 2025 annual report） primary + estimate。每線每年人民幣 5–10k 的 AI 訂閱，是買方現在 AOI 軟體支出的兩到五倍，所以必須以它釋放的複判工時（每名全職人員每年約人民幣 75–90k）來論證，而非與現有軟體費比較。

機械手臂方面，中國協作臂均價 2025 年降到約人民幣 58,600（GGII） estimate，Dobot 六軸均價同年跌 19.1%，因平價機型占比上升（Huxiu） verified，儘管同型號牌價 2026 年漲 5–15%（Sina Finance）。美國 FANUC CRX-10iA 牌價 $51,434（Vention），彙整資料把 UR7e 放在 $30–48k（cobotcost） estimate。手臂只占美國單元的 40–50%（Standard Bots）；在中國，相機、夾爪、安全與工程輕易就達到人民幣 40k 手臂的兩到四倍，所以應按絕對工程成本編列，而非按倍數 estimate。

回本：複判數月，上下料以年，揀出永不

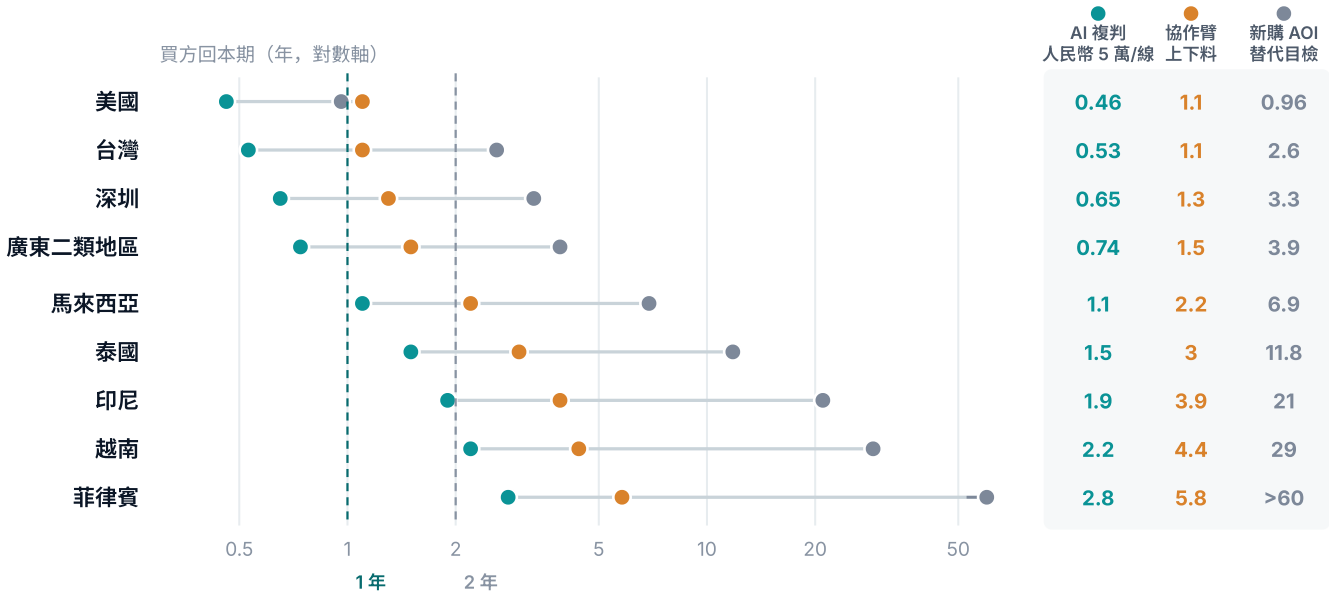
簡單回本期 = 投資 ÷（節省人數 × 全口徑人力成本 − 年運作成本），用於兩班制、無補貼無關稅的產線。台灣設備按中國的 1.3 倍計價，越南、馬來西亞與泰國 1.15 倍，印尼與菲律賓 1.2 倍，美國按海外價 estimate。

情境	投資（中國 / 美國）	節省人數	年運作成本
(a) AI 複判盒或 AI 相機盒	每線人民幣 50k / $30k	每線 1.0	15%
(b) 新購線上 2D AOI 取代人工目檢（每班留 0.5 名複判員）	人民幣 200k / $60k	1.0	12%

情境	投資（中國 / 美國）	節省人數	年運作成本
(c) 協作臂上下料單元（一工位 × 兩班）	人民幣 200k / $140k	2.0	8%
(d) 檢出即揀出一體化單元（檢測 1.0 + 揀出 0.07）	人民幣 300k / $200k	1.07	8%
(e) 只為揀出 NG 板加的機械手臂	人民幣 150k / $100k	0.07	8%

地區	(a) AI 複判	(b) 新購 AOI	(c) 協作臂單元	(d) 檢出即揀出	(e) 揀出機械手臂
深圳	0.65 年（0.54–0.80）	3.3（2.6–4.4）	**1.3**（1.1–1.6）	4.5（3.6–5.9）	永不
廣東二類	0.74	3.9	1.5	5.3	永不
台灣	0.53	2.6	1.1	3.6	永不
馬來西亞	1.1	6.9	2.2	9.2	永不
泰國	1.5	11.8	3.0	15	永不
印尼	1.9	21	3.9	25	永不
越南	2.2（1.6–4.5）	29	4.4	34	永不
菲律賓	2.8	>60（運作成本 ≈ 節省）	5.8	≈100	永不
美國	0.46	**0.96**	**1.1**	3.4	永不

FIG 19 ***各地區回本期AI 複判在美國、台灣與深圳數月回本，東南亞 1–3 年；協作臂上下料以年計。***



只為揀出不良板加裝機械手臂：所有地區都不回本
直通率 90% 以上時，每班揀出只佔 0.003–0.14 個人力（見圖 20）

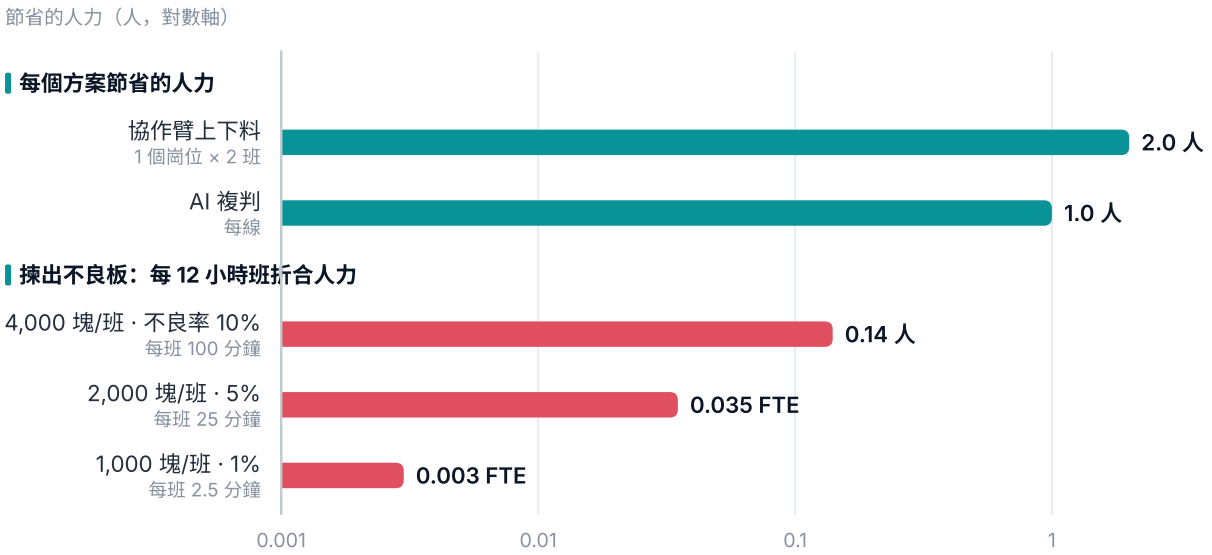
來源： 作者模型：每線人民幣 5 萬 AI 複判盒；協作臂上下料單元人民幣 20 萬或 14 萬美元，兩班制。

所有數值都是以上輸入的估計。敏感度：節省人數減半，每個回本期大致翻倍；只跑一班，(b)–(d) 翻倍；美國首年 100% 折舊縮短回本期約 20%，中國、日本或韓國設備的關稅延長約 10–40%；中國 15% 地方補貼縮短約 15%。我沒找到買方回本期門檻的獨立調查；廠商稱 AOI「通常 9–18 個月」回本（Averroes），Siemens 案例稱八個月 vendor claims。我用一年與兩年作篩選線。模型不計品質收益（保固、退貨）與流動、訓練成本，所以屬保守。

「檢出即揀出」違反直覺之處在於，人力在複判，而非把為數不多的真 NG 板拿下產線。以每塊 NG 板 15 秒處理時間 estimate：

每班板數	缺陷率	每 12 小時班揀出時間	人力
1,000	1%	2.5 分鐘	0.003
2,000	5%	25 分鐘	0.035
4,000	10%	100 分鐘	0.14

