Knowledge Representation and Reasoning for Robotics and Autonomous Agents in AI-Empowered Intelligent Engineering Systems: A Review

Isabella Martinez
Department of Electrical and Electronic Engineering, University of Melbourne, Australia

ABSTRACT

Knowledge representation and reasoning (KR&R) constitute the cognitive foundation enabling robotics and autonomous agents to operate reliably within complex, dynamic environments characteristic of intelligent engineering systems. Anchored in systems engineering theory, cyber-physical systems (CPS) principles of bidirectional physical-cyber coupling, and information-theoretic measures of uncertainty and information gain, this review synthesizes advances that transform raw sensor streams into structured, actionable knowledge. Ontologies, description logics, and neuro-symbolic architectures are examined alongside learning-augmented reasoning mechanisms that support adaptive planning and explainable decision-making. Concrete exemplars illustrate translation into practice: a digital twin of an automotive assembly cell employing reinforcement learning for real-time task rescheduling under production disturbances, and a semantic knowledge-graph framework guiding a collaborative manipulator through context-aware grasp selection and failure recovery. Benefits in operational resilience, human-robot synergy, and lifecycle optimization are weighed against limitations in computational tractability, brittleness under distributional shift, and certification challenges for safety-critical reasoning. An original insight proposes augmenting CPS reference architectures with semantically layered entropy-aware belief maintenance to optimize knowledge fidelity versus update cost. The review concludes by outlining research trajectories toward hybrid KR&R frameworks that reconcile symbolic rigor with data-driven adaptability, thereby advancing trustworthy autonomy in next-generation intelligent engineering systems.

Keywords: Knowledge Representation And Reasoning, Robotics, Autonomous Agents, Cyber-Physical Systems, Digital Twins, Neuro-Symbolic Reasoning, Ontologies, Intelligent Engineering Systems

1. INTRODUCTION

The deployment of autonomous robotic systems across manufacturing, logistics, and mobility domains has exposed the inadequacy of purely reactive or end-to-end learned controllers when confronted with partial observability, dynamic task specifications, and the need for human-interpretable justification. Knowledge representation and reasoning furnish the symbolic substrate that maintains coherent world models, supports inference over incomplete information, and generates plans whose rationale can be inspected and contested [1]. Within the broader program of artificial intelligence empowering intelligent engineering systems, KR&R functions as the integrative layer that aligns low-level perception and control loops with high-level requirements traceability and lifecycle management prescribed by systems engineering standards.

Cyber-physical systems theory emphasizes the continuous synchronization between physical processes and their computational mirrors through dense sensing and actuating feedback. Effective KR&R is indispensable for maintaining this synchronization when sensor noise, actuator wear, or environmental change invalidate nominal models. Information theory supplies quantitative guidance: the entropy of an agent's belief distribution over object states or action outcomes directly informs the value of additional sensing or reasoning steps, while mutual information between observations and task-relevant variables bounds the efficiency of knowledge-update policies [2].

Early robotic architectures relied on hand-engineered state machines or limited predicate logics that scaled poorly beyond structured laboratory settings. Subsequent surveys documented the migration toward richer ontological commitments and hybrid symbolic-subsymbolic pipelines capable of grounding abstract concepts in perceptual data. Action representations, in particular, have been taxonomized according to the level of abstraction, embodiment grounding, and acquisition method, revealing persistent gaps in handling concurrent, durative, and context-dependent behaviors [2].

This review is organized to mirror the logical progression from theory to technology to deployment. Section 2 articulates the theoretical integration of systems engineering, CPS, and information-theoretic perspectives. Section 3 surveys methodological advances, emphasizing ontologies, neuro-symbolic reasoning, and learning-augmented frameworks, each illustrated by a deployed exemplar. Section 4 examines application domains with emphasis on industrial robotics and multi-agent autonomy. Section 5 critically assesses risks and open challenges while advancing an original architectural proposal. Section 6 synthesizes findings and identifies priority research directions.

2. Theoretical Foundations of Knowledge Representation in Cyber-Physical Intelligent Engineering Systems

Systems engineering frames intelligent engineering systems as socio-technical wholes whose emergent behaviors arise from tightly coupled constituent elements whose interfaces must be explicitly modeled and verified across the lifecycle. KR&R supplies the formal vocabulary and inference machinery necessary to represent requirements, functional architectures, and verification artifacts in machine-interpretable form, thereby closing the loop between design-time models and runtime adaptation.

CPS theory deepens this perspective by insisting on bidirectional causality: physical state evolution continuously updates the cyber model, while cyber-generated decisions immediately affect physical trajectories. Knowledge representations must therefore encode not only static object properties but also dynamic relations, temporal constraints, and uncertainty distributions that evolve with sensor streams. Digital twins exemplify this requirement. In automotive body-in-white assembly, a high-fidelity digital twin of a collaborative robot cell captures geometric, kinematic, and process parameters; real-time synchronization permits the twin to forecast cycle-time deviations arising from part variation or tool wear [3]. The twin's internal knowledge model fuses geometric CAD data with ontological assertions about task precedence and resource capabilities, enabling downstream reasoning modules to diagnose root causes and propose rescheduling without halting the physical line.

