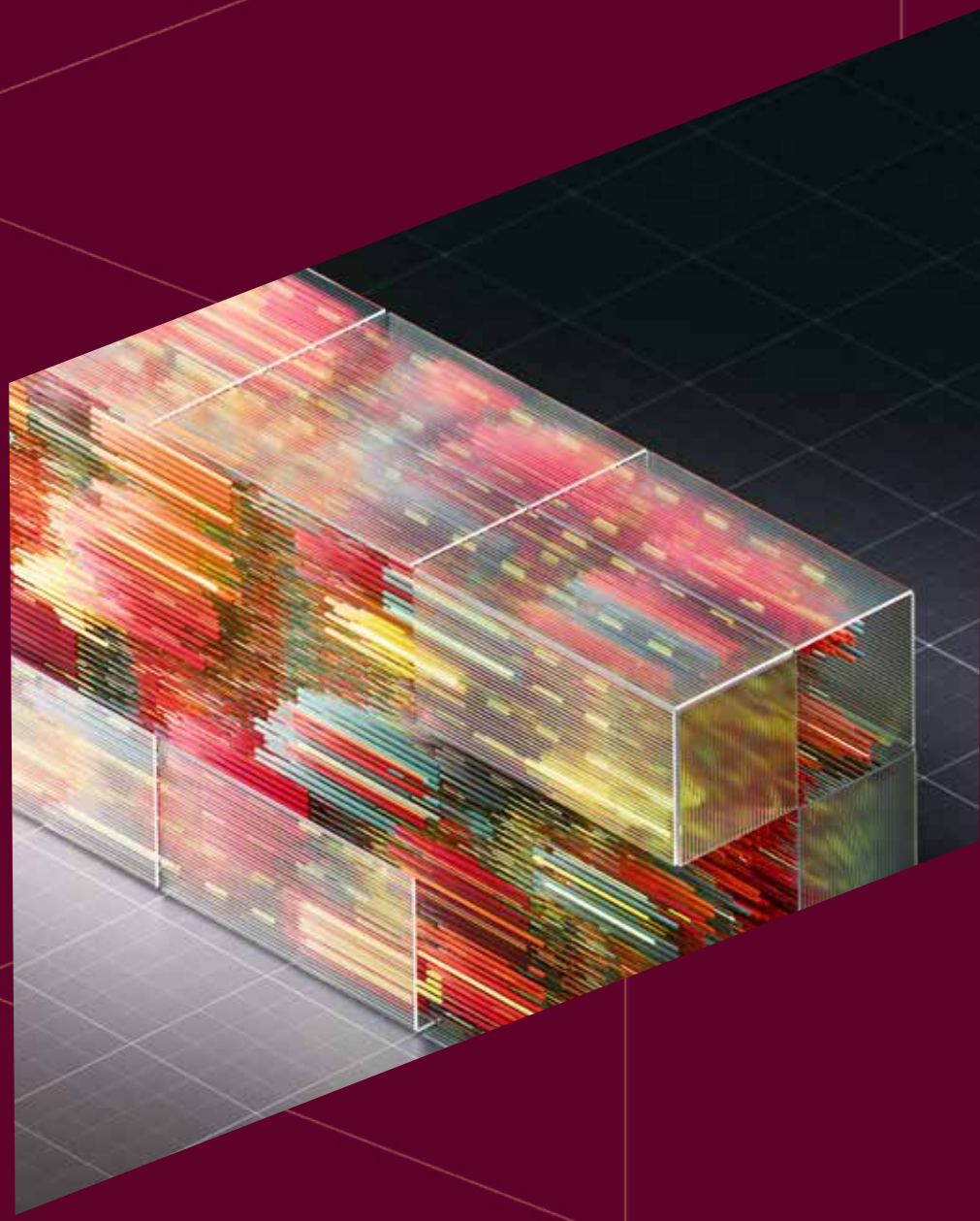
Whitepaper

Accelerating Robotics Development Through Data Readiness



Author

Mothiraj Ramalingam,
Group Practice Head, Digital Business Operations,
Tech Mahindra Business Process Services

Executive Summary

Robotics is emerging as the next wave of AI-driven transformation, with rapid adoption across manufacturing, healthcare, logistics, and autonomous systems. Unlike traditional AI, robotics operates at the intersection of physical and digital **(phygital)** ecosystems. This creates distinct challenges for enterprises in capturing realistic scenarios, managing complex environments, and building scalable operations that support continuous learning.