FIG 20 ***為什麼揀出不良板不回本*直通率 90% 以上時，揀出每班只佔 0.003–0.14 個人。**



來源： 作者估計。

基準案例（每班 2,000 板、5% 不良、兩班）是 0.07 個人，在深圳每年約人民幣 6,000，低於手臂約人民幣 12,000 的年運作成本，所以任何地方都不回本；即便最差案例（每班 4,000 板、10% 不良）在深圳也要約 13 年、美國 9 年 estimate。若擔心的只是停線揀出，標準 NG 分流器或暫存輸送帶即可解決，無需手臂。業界人士說「直通率 90% 以上，停線手工取出即可」，**這對揀出是對的，但對檢測是錯的**。因此 SMT 線上的產品設計是「檢測加標記或分流」，機械手臂留給每個零件都要搬運的地方（測試治具上下料、托盤與料匣搬運、包裝），在那裡檢測幾乎是免費的加值項，適用情境 (c)。

補貼、稅與關稅讓回本移動 15–25%，但不改變排序

中國大陸：支持多數降低融資成本；單機很少直接符合資格。 一項全國性貸款貼息在 14 條優先鏈上補貼最多兩年 1.5 個百分點利息（Yicai） verified，人民幣 30 萬的機器約人民幣 9,000；廣東最高 20% 的升級獎勵需登記專案、新增設備至少人民幣 700 萬元，且不含深圳（Guangdong rules），東莞 2026 年 7 月實際撥付 4.42%（Dongguan allocation） single source, primary。AI 檢測設備只有作為買方登記專案中「含配套軟體」的設備才獲批，所以基準案例假設無補貼。

台灣：誘因給買方。 依《產業創新條例》第 10-1 條（2025 年 5 月修正納入 AI），買方可把 2025–2029 年新購智慧機械或 AI（含 CNN/DNN 檢測）支出的 5% 抵減當年營所稅，或三年各 3%（Statute、credit rules） verified；供應商端補助對大陸資本關閉（MOEA） single source。

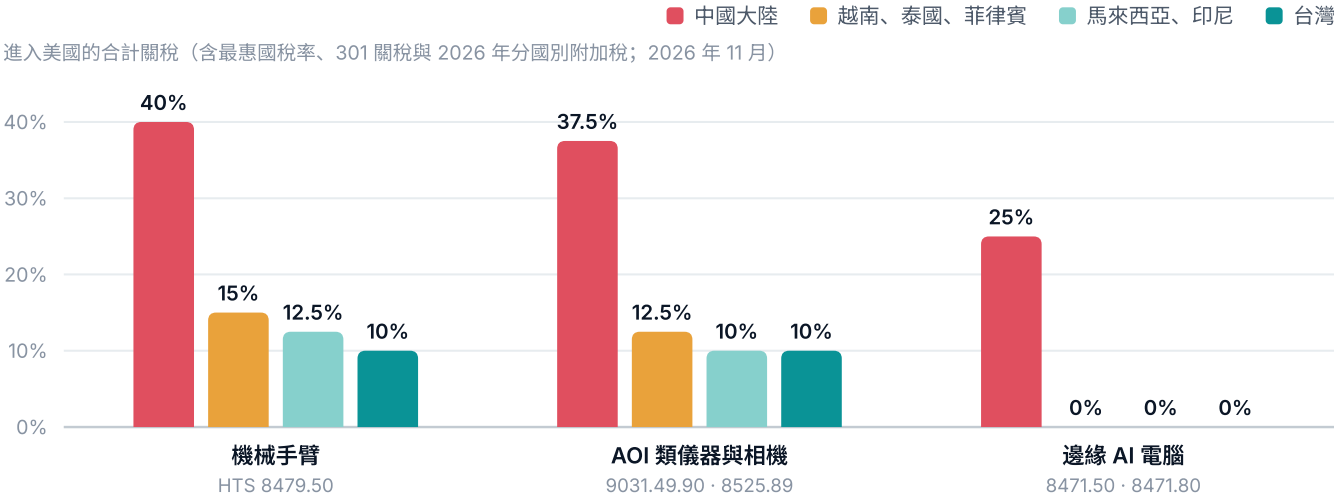
東南亞：只有馬來西亞與泰國獎勵買自動化。 馬來西亞首 RM10 百萬的 200% 自動化資本扣除，對獲利買方約值資本支出的 24%（MIDA、PwC Malaysia） verified；泰國 BOI 給予升級投資 50–100% 的三年免稅，涉及泰國自動化供應商時上限更高（BOI） single source, primary，所以進泰國最好配本地整合商。越南、印尼與菲律賓獎勵的是建廠而非買自動化（Vietnam Law Magazine、PwC Philippines）。

美國：折舊縮短回本；關稅延長。 2025 年 1 月 19 日後購入的資產享永久首年 100% 折舊，對獲利買方約 21% 的聯邦稅盾（IRS Notice 2026-11） single source, primary；機器人與工業機械的 232 調查於 2026 年 10 月 28 日以協商而非新關稅收

場，180 天內重估（White House、Mayer Brown） verified 。東南亞最終組裝可把手臂與 AOI 類儀器的附加稅從約 37.5–40% 降到 10–15%（第 1 章）。

地區	主要工具	對買方回本的影響 estimate
中國大陸	最多兩年 1.5 點貸款貼息；省市升級獎勵（廣東需 ≥人民幣 700 萬登記專案；東莞實際撥付 4.42%）；深圳 2026–2029 措施（費率不公開）	單機約 0–5%；大型專案內 4–15%
台灣	條例 §10-1 抵減，5% 或 3% × 3 年	約縮短 5%，若買方繳稅夠多
馬來西亞	200% 自動化資本扣除（首 RM10M）	獲利買方約縮短 15–25%
泰國	BOI 三年免稅，最高投資的 50–100%	免稅期外的獲利買方約縮短 15–25%
越南、印尼、菲律賓	建廠所得稅誘因；菲律賓設備加碼 20% 折舊	對單筆自動化購買幾乎無效
美國	首年 100% 折舊；大陸製設備合併關稅 ≈37.5–40%	折舊 ≈縮短 20%；關稅 ≈延長 10–40%

FIG 21 **機器人、AOI 與邊緣電腦的美國關稅 *中國大陸製造的手臂約付 40%；從東南亞或台灣出貨 10–15%。***



來源： USITC 協調關稅表第 99 章；聯邦公報；USTR（2026 年 11 月）。

分地區看：中國大陸 AI 複判盒數月回本、協作臂單元 1.3–1.5 年，而只為省人力買的新 AOI 要 3–4 年、檢出即揀出一體化 4.5–5.3 年，都超出兩年篩選線。台灣略優於深圳（複判 0.53 年、協作臂單元 1.1 年），且當地招募更難。馬來西亞與泰國 AI 盒 1–1.5 年、協作臂單元 2–3 年，獲利買家計入本地稅務誘因後快約 15–25%。印尼、越南與菲律賓只有非常便宜的 AI 盒能在 2–3 年內回本；當地需求更多來自客戶的品質要求與人力短缺。美國除了純揀出機械手臂外，每個選項都在 0.2–3.4 年內回本，主要變數是關稅、232 重估與 NRTL 認證。

05

市場與競爭

市場、競爭與買方如何付款

市場規模：公開數字差上數倍；只有自下而上的計數有用

每個市場數字都得看它的口徑。中國機器視覺市場 2024 年為人民幣 181.5 億元（Zhiyan） verified ，3C 電子是最大買家；全球機器視覺估計差約三倍、AOI 差約 6.5 倍（MarketsandMarkets、Vision Systems Design/Interact Analysis、Zhiyan）。自下而上由 Koh Young、TRI、Jutze、Sinic-Tek 與 ViTrox 的營收加總，全球 SMT 檢測設備約在 $1–20 億量級 estimate 。