Information theory provides the normative lens for deciding which assertions merit inclusion and how frequently they should be revised. Belief entropy quantifies ambiguity remaining after sensor fusion; expected information gain from a prospective observation or inference step supplies a principled stopping criterion for reasoning under resource bounds. When fused with CPS feedback loops, these measures enable agents to allocate limited onboard compute preferentially to high-entropy, high-impact knowledge fragments—an essential capability for resource-constrained mobile platforms.

An original extension proposed here augments standard CPS reference models (e.g., the 5C architecture) with a semantic entropy layer. At each abstraction stratum—perception, world model, task network—entropy is tracked explicitly and used to trigger targeted knowledge refinement or human query. This yields a quantifiable trade-off between model fidelity and communication or computation cost, directly supporting systems-engineering requirements for verifiable performance margins.

3. Methodological Approaches and Key Technologies for Integrated Reasoning

Contemporary KR&R pipelines for robotics combine symbolic structures with data-driven components to overcome the brittleness of pure logic and the opacity of pure neural controllers. Ontologies expressed in description logics or OWL provide the terminological backbone, defining classes, properties, and axioms that support automated classification and consistency checking. Neuro-symbolic architectures interleave neural perception modules with symbolic reasoners, allowing gradient-based learning to populate or revise symbolic assertions while preserving logical soundness.

A deployed illustration is the neuro-symbolic interface developed for industrial digital twins [4]. Natural-language queries from shop-floor engineers are parsed into logical forms that query a 3D scene graph maintained by the twin; answers are generated by a symbolic solver whose intermediate proof tree is rendered back into human-readable explanations. The hybrid loop demonstrably reduces mean time to diagnose fixture misalignment compared with either purely neural or purely rule-based baselines, while preserving traceability demanded by quality-management standards.

Knowledge graphs extend ontological foundations by incorporating instance-level facts extracted from perception or historical logs. In robotic manipulation, a semantic knowledge graph encodes object affordances, grasp taxonomies, and failure-mode preconditions; graph neural networks or symbolic query engines then retrieve contextually appropriate skills [5]. When the graph is maintained inside a digital twin, continuous edge updates from vision and force-torque sensors enable the robot to adapt grasp strategies to previously unseen part variants without retraining the entire policy.

Reinforcement learning augments these representations by discovering policies that optimize long-horizon objectives expressed over symbolic state abstractions. In digital-twin environments, model-based RL agents learn to adjust production schedules when simulated sensor drift exceeds entropy thresholds, thereby mitigating the reality gap that otherwise degrades transfer performance. Edge computing further constrains architectural choices. Onboard reasoning must execute within millisecond budgets for closed-loop control; consequently, lightweight description-logic fragments or compiled rule sets are deployed at the edge while heavier abductive or planning queries are offloaded to cloud twins. Computer-vision pipelines for online surface-defect detection exemplify the pattern: convolutional features are classified into symbolic defect categories that immediately trigger rule-based recovery behaviors, all within a 30 ms cycle on embedded hardware.

4. Applications in Robotics and Autonomous Agents

In smart-factory deployments, fleets of autonomous mobile robots (AMRs) coordinate via shared ontological models of layout, task allocation, and safety zones. When a conveyor blockage is detected by onboard vision, the affected AMR asserts a new spatial constraint into the shared knowledge base; a distributed reasoner recomputes task priorities and issues rerouting directives whose rationale is logged for later audit. The resulting system exhibits measurably higher throughput and fewer deadlocks than purely geometric planners, while the explicit knowledge layer facilitates root-cause analysis after incidents. Such coordination is supported by recent advances in hybrid reasoning frameworks that integrate symbolic task allocation with learned bidding strategies under digital-twin supervision [6].

Multi-agent warehouse orchestration provides a second exemplar. Each robot maintains a local belief state over tote locations and human co-worker positions; periodic symbolic synchronization merges these beliefs while respecting privacy constraints encoded as access-control axioms. Reinforcement-learned bidding policies operate over the symbolic task descriptors, enabling emergent load balancing that adapts to order-mix changes without centralized replanning.

Beyond manufacturing, autonomous aerial and ground vehicles in logistics corridors employ scene graphs that fuse LiDAR semantics with traffic-rule ontologies. When an occluded pedestrian is hypothesized, the reasoner evaluates counterfactual trajectories and selects the maneuver whose worst-case outcome remains within acceptable risk bounds derived from information-theoretic value-of-information calculations. Such hybrid pipelines have demonstrated reduced intervention rates in field trials relative to end-to-end learned baselines, while retaining the audit trails required for regulatory approval.

