Artificial Intelligence in Logistics and Supply Chain

Script: Artificial Intelligence in Logistics and Supply Chain

Authors: Jozef Ristvej, Daniel Chovanec, Boris Kollár, Jozef Kubás

European Funds for Social Development | Republic of Poland | Co-funded by the European Union | NAWA POLISH NATIONAL AGENCY FOR ACADEMIC EXCHANGE

Content

Introduction ... 5

1. SCIENTIFIC METHODS AND TECHNIQUES IN ARTIFICIAL INTELLIGENCE IN LOGISTICS AND SUPPLY CHAIN, KEY CONCEPTS AND PRINCIPLES OF ARTIFICIAL INTELLIGENCE IN LOGISTICS AND SUPPLY CHAINS ... 6

2. SUPPLY CHAIN VS. LOGISTICS CHAIN ... 11

Concise overview of supply chain issues ... 11

Historical Evolution of Supply Chain ... 11

Core Theoretical Frames ... 13

Emerging & Future Trends ... 14

Concise overview of logistics chain issues ... 15

Relationship between the Supply Chain and the Logistics Chain ... 16

Trends and Differences in the Use of Modern Technologies in Supply and Logistics Chains ... 17

Conclusion ... 18

Tasks ... 18

3. ARTIFICIAL INTELLIGENCE ... 20

Knowledge-Based Era ... 21

Foundation-Model Era ... 22

Data-Driven Approach ... 23

Artificial Intelligence Definition ... 24

Deep Learning ... 27

Neural Networks ... 28

Machine Learning ... 30

Summary of AI ... 31

Interactions of Layers ... 32

Principles of AI ... 33

Artificial Intelligence of Today ... 34

Future of AI ... 37

Limits, Costs and Safety ... 40

Conclusion ... 42

4. PREDICTIVE ANALYTICS VS. DEMAND FORECASTING ... 44

Predictive Analytics in Logistics ... 47

Demand Forecasting ... 49

Demand Forecasting in Logistics ... 52

Conclusion ... 53

5. INVENTORY MANAGEMENT AND PRODUCTION PLANNING ... 55

Overview ... 55

Historical Evolution ... 56

Artificial Intelligence in Inventory Management & Production Planning ... 60

Use Cases for AI and Advanced Technologies in Inventory Management and Production Planning ... 62

Future Directions and Research Agenda ... 63

Conclusion ... 64

6. ARTIFICIAL INTELLIGENCE IN WAREHOUSE ... 65

Overview ... 65

Historical Evolution of Warehousing and Modern Technology ... 66

AI Techniques for Warehousing ... 67

Core Warehouse Processes Enhanced by AI ... 68

Technology Infrastructure ... 71

Conclusion ... 74

Tasks ... 74

7. ARTIFICIAL INTELLIGENCE IN LOGISTICS AND IN TRANSPORT PROCESSES ... 76

Rising Competition and the Need for Efficiency ... 78

Innovation and Future ... 80

AI Use Cases in Logistics and Transport ... 81

Tasks ... 83

8. ARTIFICIAL INTELLIGENCE – ENABLED ROBOTS ... 84

Historical Evolution ... 85

Core Theoretical Frames ... 87

Technology Infrastructure in AI-Enabled Robots ... 90

Use Cases for AI-Enabled Robots ... 91

Future Trends in the Use of AI in Enabled Robots ... 92

Conclusion ... 93

Tasks ... 93

9. ARTIFICIAL INTELLIGENCE IN AUTONOMOUS VEHICLES IN TRANSPORT ... 95

General Principle ... 95

Advantages and Disadvantages of AI-Driven Autonomy in Transport ... 96

Use Cases ... 98

Conclusion 100

Tasks 101

10. VISUAL ARTIFICIAL INTELLIGENCE AND DEEP LEARNING TECHNOLOGY FOR OBJECT RECOGNITION 103

Deep Learning in Visual AI 106

Use Cases in Transport and Logistics 107

Conclusion 109

Tasks 110

11. VISUAL ARTIFICIAL INTELLIGENCE AND DEEP LEARNING TECHNOLOGY FOR GOODS COLLECTION AUTOMATION VIA INDEPENDENT PRODUCT IDENTIFICATION 112

Use cases 115

Conclusion 118

Tasks 119

12. THE FUTURE OF ARTIFICIAL INTELLIGENCE IN LOGISTICS AND SUPPLY CHAIN 120

Challenges 123

13. CRITICAL THINKING AND CASE ANALYSIS 126

Critical Thinking 126

Critical Thinking vs Artificial Intelligence 128

Case Analysis and Case Studies 129

Case Analysis in Critical Thinking 129

Conclusion 130

Conclusion 132

Bibliography 133

Introduction

The rapid advancement of Artificial Intelligence (AI) has significantly transformed logistics and supply chain management (SCM), leading to enhanced operational efficiency, reduced costs, and improved decision-making. The integration of Artificial Intelligence into logistics and supply chain management is transforming global business operations by enhancing efficiency, resilience, and decision-making processes. This scriptum explores the integration of AI technologies such as machine learning, computer vision, and natural language processing into various facets of logistics and supply chains, the methods, techniques, and principles underpinning AI applications in logistics and supply chains, while highlighting key distinctions between supply chains and logistics chains. It systematically examines predictive analytics, demand forecasting, inventory management, and production planning as foundations for intelligent decision-making. Furthermore, the study investigates the role of AI in warehouses, logistics, and transport processes, with emphasis on AI-enabled robotics and autonomous vehicles as revolutionary technologies driving automation and optimization. Special attention is given to visual AI and deep learning technologies for object recognition and goods collection automation, which enable independent product identification and streamline distribution networks. The future potential of AI in logistics and supply chains is critically analysed, considering both opportunities and challenges. Through case analyses and critical thinking exercises, this work provides a comprehensive framework for understanding, applying, and advancing AI solutions in modern supply chain ecosystems. The scriptum not only demonstrates the transformative impact of AI but also encourages reflective and analytical approaches to its adoption in practice and research. Moreover, the scriptum outlines the benefits, challenges, and future trends of AI adoption in SCM, emphasizing its pivotal role in enabling sustainable and resilient supply chains.

Keywords

Artificial Intelligence, Supply Chain Management, Logistics, Machine Learning, Automation, Demand Forecasting, Route Optimization, Predictive Analytics

1. SCIENTIFIC METHODS AND TECHNIQUES IN ARTIFICIAL INTELLIGENCE IN LOGISTICS AND SUPPLY CHAIN, KEY CONCEPTS AND PRINCIPLES OF ARTIFICIAL INTELLIGENCE IN LOGISTICS AND SUPPLY CHAINS

This chapter explains how artificial intelligence supports logistics and transport processes as a rigorously engineered decision system. The core idea is simple to state and demanding to implement. Predictions about the very near future feed directly into optimization routines that choose routes, assignments, and schedules, and the resulting actions are continuously checked against reality, so models and plans adapt. Transport operations live on short horizons, with time-varying traffic, stochastic customer arrivals, uncertain stop service times, and strict service windows. Because of this, the methodological backbone of the field is a closed loop that can be summarized as predict, optimize, execute, observe, update. When this loop is time aware and uncertainty aware, it consistently outperforms static planning or purely reactive dispatch in middle mile and last mile contexts (Adamo et al., 2024; Pillac et al., 2013; Soeffker et al., 2022). Two persistent facts motivate the scientific approach taken here. First, nonstationarity is normal in transport systems, which means measurable performance drift will occur unless data pipelines, validation protocols, and monitoring are in place to detect and respond to change. Second, failures in production AI are more often due to data quality, leakage, and broken assumptions than to a lack of algorithmic sophistication, which is why disciplined machine learning operations practices are essential to sustained value in operations (Gama et al., 2014; Paleyes et al., 2022).

Within this, the chapter focuses on a small set of high leverage components that most directly move operational performance. On the predictive side, we examine near term travel time estimation on road networks, stop level service time modelling, and short horizon demand or request arrival intensity forecasting. These models are built and evaluated with rolling origin procedures that respect temporal order, and they expose probabilistic outputs, for example quantiles or prediction intervals, rather than only point estimates. Probabilistic outputs allow dispatchers and optimizers to trade risk against penalties for early or late service. The transport literature shows that even relatively simple temporal models, when properly validated, improve short horizon accuracy, and that graph temporal models can add value in dense urban settings. The same literature stresses that what you choose to predict matters as much as how you predict it, because only a few quantities affect routing feasibility and lateness when inserted into solvers (Vlahogianni et al., 2014; Adamo et al., 2024; Soeffker et al., 2022). On the decision side, the emphasis is on time dependent vehicle routing and dispatch that re-optimizes frequently and on anticipation rather than pure reaction.

Approximate dynamic programming and reinforcement learning provide a principled way to learn the value of preserving time and capacity for likely future requests. Deployed policies then act myopically at millisecond scale but are guided by offline learned value estimates, which reduces lateness and empty miles without sacrificing dispatch latency (Ulmer, 2018; Ulmer et al., 2019; Pillac et al., 2013).

The chapter also situates road transport within intermodal operations where similar principles apply but decisive variables differ. In ports and rail terminals, accurate predictions of vessel estimated time of arrival and port call duration unlock better berth allocation, quay crane scheduling, and yard planning, and they shorten truck turn times when combined with coherent truck appointment schedules. Reviews and recent studies document how these predictions are integrated with terminal optimization so that quay, yard, and gate plans remain synchronized despite variability in arrivals and handling times (Evmides et al., 2024; Bierwirth & Meisel, 2015). Finally, evaluation and rollout are treated not as afterthoughts but as first class components. Digital twins and realistic simulators are used to test policy changes before deployment, while champion challenger releases, switchback tests, and focused monitoring catch regressions early. Because many useful estimated time arrival (ETA), dwell, and demand signals span firm boundaries, the overview closes with a brief note on privacy preserving collaboration, where federated learning and differential privacy enable joint model improvements without moving raw data (Ivanov & Dolgui, 2021; Kairouz et al., 2021; Dwork & Roth, 2014).

Artificial intelligence delivers its largest gains in logistics when it is treated as a decision system rather than a collection of disconnected models. Transport operations are dominated by time-varying conditions (traffic, arrivals, dwell, no-shows) and tight service constraints, so the central methodological pattern is a closed loop of *predict → optimize → execute → observe → update*. In this loop, models forecast the quantities that operators can actually control around the next few hours; link speeds, stop service times, request arrival intensities, port-call durations and inject those forecasts directly into routing, dispatching, and scheduling optimizers that can re-plan on short horizons. The loop then monitors reality, measures regret, and adapts to drift. Long-run experience and the research record both show that this “predict-and-optimize” framing, when engineered with re-optimization and uncertainty awareness, consistently outperforms static plans or purely reactive heuristics in middle- and last-mile transport (Adamo et al., 2024; Pillac et al., 2013; Soeffker et al., 2022). Two practical consequences follow. First, time must be in the model; networks and service processes change within the day so the optimization layer should be time-dependent. Second, uncertainty must be quantified and propagated, because detours, early/late arrivals, and cancellations are endemic in real fleets; ignoring this variability leads to brittle plans and escalating ad-hoc overrides (Gama et al., 2014; Paleyes et al., 2022). Within this loop, the highest-return modelling work focuses on a small set of operationally decisive predictions and on how their uncertainty flows into decisions. Short-term traffic forecasting provides link or corridor travel times by time-of-day and day-of-week, often with features for incidents, weather, and special calendars; even simple temporal models or gradient-boosted trees can yield reliable short-horizon accuracy when trained with careful temporal cross-validation and evaluated with rolling origins, while graph-temporal networks add value on dense urban grids. Crucially, outputs should be probabilistic (quantiles or full predictive intervals), not just point estimates, so dispatch can trade off risk and service penalties when selecting routes and appointment slots (Vlahogianni et al., 2014; Paleyes et al., 2022). A second high-leverage target is stop-level service time, which typically depends on parcel mix, building type, elevator availability, and customer behaviour; accurate service-time distributions stabilize crew plans and protect on-time performance when inserted into time-dependent VRP solvers (Adamo et al., 2024). A third is the arrival process itself: for same-day and on-demand operations, forecasting the intensity and spatial distribution of future requests enables capacity reservation, proactive load balancing across zones, and dynamic promise windows that reflect true feasibility rather than static rules (Pillac et al., 2013; Soeffker et al., 2022). Across all three prediction tasks, the engineering discipline like data lineage, drift detection, champion–challenger models often dominate the choice of algorithm, because the costliest failures in deployed AI systems usually arise from data shifts and broken assumptions rather than from model under fitting (Gama et al., 2014; Paleyes et al., 2022).

On the decision side, dynamic routing and dispatch illustrate why anticipation matters. Re-optimizing routes whenever events arrive is necessary but not sufficient; naively serving the nearest or earliest request can starve later high-value demand or cause excessive detours. Approximate dynamic programming and reinforcement learning resolve this tension by learning a value-to-go function offline; how much future opportunity a vehicle "consumes" by taking a particular job now and then using that learned value to guide fast online accept/assign/move decisions. This offline–online split has been shown to deliver non-myopic behaviour that reduces lateness and empty miles in courier and same-day settings, while still meeting strict latency constraints at dispatch time (Ulmer, 2018; Ulmer et al., 2019; Soeffker et al., 2022). The same idea underpins large-scale fleet repositioning in ride-hailing, where multi-agent deep reinforced learning improved earnings and wait times in live systems, evidence that transfers well to crowd-shipping and urban freight when properly constrained (Lin et al., 2018). In practice, the most robust transport controllers combine these learned value estimates with time-dependent VRP re-optimization, capacity guards, and simple, auditable rules for fairness and driver well-being, yielding policies that are both effective and explainable to operations teams (Adamo et al., 2024; Pillac et al., 2013).

Intermodal nodes: ports, rail ramps, and large cross-docks benefit from the same predict-and-optimize pattern, but the decisive variables differ. Accurate ETA and port-call duration predictions from AIS and operations logs allow terminals to plan berths, crane gangs, and yard equipment proactively, smoothing peaks and shortening truck turn times; these predictions then feed berth allocation, quay-crane scheduling, and yard storage/retrieval plans so that quay, yard, and gate operate in concert rather than as silos (Bierwirth & Meisel, 2010; Bierwirth & Meisel, 2015; Carlo et al., 2014). Landside, truck appointment systems and drayage dispatch achieve their performance only when appointment slot supply is tied to realistic handling capacities inferred from data and when dispatch heuristics exploit street-turns and time-window flex where feasible; learning-augmented heuristics are increasingly used to stabilize these decisions under volatile arrivals (Chen et al., 2022; Soeffker et al., 2022). The key is again closed-loop coherence: predictive models inform feasible schedules, schedules constrain promises to carriers and shippers, and monitoring detects slippage early so yard plans and appointment offers are adjusted before queues form (Bierwirth & Meisel, 2015; Carlo et al., 2014).

Because transport systems drift with seasons, networks, and behaviour, evaluation and rollout must mirror reality and safeguard service. Models and policies should be validated with rolling-origin backtests for forecasts and with counterfactual or simulation-based evaluation for control policies, ideally inside a digital-twin environment that reproduces arrival processes, capacities, and execution variability. Digital twins make it possible to probe “what-if” changes—altered time windows, different batching rules, new zoning, or revised appointment schedules before impacting live operations, and they shorten the diagnose-and-fix cycle when disruptions occur (Ivanov & Dolgui, 2021; Paleyes et al., 2022). In production, champion–challenger deployments, switchback or interleaved A/B tests, and progressive widening of coverage provide a safe path to scale; rigorous monitoring of input drift, constraint violations, and business KPIs catches degradations early and triggers retraining or rollbacks (Gama et al., 2014; Palyes et al., 2022). Finally, since high-quality ETA, dwell, and demand models often require cross-firm signals, collaboration must respect confidentiality: federated learning avoids moving raw data while still producing global models, and differential privacy provides formal guarantees when sharing metrics or releasing benchmarks, practical tools for regulated logistics ecosystems (Kairouz et al., 2021; Dwork & Roth, 2014).

Taken together, these practices amount to a compact set of principles for AI in transport logistics: model the clock and the uncertainty explicitly; predict the handful of quantities that move the operational “control knobs”; turn those predictions into decisions with time-dependent, constraint-aware optimizers; anticipate rather than merely react by learning value-to-go offline and acting fast online; evaluate end-to-end in digital twins before rollout; and engineer for drift, safety, and privacy from day one. Organizations that institutionalize this loop see fewer exceptions, steadier on-time performance, shorter dwell and turn times, and lower empty miles, not because any single model is exotic, but because the system as a whole is coherent, adaptive, and aligned with how transport actually runs (Adamo et al., 2024; Ulmer, 2018; Ulmer et al., 2019; Soeffker et al., 2022; Ivanov & Dolgui, 2021).

2. SUPPLY CHAIN VS. LOGISTICS CHAIN

Supply-chain management (SCM) and logistics represent hierarchically related yet analytically distinct constructs. Logistics coordinates the operational flow and storage of materials to meet prescribed service levels, whereas SCM architects and controls the end-to-end configuration of organisations, processes and information that convert latent demand into customer value. Recognising their structural interdependence is essential to contemporary competitive strategy. This chapter therefore traces their historical evolution, theoretical foundations, current practice and emerging trajectories, highlighting the managerial imperatives of an integrated perspective.

Concise overview of supply chain issues

Supply-chain management (SCM) is the coordinated design and control of all material, information and financial flows that transform raw resources into finished offerings and deliver them to the final customer. A widely cited academic definition frames SCM as:

> *"a set of three or more entities (organizations or individuals) directly involved in the upstream and downstream flows of products, services, finances, and/or information from a source to a customer"* (Mentzer et al., 2001).

Beyond mere cost containment, an effective supply chain is now recognised as a primary source of sustainable competitive advantage: by viewing the entire network of suppliers, manufacturers, distributors and retailers as an integrated value-creation system, firms can outpace rivals on responsiveness, reliability and customer satisfaction rather than price alone (Christopher, 2005).

Historical Evolution of Supply Chain

Before the commercial adoption of the "supply chain," the concept was embedded in the context of warfare: historians of the U.S. military called it "the art of defining and extending the possible," emphasizing that success on the battlefield depended on the synchronization of supply, transportation, and information-principles that later carried over into business networks (Southern, 2011).

1950s - Transportation Era

Business attention centred on regulated freight rates, modal economics and the new Interstate Highway System. University programmes taught "traffic management," and logistics meant little more than moving goods from A to B (Southern, 2011).

1960s - Physical Distribution

In the 1960s logistics broadened from pure transport to physical distribution and, soon after, business logistics. Outbound practice gained a professional home with the 1963 founding of the National Council of Physical Distribution Management, while inbound purchasing was championed by the National Association of Purchasing Management; their journals ultimately became today's *Journal of Supply Chain Management*. Seminal textbooks by Smykay-Bowersox-Mossman (1961) and Heskett-Ivie-Glaskowsky (1964) framed the field, as the *Transportation Journal* (1961) and the creation of the U.S. Department of Transportation (1966) kept transport issues in the spotlight. These institutions, publications and regulatory shifts laid the groundwork for modern supply-chain management (Southern, 2011).

1970s - Physical Supply & Integrated Logistics

In the 1970s, materials management (physical supply) surged in importance and was soon integrated with outbound distribution under the broader label of business logistics. Universities introduced dedicated curricula, and new journals most notably the Journal of Business Logistics (1978) alongside an expanded Transportation Journal provided a scholarly backbone. Transportation issues stayed centre-stage: rail bankruptcies spurred federal take-overs just as Congress launched deregulation of rail and air carriers. Foundational textbooks by Heskett, Ballou, Coyle and others codified these shifts, firmly establishing logistics as a comprehensive managerial discipline (Southern, 2011).

1980s - Deregulation & Business Logistics

During the 1980s U.S. transport deregulation accelerated. The Motor Carrier Act (1980) eased entry barriers and rate controls for trucking, while the Staggers Rail Act (1980) granted railroads contractual freedom and reduced Interstate Commerce Commission oversight. The Civil Aeronautics Board was dissolved in 1984, and the federal government divested its Conrail shares in 1987, collectively fostering a more competitive and adaptable transport sector. Simultaneously, professional vocabulary shifted: "physical distribution" gave way to the broader term "logistics." This transition was reflected in the retitling of Johnson and Wood's textbook to *Contemporary Physical Distribution and Logistics* (1982) and the 1985 rebranding of the National Council of Physical Distribution Management as the Council of Logistics Management (Southern, 2011).

1990s - Technology-Enabled Logistics

In the 1990s, firms embraced the total-cost approach to business logistics, exploiting carrier negotiations and systems thinking to unlock savings. Reflecting this broadened scope, Johnson & Wood's flagship text was retitled Contemporary Logistics (1990), and many transport providers began marketing themselves as "logistics solutions" rather than mere carriers. Technological leaps

Internet connectivity, electronic data interchange and the rise of third-party logistics, strategic alliances and partnerships embedded logistics in corporate strategy. Scholarship kept pace: leading textbooks by Ballou (1998), Coyle & Bardi (1999) and others codified new practices, while Journal of Business Logistics analyses showed research gravitating toward inventory control, IT applications and customer service. Parallel reviews in the Journal of Supply Chain Management charted a shift in purchasing research from legal concerns to strategic sourcing and supplier alliances, signalling the discipline's evolution toward holistic supply-chain thinking (Southern, 2011).

2000s - Transition to Supply-Chain Management

Early-21st-century discourse shows a gradual but uneven shift from logistics to supply-chain management (SCM); large corporations adopted the broader lens first, while many SMEs trailed. Professional bodies remain pivotal, especially CSCMP, which frames SCM as the integrated planning and control of sourcing, conversion and all logistics activities, enabled by shared information across vendors, manufacturers and customers. Contemporary texts such as Supply Chain Logistics Management (Bowersox, Closs & Cooper 2010) note that logistics has moved "from the warehouse floor to the boardroom," while titles by Coyle et al. now pair logistics and transportation explicitly with an SCM perspective (Southern, 2011).

Core Theoretical Frames

Modern scholarship distils supply-chain behaviour into two complementary lenses: a process framework that maps what firms do and a strategic-fit framework that prescribes how they should configure networks for different demand conditions (Fisher, 1997), (APICS, 2017).

Process lens – SCOR

The Supply-Chain Operations Reference (SCOR) model defines six canonical processes – Plan, Source, Make, Deliver, Return and Enable, and decomposes each into hierarchical activities linked to standard performance metrics. Firms can therefore benchmark, diagnose bottlenecks and align technology investments against a common vocabulary. Subsequent empirical studies show that SCOR improves cross-functional communication and shortens improvement cycles because managers can trace service, cost and carbon outcomes back to discrete workflow nodes (Stephens, 2001).

The 2022 SCOR Digital Standard (SCOR-DS) replaces the linear Plan–Source–Make–Deliver–Return scheme with a digital, infinity-loop architecture anchored on seven top-level processes – Orchestrate, Plan, Order, Source, Transform, Fulfill and Return. "Orchestrate" now embeds strategy, data governance, talent and ESG, while the split of Deliver into Order and Fulfill and the renaming of Make to Transform broaden the model to multichannel service and circular flows. SCOR-DS retains the familiar hierarchical decomposition and benchmarking logic, but hard-wires IoT, AI and real-time analytics as baseline capabilities, positioning supply chains for data-driven resilience, customer-centricity and sustainability (Becker, 2023).

Strategic-fit lens – Fisher typology

Fisher's strategic-fit typology draws a clean line between functional products (predictable demand, low margins) and innovative products (volatile demand, high margins). Functional items belong in efficient, lean supply chains that drive unit cost down through high utilisation and long runs; innovative items require responsive, agile chains that trade capacity buffers and rapid replenishment for speed and flexibility. Matching product type to chain design lifts both service and profit, whereas any "misfit" quickly erodes performance. Recent empirical work (e.g., Mahdavi et al., IJPE, 2023) reconfirms the framework's validity.

In addition to SCOR and Fisher, other supply chain frameworks currently exist. These are for example:

- Triple-A Supply Chain (agility, adaptability, alignment) links superior cost and service to three dynamic capabilities that allow networks to sense change and realign incentives.
- The Sustainable Supply Chain framework embeds economic, environmental and social metrics, framing ESG performance as a source of strategic value. Together, these lenses add capability, risk and sustainability dimensions to classic process- and demand-based views (Lee, 2004; Carter & Rogers, 2008).

Emerging & Future Trends

Supply-chain priorities are shifting from cost efficiency toward data-driven resilience and sustainability. Three technology-policy trajectories dominate the current research agenda.

Digital-twin orchestration

Systematic reviews show that virtual replicas of end-to-end networks—digital twins—now integrate real-time IoT data, physics-based simulations and predictive analytics to test "what-if" scenarios before physical execution. Early adopters report shorter disruption-recovery times and double-digit inventory reductions because planners can visualise ripple effects across sourcing, production and logistics simultaneously (Zaidi et al., 2024).

AI-enabled resilience

Recent studies model artificial-intelligence agents that learn demand shocks, optimise multi-echelon safety stocks and recommend dynamic routing during crises. The evidence indicates that AI-driven planning raises service levels 5-10 % without additional capacity by reallocating slack in real time, thereby hard-wiring agility into otherwise lean networks (Tang et al., 2025).

Circular and low-carbon networks

Research on circular-economy supply chains highlights closed-loop design, regenerative materials and cross-industry symbiosis as next-generation competitive levers. Firms are embedding life-cycle carbon metrics into sourcing and logistics decisions, aligning with net-zero mandates while unlocking secondary-market revenue streams (Fröhlich et al., 2024).

Together, digital twins provide visibility, AI delivers adaptive intelligence and circular-economy logic embeds sustainability, framing a future in which supply chains are simultaneously smart, resilient and regenerative.

Concise overview of logistics chain issues

A logistics chain is the end-to-end sequence of activities that plans, implements and controls the forward-and-reverse flow of materials, services and related information from origin to final consumption to satisfy customer requirements. It covers inbound logistics (materials management), internal handling and outbound physical distribution, treating time and place utility as its core value proposition. Unlike broader supply-chain constructs, the logistics chain zeroes in on operational execution and customer-oriented service performance (Tseng et al., 2005).

Classical analysis groups logistics-chain activities into four tightly linked pillars: transportation, inventory/warehousing, order processing and information systems. Transportation is the dominant cost driver often one-third of total logistics expenditure and determines network speed and reliability; warehousing and inventory decisions balance holding costs against service levels; order processing translates demand signals into physical flows; and real-time information synchronises the whole system. Managers therefore confront a continuous cost-service trade-off, seeking the lowest total landed cost that still meets promised availability and lead-time targets (Rushton et al., 2014).

A widely cited academic definition frame Logistic Chain as:

> "*Logistics is that part of the supply-chain process that plans, implements and controls the efficient, effective flow and storage of goods, services and related information from the point of origin to the point of consumption in order to meet customers' requirements*" (Lambert & Cooper, 2000).

Effective logistics-chain design integrates these pillars through standardised processes, shared performance metrics and enabling technologies such as routing optimisation and warehouse automation. Current best practice measures success via total logistics cost, on-time delivery, fill rate and carbon intensity, emphasising cross-functional coordination over functional silos. In sum, a well-managed logistics chain converts operational efficiency directly into market responsiveness and competitive advantage for the firm (Tseng et al., 2005; Rushton et al., 2014).

Relationship between the Supply Chain and the Logistics Chain

Scholars now view the logistics chain as a specialised component nested within the wider supply-chain framework. Earlier practice equated supply-chain management (SCM) with logistics simply stretched upstream to suppliers and downstream to customers a view reinforced by the Council of Logistics Management's 1986 definition, which already spanned flows "from origin to consumption." The overlap caused confusion because logistics can denote both an internal function and a network-wide activity. By the late 1990s, research reframed SCM as the coordination and governance of multiple inter-firm business processes far beyond physical movement, leading the CLM to specify logistics as just one element of SCM.

Consequently, logistics manages the efficient movement and storage of goods, services and information, while SCM synchronises those logistics tasks with product development, customer relations, demand planning and other cross-organisational processes. In essence, every logistics chain operates inside a broader supply chain, and competitive advantage now hinges on integrating logistics proficiency with end-to-end process management (Lambert & Cooper, 2000).

The main relationships between the supply chain and the logistics chain include:

- *Operational foundation* – Logistics governs the day-to-day physical flow of goods through transportation, warehousing, inventory and last-mile delivery; supply-chain management (SCM) builds on that base by coordinating all parties—suppliers, manufacturers, distributors and customers—to optimise network efficiency and responsiveness (Nguyen et al., 2025).
- *Subsystem relationship* – Logistics is treated as a core component embedded inside SCM; beyond logistics activities, SCM adds procurement, production planning, demand forecasting and relationship management to create end-to-end value (Nguyen et al., 2025).
- *Shared performance logic* – Effective SCM depends on accurate, timely logistics flows: practitioners agree that controlling the movement of materials and information within logistics provides the essential data and service capability on which supply-chain decisions and customer satisfaction rest (Sweeney et al., 2018).

The main differences between supply chains and logistics chains are:

- *Scope* – Logistics chains plan, execute and control the efficient flow & storage of goods and related information inside a single enterprise; supply-chain management adds sourcing, transformation and inter-firm coordination, extending those flows across multiple firms (Lummus et al., 2001).
- *Strategic focus* – Logistics is chiefly operational and cost-driven, whereas supply-chain management is strategic, seeking network-wide value creation and competitive advantage (Larson & Halldorsson, 2004).
- *Types of flows governed* – Logistics manages physical movements (and their data); supply-chain management orchestrates physical, information and financial flows while aligning relationships among partners (Sweeney et al., 2018).
- *Time horizon* – Logistics decisions are short- to medium-term and execution-oriented; supply-chain management encompasses long-term network design and continuous improvement (Larson & Halldorsson, 2004).

Trends and Differences in the Use of Modern Technologies in Supply and Logistics Chains

Network-level analytics vs. facility execution

Supply-chain teams increasingly deploy *digital-twin platforms* that simulate end-to-end networks for resilience and carbon optimisation, whereas logistics units tend to apply twin technology only to individual warehouses or fleet assets for local efficiency gains (Freese & Ludwig, 2025).

AI use-case focus

In supply chains, artificial-intelligence tools target strategic tasks such as demand forecasting, multi-echelon inventory optimisation and disruption recovery; logistics operations apply AI to real-time routing, robot coordination and order-picking automation on the shop-floor (Culot et al., 2024).

Automation intensity

Warehouse and transport functions invest heavily in AGVs, AMRs and automated storage-and-retrieval systems to speed fulfilment and cut labour, while supply-chain control towers emphasise cloud analytics and cross-partner orchestration rather than physical robotics (Ellithy et al., 2024).

Traceability technologies

Blockchain pilots in supply chains aim to secure provenance and financial flows across multiple firms; logistics chains more commonly rely on RFID/barcode and IoT sensors for item-level tracking within a single enterprise or leg (Ellithy et al., 2024; Freese & Ludwig, 2025).

Sustainability data capture

Supply-chain dashboards aggregate lifecycle carbon metrics to inform sourcing and design decisions, whereas logistics telematics focus on fuel efficiency, load consolidation and route optimisation to shrink the operational footprint (Culot et al., 2024; Ellithy et al., 2024).

Conclusion

Modern scholarship distinguishes logistics as the operational core that manages transportation, warehousing, inventory, and information flows, with supply chain management considered as an umbrella discipline that organizes these logistics functions along with supply, product fulfilment, financial coordination, and linkages to customers across multiple firms. Logistics has historically matured from a stand-alone transportation management to an integrated business practice. Supply chain thinking then extended this foundation to managing processes across businesses, emphasizing collaboration and joint performance measurement.

Current research shows that cost leadership alone is not enough. Competitive advantage now depends on synchronizing logistics accuracy with network-wide agility, risk mitigation and sustainability goals. Digital technologies - from real-time telemetry and autonomous material handling devices to predictive analytics and cloud-based control towers - enable detailed insight and rapid decision support, but their focus is diverging. Logistics applications focus on local efficiency and execution, while supply chain initiatives leverage the same data to optimize end-to-end configuration, partner alignment and carbon accountability. Effective organizations therefore integrate thorough logistics execution into a strategically coordinated, technology-enabled supply chain architecture to achieve resilience, responsiveness and responsible growth.

Tasks

1. Identify and describe other theoretical frameworks in the field of supply chains. Choose one and characterize it.
2. Find a current case of the use of modern technology in supply chains and describe it.

3. ARTIFICIAL INTELLIGENCE

lan Turing's 1950 paper "*Computing Machinery and Intelligence article*" did more than introduce the Imitation Game, it quantified the very challenge of machine intelligence. Turing opened with the question "Can machines think?" but, rather than wrangle over the definitions of "machine" and "think," he proposed a concrete experiment - the Imitation Game, now universally known as the Turing Test. In the paper he predicted: "I believe that in about 50 years' time it will be possible, to programme computers to make them play the imitation game so well that an average interrogator will not have more than 70 % chance of making the right identification after 5 min of questioning". By framing intelligence as an empirically testable property instead of a metaphysical abstraction, Turing laid the groundwork for artificial intelligence (AI) as an experimental science (Turing, 1950; Warwick & Shah, 2015).

Half a decade after Turing's essay, a quartet of young computer scientists John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon circulated "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence". In just six pages they asked the Rockefeller Foundation for enough money to bring "ten men" to Hanover, New Hampshire, for eight weeks in the summer of 1956, and in the process introduced the phrase *artificial intelligence (AI)* to the scientific lexicon. Their audacious premise was that every component of intelligence, from learning to abstraction consist in processes that could be described so precisely that it could be engineered into a machine (Dick, 2019). There is a plaque at Dartmouth College that reads:

> "In this building during the summer of 1956 John McCarthy (Dartmouth College), Marvin L. Minsky (MIT), Nathaniel Rochester (IBM), and Claude Shannon (Bell Laboratories) conducted the Dartmouth Summer Research Project on Artificial Intelligence. First use of the term "Artificial Intelligence". Founding of Artificial Intelligence as a research discipline "to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it (McCarthy et al., 1955)"".

The plaque enshrines the standard account of the history of Artificial Intelligence that it was born in 1955.

Within that milieu, Newell, Shaw & Simon's *Logic Theorist* (1956) and subsequent *General Problem Solver* formalised *means-ends* search, demonstrating that symbolic heuristics could prove 38 of the first 52 *Principia Mathematica* theorems and inspiring the search algorithms still taught today (Newell et al., 1956). Early optimism grew with Joseph Weizenbaum's *ELIZA* (1966), which showed that pattern-matching and script-based re-assembly rules could sustain open-ended dialogue, foreshadowing today's chat interfaces while also revealing the thin veneer of understanding behind syntactic tricks (Weizenbaum, 1966). Yet connectionist enthusiasm faltered when Marvin Minsky & Seymour Papert's 1969 monograph *Perceptrons* mathematically proved that single-layer perceptrons cannot solve linearly inseparable problems such as XOR, redirecting U.S. funding away from neural nets toward symbolic AI for more than a decade (Minsky & Papert, 1969).

The exclusive-OR (often written "XOR" or with the symbol ⊕) is a basic Boolean operation that returns *true* when its two inputs differ and *false* when they are the same. Put another way, XOR outputs "1" only for the input pairs (0, 1) or (1, 0); it outputs "0" for (0, 0) and (1, 1). Digital-electronics texts list XOR alongside AND, OR and NOT as one of the four canonical two-input logic gates, complete with a distinctive "double-curved" gate symbol and the truth-table just described (Elahi, 2022). Why does it matter in AI? XOR became famous in machine-learning history because it exposed a critical limitation of the 1950--60s *perceptron*, the first learning neuron. Marvin Minsky and Seymour Papert proved that a single-layer network cannot learn XOR: the two "1" cases sit on opposite corners of a square so no straight line can separate the "1"s from the "0"s. Their analysis showed that some real-world problems need *non-linear* decision boundaries and therefore deeper or otherwise more complex networks, a result that helped push neural-network research into a decade-long slump and ultimately motivated the development of multilayer back-propagation in the 1980s (Vakalapoulou et al., 2023).

Knowledge-Based Era

From the mid-1960 s to the early 1990 s, AI research entered what is now called knowledge-based era, grounded in the assumption that intelligence could be captured as explicit rules operating on symbolic representations. Progress persisted. Hans Berliner's ***BKG 9.8*** backgammon program stunned Monte Carlo in 1979 by defeating world champion Luigi Villa 7-1. It was the first time any computer had beaten a reigning human world champion at a board game, using selective search and probabilistic roll-out evaluation (Berliner, 1980).

Even with this huge progress and despite headline successes, three systemic weaknesses emerged:

1. *Brittleness* - small input deviations could trigger cascades of rule misfires, a problem documented in early pattern-recognition critiques (Musen & van der Lei, 1988).
2. *Knowledge-acquisition bottleneck* - eliciting, verifying and maintaining thousands of rules required labour-intensive interviews with scarce experts. A 1988 reassessment identified it as the costliest phase of deployment (Cullen & Bryman, 1988).

3. *Combinatorial explosion* - symbolic search grew exponentially outside narrow toy domains, a criticism formalised in Sir James Lighthill's 1973 report, whose conclusions precipitated a funding retreat in the United Kingdom and helped usher in the first AI winter (Lighthill, 1973).

Parallel-distributed processing research resurrected neural networks when Rumelhart, Hinton and Williams (1986) provided the first general-purpose proof that back-propagation could train multilayer perceptrons to learn internal representations. The shift to data-driven thinking began as they showed that the *back-propagation* algorithm could adjust every connection in a multilayer neural network so that the network built its own internal features: edges, phonemes, abstract concepts directly from examples. Their results shattered the belief that hand-engineered feature detectors were indispensable and revived interest in neural nets after a long symbolic interlude. Crucially, back-prop demonstrated that learning need not stop at the surface: stacked layers could refine rough sensory correlations into hierarchies of representation deep enough to solve problems that once seemed to require explicit logic (Rumelhart et al., 1986).

During the same period, Vapnik's support-vector machines and Bayesian networks supplied principled statistical frameworks with provable generalisation bounds, shifting emphasis from rule enumeration toward data-driven pattern induction. Vapnik's *structural-risk minimisation* principle shows that a wide margin reduces the bound on test-error even if the classroom of possible functions is huge. Bayesian networks added a complementary probabilistic grammar for reasoning under uncertainty. Together these tools replaced manual rule enumeration with principled pattern induction. Models were now selected because theory guaranteed they could generalise, not because experts enumerated every contingency (Nasien et al., 2010).