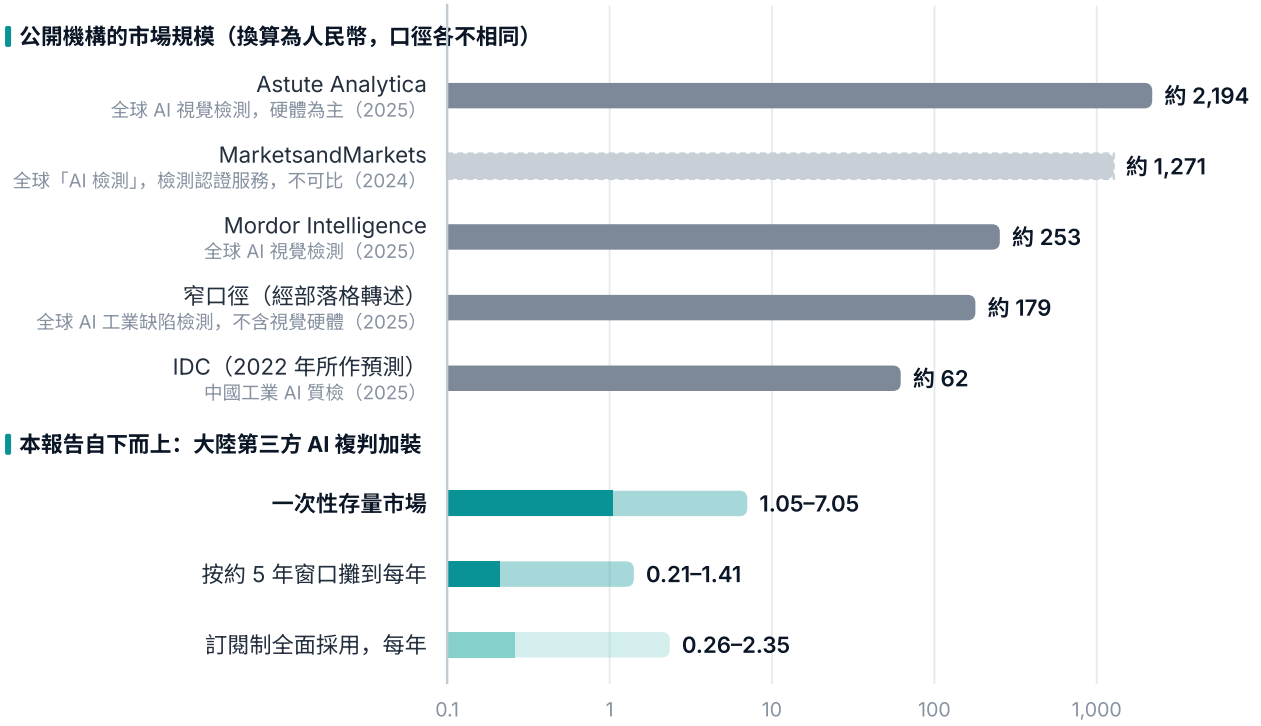
「AI 視覺檢測軟體」沒有獨立市場。公開數字差十倍以上，且都含硬體或服務：

口徑	發布者	數字
全球「AI 工業缺陷檢測」（窄口徑，不含視覺硬體）	未具名公司，經 iFactory 部落格	2025 年 $26.6 億，至 2035 年 CAGR 8.6%，電子占 34%
全球「AI 視覺檢測」	Mordor Intelligence	2025 年 $37.6 億，2031 年 $128 億，CAGR 22.67%
全球「AI 視覺檢測」（以硬體為主）	Astute Analytica	2025 年 ≈$326.6 億
全球「AI 檢測」（檢測認證服務；不可比）	MarketsandMarkets	2024 年 $189.2 億
中國「工業 AI 質檢」	IDC，2022 年所作預測	2025 年 $9.58 億（≈人民幣 62 億），91.5% 在通信、電子與另兩行業

來源：（iFactory、GII/Mordor、GII/Astute、MarketsandMarkets、IDC） each single source, mostly paid-report summaries 。

FIG 22 **市場規模相差十倍以上*只有窄口徑與自下而上計數對決策有用。***

人民幣億元（對數軸；美元按 6.7188 換算）

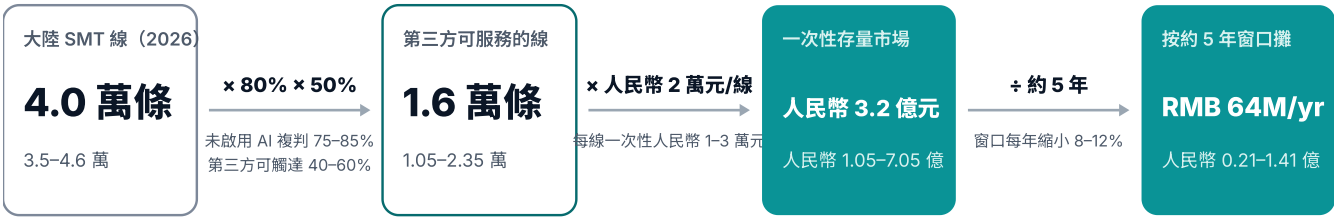


來源： Mordor Intelligence、Astute Analytica、MarketsandMarkets、IDC（2022）、iFactory；作者自下而上估計。

自下而上，SMT 線只有一份公開計數：New Venture Research 估計 2020 年全球約 73,231 條線，47% 在中國（2000 年為 19%），約 70% 高階、30% 低階（EMSNow、NVR SMT2020 summary） single source, 2020；NVR 現歸屬 IPC，新資料需付費（NVR）。以年增 3–5%、中國占 40–47% 推得 2026 年全球約 87,000–98,000 條線、中國大陸 35,000–46,000 條（基準 40,000） estimate。

輸入	低	基準	高	依據
中國大陸 SMT 線（2026）	35,000	40,000	46,000	NVR 2020 基準 × 成長 × 份額
未用 AI 複判的比例	75%	80%	85%	改造窗口，見下
第三方可觸達的比例（非自建、非 OEM AI 綁定、付得起）	40%	50%	60%	假設：≈30% 低階線加部分混牌高階線，排除 DBG 式自建
可服務的線	**10,500**	**16,000**	**23,500**	乘積
每線一次性價格	人民幣 10k	人民幣 20k	人民幣 30k	價格階梯與上限（第 4 章）
一次性可服務市場（裝機存量）	**人民幣 1.05 億**	**人民幣 3.2 億**	**人民幣 7.05 億**	乘積
約五年窗口攤到每年	人民幣 2,100 萬	人民幣 6,400 萬	人民幣 1.41 億	窗口每年縮 8–12%
訂閱（每線每年人民幣 2.5k / 5k / 10k，全面採用）	每年人民幣 2,600 萬	每年人民幣 8,000 萬	每年人民幣 2.35 億	百度 式年費錨點

FIG 23 **自下而上：4 萬條線 → 1.6 萬可服務 → ≈人民幣 3.2 億一次性*中國大陸的第三方 AI 複判利基是真實的，但很小。***



新進入者 3 年拿下可服務線的 2–5%：約 320–800 條線，每年約人民幣 200–500 萬元
訂閱制（每線每年人民幣 0.5 萬元）則為 160–400 萬元/年的經常性收入；這個細分每年只相當於 IDC 中國工業 AI 質檢預測的 0.3–2.3%

來源： NVR（2020 基準）；作者估計。

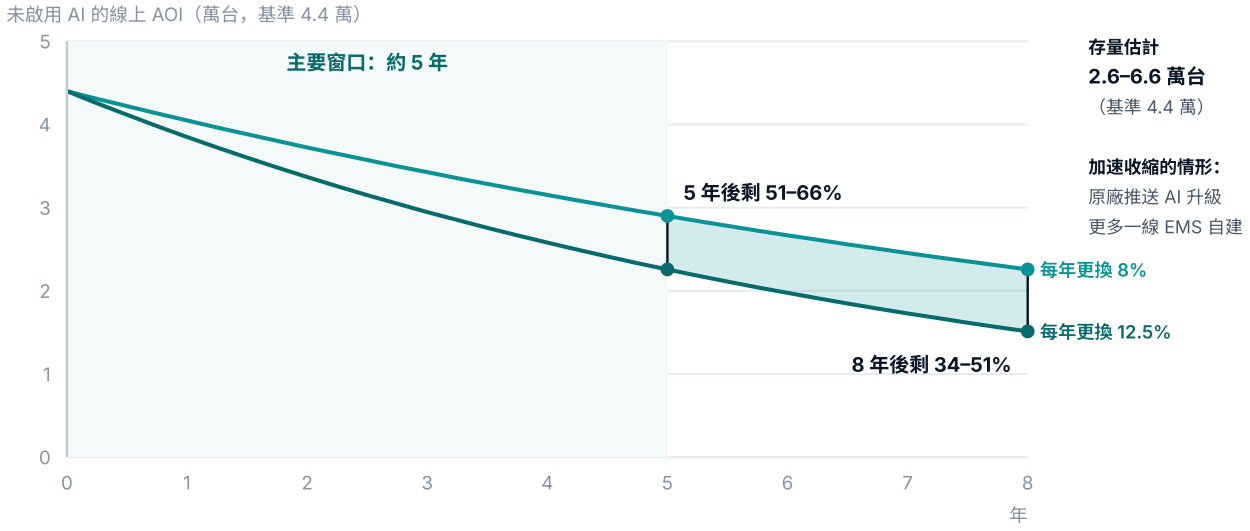
這些都是估計；份額與價格都未公開量測。這個利基每年只相當於 IDC 對 2025 年中國工業 AI 質檢全部預測（約人民幣 62 億）的 0.3–2.3%。新進者三年拿下基準可服務線的 2–5%，約 320–800 條線：每線人民幣 20,000 計共人民幣 640–1,600 萬，每年約人民幣 200–500 萬；每線人民幣 5,000 計的經常性收入每年人民幣 160–400 萬 estimate。這夠一個公司內的業務單元，但不夠一家風險投資規模的獨立公司。台灣、東南亞或美國每 1,000 條裝機線，按大陸價約值一次性人民幣 300–1,500 萬；台灣與美國價格更高，但那裡的買方更常使用帶 AI 選配的 OEM 3D AOI estimate。

改造窗口約五年

沒有 AOI 原廠公布機器壽命；EMS 公司設備折舊 3–10 年（DBG 2025 annual report），舊 AOI 在壽命終止後失去軟體更新與備件（I-Connect007） single source。我用 8–12 年（基準 10）。

以每線 1.0–1.65 台線上 AOI 計，中國大陸在役約 35,000–78,000 台線上 AOI（基準約 55,000），與已知出貨相符（Circuits Assembly、Sinic-Tek 2025 annual report）。沒有來源量測有多少跑 AI 複判；假設國產機自約 2023 年起提供、近交貨的 40–60% 已採用，**中國大陸約 26,000–66,000 台線上 AOI（基準約 44,000）未跑 AI** estimate。若每台新機都帶 AI 出貨，這個池子每年縮 8–12.5%：五年後剩 51–66%、八年後 34–51%，若原廠把 AI 升級推向較新機則更快。**機會大多落在未來五年左右**，核心是跑舊款國產 2D AOI、多品牌混用、無 AI 升級路徑的中型 EMS。