By investing early in scalable data ecosystems, strong governance, and experienced partners, enterprises can reduce development costs, improve robot performance, and scale robotics initiatives more effectively. Based on extensive experience from live robotics programs, this whitepaper presents **Tech Mahindra's Robotics Data Readiness Framework,** a structured approach to data collection, onboarding, validation, and optimization.

Table of Contents

The State of Play

Robotics is expected to drive major disruption across key industries. Beyond traditional AI use cases, it is rapidly advancing with progress in Gen AI and agentic AI solutions. With faster models and automation capabilities at hand, enterprises are focusing on building robots, autonomous vehicles, and drones as the next logical steps in enhancing services and driving technology-led innovations.

The scale of this opportunity is reflected in the market's projected growth. Market Research Future estimates that the global robotics market will grow from USD 93.5 billion in 2026 to USD 421.1 billion by 2035.[1] This trajectory demonstrates the expanding application of robotics and its potential for wider adoption.

Realizing this potential, however, depends on the data used for training the models. As with any other automation initiative, this endeavor requires a significant volume of high-quality data to enable robots to perform human-like activities. Therefore, a key challenge for enterprises is building a strong foundation of diverse, use-case-specific data for effective training.

Data Requirements for Robotics

Traditional AI data requirements are typically single-mode and primarily digital. Robotics data, on the other hand, involves a combination of both digital and physical data.

Here's a rundown on key requirements for robotics development.

Data Dimension	Core Requirements
Multimodal data	Robotics training requires multiple data types, including video, telemetry, sensors, and LiDAR.
Evolving data	Robotics demands highly dynamic data for active training as opposed to static data used in traditional AI model development.
Data processing and management	Traditional AI models rely on human-generated data and annotations for training. Robotics training requires both human-generated and system-generated data to develop robust models. This process involves collecting robotic training data, annotating it, and subsequently validating the annotations to ensure data quality.

Tech Mahindra's Key Findings from the Field

The experience across real-world robotics programs reveals three critical phases that shape data quality and training readiness.

Phase 1: Data Collection/ Acquisition

The initial phase focuses on capturing data in the physical environment. From our learnings, we suggest enterprises to:

- Build an appropriate data mix to handle seasonality-related issues such as weather, lighting, temperature, motion, and edge cases.
- Validate system alignment to avoid data recollection. For instance, human actions captured on video do not map to robotic actions. Therefore, these factors need to be accounted for.
- Capture scenarios beyond baseline behavior. Human behavior is inherently unpredictable, making representative edge cases, failures, and exceptions more difficult to capture.
- Maintain accurate metadata quality, synchronization across devices/sensors, and traceability throughout the collection process.

In robotics, the value of data comes from its context, synchronization, and traceability, not volume alone.

Phase 2: Data Onboarding

The next phase involves preparing captured data for processing and training. For this stage, enterprises should consider:

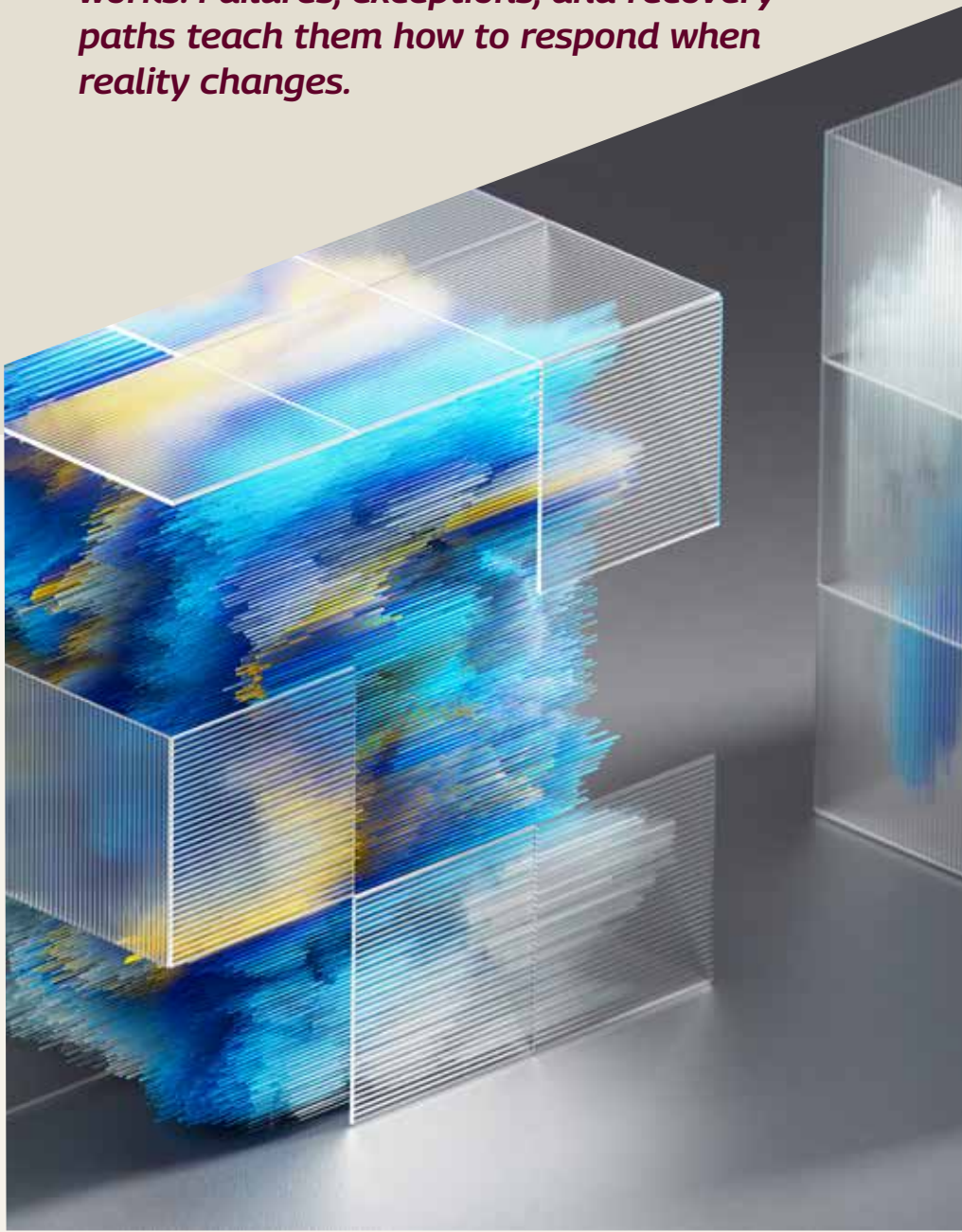
- Standardizing data capture to map ground realities accurately.
- Verifying calibration and synchronization across video, audio, sensors, and other data sources before onboarding them into the training pipeline.
- Defining relationships of data sources and mapping them to a common structure, making the information easier to process and interpret.
- Managing calibration and version drift proactively.

