machines

MDPI

Review

Artificial Intelligence for Autonomous Mobile Robots in IR4.0–IR6.0: A Unified Review from Perception and Visual Servoing to Decision-Making

Montaser N. A. Ramadan [1], Mohammed A. H. Ali [2,*] and Nik Nazri Nik Ghazali [2,*]

[1] Application and Research Center for Advanced Studies, Istanbul Aydin University, Istanbul 34295, Türkiye; montaserramadan@aydin.edu.tr
[2] Department of Mechanical Engineering, Faculty of Engineering, Universiti Malaya, Kuala Lumpur 50603, Malaysia
* Correspondence: hashem@um.edu.my (M.A.H.A.); nik_nazri@um.edu.my (N.N.N.G.)

Abstract

Reviews of artificial intelligence (AI) for mobile robots usually cover one competence—perception, SLAM, path planning, control, or reinforcement learning—and rarely show how these combine into a working system. We take the opposite view and treat autonomy as one pipeline: sensing, perception, localization and mapping, prediction, planning, visual servoing and control, high-level decision-making, and continual learning. We survey how AI has reshaped each stage for industrial and service robots across Industry 4.0, 5.0, and the emerging Industry 6.0. Using a structured, PRISMA-informed protocol with explicit search strings, inclusion criteria, and cross-embodiment transfer rules, we screen the literature, analyze a corpus drawn mainly from the last five years, and position it against prior surveys with a coverage matrix that exposes their single-block focus. Four findings stand out. Perception and localization approach engineering maturity through multimodal fusion and foundation vision models. Planning and control stay effective but computationally demanding. Decision-making, now driven by large language and vision–language–action models, is powerful yet unverifiable and fails under safety constraints. Lifelong learning is almost absent from deployed systems. The decisive weaknesses sit at the interfaces: at the perception–planning, planning–control, and control–decision handoffs the sim-to-real gap, limited on-robot compute, and scarce industrial data compound. We compare AI families by technology readiness, catalog datasets and benchmarks, examine the safety-certification barrier, and consolidate cross-cutting gaps. We close with a staged roadmap toward Industry 6.0 and a next-generation architecture coupling a foundation perception backbone, a world model and digital twin, a continual-learning memory, and a reasoning core wrapped by a safety monitor. The aim is to move from cataloging algorithms to engineering integrated autonomy.

Keywords: autonomous mobile robots; mobile robot perception; visual servoing; mobile robot path planning; mobile robot control system; decision-making; foundation models; vision-language-action models; deep reinforcement learning; Industry 5.0; simultaneous localization and mapping; sim-to-real transfer

Check for updates

Academic Editor: Marco Ceccarelli

Received: 6 July 2026
Revised: 4 August 2026
Accepted: 4 August 2026
Published: 19 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

A modern autonomous mobile robot no longer follows a buried wire. It senses a crowded aisle, maps as it moves, picks an order, and threads between a forklift and a co-worker. Every one of those behaviors now leans on a learned component. The Kiva system that brought shelves to human pickers roughly doubled pick-station throughput yet ran on a central scheduler with little onboard intelligence [1]. Warehouse robots today localize without pre-installed infrastructure and dodge dynamic obstacles using AI front-ends that did not exist a decade ago [2]. The chassis barely changed; the software did the rest.

This tracks a redefinition of the industrial mission. Industry 4.0 framed the robot as one node in a cyber-physical system, its 5C architecture turning raw signals into cognition [3,4]. Industry 5.0 (a European policy direction from around 2021) puts human centricity, sustainability, and resilience ahead of pure automation, with the force-limited cobot its emblem [5–9]. Industry 6.0 remains a proposed horizon around hyper-customization, digital twins, 6G, and generative-AI "cobot networks" [10]; we treat it as a direction, not a standard. Across all three, AI is the connective layer [11].

These eras share one chain of computation: sensors feed perception, then localization and mapping, scene understanding, prediction, planning, control, and decision-making, with lifelong learning meant to close the loop. Each block has matured in isolation: camera-only perception approaches LiDAR baselines on nuScenes [12]; visual-inertial SLAM reaches centimeter accuracy on EuRoC [13]; learned dense SLAM has cut catastrophic tracking failures [14]; learned planners prune sampling-based search [15]; and language-grounded agents select the correct skill sequence on long-horizon instructions [16]. Strong per-block numbers are what mask the integration problem.

1.1. The Problem: A Stack Reviewed in Silos

The reviews of these advances mirror the software's modularity, not the integration the robot needs. Perception surveys stop at the bounding box [17,18]; SLAM surveys leave planning and control untouched [19,20]; path-planning reviews treat tracking as another chapter [21,22]; RL surveys accept perception and planning as givens [23–25]; work on large language and vision–language–action models thins out on classical SLAM and low-level control [26–28]; and even visual servoing is reviewed as control alone [29,30]. Each survey covers one block; none covers the pipeline end to end.

That siloing is the gap. No existing review takes the whole autonomy pipeline as its unit of analysis. Integration failures, not per-block accuracy deficits, dominate real breakdowns: a detector tuned on clean imagery feeds a planner never trained against its failure modes. We advance this as a structured thesis rather than a measured fact, and support it three ways through the review: the per-block evidence of Sections 4–8, which shows each block maturing while its outputs stay mismatched to the next; the absence of any end-to-end or cross-block benchmark that would let interface errors even be scored (Section 11.3); and the direct finding that integrated hierarchical policies outperform modular sense–plan–act pipelines on long-horizon tasks [31]. The frontier lies between the blocks, and in the near-total absence of lifelong learning in deployed systems.

1.2. Contributions

This review is organized around that argument. We make five contributions:

1. **A pipeline-based taxonomy.** We organize the field by autonomy *function*, not algorithm class, mapping AI methods onto the full chain from sensors to decision-making and lifelong learning. This unified pipeline (Figure 1) is the paper's spine and the lens for every later section.

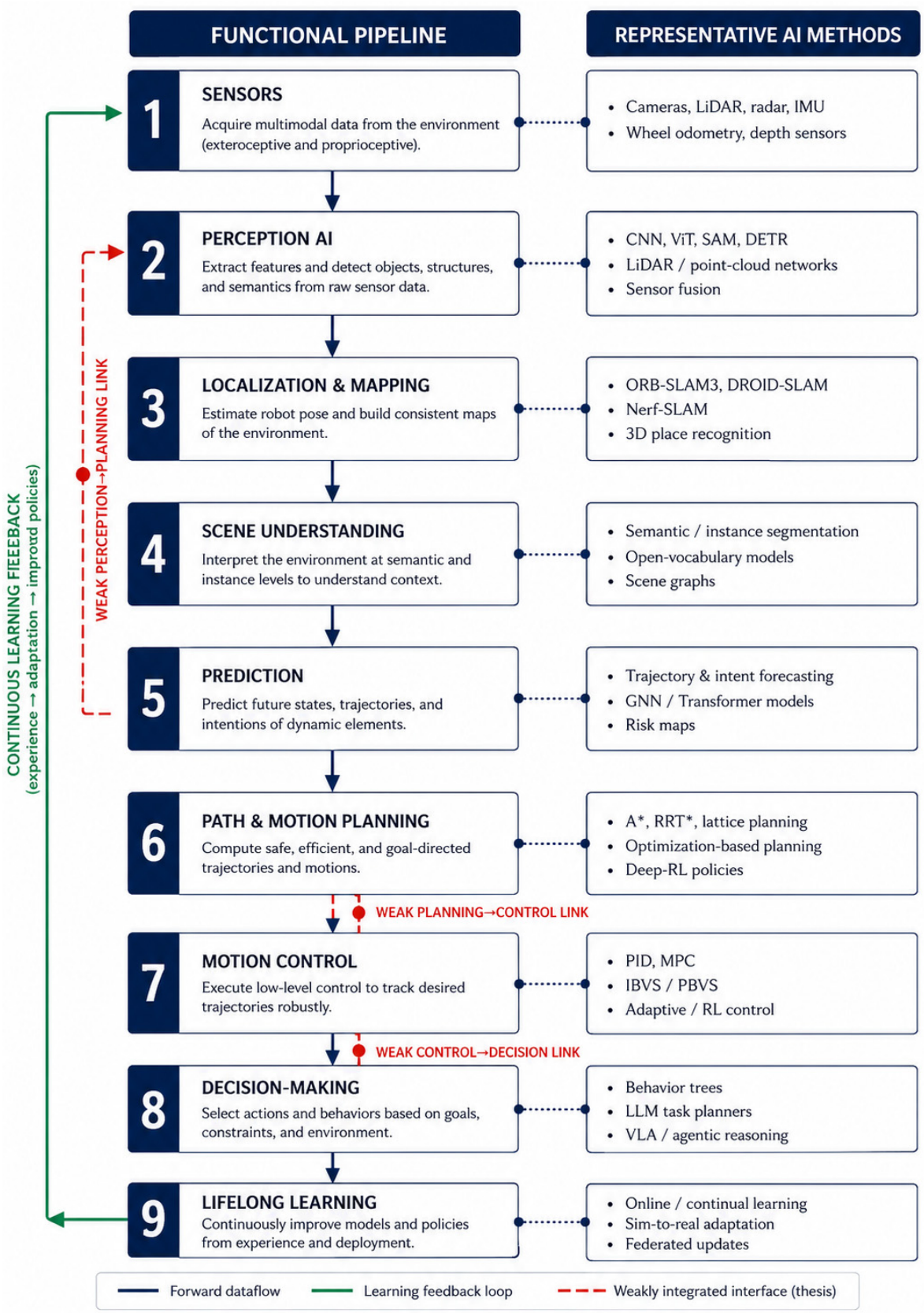


Figure 1. The unified AI autonomy pipeline for industrial mobile robots. Each block names representative AI methods; red dashed markers highlight the weakly integrated interfaces (perception–planning, planning–control, control–decision) that this review identifies as the field's real frontier.

2. **A generation analysis.** We explain *why* the field moved from AGV to AMR to collaborative to learning-based to foundation-model robots across Industry 4.0, 5.0, and the emerging 6.0, reading each transition as a response to a specific limitation of the prior generation [3,5,10].
3. **Between-block bottlenecks.** We identify the integration failures at the perception↔planning, planning↔control, and control↔decision seams that no single-block survey can see, and we position prior surveys against the pipeline in an explicit coverage matrix. We tie these failures to evidence such as the brittleness of sense–plan–act pipelines against learned policies on long-horizon tasks [31].
4. **A qualitative AI-family comparison with TRL.** We rate the major AI families (CNNs, Transformers, reinforcement learning, diffusion policies, LLM task planners, foundation/VLA models, and neuro-symbolic methods) on compute, latency, robustness, interpretability, real-time capability, and technology-readiness level, using cited ordinal ratings rather than fabricated precision.
5. **A staged roadmap toward Industry 6.0.** We lay out the advances needed to move from today's Transformer- and foundation-model robots through vision-language, embodied, digital-twin, and agentic stages, alongside enablers such as federated robot learning [32] and digital twins [33].

Three constraints run through every section: the sim-to-real gap [25,34], the edge compute budget [35], and the availability of industrial data [36].

1.3. Paper Map

Section 2 gives the review protocol and review-of-reviews matrix; Section 3 traces why each robot generation arose; Sections 4–8 walk the pipeline (perception with open-vocabulary scene understanding in Section 4.5, the visual-servoing bridge, localization and mapping, prediction and planning, control and decision-making). Section 9 gives the AI-family comparison with TRL, Section 10 maps capabilities onto applications, and Sections 11 and 12 cover benchmarks and safety-and-standards [37]. Section 13 consolidates the gaps, Section 14 presents the roadmap toward Industry 6.0, and Section 15 concludes.

2. Review Methodology

This review is integrative rather than a meta-analysis: the object is one autonomy pipeline assembled from heterogeneous evidence (benchmarks, system papers, and prior surveys), so we report our search and selection following PRISMA 2020 guidance for transparency, without pooling effect sizes. The protocol below is intended to make the corpus reproducible: the databases, dated searches, verbatim strings, explicit inclusion and exclusion rules, the screening counts of Figure S2, and the criteria by which evidence from adjacent robot embodiments is admitted.

2.1. Databases and Search Strings

We searched four databases: Scopus, Web of Science, and IEEE Xplore for peer-reviewed literature, and arXiv for the newest foundation-model systems that appear as preprints ahead of proceedings, such as RT-2 [38], OpenVLA [39], and Gemini Robotics [40]. Searches were first run in the first quarter of 2026 and refreshed immediately before submission, with the final search executed in June 2026. Each string pairs a pipeline-function term with a method term, restricted to title, abstract, and keywords. Representative strings, given verbatim, are:

- *Perception:* ("mobile robot" OR "autonomous vehicle" OR AMR OR AGV) AND ("object detection" OR "semantic segmentation" OR "BEV" OR "sensor fusion") AND ("deep learning" OR transformer OR CNN).
- *Localization and mapping:* (SLAM OR "visual odometry" OR "place recognition") AND ("deep learning" OR "neural" OR "learned") AND ("mobile robot" OR LiDAR OR "visual-inertial").
- *Prediction and planning:* ("trajectory prediction" OR "motion planning" OR "path planning" OR "social navigation") AND ("deep reinforcement learning" OR transformer OR "diffusion" OR "graph neural network").
- *Control and decision-making:* ("mobile robot" OR "wheeled robot") AND ("model predictive control" OR "reinforcement learning" OR "large language model" OR "vision-language-action") AND (navigation OR control OR planning).

The full block-by-block string set and the databases queried are reported as Table S8 in the Supplementary Materials; the four databases together returned 1914 records before duplicate removal (Figure S2). The strings are keyed to the function-based taxonomy of Figure S1.

2.2. Inclusion and Exclusion Criteria

We targeted 2021–2025 to capture the learning and foundation-model turn, while retaining a small set of seminal earlier works that remain live baselines: PRM [41], the Dynamic Window Approach [42], the Kanayama tracking law [43], LOAM [44], and the canonical image-based visual-servoing formulation [45]. A record was **included** when it (i) was a peer-reviewed journal or conference paper, an archival preprint of a widely adopted system, or a structured survey; (ii) addressed at least one autonomy-pipeline stage of a ground mobile robot, or reported a result transferable to one under the criteria of Section 2.4; (iii) reported an empirical result, a benchmark, or a systematic synthesis; and (iv) was written in English. A record was **excluded** when it was off-topic for mobile-robot autonomy, carried no AI or machine-learning contribution (purely classical work was kept only as a named baseline), was a non-primary or non-technical source such as news or a blog, or gave insufficient methodological detail to appraise. Duplicates across databases were removed before screening.

2.3. Screening, Selection, and Appraisal

The four databases returned 1914 records; reference chaining, key-venue hand-search, and citation tracking of the retrieved surveys added 47, and removing duplicates left 1488 unique records (Figure S2). Title-and-abstract screening against the criteria of Section 2.2 excluded 1002 out-of-scope records, leaving 486 for full-text assessment. Full-text assessment excluded a further 213 (off-topic, non-primary or secondary with no technical content, no AI/ML component, or low quality/insufficient reported detail), retaining 273 studies for synthesis. From these 273, overlapping or superseded reports were consolidated to a single representative citation, and, in line with the editor's and reviewers' request to condense, peripheral single-use entries were folded into their nearest surviving citation; the reference list of this revised manuscript therefore comprises 205 numbered entries. Screening and selection were carried out by the author team; borderline cases at the full-text stage were resolved by discussion against the written criteria rather than by a single reader.

Because the corpus mixes benchmark studies, system papers, and surveys rather than comparable trials, we did not apply a single numerical risk-of-bias score, which would not be meaningful across these study types. Instead each retained study was appraised on three signals—whether it reports a quantitative or benchmarked result, whether it exposes

reproducibility signals such as public code, data, or a standard benchmark, and whether an independent group corroborates it—and these signals govern the confidence language used throughout (for example, a single-paper figure is attributed to its study rather than stated as a general result). Cross-embodiment and preprint evidence is graded more cautiously on the same three signals.

2.4. Scope and Cross-Embodiment Transfer

The target platform is the ground mobile robot of industrial and service settings: wheeled or tracked AGVs, AMRs, and mobile bases of collaborative robots. Evidence from adjacent embodiments—autonomous road vehicles, fixed-base manipulators, legged robots, drones, and underwater vehicles—is admitted only under explicit conditions, because uncritical extrapolation across embodiments is a real hazard. We import such a result only when: (1) it concerns a pipeline stage whose inputs and outputs are embodiment-agnostic, so the mechanism transfers even if the platform does not—a bird's-eye-view detector [12] or a place-recognition descriptor produces the same representation for a car or an AMR; (2) the governing physical assumption still holds, which admits planar-navigation results from road vehicles to wheeled AMRs but excludes legged-locomotion dynamics and aerodynamic control from any performance claim about a wheeled base; (3) the result is used to establish an algorithmic capability or limitation, not a quantitative performance figure for industrial robots; and (4) cross-embodiment learning results, such as ManiSkill2 [46] and Open X-Embodiment [47], are read as evidence about transferability itself rather than as deployed-AMR performance. Where a platform enters purely to illustrate a bound—CERBERUS in subterranean exploration [48], ANYmal locomotion [49]—we flag it as illustrative and draw no numeric claim for wheeled robots. These rules are applied explicitly at each such citation in the sections that follow.