Foundation-Model Era

Back-prop and margin theory matured just as cheap compute and internet-scale datasets arrived. The result is the *foundation-model* era, in which hundreds of billions of learned weights capture the statistical texture of the web and can be *prompted* to write code, design molecules or steer robots. These systems remain data-driven at heart:

- Pre-training forces the network to predict missing tokens or pixels, distilling linguistic, visual and causal patterns straight from raw corpora.

- Fine-tuning and reinforcement learning from human feedback add small, task-specific traces that bend the general model toward objectives (helpfulness, harmlessness, domain style) without rewriting its entire knowledge base.
- Scaling laws, empirical curves that show error falling as a simple power of data, parameters and compute and provide the quantitative blueprint for how much new data is worth.

Because the model's competence is literally written into its weights, the decisive resource is the quality and diversity of the data itself. That insight has spawned the "data-centric AI" movement: fix mislabelled records, balance demographic slices, augment rare cases, and the very same algorithm can leap forward in performance with zero architectural change.

Data-Driven Approach

Data-driven does not mean data-only. Current research stitches the pattern-finding strengths of deep networks to symbolic reasoning, physical simulators and retrieval systems that inject grounded facts at query time. What remains non-negotiable, however, is that every new capability begins with a model that has soaked up the statistical fingerprints of a vast training corpus and tuned its billions of micro-parameters to fit that evidence. In this sense, the term data-driven captures the central dogma of contemporary AI, the road to intelligence lies through the patterns latent in data, not through the rules that human designers can enumerate.

Data-driven learning rests on the idea that the data themselves already contain the regularities needed. Instead of hand-coding rules for every situation, large collections of input and output examples such as chest X-rays with diagnoses, sentences with translations, and factory states with the next best action are gathered, and an algorithm sifts through them until it discovers an internal rule set that reproduces the patterns it sees. This discovery workflow, often called data mining, begins with collecting and cleaning raw records so missing values, duplicates, and obvious errors do not mislead the learner. Next, a learning goal is selected: classification if the records carry labels, clustering or association discovery if they do not. The algorithm then explores a vast "hypothesis space" of possible models such as decision trees, support-vector boundaries, and neural-network weight settings, retaining only those whose predictions remain reliable on a separate test split, a practice that guards against overfitting. What matters is not how many explicit rules the final system stores but how well it generalises to fresh, unseen cases, an idea captured by the standard definition that a program learns from experience E with respect to a task T and performance measure P if its performance at T, measured by P, improves with experience E. Once these hidden regularities are uncovered, the trained model can diagnose tumours, translate paragraphs, or optimise production schedules, turning raw observations into actionable knowledge without manual rule writing (Awad & Khanna, 2015; Zhao, 2012).

Artificial Intelligence Definition

Artificial intelligence is both a scientific inquiry into the nature of intelligent behaviour and an engineering discipline aimed at building artefacts that display it. John McCarthy's AI is framed as "the science and engineering of making intelligent machines, especially intelligent computer programs", while other authors extend that idea by defining AI as the study of rational agents that perceive and act in an environment. These statements anchor AI in empirical method and algorithmic design rather than philosophical speculation (McCarthy, 2007).

AI is best understood as an umbrella discipline that seeks to build computational artefacts able to carry out functions normally associated with human cognition such as perceiving the environment, reasoning about it, communicating with natural language, planning actions, analysing and learning from experience. The discipline has also become an economic force. Independent market analyses place the global AI sector between US $243 billion and US $638 billion in 2025, with compound annual growth rates ranging from 20 % to 35 % and projections exceeding US $800 billion by 2030 as enterprises embed AI in software, data-centre hardware and cloud services. Gartner further predicts that AI-optimised servers alone will help push worldwide IT spending past US $5.4 trillion in 2025 (Hale, 2025; Zoting, 2025).

Early formal definitions, such as John McCarthy's description of AI emphasise the engineering goal, while modern textbook treatments frame it as a scientific quest to understand and replicate intelligent behaviour in all its diversity. Contemporary surveys and handbook chapters organise the field into several interlocking domains, or branches, as shown in Figure 3.1, however, their exact application and interpretation depend on the specific solution:

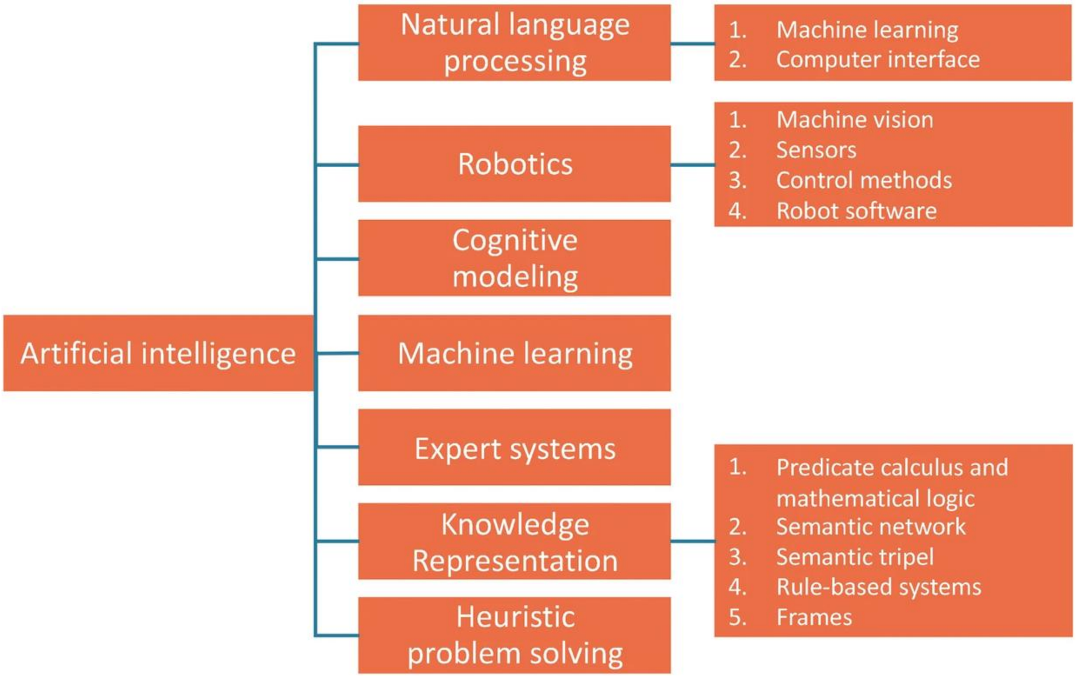

Figure 3.1 Artificial intelligence domains (Weber, 2023).

Natural language processing (NLP)

Techniques that let computers read, write and converse in human languages. Current NLP systems pair ML with specialised interfaces for text so they can translate, answer questions and hold dialogue, recent transformer architectures have pushed performance across dozens of languages (Sadiku & Musa, 2021).

Robotics & embodied AI

Combination of mechanical and electrical engineering, information technology, and computer science including artificial intelligence to design, build, operate, and apply robots. It focuses on the hardware and the computer systems that give robots perception, control, sensory feedback, and information-processing abilities. Viewed through the lens of embodied AI, robotics integrates those same perception, control, and planning modules so a machine can act autonomously in the physical world; current front-line work ranges from dexterous manipulation and agile legged locomotion to close human-robot collaboration (Baker-Brunnbauer, 2023; Weber, 2023).

Cognitive modelling

A cognitive system aims to approximate biological cognitive processes (mainly human) for the purposes of understanding and prediction. Cognitive models typically focus on a single cognitive phenomenon or process (e.g., list learning), examine how two or more processes interact (e.g., visual search and decision making), or provide behavioural predictions for a specific task or tool (e.g., how the introduction of a new software package affects productivity) (Weber, 2023).

Machine learning (ML)

Data-driven methods that induce models from examples. ML itself branches into supervised learning (classification/regression), unsupervised learning (clustering, dimensionality reduction) and reinforcement learning (sequential decision-making), with deep learning, multilayer neural networks providing today's dominant function-approximation toolkit (Weber, 2023).

Knowledge representation, expert systems & reasoning

Formal logics, ontologies and rule-based systems that support inference under certainty or uncertainty, enabling expert systems, semantic web agents and modern neuro-symbolic hybrids (Torra et al., 2018).

Heuristic problem-solving, search & planning

A problem is framed by three elements: an initial state, a goal condition, and a set of operators that move the system from one state to another. Once the problem is structured this way, search procedures seek a sequence of operators that converts the starting state into a state that satisfies the goal. When classical exhaustive search is too slow or when an exact path is unnecessary developers rely on heuristics: rules of thumb that guide the search toward promising regions of the state space, trading perfect optimality and completeness for dramatic gains in speed. These trade-offs are often pictured as a triangle whose corners represent the competing goals of optimality, completeness (accuracy) and solution time. Heuristic methods choose a balance point inside that triangle, delivering good-enough answers quickly when fully optimal solutions would be computationally prohibitive. Effective heuristics depend on solid knowledge representation and logic. This complementary field studies how to encode information about the environment facts, relationships, causal rules in a form that a computer can use to reason. Drawing on insights from psychology about human problem-solving and from formal logic about inference, it provides the symbols, ontologies and rule systems that make complex tasks such as medical diagnosis or natural-language dialogue tractable. In practice, good representations supply the structured knowledge that heuristics exploit, while logical reasoning engines can apply those representations to prune irrelevant branches, rank candidate actions and justify the solutions that heuristic search uncovers (Torra et al., 2018; Weber, 2023).

However, generally artificial intelligence is often described as a set of concentric fields, each one more specialised than the last. The broad outer layer is AI itself, the quest to give machines the ability to sense, reason and act. Within that is machine learning, where systems acquire those abilities by learning from data rather than following hand-written rules. Nested inside is the domain of neural networks, models built from layers of simple, interconnected neurons that can capture complex relationships. At the very core sits deep learning, which stacks many neural layers so the network can form its own hierarchy of features directly from raw images, audio or text. Everything done in deep learning is therefore part of neural-network research, all neural networks belong to machine learning, and machine learning is one powerful, though not exclusive approach under the wider umbrella of artificial intelligence. This representation is visualised on figure 3.2.

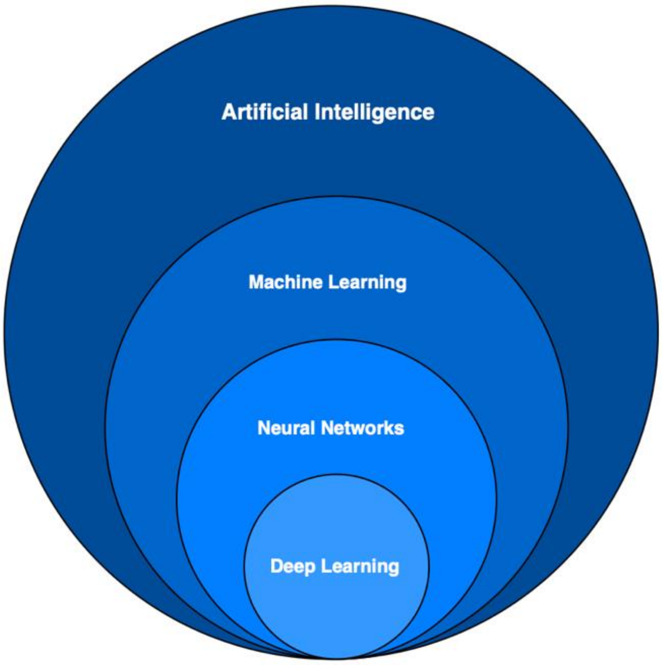

Figure 3.2 Representation of AI and its fields.

Deep Learning

Deep learning is a way of building computer programs that learn by stacking many simple neurons into dozens or even hundreds of layers. The first layers look for very small patterns, short edges in a picture or tiny sound waves in speech, while higher layers combine those pieces into larger ideas such as eyes, whole faces, or complete spoken words. Because the layers learn from one another, people no longer hand-write rules; the network discovers what matters simply by studying lots of examples. The learning process runs in a daily loop. A batch of training examples (images, sentences, data, sensor readings) is pushed through the network, which makes a guess. The guess is compared with the known answer, the difference is measured, and every connection weight is nudged so the next attempt is a little closer to correct. Millions of these tiny updates, applied across billions of weights, gradually turn raw data into a model that can recognise objects, translate languages, or finish text. This same routine often called "back-propagation with gradient descent" is the workhorse behind today's large-language chabot's, medical-image readers, and self-driving-car vision systems (LeCun et al., 2015).

Different layer designs let deep learning tackle many data types. *Convolutional layers* slide a small window across images and are ideal for vision; *recurrent and attention layers* handle sequences in speech and text; *transformer blocks* let every word or pixel compare itself to every other, giving current language models their fluency and long-range coherence. Once training finishes, the same network is switched to inference mode and can answer brand-new questions in milliseconds on phones, GPUs, or specialised AI chips (Goodfellow et al., 2016).

Deep learning became practical when three factors aligned: large public datasets, fast graphics processors for parallel maths, and clever activation functions and normalisation tricks that keep very deep stacks stable while they learn. The result is a technique that now powers face unlocks, real-time speech translation, industrial-quality defect detection, drug-discovery pipelines, and creative tools that draft images or code. Although deep models can be data-hungry and hard to interpret, their ability to turn raw experience into flexible, high-accuracy predictions have made deep learning the dominant engine of modern artificial intelligence (LeCun et al., 2015; Goodfellow et al., 2016).

Neural Networks

Neural networks are computer models built from many tiny units called artificial neurons that roughly mimic the way real brain cells pass signals. Each neuron collects numbers from the layer before it, multiplies them by its own "weight" values, adds a small bias, and runs the result through a simple gate that decides how strongly to fire. When thousands of these neurons are arranged in successive layers, the network can turn raw data at the input into ever-richer patterns inside edges become shapes, shapes become objects, letters become words, and words become ideas. The concept goes back to a 1943 paper that proved even very simple neuron-like elements, if wired together cleverly, can implement any logical rule, suggesting that reasoning itself could grow from networks of basic on/off units (McCulloch & Pitts, 1943).

Each artificial neuron takes a group of numbers, weights them, adds a bias, applies a simple non-linear gate and forwards the result. When many such neurons are arranged in layers, the network can learn a hierarchy of concepts edges, shapes, objects, phonemes, words, sentences directly from raw data. Because the computer builds these concepts from experience, engineers no longer have to list every feature or rule the program should use.

A basic network with only a few hidden layers can recognise simple patterns, but stacking dozens or hundreds of layers turns it into a deep neural network, the core engine of modern deep learning. Researchers have developed specialised layer designs for different data types: convolutional filters slide across an image to capture local texture; recurrent and attention layers keep track of word order in language; auto encoders compress data into dense codes; and long-short-term memory units store information over time. A recent review groups these architectures into five main classes deep feed-forward networks, convolutional networks, deep belief nets, auto encoders and long-short-term memory networks arguing that each solves a specific bottleneck in perception or sequence modelling (Emmert-Streib et al., 2020).

Training a neural network is an iterative loop. A batch of examples flows forward through the layers to produce a guess; that guess is compared with the correct answer; and a learning algorithm nudges every weight just enough to shrink the error. Repeating this cycle over millions of examples reshapes the entire network until it can classify images, translate sentences, steer robots or generate novel text. Because the same computation weighted sums followed by simple activations runs on every layer, modern GPUs and specialised AI chips can execute these networks in parallel, making it practical to train billion-parameter models on web-scale data.

Neural networks now underpin speech assistants, medical-image readers, recommendation engines and autonomous vehicles, yet their strengths come with trade-offs: they demand large, labelled data sets, considerable energy and careful regularisation to avoid over-fitting. Ongoing work explores sparse activations, low-precision arithmetic and neuromorphic hardware to cut the cost, while research into interpretability and robustness aims to make their decisions more transparent and reliable. Still, the central insight remains simple: by stacking many small, identical computing units and letting data shape their connections, neural networks turn raw experience into powerful models that drive much of today's artificial intelligence.

Learning happens by trial and adjustment. During training the network sees an example (such as a photo of an animal) makes a guess, and measures how far that guess is from the correct answer. An algorithm then nudges every weight just enough to shrink that error. Repeating this cycle over millions of examples lets the network reshape its internal connections until it recognises the exact thing. Because all layers are tuned together, the system learns its own features instead of relying on hand-written rules, which is why the approach works on everything from images and audio to sensor streams and text.

Once training is finished, the same network runs in inference mode: fresh data flow through the fixed weights, and the model produces an answer in milliseconds. Different layer designs target different tasks convolutional layers for vision, recurrent or attention layers for sequences, and transformer blocks that let every element compare itself with every other, yet they all rely on the same foundation of weighted sums and activation gates.

Machine Learning

Machine learning is the part of artificial intelligence that lets computers **learn useful rules directly from data instead of following rules we programmed by hand**. A learning program is given examples such as pictures, texts, sounds, tables of numbers together with feedback or goals, and it adjusts its own internal settings, so its answers gradually improve. Because the adjustment follows a mathematical recipe that reduces error step by step, the computer ends up with a model that can spot patterns, predict outcomes, or even create new content without further human guidance.

Most modern projects treat machine learning as a step-by-step process. First comes *data preparation*: collect raw records, clean obvious mistakes, balance classes, and split everything into training, validation and test portions. Next is *representation*: traditional projects build features by hand (ratios, counts, domain signals), while newer ones feed raw data into a neural network, so the network learns its own internal "embedding". Then comes *training*: an optimisation routine, usually gradient descent, repeatedly compares the model's guess with the target, measures the error, and nudges millions of weights just enough to shrink that error. After many passes, the model is frozen and switched to *inference* mode so it can answer fresh queries in milliseconds on CPUs, GPUs or edge chips. Finally, engineers run *evaluation and governance*: they measure accuracy on held-out data, check fairness across groups, study robustness to small input changes, and monitor real-world drift so the system stays reliable. This end-to-end loop – data → representation → optimisation → inference → oversight—has become the standard across industry and research (Thomas et al., 2022).

A wide toolbox of models supports these tasks. Linear and logistic regression give fast baselines; decision trees capture non-linear rules; support-vector machines separate classes with maximum margin; gradient-boosted ensembles win tabular-data competitions; and neural networks ranging from small multilayer perceptrons to giant transformers learn end-to-end on raw images, speech and text.

Machine learning now powers everyday applications, recommendation engines predict what you'll watch next, medical scanners flag suspicious lesions, and weather models refine short-term forecasts. The same techniques drive scientific discovery screening drug molecules, mapping dark-matter simulations and designing new materials by finding regularities that humans might miss.

Models can over fit noisy data, inherit social bias, leak private information, or demand large amounts of energy. Mitigations include bigger and cleaner data sets, regularisation, cross-validation, fairness metrics, differential privacy, model compression and specialised low-power hardware. Despite these hurdles, the basic promise holds; let data teach the machine**,** then keep the process transparent and well-governed so its predictions stay accurate, fair and safe.

Summary of AI

In practice, AI research has converged on a layered toolkit whose pieces nest within one another. The outermost ring is AI itself, a discipline dating to the 1950s that encompasses logic based theorem provers, planning systems that search combinatorial spaces, game tree engines, constraint solvers, and contemporary data driven methods. Within that ring sits machine learning, the family of techniques that enable computers to extract rules directly from experience rather than rely on rules written by hand. ML systems are trained on examples such as images, sounds, sensor logs, and financial tables, and they adjust their internal parameters so that the outputs they produce match the feedback they receive. When the learned rules generalise to new, unseen situations, the system is considered to have learned.

At the next layer lie neural networks. A neural network is a web of very simple processing units, artificial neurons, each of which takes a small set of numbers, multiplies them by trainable weights, adds a bias, and passes the result through a small nonlinear function. A single neuron is almost useless; thousands arranged in layers can approximate highly complex functions. Because a network's behaviour is fully determined by its weights, the learning problem is to find a set of weights that makes the network perform a specified task. That is the optimisation that machine learning algorithms carry out. Each weight update propagates through the network, so neural networks are best understood as a technique within machine learning rather than a rival to it.

Inside the circle lies deep learning. Deep learning refers to neural networks with many layers that learn hierarchical representations. Each additional layer lets the system form a more abstract view of the data: in vision, early layers respond to edges, higher layers to textures, higher still to parts and objects; in language, learned features can move from characters to words, phrases, sentences, and discourse. These representations are learned end to end from data. Engineers do not write the edge detector or the grammar rules; the network discovers useful features because doing so reduces the training loss and improves performance on held out data. GPUs, large data sets, and key algorithmic advances made this practical in the 2010s.

Interactions of Layers

1. **AI ↔ ML** - AI is the broader field that defines tasks and goals; machine learning is the subfield that learns behaviours from data when explicit programming is infeasible or brittle. Traditional AI methods still handle search and combinatorial reasoning, but when a sub-problem can be framed as learning a mapping from data, an ML module is used. The two approaches are complementary. For example, an autonomous robot may use a symbolic planner to choose a route and an ML vision model to detect obstacles.
2. **ML ↔ Neural Networks** - Neural networks use machine learning's optimisation loop but add powerful representation learning. For tabular data or domains with well-engineered features such as credit scoring or supply chain demand, linear models or tree ensembles often suffice. For raw audio, pixels, or text, neural networks excel because they learn their own features. The training loop consists of a forward pass, a loss, and gradients computed by backpropagation and used by stochastic optimisers; these come from optimisation and statistical learning. The network topology, including convolutions, attention blocks, and residual connections, is neural network engineering.
3. **Neural Networks ↔ Deep Learning** - Deep learning scales the same neuron based approach with more layers, larger data sets, and greater compute. Additional depth enables hierarchical abstractions. Residual connections and normalisation layers support stable gradient flow so very deep networks can learn, and attention mechanisms enable long range

dependencies and flexible context mixing. In turn, deep learning has delivered state of the art performance across vision, speech, language, and multimodal tasks, far beyond what shallow networks achieved (Paaß & Hecker, 2024).

Artificial intelligence can be understood as the mission, machine learning as the learning strategy, neural networks as a versatile model family within that strategy, and deep learning as the current high-capacity version of that family. All four layers follow the same principle: use data to turn simple computing parts into systems that carry out complex tasks associated with human intelligence. The outer layers set the goals, the inner layers provide stronger solutions, and progress in any layer supports progress in the others, creating a virtuous cycle in modern AI.

Principles of AI

Artificial intelligence turns experience stored in data into intelligent behaviour by running each problem through a recurring loop. It begins with data: numbers from sensors, lines of text, pixels, or audio clips are gathered and cleaned, and where needed example outputs are attached. The records then pass through a representation step that converts raw signals into internal features the computer can use. Earlier systems relied on hand crafted features, while modern systems let stacked layers of artificial neurons learn them. In the learning phase, an optimisation routine compares the model's guesses with known answers and adjusts millions of connection weights to reduce the gap. Over many passes the model learns to recognise voices, forecast prices, or complete sentences. After training, the same network runs in inference mode and applies its fixed weights to new inputs to produce a prediction, a class label, or a generated paragraph, often in milliseconds on specialised hardware. Around this core sit evaluation and governance: separate test data, checks for bias and robustness, privacy protections, and continuous monitoring help the model stay accurate and fair as conditions change. Most industry and research guides present these stages, namely data handling, representation learning, optimisation, inference, and post deployment oversight, as the standard life cycle for practical AI systems. This shows that across applications the engine under the hood is the same data driven loop (Awan, 2022).

Machine-learning methods, the engine inside most AI systems today, fall into three main families whose differences come from the kind of feedback the program receives while it learns.

- **Supervised learning** - every training example arrives with a correct label, so the model steadily adjusts its internal weights until the outputs it produces match the labels. Because the mapping from input to answer is explicit, supervised models excel at point-by-point tasks flagging fraudulent credit-card transactions, grading tumours as benign or malignant, predicting house prices, filtering spam where accuracy can be checked immediately. Typical algorithms include decision trees, support-vector machines and deep neural networks.

- **Unsupervised learning** - the algorithm sees only raw, unlabelled data and tries to organise it by similarity, compress it into a few dimensions, or spot items that do not fit. Clustering shoppers by spending patterns, grouping news stories by topic or detecting network intrusions all rely on this "find hidden structure" strategy; k-means, hierarchical clustering and principal-component analysis are common tools.
- **Reinforcement learning (RL)** is a way of teaching an agent to act in an environment by trial and error. The agent observes the current state, chooses an action, and then receives a reward that tells it how good or bad that action was. Over many time-steps the agent builds up a table or in large problems, a function linking state action pairs to the long-term rewards they tend to bring. Because the feedback may arrive only after several moves, the agent must learn to balance short-term gains against longer-term pay-offs, sometimes accepting an immediate penalty if it leads to a bigger future reward. The typical RL loop is: (1) sense the state, (2) pick an action using a decision rule, (3) execute the action in the environment, (4) receive and record the resulting reward, and (5) update the decision rule so the same state is handled better next time. This framework is well matched to tasks that can be simulated, where an explicit formula for the best action is unknown, but the agent can still gather experience by interacting with the model, for example, game playing, robot control or resource scheduling (Weber, 2023).

Together these three learning styles let developers pick the right tool for the data they have and the feedback they can obtain: labelled data calls for supervised learning, structure-seeking tasks depend on unsupervised methods, and problems that unfold over time like robot navigation, trading strategies, video-game AI fit naturally into reinforcement learning.

Artificial Intelligence of Today

The last decade's deep-learning revolution replaced past features with *end-to-end* representation learning. Convolutional layers excel at local spatial patterns, while attention mechanisms capture long-range dependencies. The watershed came in 2017 when Vaswani et al. introduced the *Transformer*, showing that self-attention alone could outperform recurrent networks in machine translation and crucially scale almost linearly on parallel hardware (Vaswani et al., 2017).

Once researchers observed that performance improves predictably with more data, parameters, and compute, scaling laws became a guiding tool for the field. Kaplan et al. reported power law relationships across several orders of magnitude, allowing teams to estimate accuracy and cost before training. That plot became a planning aid: scaling data, model size, or compute by known factors yields a roughly predictable change in loss. This insight encouraged larger training runs, specialised AI accelerators, and competition among cloud providers to build large data centre clusters. It also helped drive the rise of foundation models. GPT-3 reached 175 billion parameters in 2020, PaLM 540 billion in 2022, and GPT-4 in 2023 added multimodal capabilities with strong performance on professional exams, signalling the era of general purpose large language models (Kaplan et al., 2020).

Today's frontier systems can translate dozens of languages, generate photorealistic images, automate code synthesis, draft legal clauses and assist physicians in triaging radiology scans. They do so by learning joint embedding that align language, vision and, increasingly, action spaces, allowing a single architecture to summarise documents, caption images or plan robot trajectories with minimal task-specific re-engineering.

A 2024 study looked at the environmental impact of modern large-language models and found that recent progress in AI brings real costs for the planet. Training and running even a typical model such as Llama-3-70B uses a large amount of electricity, produces carbon emissions, and consumes fresh water for data-centre cooling. The researchers compared these figures with the energy and water a person would use to do the same work. They discovered that an AI model can be far more efficient than human labour, by factors of 40 to 150 for a mid-sized system and up to 4 400 for a very small model such as Gemma-2B-it, but the absolute resource use is still high. Once a model is deployed at scale, the energy needed to answer billions of daily queries can approach the yearly power demand of a medium-sized city. The authors conclude that AI developers should publish clear carbon and water numbers for every new system and focus on smaller or more efficient models whenever possible, so the benefits of rapid AI progress do not come with an unchecked environmental burden (Ren et al., 2024).

Looking ahead, the field is converging on hybrid neuro-symbolic architectures that couple the pattern-recognition strength of deep networks with the transparency and compositionality of symbolic logic; advances in continual learning aim to keep models up-to-date without catastrophic forgetting; and policy frameworks now require explainability, fairness and energy disclosure as first-class metrics. In short, AI has grown from an academic curiosity into a multi-trillion-dollar, society-shaping technology stack one whose inner workings marry mathematics, data and silicon in service of ever more capable (and accountable) intelligent systems.

The same machinery that recognises patterns can create them. In a generative adversarial network (GAN), one network invents images while a rival tries to spot the fakes; their competition sharpens the generator until its outputs fool both the discriminator and human observers. Autoregressive Transformers (e.g., GPT-series models) learn the statistical regularities of text so well that they can extend any prompt one token at a time with plausible continuations. Diffusion models begin with pure noise and run it through hundreds of tiny "denoise" networks, gradually revealing a coherent photograph, 3-D shape or protein backbone.

Large-scale experiments show a power-law relationship between model size, training data, compute budget and error rate: double any of the three and performance improves in a predictable, smooth curve. Armed with these scaling laws, research groups now build language models containing hundreds of billions of neurons trained on web-scale corpora. Such systems translate dozens of languages, recommend code fixes, draft drug candidates and summarise legal briefs, all with one shared set of weights.

Training GPT-class models strains global data-centre capacity, so efficiency innovations are racing the appetite for larger nets: sparsely activated layers where most neurons sleep through each inference; 8-bit and 4-bit arithmetic that halves memory traffic; and neuromorphic chips that communicate via spike-like events instead of clocked matrix math. These hardware advances aim to shrink the carbon and water footprint of AI while leaving its creative and analytical reach unhindered. The field is converging on hybrid neuro-symbolic architectures that graft logical reasoning onto deep perception, enabling systems that both see patterns and explain them. Meanwhile, continual-learning research seeks networks that grow their knowledge without forgetting past lessons, much as humans do. Whatever shapes these innovations take, the indivisible building block remains the artificial neuron, a tiny numerical gadget whose collective behaviour, when arranged at unprecedented scale, yields the astonishing breadth of modern AI (Goodfellow, 2016).

Artificial intelligence as we encounter it today is no longer hidden in research labs, it is the recommendation list on a streaming-app home page, the autocomplete that finishes e-mail, the voice assistant that schedules meetings, and the generative model that offers a first draft of a marketing campaign. These systems are powered by *foundation models*, very large neural networks trained on web-scale data that can be adapted to many tasks with only small additional tuning. In practical terms, modern AI is the capacity to learn rich statistical regularities from trillions of words, images, signals and lines of code and then reuse that internal "world model" to classify, predict, reason, converse and create (Batz et al., 2025).

A foundation model is trained in two phases. Pre-training: the network ingests massive unlabelled corpora (text, images, code, videos) and learns to predict masked words, next pixels, or future audio frames, objectives that force it to compress syntax, semantics and physical regularities into its weight matrices. Adaptation**:** the frozen model is lightly steered via instruction-tuning, reward reinforcement with human feedback, or retrieval-augmented generation to behave safely and solve domain-specific tasks. At inference time, the same self-attention mechanism lets every new token or image patch consult *all* prior context, enabling coherent long-range reasoning, style consistency and on-the-fly tool use (for example, calling a calculator or a database inside the conversation loop).

Capabilities and Limitations of AI

Capabilities include writing multi-lingual copy, drafting functional code, designing molecules, interpreting medical scans, answering questions against large document repositories, composing images, music and video, and guiding robots through natural-language instructions. Limitations persist: hallucinated facts, bias inherited from training data, vulnerability to prompt injection, high energy use and the opacity of internal representations. Research now focuses on alignment (making models follow human intent), evaluation (measuring factuality and robustness), and efficiency (sparse activations, low-precision arithmetic, neuromorphic hardware).

Future of AI

The next wave is likely to blend symbolic reasoning with neural pattern recognition, turning today's fluent but fallible chatbots into systems that can not only draft an argument but also check it against logic and trusted knowledge bases. At the same time, regulation such as the EU AI Act and corporate carbon targets are pushing developers to report compute footprints and prove safety before deployment. In other words, the spectacular capabilities of Transformer-era AI are now meeting their adulthood tests: accountability, sustainability and trust. Whether delivered through trillion-parameter clouds or low-power on-device models, the artificial intelligence we see today is both a culmination of sixty-nine years of research and the starting line for entirely new forms of digital cognition.

ChatGPT

Today's most visible face of artificial intelligence is *ChatGPT* a conversational interface layered on top of OpenAI's GPT family of large-language models (LLMs). In mid-2025 the service is processing more than 2.5 billion user prompts every day, a growth curve that has made it the fastest-adopted consumer application since the mobile phone and is beginning to compete with web search as a primary gateway to information (Roth, 2025).

LLM

An LLM is a neural network with billions or now trillions of adjustable weights, arranged in the *Transformer* architecture introduced in 2017. Inputs are broken into discrete tokens (sub-word fragments), each token is turned into a vector, and self-attention layers let every position weigh the relevance of every other position, capturing syntax, semantics and long-range dependencies in a single pass. Because the model is trained to predict the next token across a web-scale corpus, its weights become a compressed statistical map of human language and the world knowledge implicit in that language. The power of this template was quantified by Kaplan et al., who showed that model error falls along a smooth power-law as you scale parameters, data and compute an empirical "scaling law" that still guides model design (Kaplan et al., 2020).

ChatGPT's underlying LLM is built in three stages:

1. **Pre-training** - GPT ingests trillions of tokens of text (and images in its multimodal variant) and learns to predict masked or next tokens, developing a general linguistic and conceptual prior.
2. **Instruction fine-tuning** - Developers curate thousands of task-oriented prompts and demonstration answers and continue training the network so it follows explicit instructions.
3. **Reinforcement learning from human feedback (RLHF)** - Human judges rank multiple model outputs; those rankings train a reward model, and the LLM is further refined with policy-gradient methods to optimise for helpfulness, harmlessness and honesty, an approach first formalised in the InstructGPT study (Ouyang et al., 2022).

The result is a single trained model that can converse, translate, write code, reason, and generate creative text without task specific retraining. OpenAI reports that the strongest ChatGPT variants achieve human level performance on many professional and academic benchmarks, suggesting that pre-training and alignment approaches can push general language competence toward near expert levels.

Chat Turn Processing

When a user submits a message, the application assembles the full conversation context, including system instructions, prior exchanges, and the new message, and encodes it as tokens. This request is sent to servers running specialised accelerators (GPUs or TPUs). At each step, the model consumes the current token sequence and returns a probability distribution over the next token. The server samples from this distribution (for example with temperature or nucleus sampling), appends the chosen token, and repeats until an end of text marker appears or a token limit is reached. The reply is therefore generated token by token, conditioned on all prior context.

Capabilities of LLM's such as ChatGPT

- **Productivity** - Generate first-draft emails, marketing copy, legal clauses and press releases in seconds.
- **Software engineering** - Explain legacy code, suggest bug fixes, and scaffold entire applications from plain-language specs.
- **Data & research** - Summarise long PDFs, huge amount of data, translate across 100+ languages, and act as a conversational front end to databases and search APIs.
- **Creative work** - Brainstorm ad campaigns, compose poetry, design game levels, or pair with image models to script storyboards.
- **Education** - Offer personalised tutoring and Socratic dialogue, automatically adjusting explanations to a student's knowledge state.

Beyond ChatGPT, the market now offers a family of large-language systems that serve similar conversational roles but differ in how they handle data, privacy, speed and multimodal input.

- **Google Gemini** comes in several sizes, Ultra for cloud workloads, Pro for everyday use and Nano for phones and wearables. The entire line is natively multimodal: the same network processes text, images, audio and video, and the newest 2.5 release can accept prompts as large as one million tokens, letting it keep whole code-bases, textbooks or hours of video in memory during a single chat.

- **Anthropic Claude 3** focuses on safe, long-context reasoning. Its Opus tier reads book-length prompts and applies a refusal policy that filters disallowed content more strictly than ChatGPT. Developers like it for refactoring large repositories because Claude can keep thousands of source files "in mind" while proposing edits.
- **Meta Llama 3** and **Mistral Large** are open-weight models. Organisations can download the checkpoints, fine-tune them on private documents and run them in their own data-centres, avoiding vendor lock-in and cutting inference costs at the price of slightly weaker reasoning scores compared with the biggest proprietary models.
- **Special-purpose copilots** now ship with design suites, office software and programming IDEs; most wrap a smaller, domain-tuned LLM around retrieval tools so the assistant can quote company documents verbatim while generating answers.
- **Edge-oriented versions** such as Gemini Nano or Llama 3-8 B run entirely on modern smartphones, enabling offline translation, image captioning and voice control without sending sensitive data to the cloud.

ChatGPT (GPT-4/4o) still holds a lead in general reasoning and creative writing, thanks to its large training corpus, refined alignment pipeline and strong tool ecosystem. However, ChatGPT is no longer a solitary breakthrough but one flagship among dozens: cloud-scale rivals (Gemini, Claude), permissive open-source stacks (Llama, Mistral), and lightweight on-device engines all occupy distinct slices of the landscape. Choosing between them now depends less on headline benchmarks and more on practical trade-offs such as cost, privacy, context length, multimodal reach and the freedom to customise making today's LLM ecosystem diverse, rapidly evolving and highly competitive.

Limits, Costs and Safety

Even the most advanced language models share a common set of weaknesses. They can "hallucinate" produce fluent statements that are not true; they inherit social and cultural biases that were present, often invisibly, in the web-scale data used for pre-training; and they are susceptible to prompt injection, jail-break techniques that trick a model into revealing disallowed content or instructions. Because the same weights drive every downstream application, one silent flaw can propagate across many products before it is detected. These technical limits sit alongside environmental and monetary costs: training a model requires roughly huge amount of MWh of electricity and hundreds of thousands of litres of cooling water, while today's global inference load billions of requests per day already draws power comparable to that of a medium-sized city and is growing at double-digit rates each quarter.

To keep performance gains from being overshadowed by these risks, developers and hardware designers layer defences into a safety stack that spans data curation, model architecture, post-training alignment and runtime controls:

- **Pre-training hygiene** - toxin-filtering pipelines remove obvious hate or misinformation before it ever reaches the optimiser, and dataset audits aim for balanced demographic representation.
- **Alignment tuning** - techniques such as instruction fine-tuning, constitutional reinforcement and human-feedback reward models teach the network to reject unsafe prompts or flag uncertainty rather than fabricate confident answers.
- **Red-team evaluation** - specialised teams and external researchers probe for jailbreaks, bias, privacy or data leakage, and prompt injection. Their adversarial prompts and findings are distilled into automated evaluation suites that run regularly as the model evolves.
- **Policy filters and guardrails** - lightweight classifiers or rule engines sit in front of the core model, screening inputs and outputs in real time for disallowed content or privacy-sensitive data.
- **Monitoring and governance** - after deployment, telemetry tracks accuracy drift, refusal patterns and energy usage; many organisations map these checks to technology-readiness frameworks so a single dashboard shows whether the service remains compliant with corporate, regulatory and environmental targets.