FIG 24 **改造窗口約五年*未啟用 AI 的線上 AOI 在被汰換或升級的過程中每年縮小 8–12.5%。***



來源： 作者估計；裝機存量與份額未獨立核實。

競爭者已經很強：複判楔形市場被合圍

AI 複判楔形市場正從三個方向收窄。AOI 與視覺龍頭正把 AI 內建進新平台：Jutze「全面採用 AI 演算法」（Jutze investor relations, June 2026），OPT 的工業 AI 營收 2025 年成長 96.77%（OPT 2025 annual report） single source, primary ，ViTrox 宣稱自主複判準確率最高 95%（ViTrox） vendor claim 。大型製造商自建（iThome）。而風投支持的 AI 公司大量銷售：SmartMore 邊緣感測器客戶兩年從 65 家增至 251 家（36Kr） single source ，海康威視 販售觀瀾檢測模型（Xinhua），「眼、腦、手」整合已由 SmartMore、Hikrobot 與 Mech-Mind 提供（36Kr、Securities Star）。

FIG 25 **複判楔形市場三面被合圍*第三方的空間在中小 EMS 與 AOI 覆蓋薄弱的檢測利基。***



來源： 矩子、思泰克與光弘年報；海康與思謀公開資料；由作者彙整。

留給新進者的空間很窄：配備舊款 2D/3D AOI、原廠不提供 AI 升級的中型 EMS；混用多個 AOI 品牌的產線（一家改造廠商已支援 Omron、VI、Koh Young 與 Mirtec（Sohu on Hengtianyi））；SMT AOI 覆蓋薄弱的利基，如通孔與手焊焊點、三防漆、線纜、連接器與 FPC，以及整機外觀與螺絲；以及買不起人民幣 23–28 萬 AOI 的小廠。

最接近的上市可比公司給出警訊。銷售 AI 檢測設備與機器人的 Weiyi Zhizao，3C 營收 2026 年 1–5 月年減 79.9%，坦承「競爭加劇與價格壓力」，2026 年 10 月 25 日延後香港上市，並披露 15 個專案在完成前終止，其中 12 個是具身機器人專案（Weiyi Zhizao prospectus、postponement notice） single source, primary 。AInnovation 2025 年底應收帳款提列減值 29.8%（AInnovation 2025 annual report）。3C 檢測楔形市場擁擠，每線的定價權薄弱。

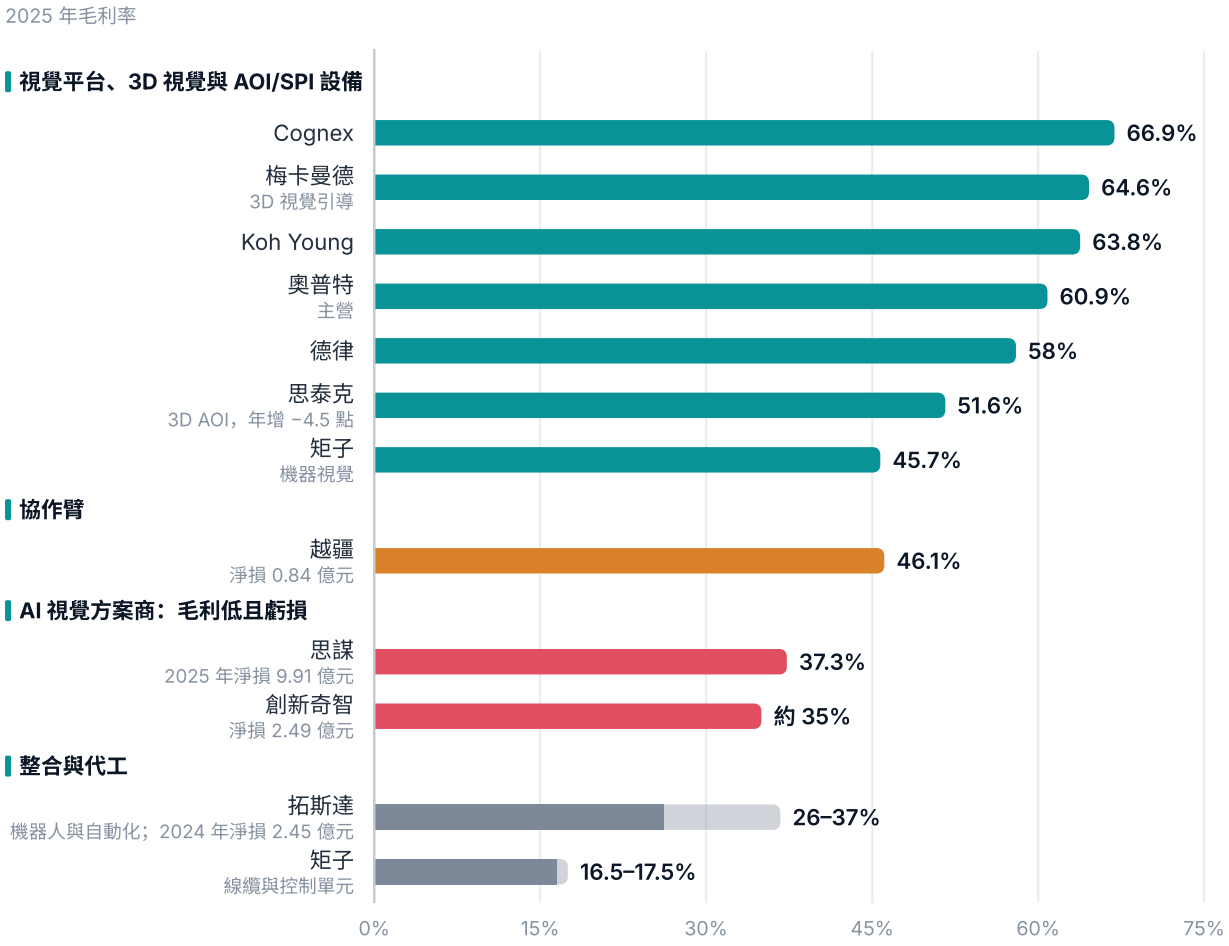
毛利在產品化硬體

價值鏈位置	代表性公司與 2025 年毛利率	獲利與收款
視覺平台	Cognex 66.9%	獲利；剝離 ≈$22M 低毛利營收
AOI/SPI 設備	Koh Young 63.8%；TRI 58%；OPT 主營 60.85%；Sinic-Tek 3D AOI 51.62%（年減 4.53 點）；Jutze 機器視覺 45.74%	獲利；應收天數 139（Jutze）到 247–274（OPT）
3D 視覺引導	Mech-Mind 64.6%	—
AI 視覺方案商	SmartMore 37.3%；AInnovation ≈35%	虧損：SmartMore 2025 年淨損人民幣 9.91 億，AInnovation 人民幣 2.49 億
協作臂	Dobot 46.1%	Dobot 淨損人民幣 8,400 萬；UR 營收從 $326M（2022）到 ≈$232M（2025 預測）

價值鏈位置	代表性公司與 2025 年毛利率	獲利與收款
系統整合	Topstar 機器人與自動化 26.2–36.7%	淨損人民幣 2.45 億（2024）；應收天數 180–240
代工硬體	Jutze 線纜與控制單元 16.5–17.5%	—

來源：（Cognex、StockAnalysis on Koh Young、Fugle on TRI、OPT 2025 annual report、Sinic-Tek 2025 annual report、Jutze 2025 annual report、Jiemian on Mech-Mind、36Kr on SmartMore、Yicai on SmartMore、TMTPost、DoNews on AInnovation、The Paper on Dobot、The Robot Report on UR、Jiemian on Topstar）。

FIG 26 ***毛利在產品化硬體視覺平台與設備商毛利率 46–67%；AI 方案商 35–37% 且虧損。***



來源： 2025 年報或招股書。

新進者只有把檢測做成產品、設備或軟體才能保住毛利；逐案調校把毛利壓到約 35% 並持續虧損。機制也對小公司不利：營收只有在現場驗收後才認列、預付很少（Jutze、OPT）、試點通常免費（A3/Mindtrace），且品牌廠常在代工廠產線上指定檢測設備。

買方如何付款：簽約 30%、驗收後 60%、10% 押一年

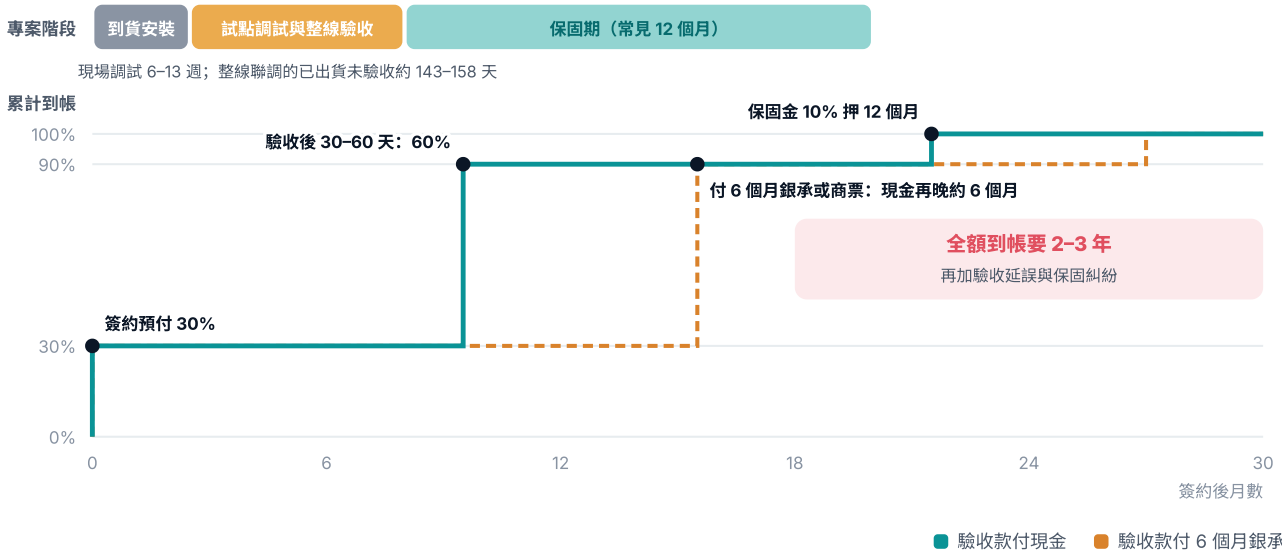
中國大陸電子廠對檢測設備的付款條件，明載於 Sinic-Tek 上市問詢回覆中披露的 13 份具名買方合約（Sinic-Tek inquiry reply 8-1-1） single source: one primary filing, several independent buyers, 2020–2022 ：