Phase 3: Data Operations

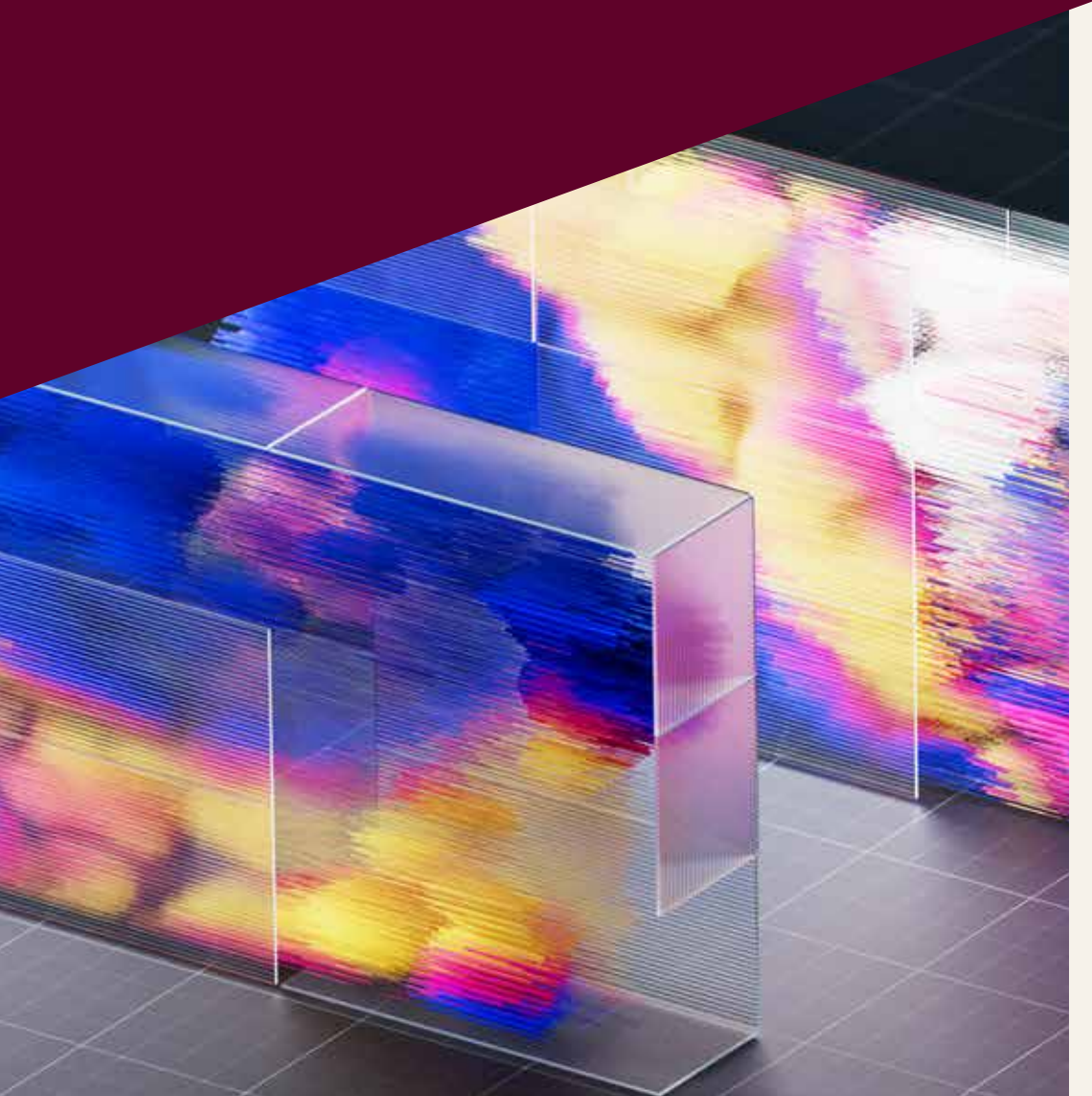
The third phase is the continuous process of preparing, annotating, validating, and improving data during model training. Effective data operations require enterprises to:

- Establish a continuous feedback loop to identify data gaps and improve training data as models evolve.
- Use tool features to capture data requirements, specifically robot training data, such as time, capture, force, placement, and trajectories.
- Introduce active learning to prioritize valuable data and reduce repetitive labeling.
- Leverage human demonstration videos to help robotics models learn complex actions and behaviors.
- Maintain control as data collection scales across multiple sites, geographies, and teams to ensure data quality and consistency.
- Capture failures and recovery behaviors alongside successful actions, as exceptions and recovery paths are equally valuable for training and validation.