Large-language models now draft legal clauses, write code and tutor students, yet three systemic limits still define the frontier. Truthfulness remains fragile because the networks learn from raw web text that mixes facts with rumours; when a prompt asks for information outside the statistical patterns the model has seen, it may fill the gap with confident sounding fabrications, a phenomenon called hallucination. Fairness and security are the next hurdles: biases embedded in the training corpus can surface as skewed advice, and hostile prompt-injection attacks can trick the model into revealing disallowed content or ignoring safety instructions. Resource demand is the third constraint. Training a single model required well over a gigawatt-hour of electricity and hundreds of thousands of litres of cooling water, while today's daily inference load billions of requests served from GPU clusters already draws power comparable to that of a medium-size city and is climbing rapidly each quarter.

Meanwhile, the hardware community tackles the resource side of the equation. Sparsely activated transformers keep most neurons idle for any single token, 4-bit quantisation and model distillation cut memory traffic and power by factors of two to four, and prototype neuromorphic accelerators promise order-of-magnitude efficiency gains by letting activations flow only when signals change. These measures, combined with renewable-powered data-centres, aim to decouple future gains in fluency from runaway energy use.

The modern LLM ecosystem advances on two intertwined fronts, capability and responsibility. Each leap in reasoning depth or context length is matched by tighter guardrails, more aggressive red-teaming and more efficient silicon, reflecting a growing consensus that useful intelligence must be delivered with verifiable truthfulness, demonstrable fairness and a shrinking environmental footprint.

The next decade of artificial intelligence is expected to pivot from single-modality, “frozen” models toward multimodal, retrieval-augmented, tool-using systems that can see, hear, plan and cite. Early surveys of Multimodal Retrieval-Augmented Generation (MRAG) point out that future assistants will pair large neural encoders for text, images and audio with fast retrieval engines that pull fresh facts from live databases, giving each reply an up-to-date evidential trail instead of relying only on what was memorised at training time. By fusing these retrieved facts with generative reasoning, the same model can explain a chart, caption a photograph, or draft a report that links directly to source documents. Tool APIs extend this idea further: the model no longer just describes a solution but calls a calculator, triggers a web query or issues a robot command, chaining multi-step actions on the user’s behalf. Parallel research in hybrid neuro-symbolic design adds formal logic modules on top of deep networks so the system can not only state an answer but prove it, while energy-efficient hardware and alignment-by-design aim to keep the ecological and social footprint proportional to the benefits. If these threads converge, tomorrow’s AI will act less like a static encyclopaedia and more like a grounded, multi-sensor co-worker able to retrieve evidence, reason across modalities and execute plans, all while showing its work (Mei et al., 2025).

Conclusion

Artificial intelligence is a broad, multidisciplinary field devoted to building computer systems that can sense, reason, learn, and act in ways that traditionally required human intelligence. Over the past decade AI has shifted from research labs into everyday life: vision models now flag suspicious lesions in medical scans, language assistants translate dozens of languages in real time, and planning algorithms continuously adjust power grids and supply chains to reduce waste. These applications show how AI can be a powerful tool saving lives through faster diagnoses, boosting productivity by automating routine tasks, and uncovering patterns that help scientists design new materials or predict weather extremes. Yet rapid progress also brings unresolved questions. Large models can generate misleading information, inherit social biases from their training data, or be manipulated by cleverly crafted prompts. Energy-hungry data centres raise environmental concerns, and autonomous systems introduce safety and accountability issues when decisions go wrong. Addressing these challenges demands a multidisciplinary approach: computer scientists to refine algorithms, ethicists and lawyers to shape fair policies, psychologists and domain experts to guide human-AI interaction, and engineers to design hardware that runs efficiently. In short, AI is both an exceptionally effective set of tools and an evolving set of open problems; harnessing its advantages while managing its risks will require continued collaboration across many fields.

Tasks

1. How can artificial intelligence be described simply?
2. Name at least three specific LLMs.
3. What is deep learning?

4. PREDICTIVE ANALYTICS VS. DEMAND FORECASTING

Predictive analytics is a data-driven discipline that converts large collections of historical and live information into numerical estimates of future events, such as the likelihood that a sensor will fail, a customer will churn, or sales will rise next quarter. A typical workflow begins with a clearly framed business question, followed by data acquisition from internal systems and external feeds; engineers then cleanse the records, reconcile time stamps, handle missing or noisy values, and build features that capture seasonality, interaction effects, and domain alerts. Modellers choose algorithms that match the structure and size of the data set: linear or logistic regression for fully tabular problems where transparency is vital, tree ensembles and gradient-boosted machines for non-linear tabular data, and recurrent, convolutional, or transformer networks when inputs include sequences, images, or language. Each candidate model is evaluated with time-aware cross-validation and hold-out tests to verify generalization; only when stability and business accuracy thresholds are met is the model wrapped in a container or function and exposed to production systems through an API. Mature teams add monitoring dashboards that watch prediction error, data drift, and feature distributions in real time, scheduling automated retraining when thresholds are breached. Applications span many sectors: manufacturers lower unplanned downtime by predicting bearing failures, banks limit fraud by flagging anomalous transactions within milliseconds, retailers refine inventory by daily demand scores at the SKU-store level, and hospitals deploy early-warning systems that alert clinicians to sepsis risk hours before symptoms escalate. Empirical reviews indicate that organisations that deploy predictive analytics at scale often report meaningful cost reductions and accuracy improvements compared with rule-based planning. The magnitude varies by context, but gains are commonly in the single to low double digit range when data quality is strong and teams include domain specialists, data engineers, and MLOps support. Persistent challenges include biased or incomplete labels, concept drift due to market shifts, and the need for explainable and accountable systems under regulations such as the EU AI Act.

Mitigation strategies involve balanced sampling, fairness audits, and techniques like SHAP that quantify feature influence for individual predictions. Research trends now aim at automated feature engineering and hyper-parameter tuning through AutoML, causal layers that move beyond correlation, privacy-preserving federated learning on edge devices, and the use of large pre-trained language or vision models as universal representation generators that speed up downstream training. Together these advances underline one fact, predictive analytics transforms passive data storage into proactive intelligence that guides everyday decisions, but lasting value depends on disciplined data governance, continuous validation, and clear alignment with business outcomes (Jamarani et al., 2024; Azmi et al., 2024).

Below is a concise catalogue of the model families used in predictive analysis based on a study from Jamarani et al., 2024:

- **Classical statistics**
  - Linear and logistic regression
  - Generalised-linear models (Poisson, negative-binomial, etc.)
  - Autoregressive-Moving-Average variants (ARIMA, SARIMA)
  - Exponential-smoothing families (Holt–Winters, ETS)
- **Probability-based learners**
  - Naïve Bayes, Bayesian networks
  - Hidden-Markov models
- **Distance and instance methods**
  - *k*-Nearest Neighbour (KNN) for classification/regression
  - *k*-means, DBSCAN and hierarchical clustering for segmentation
- **Kernel and margin algorithms**
  - Support-Vector Machines (classification and ε-SVR regression)
- **Tree ensembles**
  - Single CART/C4.5 trees for explainability
  - Random Forests (bagging)
  - Gradient-Boosted Machines (XGBoost, LightGBM, CatBoost)
- **Deep-learning architectures**
  - Multilayer perceptrons for numeric tables
  - Convolutional Neural Networks for images and sensor grids
  - Recurrent nets (LSTM, GRU) for single-series sequences

  - Transformer-based models (Informer, Temporal Fusion Transformer) for multi-horizon time series
  - Graph Neural Networks for data with explicit links

- **Hybrid and automated approaches**
  - Signal-decomposition hybrids (e.g., Variational Mode Decomposition + LSTM)
  - Stacked/voting ensembles that blend several of the above
  - AutoML frameworks that search hyper-parameters and model families automatically

Artificial intelligence amplifies predictive analytics by automating every stage between raw data and forward-looking insight. Modern AI systems begin by ingesting multi-modal inputs structured transactions, sensor streams, text, images and use deep representation learning to turn these heterogeneous signals into dense numerical vectors that preserve seasonality, context and cross-entity links. Transformer encoders, graph neural networks and convolutional filters replace manual feature engineering, uncovering non-linear dependencies that classic statistical models miss. During training, AutoML pipelines search hyper-parameter and architecture spaces at scale, while self-supervised pre-training on public corpora supplies robust priors that accelerate convergence on small proprietary datasets. In production, model ensembles run inside containerised micro-services that score new events in milliseconds; drift-detection agents monitor feature distributions and trigger automated retraining when data shifts. Edge versions of the same networks execute on gateways or mobile chips so that latency-critical forecasts machine-failure warnings on a factory floor, fraudulent-transaction flags in a point-of-sale terminal remain local even if the cloud link drops. Throughout the life-cycle, explainability layers generate feature-attribution heat maps and counterfactual examples to satisfy auditors and to fine-tune operating thresholds. Industry case studies report double-digit gains from this integration: deep networks that fill large data gaps in retail demand forecasting cut stock-outs and markdowns, and Big-Data-Predictive-Analytics frameworks raise forecast accuracy and decision speed across sectors from smart cities to healthcare. The evidence shows that AI's role is not merely to improve model fit but to make predictive analytics continuously adaptive, scalable and self-maintaining turning a once-static statistical exercise into an autonomous, learning operational capability (Jamarani et al., 2024; Riachy et al., 2025).

Predictive analytics with artificial intelligence operate through an end to end cycle that starts by defining a clear forecast target, for example time to failure for a bearing or next week demand for a product, and ends with automated predictions feeding operational systems. The first requirement is an integrated data layer: high frequency sensor logs, transactional tables, and external signals such as weather, promotions, and macroeconomic indicators are collected, aligned to common timestamps, cleaned for outliers and gaps, and tracked with version control to ensure results can be reproduced. Next is representation learning, also called feature learning. Instead of hand crafted features, modern systems rely on deep models that learn useful patterns directly from raw inputs. For time series in business and industry, transformers with attention and hybrid setups that first break signals into components and then use encoder decoder networks can capture patterns at different time scales and sudden changes in a single run. These approaches can reduce forecast error on volatile series. In asset intensive settings, graph neural networks model interactions among machines, and convolutional networks can read vibration spectrograms to spot early signs of faults.

Candidate models produced by automated model searches are validated with time aware cross validation that keeps the time order. The selected models are packaged behind REST APIs or message queues, allowing low latency scoring. Production success depends on MLOps support: feature stores to keep training and inference aligned, monitors that track drift and data quality, and automatic retraining when performance drops past a set threshold. Governance is equally important and includes role based access, audit logs, and explainability dashboards to support trust and compliance.

Case studies report that AI centric predictive maintenance can extend remaining useful life estimates and reduce unplanned stoppages, and that deep learning demand forecasting can lower stock outs and markdowns through accuracy improvements. The size of the gains varies by context and is often in the single to low double digit range, provided data volume, labelling quality, and cross functional expertise are in place.

Predictive Analytics in Logistics

Predictive analytics has become a central decision engine in modern logistics, converting streams of operational data into forward-looking insight that guides everything from strategic network design to real-time dispatch. Logistics organisations begin by consolidating heterogeneous inputs such as historical order lines, point-of-sale signals, fleet-telemetry feeds, weather and event calendars, traffic APIs and even image data from warehouse cameras into a governed lake where time stamps are synchronised and quality rules remove duplicates, fill gaps and standardise units. Modelling teams then fit a layered stack of algorithms that each solve a specific forecasting problem along the flow of goods. At the planning horizon of weeks to months, deep-learning demand forecasters built on encoder-decoder transformers augmented with Variational Mode Decomposition learn long-range seasonality and holiday shocks directly from raw series and exogenous drivers. A recent open-access study reports that the approach cuts mean-absolute-percentage error by more than ten percentage points for consumer packaged-goods SKUs compared with classical baselines. Downstream inventory optimisation engines translate these probabilistic forecasts into replenishment signals, increasingly through end-to-end neural policies or hybrid stochastic–reinforcement schemes that bypass the older "predict-then-optimise" split; although results vary by product volatility, case studies in multi-period replenishment show average service-level improvements above five points while reducing capital tied up in stock (Lei et al., 2025).

Closer to execution, predictive analytics reshapes last-mile and middle-mile routing. AI-powered optimiser platforms blend supervised demand-density prediction with reinforcement-learning search or evolutionary heuristics to produce dynamically updated tours that respect driver preferences, curb-side constraints and micro-weather anomalies; empirical work on urban e-commerce fleets demonstrates fuel savings of 8–12 percent and delivery-time reductions of roughly the same order when such models replace static distance-minimising heuristics. The same data-driven mind-set underpins emerging "route-realisation" models that learn how experienced drivers sequence stops; pointer-network architectures trained on millions of GPS traces can now replicate human choices within a few parcel drops, making auto-generated manifests more realistic and reducing manual re-sequencing in handheld apps (Paul et al., 2025; Pegado-Bardayo et al., 2024).

Fleet reliability is another beneficiary. Internet-of-Things sensors embedded in tractors, trailers and material-handling equipment stream vibration patterns, fluid diagnostics and ambient measurements in real time; gradient-boosted trees and temporal-convolutional networks convert these signals into remaining-useful-life estimates so that workshops can be scheduled before costly roadside failures occur. A 2024 study covering a mixed truck fleet showed that sensor-driven predictive maintenance lengthened service intervals by roughly seven per cent while cutting unplanned downtime by over 20 percent. These gains cascade directly into higher on-time-in-full scores and lower spare-vehicle buffers (Kansal & Ediga, 2024).

At the customer-promise frontier, machine-learning systems predict the full distribution, not just the mean, of order-fulfilment time. Conformal predictive distributions have recently been applied to millions of historical parcels, delivering valid coverage guarantees while improving the hit rate of "delivered-by" commitments by up to fourteen per cent and reducing late-delivery surprises by three-quarters in large-scale A/B tests. Such calibrated uncertainty is critical for cost-sensitive decision rules that weigh expedited shipping or split-shipment options against the risk of service failures (Ye et al., 2025).

The commercial impact of integrating these purpose-built models into a single analytic stack is well documented. Retail and third-party-logistics operators finds typical cost reductions between eight and twenty-five percent and accuracy gains of fifteen to thirty-five percent once predictive analytics replaces rule-based planning; most of the variance in outcomes is explained by data maturity, cross-functional collaboration and the presence of automated monitoring pipelines that retrain models when input distributions drift (Adedoyin et al., 2024). Best-in-class implementations therefore embed MLOps practices version-controlled feature stores, automated hyper-parameter sweeps, canary deployments, and live dashboards showing forecast accuracy, bias and latency so that models evolve with market and network change rather than degrade silently.

Technically, the model zoo ranges from lightweight baselines such as regularised regression and ARIMAX (kept for interpretability and as fall-back) to gradient-boosted decision trees that dominate tabular tasks, encoder-decoder transformers and Temporal-Fusion-Transformers that capture hierarchical seasonality, graph neural networks that propagate signals across facility, lane or product interaction graphs, and policy-gradient reinforcement learners that explore high-dimensional routing or allocation spaces. Yet algorithms alone do not guarantee value: success hinges on data richness, clear target definitions, disciplined experiment design and for regulated or mission-critical contexts explainability layers such as SHAP or attention heat maps that translate model output into operational logic understandable by dispatchers, warehouse planners and quality auditors.

In sum, predictive analytics under the AI umbrella is pulling logistics from a reactive, after-the-fact reporting paradigm into a proactive, probability-driven orchestration of assets and promises. By linking granular sensors, transactional events and external context into continually learning models, organisations achieve a tighter match between capacity and demand, faster and greener routes, higher equipment uptime and more reliable customer commitments. The next frontier, already visible in pilot projects, combines foundation models that unify language, imagery and spatial-temporal records with optimisation layers that nudge the entire supply-chain network toward orchestrated objectives such as minimum total emissions or maximum on-time-delivery under cost ceilings. However, advanced the toolkit becomes, the enduring prerequisites remain high-quality data, rigorous governance and cross-functional alignment between data scientists and frontline logistics experts.

Demand Forecasting

Demand forecasting is a disciplined procedure for translating past and present information into quantified expectations about future demand so that managers can make better informed decisions on inventory, capacity, staffing and finance. It begins with a clear statement of the decision context, continues with the careful collection and cleansing of historical demand data, incorporates relevant causal or leading indicators, and applies statistically valid techniques to produce forecasts accompanied by explicit measures of uncertainty. Armstrong emphasises two principles that underpin successful practice: first, match the complexity of the method to the complexity of the data and the cost of error; second, embed a feedback process so forecasts are routinely compared with outcomes and models are refined (Armstrong, 2001).

Operationally, the core of most quantitative demand forecasting lies in time series analysis. Box et al. (2015) detail a systematic workflow that has become standard: (1) visualise and diagnose the series for level, trend, seasonality and anomalous events; (2) difference or transform the data until stationarity is achieved; (3) identify candidate autoregressive and moving average terms with autocorrelation plots; (4) estimate model parameters, usually by maximum likelihood; (5) validate residuals to confirm that no structure remains; and (6) generate forecasts and confidence intervals. This framework remains valuable because it is transparent, statistically rigorous, and when the underlying data generating process is well captured, provides unbiased point forecasts and realistic prediction intervals that managers can translate directly into safety stock or capacity buffers. Box et al. also stress that forecast accuracy should be monitored with proper scoring rules such as mean squared error, and that models must be re-estimated whenever structural change is detected, thus ensuring the forecasting system stays current with market dynamics.

The importance of aligning method with context is underscored by Petropoulos et al. (2014). In their comparative study of dozens of methods across more than 3000 real demand series, they coined the phrase “horses for courses” to capture the finding that no single technique dominates; rather, simple exponential smoothing wins when demand is stable, damped-trend variants work better when growth is slowing, seasonal ARIMA excels for regular seasonality, and machine learning or hybrid ensembles outperform in noisy, high-frequency retail data. Their evidence shows that selecting the right model for the right data can cut forecast error by 10 to 30 percent, which converts directly into lower stock outs, reduced obsolescence and higher return on assets. They also demonstrate that combining several well-chosen models often yields further gains because different methods capture complementary aspects of the demand signal.

In practical terms, therefore, demand forecasting means establishing a repeatable process that (i) collects timely, relevant data, (ii) applies a model whose statistical assumptions are reasonably satisfied, (iii) quantifies forecast uncertainty, (iv) monitors performance in real time, and (v) adapts when business conditions change. When executed according to these evidence-based principles, demand forecasting delivers three strategic benefits. First, it reduces waste by aligning production and procurement with true market needs, thereby freeing capital and warehouse space. Second, it enhances customer service by ensuring products are available when and where needed, which strengthens loyalty and revenue. Third, it supports long-range planning by providing an objective baseline against which alternative scenarios such as new product launches, price changes, promotional campaigns can be stress-tested. Collectively, these advantages make demand forecasting a cornerstone of modern operations and supply-chain management.

Artificial intelligence now drives demand-forecast by replacing manual model-by-model tuning with systems that learn directly from data at catalogue scale, deliver full probability distributions, and reveal which business drivers matter most. A first class of solutions exemplified by DeepAR trains a single recurrent neural network on thousands of related time series and covariates such as price, weather or stock-on-hand; the network's shared parameters let weak or intermittent items borrow strength from strong sellers, while Monte-Carlo sampling from the model's likelihood produces calibrated quantiles that feed directly into service-level inventory rules and cash-flow simulations, yielding roughly fifteen-percent accuracy gains over classical baselines and eliminating the need for hand-crafted feature engineering (Salinas et al., 2019).

Complementing these black-box learners, decomposable additive models such as Prophet combine piece-wise linear or logistic trend segments with Fourier-based seasonality and holiday regressors; analysts can lock or override individual components when they know a promotion or range review is coming, yet the Bayesian engine still returns uncertainty bands and automatic change point detection, so planners enjoy both transparency and the ability to run the same configuration across tens of thousands of stock-keeping units with minimal upkeep (Taylor & Letham, 2017).

Pushing accuracy and insight further, the Temporal Fusion Transformer blends recurrent encoders for short-term dynamics with self-attention layers that learn long-range dependencies and a variable-selection network that assigns time-varying importance weights, delivering state-of-the-art multi-horizon forecasts while exposing which inputs like price, holiday flags, macro indicators drive each horizon; field studies in energy, traffic and omnichannel retail report double-digit percentage improvements over gradient-boosted trees and LSTMs, and the attention heat-maps give supply-chain teams an audit trail for governance reviews and bias checks (Lim et al., 2021).

Together, these AI approaches transform forecasting into a continuous-learning loop: live data streams update feature stores, models retrain overnight with automated back-testing, and exception dashboards highlight items where predictive distributions shift, prompting human investigation or policy intervention. The result is shorter safety-stock buffers, faster reaction to demand shocks, better alignment between merchandising and operations, and, crucially, a unified probabilistic language that lets finance, logistics and marketing optimise jointly against quantified risk instead of haggling over rival point estimates

Demand Forecasting in Logistics

Demand forecasting is the hinge on which logistics efficiency turns: every booking of vessel slots, line-haul trailers or cross-dock capacity presupposes an estimate of how many units will need to move, from which origin to which destination, and when. Logistics teams begin by cleansing historical shipment files, mapping each record to common time buckets (day, week, month) and origin-destination pairs. Baseline volume is then projected with lightweight statistical models moving averages for stable lanes, Holt-Winters for trend-seasonal flows, seasonal ARIMA for more erratic patterns because these methods are transparent, quick to recalculate and easy to explain to operations staff. The resulting lane-level forecasts are reconciled both bottom-up and top-down so that sums across SKUs match network totals and vice versa, ensuring procurement, warehousing and transport planners all work from a coherent picture.

Once the statistical baseline exists, a consensus review pulls in upcoming promotions, contract roll-offs and known disruptions; planners adjust the figures and critically record the overrides so future back-tests can distinguish model error from human judgment. Forecasts are then rolled into three horizons: strategic (one to five years) for fleet sizing and warehouse investment, tactical (one to six months) for seasonal labour and network balancing, and operational (days to weeks) for dock schedules and route assignment. Continuous monitoring closes the loop: real shipment counts stream into control charts, and any lane whose error exceeds tolerance triggers immediate model recalibration or management escalation.

The payoff is substantial. A five-percentage-point gain in forecast accuracy typically cuts total logistics cost by about two percent because better forecasts let shippers negotiate capacity that matches true demand, pre-position inventory in the correct regions, and roster labour precisely to workload reducing overtime, dwell time and emergency freight (Christopher, 2016).

By following this disciplined cycle cleanse, model, reconcile, review, monitor—logistics organisations transform forecasting from a reactive statistical chore into a proactive lever for service reliability and cost leadership.

Conclusion

Predictive analytics is the broad practice of using historical data, statistical algorithms and machine-learning models to estimate the likelihood of any future event or behaviour—late deliveries, equipment failures, fuel spikes, customer churn, and more. Demand forecasting is one specific branch of that practice, focused only on predicting how many units of a product (or shipments of a lane) will be required in a given period; its output is a volume profile rather than a probability of occurrence.

Key differences

- **Scope** - Predictive analytics can target any outcome (delay/no-delay, high/low risk, numeric volume), whereas demand forecasting targets unit counts or order lines.
- **Techniques** - Both may use similar tools (regression, neural networks, time-series models). The difference is in framing: demand forecasting always produces a quantity over time; predictive analytics can return classes, scores or quantities.
- **KPIs** - Demand forecasts are judged by volume accuracy (MAPE, RMSE). Predictive-analytics models may be judged by recall, AUC, or cost savings, depending on the question.
- **Decisions enabled** - Demand forecasts drive capacity booking, stocking and labour planning. Predictive analytics informs wider logistics choices: which shipment is likely to miss cut-off, which trailer is at failure risk, what surcharge to apply, or whether to reroute around a storm.

Using demand forecasting helps a carrier or shipper book the right amount of line-haul capacity, pre-position inventory in the correct region and roster enough pick-pack staff. Even a modest accuracy gain can trim total logistics cost through fewer emergency expedites and lower safety stock.

Using wider predictive analytics extends those savings. Late-arrival models let dispatchers re-sequence stops, failure-risk models schedule preventive trailer maintenance or fuel-price predictions lock hedging contracts at the right time. Together, the two capabilities create an end-to-end “sense-and-respond” network demand forecasting sets the baseline flow, while predictive analytics flags exceptions and optimises resources along the way.

Tasks

1. What is predictive analytics?
2. What is demand forecasting?
3. What is the difference between predictive analytics and demand forecasting?

5. INVENTORY MANAGEMENT AND PRODUCTION PLANNING

This chapter traces how inventory management and production planning evolved from Harris's EOQ and Orlicky's MRP to today's Industry 4.0 ecosystems. It distils the key deterministic, stochastic and multi-echelon theories that govern cost-service trade-offs across global supply networks. We explore how artificial intelligence like demand-sensing, reinforcement learning and digital twins augments classical models to create autonomous, resilient plans.

Overview

Inventory management and production planning (IMPP) are now framed as a cornerstone of supply-chain performance, with systematic reviews classifying them as the central logistics function that links materials, capacity and customer service goals. Current surveys of production-scheduling research reinforce this view, showing that advanced planning and scheduling (APS) systems built on machine-learning and Industry 4.0 data streams outperform stand-alone inventory or scheduling tools in volatile manufacturing environments (Munyaka & Yadavalli, 2022; Mustajab, 2023).

Effective inventory management is central to a firm's competitiveness and financial health, a link that has been demonstrated in industries ranging from retail to healthcare. The reach of inventory policies now extends well beyond profit, influencing environmental impact, social goals, and the resilience of supply networks. Recent studies tackle challenges such as reducing waste in perishables, enabling circular flows through remanufacturing, supporting humanitarian logistics, and mitigating shortages brought on by crises like the COVID-19 pandemic or geopolitical tensions. These pressures keep inventory management high on both research and managerial agendas and underline the need for fresh approaches that can handle growing complexity (de Souza et al., 2025).

Inventory management can be defined as:

"*a supply chain function that involves the systematic coordination and control of procurement, storage, and distribution of goods. Its objective is to optimize overall operational efficiency by ensuring that optimal order quantities are shipped in a timely matter*" (de Souza et al., 2025).

Researchers have introduced numerous ordering rules and inventory policies and have progressively woven related supply-chain considerations such as production planning, distribution choices, uncertainty, and market mechanisms like pricing or auctions into inventory models. Because real-world settings are so intricate, the field largely turns to optimisation frameworks to guide decisions. This focus has produced a broad toolkit that spans linear and integer programming, heuristics and metaheuristics, as well as dynamic programming approaches (de Souza et al., 2025).

Recent bibliometric evidence reveals a shift in research focus from single-item EOQ models to holistic, multi-stage formulations that embed service-level targets, stochastic demand and capacity constraints. Publication counts in inventory research grew by more than 500 % between 1998 and 2020, and contemporary reviews argue that integrated planning generates measurably lower total cost than sequential "inventory-first, scheduling-later" tactics (Munyaka & Yadavalli, 2022; Mustajab, 2023).

Production planning sets the quantities and timing of factory orders so that demand forecasts align with finite machine, labour and material capacities. Recent literature reviews show that advanced planning and scheduling platforms enriched with machine-learning forecasts and Industry 4.0 data streams boost resource utilisation and shorten lead times relative to traditional rule-based methods, making plants more responsive to variability. Conceptual studies now couple demand-driven MRP logic with digital-twin models, giving planners a real-time sandbox where alternative schedules can be stress-tested and automatically updated as shop conditions evolve. These converging strands confirm production planning's role as the pivotal layer that converts strategic sales-and-operations goals into executable, resilience-oriented shop schedules (Ma et al., 2024; Mustajab, 2023).

Production planning can be defined as:

"*a sequential ladder in the manufacturing setting to ensure that materials input (raw materials, men, money, and machine) is made available within a stipulated time frame, in the appropriate amount to produce the demanded output of goods and services based on the schedule specified*" (Sunday et al., 2021).

Recent research on AI-driven planning emphasizes real-time data flows, digital-twin simulations, and reinforcement-learning algorithms that continuously adjust reorder points and capacity-constrained schedules. Studies report double-digit gains in lead times and inventory turns over traditional rule-based methods, signalling a wider fusion of analytics, automation, and resilience goals in inventory and production management (Mustajab, 2023).

Historical Evolution

Early analytic foundations (1913 – 1960s)

The first quantitative treatments of inventory focused on single-item lot sizing: Harris's EOQ equation balanced setup and holding costs, while Wagner and Whitin later generalised the problem to time-varying demand, proving that a dynamic-programming solution yields the minimum-cost production schedule. Modern retrospectives mark these models as the starting point for formal production-inventory optimisation and still cite them when benchmarking new heuristics (Andriolo et al., 2014; Vargas, 2009).

MRP and the mainframe era (1970s)

With corporate computing power available, Orlicky packaged the idea of exploding bills of material back from a finished-goods master schedule to component orders, creating material-requirements planning (MRP). Contemporary reviews of ERP history trace nearly every modern planning module to this breakthrough, noting that MRP II later closed the loop to shop-floor capacity and cost accounting (Al-Amin et al., 2023).

Lean/JIT revolution (1980s)

Rejecting large batches, Toyota's Production System showed that small, pull-based lots and visual kanban cards could slash lead time and expose hidden process problems. Recent overviews of lean manufacturing emphasise that JIT redefined inventory from a "necessary asset" to a visible symptom of waste, profoundly influencing both Western continuous-improvement programmes and today's agile production cells (Dave, 2020; Anoop et al., 2020).

ERP, APS and global outsourcing (1990s – 2000s)

As supply chains globalised, firms adopted enterprise-resource-planning suites to unify purchasing, production and finance data. State-of-the-art reviews describe how advanced-planning-and-scheduling (APS) engines—often embedded in ERP—use mixed-integer optimisation to create capacity-feasible plans, a prerequisite for offshore manufacturing with long, risky lead times (Al-Amin et al., 2023; Stüve et al., 2025).

Industry 4.0 and the digital-twin age (2010 – present)

Recent studies position digital twins, IoT sensors and reinforcement-learning algorithms as the new frontier: virtual replicas of plants and networks stream real-time data into cloud solvers that adjust lot sizes, safety stock and machine sequences on the fly. Reviews published in 2024 outline frameworks where these twins integrate lean principles, resilience metrics and sustainability targets, marking a shift from static planning to continuously self-optimising systems (Guo & Mantravadi, 2025; Sani et al., 2024).

Core Theoretical Frames

Deterministic lot-sizing

Deterministic models assume demand, lead times and costs are known in advance, allowing planners to calculate exact order quantities and timing that minimise the sum of setup and holding costs. Classic algebraic rules such as the Economic Order Quantity (EOQ) have been extended to handle time-varying demand, multiple items and capacity limits, giving rise to dynamic-programming and mixed-integer formulations that now sit inside many enterprise planning systems. Although they ignore randomness, these models teach the fundamental cost–service logic of inventory decisions and provide useful "first-cut" benchmarks before uncertainty or multiple echelons are added (CHIU et al., 2003; Robinson et al., 2009).

Examples of well-known deterministic models that incorporate optimization, analytical statistical and simulation techniques include:

- Economic Order Quantity
- Economic Production Quantity
- Wagner – Whitin Algorithm or
- Single lot (Vidal, 2023).

Stochastic inventory theory

When demand or replenishment lead times fluctuate, policy parameters must include safety slack. Continuous-review (s, Q) and periodic-review (R, S) rules are derived by balancing the expected holding cost of extra stock against the expected penalty of a shortage. Analytical and simulation studies show that state-dependent reorder points outperform fixed ones, especially for intermittent or highly variable demand. Stochastic theory therefore underpins service-level guarantees and is essential for sizing buffers that protect downstream schedules from upstream variability (Perera & Sethi, 2023; Babai et al., 2025).

Examples of well-known stochastic models that incorporate optimization, analytical statistical and simulation techniques include:

- Markov Chains
- Dynamic Programming
- Optimization or
- Stochastic Simulation (Vidal, 2023).

Multi-echelon optimisation

Real supply chains hold inventory at several layers, plants, regional hubs and local outlets so decisions made at one node affect the rest of the network. Multi-echelon models pool risk by locating safety stock where it cushions the most variability for the lowest cost, often favouring upstream placement to exploit demand aggregation. Optimisation and heuristic techniques now build these choices into a single problem that trades off total working capital against responsiveness, enabling coordinated reorder policies rather than siloed site-by-site rules (Driessen, 2025; Raj, 2024; Henriott, 2024).

Key components involved in MEIO include:

- Inventory Levels – quantity of items sitting at every stage of the supply network
- Safety Stock – extra units kept on hand to cushion unexpected swings in demand or supply
- Service Levels – chosen probability of satisfying demand without a stock-out during one replenishment cycle
- Buffer Lanes – links that connect individual inventory buffers at different sites and channel product flow through the chain
- Statistical Data – records of past demand, lead-time values and forecast errors that feed optimisation models (Raj, 2024).

Aggregate production planning and hierarchical control

Aggregate Production Planning (APP) translates medium-term sales forecasts into monthly output, workforce and inventory targets that are feasible for existing capacity. The plan is then broken down hierarchically into a master schedule, material-requirements plans and detailed job sequences. This "top-down, bottom-up" feedback loop ensures that high-level business goals remain achievable on the shop floor while letting tactical modules react quickly to short-term changes. Linking the layers prevents over-promising, smooths resource utilisation and aligns inventory budgets with production capability (Nam & Logendran, 1992; Fakhry et al., 2025).

Pull, flow and constraint-based heuristics

While optimisation generates precise targets, lean pull systems control work in process through simple visual signals. Kanban and CONWIP cards authorize replenishment only when downstream capacity becomes free, exposing delays and encouraging small-lot flow. Drum-Buffer-Rope logic, rooted in the Theory of Constraints, releases material at the pace of the slowest resource to keep queues stable. These heuristics require far less data than algorithmic schedules and, when combined with sensors and dashboards, deliver quick wins in high-mix or data-poor environments (Covell, 2023; Herer & Shalom, 2000).

Performance metrics and trade-offs

Every theoretical frame must ultimately support measurable objectives. Traditional key performance indicators include cycle-service level, fill rate, inventory turns and capacity utilisation, while modern digital programmes add data latency, energy use and carbon footprint. Because improving one metric can worsen another, decision makers rely on cost-service (and increasingly cost-service-sustainability) frontiers to visualise trade-offs. Integrated dashboards now track these indicators across echelons, enabling continuous adjustment of lot sizes, safety stock and schedules as business priorities evolve (Govindan et al., 2022; Carpitella & Izquierdo, 2025).

Artificial Intelligence in Inventory Management & Production Planning

Artificial intelligence has moved from pilot projects to mainstream tools that complement traditional operations-research methods in inventory and production planning. Recent practitioner guides emphasise that much of the data needed for AI already reside in enterprise systems; by layering learning algorithms on these streams, firms achieve faster, more granular decisions without major system overhauls. Case evidence reports double-digit gains in stock turns and lead-time reduction when AI modules replace static, rule-based settings (Kiran, 2024; O'Keeffe, 2025).

AI-enhanced demand sensing

Machine-learning forecasters like gradient-boosting ensembles, long-short-term-memory (LSTM) networks and attention-based transformers ingest promotions, weather and even social-media cues to flag demand inflections weeks earlier than classical time-series methods. Graph neural networks push the frontier further by capturing cross-store and cross-product relationships, lifting forecast accuracy by 10–30 % in benchmark studies. Improved signal quality lets planners lower safety stock without jeopardising service, freeing working capital for other uses (T. Chen et al., 2023; Olamide Raimat Amosu et al., 2024; C.-S. Chen & Chen, 2025).

Reinforcement learning for inventory policy

Deep reinforcement-learning (DRL) agents treat each echelon's inventory position as a state, learn reorder actions through simulated interaction, and update policies as conditions change. Research shows DRL cutting total cost by up to 16 % across single- and multi-echelon networks, outperforming fixed (sq.) or base-stock rules, especially when demand is intermittent or lead times vary. Because the algorithm adapts continually, it copes well with structural breaks such as product launches or supplier disruptions (Wu et al., 2023; Geevers et al., 2024; Ziegner et al., 2025).

Machine-learning production scheduling (MLPS)

MLPS brings data-driven intelligence to the classic sequencing problem by learning directly from shop-floor records and sensor streams. Current research organises these methods along two dimensions: the *decision knowledge* they supply (what to produce, why, when and how) and the *manufacturing hierarchy* they affect (product, process, machine, system). Scheduling occupies the 'know-when'– 'system' space, where algorithms translate orders, capacities and constraints into start–finish times that satisfy customer demand and plant objectives. Supervised learning models enhance traditional heuristics by predicting uncertain inputs such as setup times, machine availability, energy use more accurately than fixed estimates. Feeding these forecasts into dispatching or metaheuristic routines creates tighter due-date promises and lowers reactive rescheduling effort, particularly in high-mix environments were standard parameters age quickly (T. Chen et al., 2023; C. Chen et al., 2023; Ben Hamou et al., 2025).

Reinforcement-learning agents push the frontier further by learning complete schedules through trial-and-error interaction with a simulated shop. Reward functions reflect multi-objective goals such as makespan, tardiness and energy consumption, allowing the agent to uncover policies that cope with stochastic arrivals, breakdowns and priority shifts without handcrafted rules. Hybrid designs embed these agents inside metaheuristic search or digital twins, combining global exploration with domain-specific neighbourhood moves for faster convergence. Successful deployment follows a four-step pipeline:

- capture high-resolution production data,
- choose an algorithm matched to task complexity and data volume,
- integrate the model with execution systems for real-time feedback and
- monitor performance using dashboards that flag drift or model decay.

Key challenges remain around data governance, interpretability and change management, but case results already report double-digit improvements in utilisation and lead-time reduction-evidence that machine-learning scheduling is moving from experimental proof to practical classroom and factory adoption (T. Chen et al., 2023; C. Chen et al., 2023; Ben Hamou et al., 2025).