買方（年份）	簽約	交貨	驗收	尾款或押款
某 海康威視 單位（2021）	15 天內 30%	—	最終驗收後 30 天內 65%	驗收後 12 個月付 5% 保固押款；部分以 6 個月以上銀承支付

買方（年份）	簽約	交貨	驗收	尾款或押款
南昌 華勤（2020–2021）	10 個工作日內 30%	—	驗收後 60 天 60%	一年保固後 10%
廣東 Yinuo（2021，3D AOI）	10 個工作日內 30%	—	30 天內 60%	無故障運行三個月後 10%
廈門 Hongxin、荊門 Hongyi（2022）	10 個工作日內 30%	—	最終驗收後 30 個工作日內 60%，6 個月票據	一年保固後 10%
AKM 番禺、蘇州（2021–2022）	7–10 個工作日內 30%	—	驗收並開票後一個月 60%，180 天票據	360 天後 10%
BYD 精密、深圳、惠州（2020）	—	到貨 30%	驗收並開票後 70%	全數 6 個月商票
深圳 H&T（2020）	—	—	驗收並開票後 90%	無故障使用六個月後 10%

其他一手申報文件顯示相同結構：Mech-Mind 大型專案簽約 30%、交貨前 30%、驗收後 30%、12–15 個月保固押款 10%（Mech-Mind prospectus）；GKG 稱設備業押款通常為合約的 5–10%，驗收後一年支付（GKG inquiry reply） single sources, primary 。所以慣例是簽約約 30%（20–30%，部分廠如 BYD 為零）、驗收後 60–65%（30–60 天後支付，常為六個月銀承或商票，使可用現金再延半年）、押款 5–10% 押 3–12 個月（12 個月最常見） verified across several contracts 。

FIG 27 ***買方如何付款簽約 30%、驗收後 60%、10% 押一年：全額到帳要 2–3 年。***



來源： 思泰克問詢回覆、梅卡曼德、GKG、JT 等申報；示意為常見條款，時程為估計。

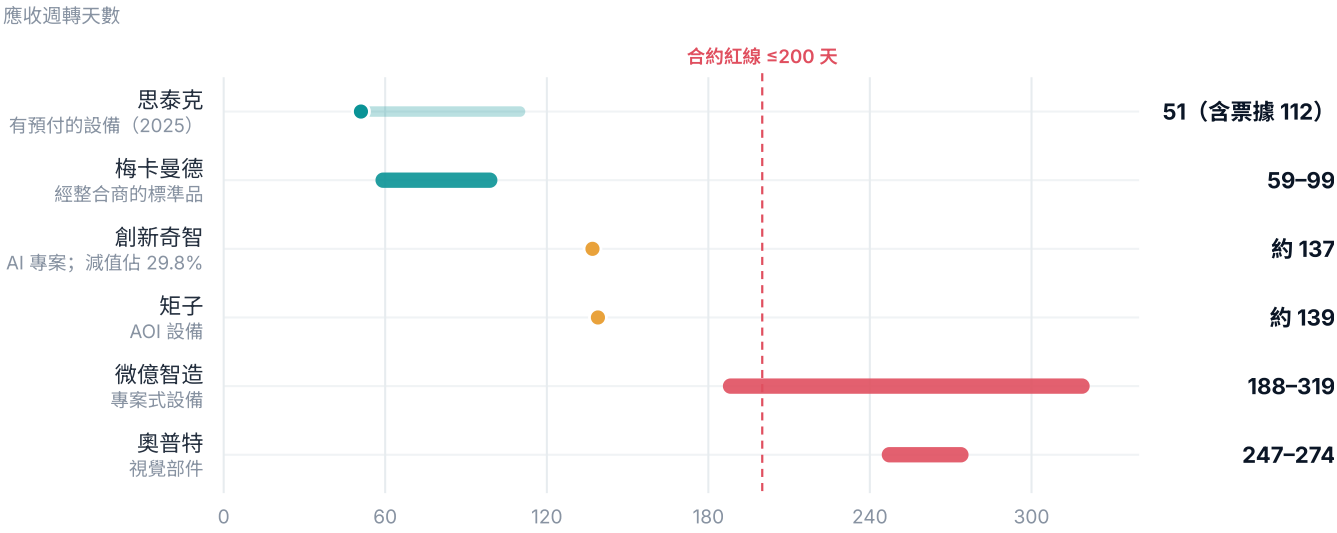
驗收與試車慢，且由賣方承擔。Sinic-Tek 已出貨未驗收的貨物周轉約 143–158 天，因大型 EMS 與品牌客戶需整線聯調與延長試運行（Sinic-Tek inquiry reply 8-1-2）；Weiyi 的 AI 專案從評估到部署要 4–8 個月（Weiyi Zhizao prospectus） single sources, primary 。AI 複判試點沒有轉化資料；Sinic-Tek 只有約 18–19% 的展示機直接轉為銷售。我對免費試點按三到五分之一、付費試點在篩選後約 80% 規劃 estimate 。

公司	商業模式	應收帳款周轉天數
Sinic-Tek（2025）	有訂金的設備	≈51（含票據 ≈112）
Mech-Mind	經整合商的標準品	59–99
Jutze	AOI 設備	≈139

公司	商業模式	應收帳款周轉天數
AInnovation（2025）	AI 專案	年底 ≈137，減值 29.8%
Weiyi Zhizao	專案式設備	188–319
OPT	視覺元件	247–274

來源：上述年報與招股書；應收帳款（含合約資產）÷ 營收 × 365 primary + estimate 。

FIG 28 ***應收週轉天數標準品 51–99 天；專案式設備 188–319 天。***



來源： 公司申報；應收帳款（含合約資產）÷ 營收 × 365。

經 AOI 原廠或整合商銷售的 AI 複判授權，應像 Mech-Mind 的 60–100 天收款；直銷專案像 Weiyi 的 190–320 天。合併成現金時間線：免費試點一到三個月，大型 ODM 資格審查最多約十二個月，建置、交貨與驗收三到七個月；30% 隨訂單到帳，60% 約在訂單後四到九個月，最後 10% 約在訂單後 16–21 個月。**從大型 EMS 首次試點到全額收款可能需要兩到三年**；無正式招標的中型 EMS 可把資格審查縮到數週 estimate 。披露合約中的驗收標準只涵蓋整線聯調、產能（一個台灣專案要求每小時 100 台）與數月無故障運行；沒有一份指明漏失或誤報率（Anda inquiry reply） single source, primary 。AI 複判的書面驗收標準必須在試點前議定。

四種失敗模式

有紀錄的失敗歸為四類：產品未可靠就出貨（Rethink（The Robot Report））；在龍頭硬體上做的軟體層，後來被龍頭生態系取代（READY Robotics（The Robot Report））；在週期底部燒光錢（Dextrous Robotics（Robots & Startups））；以及模型公司以收購式挖角收場（Covariant（GeekWire））。教訓：窄產品化而非逐案調校；為驗收制營收與長收款預留營運資金；為 24 小時運轉的工廠提供本地服務；產品未驗證前不擴張。

時機：檢測等下去沒有好處；手臂硬體會變便宜

等待的理由薄弱。2026 年硬體並未變便宜：機器人廠漲價 5–15%、部分產品 25–50%（Sina Finance） single source ，NVIDIA 套件第三次調漲（36Kr）。量產級 VLA 不會在窗口內到來（IFR、Instrumental）。同時每換一台 AOI 就少一個改造標的，每部署一個 OEM AI 就鎖住一條線的資料與流程；模型能力正在商品化，稀缺的是工廠接入權與標註過的工廠資料。本地支援也有窗口：深圳 2026 年「AI 先鋒城市」政策，AI 應用示範專案最高人民幣 200 萬、模型券補貼採購 30%，實施至 **2027 年 6 月 30 日**（Yicai, May 2026） terms single source 。

仍在快速變便宜的是人形機器人與靈巧手（手兩年從約人民幣 5 萬降到 1 萬以下（Sina Finance）），以及產品組合下移的入門協作臂；這支持推遲具身部分，而非檢測部分。**總結：檢測現在做，作為窄產品；機器人單元現在做，透過夥伴與整合；具身智慧以後做，今天只設一個研發站。**