5. Challenges, Limitations, and Future Prospects

Despite demonstrated gains, several structural limitations impede wider adoption. First, the computational complexity of sound and complete reasoning over expressive ontologies grows rapidly with axiom count; real-time guarantees remain elusive for large-scale industrial models unless aggressive compilation or approximation is accepted, introducing soundness risks. Second, distributional shift between training environments and deployment sites continues to expose brittleness: neuro-symbolic systems may correctly classify familiar objects yet fail to revise ontological commitments when novel affordances appear. Third, safety certification of reasoning components is complicated by the absence of standardized assurance cases that link entropy bounds or proof certificates to probabilistic failure rates demanded by ISO 26262 or IEC 61508.

Data quality and provenance constitute an additional vector of risk. Digital twins fed by noisy or biased sensor streams can propagate systematic errors into symbolic assertions, undermining downstream planning. Privacy and security concerns arise when shared knowledge bases span organizational boundaries, as in multi-vendor production networks.

An original prospect lies in the tighter integration of information-theoretic monitoring with symbolic maintenance. By maintaining running estimates of belief entropy and expected information gain at each layer of the CPS hierarchy, an agent can autonomously decide when to invoke expensive abductive reasoning, request human clarification, or trigger targeted data collection [7]. Such entropy-aware KR&R would directly support systems-engineering requirements for quantifiable margins and could be embedded within digital-twin governance dashboards. Parallel progress in neurosymbolic program synthesis and verified compilers for description logics offers a pathway toward certifiable yet adaptive reasoning engines [8]. Finally, the emergence of standardized robotic ontologies and scene-graph interchange formats promises to reduce the current fragmentation that forces each deployment to re-engineer core KR components.

6. Conclusion

Knowledge representation and reasoning have matured from peripheral utilities into central enablers of trustworthy autonomy within AI-empowered intelligent engineering systems. Grounded in the tripartite theoretical foundation of systems engineering, CPS bidirectional coupling, and information-theoretic resource allocation, contemporary pipelines successfully fuse ontological structure with neuro-symbolic and reinforcement-learning components. Documented deployments in automotive assembly, warehouse logistics, and autonomous mobility demonstrate concrete gains in adaptability, explainability, and operational efficiency. Persistent challenges in scalability, robustness under shift, and certification nevertheless require sustained attention. The proposed entropy-aware semantic layer offers a principled mechanism for balancing knowledge fidelity against computational and communication cost, pointing toward reference architectures that are simultaneously more capable and more verifiable. Continued convergence of symbolic rigor, data-driven adaptability, and rigorous systems-engineering discipline will determine the pace at which autonomous agents transition from laboratory curiosities to indispensable constituents of resilient, human-centric intelligent engineering systems.

REFERENCES

[1] D. Paulius and Y. Sun, "A survey of knowledge representation in service robotics," Robotics and Autonomous Systems, vol. 118, pp. 13–30, 2019, doi: 10.1016/j.robot.2019.03.005.

[2] P. Zech, E. Renaudo, S. Haller, X. Zhang, and J. Piater, "Action representations in robotics: A taxonomy and systematic classification," The International Journal of Robotics Research, vol. 38, no. 5, pp. 518–562, 2019, doi: 10.1177/0278364919835020.

[3] Bilberg and A. A. Malik, "Digital twin driven human-robot collaborative assembly," CIRP Annals, vol. 68, no. 1, pp. 499–502, 2019, doi: 10.1016/j.cirp.2019.04.011.

A. Siyaev and G.-S. Jo, "Interaction with industrial digital twin using neuro-symbolic reasoning," Sensors, vol. 23, no. 3, p. 1729, 2023, doi: 10.3390/s23031729.

[4] R. Miao, Q. Jia, F. Sun, G. Chen, H. Huang, and S. Miao, "Semantic representation of robot manipulation with knowledge graph," Entropy, vol. 25, no. 4, p. 657, 2023, doi: 10.3390/e25040657.

[5] Cronrath and B. Lennartson, "Enhancing digital twins through reinforcement learning," in Proc. 2019 IEEE 15th Int. Conf. Automation Science and Engineering (CASE), 2019, pp. 293–298, doi: 10.1109/CASE.2019.8843063.

[6] P. Zheng, Z. Wang, C. H. Chen, and L. P. Khoo, "A generic tri-model-based approach for product-level digital twin in cyber-physical systems," Int. J. Comput. Integr. Manuf., vol. 33, no. 8, pp. 753–769, 2020, doi: 10.1080/0951192X.2020.1775296.

[7] M. Sridharan, "Integrated knowledge-based reasoning and data-driven methods for robots," in Explainable Agents and Robots: Papers from the 2023 AAAI Spring Symposium, 2023