Predictive risk management and resilience analytics

AI also helps orchestrate inventory and capacity during disruptions. Graph neural networks propagate delay signals across multi-tier supply maps, identifying parts at risk before stock outs occur, while classification models rank mitigation levers such as alternate sourcing or expedited shipments. Early detection and targeted response lower emergency freight and lost-sales costs, embedding resilience into everyday planning rather than reserving it for crisis teams (C.-S. Chen & Chen, 2025; Wasi et al., 2024), (Das, 2025).

Digital twins and autonomous planning

A digital twin is a real-time, virtual replica of a factory or supply network that mirrors current states and constraints. Coupling the twin with reinforcement-learning or optimisation engines closes the loop: the model tests thousands of "what-if" scenarios, pushes the best policy to execution, then ingests fresh sensor data to start again. Firms report autonomous twins trimming finished-goods days-of-supply by more than 20 % while keeping service levels intact, signalling a shift from periodic re-planning to continuous self-optimisation. Market analysts predict rapid growth of these cognitive twins as cloud platforms make the technology affordable for midsize manufacturers (Das, 2025; Aktas and Bobba, 2025; *The Rise of Autonomous Supply Chains: Powered by Cognitive Digital Twins*, 2025).

Collectively, these AI applications extend classical inventory and production-planning theory into data-rich, rapidly changing environments. They illustrate how foundational models remain relevant but now operate within learning, adaptive systems highlighting the need to blend analytics, computing and domain knowledge to build the next generation of resilient, efficient supply chains.

Use Cases for AI and Advanced Technologies in Inventory Management and Production Planning

Amazon

Amazon is using digital twin technology to create virtual versions of its fulfilment centres and shipping systems. These digital models help Amazon simulate operations, monitor performance as they happen, and troubleshoot logistics issues. By combining data from IoT devices and machine learning, Amazon can predict changes in demand, find the best delivery routes, and improve the performance of its warehouses. This digital twin approach helps Amazon deliver orders faster, reduce costs and make customers more satisfied (*The Rise of Autonomous Supply Chains: Powered by Cognitive Digital Twins*, 2025).

BMW

BMW runs a virtual replica of its Regensburg plant that tracks live activities from frame painting to equipment movement in real time. Engineers and managers use this digital twin to trial adjustments and fine-tune workflows before touching the actual line, cutting downtime, avoiding expensive mistakes and speeding up improvements. The setup boosts efficiency and gives BMW the agility to respond quickly to shifting market needs and production requirements (*The Rise of Autonomous Supply Chains: Powered by Cognitive Digital Twins*, 2025).

Siemens

At Siemens' flagship PLC plant in Amberg, every workstation, AGV and inspection station is mirrored inside a cloud-hosted digital twin that ingests live sensor feeds. A reinforcement-learning engine runs continuous "what-if" experiments on the twin, evaluating thousands of sequencings, lot-sizing and change-over options per hour. The policy that minimises cycle time while keeping the tact buffer within ±2 % is pushed to the shop-floor execution system, and the twin then re-syncs with fresh data for the next iteration. Since roll-out in 2024 the plant's productivity has risen by 75 %, the defect rate has fallen to nearly zero parts per million, and overall line reliability now exceeds 99.998 %. Operators highlight two enablers: tight IoT integration that gives the twin sub-second visibility of machine states, and explainable dashboards that translate RL decisions into plain-language change-over and dispatching instructions for supervisors (*Introduction to Digital Twin Technology in Manufacturing with Unprecedented Efficiency and Innovation*, 2024; Jover & Sempere, 2025).

Future Directions and Research Agenda

Inventory management and production planning are poised to merge fully with cognitive digital-twin ecosystems. Live twins streamed from sensors will host reinforcement-learning or optimisation agents that self-tune lot sizes, safety stocks and machine sequences, while simultaneously stress-testing plans against disruption scenarios. Early evidence already links this closed loop to sharper visibility and faster recovery; future work must formalise twin-validation protocols, governance for shared models, and energy-efficient deployment on edge hardware (Roman et al., 2025).

A second trajectory centres on sustainability-aware optimisation. Multi-objective models are beginning to treat carbon, circular economy flows and social impacts as co-equal with cost and service. Open questions include how to monetise environmental externalities inside ERP data structures, how to position safety stock to minimise both emissions and risk, and how learning algorithms can navigate the enlarged trade-off surface without sacrificing transparency for operators. Rigorous case research mapping these methods into real-time decision support remains scarce, making it a high-impact avenue for scholars and practitioners alike (Soori et al., 2024).

Conclusion

This chapter charts the arc of inventory management and production planning from the earliest economic-order and lot-sizing formulas to today's Industry 4.0 ecosystems, showing how each wave of technology has expanded the scope and speed of decision-making. After a concise overview of why IMPP anchors cost, service and resilience, the text traces five historical milestones: classical analytics, MRP's mainframe breakthrough, the Lean/JIT revolution, ERP-enabled globalisation, and the current digital-twin age. It then lays out six core theoretical frames, moving from deterministic and stochastic models to multi-echelon optimisation, hierarchical planning, lean pull heuristics and KPI trade-offs. The narrative highlights how artificial intelligence now augments these foundations—using machine learning for demand sensing, reinforcement learning for dynamic reorder and scheduling, and cognitive twins for self-optimising factories. Real-world cases from retail, semiconductors and discrete manufacturing demonstrate double-digit gains in stock turns, lead-time and throughput. Finally, a forward-looking agenda calls for validated digital-twin governance and sustainability-aware optimisation, signalling a future where autonomous, eco-efficient supply networks continuously adapt to risk and opportunity.

Tasks

1. In a short paragraph, explain how a multi-echelon inventory system can reduce the total safety stock compared to holding reserves at each downstream location, and indicate the basic conditions that must be met for the application of this system.
2. Draw a sketch of a digital twin for a small assembly line (or your choice), label the data you would capture in real time, and the performance indicators (e.g., lead time, work-in-progress) that the twin would track.

6. ARTIFICIAL INTELLIGENCE IN WAREHOUSE

Artificial intelligence has permeated warehousing ever since the first automated guided vehicles (AGVs) were prototyped in the 1950 s, inaugurating a decades-long quest to let machines move, store and retrieve goods autonomously. Early warehouse-management systems (WMS) of the 1980 s and 1990 s combined barcode data with rule-based expert systems; today those rules have largely been supplanted by machine-learning models that discover optimal slotting, batching and routing patterns from historical data (de Assis et al., 2024; Keith & La, 2024). This chapter describes the basic theory in the field of warehousing and the use of artificial intelligence within this field.

Overview

Artificial intelligence is reshaping warehouse operations by automating decisions that were once manual, intuition-driven and error-prone. Warehouses now account for a significant share of logistics cost, yet they also generate rich data from radio-frequency scanners, sensors and autonomous mobile robots (AMRs) that AI can exploit to cut travel, labour and space waste. Recent empirical studies on AI-assisted order-picking and surveys of AMR deployments agree that data-driven optimisation has moved beyond pilots and is becoming a mainstream source of competitive advantage in fulfilment centres (Keith & La, 2024; Rad et al., 2025).

AI techniques fall into three broad groups:

- *Predictive machine-learning models* (e.g., gradient boosting and neural networks) forecast demand, receiving volumes and congestion hotspots.
- *Reinforcement-learning agents* learn routing and batching policies by interacting with digital replicas of the warehouse.
- *Robotic control algorithms* guide AMRs or articulated arms in real time, adjusting paths as conditions change.

Research on deep reinforcement learning for dynamic order picking and integrated optimisation of AMR-assisted systems shows double-digit improvements in throughput and picker travel compared with rule-based heuristics, highlighting AI's ability to handle the combinatorial complexity inherent in high-variety, high-velocity facilities (Zhen et al., 2025; Mahmoudinazlou et al., 2025).

Across core processes like receiving, put-away, slotting, picking, packing and shipping, AI delivers quantifiable gains. Predictive analytics reposition fast movers closer to consolidation lanes, vision systems flag mis-scans or damages in real time, and autonomous vehicles perform cycle counting and zone picking without downtime. Field results from AI-enabled batch-picking pilots report stock-out reductions of up to 25 % and inventory-holding cuts of around 15 %, confirming that machine-learning forecasts combined with adaptive reorder and routing policies translate directly into higher service levels and lower working capital (Rad et al., 2025; Samuels, 2025).

Successful adoption, however, depends on robust data pipelines and change management. Studies of AI rollouts emphasise the need for high-resolution event logs, tight integration between warehouse-management systems and machine-control layers, and operator training that demystifies algorithmic recommendations. Reviews of AMR programmes further warn that safety, traffic coordination and maintenance protocols must evolve alongside the technology. Addressing these organisational and technical prerequisites ensures that AI tools remain explainable, scalable and aligned with overall supply-chain objectives (Rad et al., 2025; Keith & La, 2024).

Historical Evolution of Warehousing and Modern Technology

Early warehouses were little more than sheltered storage yards, sized to match local trade flows. Mechanisation began in the 1910s with the lift-truck and, three decades later, universal pallet standards let goods be stacked higher and handled faster; the combination transformed floor-space economics and triggered the first purpose-built distribution centres in post-war North America and Europe. Academic texts of the 1960s describe these facilities largely in terms of layout geometry (aisles, racks, docks) because material-handling equipment, not information, was still the primary performance lever (Dolgui & Proth, 2010).

From the 1970s to the late 1980s, information technology joined mechanical handling. Bar-code symbology (1974) and radio-frequency data capture made it feasible to track individual cartons; automated storage and retrieval systems (AS/RS) delivered high-density unit-load retrieval; and early warehouse-management systems (WMS) running on minis and mainframes began to orchestrate put-away, replenishment and order picking. Bibliometric reconstructions of this period show a rapid rise in research that links facility design to data visibility, signalling the field's first shift from pure engineering to cyber-physical optimisation (Bottani et al., 2025).

In the 1990s and early 2000s e-commerce, mass customisation and global sourcing stretched fulfilment networks. Enterprise-resource-planning suites connected WMS to upstream planning modules, while wave picking, cross-docking and voice-directed systems emerged to cope with shorter product life cycles and higher order granularity. Smart-warehouse reviews note that this era cemented the principle of real-time control: decisions on slotting, batching and routing were no longer made once a shift but recalculated continuously as orders flowed in (Tiwari, 2023).

Industry 4.0 concepts, sensor ubiquity, cloud analytics and cyber-physical connectivity have defined the 2010s. Flexible automation moved beyond fixed conveyors to fleets of automated guided vehicles (AGVs) and, more recently, autonomous mobile robots (AMRs) that navigate with simultaneous-localisation-and-mapping algorithms. A comprehensive 2024 review of AGV adoption shows that integrating mobile robots with WMS increases picking productivity by double digits while enabling modular system scaling; the same study frames "flexible automated warehouses" as the reference architecture for volatile omnichannel demand (Ellithy et al., 2024).

Today's frontier couples those physical assets with artificial intelligence. AI-driven warehouse-automation surveys document applications ranging from gradient-boosting demand forecasts that reposition fast movers nightly, to reinforcement-learning agents that orchestrate AMR swarms in real time. Parallel work on warehouse digital twins describes virtual replicas that ingest live sensor streams, test thousands of what-if scenarios per hour and feed the best policy back to execution systems—creating a closed, self-optimising loop. Together, these advances signal a warehouse paradigm where data latency approaches zero and continuous learning supplants periodic re-engineering (Enoch Oluwademilade Sodiya et al., 2024; Drissi Elbouzidi et al., 2023).

AI Techniques for Warehousing

Artificial-intelligence tools used in warehouses fall into three broad families; predictive analytics, prescriptive optimisation and autonomous control and a recent comprehensive review classifies them along the "decision-knowledge" axis (know-what, -why, -when, -how) and the manufacturing hierarchy (product → system). This framework helps managers map each AI method to the warehouse task it can automate, from demand sensing to multi-robot coordination, and shows that the greatest performance gains emerge when several techniques collaborate in one cyber-physical loop (Enoch Oluwademilade Sodiya et al., 2024).

Predictive machine learning

Predictive machine learning improves strategic slotting and short-term labour planning. Gradient-boosting and neural-network models learn the joint effect of product dimensions, velocity and order patterns, and then recommend nightly relocations that shorten travel distance for the next day's picks. Case studies with integer-programming validators report double-digit cuts in picker kilometres, labour hours and aisle congestion when ML-based slotting replaces static ABC rules (Duque-Jaramillo et al., 2024).

Computer vision

Computer vision turns image streams into actionable data. 3-D cameras on dock doors and AMRs verify pallet condition, detect skewed loads and locate carton barcodes without human intervention. Recent vision pipelines fuse object-detection networks with lightweight attention modules so the system can run on edge devices, enabling real-time alerts that prevent product damage and unsafe stacking while feeding high-quality data back to the WMS (Chen et al., 2025).

Reinforcement learning (RL)

RL addresses the combinatorial explosion in dynamic order-picking. An RL agent views the pickers or robot's location, the unpicked orders and aisle congestion as its state, receives rewards for early completions and low travel, and iteratively improves its routing policy by simulation. Benchmarks show RL cutting makespan and tardiness by up to 20 % versus rule-based or savings-heuristic baselines, particularly when orders arrive continuously throughout the shift (Mahmoudinazlou et al., 2025).

Multi-agent path planning

Multi-agent path planning manages fleets of autonomous mobile robots (AMRs). Deep-learning policies trained in a digital twin jointly optimise collision-free routes and task allocation, adapting in milliseconds to new jobs or obstacle reports. Experimental deployments demonstrate throughput gains and smoother traffic flow compared with first-come-first-served or fixed-path controllers, confirming that AI coordination scales better as fleet size grows (L. Zhang et al., 2024).

Natural-language and voice interfaces

Natural-language and voice interfaces close the loop between AI recommendations and human operators. Voice-directed picking systems that embed natural-language understanding interpret casual speech, confirm item IDs and quantities, and adjust pick paths on the fly. Trials combining indoor-positioning, RFID and voice AI cut confirmation errors by a third and raise picker productivity, showing that conversational interfaces can deliver AI guidance without adding cognitive load (M. Xu et al., 2025).

Core Warehouse Processes Enhanced by AI

Inbound receiving and inspection

Computer-vision pipelines now sit above receiving bays, where depth cameras and point-cloud algorithms identify pallet IDs, locate label positions and spot deformations the moment freight is unloaded. The approach in *Sensors* (2023) shows millimetre-level pose accuracy on mixed pallet loads and detects damaged stringer boards with a 98 % F1 score, so defects are quarantined before they enter stock and forklift scans drop by half (Shao et al., 2023).

Smart put-away

After verification, storage decisions are generated by a learning-to-rank optimiser that weighs SKU velocity, cube utilisation and adjacency rules. A sustainability-focused study demonstrates that firework-algorithm slotting in non-traditional layouts trims forklift travel 18 % and raises rack utilisation 11 % over static ABC zoning, proving that AI can re-compute layouts nightly as demand shifts (X. Zhang et al., 2023).

Recent slotting research combines integer-programming with simulation to assign SKUs to storage locations while respecting product dimensions, demand heterogeneity and aisle layout. Tests with one year of real order data show that using ABC-based priority plus a sequential row-level-column allocation rule yields the shortest total slotting time among several strategies, confirming that data-driven placement can outperform static zoning without wholesale layout changes (Duque-Jaramillo et al., 2024). Complementary work on reinforcement-learning-based slotting models the put-away decision as a Markov process and learns policies that minimise combined travel and relocation time; simulation experiments on a mid-size warehouse report noticeable travel-distance savings and higher location-utilisation rates compared with heuristic rules, especially when SKU demand is volatile (Zhou, 2024).

Dynamic slotting for fast movers

Deep-reinforcement-learning (DRL) agents embedded in robotic-mobile-fulfilment systems (RMFS) reslot only the top 10 % of SKUs each shift. By simulating picker traffic, shelf availability and order affinity, the DRL policy cut total travel distance 22 % and held labour hours constant even as order lines climbed, outperforming rule-based heat-map methods (Zhu et al., 2024).

Order-picking orchestration

When fleets of autonomous mobile robots (AMRs) share narrow aisles, multi-agent path-planning AI recalculates collision-free routes in sub-second cycles. A 2025 review of AMR navigation trends reports throughput gains above 30 % versus fixed-path or first-come dispatch, plus smoother traffic densities that boost battery life and safety compliance (Tang et al., 2025).

Recent work has shown that the greatest gains in robot-assisted picking come from co-optimising two tightly coupled decisions: how orders are assigned to robots (task scheduling) and how those robots navigate aisles without colliding (multi-agent path finding, MAPF). A 2023 open-access study in *Complex & Intelligent Systems* formalises both choices in one integrated model and proposes an “enhanced HEFT” scheduling rule paired with a conflict-based search path-planner. Simulation tests on warehouse layouts of varying size demonstrate that the joint algorithm consistently shortens makespan and reduces robot traffic density when compared with solving scheduling and routing separately—evidence that orchestration, not isolated optimisation, is the key to unlocking higher throughput from autonomous-mobile-robot fleets (Honglin et al., 2023).

Packing and quality confirmation

Depth-camera workstations now use a lightweight YOLO-based detection pipeline to confirm that each carton contains the correct SKU and quantity before sealing. Experiments on a live fulfilment line showed a 98.6 % accuracy in detecting mis-picked items and a 97 % success rate in flagging under-filled cartons, while inspection time per box remained under 90 ms. Because verification is automatic and hands-free, operators can maintain a steady packing rhythm without pausing to scan each unit, which in turn raises overall throughput and reduces rework cost (Kim et al., 2024).

The study presents a computer-vision pipeline that continuously inspects wooden pallets as they enter an automated high-rack warehouse. By capturing images from a camera mounted in the pallet-entry tunnel and applying a YOLO-based deep-learning model, the system classifies defects such as cracked stringers and broken boards in real time, assigns a quality score to each pallet and triggers diversion of damaged units before they reach storage racks. Test runs showed that this AI-enabled gatekeeper can screen every inbound pallet without stopping conveyor flow and thereby prevents unplanned crane outages, cutting total system downtime and maintenance effort (Giner et al., 2023).

Cycle counting and inventory accuracy

Drones equipped with artificial intelligence are increasingly being used for nightly cycle counting. Equipped with RGB-D cameras and barcode readers, they follow pre-programmed routes, capture images of each pallet location, and send the data to a visual model that reconciles what the drone "sees" with warehouse management system records. Field studies collected in a 2023 systematic review of unmanned aerial vehicle (UAV) applications report high location accuracy and a reduction in working hours compared to manual counting, with operations continuing without interruption (Malang et al., 2023).

Reverse-logistics triage

Artificial intelligence tools now sort and dispose of returned goods at the dock. Machine learning classifiers analyse product images, sensor tags, and customer return reasons to predict resale value, repair costs, or recycling methods, then route items to the appropriate channel—restocking, refurbishment, secondary market, or disposal. A 2024 bibliometric study on artificial intelligence in reverse logistics shows that combining deep learning image recognition and optimization models can shorten the return-to-warehouse cycle and increase recovery value, while also providing traceability data for sustainability reporting (Bhowmik et al., 2024).

Technology Infrastructure

The technological infrastructure of an AI-enabled warehouse is built on seven interlocking layers:

1. Physical IoT devices (sensors and actuators)
2. Communication protocols (e.g. MQTT, OPC-UA)
3. Middleware to unify device interfaces
4. Edge-computing components for local AI processing
5. Digital-twin platforms for simulation and monitoring
6. AI-driven WMS modules (forecasting, vision, optimization)
7. Embedded cyber-risk analytics for security and resilience.

Each layer exposes well-defined interfaces to the next, enabling modular deployment and seamless upgrades of AI capabilities (Khan et al., 2022).

Physical IoT devices

IoT devices form the sensory and action arms of the system. In a typical deployment, weight scales, temperature gauges, barcode/RFID readers and 3-D cameras capture stock levels, environmental conditions and item identities, while conveyor belts, robotic shuttles and pick-and-place arms execute movement and handling tasks. By distributing these devices across receiving, storage, picking, packing and dispatch zones, warehouses collect a rich, high-granularity stream of operational data, feeding downstream AI services with real-time inputs (Khan et al., 2022).

Communication protocols

Communication protocols carry device data to higher layers with the reliability and performance that warehouse operations demand.

- *MQTT* (Message Queuing Telemetry Transport) is a lightweight, publish/subscribe protocol: each sensor "publishes" small data packets (messages) to a broker, and any application that "subscribes" to those topics receives updates immediately—ideal for battery-powered or bandwidth-constrained devices.
- *OPC-UA* (Open Platform Communications – Unified Architecture) provides a richer framework, modelling data in a hierarchical "address space," enforcing encryption and authentication, and standardizing data exchange among heterogeneous equipment (Khan et al., 2022).

Middleware layer

Sitting above these protocols is a middleware layer that sits between the raw IoT devices and the higher-level applications, providing a common interface regardless of the underlying hardware. The middleware performs three key functions:

- Data ingestion and normalization: it collects diverse data formats (e.g., raw scale readings, camera frames, RFID scans), converts them into a unified message structure, and tags them with metadata (timestamps, device IDs).
- Command abstraction: it exposes a standard set of control commands to actuators—so that, for example, a "move-to-location" request has the same API call whether it's sent to a conveyor belt, an AMR, or a robotic arm.
- Device management: it handles registration, health monitoring and firmware updates for all connected devices, hiding vendor-specific protocols behind a single management console.

By abstracting these details, the middleware allows new sensors or robots to be integrated simply by installing a driver that conforms to the middleware's plug-in model, rather than by rewriting dozens of proprietary interfaces dramatically reducing development effort and accelerating the rollout of AI-driven features (Farahzadia et al., 2017).

Edge Intelligence

To meet stringent latency, bandwidth and privacy requirements, AI workloads shift from the cloud to the network edge. Edge intelligence decomposes AI processing into four components:

- Edge caching for temporary storage of raw sensor streams

- Edge training to update models using local data
- Edge inference for real-time decision making close to devices
- Edge offloading to dynamically delegate tasks between edge nodes and central servers.

This architecture ensures that time-critical functions such as vision-based inventory counting or motion planning execute with minimal delay and without saturating the warehouse network (D. Xu et al., 2020).

Digital-Twin Platform

Overlaying the physical and edge layers, a digital-twin platform provides a synchronized virtual replica of the warehouse. In the *virtualization/synchronization* phase, real-time sensor feeds (e.g., IoT device streams) are fused into a high-fidelity model; during *monitoring/awareness*, analytics engines detect anomalies or impending maintenance needs; and in *decision-making/optimization*, reinforcement-learning agents and generative-design algorithms adapt layouts, task schedules, and resource allocations. By validating strategies first in simulation and then applying them to the real warehouse, operators gain agility and reduce operational risk (Sajadieh & Noh, 2025).

AI-enhanced Warehouse Management System

At the core of day-to-day control sits the AI-enhanced Warehouse Management System (WMS). Its micro service architecture embeds:

- *Demand-forecasting* modules using hybrid deep-learning models
- *Computer-vision audits* (e.g., Faster R-CNN for pallet detection)
- *Combinatorial task-scheduling* solvers augmented with learned heuristics
  Each service exposes a RESTful API to the WMS, allowing “hot-swapping” of AI components to continuously refine picking efficiency, minimize idle time and adapt to evolving order profiles (Sodiya et al., 2024).

Embedded Cyber-Risk Analytics

Finally, embedded cyber-risk analytics safeguard the entire AI warehouse stack. Real-time telemetry from RFID readers, programmable logic controllers (PLCs) and network sensors feeds into a cognition engine, which applies anomaly-detection algorithms and federated-learning based threat models to surface and mitigate attacks at the edge. By continuously monitoring device health and network traffic, this system ensures both data integrity and operational continuity in AI-powered warehouses (Radanliev et al., 2020).

Conclusion

Artificial intelligence has revolutionized warehouse operations by transforming manual, intuition-driven tasks into data-driven, automated processes. From the earliest automated guided vehicles and rule-based warehouse-management systems, AI techniques now span predictive analytics (e.g., demand forecasting with gradient boosting and neural networks), reinforcement learning for dynamic routing and batching, and real-time robotic control for autonomous mobile robots and articulated arms. These methods deliver double-digit improvements in throughput, picker travel distances, stock-out rates and inventory accuracy across core workflows; receiving, put-away, slotting, picking, packing, shipping and reverse logistics while emerging interfaces such as natural-language voice systems and UAV-based cycle counting further integrate human operators and unmanned assets into a seamless fulfilment ecosystem.

Underpinning these capabilities is a layered technological infrastructure that begins with distributed IoT devices (sensors and actuators) and industrial communication protocols (MQTT, OPC-UA), unified by middleware that normalizes data and abstracts control commands. To meet stringent latency, bandwidth and privacy requirements, AI workloads increasingly run at the network edge handling caching, training, inference and offloading locally while a synchronized digital twin offers real-time simulation and optimization. At the software core, microservices within the Warehouse Management System host modular AI components (forecasting, vision audits, scheduling) that can be hot swapped via APIs and embedded cyber-risk analytics continuously monitor and protect the entire stack. Together, these layers enable warehouses to operate as self-optimizing, resilient cyber-physical systems that adapt continuously to changing demand, resource availability and operational constraints.

Tasks

1. Describe the seven interlocking layers of the AI-enabled warehouse infrastructure and explain the primary role of each layer.
2. Compare edge intelligence and digital-twin platforms in AI-powered warehouses: how does each approach contribute to real-time decision making and operational resilience?

7. ARTIFICIAL INTELLIGENCE IN LOGISTICS AND IN TRANSPORT PROCESSES

Artificial Intelligence is reshaping logistics and transport by converting heterogeneous operational data such as orders, telematics, sensor feeds, weather, and infrastructure states into timely, auditable decisions that lift efficiency, resilience, safety, and sustainability. Under intensifying global trade and environmental constraints, organizations depend on AI to align capacity with volatile demand, anticipate bottlenecks, and orchestrate scarce resources without eroding service quality. Well-designed data pipelines, robust features, and human-in-the-loop controls let AI compress decision latency across planning through execution while supporting decarbonisation through fewer empty miles, reduced idling, and better asset utilization (Cannas et al., 2023; Culot et al., 2024).

A practical way to frame AI in this domain is as a layered capability stack - prediction, optimization, and perception embedded in closed decision loops. Prediction spans granular demand sensing, dwell estimation, and risk alerts; optimization steers network design, inventory positioning, pricing, and multi-constraint routing and scheduling; perception digitizes the physical world via computer vision and IoT to validate picks, detect hazards, or verify equipment states. When these layers are integrated with operational systems, organizations move from static plans to event-driven re-planning, reducing variability while meeting tight service-level and emissions targets. This integration depends as much on data governance and change management as on algorithms (Culot et al., 2024; Woschank et al., 2020).

Within warehousing and fulfilment, AI advances both “see” and “decide” functions. Edge vision confirms picks and packaging, reads labels in difficult lighting, and flags damage; learning-based task allocation and slotting balance travel time, congestion, and ergonomics; and digital-twin models let teams test wave designs, labour mixes, and automation settings before deployment. Reliable gains come from right-sizing models for on-device inference, curating datasets with active learning to accommodate SKU churn, and wiring detections to actions in labour and automation controllers so every alert has a measurable operational consequence (Assis et al., 2024; Villegas-Ch et al., 2024).

In road transport and urban logistics, AI enables adaptive signal control, demand-responsive services, dynamic fleet dispatch, and last-mile optimization. Reinforcement learning has matured from toy networks to corridor- and network-scale signal timing, learning phase splits, offsets, and green waves directly from data streams; pairing those policies with robust calibration and safety monitors translates algorithmic gains into reduced delay and emissions in practice. For fleets, learning-augmented routing combines accurate travel- and service-time predictions with fast re-optimization as orders stream in, supporting tighter windows and higher asset utilization. At the curb, data-driven reservation and pricing treat loading zones as optimizable assets, curbing double parking and improving safety for buses and cyclists (Elbouzidi et al., 2023; Rasheed et al., 2020).

Rail derives substantial near-term value from condition-based maintenance and energy-aware operations. Models trained on vibration, acoustic, thermal, and electrical signatures anticipate faults in tracks, points, bogies, and traction systems, enabling planned interventions that cut disruptions and extend asset life; coupling these predictions to maintenance windows, crews, and spares makes the savings realizable. On the operations side, passenger-flow and dwell-time predictions stabilize timetables, while driver advisory and speed-profile smoothing reduce energy consumption and help maintain punctuality, particularly when topography and rolling-stock constraints are respected (Tang et al., 2022; Sarp et al., 2024).

In aviation, AI strengthens trajectory prediction, sector complexity assessment, conflict detection, and proactive flow management, and supports speech technologies for controller–pilot exchanges and document processing in operational centres. Airlines and airports use ML to fuse weather, network propagation, and surface state to anticipate delays, resequencing turnarounds, allocate gates and crews dynamically, and personalize passenger handling; predictive maintenance on engines and avionics reduces unscheduled removals and improves availability. Safe adoption hinges on transparent models, workload-aware human AI teaming, and rigorous assurance processes that align with the sector's safety culture (Degas et al., 2022; Esmaeilzadeh et al., 2020; Demir et al., 2024).

Maritime transport leverages learning on AI, meteorological, and oceanographic data to improve route planning, weather routing, collision-risk assessment, and anomaly detection, while ports apply AI to Estimated Time Arrival (ETA) prediction, berth allocation, quay-crane scheduling, yard reshuffles, and gate and truck-appointment operations. Treating ETA uncertainty explicitly lets planners hedge resources against variability; combining predictive layers with prescriptive optimizers shortens ship turnaround, reduces truck idling, and raises throughput without new infrastructure delivering operational and environmental benefits together (Yang et al., 2024; Munim et al., 2020).

Multimodal and synchromodal orchestration depends on continuous forecasts of demand, capacity, travel times, and risk to re-choose modes as conditions change. Preference-learning captures shippers' trade-offs among reliability, cost, and carbon; agent-based models explore disruption responses and coordination rules; and operational re-planning integrates trucks, trains, barges, and short-sea with consistent service promises. The core principle is simple: predictive insights must be wired to contractual and platform mechanisms that trigger timely switches across modes (Alei et al., 2024; Giusti et al., 2019).

Sustainability and governance anchor responsible scaling. In practice, AI enables eco-routing, consolidation, energy-aware scheduling, and strategic inventory placement that reduce empty miles and waste; better ETA and capacity forecasts shrink buffers, while improved traceability strengthens emissions accounting especially for Scope 3. Organizations also need explainability, bias monitoring, model-risk controls, and data-privacy safeguards so automated decisions remain transparent, fair, and compliant; measuring the computational footprint and favouring frugal architectures ensure environmental gains are not offset by inefficient model operations (Cannas et al., 2023; Nikseresht et al., 2025; Woschank et al., 2020).

An implementation playbook turns pilots into enterprise impact: select use cases with clear owners and KPIs (on-time performance, cost-to-serve, CO_2 per unit); stand up a shared data layer and feature store across orders, assets, locations, weather, and events; deploy human-in-the-loop decision support so operators can override and teach the system; and “close the loop” by pushing predictions into optimizers and front-line systems where decisions are executed. Continuous monitoring for drift, uncertainty, and unintended effects, combined with change-management that upskills teams, sustains gains over time (Culot et al., 2024; Cannas et al. 2023).

Rising Competition and the Need for Efficiency

Global competition, shorter product lifecycles, and volatile demand have made static planning and manual exception-handling untenable. AI closes the gap by pairing high-frequency operational data with predictive and prescriptive models that sense demand shifts, forecast lead times and dwell, and continuously re-optimize networks under changing constraints. In practice, the same stack that forecasts orders also feed routing, slotting, and crew/asset assignment, so plans adapt in minutes rather than weeks. Companies that institutionalize this loop; data ➜ prediction ➜ optimization ➜ execution ➜ feedback, see measurable gains in service level, cost to serve, and reliability, while competitors tied to rigid rules accumulate delays and buffers. Concretely, AI improves ETA prediction, flags emerging bottlenecks, and reroutes around disruptions in real time, helping firms deliver faster at lower cost without adding capacity (Cannas et al. 2023; Pournader et al., 2021).

Cost Reduction and Operational Efficiency

Transport and warehousing are dominant cost centres, and AI attacks their biggest inefficiencies. Learning-augmented routing cuts empty miles and idling by fusing traffic, weather, curb access, and service-time predictions with fast re-optimization; energy-aware dispatch reduces fuel burn without missing time windows. Predictive maintenance models trained on vibration, temperature, and electrical signals for vehicles, lift trucks, and fixed assets detect precursors to failure, shifting repairs into planned windows and shortening mean time to repair. Inside warehouses, computer vision validates picks, reads labels in poor lighting, spots damage, and drives cycle counts; task-allocation and slotting engines then minimize travel and congestion while respecting ergonomic limits. The result is fewer unplanned stoppages, higher throughput, and a structurally lower cost base (Chen et al., 2024; Mahale et al., 2025).

Meeting Consumer Expectations

E-commerce has reset expectations toward fast, precise, and transparent delivery often with narrow windows and flexible drop options. AI helps providers meet these promises by forecasting micro-area demand, staging inventory closer to customers, and planning last-mile tours that account for time-dependent travel, parking, building access, and customer availability. When conditions shift mid-route, policies learned from historical operations generate near-instant re-plans that protect on-time performance and reduce failed attempts; at the same time, lockers, pickup points, and dynamic delivery options are priced and offered using the same predictive stack. Importantly, customer-preference modelling and service-time learning improve both reliability and perceived quality, not just speed (Giuffrida et al., 2022; Samani et al., 2025).

Handling Increasing Data Complexity

Modern networks generate torrents of multimodal data like orders, telematics, scanner events, weather, capacity notices that exceed human processing. AI and big-data analytics provide the scaffolding to ingest, clean, fuse, and learn from these streams so planners see leading indicators instead of lagging KPIs. For demand planning, machine- and deep-learning models outperform static baselines when they blend temporal patterns with promotions, holidays, and exogenous signals, and when uncertainty is quantified to inform buffers and contracts. The same data backbone supports anomaly detection for fraud and shrink, and feeds optimization layers for inventory positioning and replenishment (Maheshwari et al., 2020; Douaioui et al., 2024).

Sustainability Goals

Decarbonisation targets and fuel costs push logistics toward “do more with less movement.” AI contributes by finding consolidation opportunities, proposing eco-routes that trade a small detour for a large emissions drop, smoothing speed profiles, and sequencing operations to avoid queueing and cold-chain breaches. In planning, AI informs network design (nodes, modes, and service policies) that reduces embodied transport; in execution, it tunes set points on equipment and selects low-emission options when risk and cost allow. Traceability also improves: data-driven footprints at shipment and lane level support reporting and greener procurement (Chen et al., 2024; Naz et al., 2022).

Adaptation to Disruptions

From port closures and extreme weather to labour actions and component shortages, disruptions are now the norm rather than the exception. AI strengthens resilience by combining early-warning signals (news, weather, social, operations) with scenario generation and fast, policy-based re-optimization of sourcing, modes, and routings. During events, agent-based or reinforcement-learning policies help allocate scarce capacity, throttle promises and prioritize high-value orders; after events, post-mortems feed new features back into models to harden responses. Emerging practices also show generative tools assisting control-tower triage and communication while traditional ML handles quantitative decisions (Yang et al., 2022).

Innovation and Future

AI's value multiplies when paired with enabling infrastructure like 5G for low-latency telemetry, edge computing for on-site inference, IoT for pervasive sensing, and blockchain for tamper-evident provenance. Together they unlock smart-city logistics: real-time curb management, dynamic geofencing for zero-emission zones, cooperative traffic systems, and secure data sharing across carriers, terminals, and municipalities. Firms that design their platforms to ingest these streams and run models close to where data is born can iterate faster, orchestrate across partners, and adopt autonomy and digital-twin workflows as they mature (Idriss et al., 2024; Lagorio et al., 2022).

AI is now embedded deep in core networks. On roads, learning-based signal control coordinates corridors and networks to reduce delay and emissions, and fleet-wide policies balance reliability with energy use. In rail, condition monitoring of track, rolling stock, and power systems anticipates faults and schedules interventions that protect punctuality and asset life while supporting energy-efficient driving. In aviation, AI elevates trajectory prediction, sector complexity assessment, and flow management, and it supports speech-enabled tools that lighten controller workload always within strict safety-assurance and boundaries (Noaeen et al., 2022; Du et al., 2025).

Freight railroads have long struggled with the combinatorial "rail-car itinerary" problem deciding where each wagon should be re-classified so that overall transit time, yard congestion and crew hours are minimised. A 2024 paper proposes a graph-neural-network, deep-reinforcement-learning controller that continuously learns from live wagon flows and yard states. When tested on a year of traffic from a 30-track Asian marshalling yard the agent cut average dwell time by 14 percent, reduced re-marshalling moves by 21 percent and freed enough yard capacity to route two extra block trains per day without new infrastructure. National railways are now commercialising the concept as “AI Yard Planner” modules within Siemens Railigent and the DB Cargo digital yard suite (Zhang et al., 2025).

AI Use Cases in Logistics and Transport

Real-time fleet routing and dispatch

Carriers now fuse short-horizon demand forecasts, live traffic, curb access constraints, and driver/vehicle duty rules into learning-augmented routing loops. In practice this means predicting service and travel times at stop and link level; using those predictions inside fast heuristics or reinforcement-learning policies to re-plan routes as orders stream in; and pushing decisions to on-board/driver apps every few minutes without breaking time-window promises. The result is fewer empty miles, fewer failed attempts, and tighter windows that hold even under non-stationary demand. Fielded approaches typically amortize computation across repeated daily instances (same city, similar order geography) and keep latency low enough for on-the-fly swaps when a late pickup, closure, or storm appears (Rios et al., 2021; Ke et al., 2025).

Adaptive traffic signal control for freight corridors

Deep reinforcement learning (DRL) and multi agent reinforcement learning (MARL) have advanced from small grid simulations to corridor scale and network scale traffic signal control, where policies learn phase splits, offsets, and green wave coordination from detector and probe vehicle data. In freight heavy corridors, this can reduce queue spillback that disrupts pickup and delivery tours. Successful deployments pair DRL policies with site specific calibration, safety guardrails, and fall back modes that allow operator override in edge cases. Reported outcomes include reduced delay and lower emissions, particularly when the reward includes freight platooning or bus priority terms (Haddad et al., 2022; Swapno et al., 2024).