具身智慧

06 具身智慧：先改變機器人怎麼被教，之後才改變它們怎麼動

正在變的是感知與程式；運動保持確定性

每個機械手臂龍頭現在都在行銷「實體 AI」，但已出貨的分三層：無需教學的學習型感知（FANUC 的 AI Masterless Vision，2027 年 1 月出貨（FANUC）；KUKA 的神經抓取選擇（KUKA））；寫程式與規劃任務的大模型（KUKA 的 iiQWorks.Copilot，警告 AI 生成的程式碼應在數位孿生中驗證（KUKA）；Yaskawa 用 Gemini Robotics-ER，由「生成式 AI 決定做什麼」、機器人決定「怎麼動」（Yaskawa））；以及算力與資料平台（UR 帶 Jetson Orin 的 AI Accelerator（Universal Robots）） vendor claims 。沒有龍頭公布其 AI 功能的量產成功率，且日本龍頭與 Google、SoftBank、Microsoft 等合作而非自建前沿 VLA（FANUC news index）：學習型策略層將來自模型公司，認證手臂與控制器來自手臂原廠。

FIG 29 ***具身智慧堆疊大模型離線規劃，小型策略在單元內執行，運動與安全保持確定性。***



小企業的機會在整合與感測硬體：把學習型感知和小策略做進經認證的機器人單元，逐個工位落地

來源： 作者綜合整理（第 3 與 6 章）；ISO 10218:2025；歐盟機械法規。

離非結構化、彈性、小批量還有多遠：結構化工位 1–3 年，彈性組裝 5 年以上

真實工廠數字開始出現，但都是廠商自報、範圍窄、且來自結構化工位。AgiBot 在 Longcheer 平板最終測試線上的八台人形機器人，超過 64 小時記錄 64,828 個任務、成功率 99.99%（AgiBot、Yicai Global） vendor and customer ，但這是帶視覺的傳統手臂就能做的上下料工作。UBTech 稱其 Walker S2 工作效率為人工的 30–50%（Irish Times/FT） single source 。Figure 02 在 BMW Spartanburg 的 11 個月產出數量（1,250+ 小時、90,000+ 零件），而其節拍、成功與干預數字是目標而非結果（Figure、BMW Group） verified, numerators only 。Nucleus 稱其工廠影片約 40% 是遙操作（Interesting Engineering） vendor claim ，Reuters 調查稱部分中國人形機器人約為一名工人產出的 20%（Notebookcheck） single source 。沒有部署同時公布成功率、節拍與干預率，也沒有一個經過稽核，而這三者正是回本計算所需的。

部署	成功率	節拍	干預或遙操作	開機率或故障間隔	量產運行
Figure 02 @ BMW Spartanburg	僅目標（每班 >99%）	僅目標（84 秒）	僅目標（每班 0 次）	未揭露（前臂頂部為故障點）	1,250 小時
Agility Digit @ GXO	未揭露	未揭露	未揭露	未揭露	100,000 個周轉箱
Apptronik Apollo @ Mercedes-Benz	未揭露	未揭露	未揭露	未揭露	未揭露
Agile ONE @ Audi	未揭露	未揭露	未揭露	未揭露	未揭露
Nucleus（未具名工廠）	未揭露	未揭露	≈40% 遙操作（方法未揭露）	未揭露	≈2 小時影片
AgiBot G2 @ Longcheer	64,828 個任務的 99.99%	每步 18–20 秒	未揭露	一次通信故障，約 2 分鐘恢復	64+ 小時
Neuromeka 混合工位	322/324	比人慢 12.8%	未揭露	未揭露	324 個循環
Physical Intelligence π0.6 客戶現場	—	—	一班 96.4% 自主	未揭露	一班

來源：本章與第 3 章 mostly vendor claims 。

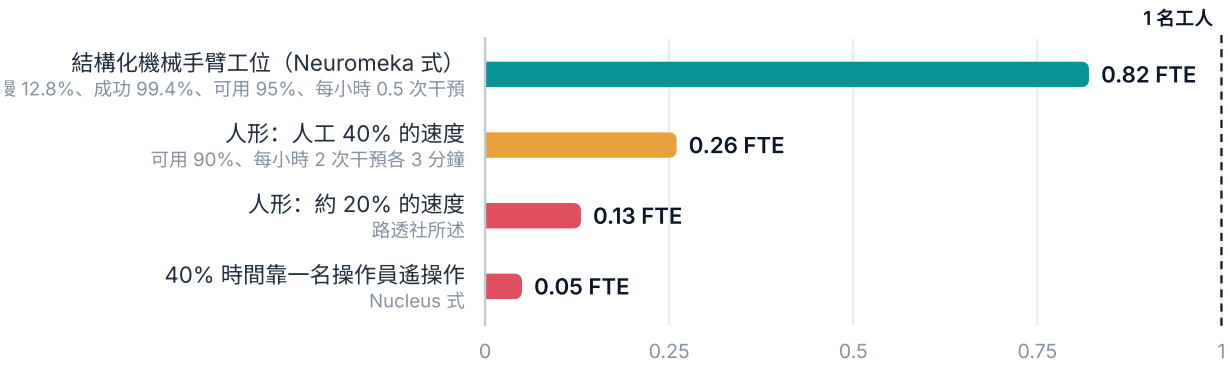
FIG 30 **廠內機器人部署實際上揭露了什麼*多數只揭露數量；沒有一個同時揭露成功率、節拍與干預率。***



來源： BMW、Figure、Agility、賓士、奧迪、智元、Neuromeka、Physical Intelligence 等公開資料。

遙操作對經濟性的影響大於成功率。每台機器人每班淨替代人力 ≈（人工節拍 ÷ 機器人節拍）× 開機率 × 成功率，減去干預時間與遙操作 estimate, my formula 。Neuromeka 式結構化工位（慢 12.8%、成功率 99.4%、開機率 95%、每小時 0.5 次各 2 分鐘的干預）約替代 0.82 人；一個人工速度 40%、開機率 90%、每小時兩次各 3 分鐘干預的人形約 0.26；Reuters 的 20% 約 0.13；而若像 Nucleus 由一名操作員 40% 時間遙操作，約 0.05，幾乎為零。以每台人形約人民幣 60 萬、大陸每人每年人民幣 85k 計，後三者要數十年才回本；只有結構化手臂工位有現實的回本（第 4 章） estimate 。

FIG 31 **每台機器人淨替代的人力*結構化手臂單元替代約 0.82 名工人；40% 時間遙操作的人形約 0.05。***



來源： 作者基於公式的估計。

結構化工位之外，通用模型在新任務上仍只 40–90%：Gemini Robotics 2 在實驗室取放達 74.2%、工具配料 78.9%、綁袋 44%（eesel/DeepMind）；在 30 個真機任務的 RoboChallenge Table30 上，π0.5 得 43.7%、最佳開源模型 50.3%（RoboChallenge、QbitAI）。

外部預測不一，但沒有一份支持兩年內普及。IFR 稱通用實體 AI「尚未達工業級」、多數人形機器人「仍是原型或試點」（IFR）；Interact Analysis 估計 2025 年生產的人形機器人僅約 10% 進入實際使用，商業轉折點「在 2032 年之後」（automation.com）；UBS 認為人形機器人的「電動車時刻」2030 年前不太可能（Yicai Global）；中國業內估計工廠級使用還需兩到三年（Gasgoo） single sources 。我的判斷 estimate ：在資源充足、廠商駐廠的電子與汽車廠，**結構化工位任務**（上下料、OK/NG 分揀、有治具插接）將在約 1–3 年內以接近人工速度達到 ≥99%；**混料配料**用通用 VLA 約 3–5 年；**自由佈線的線束與多步驟彈性組裝**則 5 年以上。人形機器人回本今天不成立：一台 Walker S2 約人民幣 60 萬（10jqka） single source ，兩班制下人工效率的 30–50%，約等於 0.6–1.0 名工人，在大陸人力成本下，未計整合、維運與停機前的簡單回本期即為 7–12 年 estimate 。