2.5. Review of Reviews: The Coverage Gap

We mapped fourteen recent reviews against the pipeline blocks each covers. Every survey covers a slice, and the slices do not overlap into a whole. Figure 2 presents this coverage matrix; its white space is the gap this paper fills.

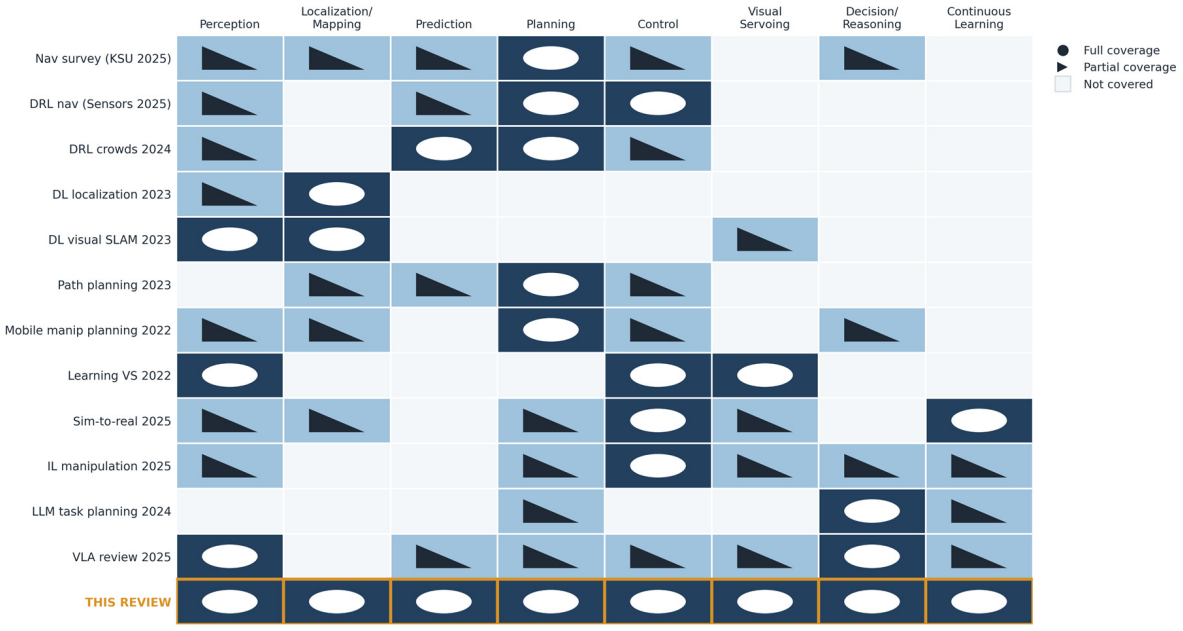


Figure 2. Coverage matrix of representative prior surveys against the autonomy-pipeline blocks. Most surveys cover only a few blocks; this review spans the full pipeline.

The siloing runs along predictable fault lines. Perception reviews stay in perception: 3D object detection [17,18] and visual localization and mapping [19,20] omit downstream stages. Planning reviews stay in planning [21,22], and DRL-navigation reviews couple planning to control while bypassing the perception front-end [24,50,51]. Control and sim-to-real reviews treat perception and planning as upstream givens [23,25,34,52,53]. The visual-servoing surveys sit on the perception–control bridge yet narrow to servo-law taxonomies [29,30,54]. The newest wave, LLM task planning [28], foundation models in deployment [26], vision–language–action models [27], and LLMs for multi-robot teams [55], adds a decision layer but treats SLAM and continuous control qualitatively. The broadest navigation surveys handle control superficially, one omitting visual servoing [56], the other foundation-model integration [57]. Socially aware navigation [58] and mobile-manipulator motion planning [59] each cover one vertical.

None of these reviews traces a claim from sensor reading to deployed decision, and none treats the weak links between blocks, or the absence of lifelong learning, as its central object. A survey-by-survey breakdown is given as Table S1 in the Supplementary Materials.

3. Evolution of AI-Enabled Mobile Robots

Robot intelligence grew in fits, each shift forced by a concrete pressure: a floor that could not be re-wired, a fleet that outgrew central scheduling, a workspace that gained people, a task nobody could hand-code. Figure 3 gives the timeline. Capability climbed the stack one block at a time, never as tightly integrated as claimed.

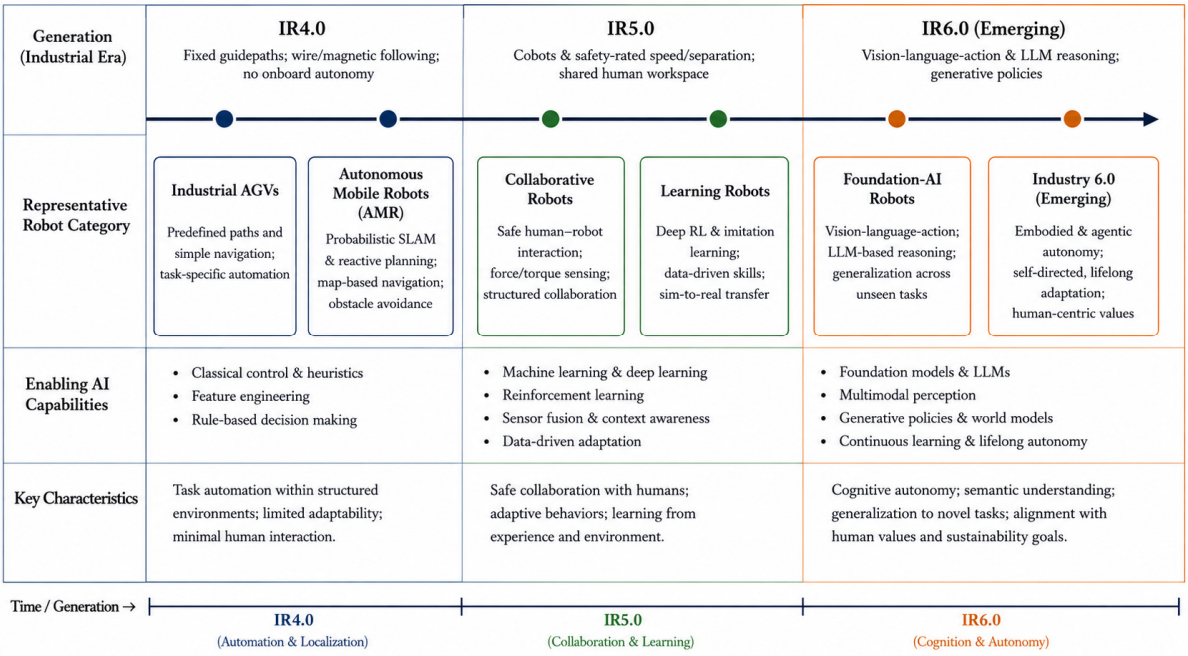


Figure 3. Evolution of AI-enabled mobile robots across Industry 4.0, 5.0, and the emerging Industry 6.0, from fixed-guidepath AGVs to embodied, agentic autonomy.

From guidepaths to onboard estimation (AGV to AMR). Automated guided vehicles followed buried wires under a central scheduler [1], the Industry 4.0 pattern of intelligence above the vehicle [3,4]. Because guidepaths break once a layout moves, the AMR put the map onboard, shifting AI toward SLAM and decentralized coordination [2]; ORB-SLAM3 [13] and LIO-SAM [60] made these stacks deployable, yet the robot still plans on a geometric abstraction that discards most of what it saw.

From navigation to coexistence (AMR to collaborative robots). The next pressure was human. Industry 5.0, EU policy since 2021, reframed the goal around human centricity,

sustainability, and resilience [5], a values-driven complement to Industry 4.0 [6,7,9] tied to edge, twins, and cobots [61,62]. Coexistence bounds contact forces via ISO/TS 15066 [8] with deep-learning detection [63]; reviews target trust, allocation, and wellbeing [64,65], yet surveys find a gap between published capability and what survives on a line [66,67], pushing navigation toward attention-based DRL [68].

From programmed to learned behavior (collaborative to learning robots). Hand-coding broke down once tasks got messy. Classical control gives asymptotic optimality [69] and Lyapunov-stable tracking [43] when the model is known, but partially observed environments resist it. Deep reinforcement learning delivered mapless sim-to-real navigation [70], collision avoidance [71], and warehouse order-picking [72]. Learning was not free: the sim-to-real gap drives domain randomization and system identification [25,52], while sparse rewards and transfer stay open [50].

From narrow policies to foundation reasoning (learning to foundation-AI robots). The recent shift trades many narrow models for a few large pre-trained ones. Vision foundation models supply transferable backbones [73,74]; SayCan grounded an LLM in skill value functions [16] and PaLM-E folded sensor observations into the language embedding space [75]. Vision–language–action models closed the instruction-to-command loop, with RT-2 transferring web knowledge into control [38] and OpenVLA running 4-bit quantized with cross-platform transfer [39,47]. But an AMR cannot host a model the size of PaLM-E, so quantization and the TinyML-to-TinyDL trajectory matter [35], and surveys flag data scarcity, real-time execution, and safety as blockers [36].

Toward Industry 6.0, stated honestly. Industry 6.0 is not settled doctrine. Articulations describe virtualized manufacturing on automated digital twins, 6G, and generative-AI cobots [10], with AI as the connective layer across the 4.0-to-5.0-to-6.0 progression [11]. For mobile robots the plausible content is lifelong, fleet-wide, privacy-preserving learning: ROS 2-native federated training already updates multi-robot models without sharing raw data [32], with twins for sim-and-deploy [33]. Every generation strengthened one block, yet the seams and missing post-deployment learning stayed put. Table 1 maps this onto the stages.

Table 1. Representative AI models mapped to each stage of the unified autonomy pipeline, with their typical inputs, outputs, and the limitation that still keeps each stage from being a solved block in industrial practice.

Pipeline Stage	Representative AI Models	Typical Inputs	Outputs	Current Limitation	Ref.
Perception	BEVFusion, YOLOv8	Camera, LiDAR/radar clouds	3D boxes, BEV features, masks	Clear-weather bias; accuracy vs. edge budget	[76,77]
Localization & Mapping	ORB-SLAM3, DROID-SLAM, LIO-SAM	Stereo/RGB-D + IMU, LiDAR	6-DoF pose, metric map	Drift, tracking failure in repetitive/dynamic scenes	[13,14,60]
Scene understanding	VLMaps, ConceptFusion	RGB-D + CLIP/foundation features	Open-vocab semantic/queryable 3D map	Web features degrade under factory domain shift	[78,79]
Prediction	Trajectron++, MTR, QCNet	Agent tracks, HD/semantic map	Multimodal trajectories, occupancy flow	Driving-tuned; sparse validation on industrial crowds	[80–82]
Planning	Motion Planning Transformers, DRL navigation	Map/occupancy, goal, range	Path/waypoints/velocity policy	Hard to certify; sim-to-real and sparse-reward gaps	[15,83]

Table 1. *Cont.*

Pipeline Stage	Representative AI Models	Typical Inputs	Outputs	Current Limitation	Ref.
Control	Kanayama tracking law, DWA + MPC	Reference trajectory, robot state	Velocity/actuator commands	Model-based; unmodeled contact, nonholonomy	[43,84]
Visual servoing (perception–control bridge)	ViT-VS, CNS	Live image features/ correspondences	Direct velocity to target	Inner-loop compute; factory-lighting transfer unknown	[85,86]
Decision-making	SayCan, ReAct	Language instruction, skill values, scene text	Skill/action sequence	LLM hallucination; weak grounding to affordances	[16,87]
Lifelong learning	Federated ROS 2 learning, DreamerV3	Multi-robot experience, sensor streams	Updated shared model/world dynamics	Absent from deployed systems; no post-install adaptation	[32,88]

4. AI for Perception

Perception holds most of a deployed robot's AI, and everything downstream inherits its errors. We read it against the sim-to-real gap, the edge budget, and label cost.

4.1. Deep 2D and 3D Detection and Segmentation

Mask R-CNN [89], YOLO [77], and YOLOv8-versus-baseline studies [90] anchor detection. SegFormer [91] and real-time BiSeNet V2 [92] handle free space, while Cityscapes [93] and SemanticKITTI [94] expose the annotation cost driving Section 4.5.

4.2. LiDAR Perception and 3D Representations

PointNet [95] and PointPillars [96] made raw and embedded point-cloud detection practical, and VoxelNeXt went fully sparse, dropping anchors and NMS to top nuScenes tracking [97]. The same features power place recognition [98] and descriptors like MinkLoc3D [99].

4.3. Radar and 4D mmWave Perception

Radar works when cameras and LiDAR fail; surveys of 4D mmWave radar report resilience in fog, rain, and snow but flag point sparsity as the weakness [100]. RadarNeXt gives real-time edge inference from sparse clouds [101], though its value comes mainly through fusion.

4.4. Multimodal Fusion and the BEV Representation

Perception has converged on a shared bird's-eye-view (BEV) frame: BEVFusion [76], TransFusion [102], camera-only BEVFormer [12], and radar-camera RCBEVDet [103] fuse there, and surveys confirm such pipelines degrade less than vision-only ones under fog, rain, and snow [104].

4.5. Semantic Perception and Foundation Vision Models

The newest thread removes the closed label set: SAM segments zero-shot [73], SAM 2 adds video memory [105], DINOv2 features transfer without fine-tuning [74], and CLIP names categories at inference [106]. The payoff is open-vocabulary mapping, where VLMaps grids CLIP-style embeddings for language landmarks [78] and ConceptFusion answers language, image, audio, and click queries over open-set 3D maps [79]. Such maps let reasoning (Section 8) query perception, and the same features reach control via ViT-VS [85]. But zero-shot strength rarely survives a dim aisle and billion-parameter backbones collide

with the edge budget, motivating distillation to MCU-class hardware [35]; surveys cite data scarcity, uncertainty, and real-time execution as barriers [36].

4.6. What Still Limits Perception?

Performance under fog, rain, snow, dust, and low light stays far weaker than on clear datasets, and the accuracy/throughput frontier of PointPillars [96], BiSeNet V2 [92], and VoxelNeXt [97] leaves foundation-scale edge perception open [35,100,104]. Closed-set detectors fail silently and transfer poorly [73,74,78,107,108], dense annotation stays the dominant expense [94,109,110], and BEV fusion [12,76] with open-vocabulary maps [79] give planning-ready representations that most systems still reduce to a brittle box list downstream (Section 5).

Table S2 summarizes how these models trade accuracy against latency, weather, and label cost.

5. AI for Visual Servoing: The Perception–Control Bridge

Visual servoing closes the loop from camera to motion with no map and often no pose, so the link is the control law.

5.1. The Classical Dichotomy and Its Mobile-Robot Complications

Servoing splits into two archetypes [45]. Image-based servoing regulates image features but hits singularities; position-based servoing drives a 3D pose error, cleaner but risking target loss [54]; photometric servoing uses raw pixels for a narrow basin later widened by learning [111].

A nonholonomic base cannot move sideways, so the six-DOF law does not transfer. MPC enforced visibility limits IBVS cannot guarantee [112], and hybrid schemes fold image and Cartesian deviation into one cost [113].

5.2. The Learning Turn

Autoencoder-latent servoing widens the basin [111]. Reinforcement learning dropped the interaction matrix: fitted Q-iteration [114] and DeepMPCVS trained unsupervised at runtime [115].

Newer work learns the controller around correspondences. CNS feeds keypoints to a graph-network controller, sub-millimeter at 40 fps with sim-to-real transfer [86]. ViT-VS drives IBVS from frozen DINOv2 correspondences [74], beating classical IBVS untrained [85]. Mwitta et al. chain detection, DeepSORT, and servoing on a weeder, lifting elimination from 47% to 72.35% [116].

5.3. What Still Limits Visual Servoing

The sim-to-real gap persists: DeepMPCVS lives in simulators [115], and CNS and ViT-VS transfer only by leaning on foundation features [85,86] untested under factory lighting, motion blur, and reflective metal, while a graph-net controller [86] or per-step MPC [117] shares one Jetson with perception and planning, and convergence comparisons stay thin [29,30,114]. The deeper limit is conceptual: a policy tuned on one line does not learn as it runs.

6. AI for Localization and Mapping

Localization and mapping sit mid-pipeline (Figure 1): every downstream stage assumes a trusted pose and map. The geometric core stays classical, learning entered at the edges, and only recently reached the center. Table S3 catalogs the SLAM systems here.

6.1. The Classical Backbone Is Still the Baseline to Beat

A few systems set the bar. ORB-SLAM3 unified visual, visual-inertial, and multi-map SLAM at centimeter accuracy [13]; DM-VIO's delayed marginalization let one camera and IMU beat stereo-inertial methods [118]; VINS-Mono set the tightly coupled sliding-window formulation [119]. On LiDAR, LOAM split odometry from mapping [44], LIO-SAM added a GPS and loop-closure factor graph [60], FAST-LIO2 used raw-point registration [120], and FAST-LIVO2 added a camera for degraded scenes [121]. These run on modest CPUs and rarely fail catastrophically.

