

ASSESSING INDUSTRIAL DEPLOYMENT READINESS OF HUMANOID ROBOTS

ZiRan Liu

International Department, Hefei No. 8 Senior High School, Hefei 230011, Anhui, China.

Abstract: Humanoid robots are increasingly tested in manufacturing and logistics, but technical demonstrations and shipment growth give an incomplete picture of deployment readiness. This paper examines the evidence needed to connect embodied capabilities with repeatable industrial operation. A focused narrative review draws on primary robotics studies, official industrial statistics, standards, and customer disclosures available through September 2026. It distinguishes demonstrated capability, task repeatability, integrated operation, and deployment transfer. The paper then develops an autonomy-adjusted cost model linking utilization, availability, relative task rate, autonomous first-pass yield, and supervision cost. A hypothetical sensitivity analysis shows how these factors jointly determine cost per autonomously accepted task. Under the stated assumptions, raising availability from 0.80 to 0.95 reduces the normalized cost from 1.111 to 0.936. Raising autonomous first-pass yield from 0.90 to 0.99 alone leaves the ratio at 1.010. These calculations describe model behavior and do not estimate commercial performance. Published factory disclosures show repeated execution of defined tasks but leave intervention rates and comparable operating costs unresolved. The paper provides a framework for interpreting deployment evidence and a proposed field evaluation protocol. The protocol is designed to test whether humanoid systems can sustain acceptable output across shifts, task variants, and sites.

Keywords: Humanoid robots; Embodied intelligence; Industrial deployment; Robot evaluation; Autonomous operation; Task cost

1 INTRODUCTION

Humanoid robots connect advances in learning and control with workplaces designed around human reach, tools, and movement. Their value in industry depends on whether this fit supports useful work under operating constraints. Embodied intelligence concerns the coupling of perception, action, and the physical environment [1]. A humanoid body is one possible platform for this coupling. Its shape alone does not establish autonomy, generalization, or economic value. A deployment assessment must therefore link the chosen body form to a defined task and a credible alternative.

Industrial policy and growing robot markets make this question relevant, but do not answer it. China's 2023 humanoid robotics guidance set milestones for component development, production, applications, and industrial ecosystems [2]. These milestones are policy objectives. Separately, the International Federation of Robotics reported approximately 542,000 industrial robot installations worldwide in 2024 [3]. This total cannot be treated as humanoid shipments. Application categories and body form describe different aspects of a robot. Neither installation counts nor component supply capacity directly measures productive autonomy.

Evaluation research already calls for tests of robot capability in a defined context [4]. Recent learning systems make deployment harder to assess because performance can depend on training coverage, embodiment, task sequencing, and human support. A robot may complete a selected demonstration but need extensive resets between attempts. A system with limited capabilities may still provide useful service when its task distribution is stable and recovery procedures work well. The research question is which operational evidence connects a demonstrated humanoid capability to repeatable industrial deployment.

This paper examines that connection from an engineering perspective. It organizes existing evidence by the claims it supports, links technical and organizational constraints to an autonomy-adjusted cost measure, and proposes a field evaluation protocol. The contribution combines established evaluation and production concepts for operational use. It does not introduce a new control algorithm or estimate market readiness from empirical data. The analysis focuses on manufacturing and logistics within industrial sites, where tasks and acceptance criteria can be specified. Domestic care, clinical assistance, and open public interaction need additional evaluation conditions and fall outside the scope.

2 EVIDENCE SCOPE AND ANALYTICAL METHOD

2.1 Source Selection and Interpretation

The study uses a narrative review with sources selected for the deployment question. Sources were retrieved from official proceedings, author manuscripts, institutional publications, standards pages, and primary company disclosures. Verification was completed on 14 September 2026. Technical searches covered vision-language-action models, robot learning across embodiments, humanoid teleoperation, and locomotion with manipulation. Industrial searches covered humanoid factory trials, deployment agreements, robot statistics, and Chinese policy. Foundational studies of embodied

intelligence and manufacturing performance provide the conceptual basis. The study is not an exhaustive systematic review.