Successful actions teach robots what works. Failures, exceptions, and recovery paths teach them how to respond when reality changes.



Tech Mahindra's Approach to Bridge Physical and Digital Data Ecosystems



Over the last five years, Tech Mahindra has helped clients on various projects involving large volumes of phygital data that help bring their robots to life. Our services dedicated to this include:

Human-Robot Interaction Safety zones, handovers, operator intent, gestures and collaborative task behavior.	**Lab / Field Protocols** Structured collection plans, environment setup, device calibration and QA sampling.	**Gesture Signals**	Hand movements, body poses, gaze, and intent cues for intuitive robot response.
		Collaborative Tasks	Human-robot handovers, shared assembly steps, and co-working task flows.
Robotics CV + Sensors Camera, depth, LiDAR, IMU, force and joint-state data captured in sync.	**Mobility Data** Route, GPS, camera, LiDAR and mobility signals for autonomous systems.	**Safety Behaviors**	Proximity, collision avoidance, emergency stops, and safe-zone interactions.
Teleoperation Sessions Human-guided robot demonstrations for grasping, sorting, navigation and recovery.	**Complex Environments** Industrial, healthcare, warehouse and controlled facility collection scenarios.	**Instruction Data**	Voice, text, and demonstration inputs that teach robots new task patterns.

Coverage across collaborative, mobile, industrial and field robotics environments

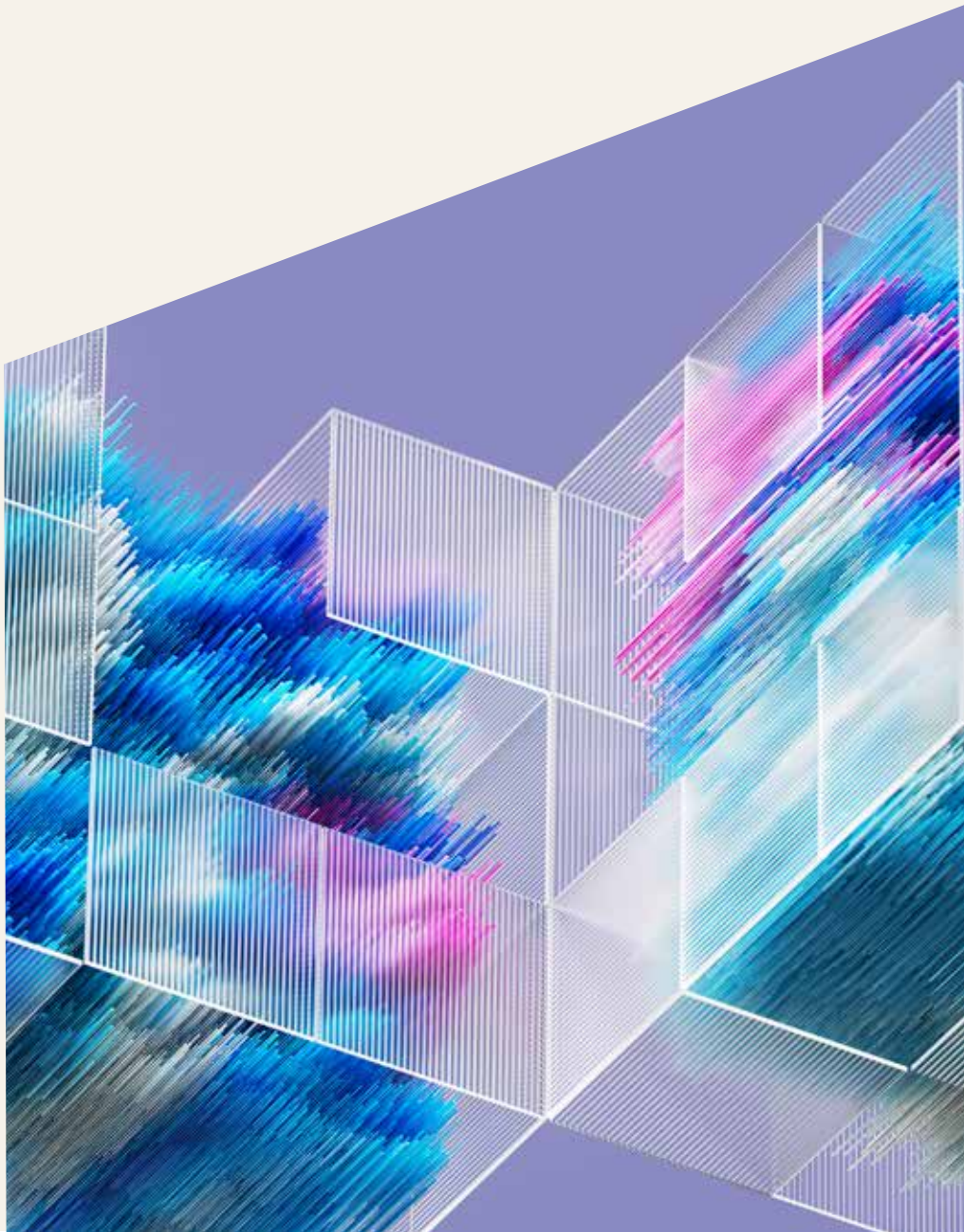
Industrial Arms Articulated, SCARA, delta; assembly, welding, inspection	**Collaborative Robots** Cobots in shared workcells; handovers, safety, gestures
AMRs / AGVs Autonomous or guided material movement; maps, routes, traffic	**Mobile Manipulators** Mobile base + arm; picking, replenishment, mobile inspection
Warehouse Systems Goods-to-person, sorting, bin picking, inventory robots	**Field Inspection** Legged, crawler, UGV; plants, utilities, mines, offshore