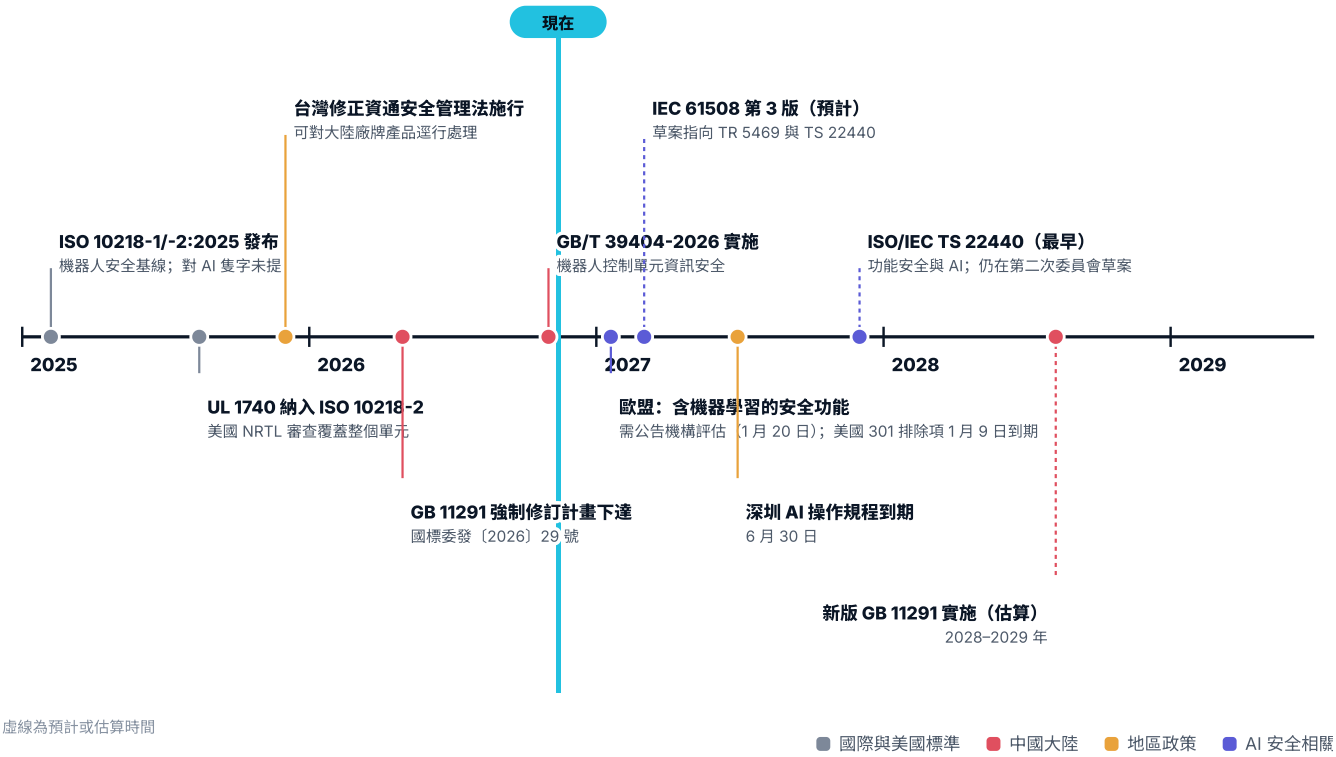
卡在哪裡

環節	2026 年的證據
資料	業界估計至少需 1,000 萬小時高品質真機資料，2026 年初僅收集約 50 萬小時（Gasgoo）；Reuters 報導約 50 萬小時對上所需約 1 億小時（Notebookcheck）；新手遙操作員每個可用動作約 300 次嘗試；工廠不共享製程資料；模擬訓練的策略在真機上急劇下滑
可靠度	結構化工位 98–99.99%（廠商說法，故障定義不明）；通用模型新任務 40–90%；未見過連接器 29–67%
速度	人形為人工速度的 30–50%（UBTech），部分約 20%（Reuters）；前沿代理每個準靜態真機任務約 280 秒；大語言模型控制環 0.2–0.4 Hz，實際需要約 83 Hz（Anthropic）
成本	Walker S2 均價約人民幣 60 萬；回本期 7–12 年 estimate
安全認證	ISO 10218-1/-2:2025 使功能安全更明確並吸收協作要求，但對 AI 隻字未提（The Robot Report）；ISO/IEC TS 22440（功能安全與 AI）2026 年 11 月仍是第二次委員會草案（ISS、ISO Open Data）；IEC 61508-3 指引建議 SIL 1 以上勿用 AI（Analog Devices）；歐盟機械法規下，含機器學習的安全功能自 2027 年 1 月 20 日起需公告機構評估（EUR-Lex）
手與工具	靈巧手每指約 30 N（≈3 kg），人類手約 10 kg（Gasgoo）；換工具仍是人工
整合	現場的感測器誤讀、協定不匹配與 MES 連結；干預率幾乎從不揭露

安全含義很直接：標準既不禁用學習型策略，也不承認它們。ISO 10218-1/-2:2025 對 AI 隻字未提；ISO/IEC TS 22440（功能安全與 AI）仍在委員會草案階段，2027 年底或 2028 年前不太可能出爐 estimate ；ISO/IEC TR 5469:2024 只是指引，建議非 AI 後備、監督約束與冗餘投票；IEC 61508 第 3 版預計 2027 年初（ISO Open Data、BSI、61508 Association） verified 。目前還沒有已發布的標準提供一條公認途徑，把機器學習感知認證為安全功能。區域法規也在動：中國正修訂 GB 11291（約 2028–2029 施行）（revision plan 20262543）；UL 1740 自 2025 年 8 月起把 ISO 10218-2 納

入美國 NRTL 審查；歐盟自 2027 年 1 月 20 日起要求含機器學習的安全功能公告機構評估；而在 ISO 25785-1 仍屬草案期間，沒有人形機器人取得無圍欄工廠使用認證（i-SCOOP）。站得住的架構：機器學習只做製程決策（「哪個零件」、「這塊板好不好」），由經認證的確定性安全層（通常 PL d / Category 3）與安全 PLC 決定速度、力與停止。任何「無圍欄安全的相機 AI」說法，在 PL d 或 SIL 2 證書指名機器學習元件之前，都應視為未經證實。

FIG 32 ***標準與法規時間線安全功能內的 AI 認證路徑在 2027–2029 年前不會底定。***



來源： ISO、IEC、UL、EUR-Lex、SAC 與各國法規資料庫。

機會在哪一層：模型歸巨頭、邊緣盒歸晶片商、小公司歸整合與感測硬體

模型層需要資本與資料，歸巨頭（AgiBot 以「數萬小時」訓練 GO-2（iFeng Tech）；最強的模型閉源或僅雲端）。邊緣盒層歸晶片商：Jetson Thor 是事實標準，大陸則有 D-Robotics 的 RDK 板（D-Robotics）與 Rockchip 協處理器（CSDN）競爭，多半未公布 VLA 成績。

小公司有兩個切入點。一個是直接買可靠度的感測硬體：觸覺把 ContactDP 的插接成功率從 58.7% 拉到 97.8%，力/力矩感測器與腕部相機做類似工作；資料收集套件也好賣（10jqka），儘管手、感測器與手臂都在降價。另一個是手臂原廠與模型公司之間的整合層，掌握每個工位的任務資料、安全包絡與 MES 連結，這是手臂原廠今天不作為產品出售的。競爭者分四群：手臂龍頭（FANUC、Yaskawa、KUKA、ABB、UR、JAKA、Dobot 等）、視覺加手臂公司（Mech-Mind、Hikrobot、SmartMore）、具身 AI 公司（AgiBot、UBTech、Unitree、Galbot、Figure、Physical Intelligence、Google DeepMind、NVIDIA）與車用晶片新進者。

買方驗收：沒有分母就不算數

由於沒有廠商公布干預率、故障間隔或長期成功數，這些應寫進合約驗收條款。統計上，3,000 次連續量產循環零失敗，以 95% 信心證明失效率至多 0.1%；3,000 中 10 次失敗只證明 ≤0.57%；零失敗下證明 ≤1%、≤0.5%、≤0.1% 與 ≤0.01% 分別需 299、598、2,995 與 29,956 個循環 estimate, exact binomial。按此標準，Neuromeka 的 322/324 有 95% 雙側區間 97.8–99.9%，不足以證明 ≥99%；AgiBot 的 64,828 個任務約 6 次失敗隱含 99.98–100%，但僅當「失敗」按買方定義 estimate。以下是我會對任何機械手臂或具身單元提出的驗收要求；門檻借鑒 Figure 在 BMW 的自身目標與 Neuromeka、AgiBot 案例，屬判斷而非標準 estimate：

項目	要求	建議通過門檻
定義	簽署循環、成功、失敗（部分成功、重新抓取、掉落、品質漏失）、干預（觸碰、遠端遙操作、重置、急停）與自主性如何計算的定義	定義先於任何指標；沒有公式就沒有百分比
干預與遙操作	每機器小時、按類型與時長；平均干預間隔；遙操作分鐘占運行時間比例；每台機器人的操作員數	結構化工位每機器小時 ≤0.5 次干預（間隔 ≥2 小時）；驗收期間零遙操作
長期成功	≥3 班、≥2 個料號、正常光源與供料變異下 ≥3,000 次連續量產循環；失敗分類；精確下限	停線線上工位下限 ≥99.5%、離線 ≥99.0%；品質漏失單獨報告
節拍	每個零件 P50/P90/P99（含恢復），對比同工位 ≥30 次人工計時循環	≤1.15 × 人工 P50，或在產線節拍內
開機率與故障	運行小時、停機日誌、按元件故障、換件時間	≥2 週開機率 ≥95%；備件在場時 MTTR ≤20 分鐘
恢復	腳本化故障注入：掉落、遮擋、光源變化、傳送帶停止、MES 或斷網、斷電、急停	≥90% 自主恢復，2 分鐘內回到生產
換型	每個新料號或治具的小時數、示範數、操作員技能與資料收集成本	結構化工位：每個新料號 ≤1 班且 ≤100 次示範
安全	手臂與單元依 ISO 10218-1/-2:2025、單元風險評估、本地法規（大陸 GB 11291；台灣 CNS 14490 加書面評估；美國 ANSI/A3 R15.06-2025 與 UL 1740）	安全功能不得由機器學習實作；經認證的安全層（通常 PL d / Cat 3）限制速度與力；ISO 25785-1 發布並評估前無無圍欄足式機器人
算力與離線行為	推論硬體、P99 延遲、無網路行為、模型版本化、每次更新後重新驗證	節拍關鍵推論不依賴雲端；任何模型、量化或引擎變更後全面重新驗收
證據	整個驗收窗口的未剪輯影片與原始日誌；由客戶品質團隊或第三方見證	沒有精華剪輯或廠商自摘
商業條款	付款里程碑綁定驗收指標；按機器小時計費時的可用性罰則；資料所有權與遠端存取規則	尾款與押款僅在 3,000 循環驗收與 ≥4 週穩定運行後釋放
經濟性	淨替代人力 = 速度比 × 開機率 × 成功率 − 干預時間 − 遙操作占比	拒絕任何淨替代無法支撐其宣稱回本的提案