Technical papers were included when they described a physical robot evaluation or a directly relevant evaluation method. Each paper was read with attention to its platform, task, human involvement, and test setting. Industrial statements were checked against the original issuing organization. They were then grouped as observed activity, contractual arrangements, or future plans. Unresolved market estimates and component localization percentages were excluded from the analytical model. Success rates were not pooled across studies because their tasks, denominators, and evaluation conditions differ. Table 1 summarizes the technical evidence used to identify deployment constraints.

Table 1 Selected Technical Evidence and Deployment Questions

Evidence	Contribution	Deployment question
VLA models [5], [6]	Semantic action learning	Control rate and unseen physical skills
Cross-robot data [7]	Manipulation transfer	Humanoid task coverage
Diffusion Policy [8]	Visuomotor action generation	Sustained integrated performance
HumanPlus and OmniH2O [9], [10]	Whole-body demonstration learning	Assistance and repeated-task reliability
EgoHumanoid [11]	Egocentric loco-manipulation learning	Transfer with documented adaptation
Locomotion and ASAP [12], [13]	Physical control and sim-to-real transfer	Manipulation and production duty cycles

Note: The studies use different tasks and are not ranked.

Figure 1 organizes the synthesis into four evidence levels. A capability demonstration establishes that a behavior can occur under specified conditions. Repeated task evaluation describes its distribution over trials. Integrated operation adds shifts, surrounding equipment, personnel, and recovery procedures. Deployment transfer tests whether acceptable performance continues across installations or changed tasks. These levels describe the scope of the evidence. They are not a certification scheme or a score for individual companies. Evidence at one level does not automatically establish the next.

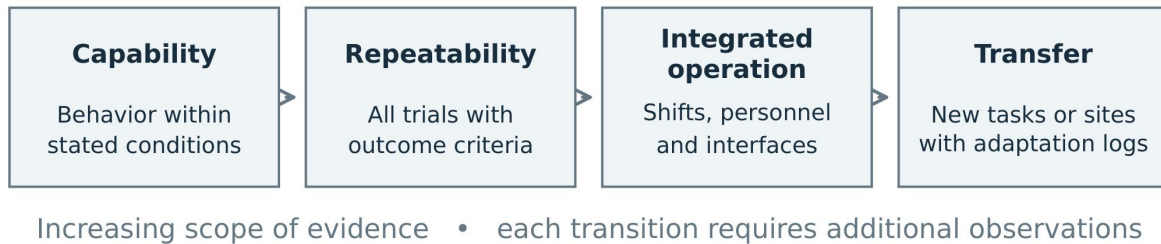


Figure 1 Evidence Levels for Deployment Claims. Each level requires additional observations. This is not a numerical readiness score.

Missing disclosure must also be distinguished from measured failure. If a company omits intervention data, this does not show poor reliability. Its public record cannot support either conclusion about reliability. Similarly, a motion-tracking study should first be judged against its stated objective before its results are considered for production use. A claim is therefore treated as supported only within its stated task, exposure, and source boundary. Incomplete records are not used to produce a numerical readiness ranking.

2.2 What Current Technical Results Establish

Vision-language-action models connect visual observations and language instructions with robot actions. RT-2 demonstrates transfer of semantic knowledge to manipulation [5]. Open X-Embodiment studies data sharing across robot platforms [7], while Diffusion Policy models action generation from demonstrations [8]. OpenVLA provides an openly described approach to generalist manipulation [6]. These studies support the relevance of transferable representations and action learning. Their manipulation results do not directly establish the balance, locomotion, payload endurance, or recovery behavior of a complete humanoid system.

Whole-body studies address part of this gap. HumanPlus combines human motion imitation with learning from humanoid demonstrations [9]. OmniH2O develops a teleoperation and learning interface for whole-body control [10]. EgoHumanoid combines robot-free human egocentric demonstrations with robot demonstrations to improve humanoid locomotion and manipulation across varied settings [11]. These results connect task learning more directly to embodiment, but deployment claims remain limited by the published evaluations. HumanPlus reports ten trials per task and conditional success rates for successive subtasks [9]. This sample describes the tested behaviors but cannot establish rare-event reliability over an industrial service life.