6.2. Learning Enters the Back-End, Then the Whole Pipeline

DROID-SLAM made the front-end differentiable, updating pose and depth in a Dense Bundle Adjustment layer with far fewer catastrophic failures [14]. Deep multimodal visual SLAM now matches single-sensor LiDAR at lower cost [19].

6.3. Neural, Gaussian, and Semantic Maps

The map itself is now learned. iMAP used a single MLP as the scene [122]. 3D Gaussian splatting swaps the MLP for an explicit cloud: Gaussian Splatting SLAM ran live monocular tracking [123] and SplaTAM improved novel-view synthesis on RGB-D [124]. These maps suit digital twins but stay GPU-hungry.

Semantics link mapping to understanding. SuMa++ labeled a surfel map to filter dynamic objects [125] and Kimera built a labeled 3D mesh on CPU [126]. Foundation models pushed this open-vocabulary: ConceptFusion [79] and VLMaps [78] fuse features and embeddings for language navigation, both on CLIP [106] (Section 4). Semantic and metric layers stay separate.

6.4. Point-Cloud Place Recognition and Relocalization

Recovering global pose without a prior is where learned descriptors have won. MinkLoc3D added sparse voxelization and hard-negative mining [99]; BEVPlace used rotation-equivariant bird's-eye features [98].

6.5. Remaining Challenges

Neural-implicit and Gaussian maps run at single-digit frame rates, off an embedded budget [123,124], and most representations assume a static world, degrading in cluttered warehouses [125–127]. Lifelong mapping is the deepest gap: repeated-route datasets expose seasonal drift [128,129], yet almost no deployed system continuously updates its map [78,79].

7. AI for Prediction and Planning

In the unified pipeline (Figure 1) prediction and planning are the tightest coupling, yet most planners get little future.

7.1. Predicting How the World Will Move

Generative and transformer models replaced pooling [80–82], with diffusion too slow for real-time control [130,131]; dense traffic uses occupancy instead [132] beside the BEV perception of Section 4 [12,76].

7.2. From Classical Planners to Learned Navigation

Learning accelerated classical planning rather than replacing it: roadmaps [41], asymptotically optimal RRT/*PRM*/RRG [69], and the Dynamic Window Approach [42] stay certifiable, and reviews sort planners into families [21,22]. Learning first guided search [15,133],

then attacked navigation directly [70], the DQN, DDPG/TD3, and PPO families recur with sparse reward and sim-to-real as standing failures [24,50], aided by LiDAR channels [83].

7.3. Social Navigation and the Prediction–Planning Seam

Crowd navigation is where prediction and planning stop being separable, moving from one-step prediction in a value network [71] to SARL [68]. Central to this thesis, feeding an explicit learned crowd-motion predictor into the loop improves success and safety [134]. SafeCrowdNav [135] adds collision-aware losses, yet a benchmark shows performance hinges on training realism and reward design more than architecture [136], while the socially-aware-navigation taxonomy still boxes prediction and motion generation apart [58].

7.4. Coupling, and Why Learned Planners Break

The strongest systems join learned and analytic worlds: DRL deciding while MPC enforces constraints [137], DWA-plus-MPC path-tracking [84], a risk-aware PPO agent under a control-barrier filter [138], FusionPlanner [139], and DRL planners for AUVs [140]. End-to-end policies absorb their simulator's dynamics, then degrade across the reality gap [24], with Habitat finding flat RL brittle where hierarchical policies held up [31]. A black-box policy gives no certificate for the velocity it chose, fatal under ISO 3691-4 (Section 12), so classical and hybrid planners still own the floor.

7.5. What Still Limits Prediction and Planning

Real-time multimodality is unsolved on an edge budget already spent on perception and SLAM [131], and sim-to-real transfer stays unreliable [50,80,134,136]. Data for forklifts and mixed traffic barely exists, so predictors import wrong priors from driving corpora [22,135], and no deployed planner yet ingests calibrated multimodal forecasts in real time under a safety case (Table S5).

8. AI for Control and Decision-Making

Control and decision-making sit at opposite ends of an abstraction ladder (Figure S3): a controller turns a trajectory into actuator commands under hard stability guarantees, while a planner reasons about goals and recovers from failure. This section spans the seam.

8.1. Low-Level Control

Kinematic tracking for wheeled robots was settled decades ago: the Kanayama controller [43] set the Lyapunov-stable baseline, sliding-mode added finite-time robustness [141], and Lyapunov-based nonlinear control stays dominant because it comes with proofs [142]; recent variants add adaptive sliding mode, backstepping observers, and fixed-time integral SMC [143–145]. Model predictive control folds dynamics and obstacle constraints into one optimization [146]; a finite-horizon optimizer is not automatically stable, so Lyapunov-derived input constraints add asymptotic stability [147] and robust MPC uses an LMI min-max for recursive feasibility [148], at a cost possibly too slow for embedded hardware [117] (Table S6). For hard-to-model dynamics, actor-critic RL learns a tracking policy online [149], and an end-to-end policy fusing proprioception and exteroception carried ANYmal through an Alpine hike [49]; because such policies train in simulation, the sim-to-real gap binds [25].

8.2. Active-Force, Fault-Tolerant, and Formation Control

Active force control cancels disturbance at the actuator via resolved-acceleration control with inertia-matrix estimation [150]. Fault tolerance is unavoidable in a fleet: adaptive compensators accommodate loss-of-effectiveness faults with no separate detection

module [151], and prescribed-performance schemes constrain error within funnels as actuators degrade [152]. As learned components enter, verification is the bottleneck: neural barrier certificates give decentralized safety, but scalable verification of learned multi-robot controllers stays open [153,154].

8.3. The Planning–Control Seam

The weakest joint is the planner-to-controller handoff under real dynamics. The industrial answer is hierarchical: RL-plus-MPC hybrids get collision-free trajectories with better sample efficiency than pure end-to-end RL [137], and a DWA-plus-MPC hierarchy embeds path-tracking constraints for shared-space navigation [84]. Visual servoing is the other bridge, increasingly cast as MPC so field-of-view and actuator limits become explicit constraints [112,113]. These hybrids work, but the split of authority stays hand-engineered.

8.4. High-Level Decision-Making

The behavior tree is the classical reactive substrate, having displaced finite state machines through modular composition and failure reaction. Large language models changed the top of the ladder, where grounding matters: SayCan multiplies LLM likelihood against learned skill value functions [16], Code-as-Policies and ProgPrompt trade language for code with assert-based precondition checks on real arms [155,156], Inner Monologue feeds success detection back into the prompt [157], and ReAct interleaves reasoning with actions to beat baselines on ALFWorld [87].

Two structural ideas address the worst failures. Formal–planner hybrids restore feasibility, with LLM+P translating the problem into PDDL for a sound solver [158], Text2Motion adding geometric feasibility checks via skill Q-functions [159], and neuro-symbolic Teriyaki interleaving planning and execution for verifiable plans [160]. Multi-robot work follows, with dialectic LLM agents negotiating collision-refined waypoints [161] and programmatic prompting decomposing instructions into fleet task allocation [55,162], where hallucination and grounding stay persistent challenges [163].

8.5. Foundation Policies and the Control–Decision Merger

Vision–language–action (VLA) models tokenize robot actions as text and co-train them with web-scale vision-language data (Table S7). The RT line set the recipe: RT-1 ran language-conditioned instructions [164], RT-2 generalized to novel objects [38], and PaLM-E injected sensor observations into an LLM for cross-embodiment transfer [75]. Openness and efficiency then improved: Open X-Embodiment pooled 22 embodiments [47], OpenVLA opened a quantized 7B model [39], Octo finetuned to new sensors on one GPU [165], VoxPoser composed 3D value maps [166], and Gemini Robotics paired a controller with an embodied-reasoning model [40]. World models such as DreamerV3 and Cosmos learn dynamics a planner rolls out in imagination (Section 14) [88,167]. Surveys name the barriers: data scarcity, safety and uncertainty quantification, and real-time execution [36].

8.6. What Still Limits Control and Decision-Making

Learned controllers and LLM planners carry no proofs where low-level control has decades of stability theory [142,154], and scalable formal verification of learned policies stays unsolved even in the multi-robot case [153,168]. Since an LLM can fabricate a non-existent skill, credible systems wrap it in classical planners, feasibility checks, or behavior trees [158,159,163], and how a slow deliberative brain and a millisecond reflexive spine share authority [137] stays open (Tables S6 and S7), a cross-cutting gap Section 13 treats.

9. Comparative Analysis of AI Families Across the Autonomy Pipeline

We compare the AI *families* (convolutional networks, Transformers/ViTs, reinforcement and deep RL, diffusion policies, LLM task planners, VLA and foundation models, neuro-symbolic methods) along the axes that decide deployment. Readiness is a Technology Readiness Level (TRL); the rest are ordinal High/Medium/Low ratings tied to a cited result. Ratings are *per-role*, since a CNN detector and an LLM planner sit at opposite ends of the stack.

To keep the comparison from reading as opinion, every cell in Figure 4 was assigned with the explicit rubric of Table 2: each rating maps an observable, cited result onto a fixed three-level scale, and TRL follows the ISO 16290/NASA scale [169] anchored to the strongest demonstrated evidence for that family in that role. The ratings are ordinal and per-role, not cardinal scores to be averaged across a row.



Figure 4. Qualitative comparison of AI families across compute, latency, robustness, interpretability, real-time capability, data efficiency, and deployment readiness (TRL). For compute cost and latency, higher values are worse; for the remaining criteria, higher is better.

Table 2. Scoring rubric for the AI-family comparison of Figure 4. Each family–role cell is assigned by mapping a cited, observable result onto the three-level scale below. For compute cost and latency a *higher* level is *worse*; for the remaining criteria a higher level is better.

Criterion	Low	Medium	High
Compute cost (higher = worse)	Runs within an embedded/edge budget (Jetson-class or below) at full rate	Needs a discrete mobile GPU	Needs server/multi-GPU or a >7B-parameter model
Latency (higher = worse)	<50 ms/inference, i.e., control-loop rate	50–500 ms	>500 ms, up to seconds (e.g., LLM planning)
Robustness	Degrades sharply off-distribution or under adverse weather/unseen sites	Tolerates moderate domain shift via fusion or fine-tuning	Graceful degradation, corroborated on independent benchmarks

Table 2. *Cont.*

Criterion	Low	Medium	High
Interpretability	Black-box end-to-end policy	Partial structure (attention maps, modular sub-tasks)	Human-readable/symbolic or certificate-backed reasoning
Real-time capability	Offline or seconds-scale	Sub-second but below control rate	Meets the sensor/control-loop rate on target hardware
Data efficiency	Needs large per-site labeled corpora	Transfers with modest fine-tuning	Zero/few-shot or frozen-feature transfer
Deployment readiness (TRL)	TRL 3–4: proof-of-concept in the lab	TRL 5–7: validated on a relevant platform in a relevant environment	TRL 8–9: fielded across an industrial fleet

Convolutional networks are the deployment workhorse and the only family unambiguously high-TRL across an industrial fleet. Compute and latency are favorable: PointPillars runs embedded 3D detection at high frame rates [96], and YOLOv8's anchor-free head suits Jetson-class platforms [77,90]. Vision-only CNNs degrade under fog, rain and snow, hence multimodal fusion [100,104]; the soft spot is data efficiency, needing large labeled corpora [93,109]. For perception and visual servoing, CNNs sit at TRL 8–9 but do not plan or reason.

Transformers and ViTs buy accuracy and multimodal flexibility at a compute premium. BEVFormer lifts camera-only detection [12], TransFusion's LiDAR–camera fusion is robust to poor illumination [102], and VoxelNeXt predicts boxes from sparse voxels [97]. SegFormer and BiSeNet V2 push toward real-time segmentation [91,92]; in prediction, MTR won the 2022 Waymo challenge [81] and QCNet ranked first on Argoverse 1 and 2 [82]. The tradeoff is quadratic attention pressuring the edge budget. TRL 7–8 for perception and prediction, lower as the backbone is asked to act.

Reinforcement and deep RL are most associated with control and local planning, and its readiness is the most overstated. Demonstrations span a mapless DRL planner transferred to unseen environments [70], SARL's crowd model on CrowdNav [68], and ANYmal's alpine hike under a learned locomotion policy [49]. Yet surveys name sim-to-real and data efficiency as recurring bottlenecks [23], and sparse rewards and transfer as the dominant open problems for path planning [50]; DayDreamer's quadruped walks in about an hour only via a world model [170]. Interpretability is low and robustness brittle off-distribution, but inference latency is excellent. Low-to-medium TRL: the reality gap and absent formal guarantees keep most learned controllers in pilots [25,52].

Diffusion policies are newer, rated mainly in prediction and emerging manipulation/control. MID casts trajectory prediction as reverse motion-indeterminacy diffusion [130], and Leapfrog skips most denoising steps to recover real-time inference [131]. For control, Pi0 attaches a flow-matching action expert to a VLM [171], and imitation-learning surveys treat diffusion policy as a first-class action representation [172]. Iterative sampling cost keeps latency medium at best. TRL 4–6.

LLM task planners sit at the decision layer, turning instructions into structured plans. SayCan grounds an LLM in learned skill value functions [16], ProgPrompt reaches state-of-the-art on VirtualHome via Pythonic plans [156], Text2Motion lifts long-horizon success via feasibility heuristics [159], and LLM+P offloads to a classical PDDL solver where vanilla LLMs fail [158]. Weaknesses are compute, latency (seconds, not milliseconds), and grounding/hallucination, the key open challenge [163]. The reasoning chain is human-readable natural language [87,157]. TRL 4–5.

VLA and foundation models are the most general and least mature for industrial duty. OpenVLA runs quantized inference at a fraction of larger models' size [39], RT-2 transfers web knowledge into emergent control [38], Open X-Embodiment shows positive cross-embodiment transfer from pooled data [47], and Octo finetunes to new robots in hours on one consumer GPU [165]. Data scarcity, uncertainty quantification and real-time execution remain the principal barriers [36]. TRL 3–5, at the decision/whole-stack level in Figure 4.

Neuro-symbolic methods are the smallest family but the most relevant to the interpretability and certification gaps. Teriyaki solves most of a 1000-problem planning benchmark and yields shorter plans than classical symbolic planners [160]; learned-certificate work (neural Lyapunov, barrier, contraction) offers a path to formal guarantees on learned controllers [153,154]. Interpretability is high by construction, but the cost is engineering effort and brittleness when the abstraction breaks. TRL 3–4.

Three constraints cut across the table. The **sim-to-real gap** hits the learning-heavy families hardest (RL, diffusion, VLA) and barely touches CNN perception [25,52,173]. The **edge/compute budget** inverts the accuracy ranking: the families highest on capability score lowest on deployability, hence TinyDL-style compression and quantized VLAs [35,39]. **Data availability** rewards the foundation/cross-embodiment route [47] and punishes anything needing per-site labels [109].

No single family spans the stack: CNNs perceive without reasoning, LLMs reason but cannot perceive in real time, and RL policies come undone in a new plant. The low-readiness regions sit at the *seams* between blocks and at the lifelong-learning column, empty for every family. Section 13 takes up these seams and that gap.

10. Industrial and Service Applications, Organized by Required AI Capability

Sorting applications by the AI capability that decides success shows where the pipeline breaks. Each domain leans on one or two blocks and exposes the weak seam between them.

Warehouse and logistics: planning and coordination. The one profitable domain earned that by solving coordination. Kiva coordinated engineered floors [1]; the frontier moved to AMRs that decide onboard [2]. Conflict-based search wins for small fleets while learned multi-agent RL scales with agent count [174], and hierarchical MARL beats order-picking heuristics [72] to meet Industry 5.0's human-centric pillar [5]. A planner optimal for ten robots is intractable for a thousand.

Inspection, mining, and subterranean: perception and robustness. YOLOv8-class detectors carry power-line defect detection [175], unsupervised methods target scarce defects on MVTec AD [176], Spot ran AutoInspect inside the Joint European Torus [177], and LiDAR UAVs plan wind-turbine trajectories online [178]; the binding constraint is domain shift [107]. Reaching stepped or multi-level structures a wheeled base cannot forces a different platform: stair-climbing robots that recognize and gauge each step before committing to the climb [179]. A neighboring role is environmental monitoring, where multi-pollutant air-quality prediction inside industrial buildings leans on adaptive sensor fusion to hold accuracy as individual gas sensors drift [180]. Underground, dust and darkness defeat perception and localization at once [181]. Fused 3D LiDAR and SLAM holds pose in a mine [182], FusionPlanner couples LiDAR and camera on MiningNav [139], and CERBERUS won the DARPA Subterranean Challenge by fusing camera, LiDAR, and inertial data [48], the same complementarity that resists fog in driving [104].

Healthcare and agriculture: interaction and drift. Hospitals stress safe, legible interaction, yet benchmarks stay rare: one measures safety margins around staff [183], and nursing robots span guiding, disinfection, and transfer [184]. Deployment breaks at

the planning-HRC seam, where pedestrian prediction beats reactive avoidance [68,134] inside the ISO/TS 15066 envelope [8] against which human detectors must be judged [63]. Agriculture fuses perception with degraded GNSS: GREENBOT benchmarks greenhouse SLAM [127], RTK-GPS fuses with visual crop-row following when fixes drop [185], a laser weeder combines detection, DeepSORT tracking, and visual servoing [116], and steer-and-drive robots learn DRL policies [186].