來源：（Figure、Neuromeka、Yicai Global、i-SCOOP）；門檻為我的建議。

07 待解問題
公開紀錄無法回答的問題，以及試點如何釐清

公開來源無法回答的問題分五群：AI 複判實際賣多少、買方願付多少；AOI 原廠如何搭售 AI 並升級舊機；在役 AOI 機齡多老、有多少已跑 AI；每線複判實際用多少人力與工時；以及試點如何轉化、買方如何付款。此外還有最新模型、量化版與機械手臂在特定工廠條件下的實際表現。每一項都能變成一個有明確門檻的測試。以下門檻是我從本報告證據推導的決策規則 estimate；首批工廠資料一到就重算。

待解問題	如何釐清	好的結果長什麼樣
AI 複判的價格與付費意願	一份標準詢價（例：四條線，混用舊 2D 與新 3D AOI、兩班、每班約 2,000 板、12 類缺陷、僅本地、MES 或 IPC-CFX 整合）發給多家第三方廠商；對 20–30 家 EMS 廠做四題價格敏感度調查	多數報價在四條線一次性人民幣 40–120k 或每年 20–40k；至少半數廠接受每線人民幣 10–30k、12 個月內回本
AOI 原廠的 AI 搭售與舊機升級	對約十幾家 AOI 原廠與經銷商做短結構化訪談	2025–26 機型的 AI 複判是標配、選配還是缺席；2020 年前的機器能否升級、價格多少
在役 AOI 機齡與跑 AI 的比例	對約 30 家 EMS 廠做現場調查，按規模與地區分層，在保密協定下審設備清單	中位 AOI 機齡 ≥5 年、≥60% 未用 AI，則改造池真實存在
複判人力與工時	對兩家以上工廠做訪談與 2–4 週的 AOI 警報、複判、維修與班別日誌，加工作抽樣	每線每班 ≥1 名專職複判員、每線每班 ≥3 複判工時，且廠方願調度每線至少 0.5 FTE
AI 複判本身的表現	金標準集加真實產線影子模式（第 3 章）：第 0 週資料協議，第 1–2 週基線與標註，第 2–4 週影子運行，第 4–8 週選配輔助運行	誤報或複判工時削減 ≥50–60%；逐類漏失界限不劣於人工基線；關鍵類別零漏檢；p95 延遲 ≤½ 節拍；自動失效即放行給人
前沿模型的老師標籤品質	同一金標準集上跑兩到三個候選，各三種提示措辭、一張好參考圖、≥42–48 px 的單元件裁剪圖	僅對 ≥299 個缺陷零漏檢且達人工級一致性的類別自動採納；否則僅預標註與分診
量化版	只改精度；鎖定權重雜湊、模板、前處理、解碼、引擎與批次大小；停用或重置前綴快取；檢查參考版重現廠商結果；在金標準集與漂移切片上成對評分	關鍵類別無新漏檢；1,000 缺陷中 ≤1 個新漏檢；清除率下降 ≤2 點；三次運行 ≥99.5% 決策一致；否則退回 FP8/INT8
產線側硬體、潔淨與靜電	每個候選盒 ≥2 台，以兩倍預期警報率壓力測試；真實機櫃氣流下 72 小時浸泡；客戶靜電協調員查核；ISO 7 及更潔淨客戶做 ISO 14644-14 測試	p95 ≤½ 節拍；72 小時無當機或降頻；靜電簽核；正確等級的微粒報告
付款條件、試點規則與轉化	受訪工廠近期檢測設備合約的付款條款；試點規則、驗收標準與資格審查時間訪談	約 30/60/10、驗收後 60 天內現金；試點 ≤8–12 週並有具名簽核負責人；≥1/3 試點轉化；首年框架覆蓋 ≥5–10 條線
資料、IT 與責任	訪談與法務審查合約	本地盒可連線並存取資料；責任限於軟體缺陷
出海前的區域合規	台灣：客戶在資通安全管理法下的書面身分、E187 自評；美國：單元分類的海關裁定、NRTL 報價、ITAR 查核；東南亞：BOI 本地連結與馬來西亞扣除資格、客戶母公司；大陸：捆綁 PC 與電源的 CCC 範圍	客戶非受控機構，且含關稅與認證的回本期 ≤2 年
機器人或具身單元的驗收	第 6 章的 12 項驗收表：簽署定義、≥3 班與 ≥2 料號的 ≥3,000 次連續量產循環、腳本化故障注入、≥30 次人工計時循環、見證	線上下限 ≥99.5%（離線 ≥99.0%）；節拍 ≤1.15 × 人工；每小時 ≤0.5 次干預且零遙操作；開機率 ≥95%；MTTR ≤20 分鐘；恢復 ≥90%

待解問題	如何釐清	好的結果長什麼樣
中國邊緣晶片上的 ACT 與 VLA 策略	四週：移植與微基準（≥1,000 次推論、延遲分布、功耗、60 分鐘浸泡）；注入量測延遲的閉環模擬；對 GPU 的真機 A/B；8–24 小時耐久（≥3,000 循環）	ACT 塊 p95 ≤100 ms、VLA ≤300 ms；真機成功率與 GPU 相差 3 點內、節拍 ≤1.2 × 人工；否則策略留 Jetson 或 GPU
協作臂、AMR 與整合價格	多家協作臂經銷商逐型號報價，及兩三家整合商對測試治具上下料（一工位、兩班）的整單元報價	整單元 ≤人民幣 200k，深圳兩班回本期 ≤1.5 年
台灣、東南亞與美國的 SMT 線數	IPC（原 New Venture Research）資料或本地設備經銷商裝機存量	每區可服務線數足以支撐本地應用工程

其中四項決定本報告最敏感的數字：買方願付的價格、AI 複判實際釋放的人力、AI 複判是否真的不比人漏檢更多，以及現金多快回流。無論是考慮採購的工廠，還是考慮產品的供應商，都值得在重大投入前先釐清。

對雙方都管用的試點順序

對工廠，與對想要可信參考案例的廠商，順序相同：

1. **第 0–2 週：議定資料與定義。** 簽資料協議、定義缺陷類別與何謂漏失，並記錄基線：每板誤報數、每次複判秒數、下游發現的漏失。
2. **第 1–4 週：建立金標準集**，取自 AOI 存檔與保留的 NG 板，盲標並裁決（第 3 章）。
3. **第 2–4 週（或更久）：影子模式。** 模型記錄其決策，人照舊決策；逐類比較。
4. **第 4–8 週：隨機班別輔助運行**，模型只清除通過門檻的類別，其餘全交人工。
5. **擴展前書面驗收：** 逐類結果、延遲、失效即放行行為，以及模型與標籤歸誰。逐線擴展，並在每個新產品、模型版本、量化或引擎變更時重新驗證。

本報告的限制

這是案頭研究。本版我未對工廠做一手訪談；大量產線資料來自廠商且未經獨立稽核。數個政府網站、招標資料庫與招聘平台無法存取，因此深圳 2026 年升級補貼費率、長江三角洲工資指引、AI 複判成交價、每線複判員數與試點轉化率都是估計或類比；每一項都以上述待解問題形式呈現。匯率為中國人民銀行 2026 年 11 月 9 日中間價（1 美元 = 人民幣 6.7188），台灣為 SMBC 同日定價（1 美元 = 新台幣 32.12）；越南盾用 2026 年 5 月資料。所有門檻都是從所引證據推導的決策規則，不是標準。

關於作者

Tan Shuai

Tan Shuai 是一位技術人與創辦人，擁有二十多年橫跨晶片、硬體與系統軟體的經驗，並持有 20 項已授權專利。他的工作足跡遍及矽谷、紐約、芝加哥、台灣、大灣區（深圳與香港）、越南與馬來西亞，長期專注於先進製造與供應鏈。他目前擔任 AI 研究員與開發者，專注於 AI 在工廠現場的實際應用。

無論是組建團隊、尋找夥伴還是交流心得，他都歡迎與從事製造業 AI 的人交流：i@tanshuai.com · tanshuai.com

註記與免責聲明

本報告是基於截至 2026 年 11 月 11 日公開文件的獨立案頭研究，僅供參考，不構成投資、法律、稅務或工程建議。產品名稱、價格、模型發布、出口分類、關稅與法規變化很快；任何賴以決策的資訊請以一手來源為準核實。標記 vendor claim 或 estimate 的數字應照此理解。公司與產品名稱歸其所有權人所有，僅供識別之用；沒有任何公司審閱或贊助本報告。

© 2026 Tan Shuai。你可註明出處後原樣分享本報告。

電子製造 · AI 檢測 × 機器人

電子工廠的 AI 視覺檢測與機械手臂：2026 年評估

每個關鍵數字都附有其來源與可信度標籤：verified、single source、vendor claim 或 estimate。公開紀錄無法回答的問題，都收錄在第 7 章，並說明試點如何逐一釐清。

Tan Shuai
tanshuai.comi@tanshuai.com

公開版 · 資料截至 2026 年 11 月 11 日