Work on real-world humanoid locomotion and alignment between simulated and physical dynamics adds to the available technical approaches [12-13]. For deployment, the key question is how a policy behaves when perception, contact, and task progress become uncertain. Table 1 therefore pairs each technical contribution with a deployment question. The published RT-2 and OpenVLA models identify inference throughput as a limitation for high-frequency control [5-6]. RT-2 also distinguishes semantic generalization from acquiring physical skills outside its robot training distribution [5]. These effects must be measured in the complete system. They cannot be inferred from model size or a selected success rate.

3 DEPLOYMENT CONSTRAINTS AND INDUSTRIAL EVIDENCE

3.1 Coupled Technical and Operational Constraints

The useful output of a humanoid system depends on how its hardware, control, task policy, and operating environment interact. Actuator limits constrain feasible motions. Contact sensing and state estimation affect manipulation and balance. Power and thermal conditions limit sustained duty cycles. Task planning must respect these physical limits. Figure 2 shows these interactions and the operational feedback needed to diagnose failures. The diagram is a proposed framework. It does not describe the architecture of a tested robot.

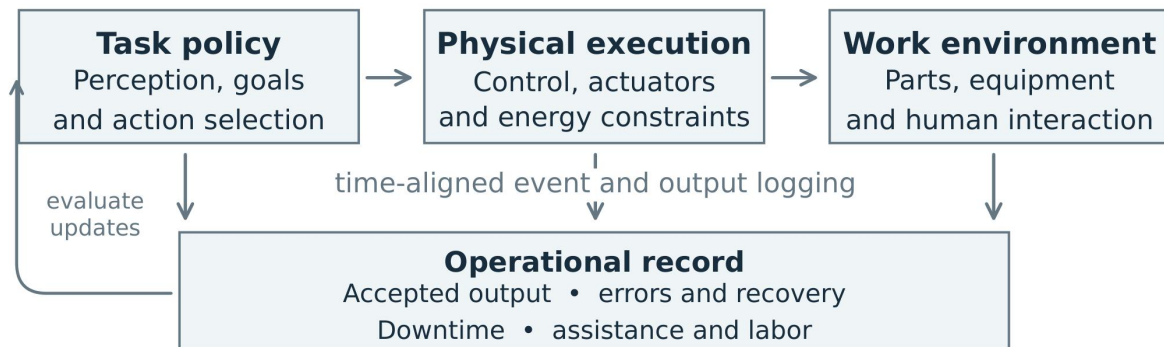


Figure 2 Links between the Robot System and Its Operating Records. Failures guide changes that must be tested again.

Integration adds constraints that component specifications do not capture. Parts must arrive in usable poses, surrounding equipment must receive task-completion signals, and production staff need workable responses to interruptions. In shared workspaces, robot performance also interacts with team performance. Manufacturing research on overall equipment effectiveness supports separating availability, execution rate, and accepted output [14]. Human-robot evaluation likewise calls for measuring the interaction and its context, including the operator's role [15]. These approaches support evaluating the robot cell together with its support labor.

Data collection should use the same system boundary. A failed grasp, a balance recovery, a missing part, and an operator pause have different causes and remedies. Recording only final success loses this distinction. Teleoperation can help collect demonstrations or recover a task, but its labor requirements must remain visible. Training on deployment failures may improve coverage if updated policies are evaluated on held-out shifts and task variants. Otherwise, repeated exposure to a small task set can resemble generalization without showing transfer.

The choice of body form is part of this analysis. A humanoid may avoid some workstation changes but require additional balance control, collision management, or maintenance. A fixed manipulator may need fixtures and achieve a different duty cycle. The comparison must include these integration effects. Human resemblance is not a performance metric. Task access, payload, reach, and permissible layout changes should determine which alternatives are evaluated. Better performance by one platform in a particular workflow would not establish its superiority across other tasks.

3.2 Evidence from Production and Logistics

Public customer reports provide evidence beyond laboratory demonstrations, although they disclose only part of the operating record. BMW reported that its 2025 Figure 02 pilot at Spartanburg moved more than 90,000 sheet-metal components over approximately 1,250 operating hours and supported production of more than 30,000 vehicles [16]. The report also describes involvement from production IT, occupational safety, process management, and logistics. These observations show repeated activity in a production setting. They do not reveal the number of attempts, intervention frequency, maintenance burden, or an independently comparable cost per task.