Connected vehicles, underwater, and aerospace. V2V4Real frames cooperative detection and Sim2Real as open problems [187] on the BEV machinery of Sections 4 and 7 [12,81], bounded by certifying a black-box stack. Underwater DRL planners must reach goals without GPS or a dynamics model [140,188]. On-orbit servicing exercises the perception–control bridge via visual servoing, with no second attempt and no large real dataset [189]. Table 3 summarizes the mapping.

Table 3. Industrial and service application domains for mobile robots mapped to the autonomy-pipeline blocks they stress most, the dominant deployment constraint in each, and representative systems or methods.

Domain	Primary AI Tasks	Dominant Challenge	Representative Methods/Systems	Ref.
Warehouse/logistics	Multi-robot path finding, fleet coordination, onboard SLAM	Coordination at fleet scale (planning↔coordination seam)	Kiva 500+ units; CBS vs. MARL; hierarchical MARL picking	[1,2,72,174]
Inspection	Defect/object detection, anomaly detection, autonomous trajectory generation	Data availability under cross-domain shift	YOLOv8 power-line; MVTec AD; AutoInspect (Spot)	[107,175–178]
Mining/subterranean	3D detection, robust localization, multi-sensor fusion planning	Perceptual degradation; robustness as integration property	3D LiDAR+SLAM mine; FusionPlanner (MiningNav); CERBERUS (SubT)	[48,139,181,182]
Healthcare/hospitals	Pedestrian-aware navigation, human detection, HRC	Safe, certifiable interaction in human spaces	Hospital nav benchmark; nursing-robot survey; ISO/TS 15066 envelope	[8,63,68,183,184]
Agriculture	Crop/weed detection, tracking, visual servoing, GNSS-degraded localization	Perception drift; sensor fusion under GNSS loss	Laser-weeding (DeepSORT+servoing); RTK-GPS+visual rows; GREENBOT SLAM	[116,127,185,186]
Autonomous/connected vehicles	BEV detection, motion prediction, cooperative perception	Safety certification of black-box full stack	V2V4Real; BEVFormer	[12,187]
Underwater/aerospace	DRL path planning, visual servoing for approach/contact	Sim-to-real gap; no localization infra, no second attempt	DRL AUV planner; adaptive AUV path/control; on-orbit visual servoing	[140,188,189]

11. Datasets, Benchmarks and Evaluation

Autonomy is scored against benchmarks built one pipeline block at a time, so the seams mark where it is weak.

11.1. The Benchmark Stack, Block by Block

Perception has the most mature data. KITTI [190], nuScenes [109], Waymo Open [191], SemanticKITTI [94], and Cityscapes' mean-IoU convention [93] feed BEVFormer [12], PointPillars [96], and VoxelNeXt [97].

Localization and mapping report trajectory error on EuRoC [192], year-long Oxford RobotCar [129], and multi-season Boreas [128]. ORB-SLAM3 reports centimeter ATE on EuRoC [13] and DROID-SLAM fails far less often [14], but that says nothing about whether the map survives downstream.

Embodied simulators sit nearest industrial constraints. Habitat 2.0 reached 25,000+ steps/s, where hierarchical RL beat sense–plan–act pipelines (Section 7) [31,193]. Matterport3D [194], RoboTHOR [108], and BEHAVIOR-1K [195] extend scenes and activities. Manipulation adds ManiSkill2 [46] and Open X-Embodiment, whose 22 embodiments transferred once formats agreed [47]. Industrial sets exist thinly: MVTec AD [176], GREENBOT [127], V2V4Real [187], UniOcc [132], and one hospital benchmark [183] are isolated points.

11.2. What Industrial Mobile Robots Actually Need to Measure

No lifelong benchmark. Every dataset scores a fixed-weight model on a held-out split. Oxford RobotCar and Boreas span a year but test seasonal change, not whether a robot keeps learning without forgetting, so continual learning goes unmeasured and undeployed.

Per-block, never end-to-end. Nothing scores the handoffs, whether a perception output yields a safe plan or a map survives a planner. Habitat 2.0's hierarchical-versus-flat result is a rare cross-block hint.

Unstandardized safety. ISO 3691-4 [37] and ISO/TS 15066 [8] set person-detection and biomechanical limits, yet no dataset scores a learned planner against those clauses. SafeCrowdNav [135] and risk-aware barrier schemes [138] use their own protocols, and none enumerates the SOTIF unsafe-known/unknown split of ISO 21448 [110].

Scarce real-deployment data. Month-scale inspection runs and a degraded-underground traverse (Section 10) [48,177] are not shared benchmarks. Surveys flag cross-domain detector degradation [107], open sim-to-real transfer [25,52], and data-scarce foundation-model deployment lacking uncertainty-aware evaluation [36]. Table S4 in the Supplementary Materials catalogs these datasets.

11.3. Reported Compute, Latency, Energy, and Validation Across the Pipeline

Pulling the reported figures for representative systems into one place (Table 4) exposes a pattern that no single-block benchmark shows. Compute and accuracy are reported almost everywhere; latency is reported often but on heterogeneous hardware, so cross-system comparison is unsafe; and *energy per inference is essentially never reported* (NR below), even though the on-robot power budget is one of the three constraints this review foregrounds. Real-world validation thins sharply from perception (dataset-scored) to decision-making (few real-robot trials), and safety is validated against bespoke protocols rather than the ISO clauses of Section 12. The blank cells are the finding: the field cannot presently be compared on the axes an industrial integrator actually cares about, which is itself an argument for the standardized, end-to-end, energy-and-safety-aware benchmark this review calls for.

Table 4. Reported hardware, latency, energy, real-world validation, and dominant failure mode for representative systems across the autonomy pipeline. Values are as stated by the cited source; NR = not reported. Latencies are on heterogeneous hardware and are not directly comparable. The pervasive NR columns, especially energy, are discussed in the text.

Representative System (Role)	Reported Hardware/Model Size	Latency or Throughput	Energy/ Inference	Real-World Validation	Dominant Failure Mode	Ref.
PointPillars (3D detection)	Desktop GPU (GTX 1080 Ti)	62 Hz	NR	KITTI dataset (no fleet trial)	Single-class AP; degrades in fog/rain/snow	[96,100]

Table 4. *Cont.*

Representative System (Role)	Reported Hardware/Model Size	Latency or Throughput	Energy/ Inference	Real-World Validation	Dominant Failure Mode	Ref.
BEVFormer (camera BEV detection)	Server GPU	NR (attention-heavy)	NR	nuScenes dataset, 56.9% NDS	Compute premium; no real-fleet validation	[12]
YOLOv8 (2D detection)	Jetson-class edge	Real-time on edge	NR	Field navigation study	Data-hungry; cross-site domain shift	[77,90,107]
ORB-SLAM3 (SLAM)	CPU, no GPU	Real-time	NR	EuRoC (real MAV), ~3.6 cm ATE	Low-texture/dynamic-scene degradation	[13]
FAST-LIO2 (LiDAR-inertial odometry)	Embedded CPU	Real-time, high-rate	NR	Real UAV/UGV runs	Fails without geometric structure	[120]
SARL/ SafeCrowdNav (social nav, DRL)	GPU (training)	Real-time inference	NR	Simulation (CrowdNav); own safety protocol	Brittle off-distribution; not ISO-scored	[68,135]
SayCan (LLM task planner)	Server LLM	Seconds	NR	Real kitchen, mobile manipulator	Grounding/ hallucination	[16]
OpenVLA (VLA policy)	7B params, 4-bit quant, GPU	Below control-rate	NR	Real robot manipulation suite	Data scarcity; real-time execution	[39]
Teriyaki (neuro-symbolic planner)	CPU solver	Plan-wait −61.4%	NR	Simulation, 95.5% of 1000 solved	Brittle when abstraction breaks	[160]

12. Safety, Certification, and Standards

Industry measures robots by a yardstick beyond accuracy: can you prove the machine will not hurt anyone when it fails? A pipeline block whose behavior cannot be bounded makes the whole chain uncertifiable, which keeps foundation models and end-to-end policies at low technology-readiness levels.

ISO 3691-4:2023 governs driverless industrial trucks, setting person-detection performance levels, braking, and test methods a manufacturer must pass before deployment [37]. ISO 26262 covers hazards from electrical and electronic faults, while ISO 21448 (SOTIF) covers hazards that arise when nothing has broken [110]. A misclassifying detector violates no hardware standard yet can drive the robot into a person, so modern perception lives under SOTIF.

Certification assumes traceability from specification to behavior, which learning-based components lack. A Lyapunov-stable controller comes with a proof [43,141]; a deep policy such as OpenVLA comes only with a benchmark number [39]. Rana and co-authors map the ML lifecycle onto the ISO 26262 V-model and SOTIF and find no slot for a data-driven component [196].

Verifiers like Marabou 2.0, driven by VNN-COMP, prove robustness of a small control network but not a billion-parameter VLA [197]. A control-barrier-function safety filter instead projects the action onto the safe set [138,198]. Dawson and colleagues unify neural Lyapunov, barrier, and contraction certificates into one taxonomy [154]; Garg and co-authors flag scalable verification of learned fleet controllers as open [153]. A structured assurance argument then states auditably why the AI is acceptably safe, as in the LAISC methodology [199]. Explainability differs from generic XAI because embodied actions have

physical consequences [200], and almost no work tests whether an explanation changed an assessor's decision [201].

AutoInspect deployments lean on conservative navigation, not foundation-model planners [177], and ISO/TS 15066 governs human–robot collaboration, fixing biomechanical force and pressure limits and four collaborative modes that deep-learning perception must respect [8,63].

This certification gap is the sharpest cross-cutting bottleneck this review identifies. The evidence deep policies generate does not match what ISO 3691-4, ISO 26262, and SOTIF demand, and a larger model will not close it. The most advanced AI arrives wrapped in a classical safety cage, and the cage gets certified.

13. Cross-Cutting Research Gaps

Each block has matured on its own: real-time detectors [77,96], centimeter SLAM [13], planners with optimality proofs [69], learned controllers [149]. But a deployed robot is one closed loop, only as strong as its seams. Six gaps sit between blocks (Figure 5), each hitting the same three constraints: sim-to-real, the edge budget, and scarce labeled data.

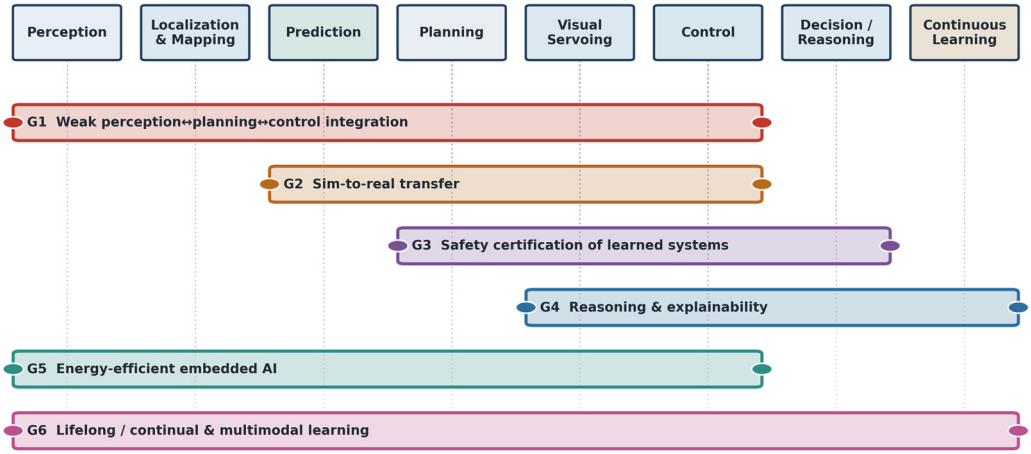


Figure 5. Cross-layer research gaps. Each bar spans the pipeline blocks the gap affects, emphasizing that the most consequential gaps are cross-block.

Perception–planning–control integration remains weak. Most work optimizes one block and treats neighbors as fixed; surveys mirror this, stopping at perception [17,18], taking SLAM as given [23,51], or omitting planning and control [19]. Perception reports accuracy, not how its errors shrink a controller's feasibility set. Barrier-filtered MPC over deep-RL cuts collisions while holding success [137,138]; hybrid MPC-IBVS bridges the gap [45,112,113].

Sim-to-real transfer keeps failing in industrial settings. Sim-trained policies have gone zero-shot to a quadruped [202] and across a 2.2 km alpine route [49], but these wins rely on randomized dynamics needing per-platform hand-tuning [25,34,203]. Environments drift outside the simulator's distribution [127,182], and the gap is perceptual too, since fog, rain, and snow degrade vision more than radar [100,104]. It [50,52,108,193] stays under-treated because real validation data is costly and proprietary. Randomization from small offline datasets [173] and physics-grounded world-foundation models [167] point the way; missing are industrial benchmarks with the coverage Boreas brought to driving [128].

Safety certification of learned systems is immature. ISO 3691-4 sets detection-and-braking levels for driverless trucks [37] and ISO/TS 15066 fixes collaborative force limits [8], but neither anticipates a black-box planner. The deeper problem is SOTIF (ISO 21448): learned perception fails through functional insufficiency in an unenumerated sce-

nario [110,196]. Formal work in robustness verifiers [197], Lyapunov/barrier/contraction certificates [154,198], and assurance cases [199] does not scale, and learned multi-robot controllers stay uncertified [153]. Runtime assurance pairs the policy with a verified safety filter [138].

Reasoning and explainability are still thin. LLM-driven planners decompose instructions, ground them in affordances [16], and self-correct in a closed loop [157]; neuro-symbolic frameworks solve 95.5% of planning problems with shorter plans than classical solvers [160]. Yet reasoning stays brittle, with hallucination and grounding the flagged open challenges [36,163]. Explanation lags too: most robot-explainability methods are untested for their effect on operator trust [200,201], which matters where Industry 5.0 puts a human beside the robot [5,64]. Routing language through feasibility-guaranteeing planners is the path [158,159].

Energy-efficient embedded AI is under-explored. An AMR runs on a battery and a small SoC, yet most models target server GPUs, and foundation-model surveys name real-time execution and onboard compute as principal blockers [36,204]. The budget can be met deliberately, through 4-bit inference [39], single-GPU fine-tuning [165], PointPillars at 62 Hz [96], and TinyML-to-TinyDL quantization, pruning, and NAS for MCU-class hardware [35]. Still, efficiency is reported as FPS rather than a joint accuracy–latency–energy objective, since benchmarks ignore watts.

Lifelong and multimodal learning is largely absent from deployment. Almost every deployed industrial system is trained once and frozen. The ingredients exist in the literature but not on the floor: cross-embodiment models over 22 robot types [47], federated training without raw-data sharing [32,205], world models learning a gait from one hour of real experience [170], and multimodal fusion of vision, language, 3D, and tactile channels [75,204]. The barrier is structural: continual learning risks catastrophic forgetting and resists certification, fleet data is siloed, and no benchmark rewards adaptation. Federated lifelong learning over fleets, inside the runtime safety envelope above, is the frontier this review points toward. Table 5 draws these threads together, stating the maturity of each pipeline block, the cross-cutting challenge that dominates it, and a promising direction grounded in recent work.

Table 5. Pipeline-block maturity, the dominant cross-cutting challenge at each block, and a promising future direction grounded in recent work. Maturity is stated on the TRL rubric of Table 2, taking the strongest fielded evidence for each block.

Pipeline Block	Current State	Main Challenge	Promising Future Direction	Ref.
Perception	Good	Weather/edge robustness; closed-set silent failure; uncertainty-free handoff	Frozen foundation backbones for open-vocabulary, label-free perception passing calibrated confidence	[74]
Localization/SLAM	Good	Semantic-metric fusion; lifelong map maintenance under change	Open-vocabulary semantic maps queryable by a language planner, updated over a deployment's life	[79]
Prediction	Emerging	Real-time multimodal forecasts at the edge, rarely consumed by the planner	Dynamically feasible, plan-conditioned forecasts wired into the planning loop	[80]
Planning	Emerging	End-to-end learned planners are fragile and uncertifiable	Learned search guidance with a classical/MPC safety net	[15]
Control	Mature	Learned controllers lack stability/safety proofs; may not fit the embedded budget	Pair learned policies with certified safety filters and neural Lyapunov/barrier certificates	[154]

Table 5. *Cont.*

Pipeline Block	Current State	Main Challenge	Promising Future Direction	Ref.
Visual servoing	Emerging	Perception–control loop solved per-act; untested under factory lighting/reflective metal	Foundation-feature servoing reusing the same frozen DINOv2 backbone as perception	[85]
Decision-making	Emerging	LLM hallucination and ungrounded plans become safety events on a factory floor	Neuro-symbolic planning routing LLM reasoning through a sound classical/PDDL planner	[160]
Lifelong learning	Weak	Train-once-and-freeze; catastrophic forgetting; siloed fleet data; no certifiable on-job updates	Federated continual learning across robot fleets, rehearsed and validated in digital twins	[32]

14. Future Research Roadmap

Each stage removes a bottleneck the next depends on (Figure 6), ending in the target architecture of Figure 7. Three constraints decide which ideas survive a real factory: the sim-to-real gap, edge compute, and labeled data.