Curb and loading-zone management

Cities are treating the curb as a programmable resource. AI models forecast commercial vehicle arrivals by block face and time of day, then inform reservation and dynamic pricing policies that maintain turnover and reduce double parking. For carriers, the benefits include fewer missed deliveries, shorter dwell times, and more reliable last mile schedules; for cities, improved bus reliability and safer conditions for cyclists and pedestrians. Recent work also frames curb allocation as a stochastic optimisation problem with uncertainty in arrivals and dwell times, allowing operators to hedge supply while preserving access for priority users (Lim & Masoud, 2024; Burns et al., 2024).

Truck appointment systems and gate operations at ports

Container terminals use AI-enhanced truck appointment systems (TAS) to smooth peaks at gates and in yards. Forecasts of vessel ETAs, crane moves, and yard occupancy inform quota planning by time slot and container class; optimization then assigns and reschedules appointments to minimize turnaround time while avoiding yard congestion. On the carrier side, integration with dispatch lets fleets sequence drays to chain import/export moves and reduce empties. Empirical studies report shorter queues, higher gate throughput, and more predictable turn times when TAS are co-designed with data-driven quotas and simulation-backed rescheduling (Caballini et al., 2020; Bett et al., 2024).

Real-time terminal orchestration

Inside the terminal, predictive ETA distributions and workload forecasts feed prescriptive models that co-optimize berth choice, start times, and quay-crane sequences. Recent contributions use learning-based or robust methods to handle uncertain arrivals, tides, and energy limits, enabling mid-shift adjustments without cascading delays. Combined with truck appointments, these schedulers compress vessel turnaround, reduce idling of expensive equipment, and lower energy consumption per move (Chargui et al., 2023; Lv et al., 2024).

Airline gate/stand assignment and crew-related re-planning

Airports and airlines increasingly take a “predict-then-optimize” approach: high-resolution arrival predictions (accounting for weather, surface states, and network propagation) feed gate/stand assignment to reduce towing, conflicts, and passenger walking, while keeping critical connections. On the airline side, planners blend learned preference models with classical crew-pairing/rostering to reflect hard rules plus tacit shop-floor knowledge (e.g., station familiarity), improving downstream recoverability when delays hit. The operational payoffs are fewer last-minute gate swaps, more stable turnarounds, and fewer legality violations during disruptions (Cao et al., 2024; Gattermann-Itschert et al., 2023).

Ride-hailing / ride-pooling dispatch and pricing (lessons for on-demand logistics)

Large-scale platforms have shown that learning-based dispatch and spatio-temporal pricing can run in production and improve both utilization and service metrics. These systems learn assignment and repositioning policies that anticipate future supply–demand imbalances, then price and match accordingly, logic that parcel and instant-delivery operators are adapting for courier fleets. Documented deployments report measurable gains in driver income/utilization and order fulfilment, demonstrating how reinforcement-learning control loops can operate reliably at city scale (Qin et al., 2020; Huang et al., 2022).

Energy-aware driving and automatic train regulation

Urban rail and metro operators use AI both to advise drivers (or ATO systems) on energy-optimal speed profiles and to regulate headways in real time when dwell times fluctuate. Models incorporate gradients, limits, regenerative braking, and passenger loads; regulators then make small timing and speed adjustments to stabilize intervals while saving traction energy. When integrated with timetable control and power constraints, these tools have shown simultaneous gains in punctuality and energy efficiency (Peng et al., 2024; Zhang et al., 2025).

Tasks

1. Why is it important to implement AI in logistics and transport today?
2. Choose one case study from the text where AI was implemented in logistics or transport and describe it.

8. ARTIFICIAL INTELLIGENCE – ENABLED ROBOTS

AI-enabled robots combine advanced mechanical systems with artificial intelligence to perform tasks with greater autonomy, adaptability, and efficiency than traditional robots. By integrating machine learning, computer vision, and natural language processing, these systems can perceive their environment, make decisions, and interact with humans in dynamic and unstructured settings. Their applications span industries such as manufacturing, healthcare, logistics, and service sectors, driving innovation and transforming the way humans and machines collaborate.

AI-enabled robotic systems in logistics can be classified into several main categories, including automated guided vehicles (AGVs), which follow predefined paths using markers or sensors; autonomous mobile robots (AMRs), which navigate dynamically using environmental mapping and obstacle detection; and robotic arms or manipulators, which perform precision tasks such as picking and sorting. These systems integrate sensing technologies, adaptive decision-making, and spatial awareness to carry out material transport, order fulfilment, and internal routing with minimal human oversight. Their deployment reduces manual handling, increases process consistency, and enhances operational safety in complex warehouse and distribution environments (Bernardo et al., 2022).

Beyond pure automation, AI-enabled robots are central to achieving resilient and efficient supply chains. By working within digitized orchestration platforms and warehouse management systems, they contribute to flexible task allocation, rapid response to operational disruptions, and sustainable performance through intelligent maintenance and optimized deployment. This integration supports adaptability to fluctuating demand and evolving supply chain configurations, ensuring that logistics networks remain both responsive and cost-effective (Ferreira & Reis, 2023).

The future development of AI-enabled robots in logistics is closely linked to advances in connectivity, interoperability, and data-driven decision support. Emerging directions include multi-robot cooperation (including swarm-inspired coordination) for cooperative transport, tighter integration with IoT-enabled sensing and platforms for real-time status monitoring, and predictive analytics for maintenance and workload forecasting. As these capabilities mature, they support more autonomous material flows, reduce downtime, and align operations with energy-efficiency and sustainability objectives (Gielis et al., 2022; Shahzad et al., 2023; Idrissi et al., 2024).

Automated Guided Vehicle

An AGV is an unmanned vehicle that automatically follows marked paths such as wires, magnets, or visual cues to transport materials within industrial or warehousing environments. It operates without a human driver yet relies on predefined routing (Wilson, 2015).

Autonomous Mobile Robots

AMRs are intelligent machines designed for independent operation in dynamic environments. They stand out from traditional robots by their capacity for autonomous navigation and real-time decision-making, facilitated by advanced sensors and software. Unlike fixed transport industrial robots, AMRs possess the ability to adapt and modify their routes and tasks as needed, enhancing their versatility for applications in dynamic settings (Isbell, 2025).

Historical Evolution

Foundations of Industrial Robotics (1950s–1960s)

The modern era of robotics in industry took shape with early programmable manipulators that formalized the idea of repeatable, automated motion on factory lines. A key inflection point was the installation of Unimate on a General Motors line in 1961, widely recognized as the first industrial robot deployed in production. Early systems in this period established core principles; programmable motion, reliable actuation, and cycle-time consistency that later logistics technologies would inherit. Subsequent developments through the 1960s expanded capabilities in actuation and control, setting a technical foundation for material handling and repetitive tasks that would become central to warehousing (Hockstein et al., 2007).

Guided Vehicle Era in Material Handling (AGVs)

As factory automation scaled, Automated Guided Vehicles (AGVs) emerged as a standard solution for internal transport in manufacturing, distribution, and warehousing. Scholarly surveys of the period consolidate AGV practice around several design and control decisions: guide-path design, fleet sizing, vehicle scheduling/dispatching, idle-vehicle positioning, battery and failure management, and routing/deadlock resolution. Together, these decisions characterize the guided-infrastructure paradigm vehicles follow predefined paths under centralized control to deliver predictable throughput and safe, repeatable moves in structured layouts. This guided model became the backbone of early intralogistics mechanization and informed later transitions to more flexible systems (Le-Anh & De Koster, 2006).

An example of a system for guided vehicles in a distribution centre (published in: A review of design and control of automated guided vehicle systems) is shown in Figure 8.1 (Le-Anh & De Koster, 2006).

European Funds for Social Development
Republic of Poland
Co-funded by the European Union
NAWA POLISH NATIONAL AGENCY FOR ACADEMIC EXCHANGE

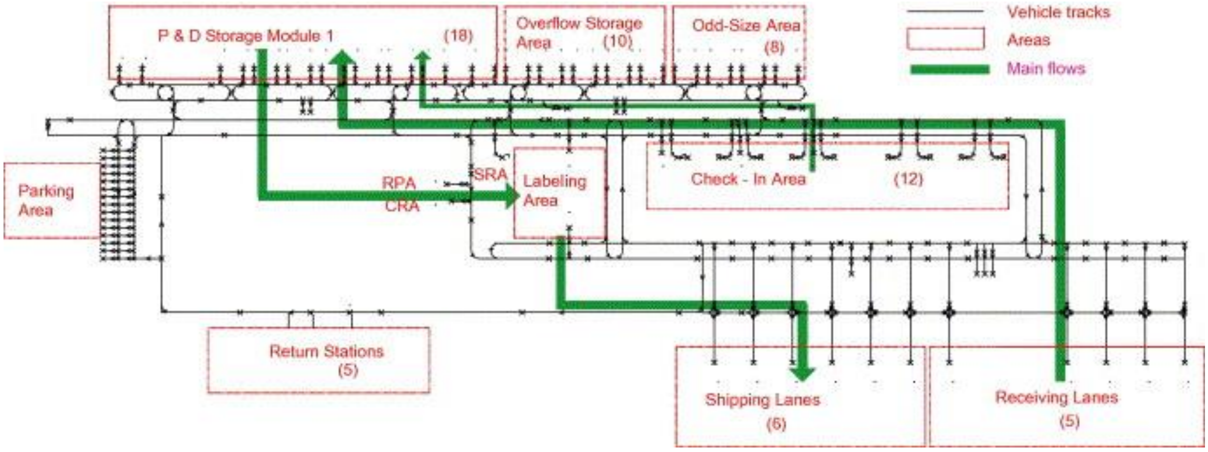

Figure 8.1 The guided-vehicle system of a distribution centre (Le-Anh & De Koster, 2006).

Robotized Warehouses and Goods-to-Person Systems

With the rise of e-commerce, warehouses adopted robotized handling systems that reconfigured order fulfilment flows. Reviews of this era document categories such as shuttle-based storage and retrieval, compact storage, and Robotic Mobile Fulfilment Systems (RMFS), in which mobile robots bring inventory racks or pods to stationary pick stations shifting from picker-to-parts to goods-to-person. These systems emphasize 24/7 operation, space efficiency, and responsiveness to demand variability, and they catalysed new planning/operations research for layout, storage assignment, order batching, and integrated control across subsystems. The robotized-warehouse paradigm thus bridged traditional AGV lines with flexible, data-driven fulfilment workflows (Azadeh et al., 2019).

Autonomous Mobile Robots for Intralogistics

The next milestone is the shift from guided vehicles to Autonomous Mobile Robots (AMRs) optimized for intralogistics. Research synthesizes AMR planning and control decisions across environment mapping, navigation and collision avoidance, task assignment, traffic management, and fleet coordination reflecting robots that operate in dynamic layouts with local sensing and decision-making rather than fixed guide paths. This transition broadens deployment options (e.g., mixed-traffic aisles, evolving slotting plans) and reframes managerial choices with frameworks that align AMR capabilities to service levels and throughput targets. In effect, AMRs generalize the mobility layer in robotized warehouses, enabling re-configurability and responsiveness under real-time conditions (Fragapane et al., 2021).

AI-Enabled Perception, Decision Support, and Warehouse Digital Twins- Present

The integration of artificial intelligence with warehouse digital twins is increasingly linked to the deployment of advanced robotics, creating tightly coupled physical-virtual systems. In these environments, mobile robots, robotic arms, and automated handling equipment act as real-time data sources, feeding operational states such as location, load status, and task completion directly into the digital twin. AI techniques process this continuous stream to forecast system behaviour, optimize task scheduling, and adapt robot actions dynamically in response to disruptions or demand changes. This bidirectional exchange enables digital twins to function not only as planning and simulation tools but also as active control layers, issuing instructions back to robots for route adjustments, workload balancing, or collaborative task execution. By merging AI-driven decision-making with robotic mobility and manipulation capabilities, warehouses can achieve higher autonomy, adaptability, and efficiency, aligning physical operations with optimized virtual models in near real time (Drissi Elbouzidi et al., 2023).

Learning-Based Robotic Manipulation for Picking and Handling (Ongoing)

Advances in learning-based robotic grasping are redefining how warehouse robots handle objects of varying shapes, sizes, and material properties. Unlike traditional rule-based approaches that rely on fixed grasp points or pre-programmed motions, modern systems integrate machine learning to continuously improve grasp planning and execution. Using visual perception, depth sensing, and tactile feedback, these robots can identify an object's geometry, estimate its pose, and select an appropriate grasp strategy in real time. Learning methods ranging from analytic models to deep neural networks enable adaptation to new items without extensive manual reprogramming, a crucial capability for fulfilment centres managing large and frequently changing assortments. Furthermore, the combination of simulation environments and sim-to-real transfer allows robots to train on vast numbers of virtual grasping scenarios before deployment, accelerating performance gains while reducing on-site trial-and-error. These developments make robotic manipulators increasingly capable of operating in unstructured warehouse settings, supporting automated picking, sorting, and packing as integral parts of the logistics workflow (Kleeberger et al., 2020).

Core Theoretical Frames

Cyber-Physical Systems (CPS) and Digital Twins

At the foundation of AI-powered logistics robotics lies the concept of *cyber-physical systems* (CPS) integrated architectures that tightly bind computational algorithms, communication networks, and physical robotic entities. In modern warehousing, robots (mobile and manipulator units) serve as physical agents, while CPS frameworks coordinate sensing, actuation, and real-time control within decentralized environments. *Digital twins* act as virtual replicas of these systems, mirroring robot state and warehouse flows to enable simulation, monitoring, and dynamic adjustment of robotic operations. This synergy empowers pre-emptive decision support, virtual testing, and proactive anomaly detection across robot fleets and material flows (Kosacka-Olejnik et al., 2021; Wu et al., 2025).

Machine Learning and Robotic Control

AI in logistics robotics leverages machine learning to enhance perception, grasping, and navigation. Sampling-based and model-free learning algorithms enable robots to generalize to varied object shapes and cluttered environments. Reinforcement learning (RL) models train robotic agents on trial-based feedback loops, enabling autonomous improvement in grasping and motion policies over time. Modern systems often apply deep learning and RL methods to multi-object manipulation, sim-to-real transfer, and real-time sensor fusion vital for autonomous operation in dynamic logistic settings (Jahanshahi and Zhu, 2024).

Human–Robot Collaboration and Interactive Robotics

As logistics environments evolve, robotics increasingly must operate alongside human workers. Theoretical frames for *human-cantered CPS* emphasize adaptability, safety, and seamless interaction. Motivated by frameworks in manufacturing, these models advocate shared workspaces where robots adjust trajectory, speed, and task allocation based on human presence and intent, mediated via sensor systems and feedback loops. This perspective supports ergonomic efficiency, conflict avoidance, and flexible cooperation in warehouse throughput (Colombathanthri et al., 2025).

Logistics Network Theory and Performance Metrics

Integrating robots into logistics and supply chains requires a solid theoretical foundation to ensure that systems are both efficient and adaptable. Mathematical principles provide the analytical tools needed to predict performance, identify bottlenecks, and optimize operations. Without these models, it is difficult to balance robot fleet size, coordinate navigation, manage queues at workstations, or forecast the impact of demand fluctuations. By formalizing relationships between input variables (e.g., number of robots, travel distances, task types) and output measures (e.g., throughput, wait time, utilization), these principles support data-driven decision-making and long-term system resilience.

Material flow in warehouses can be analysed by structuring operations into core functions; receiving, storage, order picking, and shipping and then modelling how design and control decisions (e.g., layout, storage assignment, order batching, and routing) affect flows and performance across these stages. This framework supports simulation and analytical evaluation of congestion, travel, and service trade-offs, providing a rigorous basis for configuring robotized systems to meet throughput and service targets (Gu et al., 2007).

Queuing theory is essential for evaluating how robotic systems handle tasks when multiple orders, picking stations, or transport requests compete for resources. By modelling the arrival rates of tasks and the service rates of robots, it becomes possible to predict congestion points, optimize workstation allocation, and reduce delays. In practice, queuing models are used to dimension buffer areas, evaluate service levels, and determine the optimal number of robots needed for a target throughput (Boysen et al., 2019).

Throughput analysis gauges the sustainable output of a robotized fulfilment system and links it to key design/control choices. In robotic mobile fulfilment systems (RMFS), analytical models estimate maximum order throughput, average order cycle time, and robot utilization as functions of factors such as the number of robots, number of picking stations, and warehouse layout/zoning. These results support decisions on fleet sizing and workstation configuration by revealing where capacity constraints arise and how policy changes affect performance (Lamballais et al., 2017).

Multi-Agent Systems and Swarm-Inspired Coordination

An essential theoretical lens for robot fleets involves multi-agent systems (MAS) and swarm intelligence, where multiple robots operate under decentralized control to collectively achieve shared objectives, such as collision-free movement, balanced task distribution, and responsive scalability. Swarm-inspired models emphasize robustness, emergent behaviour, and adaptability robots rely on local cues (like congestion or task signals) and simple interaction rules rather than central oversight. This yields flexible, fault-tolerant logistics systems capable of dynamic response within dense warehouse environments.

A comprehensive review of swarm intelligence in multi-robot systems highlights how biological principles such as those observed in ant or bee colonies inform control algorithms that enhance efficiency, robustness, and scalability. This review frames the theoretical underpinnings of MAS coordination and its promising applications to logistics and warehousing robotics (Nguyen, 2024).

Simulation, Digital Twins, and Learning Loops

Finally, a theoretical pillar combines simulation-driven learning and digital twins for iterative improvement. Robots perform tasks within simulated environments (the digital twin), enabling policy testing, fault simulation, and control refinement before deployment. AI-supported feedback loops compare digital predictions with physical performance, continually tuning control algorithms and updating the twin. This convergent physical-virtual loop underpins safe, scalable learning in logistics robotics (Guerra-Zubiaga et al., 2025).

Technology Infrastructure in AI-Enabled Robots

The technology infrastructure of AI-enabled robots in logistics consists of two interdependent layers: *hardware* and *software*. Hardware provides the physical capacity for movement, perception, and manipulation, while software translates data into decisions and actions. Artificial intelligence integrates these layers, enabling autonomous behaviour, adaptability, and continuous performance improvement in dynamic logistics environments (Fragapane et al., 2021).

Hardware Components

The *locomotion system* forms the foundation of mobility. AMRs and AGVs may use differential drive, omnidirectional wheels, or track systems, chosen based on manoeuvrability requirements, load capacity, and flooring conditions. The *sensing suite* is central to environmental perception, typically combining LiDAR for mapping and obstacle detection, RGB or depth cameras for object recognition, ultrasonic sensors for proximity detection, and IMUs for motion tracking. These sensors operate in fusion to create an accurate and dynamic map of the environment (Fragapane et al., 2021; Liu et al., 2024).

The *manipulation system*, where present, consists of robotic arms or grippers designed to handle specific types of items ranging from standardized totes to irregularly shaped parcels. Gripper design is critical for handling delicate or varied objects and may include parallel, vacuum, or adaptive mechanisms (Kleeberger et al., 2020). *Power infrastructure* is equally important, with robots typically powered by lithium-ion batteries supported by charging stations or inductive charging pads to ensure continuous operation in multi-shift warehouses. The *on-board computing unit* processes sensor data and executes control commands in real time, enabling immediate response to environmental changes (Fragapane et al., 2021).

Software Components

The software layer governs perception, navigation, planning, and task execution. *Perception modules* process raw sensor data, performing tasks such as feature extraction, object recognition, and environment mapping. Algorithms in simultaneous localization and mapping (SLAM) allow robots to continuously update their position and generate navigable maps (Cadena et al., 2016). *Planning and* *control systems* determine how robots move and execute tasks. Path-planning algorithms calculate collision-free routes, while motion control ensures precise speed, acceleration, and turning in line with operational constraints. *Task management software* integrates with warehouse management systems (WMS), allocating tasks to robots based on priorities, location, and workload. Middleware such as ROS (Robot Operating System) facilitates communication between hardware components and higher-level applications, enabling modularity, scalability, and easier integration of AI functions (Fragapane et al., 2021).

Artificial Intelligence Integration

Artificial intelligence enhances both hardware and software capabilities. In hardware, AI improves perception by enhancing object detection, semantic segmentation, and multi-sensor data fusion. In software, AI enables predictive navigation, adaptive routing, and task allocation based on real-time demand and environmental changes. Machine learning models can forecast congestion zones, adjust robot deployment dynamically, and optimize energy consumption. Reinforcement learning supports continuous improvement by allowing robots to learn from operational feedback and adapt to new layouts or workflows without extensive reprogramming (Fragapane et al., 2021; Kleeberger et al., 2020).

Use Cases for AI-Enabled Robots

Amazon Robotics

Amazon reports over 1 million robots deployed across its operations, including *Hercules* robots that move inventory pods, *Pegasus* units that convey individual packages, and *Proteus*, the company's first fully autonomous mobile robot that can safely navigate in open areas around employees. Amazon also describes launching a new AI foundation model to power its robotics fleet, supporting safer navigation, tasking, and efficiency in fulfilment workflows (Dresser, 2025).

Ocado – "Second-Hands" Collaborative Robot

Ocado Technology coordinated the Second-Hands project to develop an autonomous humanoid collaborative robot that uses artificial intelligence, machine learning, and advanced 3D vision to perceive human technicians' needs and proactively assist with maintenance tasks. The robot is designed to recognize when help is required, hand tools, and manipulate objects such as ladders, pneumatic cylinders, and bolts, aiming to increase safety, efficiency, and productivity in the workplace through natural human–robot collaboration (Ocado Technology, 2015).

DHL – Computer Vision for Logistics Robotics

DHL is applying computer vision to enhance the performance of AI-enabled robotics in its logistics operations. By combining cameras, sensors, and AI algorithms, robots can identify, classify, and track packages with high accuracy in real time. This capability enables automated sortation, improved quality control, and faster exception handling within warehouses. Computer vision systems also allow robots to detect irregularities, read labels in varied conditions, and operate effectively alongside human workers, creating a more efficient and reliable supply chain (Trend report, 2025).

Future Trends in the Use of AI in Enabled Robots

The future of AI-enabled robots in logistics and supply chain management is shaped by emerging trends that enhance autonomy, coordination, and adaptability. Two interrelated directions; multi-robot intelligence via AI-enhanced task allocation, and workflow learning through transformer and graph neural network (GNN) integration are especially promising for advancing operational efficiency and system robustness.

AI-Driven Task Allocation in Robot Fleets

As autonomous mobile robots (AMRs) become more prevalent, efficient coordination across fleets becomes vital. AI-based task allocation methods are emerging to assign tasks dynamically, optimizing both energy consumption and the number of robots needed. These models use machine learning techniques to adapt to evolving workloads, fleet performance, and environmental constraints, ensuring sustainable and cost-effective deployment in logistics operations (Valenzuela & Noguera, 2025).

Transformers & GNNs for Enhanced Route Optimization

Another key innovation is the integration of advanced AI architectures specifically, *Transformer models* and *Graph Neural Networks (GNNs)* for route optimization. By representing warehouse layouts and robot tasks as graph structures and using attention-driven Transformer layers to interpret them, these methods significantly reduce travel distances and time, while improving energy efficiency. Early experiments report improvements of approximately 15–20% in path efficiency and 10% in energy savings, signalling a powerful tool for smarter robotic movement (Luo et al., 2025).

Looking Ahead

Together, these trends point toward a future where AI not only facilitates autonomy but also strategic, performance-driven coordination across entire robot fleets. Dynamic task allocation ensures minimal idle time and energy use, while AI-based route planning refines mobility with precision and context-awareness. As these technologies mature and integrate with real-world systems, logistics environments can expect smarter, more flexible, and efficient robotic operations that continuously learn and adapt to evolving demands.

Conclusion

This chapter provides a comprehensive overview of AI-enabled robots, detailing their definitions, classifications, historical development, theoretical frameworks, technological infrastructure, and real-world applications. It begins by explaining that AI-enabled robots integrate mechanical systems with artificial intelligence, using technologies like machine learning, computer vision, and natural language processing to operate autonomously in dynamic environments. In logistics, they are categorized into Automated Guided Vehicles (AGVs), Autonomous Mobile Robots (AMRs), and robotic arms, each contributing to efficiency, safety, and flexibility. The historical review traces developments from early industrial robots like Unimate in the 1960s, through AGV-based material handling, robotized warehouses with goods-to-person systems, the emergence of AMRs, and the integration of AI with warehouse digital twins for real-time optimization. Learning-based robotic manipulation is highlighted as an ongoing advancement in handling diverse objects autonomously.

The chapter also outlines key theoretical frameworks underpinning AI-enabled robotics, such as Cyber-Physical Systems, machine learning-driven robotic control, human–robot collaboration models, and mathematical principles including material flow theory, queuing theory, throughput analysis, and multi-agent systems. It emphasizes technology infrastructure in terms of hardware (locomotion, sensing, manipulation, power, on-board computing) and software (perception, SLAM, planning, control, task management), with AI enhancing both layers. Real-world use cases include Amazon Robotics' large-scale AMR deployment, Ocado's Second-Hands collaborative robot, and DHL's computer vision-assisted warehouse robots. The closing section discusses future trends such as AI-based task allocation in robot fleets and advanced route optimization using Transformers and Graph Neural Networks, pointing toward increasingly adaptive, efficient, and autonomous logistics operations.

Tasks

1. Based on the definitions and descriptions in the text, prepare a comparative table of automated guided vehicles (AGVs) and autonomous mobile robots (AMRs). List at least five

different criteria. Briefly explain which type would be more suitable for a warehouse whose layout changes frequently, and why?

2. Identify the basic technological elements of robots powered by artificial intelligence.
3. Find another example of the use of artificial intelligence in robots used in logistics or supply chains. Describe: the purpose of use, the technology, the impact on processes.

9. ARTIFICIAL INTELLIGENCE IN AUTONOMOUS VEHICLES IN TRANSPORT

Autonomous vehicles (AV) in logistics and transport are vehicles guided by an on-board AI "brain" that constantly looks at the surroundings, makes sense of what it sees, predicts what nearby people and machines might do, chooses a safe and efficient path, and then smoothly steers, accelerates, and brakes. Crucially, "autonomous" is scoped: systems are approved for specific roads, yards, speeds, weather, or indoor areas an operational design domain (ODD) and automation levels describe how much of the driving task the system takes over and how much a human must supervise. In real operations this spans highway tractors on predictable lanes, self-driving yard tractors and terminal tractors inside fenced areas, automated guided vehicles (AGVs) that shuttle containers between quay and yard, driverless shuttles in campuses, rail systems with automatic train operation, and autonomous mobile robots (AMRs) that move totes and pallets between dock doors and staging points. The common thread is fit-for-purpose autonomy that improves safety and reliability while staying within clearly defined bounds (Fragapane et al., 2021).

General Principle

An AV runs a fast loop of *see → understand → predict → plan → act*. Cameras, radar, lidar, GNSS/IMU, and wheel encoders provide complementary "eyes" for daylight, darkness, rain, and tight indoor aisles. Perception software turns raw signals into lane markings, traffic lights, signs, forklifts, pedestrians, and other vehicles with positions and speeds. Localization keeps the vehicle exactly where it thinks it is, either by aligning to high-definition maps (lanes, crossings, speed limits, landmarks) or by rebuilding a local map on the fly in warehouses and terminals. Prediction looks a few seconds ahead who will turn, cross, reverse, or merge and attaches probabilities, so the planner can keep safe margins. Planning and control then generate a smooth, rule-respecting path that the actuators can really follow, with monitors and fall backs (slow down, pull over, stop) if anything looks off. The same loop applies indoors and outdoors: AMRs share the *predict–plan–act* logic with highway tractors and port AGVs, just with different speeds, sensors, and rules (Fragapane et al., 2021; Sánchez-Ibáñez et al., 2021).

Concrete logistics applications include the following. On highways, autonomous trucking and truck platooning maintain short, coordinated gaps to reduce aerodynamic drag and fuel consumption while meeting time windows. In terminals, automated guided vehicles and electric yard tractors move containers between quay cranes and stacks, with scheduling that accounts for battery levels, crane workloads, and uncertain arrivals. In rail, automatic train operation and eco-driving compute energy-saving speed profiles while maintaining punctuality on congested corridors. These are operational systems, not lab prototypes: fuel and energy benefits have been measured for platooning and train eco-driving, and modern AGV dispatchers co-optimise routes and battery swaps to shorten vessel turnaround and reduce truck dwell time (Sun et al., 2021; Li et al., 2023; Fernández-Rodríguez et al., 2020).

Advantages include fewer human-factor errors in repetitive, fatigue-prone tasks; smoother traffic and yard flows that cut idling, braking, and energy use; tighter and more reliable schedules on fixed lanes and fenced sites; and better utilization of scarce assets such as tractors, cranes, and train paths. Achieving those gains depends on disciplined scoping of the ODD, staged validation and clear handover rules between automation and human oversight, and integration with day-to-day systems and terminal operating systems so that predictions trigger real dispatch and scheduling decisions. In other words, AI-driven autonomy delivers value when it is engineered as part of the logistics system, not as a gadget bolted onto it. (ITF, 2021).

Advantages and Disadvantages of AI-Driven Autonomy in Transport

When the autonomous driving stack is integrated end to end, with clean data as input, reliable perception and prediction, rule aware planning, and robust control as output, the benefits for transport systems are tangible. Safety can improve because automation removes major human error pathways (distraction, fatigue, inattention) and reacts consistently to hazards. Efficiency can rise as vehicles keep smoother headways, avoid unnecessary stops, and coordinate merges. Logistics providers gain more reliable schedules and higher asset utilisation on predictable lanes and fenced sites, and passengers who cannot drive gain dependable access. These gains do not come for free. Organisations realise them when autonomy is scoped to an appropriate operational design domain (for example specific roads, yards, and weather conditions), connected to operations platforms such as transport management systems (TMS), yard management systems (YMS), and terminal systems, and backed by disciplined validation so that performance holds across locations, seasons, and traffic cultures. Evidence in the literature points to potential reductions in crashes and delays, increased effective road capacity, and system level cost and energy benefits under the right deployment conditions (Fagnant and Kockelman, 2015).

Safety is the headline promise, but it depends as much on engineering assurance as on AI accuracy. Well-designed autonomous systems combine redundancy (sensors, compute, power), health monitoring, conservative fall backs (slow, pull over, stop), and clear human oversight policies with staged testing from simulation to test track to on road pilots. Demonstrating “better than human” performance is a statistical challenge because serious crashes are rare events; therefore, safety cases combine coverage of hazardous scenarios, fault tolerance arguments, and evidence that the system remains within its defined operational design domain. In practice, safety benefits materialise when capabilities and context are aligned, while risks grow when systems are pushed outside their domain (for example, poor weather or ambiguous road works) or when mixed fleets create unexpected interactions. The conclusion is two sided: autonomy can reduce exposure to human error, but only with systematic validation and runtime governance (Milakis et al., 2018; Koopman & Wagner, 2016).

Network efficiency and energy outcomes can improve when automated vehicles damp stop and go waves and maintain stable car following gaps. Modelling and experiments show that even partial penetration of connected and autonomous vehicles can improve string stability and increase throughput, leading to fewer shockwaves, shorter travel times, and smoother traffic. These dynamics, together with gentler acceleration and braking, translate into real fuel and energy savings, especially on freeways and freight corridors. However, system level energy use also depends on behaviour outside the vehicle: if automation lowers the generalised cost of travel (comfort, time, attention), total vehicle kilometres travelled may rise, offsetting some efficiency gains. Both effects can be true at once; deployment choices such as pricing, lane management, and ride pooling determine which dominates in practice (Talebpour & Mahmassani, 2016; Wadud et al., 2016).

In freight logistics, advantages appear in day to day operations. Truck platooning reduces aerodynamic drag through tightly coordinated gaps, lowering fuel use and emissions while maintaining safety through automated control. Planning tools now co optimise which trucks to pair, when to pair them, and on which links, given schedules and regulations. In and around ports and terminals, autonomous yard tractors and automated guided vehicles (AGVs) move containers predictably, supporting tighter quay crane and gate schedules. When dispatchers account for battery state, crane workloads, and uncertain arrivals, terminals shorten vessel turnaround and truck dwell. Across these use cases, the common thread is higher asset productivity: more moves per hour, less idling, and better on time performance with the same physical infrastructure (Zhang et al., 2020; Norozoliaee et al., 2020; Bhoopalam et al., 2018).

Accessibility is another concrete advantage. By removing the need to drive, autonomous services can expand mobility for older adults and people with travel-restrictive medical conditions, enabling more trips to work, healthcare, and social activities. Studies estimate sizable latent demand among non-drivers that autonomy could unlock, which is both a social benefit and an operational planning signal: fleets and cities should expect different travel patterns (times of day, trip lengths) when access barriers fall. Designing vehicles and services for accessibility (step-free entry, securement, human assistance protocols) is therefore central to realizing these benefits safely and equitably (Harper et al., 2016; Kassens-Noor et al., 2021).

Balanced against these upsides are important disadvantages and risks that policy and design must address. On the demand side, automation can induce extra travel and empty repositioning, which may increase congestion and emissions unless safeguards are in place; travel can also shift away from public transport if door to door service becomes cheaper or more convenient. On the safety and governance side, rare edge cases, data drift, and unclear human machine handovers can undermine reliability unless systems are tightly scoped and continuously monitored. On the security side, increased connectivity expands the attack surface, from sensor spoofing and GPS interference to vehicle to everything (V2X) message manipulation, making cybersecurity a first order safety concern rather than an afterthought. These drawbacks are manageable, but they are real. Meeting them requires pricing and curb policies to limit rebound travel, formal safety assurance processes, and defence in depth across the automated vehicle (AV) stack and its interfaces (Wadud et al., 2016; Petit and Shladover, 2015; Milakis et al., 2017).

Finally, trustworthy deployment requires evidence at scale. The field now combines scenario-based simulation, hardware-in-the-loop, closed-course testing, and staged public pilots to accumulate exposure to rare but critical situations, assurance arguments weave together metrics, coverage, and defect discovery rates across perception, prediction, planning, and control. Standards and guidance connect AI development to transport safety practice, linking data management, model risk controls, and runtime monitors to the legal and societal expectations placed on vehicles that operate among people (Tang et al., 2023; Expósito Jiménez et al., 2024).

Use Cases

Automated container terminals (ACTs) with autonomous yard transport. Modern seaports operate fleets of automated guided vehicles and electric yard tractors that shuttle containers between quay cranes and yard blocks under a terminal operating system (TOS). Recent studies show that AI based scheduling and speed control can reduce delays and emissions by jointly assigning container moves and battery swap stops, modelling congestion at swap stations, and varying AGV speeds across the apron, transfer lanes, and narrow yard aisles. These models reflect layouts now in service, including Tuas in Singapore and Haifa in Israel, and report measurable gains in vessel turnaround time and CO_2 when dispatchers co-optimise moves and energy (Yang et al., 2023; Zheng et al., 2022).

European Funds for Social Development
Republic of Poland
Co-funded by the European Union
NAWA POLISH NATIONAL AGENCY FOR ACADEMIC EXCHANGE

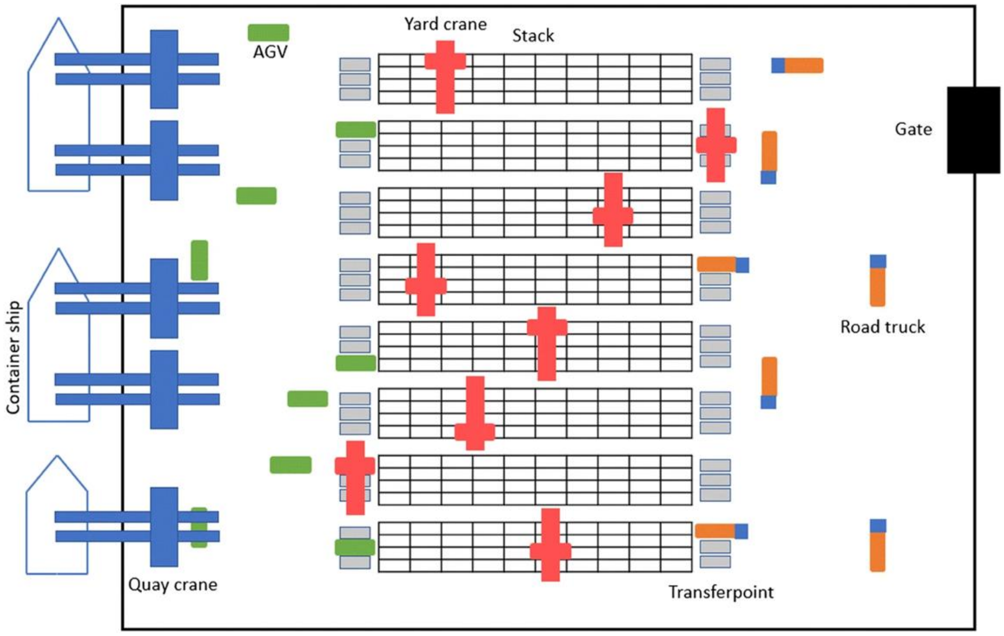

Figure 9.1 AGV system architecture in a container terminal (Jonker et al., 2021).

This figure is a schematic of autonomous operations in a modern automated container terminal. On the waterside, quay cranes (QCs) load and unload containers. Automated guided vehicles (AGVs), shown in green, travel on fixed lanes to yard transfer points. Yard cranes (YCs), shown in red, stack and retrieve containers, and road trucks interface at the gate. The paper explains that terminal performance depends on coordinating QCs, AGVs, and YCs together rather than optimising each in isolation. It formulates a hybrid flow shop scheduling model that captures realistic constraints such as bidirectional flows and twin lift moves treated as paired jobs, and solves it with a simulated annealing metaheuristic tuned for near real time re-planning. In practice, the AI and operations research scheduler decides which AGV serves which crane, in what sequence, and when, so that cranes do not wait idly and vehicles do not queue at transfer pads, increasing quay productivity and shortening vessel turn times (Jonker et al., 2021).

Truck platooning and learning-augmented convoy planning

Freight corridors are trialling AI coordinated platoons, where trucks travel with short, automated gaps to reduce aerodynamic drag and fuel consumption while maintaining safety through cooperative control. Evidence from field experiments and network level planning studies points to two factors behind the benefits: (1) on road control that stabilises car following and damps stop and go waves, and (2) itinerary optimisation that decides which trucks should meet, where, and when to form fuel saving platoons without violating time windows. Together these approaches show fuel and throughput improvements, subject to operational constraints such as driver rules, weather, and pairing logistics (Stern et al., 2018; Abdolmaleki et al., 2021).