GXO's 2024 agreement with Agility Robotics provides another example of a commercial arrangement [17]. Following a pilot, the companies announced a multi-year Robots-as-a-Service arrangement for Digit to transfer totes from mobile robots onto conveyors. This establishes a defined workflow and a service contract. It does not show that leasing is economically better than ownership or that the arrangement represents the whole sector. Table 2 separates what each disclosure establishes from the operational evidence still needed for comparison.

Table 2 Scope of Customer Deployment Reports

Case	Public evidence	Unresolved comparison
BMW Figure 02 [16]	2025 pilot, over 90,000 components, approximately 1,250 operating hours	Attempts, assistance, availability, cost
GXO Digit [17]	2024 multi-year RaaS agreement for tote transfer	Scale, throughput, support labor, cost
BMW Figure 03 [18]	June 2026 logistics sequencing project announcement	Measured outcomes of the new workflow

Note: Company reports describe activity or agreements, not independent validation.

BMW's June 2026 Figure 03 announcement describes a later logistics sequencing project [18]. It explores a different workflow that connects component sorting to existing transport infrastructure. These planned activities should be kept separate from the earlier Figure 02 operating record. Together, the cases support assessing deployment readiness for each task. They do not support a universal commercialization date, a ranking of national capability, or a fixed purchase-price threshold for profitable operation.

3.3 Safety as a Condition of Useful Output

Cost comparisons are meaningful only within acceptable operating conditions. ISO 10218-1 and ISO 10218-2 address industrial robots and their applications and cells, respectively [19-20]. Their scope must be checked against the actual humanoid deployment. In particular, ISO 10218-1 does not cover mobility hazards associated with integration on mobile platforms [19]. The assessment must therefore examine relevant mobility, contact, fall, and application hazards alongside the industrial robot requirements. A low calculated cost cannot compensate for an unacceptable operating condition. Safety-related stops should remain in the evaluation record. Their causes and downtime should be separated from other interruptions. A trial with no observed injury does not establish a sufficiently low rare-event risk. The proposed protocol therefore requires safety review and documented operating limits before evaluation. Protective interventions remain part of the evidence about the conditions under which production is feasible.

4 AN AUTONOMY ADJUSTED COST MODEL

4.1 Output Accounting and System Boundary

Consider one robot cell performing a specified task family over one reference year. Let H be its scheduled service hours during that year, u the fraction allocated to the task, and a the available fraction of allocated time. Let r be the number of initiated work-unit episodes per available hour and p the fraction accepted on the first pass without task-level human assistance. Remote commands that complete or correct an episode count as assistance. Passive monitoring does not, but its labor cost is included. Table 3 defines the variables and measurement requirements.

Table 3 Variables in the Annual Task Model

Quantity	Definition	Unit
H	Scheduled service time	hours/year
u, a	Task allocation and conditional availability	dimensionless
r, rB	Robot episode rate and baseline accepted-task rate	tasks/hour
p	Autonomous first-pass accepted episodes / all initiated episodes	dimensionless
K, I	Acquisition and integration expenditure	currency
g	Annualization factor	1/year
F, C	Other operating cost and total annual cost	currency/year
s, w	Support person-hours per cell hour and loaded wage	person-hours/hour, currency/person-hour
b, p	Relative rate and normalized annual cost	dimensionless

Note: Use the same reporting period and accounting rules throughout.

Equation (1) gives the number of autonomously accepted tasks.

$$Q_A = Huar p. \quad (1)$$

The factors form an accounting identity when calculated from the same complete log with the stated time and count denominators. Their multiplication does not require statistical independence. Charging, maintenance, and recovery downtime must each be assigned once to a predefined category. Attempts and accepted outputs need consistent identifiers so that retries cannot inflate the output count. For different tasks within a task family, calculate output by task or use a stated common output unit.

Each attempt should represent one work unit under a prespecified episode rule. Failure or assistance must not create a new identifier that turns the same unit into an autonomous first-pass success. Any automatic retry allowed within an episode must be declared, and its time must be included. A failed attempt can reduce yield, while a later stopped recovery reduces availability. These are distinct losses. Once stopped recovery has reduced a , its duration must not be subtracted again when calculating r from available time.