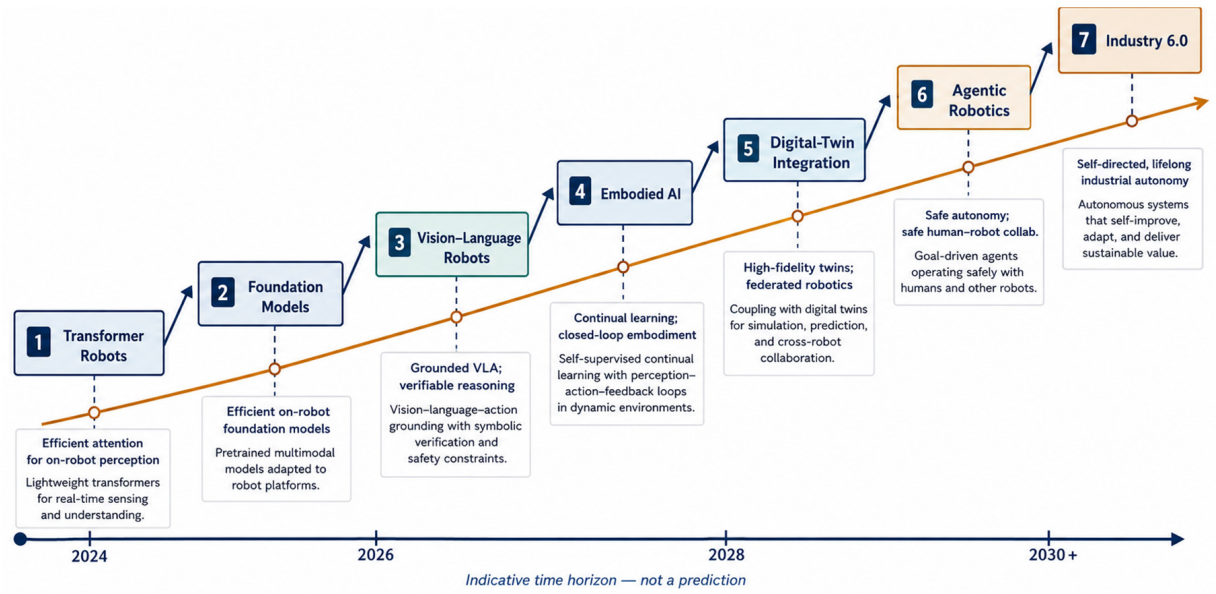


Figure 6. Indicative (not predictive) staged roadmap toward Industry 6.0, with the key technical advance required at each transition.

We label each stage by its evidence tier and, deliberately, attach no dates. Stages 1 and 2 are *evidence-based and near-term*: they extend techniques already demonstrated on robots and could reach industrial practice with engineering effort. Stages 3 and 4 are *emerging*: the components exist in research but their integration on a fielded mobile robot is unproven. Stages 5 through 7 are *speculative*, describing the Industry 6.0 vision as a direction of travel, not a forecast. The ordering is logical dependency, not a predicted timeline; readers should treat the later stages as hypotheses about what integration would require, and the whole roadmap as indicative rather than predictive.

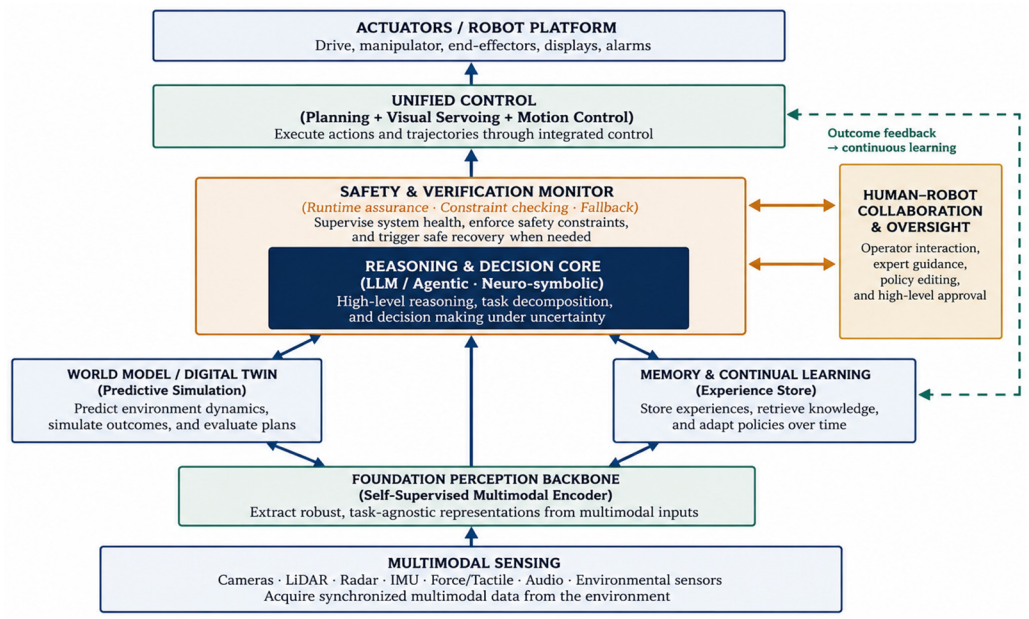


Figure 7. Proposed next-generation AI architecture for industrial mobile robots, integrating a foundation perception backbone, world model/digital twin, continual-learning memory, a reasoning and decision core wrapped by a safety monitor, and unified control. Solid arrows show the upward data path from sensing to actuation; double-headed arrows show two-way exchange with the world model and the memory store; orange double-headed arrows show human oversight; and the dashed arrow is the outcome-feedback loop that drives continual learning.

Stage 1—Transformers as the connective tissue. *(Evidence-based, near-term.)* Attention already replaces bespoke architectures across BEV perception [12], semantic segmentation [91], differentiable-SLAM [14], and motion forecasting [81]. This stage owes industry efficiency: distillation, quantization, and architecture search along the TinyML-to-TinyDL line [35], with detectors trading accuracy for latency [77], until the stack fits an AMR's envelope.

Stage 2—From per-block transformers to foundation models. *(Evidence-based, emerging.)* One model replaces one network per task, and frozen self-supervised features transfer without fine-tuning: DINOv2 [74], SAM 2 [105], and CLIP [106]. Cross-embodiment transfer [47] and OpenVLA's within-budget policy [39] point toward the missing *on-robot* foundation model, adapting to new sensors via parameter-efficient tuning [165]. Blockers are data scarcity, uncertainty quantification, and real-time execution [36]; scarcity is sharpest, motivating world models that synthesize physics-grounded data [167,170].

Stage 3—Vision-language robots. *(Emerging.)* Language becomes the tasking interface: RT-2 and PaLM-E fold web knowledge and sensor observations into control [38,75]. This yields open-vocabulary navigation: VLMaps fuses CLIP into a spatial grid [78], ConceptFusion builds language-queryable open-set 3D maps [79], and NaVILA emits metric commands an RL policy executes [168]. Grounding without hallucination is the open problem [163]; an ungrounded instruction can steer a vehicle towards a worker.

Stage 4—Embodied AI and closing the perception–action loop. *(Emerging to speculative.)* The robot must act under contact, latency, and partial observation. Pi0 adds a flow-matching action expert on a VLM backbone [171], 3D-VLA couples a 3D language model to embodied diffusion [206], and DreamerV3 masters many tasks from one configuration [88], yet BEHAVIOR-1K's long-horizon activities stay beyond current learning [195] and flat RL stays brittle next to hierarchical policies [31]. Two advances gate this stage: lifelong learning, since deployed systems never learn after commissioning (the largest gap this review identifies); and sim-to-real, the recurring DRL bottleneck [23], via domain

randomization from a small real dataset [173,203] and zero-shot transfer on hardware like ANYmal [49,202].

Stage 5—Digital-twin integration. *(Speculative.)* A bidirectionally synchronized twin gives lifelong learning somewhere safe to rehearse a policy update, then validate it against hazards too dangerous to stage. Platforms already synchronize physical and virtual state in real time [207], and reviews map that exchange onto autonomy [33]. The twin is where Industry 5.0 and 6.0 converge [61], running over 6G [10]. Its fidelity must be certified as safety-case evidence, connecting to the SOTIF functional-insufficiency split [110], ISO 26262 and 21448 [196], and ISO 3691-4 AGV person-detection and braking [37].

Stage 6—Agentic robotics. *(Speculative.)* A grounded, twin-validated policy can now take autonomy over its own task structure, planning, acting, and self-correcting: ReAct beats imitation and RL baselines [87], Voyager runs a self-verifying skill curriculum [208], and Inner Monologue closes the loop with success detection [157]; for teams, RoCo coordinates arms through LLM dialogue [161] and SMART-LLM handles task decomposition [162]. Verifiable reasoning is the missing piece. Routing problems through a classical PDDL planner [158,160], then pairing learned reasoning with verifiers [197] and learned certificates such as neural Lyapunov and barrier functions [153,154], is the most credible path to certifiable autonomy. Federated learning supplies the distribution layer, a fleet that learns without exporting plant data [32,205].

Stage 7—Toward Industry 6.0. *(Speculative, long-horizon vision.)* The endpoint is speculative: hyper-customized, sustainable production on automated twins, 6G, AI cobot networks, and the Internet-of-Anything [10], with AI as the connective layer across the 4.0→5.0→6.0 progression [11]. The deliverable is a fleet that perceives, reasons, acts, and *keeps learning* under verifiable safety, Industry 5.0's human-centric pillar [5] enforced by ISO/TS 15066 perception [8] and proximity benchmarks [183].

What matters is what each interface carries. A frozen backbone emits a shared representation consumed without a hand-tuned adapter; the twin returns rollouts trusted only when twin-predicted and on-hardware trajectories agree; a lifelong-learning memory writes back corrected behavior without forgetting; a verification gate stands between every learned decision and any actuator; and a federated link shares updates, not raw data.

15. Conclusions

This review read the field through one lens: a single autonomy pipeline running Sensors → Perception → Localization and Mapping → Scene Understanding → Prediction → Planning → Control → Decision-Making, with Lifelong Learning meant to close the loop. Organizing the work by autonomy function rather than algorithm class exposed uneven blocks, with progress mapping onto where the easy wins were.

Perception and localization are mature by research standards; the connective tissue between blocks is not. Camera-only BEV transformers approach LiDAR-grade detection [12], multimodal fusion degrades gracefully under fog and rain [104], and foundation backbones such as DINOv2 and Segment Anything transfer zero-shot to unseen scenes [73,74]. What stays fragile are the interfaces: perception hands planning a representation it was not designed to consume, planning hands control trajectories that ignore actuator and stability limits, and control hands decision-making a state high-level reasoners cannot ground reliably [163].

The second gap is the near-total absence of lifelong learning in fielded systems. Cross-embodiment transfer is real [47], and world models let robots learn on physical hardware in about an hour [170], yet deployed AMRs and AGVs still ship frozen policies. Three constraints explain why: the sim-to-real gap, the edge compute budget, and the scarcity of labeled industrial data. These are one problem seen through perception, compute, and data.

The path to Industry 6.0 follows from this diagnosis. The next gains come from tightening the seams: representations that survive the perception-to-planning handoff, controllers that expose their feasibility envelope upward, grounded reasoners, and learning that continues after commissioning. Foundation models, embodied agents, neuro-symbolic planning, digital twins, and human-centric collaboration under Industry 5.0 values [5] are the ingredients, but only if integrated as a stack. Certification adds a deferred constraint: a black-box policy that cannot explain a stop is hard to field where ISO collaborative-safety limits apply [8].

We offered this review as a map. The real frontier sits in the connective tissue and in the missing capacity to keep learning after deployment. Industrial mobile robots will become genuinely intelligent when the pipeline behaves as one adaptive whole.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/machines14080950/s1, Table S1: Coverage of prior surveys against the autonomy pipeline; Table S2: Representative perception and computer-vision models; Table S3: Representative SLAM and localization systems; Table S4: Representative datasets and benchmarks; Table S5: Representative path- and motion-planning methods; Table S6: Controller-family comparison; Table S7: Foundation and vision–language–action models; Table S8: Verbatim search strings by pipeline block. Figure S1: Function-based AI taxonomy; Figure S2: PRISMA-style literature-selection flow; Figure S3: Decision-abstraction hierarchy; Figure S4: Applications-to-capabilities map.

Author Contributions: Conceptualization, methodology, investigation, and writing—original draft preparation, M.N.A.R.; writing—review and editing, supervision, and project administration, M.A.H.A.; writing—review and editing and supervision, N.N.N.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wurman, P.R.; D'Andrea, R.; Mountz, M. Coordinating Hundreds of Cooperative, Autonomous Vehicles in Warehouses. *AI Mag.* **2008**, *29*, 9–19. [CrossRef]
2. Keith, R.; La, H.M. Review of Autonomous Mobile Robots for the Warehouse Environment. *arXiv* **2024**, arXiv:2406.08333.
3. Hermann, M.; Pentek, T.; Otto, B. Design Principles for Industrie 4.0 Scenarios. In Proceedings of the 2016 49th Hawaii International Conference on System Sciences (HICSS), Koloa, HI, USA, 5–8 January 2016. [CrossRef]
4. Lee, J.; Bagheri, B.; Kao, H.-A. A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems. *Manuf. Lett.* **2015**, *3*, 18–23. [CrossRef]
5. Breque, M.; De Nul, L.; Petridis, A. *Industry 5.0: Towards a Sustainable, Human-Centric and Resilient European Industry*; Directorate-General for Research and Innovation (Policy Brief/White Paper); European Commission: Brussel, Belgium, 2021.
6. Xu, X.; Lu, Y.; Vogel-Heuser, B.; Wang, L. Industry 4.0 and Industry 5.0—Inception, conception and perception. *J. Manuf. Syst.* **2021**, *61*, 530–535. [CrossRef]
7. Leng, J.; Sha, W.; Wang, B.; Zheng, P.; Zhuang, C.; Liu, Q.; Wuest, T.; Mourtzis, D.; Wang, L. Industry 5.0: Prospect and retrospect. *J. Manuf. Syst.* **2022**, *65*, 279–295. [CrossRef]
8. *ISO/TS 15066:2016*; Robots and Robotic Devices—Collaborative Robots. ISO Technical Specification. International Organization for Standardization (ISO): Geneva, Switzerland, 2016.
9. Nahavandi, S. Industry 5.0—A Human-Centric Solution. *Sustainability* **2019**, *11*, 4371. [CrossRef]
10. Chander, B.; Bahrami, A.; Badawi, M.; Elsamak, G.; Emara, M.; Abdallah, A.M. Industry 6.0: Vision, technical landscape, and opportunities. *Ain Shams Eng. J.* **2025**, *16*, 103354. [CrossRef]