Automatic Train Operation (ATO) and eco-driving in rail

Passenger and freight railways use AI-backed ATO to compute energy-efficient speed profiles that respect gradients, signal blocks, dwell variability, and regenerative braking while holding punctuality. Recent formulations solve the driving problem on real networks (including high-speed lines), showing how discrete, rule-aware optimization trims traction energy without reducing headways; many systems are already deployed on metros and are being extended to mainline corridors with strong energy and stability results (Wei et al., 2022).

Autonomous shuttles in mixed traffic

Cities and universities have piloted low speed, geo-fenced driverless shuttles to serve first and last mile links. A recent United States campus pilot completed hundreds of trips resembling revenue service on a short, mixed traffic route and sets out the operational approach: a conservative operational design domain (speed, route, conditions), remote support, staged testing, and clear incident procedures. Broader reviews synthesise lessons from multi-site European pilots, including service design, safety, user acceptance, and integration with public transport. These deployments show stable operations today when routes, speeds, and supervision are tightly specified (Golchin et al., 2024; Nesheli et al., 2021).

Road traffic smoothing with a small share of automated vehicles

Controlled field trials show that even a single well behaved automated vehicle in mixed traffic can damp stop and go waves, the self-propagating cycles of braking and acceleration that travel upstream through traffic, cutting fuel use and braking events for the vehicles behind it. This measured effect, not just a simulation result, motivates corridor policies and connected vehicle controls designed to damp disturbances and recover capacity on busy freeways (Stern et al., 2018).

Conclusion

AI driven autonomy matters in transport because it can convert messy, fast changing signals into safe, repeatable motion within clearly defined operating bounds. When the full stack of perception, prediction, planning, and control is integrated into day to day operations, for example transport management systems (TMS), yard management systems (YMS), terminal operating systems (TOS), and automatic train supervision (ATS), and the deployment is scoped to a suitable operational design domain (ODD), the gains can be concrete: fewer human factor incidents, smoother traffic and yard flows, tighter schedules on predictable lanes, and better utilisation of expensive assets across roads, rail, ports, and fenced logistics sites. In short, autonomy becomes a systems tool, not a gadget; it raises service quality while supporting cost discipline and lower energy use.

Those benefits appear only under disciplined engineering. Success depends on clean, well governed data; rigorous validation that combines simulation, closed course tests, and staged on road pilots; explicit ODD definitions (where, when, and with which other road users the vehicle may operate) and automated checks that keep operations within those bounds; human in the loop oversight with clear handoff rules; and explainable decisions so incidents can be understood and prevented. Autonomy must be integrated with the rest of the transport system, including dispatch, maintenance, energy management, and customer promises, so predictions trigger real, auditable actions rather than dashboards that go unused.

Caution is essential on three fronts. First, safety assurance: rare edge cases, sensor faults, and model drift will occur, so redundancy, runtime monitoring, graceful fall backs, and post incident learning loops are essential. Second, system effects: automation can lower the perceived cost of travel, potentially increasing vehicle kilometres through induced demand or empty repositioning; without policy guardrails such as pricing, pooling, and lane management, network energy and congestion gains can be diluted. Third, security and trust: connected automated vehicles expand the attack surface, including in vehicle networks, V2X links, and cloud back ends, and they handle sensitive data; defence in depth, privacy by design, and transparent accountability are as central to safety as perception accuracy.

The practical takeaway for operators and public agencies is a phased, evidence led roadmap: start where the environment is controllable and payoffs are clear (highway line haul, terminals, yard and port movements, shuttle corridors); measure what matters (on time performance, dwell, incident rates, and energy or CO_2 per move); invest in people, processes, and machine learning operations (MLOps) as much as in sensors; and couple deployment with policy that steers benefits toward public goals. Done this way, AI enabled autonomy is not an end in itself but a dependable lever to make transport safer, cleaner, and more reliable at the scale that real logistics requires.

Tasks

1. Explain at least three advantages and three disadvantages of AI-driven autonomy in transport logistics.
2. Describe one case study of autonomous vehicles with integrated AI in transport.

10. VISUAL ARTIFICIAL INTELLIGENCE AND DEEP LEARNING TECHNOLOGY FOR OBJECT RECOGNITION

Visual AI for object recognition refers to deep learning methods that convert raw images and video from cameras mounted on vehicles, equipment, facilities, or drones into structured facts about the physical world, such as "person entering a crossing," "vehicle or asset identifier recognised," "object present at a designated location," or "unauthorised entry into a protected area." In logistics and transport, these machine-readable facts feed dispatch systems, traffic control centres, terminal operating systems, and safety interlocks so that identification, counting, inspection, and alerts use consistent criteria with recorded outcomes. Deep learning displaced rule-based vision because learned detectors adapt better to lighting changes, cluttered backgrounds, and nonstandard viewpoints, which is why video analytics now supports roadway monitoring and asset identity or condition verification in ports, yards, and rail terminals (Ceccarelli et al., 2023; Boukerche and Hou, 2021).

At a high level, Visual Artificial Intelligence and deep learning for object recognition combine cameras, edge computing, trained models, and operational software to automate routine visual checks and trigger actions in near real time. The promise is consistent decisions at speed and at scale, with traceable outcomes; the practical considerations are data quality, camera placement, lighting and weather, latency, privacy and governance, and ongoing monitoring and retraining. The right design depends on the site, the risks, and the required response times.

Deep learning in this context follows a training and inference workflow. During training, detectors optimise two coupled objectives: classification (assigning an object category) and localisation (regressing bounding boxes), typically with multi-scale feature extraction so that small distant objects and large nearby objects are both represented. Architectures are organised into a backbone (feature extractor), a neck (feature aggregation across scales), and a head (outputs of boxes, classes, and confidence). Two-stage families often prioritise precision by generating candidate regions before classification, while one-stage families often prioritise speed by predicting boxes and classes in a single pass; transformer-based detectors add global attention for longer-range context. At inference, the trained network processes frames on edge hardware to produce detections in milliseconds, and non-maximum suppression or learned de-duplication removes overlaps. These design choices are standard across surveys of modern object detection and matter in transport because they determine whether recognition can meet latency budgets at gates, intersections, and on moving platforms (Zaidi et al., 2022; Mittal, 2024).

Operational systems extend detection with tracking and an application layer. A multi-object tracker associates detections over time and across cameras to maintain stable identities for counting, dwell estimation, and "who went where" analytics; re-identification features help maintain identities when objects leave and re-enter the field of view. Application logic turns recognised events into actions and records, for example opening a barrier when an asset ID matches an order, dispatching an AGV when a crane completes a lift, or issuing a near-miss alert when a vulnerable road user enters a conflict zone. To achieve low latency and reduce backhaul traffic, inference is deployed at the edge on ruggedized processors mounted near cameras or on vehicles; only compact event messages or anonymised metadata flow to central systems. Maintaining performance after deployment requires MLOps practices: health monitoring, drift detection, periodic relabelling of hard examples, and controlled retraining so models remain calibrated as seasons, camera positions, or site layouts change. These edge and lifecycle patterns are emphasised in intelligent-transport surveys of edge intelligence and are directly applicable to transport operations (Gong et al., 2023).

Robustness in real sites depends on sensor fusion and calibration. Camera-only systems degrade in rain, fog, glare, and low light; fusing cameras with lidar or radar combines complementary geometry and velocity cues with appearance cues so that detection and tracking remain stable in adverse conditions and under partial occlusion. Fusion can be implemented early (feature-level), late (decision-level), or in learned cross-modal attention blocks; all approaches require precise time synchronization and extrinsic calibration so that sensor measurements align in space and time. Reviews of multi-modality object detection for autonomous driving document consistent improvements from lidar–camera and radar–camera fusion, which carry over to mixed indoor–outdoor logistics sites where lighting and backgrounds vary strongly between apron, yard aisles, and gate canopies (Tang et al., 2023; Yeong et al., 2021).

Model transfer across sites and cameras is a recurring challenge. Detectors trained in one domain (daylight, specific cameras, specific signage or paint) often underperform in another domain with different textures, weather, or viewpoints, which is known as domain shift. Two families of methods address this problem in transport deployments: (1) transfer learning, which fine-tunes pre-trained detectors on a small, site-specific labelled dataset; and (2) domain adaptation, which aligns features between source and target domains to reduce performance loss when labels are limited. Comprehensive studies in Computer Vision and Image Understanding report that multi-level adversarial alignment and consistency regularization can recover much of the lost accuracy without full re-labelling. Where rare conditions or defects are under-represented, synthetic data augments training; for example, in ports, synthetic labelled images of containers with varied paints, corrosion, and lighting improve defect and code-spotting performance before on-site fine-tuning with real samples (Marnissi et al., 2023; Delgado et al., 2023).

On public roads, visual AI detects and tracks vehicles, pedestrians, cyclists, buses, and roadway elements for incident detection, near-miss analysis, sign and signal understanding, and traffic performance measurement. Long-horizon monitoring programs integrate edge detectors and trackers to turn continuous video into counts, speeds, queue lengths, and surrogate safety indicators that support evaluation of street changes and tuning of signal timing or priority measures. Integrating structured video analytics into operations decisions has been documented in transportation research and demonstrates that deep learning can operate at city scale with stable accuracy when models are maintained and evaluated against ground truth (Ceccarelli et al., 2023).

At gates, yards, ports, and rail terminals, object recognition supports identity and condition recognition under non-ideal imaging. A common workflow is container code reading: a detector first isolates the container area despite dirt, dents, skew, and motion blur; then an OCR module reads the BIC/ISO code with sequence modelling and quality control; finally, the event is reconciled against a job in the terminal system. Recent work shows that line-segmentation masks and improved sequence modelling increase read rates under real gate conditions, reducing manual re-keying and dwell. A related and mature application is automatic number-plate recognition for trucks and trailers; comparative studies summarize performance trade-offs across illumination and camera placements and motivate the use of infrared illumination and robust detection pipelines to maintain accuracy in night operations. Together, these systems shorten gate cycles and standardize inspection quality, and they are now routine building blocks in automated terminal processes (Zhang et al., 2024; Lubna et al., 2021).

Advantages in logistics and transport are measurable when recognition outputs are integrated into operational systems. Continuous, non-intrusive sensing turns existing cameras into telemetry for throughput, dwell, and safety indicators; automated identification reduces transcription errors and manual handling; synchronized recognition across cranes, AGVs, trucks, and stacks improves hand-off reliability; and edge processing reduces latency and network load. Surveys of edge intelligence for ITS and lightweight detection on edge devices describe design patterns; model quantization, pruning, and compact backbones that meet latency and power constraints without unacceptable loss of accuracy, which is necessary for smart poles, vehicle computers, and gate cabinets (Gong et al., 2023; Mittal, 2024).

Limitations

There are limitations that require explicit management. Domain shift can lower accuracy after camera moves or seasonal changes; rare events are under-represented, which increases miss rates for important but infrequent hazards; weather and occlusion reduce image quality; and small or distant targets remain challenging. Addressing these limitations involves site-specific fine-tuning with targeted data collection, fusion where camera-only performance is insufficient, and disciplined model lifecycle management with monitoring and controlled re-training. In addition, privacy and governance must be handled when faces and plates enter the frame; storing only event metadata or applying on-edge redaction are established practices. These issues are widely reported in transportation video analytics and domain-adaptation research and should be accounted for in deployment plans and KPIs (Gong et al., 2023; Marnissi et al., 2023).

Deep Learning in Visual AI

Visual AI for object recognition uses deep learning to turn camera streams on vehicles, cranes, gates, poles, and drones into structured data that transport systems can act on. In road traffic it supports detection and tracking of vehicles, pedestrians, cyclists, buses, signs, and signals; in logistics facilities it supports identification and condition assessment of cargo units and vehicles. The main benefit is replacing manual observation and rule-based vision with learned models that sustain accuracy under changing lighting, clutter, and viewpoints, so that counts, violations, identities, and safety events feed operational tools such as traffic control centres and terminal operating systems. A recent journal survey focused on traffic video analytics outlines these roles and documents why object detection has become the primary perception step in intelligent transportation workflows (Ghahremannezhadm et al., 2023).

During training, detectors optimize classification and localization objectives using a backbone–neck–head architecture; one-stage designs prioritize speed for real-time control, while two-stage detectors emphasize precision, and transformer-based detectors add global context. At inference, multi-object tracking maintains identities across frames and cameras for dwell and flow metrics, and application logic converts events into actions (alerts, dispatch decisions, barrier control). Edge deployment keeps latency low and reduces backhaul, while model lifecycle processes (monitoring, data selection, re-labelling, controlled retraining) maintain accuracy as seasons and layouts change. Edge intelligence surveys in transportation and traffic-video detection surveys describe these patterns and show how they are applied at scale (Ghahremannezhad et al., 2023; Ristić-Durrant et al., 2021).

Two techniques support robustness across transport settings: sensor fusion and transfer learning. Lidar or camera-radar fusion combines appearance, geometric, and velocity cues for stable detection in rain, fog, at night, and under partial occlusion; careful spatial and temporal calibration is required so fused features align. Transfer learning and domain adaptation fine-tune pre-trained detectors on site-specific data to counter domain shift caused by different camera models, lighting, markings and signage, or backgrounds, and synthetic data are used to cover rare defects or conditions before deployment. Reviews of multimodal perception and domain adaptation for intelligent vehicles report consistent accuracy gains from fusion and adaptation, and these gains carry into mixed indoor and outdoor logistics settings such as aprons, yard aisles, and gate canopies (Wang et al., 2023; Delgado et al., 2023).

Use Cases in Transport and Logistics

Road network operations: traffic surveillance, safety analytics, and control

Visual AI detects and tracks road users in live video to produce counts, speeds, queue lengths, lane occupancy, and surrogate safety indicators such as near misses and conflicts. These outputs support incident detection, protection of vulnerable road users, sign and signal understanding, and data driven policies such as adaptive signal timing or bus priority. A recent survey reports that modern deep detectors and multi object trackers can meet city scale accuracy and throughput targets when models are monitored for drift and maintained over time, enabling more reliable and responsive monitoring. The same review explains that surrogate safety is typically computed from tracked trajectories and highlights practical challenges such as occlusion and changing illumination that should be considered when deploying at scale (Ghahremannezhad et al., 2023).

Gate automation: container ID and plate recognition

At truck and rail gates, object detection localizes the container surface or license plate; an OCR module then reads the code under dirt, dents, skew, and motion blur; application logic reconciles the ID with a job order and opens the barrier. Recent work integrates a two-stage pipeline (detector + sequence recognizer) tailored for port conditions and reports higher read rates at realistic gate viewpoints, while ANPR studies document real-time, edge-inference pipelines that maintain performance under low resolution and noise. These functions shorten gate cycles and reduce manual re-keying (Yu et al., 2024; Ammar et al., 2023).

Cargo condition assessment: container damage detection

Visual AI extends beyond identity to condition. Transfer learning models with lightweight backbones can detect multiple damage types on container surfaces, for example dents and corrosion, and synthetic data sets are used to cover rare defect cases before on site fine tuning. Studies report that adding synthetic images and class balancing improves generalisation and reduces false negatives in port environments, supporting more consistent inspection outcomes with traceable decisions (Wang et al., 2023; Delgado et al., 2023).

Rail transport: on-board obstacle detection and scene understanding

In rail, on-board cameras with deep detectors monitor the track ahead to identify obstacles, intrusions, and unexpected objects at level crossings and on open track. A study presents a deep learning method tailored to rail scenes and shows improved detection in defined regions of interest, while a broader review summarizes vision-based railway obstacle detection and distance estimation, motivations, and open issues such as adverse weather and long-range small targets. These systems underpin warnings and automatic braking assistance in rail safety workflows (He et al., 2021; Ristić-Durrant et al., 2021).

Air transport: foreign object debris (FOD) detection on runways and aprons

Runway and apron safety programmes deploy fixed or vehicle-mounted cameras with deep learning detectors to identify foreign object debris (FOD) that threatens operations. Recent surveys and studies on AI-based FOD detection note that visual methods can reduce reliance on manual inspection and improve detection of small, low-contrast objects when supported by appropriate data and model selection. These systems generate actionable events for airside operations and maintenance teams (Shan et al., 2025; Mo et al., 2024).

Intermodal yards and terminals: flow monitoring and hand-off verification

Infrastructure cameras at transfer pads, apron lanes, and yard aisles detect vehicles and equipment, verify handoffs between cranes, automated guided vehicles (AGVs), and trucks, and measure dwell and queue times. When combined with identification and optical character recognition (OCR) events, multi object tracking enables “who went where” analytics across gate → yard → dock, which terminal systems use to adjust sequencing and allocate resources. Surveys of transport video analytics and domain specific visual inspection document these applications and the associated integration patterns with terminal operating systems (Ghahremannezhad et al., 2023; Delgado et al., 2023).

These use cases illustrate a common pattern: visual AI produces timely, auditable events that downstream systems can optimize against. On roads, this means more reliable detection for safety and control; at gates and terminals, this means faster identification and standardized inspection; in rail and aviation, this means earlier hazard detection with defined response workflows. The literature shows that when recognition outputs are connected to operations platforms and maintained through edge deployment and model-lifecycle controls, organizations can set KPIs (throughput, dwell, incident rate) and measure impact, while managing known limitations such as domain shift, small or distant targets, and adverse weather through fusion and adaptation (Ghahremannezhad et al., 2023; Ristić-Durrant et al., 2021).

Conclusion

Visual AI for object recognition has moved from experimental pilots to a dependable part of transport and logistics operations. Deep-learning detectors and trackers now turn camera streams into consistent, auditable events (identities, counts, violations, and condition flags) that feed traffic control centres and terminal systems. In road networks, this means stables city-scale analytics for safety and performance; in freight nodes (gates, yards, ports, rail terminals), it means faster identification, standardized inspections, and better orchestration of hand-offs across trucks, AGVs, cranes, and trains. Reviews focused on transportation video analytics document that modern object detection is the primary perception step in intelligent transport workflows, and that its accuracy and throughput meet operational needs when models are maintained (Ghahremannezhad et al., 2023).

The implementation pattern that works is clear: run inference at the edge to meet latency and bandwidth constraints; keep a narrow event and metadata interface into transport management systems (TMS), yard management systems (YMS), terminal operating systems (TOS), and traffic platforms; and engineer a full model lifecycle with data quality checks, drift monitoring, selective relabelling, and controlled retraining. At the sensing layer, robustness comes from multi-sensor fusion, combining cameras with lidar and/or radar to stabilise detection in rain, glare, fog, and at night, and from calibrated spatial and temporal alignment across modalities. Recent transport-focused surveys report consistent gains from fusion architectures (early/feature-level, late/decision-level, and learned cross-modal attention) and underline that edge intelligence is the right compute placement for real time intelligent transport systems (ITS). Together, these choices make recognition outputs reliable enough to drive key performance indicators (KPIs) such as throughput, dwell, incident rate, and queue length (Gong et al., 2023; Tang et al., 2023; Wang et al., 2024).

There are known limits that require active management. Models trained in one site or season often degrade in another (domain shift); rare hazards are under-represented; small or distant targets remain challenging; and adverse weather and occlusion reduce image quality. The evidence base points to three practical mitigations: (1) transfer learning or domain adaptation to align features between source and target domains when labels are scarce; (2) targeted data collection and hard-example mining as layouts, cameras, or seasons change; and (3) use of synthetic data to cover rare defects and corner cases before on site fine tuning. In container logistics, for example, synthetic labelled images of surfaces and damage types improve detection performance and inspection consistency; well-engineered multi-view OCR pipelines at port gates show how robust identity reading reduces manual rekeying and delays. These are mature, peer reviewed strategies with documented gains in transport settings (Delgado et al., 2023; Oza et al., 2021; Yoon et al., 2016).

Governance and risk controls must accompany deployment. Where feeds can include faces and plates, privacy and data-minimization patterns (on-edge inference, metadata only, or on-edge redaction) are appropriate. For identification functions, surveys catalogue how illumination, mounting, motion blur, and plate variability affect performance and recommend evaluation protocols and hardware choices to sustain accuracy; these same studies also highlight the need for transparent testing and bias controls before scale-up. In parallel, ITS surveys of edge intelligence discuss federated learning and distributed training as ways to improve models without centralizing sensitive video. These measures keep systems compliant while preserving operational value (Gong et al., 2023; Lubna et al., 2021).

Looking ahead, practice will focus on wider, standard use rather than brand new ideas. On roads, expect more multi-camera tracking, transformer-based perception, and continuous near-miss analytics that feed safety programs. In terminals and yards, expect closer links between identification and condition recognition and the scheduling and resource allocation tools, plus more sensor fusion where cameras alone are not enough. Across both areas, the emphasis will be on solid machine learning operations, regular domain adaptation, and edge-first deployments to keep latency low and accuracy stable across seasons and sites (Wang et al., 2023; Ghahremannezhad et al., 2023).

Visual AI is already delivering measurable improvements in safety, reliability, and throughput in transport and logistics. The organizations that benefit most will treat perception as an engineered system: fused where needed, deployed at the edge, governed for privacy, and maintained with data- and model-quality controls that are tied to business KPIs.

Tasks

1. What does visual AI refer to?

2. Describe one of the case studies in the context of visual AI and deep learning technology for object recognition.

11. VISUAL ARTIFICIAL INTELLIGENCE AND DEEP LEARNING TECHNOLOGY FOR GOODS COLLECTION AUTOMATION VIA INDEPENDENT PRODUCT IDENTIFICATION

Independent product identification means recognising items by their visual appearance (shape, texture, logos, and layout) without relying on line of sight barcodes or radio frequency identification (RFID) tags. In logistics and transport, this capability supports automated collection and verification wherever items are picked, consolidated, checked in or out, or routed across modes. Visual AI for this task uses deep learning to turn camera frames from stations, counters, conveyors, totes, or vehicle mounted cameras into structured outputs such as number of items in view or object embedding vector for catalogue matching. Compared with rule-based vision, learned models handle changes in lighting, cluttered backgrounds, varied viewing angles, and partial occlusions more robustly, and they can be fine-tuned as assortments or packaging designs change. Surveys of retail product recognition and store vision systems explain why deep detectors and fine-grained classifiers have become the standard perception layer for item identification in complex scenes (Santra and Mukherjee, 2019; Wei et al., 2020).

Modern recognition solutions follow a training and inference workflow. During training, models optimise two linked goals: classification (which item it is) and localisation (where it is in the image). A practical way to describe the model is in three stages: a feature extraction stage that converts the image into useful multi scale features; a feature fusion stage that combines coarse and fine details so both small and large items are captured; and a prediction stage that outputs bounding boxes, classes, and confidence scores. One stage detectors process the image in a single pass and are often chosen for high throughput at real time stations. Two stage detectors and transformer-based designs can improve precision or use longer range context when classes look very similar. At inference time, the trained network runs on edge hardware (processors near the cameras) to keep delay low, and post processing such as non-maximum suppression or learned de duplication removes overlapping detections. Real sites often budget only tens of milliseconds per frame, which is feasible with lightweight models and hardware aware optimisation. Surveys on lightweight detection and edge AI (running models on local devices) describe practical tools such as quantisation, pruning, and operator fusion that preserve accuracy while meeting power and latency limits (Mittal, 2024; Singh and Gill, 2023).

There are two common strategies for recognition. In closed set classification, the output must be one of K known stock keeping units (SKU), where K is the number of products currently in the catalogue, and the classifier is retrained as the catalogue evolves. In embedding based identification, the model produces a vector representation called an embedding, and the system searches a gallery of reference images using an approximate nearest neighbour index. This scales to large catalogues and allows incremental updates without full retraining. The retail vision literature covers fine grained recognition where different SKUs look very similar and reports gains from metric learning losses (for example margin based, contrastive, or proxy objectives) and from re ranking in the embedding space to refine the top matches. Reviews of retail product recognition and grocery shelf methods provide taxonomies that include both classifier based and retrieval pipelines for dense, cluttered scenes (Melek et al., 2024; Santra and Mukherjee, 2019).

Goods collection automation often faces fine grained classes such as flavour and size variants, and frequent catalogue change from new packaging and seasonal bundles. Accuracy can be improved with multi scale detection, high resolution crops, and fine grained models trained on shelf or checkout data sets that contain many near duplicate items. Packaging changes motivate multimodal fusion: image features are combined with optical character recognition text and simple layout descriptors from the front face of the package to reduce confusion among visually similar stock keeping units. Recent work in computer vision shows that combining vector representations from images with text extracted from packaging improves disambiguation in fine grained product sets, and that this fusion can be implemented with late fusion classifiers or multimodal transformer models (Melek et al., 2024; Pettersson et al., 2024).

Dataset design matters for product recognition. Retail vision benchmarks typically include dense shelf photos and automatic checkout scenes with both clean single item “gallery” images and multi item “tray” images. Reviews list common datasets, class groupings, and testing rules. Measure both perception quality and station outcomes: use mean average precision to judge detections and classifications, top one and top k accuracy to see how often the right class is chosen, and the identification score for tracking over time; also track precision and recall for “verify versus reject,” along with false accept rate and false reject rate. Grocery scenes are challenging because objects are small, often occluded, and class counts are uneven, so methods that handle small objects, occlusion, and class imbalance are important. Just as important is a fair and consistent test setup using the same training and test splits and clear treatment of rare classes so results can be reproduced and compared honestly (Melek et al., 2024).

Performance drops when moving between stores, cameras, seasons, or packaging updates. Transfer learning fine-tunes pre-trained detectors on small site-specific sets, while domain adaptation aligns features across source and target domains when labels are scarce. Reviews report that adversarial alignment, self-training, and consistency regularization recover much of the lost accuracy without relabelling entire catalogues. Where rare items or defect states are under-represented, synthetic data and domain randomization expand coverage before on-site fine-tuning; transport and retail studies document gains from mixing synthetic and real images for product and surface inspection (Melek et al., 2024; Tian et al., 2024).

Operational systems must recognize when a presented item is unknown (not in the training set) and either fall back to a secondary check or request labelling. Open-set and out-of-distribution (OOD) detection methods provide reject options by modelling score distributions, energy-based criteria, or generative boundaries; a recent synthesis clarifies the relationship between OOD detection and open-set recognition for safe deployment. For few-shot updates and class-incremental catalogues, surveys show that prototypes, distillation, and regularization can add new SKUs with limited forgetting. These capabilities are important for goods collection because catalogues change, and rejecting or on boarding new SKUs quickly is often more valuable than over-confidently misclassifying them (Tian et al., 2024; Wang et al., 2025).

Camera-only systems degrade in rain, glare, low light, or when objects are partially occluded. Fusion with depth or weight sensors at the station provides additional constraints for verification; more generally, camera–lidar or camera–radar fusion improves detection stability under adverse conditions. Multi-modality reviews in autonomous perception show consistent gains from early, late, or learned cross-modal fusion, and stress the importance of accurate time synchronization and extrinsic calibration so that measurements align in space and time. These principles transfer directly to mixed indoor–outdoor logistics sites and to collection cells exposed to variable illumination (Sun & Dong, 2023).

Goods-collection systems operate under strict latency budgets and bandwidth limits. Surveys of edge AI explain how on-device inference reduces backhaul, improves resilience, and enables privacy-preserving designs where only anonymized events or embedding leave the station. Maintaining accuracy over time requires data quality checks, drift detection, targeted relabelling of hard examples, versioned retraining, and calibration monitoring so confidence scores remain meaningful. Lightweight-detector surveys detail compression, quantization, and distillation techniques that keep models within power and compute envelopes of embedded CPU/NPU without unacceptable accuracy loss (Mittal, 2024; Singh & Gill, 2023).

Barcodes are inexpensive and widely used, but they need line of sight and require manual or robotic presentation. Radio frequency identification (RFID) uses small tags and readers to identify items without line of sight, often in bulk, which can give strong inventory visibility. However, tag cost, radio interference, and mixed item environments can limit use at the single item level. Reviews in supply chain research find that RFID together with the Internet of Things (IoT) improves visibility and agility, but many operations keep barcodes because of total cost and legacy systems. In practice, visual identification with AI can complement both by verifying contents when labels are occluded or missing and by providing image evidence for audit (Tan and Sidhu, 2022; Zhang et al., 2025).

Advantages of independent visual identification include label-agnostic operation, resilience to partial occlusion, rapid on boarding of new SKUs via fine-tuning or gallery updates, and direct auditability through stored images or embedding. Limitations include domain shift across sites and seasons, class imbalance with long-tail SKUs, small or highly similar items that require high-resolution crops and fine-grained models, and the need for ongoing dataset curation and calibration checks. The recent literature provides mitigation strategies; domain adaptation, multimodal fusion with OCR, synthetic augmentation, and edge-first deployments that make performance predictable and maintainable across the lifecycle (Melek et al., 2024; Pettersson et al., 2024).

Independent product identification removes a structural dependency on label presentation, enabling automation to proceed even when barcodes are damaged or not visible and when RFID is not economical. Technically, the field has matured: there are established model families for dense scenes, documented datasets and metrics for SKU-level recognition, robust methods for domain shift and catalogue churn, and deployment practices for edge inference and privacy-preserving operations. These ingredients explain why visual AI is now a viable core component for automating goods collection tasks across logistics and transport, with measurable gains when the perception layer is engineered together with data governance and lifecycle processes (Santra & Mukherjee, 2019; Melek et al., 2024).

Use cases

Independent product identification is already operating at several points in logistics and retail flows. In mature deployments the station is built around one or more fixed cameras (overhead or angled), controlled lighting, and an edge computer that runs a detector/classifier and optional tracker. The software emits structured events and pushes them to WMS/TMS/POS via a narrow API so the rest of the process (pick confirmation, tote association, sort decision, exception routing) happens automatically. Two patterns dominate in practice: (1) closed-set classification for limited product ranges with frequent fine-tuning, and (2) embedding/retrieval for large catalogues where a nearest-neighbour search over reference images makes incremental SKU on boarding straightforward. For visually similar items and new packaging, systems increasingly add multimodal cues: image features are fused with OCR tokens from the pack face, which has been shown to reduce confusion among look-alike SKUs in grocery-scale recognition (Pettersson et al., 2024).

Automatic checkout without scanning

Camera pods mounted above a tray or counter detect all items in view and tally categories and counts, reducing the need for manual barcode scans. Production oriented work on Automatic Check Out (ACO) uses the Retail Product Checkout (RPC) benchmark and addresses the training to testing domain gap (single product images versus multi item checkout scenes) with data priming and synthetic augmentation. The Data Priming Network reported 80.51 percent checkout accuracy on RPC, compared with 56.68 percent for a strong baseline, by jointly learning detection and counting and then fine tuning on automatically selected, reliable samples from deployment images. Subsequent work increased accuracy further with prototype based classifiers and re ranking tailored to fine grained retail categories. These methods map directly to automated checkout counters and staffed assisted checkout desks because they produce the outputs a point of sale system needs: a priced line item list with counts and confidence scores, plus an "unknown" route for exceptions (Chen et al., 2022; Li et al., 2019).

Pick verification

In goods to person cells and store picker stations, an overhead camera verifies each item as it enters the tote. The technical challenge is dense, cluttered scenes with many near duplicate stock keeping units. A paper, *Precise Detection in Densely Packed Scenes,* introduced the SKU 110K shelf benchmark and a detector design that improves localisation and separation of overlapping detections in tightly packed images, which matches conditions at pick faces. Recent production oriented systems pair dense detection with shelf row localisation and image retrieval based identification to assign each detection to a catalogue SKU and shelf position, forming the basis for on-the-fly pick confirmation and later audit. Together, these components demonstrate the accuracy and reliability needed to run verification at pick speed while keeping false accepts low (Goldman et al., 2019; Pietrini et al., 2024).

Parcel belts and high-speed sort feeds

In cross dock and express hubs, overhead cameras above in feed belts detect parcels, estimate their centre and footprint, and classify package types so programmable logic controllers (PLCs) can time transfers to cross belt carts. A 2024 study replaced light curtain positioning with a monocular camera pipeline based on a lightweight version of You Only Look Once version 5 (YOLOv5) and an enhanced Complete Intersection over Union (ECIoU) loss, then ran controlled trials with 10,000 parcels per method. The vision pipeline reduced parcel misplacement error rate and average supply time compared with the light curtain baseline, with clear guidance on where a single camera still needs pre-sorting for tall items. Follow on work in 2025 added fast, camera based three dimensional box positioning to meet higher line speeds. Together these results show a viable, label agnostic path to stabilise feed accuracy at modern sorter throughputs using commodity cameras and edge inference (Dai et al., 2025; Lu et al., 2024).

Returns intake

At service desks, the task is to confirm that the presented item matches the order or the return authorisation even when labels are occluded or the item is re boxed. Practical pipelines detect the product, compute a compact visual embedding, and optionally run optical character recognition (OCR) on stable text regions such as brand, variant, or net weight before scoring against the catalogue. Recent work on multimodal, fine grained grocery recognition shows that fusing visual embedding with OCR tokens from packaging improves discrimination among near identical stock keeping units, which reduces counter exceptions and speeds hand offs. Unified retail perception systems further demonstrate end to end detection, reading, and matching in real shelf and checkout imagery, providing a blueprint for stations that must both recognise and verify products in clutter (Chen et al., 2022; Pettersson et al., 2024).

Auditing re-used for goods collection quality control

Many operators reuse mature shelf-monitoring pipelines (row detection, product detection, and identification) as a quality layer for goods collection. The same camera that verifies picks can also calculate on-shelf availability, detect misplaced items, and flag planogram deviations that lead to later picking errors. A system description presents stock-keeping unit level detection and position assignment designed for edge deployment; the same components support pick verification and restock prompts without additional hardware (Pietrini et al., 2024).

Engineering playbook that shows up in working systems

Stations that perform reliably follow a consistent design: a fixed field of view and stable lighting; short exposure times to limit blur; inference at the edge (processing on devices near the cameras) to keep latency in the tens of milliseconds; a reject path for open set or out of distribution cases that routes unknowns to a scan or a second view; and a catalogue pipeline that supports incremental updates, for example adding new stock-keeping units with a small number of reference images or limited-example fine tuning. When conditions vary strongly, such as glare at dock doors or night operations, teams add a low-cost second signal, such as a scale plate or depth, to cross-check visual identification. For larger networks, surveys of edge intelligence in transport recommend distributed deployment with central model orchestration, which reduces bandwidth and supports privacy policies where only event metadata or compact feature vectors leave the site (Gong et al., 2023; Singh and Gill, 2023).

Accuracy and throughput at a station depend on three controllable factors: (1) data and catalogue curation, meaning representative reference images and periodic mining of difficult cases; (2) domain adaptation, that is lightweight fine tuning when cameras move, seasons change, or packaging updates are rolled out; and (3) multimodal fusion where needed, for example vision plus optical character recognition for fine-grained stock-keeping units, or vision plus simple sensors for belt timing. Published results show that these steps bring barcode-free identification to practical accuracy at checkout counters, pick benches, sort in feeds, and intake desks, while producing the auditable, structured events that warehouse management systems, transport management systems, and point-of-sale platforms require (Goldman et al., 2019; Pettersson et al., 2024).

Conclusion

Visual AI has matured into a reliable perception layer for automating goods collection without relying on barcodes or radio-frequency identification. Deep learning detectors and classifiers convert camera frames into auditable item identities and counts at the latency and accuracy needed at pick benches, checkout counters, parcel belts, and intake desks. The technical foundations are established: detectors tuned for dense, fine-grained products; tracking and re-identification for consistent counts; and edge deployment so that only compact event messages flow to warehouse management systems, transport management systems, and point-of-sale systems. This approach has been validated on domain data sets such as SKU-110K for tightly packed shelves and RPC for automatic checkout, and consolidated by surveys explaining why learned detectors replaced rule-based vision in retail and logistics recognition tasks (Goldman et al., 2019; Wei et al., 2019; Wei et al., 2020).

The advantages are operational and measurable. Independent visual identification reduces scan failures and manual keying, shortens dwell at verification points, lowers mis-pick and wrong-item rates, and creates image-grounded records that improve traceability and claims handling. Robustness improves further with well-understood practices: multi-sensor fusion where lighting and occlusion are difficult; transfer learning and domain adaptation to handle site and season changes; and machine learning operations, including drift monitoring, selective relabelling, and controlled retraining, to keep models calibrated over time. Edge placement is now a common default in transport contexts because it meets latency budgets and reduces backhaul while supporting privacy by keeping raw video on-site (Gong et al., 2023; Oza et al., 2021; Singh & Gill, 2023).

Looking ahead, the direction is broader standardisation rather than wholesale reinvention. Expect wider use of embedding-based identification that scales to large, changing catalogues; multimodal fusion that combines image features with optical character recognition from packaging to separate look-alike stock-keeping units; and more systematic domain adaptation so models transfer across stores, cameras, and seasons with minimal relabelling. As autonomy and connected operations expand, visual AI will integrate more tightly with scheduling and control systems, with edge computing in intelligent transport systems and in facilities providing the compute platform and lifecycle tooling. Visual identification will also coexist with barcodes and radio-frequency identification in hybrid flows, using vision to verify contents when labels are occluded or tags are absent, while radio-frequency identification continues to serve pallet and case tracking where the economics are favourable. The current evidence base in retail recognition, edge intelligent transport systems, and supply-chain radio-frequency identification suggests these directions are feasible and that organisations can capture value by treating perception as an engineered, monitored system tied to clear key performance indicators (Gong et al., 2023; Sarac et al., 2010; Wei et al., 2020).

Tasks

1. How is visual AI used in practice for goods collection?

12. THE FUTURE OF ARTIFICIAL INTELLIGENCE IN LOGISTICS AND SUPPLY CHAIN

In the previous chapters it was examined how AI is already embedded in transport and logistics forecasting and ETA prediction, routing and scheduling, terminal and yard coordination, visual recognition for identity and inspection, and autonomy within tightly defined domains. Looking ahead, progress will be fast and uneven: capabilities, regulations, costs, and hardware constraints will shift rapidly, and shocks will continue to test even well-designed systems. A sensible stance is optimistic but cautious: recognize where AI is ready to compound value, and design for adaptation because data, models, and operating contexts will keep changing. Recent reviews across operations and supply chain management conclude that AI is moving from point tools to end-to-end decision support, but they also document implementation frictions in data quality, change management, and governance signals to plan for as you scale (Cannas et al. 2023; Culot et al., 2024).