Let N count all unique initiated episodes, including failures and interruptions. Then r equals N divided by Hua , and p equals Q_A divided by N . These denominators allow Equation (1) to be reconstructed from the log.

Let C be the annual equivalent cost of the cell. Equation (2) separates annualized acquisition and integration costs from operating and supervision costs.

$$C = g(K + I) + F + Hsw. \quad (2)$$

K is the acquisition cost, I the integration cost, g a stated annualization factor, and F the remaining annual operating cost. The supervision intensity s is paid person-hours per scheduled cell hour, and w is the full hourly labor cost. F includes energy, maintenance, software, and other costs not already counted elsewhere. When one operator supervises several robots, labor allocation and the ability to respond must be measured. Annual lease payments can replace the relevant ownership costs without changing the output accounting.

The annualization rule should state assumptions about service life, discounting, and residual value. A short pilot does not determine these quantities. Counting only hands-on minutes understates supervision when dedicated coverage prevents an operator from doing other work. Support labor already paid through a service contract must not be counted again. These accounting choices must be reported because hardware prices alone cannot determine them. Applications of the model should examine how service life and labor allocation affect the result.

For shorter field observations, costs and outputs must cover the same period. Estimates of annual operation based on these observations must state assumptions about scheduling, maintenance, and stable operating conditions.

4.2 A Conservative Autonomous Delivery Comparison

The autonomy-adjusted cost c_A equals C divided by Q_A . It assigns the full cell cost to autonomously accepted first-pass output. This is a screening measure for autonomous delivery, not ordinary production unit cost. If assisted or reworked tasks produce saleable output, also report C divided by all accepted output. The autonomous first-pass share is Q_A divided by all accepted work units, each counted once. Failing the autonomous comparison alone does not show that the cell is commercially uneconomic.

Let r_B be the baseline's accepted-task rate per scheduled service hour under stated reference conditions, and c_B its cost per accepted task. Define b as the robot's initiated-episode rate r divided by this accepted-output reference rate. Define normalized annual cost ρ as C divided by $H r_B c_B$. Equation (3) gives the normalized comparison.

$$\frac{c_A}{c_B} = \frac{\rho}{uabp}. \quad (3)$$

The condition c_A no greater than c_B is equivalent to Equation (4).

$$\rho \leq uabp. \quad (4)$$

The baseline must meet the same output quality and service conditions. It may be the existing manual process, a fixed robot cell, or a mobile manipulator. Its cost calculation must cover a comparable system. The normalization denominator represents full-schedule baseline spending. It is a reference scale and does not require the baseline to operate continuously. The product on the right is an effective autonomous throughput ratio. It can exceed one if relative task speed exceeds one, so it is not a probability.

No single robot purchase-price threshold applies to all deployments. The same acquisition cost can lead to different costs per accepted task because task demand, availability, integration effort, and supervision differ. A lower purchase price changes only part of annual cost. Faster nominal motion may be offset by more rejected attempts or longer recovery time. These quantities must be measured before the economic effect of either change can be assessed.

Matching the baseline cost per task does not ensure enough output to meet the required schedule. A cell may have a low cost per task but complete too few tasks. Unused capacity does not create value when demand is limited, even if that capacity is inexpensive. A deployment decision must therefore consider required output and delivery timing alongside the cost ratio. For a robot in a production line, improvements to its task must be checked against upstream and downstream constraints before they are treated as gains in line output.

4.3 Illustrative Sensitivity Analysis

Table 4 presents four hypothetical cases. The values illustrate the equations and are not fitted to a company, robot, wage level, or market forecast. Scheduled task allocation is 0.75, and relative rate is 0.80. The reference case uses availability of 0.80, autonomous first-pass yield of 0.90, and normalized annual cost of 0.48. Each comparison changes one quantity while keeping the others fixed.

Table 4 Hypothetical Cases with $u = 0.75$ and $b = 0.80$

Case	ρ	a	p	$uabp$	c_A / c_B
Reference	0.48	0.80	0.90	0.4320	1.1111
Lower annual cost	0.40	0.80	0.90	0.4320	0.9259
Higher availability	0.48	0.95	0.90	0.5130	0.9357
Higher first-pass yield	0.48	0.80	0.99	0.4752	1.0101

Note: Hypothetical inputs. Intervention costs and effects between parameters are held fixed.