What We Collect Across Robot Types

- Multi-camera video + depth streams
- LiDAR / SLAM maps and trajectories
- Force, torque, IMU and joint-state logs
- Teleoperation and demonstration traces
- Task states, failures and recovery events
- Human proximity, safety zones and HRI signals

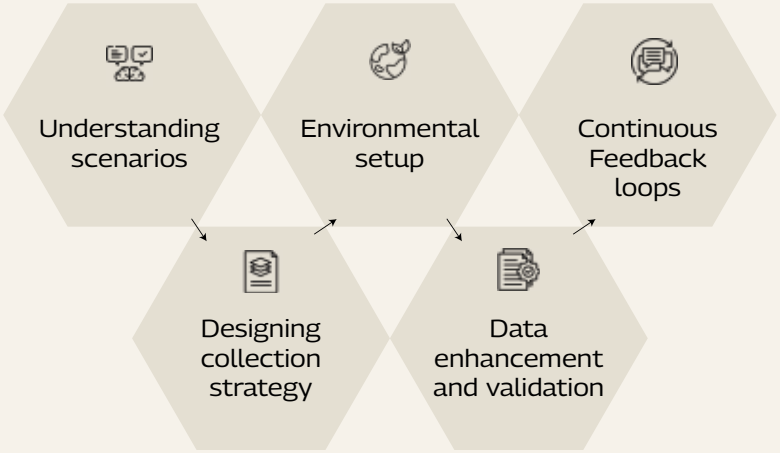
Core today: Industrial arms · Cobots · AMRs

Emerging: Humanoids/ Embodied AI programs



Data Readiness Framework for Robotics

With extensive experience in the robotics data domain, Tech Mahindra has developed a data readiness framework that supports data training for successful robotic project execution. Here is an overview of the framework.



- **Understanding Scenarios**

  For robust data readiness, teams should start by collaborating with clients to understand the intent of the robotics model. This includes where the robot will operate, what objects and conditions it will encounter, and what the expected end outcome is. Clearly mapping these parameters is essential for data collection and preparation.