The first foundation of that future is the data substrate: moving from siloed/batch integrations to streaming, standardized event feeds from orders, inventory states, locations, telematics, and emissions factors, plus edge devices (vehicles, cranes, gates, cameras) that emit structured events rather than raw data. On the compute side, edge intelligence will be the default for time-critical functions; running inference close to sensors to meet latency, bandwidth, and privacy constraints while synchronizing models centrally. In parallel, digital twins of facilities and networks will mature from visualization tools to decision engines that continuously absorb live data and support “what-if” analysis for staffing, slotting, routing, and energy. Survey evidence in intelligent transportation and logistics shows consistent benefits from edge placement and twin-driven planning, provided data standards and model lifecycle processes are in place (Ghasemi et al., 2025; Gong et al., 2023; Le & Fan, 2024).

The second foundation is tighter prediction to optimization coupling. Tomorrow’s differentiator is not only forecasting demand or travel time but choosing good actions under uncertainty; learning-augmented solvers that allocate capacity, set prices, plan multimodal routes, or schedule cranes and crews while respecting constraints and risk. Methodologically, this pulls in reinforcement learning (single- and multi-agent) to handle sequential decisions and interactions, and causal inference to separate correlation from real impact so interventions generalize beyond historical data. Transportation and operations literature show momentum on both fronts: reinforced learning (RL) in transportation and new work on causal ML for supply-chain risk and intervention design lay out what works today and where gaps remain (e.g., safety constraints, data coverage, and off-policy evaluation) (Farazi et al., 2021; Lai et al., 2024; Wyrembek et al., 2025).

The human AI interface will evolve from dashboards to decision co-pilots embedded in planners' and supervisors' tools systems that surface recommendations with uncertainty, explain trade-offs, log overrides as learning signals, and support-controlled tests before policy rollouts. To make this safe at scale, organizations will formalize model risk management (data lineage, bias/drift testing, calibration monitoring, incident playbooks) and adopt privacy preserving learning where data cannot be centralized. Surveys and editorials across operations and federated learning highlight both the need for rigorous governance and practical mechanisms; federated and edge learning, privacy and fairness constraints to keep performance and compliance aligned (Ho et al., 2017; Li et al., 2024; Rafi et al., 2024).

Sustainability will move from annual reporting to shipment-level, auditable optimization. Carbon-aware planning will co-optimize service, cost, and emissions by selecting modes, consolidations, speeds, charge/energy plans, and time windows that meet targets while respecting operations. Bibliometric and modelling studies point to growing data availability for Scope-3 transparency and to optimization methods that embed emissions directly into logistics network design and control, tools that will become mandatory for success as disclosure and performance rules tighten (Masanta et al., 2025; Ouassou et al., 2024).

Resilience will be designed in rather than bolted on. AI systems will fuse weather, news, supplier, and telemetry signals to raise early warnings; scenario generators and digital twins will produce action playbooks; and policy-based re-planning will throttle promises, re-route shipments, or shift modes as constraints bind. New research applies causal ML to risk mitigation, formalising how to prioritise interventions and estimate their effect sizes. This is especially important when disruptions are frequent and heterogeneous (Shahsavari et al., 2025; Wyrembek et al., 2025).

Autonomy is likely to expand first where environments are structured and supervision is feasible: fenced yards and terminals; rail operations with automatic train operation and energy-optimised driving; supervised highway autonomy for long-distance haulage; and inspection by unscrewed aerial vehicles. The focus will be on clear operational design domains (explicit definitions of where and when the system may operate), continuous runtime monitoring, and formal safety assurance, rather than attempting unconstrained, fully autonomous operation. As autonomy and perception are integrated more tightly with scheduling and maintenance, benefits can include smoother flows, fewer human-factor incidents, and better asset utilisation, consistent with findings from intelligent transport edge-computing surveys and operations reviews (Cannas et al., 2023; Gong et al., 2023).

There are limits the future will not erase. Data quality and latency will still cap performance, models will still face domain shift when moved across sites, seasons, or partners; and privacy, security, and regulatory requirements will keep evolving. Edge compute and energy budgets will continue to constrain model size and frequency; rare but critical events will remain hard to learn from observational data alone. The prudent response is architectural, standardize data and identity, push time-critical inference to the edge, instrument models for monitoring and safe rollback, and keep option value via modular designs that can swap models or policies without re-wiring the enterprise (Ghasemi et al., 2025; Culot et al., 2024).

The next wave of artificial intelligence will build on streaming, standardised events from orders, inventories, telematics, and energy systems. These streams will be processed close to the sensors on edge devices and mirrored in digital twins that support continuous "what if" analysis. Running models at the edge reduces latency and bandwidth use, and logistics digital twins are moving from visualisation tools to decision engines when they are fed by live data and connected to planning systems (Gong et al., 2023; Le & Fan, 2024).

Deep learning is likely to broaden from bespoke estimated-time-of-arrival or demand models towards foundation models trained across many time series and contexts, then adapted to a specific corridor, lane, or set of products. These pre trained architectures can improve zero shot and few shot forecasting, anomaly detection, and control signals, giving planners stronger baselines that adapt more quickly when conditions change. Further gains are expected from models that fuse tabular data, text, and vision signals into a single system (Liang et al., 2024).

Because supply chains are naturally represented as graphs of facilities, lanes, and contracts, graph neural networks and graph convolutional networks are well suited to tasks such as lead time prediction, risk propagation, capacity allocation, and disruption routing. These methods capture spatial and relational structure better than flat models and, in early studies, outperform classical machine learning on classification, regression, and anomaly detection. As toolchains mature, graph models are expected to feed decisions back into routing, inventory, and scheduling optimisers, not only into dashboards (Monemi et al., 2025).

The differentiator will be turning predictions into actions under uncertainty. Reinforcement learning is growing in use for inventory positioning, dynamic pricing, appointment scheduling, and fleet control. A near term pattern is learning augmented optimisation: reinforcement learning or learned value functions guide classical solvers, with safety constraints and off policy evaluation so that policies can be checked before deployment. Multi agent reinforcement learning is relevant where carriers, shippers, and facilities interact strategically (Rolf et al., 2023; Yan et al., 2022). To avoid overfitting to yesterday's patterns, causal inference will complement deep prediction by estimating the effects of actions such as congestion fee changes, tender policies, or safety programmes, and by stress testing them in digital twins before rollout. Causal machine learning and matching or synthetic control methods provide credible "what if" estimates for risk and policy when shocks and regimes shift (Yılmaz et al., 2024).

Because much of the best data is sensitive, federated learning and federated transfer learning will be used to train models across carriers, cities, and facilities without pooling raw data. These approaches support traffic prediction, perception, and edge analytics while meeting data sovereignty constraints, which is directly relevant for shipper and carrier collaboration and multi-tenant logistics hubs (Pandya et al., 2023; Guo et al., 2024).

As models move into daily planning and control, the interface will look like decision co-pilots that present recommendations with uncertainty ranges, learn from operator overrides, and run controlled A/B tests safely. Value scales when data quality, model monitoring, calibration, and change management are in place. Expect formal model risk processes that cover lineage, drift, bias testing, and rapid rollback to become routine requirements for supply chain artificial intelligence programmes (Culot et al., 2024).

In practice, this points to edge-hosted perception and analytics linked to digital twins, foundation models for time series that provide reusable baselines, graph neural networks to capture network effects, reinforcement learning with learning augmented optimisation for constrained, sequential decisions, and federated training across partners. The direction is to build systems that learn faster, decide closer to the action, share data safely, and explain their recommendations, which will perform better under volatility and tighter sustainability mandates (Gong et al., 2023; Le & Fan, 2024).

Challenges

Large language models will only be trusted in planning and execution if they ground answers in verifiable data, expose their own uncertainty, and leave audit trails. Three persistent issues must be managed: hallucination (producing plausible but false content), poor calibration (answers that are over- or under-confident), and incomplete evaluation. A practical path forward is retrieval-augmented generation to tie outputs to enterprise facts, combined with uncertainty quantification and selective answering for safety-critical actions such as promising a delivery slot or altering a plan (Fan et al., 2024; Huang et al., 2025). In the near term, large language models will call tools, follow multi-step plans, and coordinate with other agents. Typical tools include Structured Query Language queries, optimisation solvers, and application programming interfaces for transport and warehouse management systems. Two lines are converging: agent frameworks that plan, critique, and retry; and learning-augmented optimisation, where a language model proposes formulations, heuristics, or code that feed exact or metaheuristic solvers for routing, scheduling, pricing, and allocation. This is how language models move from advice to closed-loop decision support in supply chains (Brahmachary et al., 2025; Du et al., 2025; Huang et al., 2024).

Future supply-chain co-pilots will not be text-only. Multimodal large language models will combine text with images (for example damage photos and labels), audio, and sensor or telematics streams to improve incident triage, claims handling, and visual inspections. In parallel, foundation models for time series and graph-native models will complement language models for estimated time of arrival and network effects. The key direction on the language side is systems that can reason over, and take actions from, mixed modalities inside logistics workflows (Yin et al., 2024). Much of the highest-value data is sensitive. Time-critical inference should run near sensors at the edge, and partners can collaborate without centralising raw data by using federated learning or privacy-preserving analytics. Expect language-model pipelines to combine on-premises retrieval-augmented indexes, smaller fine-tuned models at the edge, and periodic central coordination (Gong et al., 2023). Training and serving large models consume significant power. Practical guardrails include algorithmic and systems-level efficiency: distillation, quantisation, routing queries to appropriately sized models, and greener compute. In logistics, where many queries are repetitive and structured, routing to smaller specialists and caching at the edge will be standard to control cost and carbon (Jiang et al., 2024).

Programmes will need model-risk playbooks: data lineage, drift and bias testing, calibration monitoring, incident response, and human-in-the-loop approvals for high-impact actions such as repricing or service-level promises. These organisational foundations are prerequisites for scale (Toorajipour et al., 2021).

The biggest shift is agent-like, multimodal large language model systems that can read enterprise context, call tools, and execute end-to-end workflows. This stack is technically feasible and advancing quickly. Answers are grounded with retrieval augmented generation (RAG). Plans are broken into verifiable steps, optimisation solvers provide guarantees and human approvals are captured as learning signals. The result is a decision co-pilot for each role (planner, dispatcher, buyer, supervisor etc.) that handles routine analysis, documentation, and first pass decisions, shifting human work toward exceptions, negotiation, and policy (Du et al., 2025; Fan et al., 2024; Huang et al., 2024).

Key challenges to address are:

- Data readiness and hygiene: maintain versioned, trustworthy datasets and clear provenance; require outputs to cite or link to their sources.
- Evaluation and safety: use task specific evaluations that include calibration checks and selective abstention; run red team tests for hallucination and prompt injection.
- Integration with optimisation solvers: decide which decisions require solver guarantees and which can rely on heuristics or agent generated proposals.
- Edge and federated deployment: split workloads between on premises or edge and cloud, use federated learning or other privacy preserving training when data cannot be centralised.
- Cost and energy control: use model distillation, quantisation, and routing of prompts to appropriately sized models.

Large language models are likely to move from assistants to grounded, tool using agents embedded in logistics operations. Success depends on pairing them with retrieval augmented generation and digital twins for grounding, operations research solvers for guarantees, edge and federated patterns for privacy and latency, and disciplined governance for safety and cost. Organisations that put these foundations in place will be better positioned as agentic artificial intelligence matures and begins to reshape day to day work in supply chains.

13. CRITICAL THINKING AND CASE ANALYSIS

Critical thinking is a separate area that needs to be addressed in relation to artificial intelligence. Artificial intelligence and its tools currently provide quick and, to a large extent, correct answers to a whole range of questions and issues. Artificial intelligence can be a tool that promotes critical thinking, but it can also be a tool that "kills" critical thinking in young people. Artificial intelligence should be a "tool" or "aid" for the general public and the academic community to achieve answers and research goals. This chapter therefore focuses on the benefits of artificial intelligence in promoting critical thinking, but also on the negatives associated with its use.

Case analysis is also part of critical thinking and artificial intelligence. Case analysis reflects some of the features of critical thinking. During case analysis, scientists, students, and researchers focus on identifying key problems, evaluating evidence, and recommending solutions. This process reflects the steps of critical thinking, in which the individual must question assumptions, uncover biases, consider multiple perspectives, and develop reasoned arguments.

Critical Thinking

Critical thinking is conceptualized as a form of higher-order thinking activity, central to deciding what to believe or do. It encompasses cognitive capabilities such as interpretation, analysis, evaluation, inference, and explanation, as well as traits like self-regulation, open-mindedness, prudent judgment, and intellectual reflexivity (Dong et al., 2023).

Instructional Strategies Supporting Critical Thinking

Problem-Based Learning (PBL), when adapted with a focus on critical thinking, has proven effective in cultivating critical thinking skills across educational contexts. Systematic adaptations of PBL emphasize activities that specifically target learners' engagement in critical reasoning, evaluation, and self-reflection, thereby strengthening their capacity to apply critical thinking in complex, real-world scenarios (Yu and Zin, 2023).

A *meta-analysis* of educational interventions in higher education shows that when critical thinking skills are taught directly and embedded across different subject areas, they have a noticeable positive effect on students’ ability to think critically. However, the degree of improvement can differ depending on the teaching approach and the learning environment (Abrami et al., 2015).

Active learning techniques, including the use of case studies, written tasks, rigorous questioning, and structured debates, offer effective means of engaging learners in higher-level reflection and analysis.

Such methods provoke critical thinking by prompting students to articulate reasoning, evaluate alternatives, and defend positions within interactive environments (Walker, 2003).

According to Facione (2013), critical thinking can be characterized as follows:

“We understand critical thinking to be purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based. CT is essential as a tool of inquiry. As such, CT is a liberating force in education and a powerful resource in one’s personal and civic life. While not synonymous with good thinking, CT is a pervasive and self-rectifying human phenomenon. The ideal critical thinker is habitually inquisitive, well informed, trustful of reason, open-minded, flexible, fair-minded in evaluation, honest in facing personal biases, prudent in making judgments, willing to reconsider, clear about issues, orderly in complex matters, diligent in seeking relevant information, reasonable in the selection of criteria, focused in inquiry, and persistent in seeking results which are as precise as the subject and the circumstances of inquiry permit. Thus, educating strong critical thinkers means working toward this ideal. It combines developing CT skills with nurturing those dispositions which consistently yield useful insights, and which are the basis of a rational and democratic society”.

In 2023, a bibliometric analysis of keywords focused on the field of critical thinking research was conducted. The following figure presents the authors' (Dong et al., 2023) output from this bibliometric analysis. The figure represents the main areas of research related to critical thinking. The larger the node, the more publications focused on that keyword. The figure shows a current example of areas of critical thinking research (Dong et al., 2023).

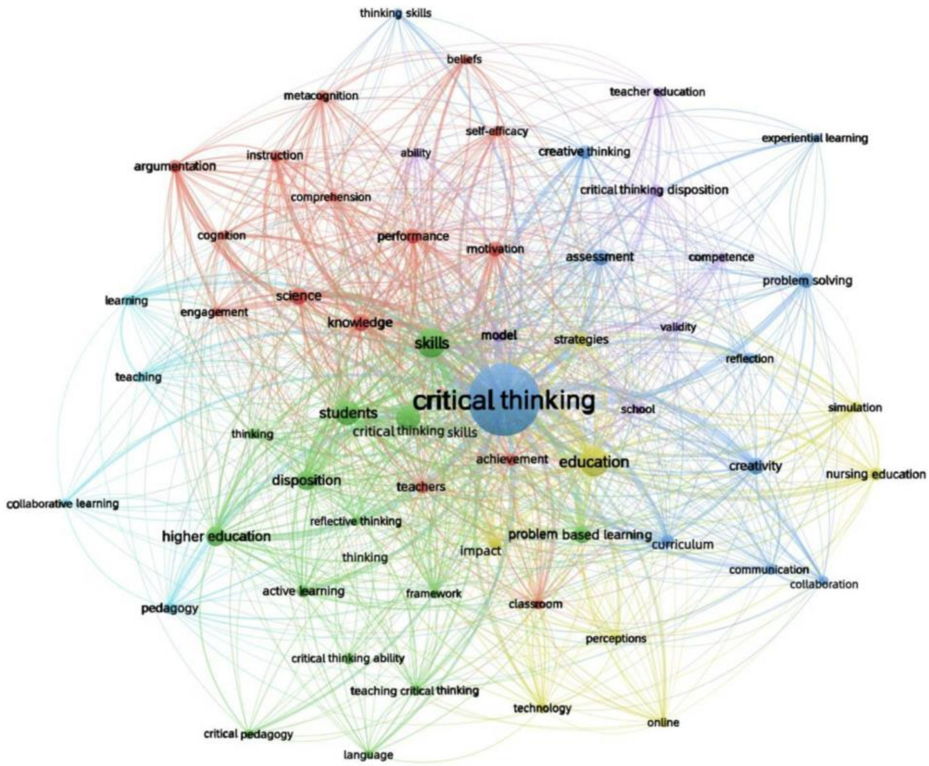

Figure 13.1 Network map of keyword co-occurrence for critical thinking research (Dong et al., 2023).

Critical Thinking vs Artificial Intelligence

The rise of AI brings with it not only new tools but also concerns about how these tools might change our thinking habits. A systematic review has shown that students can become overly dependent on AI dialogue systems especially those with generative capabilities when they accept AI responses without questioning them. This over-reliance can hinder critical cognitive functions, including analytical reasoning, decision-making, and the capacity for deep reflection. In such scenarios, users often favour quick but superficial cognitive shortcuts, weakening their ability to engage in rigorous critical thinking (Zhai et al., 2024).

On the other hand, not all AI interaction dampens thinking. Emerging frameworks like "extraheric AI" intentionally enhance human cognitive engagement by resisting easy answers. Rather than supplying direct solutions, extraheric AI offers alternative perspectives or questions that prompt users to think more deeply and creatively. By designing AI systems that foster exploration and reflection instead of delivering ready-made responses, developers can support higher-order skills such as critical analysis and problem-solving (Yatani et al., 2024).

A balanced perspective emerges in educational contexts: AI can be beneficial when tied to purposeful pedagogical use but risky if used as a crutch. Evidence from a randomized controlled trial in nursing education shows that LLM-assisted PBL yielded greater gains in students' critical thinking than traditional PBL. In this design, the LLM scaffolded guiding questions and supported group discussion so students could compare, contrast, and refine their ideas using AI outputs as a starting point for analysis rather than the final word (Kejingyun and Mingjun, 2025).

Case Analysis and Case Studies

Case analysis is widely recognized as a valuable scientific and educational method. In research, it is often used to demonstrate findings, verify proposed models, or explore relationships between variables influencing complex events. Case analysis serves two main purposes: as a detailed description of past events and as a research method for identifying causal links and testing theoretical assumptions. Because direct experience with complex situations is often difficult to gain in educational settings, the use of case analysis allows learners and professionals to study real situations and better understand the processes and decisions involved. This makes it a useful tool both for academic training and for deepening knowledge in professional practice (Ristvej et al., 2024).

The process of conducting a case analysis requires systematic steps to ensure validity and reliability. It begins with formulating clear research questions—most often of the "how" and "why" type—followed by careful selection of cases and methods of data collection. Researchers may examine unique or typical cases and collect data from multiple sources such as interviews, documents, observations, or statistics. This variety of sources ensures a comprehensive perspective and allows triangulation of evidence. The analysis phase involves identifying problems, evaluating alternative solutions, and interpreting results either through targeted criteria or by exploring the full complexity of the case. Finally, conclusions are summarized in a report, which documents the entire process and provides insights that may guide both theoretical development and practical applications (Ristvej et al., 2024).

Case Analysis in Critical Thinking

Case analysis serves as a dynamic and interactive teaching and learning approach that plays a crucial role in fostering critical thinking and problem-solving skills. Learners actively engage with real-world scenarios framed as cases, which require them to analyse information, assess multiple perspectives, and reason through complex situations. This method has been shown to enhance students' capacity for independent thinking, collaboration, and effective communication, linking theoretical understanding with practical application through reflective and critical engagement (Yazeedi et al., 2024).

Beyond its interactive benefits, case analysis promotes the application of analytical reasoning frameworks and the formulation of actionable recommendations. Unlike passive study methods, case analysis challenges students to go beyond observing and understanding they must evaluate decisions, apply relevant theories, and propose informed solutions. This transition from descriptive learning to active, theory-informed critical reasoning cultivates higher-order thinking skills such as evaluation, synthesis, and judgment—traits essential to professional competence and academic abilities (Catayoc, 2025).

Conclusion

The development of artificial intelligence brings new opportunities and challenges for critical thinking. On the one hand, it provides tools that enable rapid information gathering, encourage in-depth analysis, and support the learning process. On the other hand, however, there is a risk that overreliance on AI may weaken the ability to independently evaluate, analyse, and draw one's own conclusions. The connection between AI and critical thinking should therefore be seen as a balance between the use of technology and the preservation of human autonomous cognitive abilities.

Advantages of using artificial intelligence to promote critical thinking:

- Accelerates access to information: AI tools can provide quick and largely accurate answers, helping users gather relevant data for analysis and decision-making.
- Supports deeper reflection when designed well: Emerging frameworks such as *extraheric AI* are built to resist offering immediate answers and instead pose alternative perspectives or questions, encouraging users to think more critically and creatively.
- Enhances learning when integrated purposefully: Studies show that combining AI with structured pedagogical methods, such as Problem-Based Learning (PBL), can strengthen critical thinking skills by guiding students in questioning, discussion, and collaborative reasoning.
- Acts as a research and educational aid: When used appropriately, AI can serve as a supportive tool to achieve research goals and facilitate the development of critical reasoning in academic contexts.

Disadvantages of using artificial intelligence for the development of critical thinking:

- Risk of over-reliance: Users, especially students, may accept AI responses uncritically, leading to dependence that weakens their ability to analyse, evaluate, and make independent judgments.
- Encourages superficial shortcuts: Overuse of generative AI can promote quick but shallow cognitive processes, reducing engagement with complex problem-solving.
- Potential to diminish critical habits: If relied upon as a crutch rather than a tool, AI can undermine intellectual dispositions such as self-regulation, open-mindedness, and careful evaluation.
- Possible reduction in reflective depth: Dependence on ready-made answers discourages deep reasoning, creative exploration, and the development of higher-order skills necessary for robust critical thinking.

European Funds for Social Development
Republic of Poland
Co-funded by the European Union
NAWA POLISH NATIONAL AGENCY FOR ACADEMIC EXCHANGE

Conclusion

Artificial intelligence (AI) is reshaping logistics and supply chains by addressing complex, time-sensitive decisions with data-driven methods. Across planning, execution and control, AI supports improvements in cost efficiency, service quality, responsiveness and resilience, while creating new opportunities for operational visibility and continuous optimisation.

This script provides a practice-oriented overview of AI methods and applications relevant to logistics and supply-chain contexts. The emphasis is on clear problem framing, disciplined use of data, experimental thinking and rigorous evaluation of outcomes. Coverage spans strategic and operational themes: distinctions between supply-chain and logistics perspectives; core AI capabilities; the relationship between predictive analytics and demand forecasting; and the application of AI to inventory management, production planning, warehouse operations and transport processes. Attention is given to network-level challenges such as routing, capacity allocation, fleet management and real-time monitoring.

The integration of digital intelligence with physical work is examined through AI-enabled robotics and autonomous systems. Visual AI and deep learning for object recognition are introduced as enablers of quality inspection, safety and goods-handling automation, including independent product identification. A forward-looking view considers the trajectory of AI in logistics and supply chains, the interaction with adjacent digital technologies and the organisational conditions for effective adoption. Governance, safety, ethics and human–machine collaboration are treated as practical design constraints rather than afterthoughts.

Artificial intelligence has the potential to materially improve decision quality and operational performance across logistics and supply chains by combining large volumes of data with adaptive models for prediction, optimisation and control. It enables faster, more consistent responses to volatility and disruption, automates routine tasks to release human capacity for higher-value work, and enhances end-to-end visibility for proactive risk management. As customer expectations rise and cost, sustainability and safety pressures intensify, AI provides a practical route to tighter resource utilisation, lower error rates and measurable service gains. Focusing on this topic is therefore important not only for competitive advantage but also for building resilient, transparent and auditable operations, developing a workforce capable of effective human–machine collaboration, and ensuring that governance and ethics are embedded as capabilities scale.

Bibliography

Abdolmaleki, M., Shahabi, M., Yin, Y., Masoud, N., 2021. Itinerary planning for cooperative truck platooning. Transportation Research Part B: Methodological 153, 91–110. https://doi.org/10.1016/j.trb.2021.08.016

Abrami, P.C., Bernard, R.M., Borokhovski, E., Waddington, D.I., Wade, C.A., Persson, T., 2015. Strategies for Teaching Students to Think Critically: A Meta-Analysis. Review of Educational Research 85, 275–314. https://doi.org/10.3102/0034654314551063

Adamo, T., Gendreau, M., Ghiani, G., Guerriero, E., 2024. A review of recent advances in time-dependent vehicle routing. European Journal of Operational Research 319, 1–15. https://doi.org/10.1016/j.ejor.2024.06.016

Adedoyin Tolulope Oyewole, Chinwe Chinazo Okoye, Onyeka Chrisanctus Ofodile, Emuesiri Ejairu, 2024. Reviewing predictive analytics in supply chain management: Applications and benefits. World J. Adv. Res. Rev. 21, 568–574. https://doi.org/10.30574/wjarr.2024.21.3.0673

Aktas, E., Bobba, S., 2025. Managing supply chains in times of uncertainty: The emergence of digital twin technologies.

Alaei, S., Durán-Micco, J., Macharis, C., 2024. Synchromodal transport re-planning: an agent-based simulation approach. Eur. Transp. Res. Rev. 16, 1. https://doi.org/10.1186/s12544-023-00624-y

Al-Amin, Md., Hossain, Md.T., Islam, Md.J., Biwas, S.K., 2023. History, Features, Challenges, and Critical Success Factors of Enterprise Resource Planning (ERP) in The Era of Industry 4.0. https://doi.org/10.22541/essoar.168056807.74820210/v1

Ammar, A., Koubaa, A., Boulila, W., Benjdira, B., Alhabashi, Y., 2023. A Multi-Stage Deep-Learning-Based Vehicle and License Plate Recognition System with Real-Time Edge Inference. Sensors 23, 2120. https://doi.org/10.3390/s23042120

Andriolo, A., Battini, D., Grubbström, R.W., Persona, A., Sgarbossa, F., 2014. A century of evolution from Harris's basic lot size model: Survey and research agenda. International Journal of Production Economics 155, 16–38. https://doi.org/10.1016/j.ijpe.2014.01.013

APICS, 2017. Supply Chain Operations Reference Model SCOR.

Arbabkhah, H., Sedaghat, A., Jafari Kang, M., Hamidi, M., 2024. Automatic Identification System-Based Prediction of Tanker and Cargo Estimated Time of Arrival in Narrow Waterways. JMSE 12, 215. https://doi.org/10.3390/jmse12020215

Armstrong, J.S. (Ed.), 2004. Principles of forecasting: a handbook for researchers and practitioners, 4. printing. ed, International series in operations research & management science. Kluwer Academic, Boston.

Awad, M., Khanna, R., 2015. Machine Learning, in: Efficient Learning Machines. Apress, Berkeley, CA, pp. 1–18. https://doi.org/10.1007/978-1-4302-5990-9_1

Awan, A.A., 2022. The Machine Learning Life Cycle Explained. URL https://www.datacamp.com/blog/machine-learning-lifecycle-explained?dc_referrer=https%3A%2F%2F

Azadeh, K., De Koster, R., Roy, D., 2019. Robotized and Automated Warehouse Systems: Review and Recent Developments. Transportation Science 53, 917–945. https://doi.org/10.1287/trsc.2018.0873

Babai, M.Z., Syntetos, A.A., Teunter, R.H., 2025. Fifty years of inventory research from a forecasting perspective. European Journal of Operational Research S0377221725005272. https://doi.org/10.1016/j.ejor.2025.07.003

Baker-Brunnbauer, J., 2023. Artificial Intelligence, in: Trustworthy Artificial Intelligence Implementation, Business Guides on the Go. Springer International Publishing, Cham, pp. 7–29. https://doi.org/10.1007/978-3-031-18275-4_2

Batz, A., D'Croz-Barón, D.F., Vega Pérez, C.J., Ojeda-Sanchez, C.A., 2025. Integrating machine learning into business and management in the age of artificial intelligence. Humanit Soc Sci Commun 12, 352. https://doi.org/10.1057/s41599-025-04361-6

Becker, T., 2023. Improving the SCOR model: Is the new SCOR Digital Standard a better process reference model? JSCM 6, 155. https://doi.org/10.69554/JFSP8752

Ben Hamou, K.A., Jarir, Z., Elfirdoussi, S., 2025. Using machine learning for production scheduling problems in the supply chain: A review. Computers & Industrial Engineering 206, 111243. https://doi.org/10.1016/j.cie.2025.111243

Berliner, H., 1980. Computer Backgammon. Sci Am 242, 64–72. https://doi.org/10.1038/scientificamerican0680-64

Bernardo, R., Sousa, J.M.C., Gonçalves, P.J.S., 2022. Survey on robotic systems for internal logistics. Journal of Manufacturing Systems 65, 339–350. https://doi.org/10.1016/j.jmsy.2022.09.014

Bett, D.K., Ali, I., Gheith, M., Eltawil, A., 2024. Simulation-Based Optimization of Truck Appointment Systems in Container Terminals: A Dual Transactions Approach with Improved Congestion Factor Representation. Logistics 8, 80. https://doi.org/10.3390/logistics8030080

Bhoopalam, A.K., Agatz, N., Zuidwijk, R., 2018. Planning of truck platoons: A literature review and directions for future research. Transportation Research Part B: Methodological 107, 212–228. https://doi.org/10.1016/j.trb.2017.10.016

Bhowmik, O., Chowdhury, S., Ashik, J.H., Mahmud, G.I., Khan, M.M., Hossain, N.U.I., 2024. Application of artificial intelligence in reverse logistics: A bibliometric and network analysis. Supply Chain Analytics 7, 100076. https://doi.org/10.1016/j.sca.2024.100076

Bierwirth, C., Meisel, F., 2015. A follow-up survey of berth allocation and quay crane scheduling problems in container terminals. European Journal of Operational Research 244, 675–689. https://doi.org/10.1016/j.ejor.2014.12.030

Bierwirth, C., Meisel, F., 2010. A survey of berth allocation and quay crane scheduling problems in container terminals. European Journal of Operational Research 202, 615–627. https://doi.org/10.1016/j.ejor.2009.05.031

Bottani, E., Casella, G., Monferdini, L., 2025. Tracing the evolution of warehouse technologies: a bibliometric study. Procedia Computer Science 253, 2951–2960. https://doi.org/10.1016/j.procs.2025.02.019

Boukerche, A., Hou, Z., 2022. Object Detection Using Deep Learning Methods in Traffic Scenarios. ACM Comput. Surv. 54, 1–35. https://doi.org/10.1145/3434398

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M., 2016. Time series analysis: forecasting and control, Fifth edition. ed, Wiley series in probability and statistics. Wiley, Hoboken, New Jersey.

Boysen, N., De Koster, R., Weidinger, F., 2019. Warehousing in the e-commerce era: A survey. European Journal of Operational Research 277, 396–411. https://doi.org/10.1016/j.ejor.2018.08.023

Brahmachary, S., Joshi, S.M., Panda, A., Koneripalli, K., Sagotra, A.K., Patel, H., Sharma, A., Jagtap, A.D., Kalyanaraman, K., 2025. Large language model-based evolutionary optimizer: Reasoning with elitism. Neurocomputing 622, 129272. https://doi.org/10.1016/j.neucom.2024.129272

Burns, A.J., Michalek, J.J., Samaras, C., 2024. Estimating the potential for optimized curb management to reduce delivery vehicle double parking, traffic congestion and energy consumption.

Transportation Research Part E: Logistics and Transportation Review 187, 103574. https://doi.org/10.1016/j.tre.2024.103574

Caballini, C., Gracia, M.D., Mar-Ortiz, J., Sacone, S., 2020. A combined data mining – optimization approach to manage trucks operations in container terminals with the use of a TAS: Application to an Italian and a Mexican port. Transportation Research Part E: Logistics and Transportation Review 142, 102054. https://doi.org/10.1016/j.tre.2020.102054

Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., Leonard, J.J., 2016. Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. IEEE Trans. Robot. 32, 1309–1332. https://doi.org/10.1109/TRO.2016.2624754

Cannas, V.G., Ciano, M.P., Saltalamacchia, M., Secchi, R., 2024. Artificial intelligence in supply chain and operations management: a multiple case study research. International Journal of Production Research 62, 3333–3360. https://doi.org/10.1080/00207543.2023.2232050

Cao, F., Tang, T., Gao, Y., Michler, O., Schultz, M., 2024. Predicting flight arrival times with deep learning: A strategy for minimizing potential conflicts in gate assignment. Transportation Research Part C: Emerging Technologies 169, 104866. https://doi.org/10.1016/j.trc.2024.104866

Carlo, H.J., Vis, I.F.A., Roodbergen, K.J., 2014. Storage yard operations in container terminals: Literature overview, trends, and research directions. European Journal of Operational Research 235, 412–430. https://doi.org/10.1016/j.ejor.2013.10.054

Carpitella, S., Izquierdo, J., 2025. Trends in Sustainable Inventory Management Practices in Industry 4.0. Processes 13, 1131. https://doi.org/10.3390/pr13041131

Carter, C.R., Rogers, D.S., 2008. A framework of sustainable supply chain management: moving toward new theory. Int Jnl Phys Dist & Log Manage 38, 360–387. https://doi.org/10.1108/09600030810882816

Catayoc, R.B., 2025. The Distinction between Case Study and Case Analysis in Educational and Professional Contexts. CJMS 5, 287–298. https://doi.org/10.47760/cognizance.2025.v05i01.024

Ceccarelli, G., Messa, F., Gorrini, A., Presicce, D., Choubassi, R., 2023. Deep learning video analytics for the assessment of street experiments: The case of Bologna. Journal of Urban Mobility 4, 100067. https://doi.org/10.1016/j.urbmob.2023.100067

Chargui, K., Zouadi, T., Sreedharan, V.R., 2023. Berth and quay crane allocation and scheduling problem with renewable energy uncertainty: A robust exact decomposition. Computers & Operations Research 156, 106251. https://doi.org/10.1016/j.cor.2023.106251

Chen, C., Lee Kong, T., Kan, W., 2023. Identifying the promising production planning and scheduling method for manufacturing in Industry 4.0: a literature review. Production & Manufacturing Research 11, 2279329. https://doi.org/10.1080/21693277.2023.2279329

Chen, C., Liu, J., Yin, H., Huang, B., 2025. A Vision-Based Method for Detecting the Position of Stacked Goods in Automated Storage and Retrieval Systems. Sensors 25, 2623. https://doi.org/10.3390/s25082623

Chen, C.-S., Chen, Y.-J., 2025. Optimizing Supply Chain Networks with the Power of Graph Neural Networks. https://doi.org/10.48550/arXiv.2501.06221

Chen, H., Wei, X.-S., Zhang, F., Shen, Y., Xu, H., Xiao, L., 2022. Automatic Check-Out via Prototype-Based Classifier Learning from Single-Product Exemplars, in: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (Eds.), Computer Vision – ECCV 2022, Lecture Notes in Computer Science. Springer Nature Switzerland, Cham, pp. 277–293. https://doi.org/10.1007/978-3-031-19806-9_16

Chen, T., Sampath, V., May, M.C., Shan, S., Jorg, O.J., Aguilar Martín, J.J., Stamer, F., Fantoni, G., Tosello, G., Calaon, M., 2023. Machine Learning in Manufacturing towards Industry 4.0: From 'For Now' to 'Four-Know.' Applied Sciences 13, 1903. https://doi.org/10.3390/app13031903

Chen, W., Men, Y., Fuster, N., Osorio, C., Juan, A.A., 2024. Artificial Intelligence in Logistics Optimization with Sustainable Criteria: A Review. Sustainability 16, 9145. https://doi.org/10.3390/su16219145

Chiu, H.N., Chen, H.M., Weng, L.C., 2003. DETERMINISTIC TIME-VARYING DEMAND LOT-SIZING MODELS WITH LEARNING AND FORGETTING IN SETUPS AND PRODUCTION. Production and Operations Management 12, 120–127. https://doi.org/10.1111/j.1937-5956.2003.tb00201.x

Christopher, M., 2016. Logistics & supply chain management, Fifth Edition. ed, Always learning. Pearson Education, Harlow, England : New York.

Christopher, M., 2009. Logistics and supply chain management: creating value-adding networks, 3. ed., [Nachdr.]. ed. Financial Times Prentice Hall, Harlow Munich.

Colombathanthri, A., Jomaa, W., Chinniah, Y.A., 2025. Human-centered cyber-physical systems in manufacturing industry: a systematic search and review. Int J Adv Manuf Technol 136, 2107–2141. https://doi.org/10.1007/s00170-024-14959-w

Covell, D.D., 2023. Preventing WIPlash: Implementation of a Controlled Release Strategy to Improve Shop Performance. MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

Cullen, J., Bryman, A., 1988. The Knowledge Acquisition Bottleneck: Time for Reassessment? Expert Systems 5, 216–225. https://doi.org/10.1111/j.1468-0394.1988.tb00065.x

Culot, G., Podrecca, M., Nassimbeni, G., 2024. Artificial intelligence in supply chain management: A systematic literature review of empirical studies and research directions. Computers in Industry 162, 104132. https://doi.org/10.1016/j.compind.2024.104132

Dai, N., Chen, J., Hu, X., Yuan, Y., 2025. Research on dimension measurement algorithm for parcel boxes in high-speed sorting system. Sci Rep 15, 21980. https://doi.org/10.1038/s41598-025-07730-y

Das, S., 2025. What is Autonomous and Resilient Supply Chain Planning? Concepts, Benefits & Implementation. URL https://3scsolution.com/insight/autonomous-resilient-supply-chain-planning

Dave, P., 2020. The History of Lean Manufacturing by the view of Toyota-Ford. International Journal of Scientific and Engineering Research 11, 1598–1602.