The reference effective throughput ratio is 0.432, giving an autonomy-adjusted cost ratio of 1.111. Reducing normalized annual cost to 0.40 gives 0.926. Raising availability to 0.95 gives 0.936. Raising autonomous first-pass yield to 0.99 gives 1.010. Figure 3 shows the joint availability and yield boundary for the reference cost. With the original allocation and task rate, and availability fixed at 0.80, the yield required to reach parity is exactly 1.00. Yield improvement alone therefore cannot produce a ratio below one in this example.

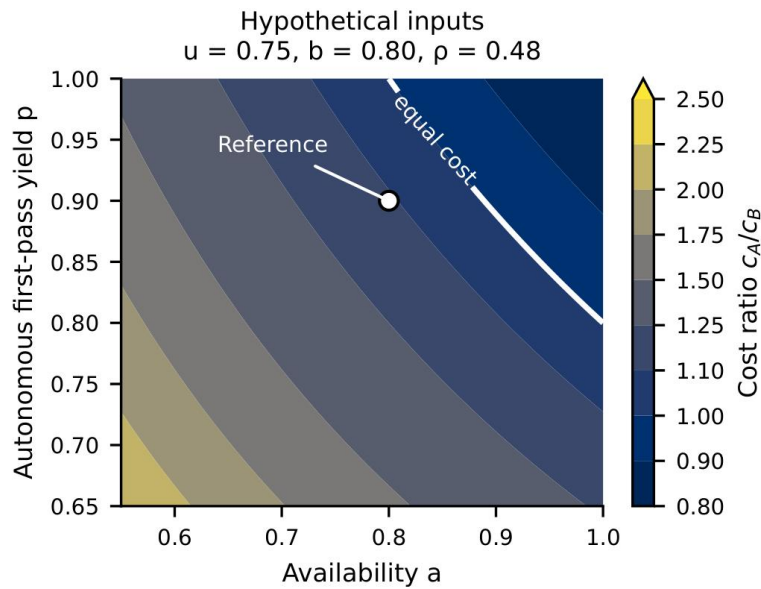


Figure 3 Calculated Cost Ratio under Hypothetical Fixed Inputs. The white line marks equal cost. No commercial performance is estimated.

These calculations show how the factors affect cost together. They do not rank engineering investments. Improving availability may require more maintenance spending, redundancy, or slower operation. Better autonomous yield may reduce support labor and change both the numerator and denominator. The comparisons keep these effects fixed to show the arithmetic clearly. A deployment decision needs measurements of their joint changes and uncertainty ranges. A fuller investment analysis must also consider financing, residual value, demand risk, and revenue from assisted output.

5 A PROPOSED FIELD EVALUATION PROTOCOL

A field test of deployment readiness should begin with a task contract. It should specify incoming parts, allowed variation, completion criteria, cycle deadlines, assistance rules, and operating limits. The same contract must apply to the baseline. Task selection should test where a humanoid body is useful, such as work at multiple stations designed for humans. The evaluation should also include simpler automation. Compatibility with human infrastructure is a proposed benefit to test, and does not justify leaving out alternatives.

Figure 4 presents a proposed evaluation sequence. The first test measures behavior on a fixed task distribution. Later shift trials record every attempted cycle and interruption. Transfer tests change a stated variable, such as part presentation, task order, or workstation, and record the effort needed to adapt. Testing at a separate site or during a held-out period checks whether the result depends on local tuning. This study proposes the sequence but has not carried it out.



Across every stage • timestamps, policy versions, attempts, acceptance, assistance and costs

Figure 4 Proposed Field Evaluation Steps. This study did not carry out the protocol.

The record should connect robot events with accepted output and support labor. Each attempt needs timestamps, task and part identifiers, policy version, result, assistance status, and recovery events. Time records should separate lack of task demand from unavailability. Operator effort should include monitoring, intervention, resets, and maintenance coordination. A policy change starts a new evaluation segment. Report scheduled exposure, attempts, autonomously accepted tasks, all accepted tasks, downtime, and supervision hours before calculating ratios.