- **Designing Collection Strategy**

  Once the end scenario is mapped, the next priority is to determine complex data requirements, including the data types, collection methodologies, and other variables. Also, in the process, teams must focus on collecting relevant, high-quality data over sheer volume.

- **Environmental Setup**

  To execute the data collection strategy, teams must prepare the data ecosystem involving both physical data collection and digital data management. Additionally, relevant stakeholders must be involved to capture, validate, and enhance data requirements during the collection phase.

- **Data Enhancement and Validation**

  The captured data is then validated and qualified for model training. This process includes data annotation, quality checks, and enhancements to ensure a seamless end-user experience.

- **Continuous Feedback Loops**

  Evaluate robot performance to identify gaps and enhance data training through a continuous feedback loop. Repeat the cycle until the model achieves optimal performance.

Key Recommendations for Robotics Data Training

Implementing the proposed framework for data readiness requires deliberate and disciplined execution. The following considerations help enterprises establish a strong data foundation for robotics.

- Build a scalable data ecosystem.
- Establish systems that support data handling and governance.
- Design the data ecosystem based on the use-case requirements from the outset.
- Collaborate early with a partner who has the right phygital data expertise.

The Final Word

Robotics is expanding beyond established manufacturing applications into collaborative workplaces, surgical systems, and e-commerce fulfillment. Mordor Intelligence forecasts collaborative robots to grow at a CAGR of 25.64% through 2031, while the market for surgical robotics platforms is expected to reach USD 8.11 billion over the same period.[2] This momentum points to a future in which robots take an increasingly significant role and operate across complex tasks and environments.

To support this expansion and development of robotics, enterprises need data quality and readiness for effective training. By prioritizing the proposed use-case-specific data collection strategies, disciplined processes, fit-for-purpose tools, strong governance, and continuous feedback loops, enterprises can reduce rework and development costs, improve model performance, and deliver reliable robotics outcomes. The path to capable robots begins with quality training data.

End Note

1. Market Research Future. (2026, July 2). Robotics market size, share, and research report by type, by environment, by mobility, by application, by end user, and by region - Industry forecast till 2035 (Report No. MRFR/SEM/3310-CR). https://www.marketresearchfuture.com/reports/robotics-market-4732

2. Mordor Intelligence. (July 31, 2026). Robotics Market Size & Share Analysis - Growth Trends and Forecast (2026-2031), https://www.mordorintelligence.com/industry-reports/robotics-market.

Mothiraj
Ramalingam

Group Practice Head,
Digital Business Operations,
Tech Mahindra Business
Process Services

About the Author

Mothiraj is a seasoned business leader with over 24 years of experience managing client engagements across diverse service lines. His expertise spans delivery operations, center-of-excellence establishment, solution design, and implementation of industry best practices.

As the leader of the Digital Data Services practice, Mothiraj provides strategic direction and drives the vision for delivering transformative AI and machine learning data solutions. By fostering strong partnerships with executive stakeholders, clients, and cross-functional teams across diverse industries, Mothiraj champions innovation, accelerates business growth, and ensures operational excellence while building high-performing teams and scalable, future-ready capabilities.

About **Tech Mahindra**

Tech Mahindra (NSE: TECHM, BSE: 532755) offers technology consulting and digital solutions to global enterprises across industries, enabling transformative scale at unparalleled speed. With 146,000+ professionals across 90 countries, Tech Mahindra provides a full spectrum of services including consulting, information technology, enterprise applications, business process services, engineering services, network services, customer experience & design, AI & analytics, and cloud & infrastructure services. It is the first Indian company in the world to have been awarded the Sustainable Markets Initiative's Terra Carta Seal, which recognizes global companies that are actively leading the charge to create a climate and nature-positive future. Tech Mahindra is part of the Mahindra Group, founded in 1945, one of the largest and most admired multinational federation of companies. For more information on how TechM can partner with you to meet your Scale at Speed™ imperatives, please visit https://www.techmahindra.com/.

TECH mahindra

www.techmahindra.com
www.linkedin.com/company/techmahindra
www.x.com/Tech_Mahindra