11. Fernández-Miguel, A.; García-Muiña, F.E.; Ortíz-Marcos, S.; Jiménez-Calzado, M.; Fernández del Hoyo, A.P.; Settembre-Blundo, D. AI-Driven Transformations in Manufacturing: Bridging Industry 4.0, 5.0, and 6.0 in Sustainable Value Chains. *Future Internet* **2025**, *17*, 430. [CrossRef]
12. Li, Z.; Wang, W.; Li, H.; Xie, E.; Sima, C.; Lu, T.; Qiao, Y.; Dai, J. BEVFormer: Learning Bird's-Eye-View Representation from Multi-Camera Images via Spatiotemporal Transformers. In *European Conference on Computer Vision (ECCV)*; Springer: Cham, Switzerland, 2022. [CrossRef]
13. Campos, C.; Elvira, R.; Rodriguez, J.J.G.; Montiel, J.M.M.; Tardos, J.D. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap SLAM. *IEEE Trans. Robot.* **2021**, *37*, 1874–1890. [CrossRef]
14. Teed, Z.; Deng, J. DROID-SLAM: Deep Visual SLAM for Monocular, Stereo, and RGB-D Cameras. *Adv. Neural Inf. Process. Syst. (NeurIPS)* **2021**, *34*, 16558–16569.
15. Johnson, J.J.; Kalra, U.S.; Bhatia, A.; Li, L.; Qureshi, A.H.; Yip, M.C. Motion Planning Transformers: A Motion Planning Framework for Mobile Robots. *arXiv* **2022**, arXiv:2106.02791.
16. Ahn, M.; Brohan, A.; Brown, N.; Chebotar, Y.; Cortes, O.; David, B.; Finn, C.; Fu, C.; Gopalakrishnan, K.; Hausman, K.; et al. Do As I Can, Not As I Say: Grounding Language in Robotic Affordances. In Proceedings of the Conference on Robot Learning (CoRL), Auckland, New Zealand, 14–18 December 2022.
17. Mao, J.; Shi, S.; Wang, X.; Li, H. 3D Object Detection for Autonomous Driving: A Comprehensive Survey. *Int. J. Comput. Vis.* **2023**, *131*, 1909–1963. [CrossRef]
18. Valverde, M.; Moutinho, A.; Zacchi, J.V. A Survey of Deep Learning-Based 3D Object Detection Methods for Autonomous Driving Across Different Sensor Modalities. *Sensors* **2025**, *25*, 5264. [CrossRef] [PubMed]
19. Chen, C.; Wang, B.; Lu, C.X.; Trigoni, N.; Markham, A. Deep Learning for Visual Localization and Mapping: A Survey. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *35*, 17000–17020. [CrossRef] [PubMed]
20. Zhang, Y.; Wu, Y.; Tong, K.; Chen, H.; Yuan, Y. Review of Visual Simultaneous Localization and Mapping Based on Deep Learning. *Remote Sens.* **2023**, *15*, 2740. [CrossRef]
21. Liu, L.; Wang, X.; Yang, X.; Liu, H.; Li, J.; Wang, P. Path planning techniques for mobile robots: Review and prospect. *Expert Syst. Appl.* **2023**, *227*, 128423. [CrossRef]
22. Tang, Y.; Zakaria, M.A.; Younas, M. Path Planning Trends for Autonomous Mobile Robot Navigation: A Review. *Sensors* **2025**, *25*, 1206. [CrossRef] [PubMed]
23. Tang, C.; Abbatematteo, B.; Hu, J.; Chandra, R.; Martín-Martín, R.; Stone, P. Deep Reinforcement Learning for Robotics: A Survey of Real-World Successes. *Annu. Rev. Control Robot. Auton. Syst.* **2025**, *8*, 153–188. [CrossRef]
24. Zhu, Y.; Wan Hasan, W.Z.; Harun Ramli, H.R.; Norsahperi, N.M.H.; Mohd Kassim, M.S.; Yao, Y. Deep Reinforcement Learning of Mobile Robot Navigation in Dynamic Environment: A Review. *Sensors* **2025**, *25*, 3394. [CrossRef] [PubMed]
25. Zhu, W.; Guo, X.; Owaki, D.; Kutsuzawa, K.; Hayashibe, M. A Survey of Sim-to-Real Transfer Techniques Applied to Reinforcement Learning for Bio-Inspired Robots. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *34*, 3444–3459. [CrossRef] [PubMed]
26. Kawaharazuka, K.; Matsushima, T.; Gambardella, A.; Guo, J.; Paxton, C.; Zeng, A. Real-world robot applications of foundation models: A review. *Adv. Robot.* **2024**, *38*, 1232–1254. [CrossRef]
27. Kawaharazuka, K.; Oh, J.; Yamada, J.; Posner, I.; Zhu, Y. Vision-Language-Action Models for Robotics: A Review Towards Real-World Applications. *IEEE Access* **2025**, *13*, 162467–162504. [CrossRef]
28. Zhai, W.; Liao, J.; Chen, Z.; Su, B.; Zhao, X. A Survey of Task Planning with Large Language Models. *Intell. Comput.* **2025**, *4*. [CrossRef]
29. Wu, J.; Jin, Z.; Liu, A.; Yu, L.; Yang, F. A survey of learning-based control of robotic visual servoing systems. *J. Frankl. Inst.* **2022**, *359*, 556–577. [CrossRef]
30. Cong, V.D.; Hanh, L.D. A review and performance comparison of visual servoing controls. *Int. J. Intell. Robot. Appl.* **2023**, *7*, 65–90. [CrossRef]
31. Szot, A.; Clegg, A.; Undersander, E.; Wijmans, E.; Zhao, Y.; Turner, J.; Maestre, N.; Mukadam, M.; Chaplot, D.; Maksymets, O.; et al. Habitat 2.0: Training Home Assistants to Rearrange their Habitat. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 251–266.
32. Gutierrez, M.; Rincon, J.A.; Julian, V. Federated Learning for Collaborative Robotics: A ROS 2-Based Approach. *Electronics* **2025**, *14*, 1323. [CrossRef]
33. Mazumder, A.; Sahed, M.F.; Tasneem, Z.; Das, P.; Badal, F.; Ali, M.; Ahamed, M.; Abhi, S.; Sarker, S.; Das, S.; et al. Towards Next Generation Digital Twin in Robotics: Trends, Scopes, Challenges, and Future. *Heliyon* **2023**, *9*, e13359. [CrossRef] [PubMed]
34. Zhao, W.; Queralta, J.P.; Westerlund, T. Sim-to-Real Transfer in Deep Reinforcement Learning for Robotics: A Survey. In Proceedings of the 2020 IEEE Symposium Series on Computational Intelligence (SSCI), Canberra, Australia, 1–4 December 2020. [CrossRef]
35. Somvanshi, S.; Islam, M.; Chhetri, G.; Chakraborty, R.; Mimi, M.S.; Shuvo, S.A.; Islam, K.S.; Javed, S.; Rafat, S.A.; Dutta, A.; et al. From Tiny Machine Learning to Tiny Deep Learning: A Survey. *ACM Comput. Surv.* **2025**, *58*, 1–33. [CrossRef]

36. Firoozi, R.; Tucker, J.; Tian, S.; Majumdar, A.; Sun, J.; Liu, W.; Zhu, Y.; Song, S.; Kapoor, A.; Hausman, K.; et al. Foundation Models in Robotics: Applications, Challenges, and the Future. *Int. J. Robot. Res.* **2024**, *44*, 701–739. [CrossRef]
37. *ISO 3691-4:2023*; Industrial Trucks—Safety Requirements and Verification—Part 4: Driverless Industrial Trucks and Their Systems. International Organization for Standardization (ISO): Geneva, Switzerland, 2023.
38. Brohan, A.; Brown, N.; Carbajal, J.; Chebotar, Y.; Chen, X.; Choromanski, K.; Ding, T.; Driess, D.; Dubey, A.; Finn, C.; et al. RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control. In Proceedings of the Conference on Robot Learning (CoRL), Atlanta, GA, USA, 6–9 November 2023.
39. Kim, M.J.; Pertsch, K.; Karamcheti, S.; Xiao, T.; Balakrishna, A.; Nair, S.; Rafailov, R.; Foster, E.; Lam, G.; Sanketi, P.; et al. OpenVLA: An Open-Source Vision-Language-Action Model. In Proceedings of the Conference on Robot Learning (CoRL), Munich, Germany, 6–9 November 2024.
40. Gemini Robotics Team; Abeyruwan, S.; Ainslie, J.; Alayrac, J.-B.; Arenas, M.G.; Armstrong, T.; Balakrishna, A.; Baruch, R.; Bauza, M.; Blokzijl, M.; et al. Google DeepMind Gemini Robotics: Bringing AI into the Physical World. *arXiv* **2025**, arXiv:2503.20020.
41. Kavraki, L.E.; Svestka, P.; Latombe, J.-C.; Overmars, M.H. Probabilistic Roadmaps for Path Planning in High-Dimensional Configuration Spaces. *IEEE Trans. Robot. Autom.* **1996**, *12*, 566–580. [CrossRef]
42. Fox, D.; Burgard, W.; Thrun, S. The Dynamic Window Approach to Collision Avoidance. *IEEE Robot. Autom. Mag.* **1997**, *4*, 23–33. [CrossRef]
43. Kanayama, Y.; Kimura, Y.; Miyazaki, F.; Noguchi, T. A stable tracking control method for an autonomous mobile robot. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Cincinnati, OH, USA, 13–18 May 1990. [CrossRef]
44. Zhang, J.; Singh, S. LOAM: Lidar Odometry and Mapping in Real-time. In Proceedings of the Robotics: Science and Systems (RSS), Berkeley, CA, USA, 12–14 July 2014. [CrossRef]
45. Chaumette, F.; Hutchinson, S. Visual servo control, Part I: Basic approaches. *IEEE Robot. Autom. Mag.* **2006**, *13*, 82–90. [CrossRef]
46. Gu, J.; Xiang, F.; Li, X.; Ling, Z.; Liu, X.; Mu, T.; Tang, Y.; Tao, S.; Wei, X.; Yao, Y.; et al. ManiSkill2: A Unified Benchmark for Generalizable Manipulation Skills. In Proceedings of the International Conference on Learning Representations (ICLR), Kigali, Rwanda, 1–5 May 2023.
47. O'neill, A.; Rehman, A.; Maddukuri, A.; Gupta, A.; Padalkar, A.; Lee, A.; Pooley, A.; Gupta, A.; Mandlekar, A.; Jain, A.; et al. Open X-Embodiment: Robotic Learning Datasets and RT-X Models. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)—Best Paper Award, Yokohama, Japan, 13–17 May 2024. [CrossRef]
48. Tranzatto, M.; Mascarich, F.; Bernreiter, L.; Godinho, C.; Camurri, M.; Khattak, S.; Dang, T.; Reijgwart, V.; Loeje, J.; Wisth, D.; et al. CERBERUS: Autonomous Legged and Aerial Robotic Exploration in the Tunnel and Urban Circuits of the DARPA Subterranean Challenge. *Field Robot.* **2022**, *2*, 274–324. [CrossRef]
49. Miki, T.; Lee, J.; Hwangbo, J.; Wellhausen, L.; Koltun, V.; Hutter, M. Learning robust perceptive locomotion for quadrupedal robots in the wild. *Sci. Robot.* **2022**, *7*, eabk2822. [CrossRef] [PubMed]
50. Singh, R.; Ren, J.; Lin, X. A Review of Deep Reinforcement Learning Algorithms for Mobile Robot Path Planning. *Vehicles* **2023**, *5*, 1423–1451. [CrossRef]
51. Le, H.; Saeedvand, S.; Hsu, C.C. A Comprehensive Review of Mobile Robot Navigation Using Deep Reinforcement Learning Algorithms in Crowded Environments. *J. Intell. Robot. Syst.* **2024**, *110*, 158. [CrossRef]
52. Tiwari, R.; Khapre, S.; Singh, A. Reinforcement learning in robotic systems: A review on sim-to-real transfer. *Robot. Auton. Syst.* **2026**, *198*, 105327. [CrossRef]
53. Kober, J.; Bagnell, J.A.; Peters, J. Reinforcement Learning in Robotics: A Survey. *Int. J. Robot. Res.* **2013**, *32*, 1238–1274. [CrossRef]
54. Hafez, H.; Marey, M.; Tolbah, F.; Abdelhameed, M. Survey of Visual Servoing Control Schemes for Mobile Robot Navigation. *Sci. J. Oct. 6 Univ.* **2017**, *3*, 41–50. [CrossRef]
55. Li, P.; An, Z.; Abrar, S.; Zhou, L. Large language models for multi-robot systems: A survey. *Auton. Robot.* **2026**, *50*, 30. [CrossRef]
56. Waga, A.; Benhlima, S.; Bekri, A.; Abdouni, J.; Saber, F.Z. A survey on autonomous navigation for mobile robots: From traditional techniques to deep learning and large language models. *J. King Saud Univ. Comput. Inf. Sci.* **2025**, *37*, 198. [CrossRef]
57. Golroudbari, A.A.; Sabour, M.H. Recent Advancements in Deep Learning Applications and Methods for Autonomous Navigation: A Comprehensive Review. *arXiv* **2023**, arXiv:2302.11089.
58. Singamaneni, P.T.; Bachiller-Burgos, P.; Manso, L.J.; Garrell, A.; Sanfeliu, A.; Spalanzani, A.; Alami, R. A survey on socially aware robot navigation: Taxonomy and future challenges. *Int. J. Robot. Res.* **2024**, *43*, 1533–1572. [CrossRef]
59. Sandakalum, T.; Ang, M.H. Motion Planning for Mobile Manipulators—A Systematic Review. *Machines* **2022**, *10*, 97. [CrossRef]
60. Shan, T.; Englot, B.; Meyers, D.; Wang, W.; Ratti, C.; Rus, D. LIO-SAM: Tightly-coupled Lidar Inertial Odometry via Smoothing and Mapping. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 24–30 October 2020. [CrossRef]
61. Maddikunta, P.K.R.; Pham, Q.-V.; Deepa, N.; Dev, K.; Gadekallu, T.R.; Ruby, R.; Liyanage, M. Industry 5.0: A survey on enabling technologies and potential applications. *J. Ind. Inf. Integr.* **2021**, *26*, 100257. [CrossRef]

62. Mentzas, G.; Hribernik, K.; Stahre, J.; Romero, D.; Soldatos, J. Editorial: Human-Centered Artificial Intelligence in Industry 5.0. *Front. Artif. Intell.* **2024**, *7*, 1429186. [CrossRef] [PubMed]
63. Bricher, D.; Müller, A.M. Analysis of Deep-Learning Methods in an ISO/TS 15066–Compliant Human–Robot Safety Framework. *Sensors* **2025**, *25*, 7136. [CrossRef] [PubMed]
64. Dhanda, M.; Rogers, B.A.; Hall, S.; Dekoninck, E.; Dhokia, V. Reviewing human-robot collaboration in manufacturing: Opportunities and challenges in the context of industry 5.0. *Robot. Comput. Manuf.* **2024**, *93*, 102937. [CrossRef]
65. Othman, U.; Yang, E. Human–Robot Collaborations in Smart Manufacturing Environments: Review and Outlook. *Sensors* **2023**, *23*, 5663. [CrossRef] [PubMed]
66. Montini, E.; Daniele, F.; Agbomemewa, L.; Confalonieri, M.; Cutrona, V.; Bettoni, A.; Rocco, P.; Ferrario, A. Collaborative Robotics: A Survey From Literature and Practitioners Perspectives. *J. Intell. Robot. Syst.* **2024**, *110*, 117. [CrossRef]
67. Gualtieri, L.; Rauch, E.; Vidoni, R. Emerging research fields in safety and ergonomics in industrial collaborative robotics: A systematic literature review. *Robot. Comput. Integr. Manuf.* **2021**, *67*, 101998. [CrossRef]
68. Chen, C.; Liu, Y.; Kreiss, S.; Alahi, A. Crowd-Robot Interaction: Crowd-aware Robot Navigation with Attention-based Deep Reinforcement Learning. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019. [CrossRef]
69. Karaman, S.; Frazzoli, E. Sampling-based Algorithms for Optimal Motion Planning. *Int. J. Robot. Res.* **2011**, *30*. [CrossRef]
70. Tai, L.; Paolo, G.; Liu, M. Virtual-to-Real Deep Reinforcement Learning: Continuous Control of Mobile Robots for Mapless Navigation. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017. [CrossRef]
71. Chen, Y.F.; Liu, M.; Everett, M.; How, J.P. Decentralized Non-communicating Multiagent Collision Avoidance with Deep Reinforcement Learning. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017. [CrossRef]
72. Krnjaic, A.; Steleac, R.D.; Thomas, J.D.; Papoudakis, G.; Schäfer, L.; To, A.W.K.; Lao, K.-H.; Cubuktepe, M.; Haley, M.; Börsting, P.; et al. Scalable Multi-Agent Reinforcement Learning for Warehouse Logistics with Robotic and Human Co-Workers. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Dhabi, United Arab Emirates, 14–18 October 2024. [CrossRef]
73. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.-Y.; et al. Segment Anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 1–6 October 2023. [CrossRef]
74. Oquab, M.; Darcet, T.; Moutakanni, T.; Vo, H.; Szafraniec, M.; Khalidov, V.; Fernandez, P.; Haziza, D.; Massa, F.; El-Nouby, A.; et al. DINOv2: Learning Robust Visual Features without Supervision. *Trans. Mach. Learn. Res.* **2024**. Available online: https://openreview.net/forum?id=a68SUt6zFt (accessed on 3 August 2026). [CrossRef]
75. Driess, D.; Xia, F.; Sajjadi, M.S.M.; Lynch, C.; Chowdhery, A.; Ichter, B.; Wahid, A.; Tompson, J.; Vuong, Q.; Yu, T.; et al. PaLM-E: An Embodied Multimodal Language Model. In Proceedings of the International Conference on Machine Learning (ICML), Honolulu, HI, USA, 23–29 July 2023.
76. Liu, Z.; Tang, H.; Amini, A.; Yang, X.; Mao, H.; Rus, D.; Han, S. BEVFusion: Multi-Task Multi-Sensor Fusion with Unified Bird's-Eye View Representation. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), London, UK, 29 May–2 June 2023. [CrossRef]
77. Yaseen, M. What is YOLOv8: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector. *arXiv* **2024**, arXiv:2408.15857.
78. Huang, C.; Mees, O.; Zeng, A.; Burgard, W. Visual Language Maps for Robot Navigation (VLMaps). In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), London, UK, 29 May–2 June 2023. [CrossRef]
79. Jatavallabhula, K.; Kuwajerwala, A.; Gu, Q.; Omama, M.; Iyer, G.; Saryazdi, S.; Chen, T.; Maalouf, A.; Li, S.; Keetha, N.; et al. ConceptFusion: Open-set Multimodal 3D Mapping. In Proceedings of the Robotics: Science and Systems (RSS), Daegu, Republic of Korea, 10–14 July 2023. [CrossRef]
80. Salzmann, T.; Ivanovic, B.; Chakravarty, P.; Pavone, M. Trajectron++: Dynamically-Feasible Trajectory Forecasting with Heterogeneous Data. In *European Conference on Computer Vision—ECCV 2020*; Springer: Cham, Switzerland, 2020. [CrossRef]
81. Shi, S.; Jiang, L.; Dai, D.; Schiele, B. Motion Transformer with Global Intention Localization and Local Movement Refinement (MTR). In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), New Orleans, LA, USA, 28 November–9 December 2022.
82. Zhou, Z.; Wang, J.; Li, Y.-H.; Huang, Y.-K. Query-Centric Trajectory Prediction (QCNet). In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 17863–17873. [CrossRef]
83. Ou, Y.; Cai, Y.; Sun, Y.; Qin, T. Autonomous Navigation by Mobile Robot with Sensor Fusion Based on Deep Reinforcement Learning. *Sensors* **2024**, *24*, 3895. [CrossRef] [PubMed]