Uncertainty estimates should account for where variation occurs. Pooling cycles can hide dependence within shifts or repeated work on the same parts. Shift-level summaries and intervals based on independent operating blocks are preferable when such blocks are available. Rare failures require enough observation time and a separate hazard analysis. No single minimum trial count establishes readiness for every application. Both successful and unsuccessful shifts should remain in the analysis. Any exclusions should be specified before performance is examined.

A reviewer should be able to reconstruct the ratios without access to proprietary model weights. The minimum record should include an anonymized event log, variable definitions, cost allocation rules, and an analysis script with a recorded version. If commercial restrictions prevent cost disclosure, a documented normalization and parameter ranges still allow

part of the analysis to be checked. These records cannot replace independent observation. They make changes in denominators and selective exclusions easier to detect, and support comparisons that use the same stated task contract. Future field evaluations can test three hypotheses. First, a fixed robot configuration evaluated across full shifts with prespecified input variation has lower autonomous first-pass yield than a fixed-condition demonstration benchmark. Second, transfer to an untuned workstation with prespecified changes in part presentation reduces autonomously accepted tasks per assigned hour, even when nominal manipulation speed is unchanged. Third, at fixed staffing and matched average assistance demand, requests clustered in time increase response delays and reduce availability relative to dispersed requests. Each comparison requires the stated task logs and a declared protocol. None has been tested in this study.

6 DISCUSSION AND CONCLUSION

The synthesis connects industrial adoption to evidence about defined tasks, sustained operation, and transfer. Technical studies establish useful learning and control capabilities. Customer disclosures add evidence of production activity and commercial arrangements. Neither source type alone provides all the quantities needed to compare autonomous delivery with an existing process. The framework identifies the missing quantities and distinguishes accepted output from the autonomy of its delivery.

The analysis has three main limitations. Narrative source selection may omit relevant systems and unsuccessful deployments. Corporate reports reflect companies' incentives about what to disclose and provide limited operating detail. The cost model is an accounting framework whose ability to explain or predict deployment performance has not been tested on operating logs. Its sensitivity cases show the effects of assumptions, not observed gains. The model needs application-specific extensions for multi-robot queues, correlated failures, peaks in human workload, and tasks whose outputs cannot be expressed in comparable units.

Integrators should collect task, downtime, and assistance data alongside hardware specifications. Researchers should explain how training coverage, recovery, and adaptation affect accepted output over time. Customers need comparisons of a defined service delivered under acceptable operating conditions. Progress toward industrial humanoid deployment can then be assessed through repeatable evidence, rather than inferred from isolated demonstrations or aggregate shipment growth. The next research step is a planned comparison using shared task definitions and operating records that others can independently review.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

- [1] Brooks R A. Intelligence without representation. *Artificial Intelligence*, 1991, 47: 139-159. Available at: <https://people.csail.mit.edu/brooks/papers/representation.pdf>.
- [2] Ministry of Industry and Information Technology of the People's Republic of China. Guiding Opinions on the Innovative Development of Humanoid Robots. Ministry of Industry and Information Technology of the People's Republic of China, 2023: 193. Available at: https://www.miit.gov.cn/zwgk/zcwj/wjfb/tz/art/2023/art_48fe01d562644aedb7ea3f4256df8190.html.
- [3] International Federation of Robotics. World Robotics 2025. Statistical summary and Industrial Robots Sources and Methods. International Federation of Robotics, 2025. Available at: <https://ifr.org/worldrobotics/report-2025>.
- [4] Norton A, Messina E R, Yanco H A. Advancing Capabilities of Industrial Robots Through Evaluation, Benchmarking, and Characterization. *Recent Advances in Industrial Robotics*. World Scientific, 2021. DOI: 10.1142/9789811222849_0013.
- [5] Brohan A, Brown N, Carbajal J, et al. RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control. *Proceedings of the 7th Conference on Robot Learning*. *Proceedings of Machine Learning Research*, 2023, 229: 2165-2183. Available at: <https://proceedings.mlr.press/v229/zitkovich23a/zitkovich23a.pdf>.
- [6] Kim M J, Pertsch K, Karamcheti S, et al. OpenVLA: An Open-Source Vision-Language-Action Model. *Proceedings of the 8th Conference on Robot Learning*. *Proceedings of Machine Learning Research*, 2025, 270: 2679-2713. Available at: <https://proceedings.mlr.press/v270/kim25c.html>.
- [7] Open X-Embodiment Collaboration. Open X-Embodiment: Robotic Learning Datasets and RT-X Models. 2024 IEEE International Conference on Robotics and Automation, 2024: 6892-6903. DOI: 10.1109/ICRA57147.2024.10611477.
- [8] Chi C, Feng S, Du Y, et al. Diffusion Policy: Visuomotor Policy Learning via Action Diffusion. *Robotics: Science and Systems XIX*, 2023. DOI: 10.15607/RSS.2023.XIX.026.
- [9] Fu Z, Zhao Q, Wu Q, et al. HumanPlus: Humanoid Shadowing and Imitation from Humans. *Proceedings of the 8th Conference on Robot Learning*. *Proceedings of Machine Learning Research*, 2025, 270: 2828-2844. Available at: <https://proceedings.mlr.press/v270/fu25a.html>.
- [10] He T, Luo Z, He X, et al. OmniH2O: Universal and Dexterous Human-to-Humanoid Whole-Body Teleoperation and Learning. *Proceedings of the 8th Conference on Robot Learning*. *Proceedings of Machine Learning Research*, 2025, 270: 1516-1540. Available at: <https://proceedings.mlr.press/v270/he25b.html>.

- [11] Shi M, Peng S, Chen J, et al. Unlocking In-the-Wild Loco-Manipulation with Robot-Free Egocentric Demonstration. *Robotics: Science and Systems XXII*, Sydney, Australia, 2026: 204. DOI: 10.15607/RSS.2026.XXII.204.
- [12] Radosavovic I, Xiao T, Zhang B, et al. Real-world humanoid locomotion with reinforcement learning. *Science Robotics*, 2024, 9(89): eadi9579. DOI: 10.1126/scirobotics.adi9579.
- [13] He T, Gao J, Xiao W, et al. ASAP: Aligning Simulation and Real-World Physics for Learning Agile Humanoid Whole-Body Skills. *Robotics: Science and Systems XXI*, 2025. DOI: 10.15607/RSS.2025.XXI.066.
- [14] Helu M, Hedberg T Jr. Recommendations for Collecting, Curating, and Re-Using Manufacturing Data. *National Institute of Standards and Technology, Advanced Manufacturing Series 300-11*, 2020. DOI: 10.6028/NIST.AMS.300-11.
- [15] Marvel J A, Bagchi S, Zimmerman M, et al. Towards Effective Interface Designs for Collaborative HRI in Manufacturing: Metrics and Measures. *ACM Transactions on Human-Robot Interaction*, 2020, 9(4): 25. DOI: 10.1145/3385009.
- [16] Benedikt Torka. BMW Group to deploy humanoid robots in production in Germany for the first time. BMW Group, Munich, 2026. Available at: <https://www.press.bmwgroup.com/global/article/detail/T0455864EN/bmw-group-to-deploy-humanoid-robots-in-production-in-germany-for-the-first-time>.
- [17] GXO Logistics, Inc. GXO Signs Industry-First Multi-Year Agreement with Agility Robotics. Greenwich, Conn, 2024. Available at: <https://investors.gxo.com/news-releases/news-release-details/gxo-signs-industry-first-multi-year-agreement-agility-robotics/>.
- [18] Benedikt Torka. BMW Group advances the use of Physical AI in production with Figure 03 project in Spartanburg. BMW Group, Munich / Spartanburg, USA, 2026. Available at: <https://www.press.bmwgroup.com/canada/article/detail/T0458794EN/bmw-group-advances-the-use-of-physical-ai-in-production-with-figure-03-project-in-spartanburg>.
- [19] International Organization for Standardization. ISO 10218-1: 2025, Robotics—Safety requirements—Part 1: Industrial robots (3rd ed.), 2025. Available at: <https://www.iso.org/standard/73933.html>.
- [20] International Organization for Standardization. ISO 10218-2: 2025, Robotics—Safety requirements—Part 2: Industrial robot applications and robot cells (2nd ed.), 2025. Available at: <https://www.iso.org/standard/73934.html>.