84. Wang, Z.; Wang, S.; Xie, Y.; Xiong, T.; Wang, C. Autonomous Navigation of Mobile Robots: A Hierarchical Planning–Control Framework with Integrated DWA and MPC. *Sensors* **2025**, *25*, 2014. [CrossRef] [PubMed]
85. Scherl, A.; Thalhammer, S.; Neuberger, B.; Wöber, W.; García-Rodríguez, J. ViT-VS: On the Applicability of Pretrained Vision Transformer Features for Generalizable Visual Servoing. In Proceedings of the 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hangzhou, China, 19–25 October 2025. [CrossRef]
86. Chen, A.; Yu, H.; Wang, Y.; Xiong, R. CNS: Correspondence Encoded Neural Image Servo Policy. In Proceedings of the 2024 IEEE International Conference on Robotics and Automation (ICRA), Yokohama, Japan, 13–17 May 2024. [CrossRef]
87. Yao, S.; Zhao, J.; Yu, D.; Du, N.; Shafran, I.; Narasimhan, K.; Cao, Y. ReAct: Synergizing Reasoning and Acting in Language Models. In Proceedings of the International Conference on Learning Representations (ICLR), Kigali, Rwanda, 1–5 May 2023.
88. Hafner, D.; Pasukonis, J.; Ba, J.; Lillicrap, T. Mastering Diverse Control Tasks through World Models (DreamerV3). *Nature* **2025**, *640*, 647–653. [CrossRef] [PubMed]
89. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017. [CrossRef]
90. Alreshidi, I.; Alatawi, H.; Binnouh, A.; Duwayriat, L.; Alhmiedat, T.; Alia, O.M. Deep Learning-Based Vision Systems for Robot Semantic Navigation: An Experimental Study. *Technologies* **2024**, *12*, 157. [CrossRef]
91. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. *Adv. Neural Inf. Process. Syst. (NeurIPS)* **2021**, *34*, 12077–12090.
92. Yu, C.; Gao, C.; Wang, J.; Yu, G.; Shen, C.; Sang, N. BiSeNet V2: Bilateral Network with Guided Aggregation for Real-Time Semantic Segmentation. *Int. J. Comput. Vis.* **2021**, *129*, 3051–3068. [CrossRef]
93. Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The Cityscapes Dataset for Semantic Urban Scene Understanding. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. [CrossRef]
94. Behley, J.; Garbade, M.; Milioto, A.; Quenzel, J.; Behnke, S.; Stachniss, C.; Gall, J. SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27–28 October 2019. [CrossRef]
95. Qi, C.R.; Su, H.; Mo, K.; Guibas, L.J. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. [CrossRef]
96. Lang, A.H.; Vora, S.; Caesar, H.; Zhou, L.; Yang, J.; Beijbom, O. PointPillars: Fast Encoders for Object Detection from Point Clouds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Beach, CA, USA, 15–20 June 2019. [CrossRef]
97. Chen, Y.; Liu, J.; Zhang, X.; Qi, X.; Jia, J. VoxelNeXt: Fully Sparse VoxelNet for 3D Object Detection and Tracking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023. [CrossRef]
98. Luo, L.; Zheng, S.; Li, Y.; Fan, Y.; Yu, B.; Cao, S.-Y.; Li, J.; Shen, H.-L. BEVPlace: Learning LiDAR-based Place Recognition using Bird's Eye View Images. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 1–6 October 2023. [CrossRef]
99. Komorowski, J. MinkLoc3D: Point Cloud Based Large-Scale Place Recognition. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–8 January 2021. [CrossRef]
100. Fan, L.; Wang, J.; Chang, Y.; Li, Y.; Wang, Y.; Cao, D. 4D mmWave Radar for Autonomous Driving Perception: A Comprehensive Survey. *IEEE Trans. Intell. Veh.* **2024**, *9*, 4606–4620. [CrossRef]
101. Jia, L.; Guan, R.; Zhao, H.; Zhao, Q.; Man, K.L.; Smith, J.; Yu, L.; Yue, Y. RadarNeXt: Lightweight and Real-Time 3D Object Detector Based on 4D mmWave Imaging Radar. *EURASIP J. Adv. Signal Process.* **2025**, *2025*, 65. [CrossRef]
102. Bai, X.; Hu, Z.; Zhu, X.; Huang, Q.; Chen, Y.; Fu, H.; Tai, C.-L. TransFusion: Robust LiDAR-Camera Fusion for 3D Object Detection with Transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022. [CrossRef]
103. Lin, Z.; Liu, Z.; Xia, Z.; Wang, X.; Wang, Y.; Qi, S.; Dong, Y.; Dong, N.; Zhang, L.; Zhu, C. RCBEVDet: Radar-Camera Fusion in Bird's-Eye-View for 3D Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–22 June 2024. [CrossRef]
104. Qian, H.; Wang, M.; Zhu, M.; Wang, H. A Review of Multi-Sensor Fusion in Autonomous Driving. *Sensors* **2025**, *25*, 6033. [CrossRef] [PubMed]
105. Ravi, N.; Gabeur, V.; Hu, Y.-T.; Hu, R.; Ryali, C.; Ma, T.; Khedr, H.; Radle, R.; Rolland, C.; Gustafson, L.; et al. SAM 2: Segment Anything in Images and Videos. In Proceedings of the International Conference on Learning Representations (ICLR), Singapore, 24–28 April 2025.

106. Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. Learning Transferable Visual Models from Natural Language Supervision. In Proceedings of the International Conference on Machine Learning (ICML), Online, 18–24 July 2021.
107. Faisal, M.A.A.; Mecheter, I.; Qiblawey, Y.; Hernandez Fernandez, J.; Chowdhury, M.E.H.; Kiranyaz, S. Deep Learning in Automated Power Line Inspection: A Review. *Appl. Energy* **2025**, *385*, 125507. [CrossRef]
108. Deitke, M.; Han, W.; Herrasti, A.; Kembhavi, A.; Kolve, E.; Mottaghi, R.; Salvador, J.; Schwenk, D.; VanderBilt, E.; Wallingford, M.; et al. RoboTHOR: An Open Simulation-to-Real Embodied AI Platform. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020. [CrossRef]
109. Caesar, H.; Bankiti, V.; Lang, A.H.; Vora, S.; Liong, V.E.; Xu, Q.; Krishnan, A.; Pan, Y.; Baldan, G.; Beijbom, O. nuScenes: A Multimodal Dataset for Autonomous Driving. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020. [CrossRef]
110. Wang, H.; Shao, W.; Sun, C.; Yang, K.; Cao, D.; Li, J. A Survey on an Emerging Safety Challenge for Autonomous Vehicles: Safety of the Intended Functionality (SOTIF). *Engineering* **2024**, *33*, 17–34. [CrossRef]
111. Felton, S.; Brault, P.; Fromont, E.; Marchand, E. Visual Servoing in Autoencoder Latent Space. *IEEE Robot. Autom. Lett.* **2022**, *7*, 3234–3241. [CrossRef]
112. Ke, F.; Li, Z.; Xiao, H.; Zhang, X. Visual Servoing of Constrained Mobile Robots Based on Model Predictive Control. *IEEE Trans. Syst. Man Cybern. Syst.* **2016**, *47*, 1428–1438. [CrossRef]
113. Gu, H.; Qin, Q.; Mao, J.; Sun, X.; Huang, Y. Hybrid Model Predictive Control with Penalty Factor Based on Image-Based Visual Servoing for Constrained Mobile Robots. *Electronics* **2023**, *12*, 3186. [CrossRef]
114. Lee, A.X.; Levine, S.; Abbeel, P. Learning Visual Servoing with Deep Features and Fitted Q-Iteration. In Proceedings of the International Conference on Learning Representations (ICLR), Toulon, France, 24–26 April 2017.
115. Katara, P.; Harish, Y.V.S.; Pandya, H.; Gupta, A.; Sanchawala, A.M.; Kumar, G.; Bhowmick, B.; Madhava Krishna, K. DeepMPCVS: Deep Model Predictive Control for Visual Servoing. In Proceedings of the Conference on Robot Learning (CoRL), Virtual, 16–18 November 2020.
116. Mwitta, C.; Rains, G.C.; Prostko, E.P. Autonomous diode laser weeding mobile robot in cotton field using deep learning, visual servoing and finite state machine. *Front. Agron.* **2024**, *6*, 1388452. [CrossRef]
117. Panta, L. Comparative Analysis of NMPC and Fuzzy PID Controllers for Trajectory Tracking in Omni-Drive Robots: Design, Simulation, and Performance Evaluation. *Int. J. Fuzzy Syst.* **2024**, *27*, 1691–1701. [CrossRef]
118. von Stumberg, L.; Cremers, D. DM-VIO: Delayed Marginalization Visual-Inertial Odometry. *IEEE Robot. Autom. Lett.* **2022**, *7*, 1408–1415. [CrossRef]
119. Qin, T.; Li, P.; Shen, S. VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator. *IEEE Trans. Robot.* **2018**, *34*, 1004–1020. [CrossRef]
120. Xu, W.; Cai, Y.; He, D.; Lin, J.; Zhang, F. FAST-LIO2: Fast Direct LiDAR-Inertial Odometry. *IEEE Trans. Robot.* **2022**, *38*, 2053–2073. [CrossRef]
121. Zheng, C.; Xu, W.; Zou, Z.; Hua, T.; Yuan, C.; He, D.; Zhou, B.; Liu, Z.; Lin, J.; Zhu, F.; et al. FAST-LIVO2: Fast, Direct LiDAR-Inertial-Visual Odometry. *IEEE Trans. Robot.* **2024**, *41*, 326–346. [CrossRef]
122. Sucar, E.; Liu, S.; Ortiz, J.; Davison, A.J. iMAP: Implicit Mapping and Positioning in Real-Time. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021. [CrossRef]
123. Matsuki, H.; Murai, R.; Kelly, P.H.J.; Davison, A.J. Gaussian Splatting SLAM. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–22 June 2024. [CrossRef]
124. Keetha, N.; Karhade, J.; Jatavallabhula, K.M.; Yang, G.; Scherer, S.; Ramanan, D.; Luiten, J. SplaTAM: Splat, Track & Map 3D Gaussians for Dense RGB-D SLAM. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–22 June 2024. [CrossRef]
125. Chen, X.; Milioto, A.; Palazzolo, E.; Giguere, P.; Behley, J.; Stachniss, C. SuMa++: Efficient LiDAR-based Semantic SLAM. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 3–8 November 2019. [CrossRef]
126. Rosinol, A.; Abate, M.; Chang, Y.; Carlone, L. Kimera: An Open-Source Library for Real-Time Metric-Semantic Localization and Mapping. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020. [CrossRef]
127. Canadas-Aranega, F.; Blanco-Claraco, J.L.; Moreno, J.C.; Rodriguez-Diaz, F. Multimodal Mobile Robotic Dataset for a Typical Mediterranean Greenhouse: The GREENBOT Dataset. *Sensors* **2024**, *24*, 1874. [CrossRef] [PubMed]
128. Burnett, K.; Yoon, D.J.; Wu, Y.; Li, A.Z.; Zhang, H.; Lu, S.; Qian, J.; Tseng, W.-K.; Lambert, A.; Leung, K.Y.K.; et al. Boreas: A Multi-Season Autonomous Driving Dataset. *Int. J. Robot. Res.* **2023**, *42*, 33–42. [CrossRef]
129. Maddern, W.; Pascoe, G.; Linegar, C.; Newman, P. 1 Year, 1000 km: The Oxford RobotCar Dataset. *Int. J. Robot. Res.* **2016**, *36*, 3–15. [CrossRef]

130. Gu, T.; Chen, G.; Li, J.; Lin, C.; Rao, Y.; Zhou, J.; Lu, J. Stochastic Trajectory Prediction via Motion Indeterminacy Diffusion (MID). In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Orleans, LA, USA, 18–24 June 2022. [CrossRef]
131. Mao, W.; Xu, C.; Zhu, Q.; Chen, S.; Wang, Y. Leapfrog Diffusion Model for Stochastic Trajectory Prediction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023. [CrossRef]
132. Wang, Y.; Huang, X.; Sun, X.; Yan, M.; Xing, S.; Tu, Z.; Li, J. UniOcc: A Unified Benchmark for Occupancy Forecasting and Prediction in Autonomous Driving. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Honolulu, HI, USA, 19–25 October 2025. [CrossRef]
133. Yonetani, R.; Taniai, T.; Barekatain, M.; Nishimura, M.; Kanezaki, A. Path Planning using Neural A* Search. In Proceedings of the International Conference on Machine Learning (ICML), Online, 18–24 July 2021.
134. Poddar, S.; Mavrogiannis, C.; Srinivasa, S.S. From Crowd Motion Prediction to Robot Navigation in Crowds. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Detroit, MI, USA, 1–5 October 2023. [CrossRef]
135. Xu, J.; Zhang, W.; Cai, J.; Liu, H. SafeCrowdNav: Safety Evaluation of Robot Crowd Navigation in Complex Scenes. *Front. Neurorobot.* **2023**, *17*, 1276519. [CrossRef] [PubMed]
136. Alyassi, R.; Cadena, C.; Riener, R.; Paez-Granados, D. Social Robot Navigation: A Review and Benchmarking of Learning-Based Methods. *Front. Robot. AI* **2025**, *12*, 1658643. [CrossRef] [PubMed]
137. Zhang, Z.; Cai, Y.; Ceder, K.; Enliden, A.; Eriksson, O.; Kylander, S.; Sridhara, R.; Åkesson, K. Collision-free trajectory planning of mobile robots by integrating deep reinforcement learning and model predictive control. In Proceedings of the IEEE International Conference on Automation Science and Engineering (CASE), Auckland, New Zealand, 26–30 August 2023. [CrossRef]
138. Guo, B.; Wang, G.; Chen, Y.; Gao, Y.; Xie, Q. Risk-Aware Reinforcement Learning with Dynamic Safety Filter for Collision Risk Mitigation in Mobile Robot Navigation. *Sensors* **2025**, *25*, 5488. [CrossRef] [PubMed]
139. Teng, S.; Li, L.; Li, Y.; Hu, X.; Li, L.; Ai, Y.; Chen, L. FusionPlanner: A Multi-task Motion Planner for Mining Trucks via Multi-sensor Fusion. *Mech. Syst. Signal Process.* **2024**, *208*, 111051. [CrossRef]
140. Tang, Z.; Cao, X.; Zhou, Z.; Zhang, Z.; Xu, C.; Dou, J. Path planning of autonomous underwater vehicle in unknown environment based on improved deep reinforcement learning. *Ocean Eng.* **2024**, *301*, 117547. [CrossRef]
141. Yang, J.-M.; Kim, J.-H. Sliding mode control for trajectory tracking of nonholonomic wheeled mobile robots. *IEEE Trans. Robot. Autom.* **1999**, *15*, 578–587. [CrossRef]
142. Borkar, K.K.; Aljrees, T.; Pandey, S.K.; Kumar, A.; Singh, M.K.; Sinha, A.; Singh, K.U.; Sharma, V. Stability Analysis and Navigational Techniques of Wheeled Mobile Robot: A Review. *Processes* **2023**, *11*, 3302. [CrossRef]
143. Zheng, Y.; Zheng, J.; Shao, K.; Zhao, H.; Xie, H.; Wang, H. Adaptive Trajectory Tracking Control for Nonholonomic Wheeled Mobile Robots: A Barrier Function Sliding Mode Approach. *IEEE/CAA J. Autom. Sin.* **2024**, *11*, 1007–1021. [CrossRef]
144. Zhu, C.; Li, B.; Zhao, C.; Wang, Y. Trajectory Tracking Control of Car-like Mobile Robots Based on Extended State Observer and Backstepping Control. *Electronics* **2024**, *13*, 1563. [CrossRef]
145. Ma, L.; Wang, C.; Ge, C.; Liu, H.; Li, B. Fixed-time integral sliding mode tracking control of a wheeled mobile robot. *Complex Eng. Syst.* **2023**, *2*, 14. [CrossRef]
146. Dong, D.; Ye, H.; Luo, W.; Wen, J.; Huang, D. Collision Avoidance Path Planning and Tracking Control for Autonomous Vehicles Based on Model Predictive Control. *Sensors* **2024**, *24*, 5211. [CrossRef] [PubMed]
147. Zhang, J.-J.; Fang, Z.-L.; Zhang, Z.-Q.; Gao, R.-Z.; Zhang, S.-B. Trajectory Tracking Control of Nonholonomic Wheeled Mobile Robots Using Model Predictive Control Subjected to Lyapunov-based Input Constraints. *Int. J. Control Autom. Syst.* **2022**, *20*, 1640–1651. [CrossRef]
148. Gao, X.; Su, X.; An, A.; Zhang, H. Robust model predictive tracking control for the wheeled mobile robot with boundary uncertain based on linear matrix inequalities. *IET Cyber-Syst. Robot.* **2023**, *5*, e12086. [CrossRef]
149. Zhang, D.; Wang, G.; Wu, Z. Reinforcement Learning-Based Tracking Control for a Three Mecanum Wheeled Mobile Robot. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *35*, 1445–1452. [CrossRef]
150. Pitowarno, E.; Mailah, M.; Jamaluddin, H. Intelligent active force control for mobile manipulator. *J. Teknol. (UTM)* **2001**, *35*, 85–98.
151. Jin, X.Z.; Zhao, Y.; Wang, H.; Zhao, Z.; Dong, X. Adaptive fault-tolerant control of mobile robots with actuator faults and unknown parameters. *IET Control Theory Appl.* **2019**, *13*, 1665–1672. [CrossRef]
152. Liu, P.; Tong, S. Fuzzy Adaptive Fault-tolerant Control for Underactuated Wheeled Mobile Robot with Prescribed Performance. *Int. J. Control Autom. Syst.* **2024**, *22*, 1998–2006. [CrossRef]
153. Garg, K.; Zhang, S.; So, O.; Dawson, C.; Fan, C. Learning Safe Control for Multi-Robot Systems: Methods, Verification, and Open Challenges. *Annu. Rev. Control* **2024**, *57*, 100948. [CrossRef]
154. Dawson, C.; Gao, S.; Fan, C. Safe Control with Learned Certificates: A Survey of Neural Lyapunov, Barrier, and Contraction Methods for Robotics and Control. *IEEE Trans. Robot.* **2023**, *39*, 1749–1767. [CrossRef]

155. Liang, J.; Huang, W.; Xia, F.; Xu, P.; Hausman, K.; Ichter, B.; Florence, P.; Zeng, A. Code as Policies: Language Model Programs for Embodied Control. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), London, UK, 29 May–2 June 2023. [CrossRef]
156. Singh, I.; Blukis, V.; Mousavian, A.; Goyal, A.; Xu, D.; Tremblay, J.; Fox, D.; Thomason, J.; Garg, A. ProgPrompt: Generating Situated Robot Task Plans using Large Language Models. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), London, UK, 29 May–2 June 2023. [CrossRef]
157. Huang, W.; Xia, F.; Xiao, T.; Chan, H.; Liang, J.; Florence, P.; Zeng, A.; Tompson, J.; Mordatch, I.; Chebotar, Y.; et al. Inner Monologue: Embodied Reasoning through Planning with Language Models. In Proceedings of the Conference on Robot Learning (CoRL), Auckland, New Zealand, 14–18 December 2022.
158. Liu, B.; Jiang, Y.; Zhang, X.; Liu, Q.; Zhang, S.; Biswas, J.; Stone, P. LLM+P: Empowering Large Language Models with Optimal Planning Proficiency. *arXiv* **2023**, arXiv:2304.11477.
159. Lin, K.; Agia, C.; Migimatsu, T.; Pavone, M.; Bohg, J. Text2Motion: From Natural Language Instructions to Feasible Plans. *Auton. Robot.* **2023**, *47*, 1345–1365. [CrossRef]
160. Capitanelli, A.; Mastrogiovanni, F. A Framework for Neurosymbolic Robot Action Planning Using Large Language Models. *Front. Neurorobot.* **2024**, *18*, 1342786. [CrossRef] [PubMed]
161. Mandi, Z.; Jain, S.; Song, S. RoCo: Dialectic Multi-Robot Collaboration with Large Language Models. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Yokohama, Japan, 13–17 May 2024. [CrossRef]
162. Kannan, S.S.; Venkatesh, V.L.N.; Min, B.-C. SMART-LLM: Smart Multi-Agent Robot Task Planning using Large Language Models. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Abu Dhabi, United Arab Emirates, 14–18 October 2024. [CrossRef]
163. Kim, Y.; Kim, D.; Choi, J.; Park, J.; Oh, N.; Park, D. A Survey on Integration of Large Language Models with Intelligent Robots. *Intell. Serv. Robot.* **2024**, *17*, 1091–1107. [CrossRef]
164. Brohan, A.; Brown, N.; Carbajal, J.; Chebotar, Y.; Dabis, J.; Finn, C.; Gopalakrishnan, K.; Hausman, K.; Herzog, A.; Hsu, J.; et al. RT-1: Robotics Transformer for Real-World Control at Scale. In Proceedings of the Robotics: Science and Systems (RSS), Daegu, Republic of Korea, 10–14 July 2023. [CrossRef]
165. Ghosh, D.; Walke, H.; Pertsch, K.; Black, K.; Mees, O.; Dasari, S.; Hejna, J.; Kreiman, T.; Xu, C.; Luo, J.; et al. Octo: An Open-Source Generalist Robot Policy. In Proceedings of the Robotics: Science and Systems (RSS), Delft, The Netherlands, 15–19 July 2024. [CrossRef]
166. Huang, W.; Wang, C.; Zhang, R.; Li, Y.; Wu, J.; Fei-Fei, L. VoxPoser: Composable 3D Value Maps for Robotic Manipulation with Language Models. In Proceedings of the Conference on Robot Learning (CoRL), Atlanta, GA, USA, 6–9 November 2023.
167. Agarwal, N.; Ali, A.; Bala, M.; Balaji, Y.; Barker, E.; Cai, T.; Chattopadhyay, P.; Chen, Y.; Cui, Y.; Ding, Y.; et al. Cosmos World Foundation Model Platform for Physical AI. *arXiv* **2025**, arXiv:2501.03575.
168. Cheng, A.-C.; Ji, Y.; Yang, Z.; Gongye, Z.; Zou, X.; Kautz, J.; Biyik, E.; Yin, H.; Liu, S.; Wang, X. NaVILA: Legged Robot Vision-Language-Action Model for Navigation. In Proceedings of the Robotics: Science and Systems (RSS), Los Angeles, CA, USA, 21–25 June 2025. [CrossRef]
169. *ISO 16290:2013*; Space Systems—Definition of the Technology Readiness Levels (TRLs) and Their Criteria of Assessment. International Organization for Standardization: Geneva, Switzerland, 2013.
170. Wu, P.; Escontrela, A.; Hafner, D.; Abbeel, P.; Goldberg, K. DayDreamer: World Models for Physical Robot Learning. In Proceedings of the Conference on Robot Learning (CoRL), Auckland, New Zealand, 14–18 December 2022.
171. Black, K.; Brown, N.; Driess, D.; Esmail, A.; Equi, M.; Finn, C.; Fusai, N.; Groom, L.; Hausman, K.; Ichter, B.; et al. Pi0: A Vision-Language-Action Flow Model for General Robot Control. *arXiv* **2024**, arXiv:2410.24164.
172. Li, Z.M.; Chapin, A.; Xiang, E.; Yang, R.; Machado, B.; Lei, N.; Dellandrea, E.; Huang, D.; Chen, L. Robotic Manipulation via Imitation Learning: Taxonomy, Evolution, Benchmark, and Challenges. *arXiv* **2025**, arXiv:2508.17449.
173. Tiboni, G.; Arndt, K.; Kyrki, V. DROPO: Sim-to-Real Transfer with Offline Domain Randomization. *Robot. Auton. Syst.* **2023**, *166*, 104432. [CrossRef]
174. Mishra, S.; Dwivedi, R.K. Designing Automation for Pickup and Delivery Tasks in Modern Warehouses Using Multi Agent Path Finding (MAPF) and Multi Agent Reinforcement Learning (MARL) Based Approaches. *EAI Endorsed Trans. AI Robot.* **2024**, *3*. [CrossRef]
175. Santos, T.; Cunha, T.; Dias, A.; Moreira, A.P.; Almeida, J. UAV Visual and Thermographic Power Line Detection Using Deep Learning. *Sensors* **2024**, *24*, 5678. [CrossRef] [PubMed]
176. Bergmann, P.; Fauser, M.; Sattlegger, D.; Steger, C. MVTec AD—A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019. [CrossRef]
177. Staniaszek, M.; Flatscher, T.; Rowell, J.; Niu, H.; Liu, W.; You, Y.; Skilton, R.; Fallon, M.; Hawes, N. AutoInspect: Toward Long-Term Autonomous Inspection and Monitoring. *IEEE Trans. Field Robot.* **2025**, *2*, 529–548. [CrossRef]

178. Castelar Wembers, C.; Pflughaupt, J.; Moshagen, L.; Kurenkov, M.; Lewejohann, T.; Schildbach, G. LiDAR-based automated UAV inspection of wind turbine rotor blades. *J. Field Robot.* **2024**, *41*, 1116–1132. [CrossRef]
179. Ramadan, M.N.A.; Hilles, S.M.S.; Alkhedher, M. Design and Study of an AI-Powered Autonomous Stair Climbing Robot. *El-Cezerî J. Sci. Eng.* **2023**, *10*, 571–585. [CrossRef]
180. Ramadan, M.N.A.; Ali, M.A.H.; Alkhedher, M. Intelligent Multi-Pollutant Prediction for Indoor Air Quality Management in Industrial Buildings Using Adaptive Sensor Fusion. *J. Build. Eng.* **2025**, *118*, 115048. [CrossRef]
181. Essien, E.; Frimpong, S. Enhancing Autonomous Truck Navigation in Underground Mines: A Review of 3D Object Detection Systems, Challenges, and Future Trends. *Drones* **2025**, *9*, 433. [CrossRef]
182. Inostroza, F.; Parra-Tsunekawa, I.; Ruiz-del-Solar, J. Robust Localization for Underground Mining Vehicles: An Application in a Room and Pillar Mine. *Sensors* **2023**, *23*, 8059. [CrossRef] [PubMed]
183. Rondoni, C.; Scotto di Luzio, F.; Tamantini, C.; Tagliamonte, N.L.; Chiurazzi, M.; Ciuti, G.; Zollo, L. Navigation benchmarking for autonomous mobile robots in hospital environment. *Sci. Rep.* **2024**, *14*, 18334. [CrossRef] [PubMed]
184. Wang, B.; Chen, S.; Xiao, G. Advancing healthcare through mobile collaboration: A survey of intelligent nursing robots research. *Front. Public Health* **2024**, *12*, 1368805. [CrossRef] [PubMed]
185. Mwitta, C.; Rains, G.C. The integration of GPS and visual navigation for autonomous navigation of an Ackerman steering mobile robot in cotton fields. *Front. Robot. AI* **2024**, *11*, 1359887. [CrossRef] [PubMed]
186. Baby, T.; Gohil, M.K.; Bhattacharya, B. Autonomous Navigation of 4WIS4WID Agricultural Field Mobile Robot using Deep Reinforcement Learning. In Proceedings of the ASME International Mechanical Engineering Congress and Exposition (IMECE India), Hyderabad, India, 10–13 September 2025. [CrossRef]
187. Xu, R.; Xia, X.; Li, J.; Li, H.; Zhang, S.; Tu, Z.; Meng, Z.; Xiang, H.; Dong, X.; Song, R.; et al. V2V4Real: A Real-World Large-Scale Dataset for Vehicle-to-Vehicle Cooperative Perception. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023. [CrossRef]
188. Hadi, B.; Khosravi, A.; Sarhadi, P. Deep reinforcement learning for adaptive path planning and control of an autonomous underwater vehicle. *Appl. Ocean Res.* **2022**, *129*, 103326. [CrossRef]
189. Amaya-Mejia, L.M.; Ghita, M.; Dentler, J.; Olivares-Mendez, M.; Martinez, C. Visual Servoing for Robotic On-Orbit Servicing: A Survey. In Proceedings of the 2024 International Conference on Space Robotics (iSpaRo), Luxembourg, 24–27 June 2024. [CrossRef]
190. Geiger, A.; Lenz, P.; Stiller, C.; Urtasun, R. Vision meets Robotics: The KITTI Dataset. *Int. J. Robot. Res.* **2013**, *32*, 1231–1237. [CrossRef]
191. Sun, P.; Kretzschmar, H.; Dotiwalla, X.; Chouard, A.; Patnaik, V.; Tsui, P.; Guo, J.; Zhou, Y.; Chai, Y.; Caine, B.; et al. Scalability in Perception for Autonomous Driving: Waymo Open Dataset. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020. [CrossRef]
192. Burri, M.; Nikolic, J.; Gohl, P.; Schneider, T.; Rehder, J.; Omari, S.; Achtelik, M.W.; Siegwart, R. The EuRoC Micro Aerial Vehicle Datasets. *Int. J. Robot. Res.* **2016**, *35*, 1157–1163. [CrossRef]
193. Savva, M.; Kadian, A.; Maksymets, O.; Zhao, Y.; Wijmans, E.; Jain, B.; Straub, J.; Liu, J.; Koltun, V.; Malik, J.; et al. Habitat: A Platform for Embodied AI Research. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019. [CrossRef]
194. Chang, A.; Dai, A.; Funkhouser, T.; Halber, M.; Niessner, M.; Savva, M.; Song, S.; Zeng, A.; Zhang, Y. Matterport3D: Learning from RGB-D Data in Indoor Environments. In Proceedings of the 2017 International Conference 3D Vision (3DV), Qingdao, China, 10–12 October 2017. [CrossRef]
195. Li, C.; Zhang, R.; Wong, J.; Gokmen, C.; Srivastava, S.; Martín-Martín, R.; Wang, C.; Levine, G.; Ai, W.; Martinez, B.; et al. BEHAVIOR-1K: A Human-Centered, Embodied AI Benchmark with 1,000 Everyday Activities and Realistic Simulation. *arXiv* **2024**, arXiv:2403.09227.
196. Radlak, K.; Szczepankiewicz, M.; Jones, T.; Serwa, P. Organization of Machine Learning Based Product Development as per ISO 26262 and ISO/PAS 21448. In Proceedings of the IEEE Pacific Rim International Symposium on Dependable Computing (PRDC), Perth, Australia, 1–4 December 2020. [CrossRef]
197. Wu, H.; Isac, O.; Zeljic, A.; Tagomori, T.; Daggitt, M.; Kokke, W.; Refaeli, I.; Amir, G.; Julian, K.; Bassan, S.; et al. Marabou 2.0: A Versatile Formal Analyzer of Neural Networks. In *Computer Aided Verification (CAV), LNCS*; Springer: Cham, Switzerland, 2024. [CrossRef]
198. Guerrier, M.; Fouad, H.; Beltrame, G. Learning Control Barrier Functions and Their Application in Reinforcement Learning: A Survey. *arXiv* **2024**, arXiv:2404.16879.
199. Schnitzer, R.; Kilian, L.; Roessner, S.; Theodorou, K.; Zillner, S. Landscape of AI Safety Concerns—A Methodology to Support Safety Assurance for AI-Based Autonomous Systems. In Proceedings of the International Conference on System Reliability and Safety (ICSRS), Sicily, Italy, 20–22 November 2024. [CrossRef]
200. Sakai, T.; Nagai, T. Explainable Autonomous Robots: A Survey and Perspective. *Adv. Robot.* **2022**, *36*, 219–238. [CrossRef]

201. Sobrin-Hidalgo, D.; Guerrero-Higueras, Á.M.; Matellán-Olivera, V. Generating Explanations for Autonomous Robots: A Systematic Review. *IEEE Access* **2025**, *13*, 20413–20426. [CrossRef]
202. Hwangbo, J.; Lee, J.; Dosovitskiy, A.; Bellicoso, D.; Tsounis, V.; Koltun, V.; Hutter, M. Learning Agile and Dynamic Motor Skills for Legged Robots. *Sci. Robot.* **2019**, *4*, eaau5872. [CrossRef] [PubMed]
203. Muratore, F.; Ramos, F.; Turk, G.; Yu, W.; Gienger, M.; Peters, J. Robot Learning From Randomized Simulations: A Review. *Front. Robot. AI* **2022**, *9*, 799893. [CrossRef] [PubMed]
204. Mehta, V.; Sharma, C.; Thiyagarajan, K. Large Language Models and 3D Vision for Intelligent Robotic Perception and Autonomy. *Sensors* **2025**, *25*, 6394. [CrossRef] [PubMed]
205. Zaland, O.; Elmroth, E.; Bhuyan, M. MTF-Grasp: A Multi-tier Federated Learning Approach for Robotic Grasping. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC), Vienna, Austria, 5–8 October 2025. [CrossRef]
206. Zhen, H.; Qiu, X.; Chen, P.; Yang, J.; Yan, X.; Du, Y.; Hong, Y.; Gan, C. 3D-VLA: A 3D Vision-Language-Action Generative World Model. In Proceedings of the International Conference on Machine Learning (ICML), Vienna, Austria, 21–27 July 2024.
207. Chang, W.; Sun, W.; Chen, P.; Xu, H. Construction and Validation of a Digital Twin-Driven Virtual-Reality Fusion Control Platform for Industrial Robots. *Sensors* **2025**, *25*, 4153. [CrossRef] [PubMed]
208. Wang, G.; Xie, Y.; Jiang, Y.; Mandlekar, A.; Xiao, C.; Zhu, Y.; Fan, L.; Anandkumar, A. Voyager: An Open-Ended Embodied Agent with Large Language Models. *Trans. Mach. Learn. Res.* **2024**. Available online: https://openreview.net/forum?id=ehfRiF0R3a (accessed on 3 August 2026). [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.