De Assis, R.F., Faria, A.F., Thomasset-Laperrière, V., Santa-Eulalia, L.A., Ouhimmou, M., Ferreira, W.D.P., 2024. Machine Learning in Warehouse Management: A Survey. Procedia Computer Science 232, 2790–2799. https://doi.org/10.1016/j.procs.2024.02.096

De Souza, E.A.G., Mlinar, T., Van Den Broeke, M., Creemers, S., 2025. Dynamic programming in inventory management: A review. Computers & Operations Research 183, 107164. https://doi.org/10.1016/j.cor.2025.107164

Degas, A., Islam, M.R., Hurter, C., Barua, S., Rahman, H., Poudel, M., Ruscio, D., Ahmed, M.U., Begum, S., Rahman, M.A., Bonelli, S., Cartocci, G., Di Flumeri, G., Borghini, G., Babiloni, F., Aricó, P., 2022. A Survey on Artificial Intelligence (AI) and eXplainable AI in Air Traffic Management: Current Trends and Development with Future Research Trajectory. Applied Sciences 12, 1295. https://doi.org/10.3390/app12031295

Delgado, G., Cortés, A., García, S., Loyo, E., Berasategi, M., Aranjuelo, N., 2023. Methodology for generating synthetic labeled datasets for visual container inspection. Transportation Research Part E: Logistics and Transportation Review 175, 103174. https://doi.org/10.1016/j.tre.2023.103174

Demir, G., Moslem, S., Duleba, S., 2024. Artificial Intelligence in Aviation Safety: Systematic Review and Biometric Analysis. Int J Comput Intell Syst 17, 279. https://doi.org/10.1007/s44196-024-00671-w

Dick, S., 2019. Artificial Intelligence. Harvard Data Science Review. https://doi.org/10.1162/99608f92.92fe150c

Dolgui, A., Proth, J.-M., 2010. Supply Chain Engineering. Springer London, London. https://doi.org/10.1007/978-1-84996-017-5

Dong, M., Li, F., Chang, H., 2023. Trends and hotspots in critical thinking research over the past two decades: Insights from a bibliometric analysis. Heliyon 9, e16934. https://doi.org/10.1016/j.heliyon.2023.e16934

Douaioui, K., Oucheikh, R., Benmoussa, O., Mabrouki, C., 2024. Machine Learning and Deep Learning Models for Demand Forecasting in Supply Chain Management: A Critical Review. ASI 7, 93. https://doi.org/10.3390/asi7050093

Dresser, S., 2025. Amazon launches a new AI foundation model to power its robotic fleet and deploys its 1 millionth robot. URL https://www.aboutamazon.com/news/operations/amazon-million-robots-ai-foundation-model

Drissi Elbouzidi, A., Ait El Cadi, A., Pellerin, R., Lamouri, S., Tobon Valencia, E., Bélanger, M.-J., 2023. The Role of AI in Warehouse Digital Twins: Literature Review. Applied Sciences 13, 6746. https://doi.org/10.3390/app13116746

Du, S., Zhao, J., Shi, J., Xie, Z., Jiang, X., Bai, Y., He, L., 2025. A Survey on the Optimization of Large Language Model-based Agents. https://doi.org/10.48550/ARXIV.2503.12434

Du, Z., Wu, J., Leng, Y., Wandelt, S., 2025. AI4ATM: A review on how Artificial Intelligence paves the way towards autonomous Air Traffic Management. Journal of the Air Transport Research Society 5, 100077. https://doi.org/10.1016/j.jatrs.2025.100077

Duque-Jaramillo, J.C., Cogollo-Flórez, J.M., Gómez-Marín, C.G., Correa-Espinal, A.A., 2024. Warehouse management optimization using a sorting-based slotting approach. JIEM 17, 133. https://doi.org/10.3926/jiem.5661

Dwork, C., Roth, A., 2013. The Algorithmic Foundations of Differential Privacy. FNT in Theoretical Computer Science 9, 211–407. https://doi.org/10.1561/0400000042

Elahi, A., 2022. Boolean Logics and Logic Gates, in: Computer Systems. Springer International Publishing, Cham, pp. 33–50. https://doi.org/10.1007/978-3-030-93449-1_2

Ellithy, K., Salah, M., Fahim, I.S., Shalaby, R., 2024. AGV and Industry 4.0 in warehouses: a comprehensive analysis of existing literature and an innovative framework for flexible automation. Int J Adv Manuf Technol 134, 15–38. https://doi.org/10.1007/s00170-024-14127-0

Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., Dehmer, M., 2020. An Introductory Review of Deep Learning for Prediction Models With Big Data. Front. Artif. Intell. 3, 4. https://doi.org/10.3389/frai.2020.00004

Enoch Oluwademilade Sodiya, Uchenna Joseph Umoga, Olukunle Oladipupo Amoo, Akoh Atadoga, 2024. AI-driven warehouse automation: A comprehensive review of systems. GSC Adv. Res. Rev. 18, 272–282. https://doi.org/10.30574/gscarr.2024.18.2.0063

Esmaeilzadeh, E., Mokhtarimousavi, S., 2020. Machine Learning Approach for Flight Departure Delay Prediction and Analysis. Transportation Research Record: Journal of the Transportation Research Board 2674, 145–159. https://doi.org/10.1177/0361198120930014

Evmides, N., Aslam, S., Ramez, T.T., Michaelides, M.P., Herodotou, H., 2024. Enhancing Prediction Accuracy of Vessel Arrival Times Using Machine Learning. JMSE 12, 1362. https://doi.org/10.3390/jmse12081362

Expósito Jiménez, V.J., Macher, G., Watzenig, D., Brenner, E., 2024. Safety of the Intended Functionality Validation for Automated Driving Systems by Using Perception Performance Insufficiencies Injection. Vehicles 6, 1164–1184. https://doi.org/10.3390/vehicles6030055

Facione, P., 2013. Critical Thinking: What It Is and Why It Counts.

Fagnant, D.J., Kockelman, K., 2015. Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. Transportation Research Part A: Policy and Practice 77, 167–181. https://doi.org/10.1016/j.tra.2015.04.003

Fakhry, D., Oger, R., Lauras, M., Pellegrin, V., 2025. Managing uncertainties within sales and operations planning (S&OP): a systematic literature review. Supply Chain Forum: An International Journal 1–18. https://doi.org/10.1080/16258312.2025.2526321

Fan, W., Ding, Y., Ning, L., Wang, S., Li, H., Yin, D., Chua, T.-S., Li, Q., 2024. A Survey on RAG Meeting LLMs: Towards Retrieval-Augmented Large Language Models, in: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Presented at the KDD '24: The 30th

ACM SIGKDD Conference on Knowledge Discovery and Data Mining, ACM, Barcelona Spain, pp. 6491–6501. https://doi.org/10.1145/3637528.3671470

Farahzadi, A., Shams, P., Rezazadeh, J., Farahbakhsh, R., 2018. Middleware technologies for cloud of things: a survey. Digital Communications and Networks 4, 176–188. https://doi.org/10.1016/j.dcan.2017.04.005

Fernández-Rodríguez, A., Cucala, A.P., Fernández-Cardador, A., 2020. An Eco-Driving Algorithm for Interoperable Automatic Train Operation. Applied Sciences 10, 7705. https://doi.org/10.3390/app10217705

Ferreira, B., Reis, J., 2023. A Systematic Literature Review on the Application of Automation in Logistics. Logistics 7, 80. https://doi.org/10.3390/logistics7040080

Fisher, M., 1997. What Is the Right Supply Chain for Your Product? URL https://hbr.org/1997/03/what-is-the-right-supply-chain-for-your-product

Fragapane, G., De Koster, R., Sgarbossa, F., Strandhagen, J.O., 2021. Planning and control of autonomous mobile robots for intralogistics: Literature review and research agenda. European Journal of Operational Research 294, 405–426. https://doi.org/10.1016/j.ejor.2021.01.019

Freese, F., Ludwig, A., 2025. A conceptual framework for supply chain digital twins – development and evaluation. International Journal of Logistics Research and Applications 28, 676–698. https://doi.org/10.1080/13675567.2024.2324895

Fröhlich, E., Steinbiß, K., Schmiedeknecht, M.H., Loza Adaui, C.R. (Eds.), 2024. Circular Economy in Sustainable Supply Chains: A Global Perspective on Challenges, Concepts and Cases, CSR, Sustainability, Ethics & Governance. Springer Nature Switzerland, Cham. https://doi.org/10.1007/978-3-031-70749-0

G M, A., 2020. A Brief Overview on Toyota Production System (TPS). IJRASET 8, 2505–2509. https://doi.org/10.22214/ijraset.2020.5415

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., Bouchachia, A., 2014. A survey on concept drift adaptation. ACM Comput. Surv. 46, 1–37. https://doi.org/10.1145/2523813

Gattermann-Itschert, T., Poreschack, L.M., Thonemann, U.W., 2023. Using Machine Learning to Include Planners' Preferences in Railway Crew Scheduling Optimization. Transportation Science 57, 796–812. https://doi.org/10.1287/trsc.2022.1196

Geevers, K., Van Hezewijk, L., Mes, M.R.K., 2024. Multi-echelon inventory optimization using deep reinforcement learning. Cent Eur J Oper Res 32, 653–683. https://doi.org/10.1007/s10100-023-00872-2

Ghahremannezhad, H., Shi, H., Liu, C., 2023. Object Detection in Traffic Videos: A Survey. IEEE Trans. Intell. Transport. Syst. 24, 6780–6799. https://doi.org/10.1109/TITS.2023.3258683

Ghasemi, A., Keshavarzi, A., M. Abdelmoniem, A., Nejati, O.R., Derikvand, T., 2025. Edge Intelligence for Intelligent Transport Systems: Approaches, challenges, and future directions. Expert Systems with Applications 280, 127273. https://doi.org/10.1016/j.eswa.2025.127273

Gielis, J., Shankar, A., Prorok, A., 2022. A Critical Review of Communications in Multi-robot Systems. Curr Robot Rep 3, 213–225. https://doi.org/10.1007/s43154-022-00090-9

Giuffrida, N., Fajardo-Calderin, J., Masegosa, A.D., Werner, F., Steudter, M., Pilla, F., 2022. Optimization and Machine Learning Applied to Last-Mile Logistics: A Review. Sustainability 14, 5329. https://doi.org/10.3390/su14095329

Giusti, R., Manerba, D., Bruno, G., Tadei, R., 2019. Synchromodal logistics: An overview of critical success factors, enabling technologies, and open research issues. Transportation Research Part E: Logistics and Transportation Review 129, 92–110. https://doi.org/10.1016/j.tre.2019.07.009

Golchin, M., Grandhi, A., Gore, N., Pulugurtha, S.S., Ghasemi, A., 2024. UNC Charlotte Autonomous Shuttle Pilot Study: An Assessment of Operational Performance, Reliability, and Challenges. Machines 12, 796. https://doi.org/10.3390/machines12110796

Goldman, E., Herzig, R., Eisenschtat, A., Goldberger, J., Hassner, T., 2019. Precise Detection in Densely Packed Scenes, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Long Beach, CA, USA, pp. 5222–5231. https://doi.org/10.1109/CVPR.2019.00537

Gong, T., Zhu, L., Yu, F.R., Tang, T., 2023. Edge Intelligence in Intelligent Transportation Systems: A Survey. IEEE Trans. Intell. Transport. Syst. 24, 8919–8944. https://doi.org/10.1109/TITS.2023.3275741

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep learning, Adaptive computation and machine learning. The MIT Press, Cambridge, Massachusetts.

Govindan, K., Kannan, D., Jørgensen, T.B., Nielsen, T.S., 2022. Supply Chain 4.0 performance measurement: A systematic literature review, framework development, and empirical evidence.

Transportation Research Part E: Logistics and Transportation Review 164, 102725. https://doi.org/10.1016/j.tre.2022.102725

Gu, J., Goetschalckx, M., McGinnis, L.F., 2007. Research on warehouse operation: A comprehensive review. European Journal of Operational Research 177, 1–21. https://doi.org/10.1016/j.ejor.2006.02.025

Guerra-Zubiaga, D.A., Aksu, M., Richards, G., Kuts, V., 2025. Integrating Digital Twin Software Solutions with Collaborative Industrial Systems: A Comprehensive Review for Operational Efficiency. Applied Sciences 15, 7049. https://doi.org/10.3390/app15137049

Guo, D., Mantravadi, S., 2025. The role of digital twins in lean supply chain management: review and research directions. International Journal of Production Research 63, 1851–1872. https://doi.org/10.1080/00207543.2024.2372655

Guo, W., Zhuang, F., Zhang, X., Tong, Y., Dong, J., 2024. A comprehensive survey of federated transfer learning: challenges, methods and applications. Front. Comput. Sci. 18, 186356. https://doi.org/10.1007/s11704-024-40065-x

Haddad, T.A., Hedjazi, D., Aouag, S., 2022. A deep reinforcement learning-based cooperative approach for multi-intersection traffic signal control. Engineering Applications of Artificial Intelligence 114, 105019. https://doi.org/10.1016/j.engappai.2022.105019

Hale, C., 2025. Global AI adoption to push IT spending beyond $5.4 trillion in 2025.

Harper, C.D., Hendrickson, C.T., Mangones, S., Samaras, C., 2016. Estimating potential increases in travel with autonomous vehicles for the non-driving, elderly and people with travel-restrictive medical conditions. Transportation Research Part C: Emerging Technologies 72, 1–9. https://doi.org/10.1016/j.trc.2016.09.003

He, D., Zou, Z., Chen, Y., Liu, B., Yao, X., Shan, S., 2021. Obstacle detection of rail transit based on deep learning. Measurement 176, 109241. https://doi.org/10.1016/j.measurement.2021.109241

Herer, Y.T., Shalom, L., 2000. The Kanban assignment problem – A non-integral approach. European Journal of Operational Research 120, 260–276. https://doi.org/10.1016/S0377-2217(99)00155-1

Ho, T.-H., Lim, N., Reza, S., Xia, X., 2017. OM Forum—Causal Inference Models in Operations Management. M&SOM 19, 509–525. https://doi.org/10.1287/msom.2017.0659

Hockstein, N.G., Gourin, C.G., Faust, R.A., Terris, D.J., 2007. A history of robots: from science fiction to surgical robotics. J Robotic Surg 1, 113–118. https://doi.org/10.1007/s11701-007-0021-2

Honglin, Z., Yaohua, W., Jinchang, H., Yanyan, W., 2023. Collaborative optimization of task scheduling and multi-agent path planning in automated warehouses. Complex Intell. Syst. 9, 5937–5948. https://doi.org/10.1007/s40747-023-01023-5

Huang, J., Huang, L., Liu, M., Li, H., Tan, Q., Ma, X., Cui, J., Huang, D.-S., 2022. Deep Reinforcement Learning-based Trajectory Pricing on Ride-hailing Platforms. ACM Trans. Intell. Syst. Technol. 13, 1–19. https://doi.org/10.1145/3474841

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., Liu, T., 2025. A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. ACM Trans. Inf. Syst. 43, 1–55. https://doi.org/10.1145/3703155

Huang, S., Yang, K., Qi, S., Wang, R., 2024. When large language model meets optimization. Swarm and Evolutionary Computation 90, 101663. https://doi.org/10.1016/j.swevo.2024.101663

Idrissi, Z.K., Lachgar, M., Hrimech, H., 2024. Blockchain, IoT and AI in logistics and transportation: A systematic review. Transport Economics and Management 2, 275–285. https://doi.org/10.1016/j.team.2024.09.002

Introduction to Digital Twin Technology in Manufacturing with Unprecedented Efficiency and Innovation, 2024.

Isbell, Z., 2025. A Guide to Autonomous Mobile Robots (AMRs). URL https://www.autostoresystem.com/insights/a-guide-to-autonomous-mobile-robots-amrs

ITF, 2021. Container Port Automation Impacts and Implications.

Ivanov, D., Dolgui, A., 2021. A digital supply chain twin for managing the disruption risks and resilience in the era of Industry 4.0. Production Planning & Control 32, 775–788. https://doi.org/10.1080/09537287.2020.1768450

Jahanshahi, H., Zhu, Z.H., 2024. Review of machine learning in robotic grasping control in space application. Acta Astronautica 220, 37–61. https://doi.org/10.1016/j.actaastro.2024.04.012

Jamarani, A., Haddadi, S., Sarvizadeh, R., Haghi Kashani, M., Akbari, M., Moradi, S., 2024. Big data and predictive analytics: A systematic review of applications. Artif Intell Rev 57, 176. https://doi.org/10.1007/s10462-024-10811-5

Jiang, P., Sonne, C., Li, W., You, F., You, S., 2024. Preventing the Immense Increase in the Life-Cycle Energy and Carbon Footprints of LLM-Powered Intelligent Chatbots. Engineering 40, 202–210. https://doi.org/10.1016/j.eng.2024.04.002

Jianjun Zhou, J.Z., 2024. Optimization Algorithm of Intelligent Warehouse Management System Based on Reinforcement Learning. jes 20, 219–231. https://doi.org/10.52783/jes.678

Jonker, T., Duinkerken, M.B., Yorke-Smith, N., De Waal, A., Negenborn, R.R., 2021. Coordinated optimization of equipment operations in a container terminal. Flex Serv Manuf J 33, 281–311. https://doi.org/10.1007/s10696-019-09366-3

Jover, V., Sempere, S., 2025. Creating Digital Twins to Celebrate Commemorative Events in the Metaverse. Computers 14, 273. https://doi.org/10.3390/computers14070273

Kairouz, P., McMahan, H.B., Avent, B., Bellet, A., Bennis, M., Nitin Bhagoji, A., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R.G.L., Eichner, H., El Rouayheb, S., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P.B., Gruteser, M., Harchaoui, Z., He, C., He, L., Huo, Z., Hutchinson, B., Hsu, J., Jaggi, M., Javidi, T., Joshi, G., Khodak, M., Konecný, J., Korolova, A., Koushanfar, F., Koyejo, S., Lepoint, T., Liu, Y., Mittal, P., Mohri, M., Nock, R., Özgür, A., Pagh, R., Qi, H., Ramage, D., Raskar, R., Raykova, M., Song, D., Song, W., Stich, S.U., Sun, Z., Suresh, A.T., Tramèr, F., Vepakomma, P., Wang, J., Xiong, L., Xu, Z., Yang, Q., Yu, F.X., Yu, H., Zhao, S., 2021. Advances and Open Problems in Federated Learning. FNT in Machine Learning 14, 1–210. https://doi.org/10.1561/2200000083

Kansal, L., Ediga, P., 2024. IoT-Enabled predictive maintenance for sustainable transportation fleets. MATEC Web Conf. 392, 01189. https://doi.org/10.1051/matecconf/202439201189

Kaplan, J., McCandlish, S., Henighan, T., Brown, T.B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., Amodei, D., 2020. Scaling Laws for Neural Language Models. https://doi.org/10.48550/ARXIV.2001.08361

Kassens-Noor, E., Cai, M., Kotval-Karamchandani, Z., Decaminada, T., 2021. Autonomous vehicles and mobility for people with special needs. Transportation Research Part A: Policy and Practice 150, 385–397. https://doi.org/10.1016/j.tra.2021.06.014

Ke, Z., Zou, Q., Liu, J., Qian, S., 2025. Real-time system optimal traffic routing under uncertainties — Can physics models boost reinforcement learning? Transportation Research Part C: Emerging Technologies 173, 105040. https://doi.org/10.1016/j.trc.2025.105040

Keith, R., La, H.M., 2024. Review of Autonomous Mobile Robots for the Warehouse Environment. https://doi.org/10.48550/arXiv.2406.08333

Kejingyun, S., Mingjun, R., 2025. Randomized Controlled Study on the Impact of Problem-Based Learning Combined With Large Language Models on Critical Thinking Skills in Nursing Students. Nurse Educator 50, 216–220. https://doi.org/10.1097/NNE.0000000000001879

Khan, M.G., Huda, N.U., Zaman, U.K.U., 2022. Smart Warehouse Management System: Architecture, Real-Time Implementation and Prototype Design. Machines 10, 150. https://doi.org/10.3390/machines10020150

Kim, S., Truong, V.-D., Lee, K.-H., Yoon, J., 2024. Revolutionizing Robotic Depalletizing: AI-Enhanced Parcel Detecting with Adaptive 3D Machine Vision and RGB-D Imaging for Automated Unloading. Sensors 24, 1473. https://doi.org/10.3390/s24051473

Kiran, P., 2024. Leveraging AI for Enhanced Inventory Management in 2024. URL https://hypersonix.ai/blogs/leveraging-ai-for-enhanced-inventory-management-in-2024/

Kleeberger, K., Bormann, R., Kraus, W., Huber, M.F., 2020. A Survey on Learning-Based Robotic Grasping. Curr Robot Rep 1, 239–249. https://doi.org/10.1007/s43154-020-00021-6

Koopman, P., Wagner, M., 2017. Autonomous Vehicle Safety: An Interdisciplinary Challenge. IEEE Intell. Transport. Syst. Mag. 9, 90–96. https://doi.org/10.1109/MITS.2016.2583491

Kosacka-Olejnik, M., Kostrzewski, M., Marczewska, M., Mrówczyńska, B., Pawlewski, P., 2021. How Digital Twin Concept Supports Internal Transport Systems?—Literature Review. Energies 14, 4919. https://doi.org/10.3390/en14164919

Lagorio, A., Cimini, C., Pinto, R., Cavalieri, S., 2023. 5G in Logistics 4.0: potential applications and challenges. Procedia Computer Science 217, 650–659. https://doi.org/10.1016/j.procs.2022.12.261

Lai, X., Yang, Z., Xie, J., Liu, Y., 2024. Reinforcement learning in transportation research: Frontiers and future directions. Multimodal Transportation 3, 100164. https://doi.org/10.1016/j.multra.2024.100164

Lamballais, T., Roy, D., De Koster, M.B.M., 2017. Estimating performance in a Robotic Mobile Fulfillment System. European Journal of Operational Research 256, 976–990. https://doi.org/10.1016/j.ejor.2016.06.063

Lambert, D.M., Cooper, M.C., 2000. Issues in Supply Chain Management. Industrial Marketing Management 29, 65–83. https://doi.org/10.1016/S0019-8501(99)00113-3

Larson, P.D., Halldorsson, A., 2004. Logistics versus supply chain management: An international survey. International Journal of Logistics Research and Applications 7, 17–31. https://doi.org/10.1080/13675560310001619240

Le, T.V., Fan, R., 2024. Digital twins for logistics and supply chain systems: Literature review, conceptual framework, research potential, and practical challenges. Computers & Industrial Engineering 187, 109768. https://doi.org/10.1016/j.cie.2023.109768

Le-Anh, T., De Koster, M.B.M., 2006. A review of design and control of automated guided vehicle systems. European Journal of Operational Research 171, 1–23. https://doi.org/10.1016/j.ejor.2005.01.036

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539

Lee, H.L., 2004. The Triple-A Supply Chain.

Lei, C., Zhang, H., Wang, Z., Miao, Q., 2025. Deep Learning for Demand Forecasting: A Framework Incorporating Variational Mode Decomposition and Attention Mechanism. Processes 13, 594. https://doi.org/10.3390/pr13020594

Li, C., Du, D., Zhang, L., Luo, T., Wu, Y., Tian, Q., Wen, L., Lyu, S., 2019. Data Priming Network for Automatic Check-Out, in: Proceedings of the 27th ACM International Conference on Multimedia. Presented at the MM ’19: The 27th ACM International Conference on Multimedia, ACM, Nice France, pp. 2152–2160. https://doi.org/10.1145/3343031.3350989

Li, C., Guo, S., Zhou, Q., Xu, C., Chen, G., 2024. Lower Limits of Petrophysical Properties Allowing Natural Gas Accumulation in Marine Sandstones: An Example from the Qiongdongnan Basin, Northern South China Sea. JMSE 12, 735. https://doi.org/10.3390/jmse12050735

Li, H., Ge, L., Tian, L., 2024. Survey: federated learning data security and privacy-preserving in edge-Internet of Things. Artif Intell Rev 57, 130. https://doi.org/10.1007/s10462-024-10774-7

Li, L., Li, Y., Liu, R., Zhou, Y., Pan, E., 2023. A Two-stage Stochastic Programming for AGV scheduling with random tasks and battery swapping in automated container terminals. Transportation Research Part E: Logistics and Transportation Review 174, 103110. https://doi.org/10.1016/j.tre.2023.103110

Liang, Y., Wen, H., Nie, Y., Jiang, Y., Jin, M., Song, D., Pan, S., Wen, Q., 2024. Foundation Models for Time Series Analysis: A Tutorial and Survey, in: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Presented at the KDD ’24: The 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, ACM, Barcelona Spain, pp. 6555–6565. https://doi.org/10.1145/3637528.3671451

Lighthill, J., 1973. Lighthill Report: Artificial Intelligence: a paper symposium.

Lim, B., Arık, S.Ö., Loeff, N., Pfister, T., 2021. Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting 37, 1748–1764. https://doi.org/10.1016/j.ijforecast.2021.03.012

Lim, J., Masoud, N., 2024. Dynamic Usage Allocation and Pricing for Curb Space Operation. Transportation Science 58, 1252–1276. https://doi.org/10.1287/trsc.2024.0507

Lin, K., Zhao, R., Xu, Z., Zhou, J., 2018. Efficient Large-Scale Fleet Management via Multi-Agent Deep Reinforcement Learning, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Presented at the KDD '18: The 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, London United Kingdom, pp. 1774–1783. https://doi.org/10.1145/3219819.3219993

Liu, Y., Wang, S., Xie, Y., Xiong, T., Wu, M., 2024. A Review of Sensing Technologies for Indoor Autonomous Mobile Robots. Sensors 24, 1222. https://doi.org/10.3390/s24041222

Lu, Z., Dai, N., Hu, X., Xu, K., Yuan, Y., 2024. Research on high-speed classification and location algorithm for logistics parcels based on a monocular camera. Sci Rep 14, 15901. https://doi.org/10.1038/s41598-024-66941-x

Lubna, Mufti, N., Shah, S.A.A., 2021. Automatic Number Plate Recognition:A Detailed Survey of Relevant Algorithms. Sensors 21, 3028. https://doi.org/10.3390/s21093028

Lummus, R.R., Krumwiede, D.W., Vokurka, R.J., 2001. The relationship of logistics to supply chain management: developing a common industry definition. Industrial Management & Data Systems 101, 426–432. https://doi.org/10.1108/02635570110406730

Luo, H., Wei, J., Zhao, S., Liang, A., Xu, Z., Jiang, R., 2025. Intelligent logistics management robot path planning algorithm integrating transformer and GCN network. https://doi.org/10.48550/ARXIV.2501.02749

Lv, Y., Zou, M., Li, J., Liu, J., 2024. Dynamic berth allocation under uncertainties based on deep reinforcement learning towards resilient ports. Ocean & Coastal Management 252, 107113. https://doi.org/10.1016/j.ocecoaman.2024.107113

Ma, G., Da Silva, E.R., Møller, C., 2024. Digital Twin-based Demand-Driven Production Planning and Scheduling System: A New Conceptual Framework. Procedia CIRP 130, 705–710. https://doi.org/10.1016/j.procir.2024.10.152

Mahale, Y., Kolhar, S., More, A.S., 2025. A comprehensive review on artificial intelligence driven predictive maintenance in vehicles: technologies, challenges and future research directions. Discov Appl Sci 7, 243. https://doi.org/10.1007/s42452-025-06681-3

Maheshwari, S., Gautam, P., Jaggi, C.K., 2021. Role of Big Data Analytics in supply chain management: current trends and future perspectives. International Journal of Production Research 59, 1875–1900. https://doi.org/10.1080/00207543.2020.1793011

Mahmoudinazlou, S., Sobhanan, A., Charkhgard, H., Eshragh, A., Dunn, G., 2025. Deep reinforcement learning for dynamic order picking in warehouse operations. Computers & Operations Research 182, 107112. https://doi.org/10.1016/j.cor.2025.107112

Malang, C., Charoenkwan, P., Wudhikarn, R., 2023. Implementation and Critical Factors of Unmanned Aerial Vehicle (UAV) in Warehouse Management: A Systematic Literature Review. Drones 7, 80. https://doi.org/10.3390/drones7020080

Marnissi, M.A., Fradi, H., Sahbani, A., Essoukri Ben Amara, N., 2023. Improved domain adaptive object detector via adversarial feature learning. Computer Vision and Image Understanding 230, 103660. https://doi.org/10.1016/j.cviu.2023.103660

Masanta, M., Singh, U.P., Selvan, K., Das, P., 2025. Optimization of green logistics network in a multi-echelon multi-product supply chain. Journal of Industrial and Production Engineering 1–21. https://doi.org/10.1080/21681015.2025.2509909

McCarthy, J., 2007. WHAT IS ARTIFICIAL INTELLIGENCE?

McCarthy, J., Minsky, M., Rochester, N., Shannon CE, 1955. A proposal for the Dartmouth summer research project on artificial intelligence.

McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics 5, 115–133. https://doi.org/10.1007/BF02478259

Mei, L., Mo, S., Yang, Z., Chen, C., 2025. A Survey of Multimodal Retrieval-Augmented Generation. https://doi.org/10.48550/ARXIV.2504.08748

Melek, C.G., Battini Sönmez, E., Varlı, S., 2024. Datasets and methods of product recognition on grocery shelf images using computer vision and machine learning approaches: An exhaustive literature review. Engineering Applications of Artificial Intelligence 133, 108452. https://doi.org/10.1016/j.engappai.2024.108452

Mentzer, J.T., DeWitt, W., Keebler, J.S., Min, S., Nix, N.W., Smith, C.D., Zacharia, Z.G., 2001. DEFINING SUPPLY CHAIN MANAGEMENT. J of Business Logistics 22, 1–25. https://doi.org/10.1002/j.2158-1592.2001.tb00001.x

Milakis, D., Van Arem, B., Van Wee, B., 2017. Policy and society related implications of automated driving: A review of literature and directions for future research. Journal of Intelligent Transportation Systems 21, 324–348. https://doi.org/10.1080/15472450.2017.1291351

Minsky, M., Papert, S.A., 2017. Perceptrons: An Introduction to Computational Geometry. The MIT Press. https://doi.org/10.7551/mitpress/11301.001.0001

Mittal, P., 2024. A comprehensive survey of deep learning-based lightweight object detection models for edge devices. Artif Intell Rev 57, 242. https://doi.org/10.1007/s10462-024-10877-1

Mo, Y., Wang, L., Hong, W., Chu, C., Li, P., Xia, H., 2024. Small-Scale Foreign Object Debris Detection Using Deep Learning and Dual Light Modes. Applied Sciences 14, 2162. https://doi.org/10.3390/app14052162

Munim, Z.H., Dushenko, M., Jimenez, V.J., Shakil, M.H., Imset, M., 2020. Big data and artificial intelligence in the maritime industry: a bibliometric review and future research directions. Maritime Policy & Management 47, 577–597. https://doi.org/10.1080/03088839.2020.1788731

Munyaka, J.-C.B., Yadavalli, S.V., 2022. INVENTORY MANAGEMENT CONCEPTS AND IMPLEMENTATIONS: A SYSTEMATIC REVIEW. SAJIE 32. https://doi.org/10.7166/33-2-2527

Musen, M.A., Van Der Lei, J., 1988. Of Brittleness and Bottlenecks: Challenges in the Creation of Pattern-Recognition and Expert-System Models, in: Machine Intelligence and Pattern Recognition. Elsevier, pp. 335–352. https://doi.org/10.1016/B978-0-444-87137-4.50029-1

Mustajab, D., 2023. Effective Production Planning and Scheduling: Literature Review. amar 1, 54–72. https://doi.org/10.70184/h9wy6f46

Nam, S., Logendran, R., 1992. Aggregate production planning — A survey of models and methodologies. European Journal of Operational Research 61, 255–272. https://doi.org/10.1016/0377-2217(92)90356-E

Nasien, D., Yuhaniz, S.S., Haron, H., 2010. Statistical Learning Theory and Support Vector Machines, in: 2010 Second International Conference on Computer Research and Development. Presented at the

2010 Second International Conference on Computer Research and Development, IEEE, Kuala Lumpur, Malaysia, pp. 760–764. https://doi.org/10.1109/ICCRD.2010.183

Naz, F., Agrawal, R., Kumar, A., Gunasekaran, A., Majumdar, A., Luthra, S., 2022. Reviewing the applications of artificial intelligence in sustainable supply chains: Exploring research propositions for future directions. Bus Strat Env 31, 2400–2423. https://doi.org/10.1002/bse.3034

Neamatian Monemi, R., Gelareh, S., González, P.H., Cui, L., Bouamrane, K., Dai, Y.-H., Maculan, N., 2025. Graph Convolutional Networks for logistics optimization: A survey of scheduling and operational applications. Transportation Research Part E: Logistics and Transportation Review 197, 104083. https://doi.org/10.1016/j.tre.2025.104083

Nesheli, M., Li, L., Palm, M., Shalaby, A., 2021. Driverless shuttle pilots: Lessons for automated transit technology deployment. Case Studies on Transport Policy 9, 723–742. https://doi.org/10.1016/j.cstp.2021.03.010

Newell, A., Simon, H., 1956. The logic theory machine--A complex information processing system. IEEE Trans. Inform. Theory 2, 61–79. https://doi.org/10.1109/TIT.1956.1056797

Nguyen, L.V., 2024. Swarm Intelligence-Based Multi-Robotics: A Comprehensive Review. AppliedMath 4, 1192–1210. https://doi.org/10.3390/appliedmath4040064

Nguyen, N.-A.-T., Wang, C.-N., Dang, T.-T., 2025. Advanced Process Optimization in Logistics and Supply Chain Management. Processes 13, 1864. https://doi.org/10.3390/pr13061864

Nikseresht, A., Shokouhyar, S., Tirkolaee, E.B., Shokoohyar, S., Pishva, N., Simic, V., 2025. Machine learning and data-driven models in sustainable supply chain management: a systematic literature review with content analysis. International Journal of Logistics Research and Applications 1–40. https://doi.org/10.1080/13675567.2025.2517639

Noaeen, M., Naik, A., Goodman, L., Crebo, J., Abrar, T., Abad, Z.S.H., Bazzan, A.L.C., Far, B., 2022. Reinforcement learning in urban network traffic signal control: A systematic literature review. Expert Systems with Applications 199, 116830. https://doi.org/10.1016/j.eswa.2022.116830

Noruzoliaee, M., Zou, B., Zhou, Y. (Joann), 2021. Truck platooning in the U.S. national road network: A system-level modeling approach. Transportation Research Part E: Logistics and Transportation Review 145, 102200. https://doi.org/10.1016/j.tre.2020.102200

Ocado Technology, 2015. Ocado Technology leads development of one of the world's most advanced collaborative robots. URL https://www.ocadogroup.com/media/news/ocado-technology-leads-development-one-worlds-most-advanced

Ojeda Rios, B.H., Xavier, E.C., Miyazawa, F.K., Amorim, P., Curcio, E., Santos, M.J., 2021. Recent dynamic vehicle routing problems: A survey. Computers & Industrial Engineering 160, 107604. https://doi.org/10.1016/j.cie.2021.107604

O'Keeffe, S., 2025. 5 ways manufacturers can embark on their AI journey. URL https://www.techradar.com/pro/5-ways-manufacturers-can-embark-on-their-ai-journey

Olamide Raimat Amosu, Praveen Kumar, Yewande Mariam Ogunsuji, Segun Oni, Oladapo Faworaja, 2024. AI-driven demand forecasting: Enhancing inventory management and customer satisfaction. World J. Adv. Res. Rev. 23, 708–719. https://doi.org/10.30574/wjarr.2024.23.2.2394

Ouassou, E.H., Onyeaka, H., Tamasiga, P., Bakwena, M., 2024. Carbon transparency in global supply chains: The mediating role of institutional and innovative capacity. Energy Strategy Reviews 53, 101405. https://doi.org/10.1016/j.esr.2024.101405

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C.L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., Lowe, R., 2022. Training language models to follow instructions with human feedback. https://doi.org/10.48550/ARXIV.2203.02155

Oza, P., Sindagi, V.A., VS, V., Patel, V.M., 2021. Unsupervised Domain Adaptation of Object Detectors: A Survey. https://doi.org/10.48550/ARXIV.2105.13502

Paaß, G., Hecker, D., 2024. Artificial Intelligence: What Is Behind the Technology of the Future? Springer Nature Switzerland, Cham. https://doi.org/10.1007/978-3-031-50605-5

Paleyes, A., Urma, R.-G., Lawrence, N.D., 2023. Challenges in Deploying Machine Learning: A Survey of Case Studies. ACM Comput. Surv. 55, 1–29. https://doi.org/10.1145/3533378

Pandya, S., Srivastava, G., Jhaveri, R., Babu, M.R., Bhattacharya, S., Maddikunta, P.K.R., Mastorakis, S., Piran, Md.J., Gadekallu, T.R., 2023. Federated learning for smart cities: A comprehensive survey. Sustainable Energy Technologies and Assessments 55, 102987. https://doi.org/10.1016/j.seta.2022.102987

Parvez Farazi, N., Zou, B., Ahamed, T., Barua, L., 2021. Deep reinforcement learning in transportation research: A review. Transportation Research Interdisciplinary Perspectives 11, 100425. https://doi.org/10.1016/j.trip.2021.100425

Paul, J., Alli, D.O., Adegbola, J., n.d. AI-Powered Route Optimization Reducing Costs and Improving Delivery Efficiency.

Pegado-Bardayo, A., Lorenzo-Espejo, A., Muñuzuri, J., Onieva, L., 2024. A predictive framework for last-mile delivery routes considering couriers' behavior heterogeneity. Computers & Industrial Engineering 198, 110665. https://doi.org/10.1016/j.cie.2024.110665

Peng, Y., Lu, S., Chen, F., Liu, X., Tian, Z., 2024. Energy-efficient train control incorporating inherent reduced-power and hybrid braking characteristics of railway vehicles. Transportation Research Part C: Emerging Technologies 163, 104626. https://doi.org/10.1016/j.trc.2024.104626

Perera, S.C., Sethi, S.P., 2023. A survey of stochastic inventory models with fixed costs: Optimality of (*s* , *S*) and (*s* , *S*)-type policies—Continuous-time case. Production and Operations Management 32, 154–169. https://doi.org/10.1111/poms.13819

Petit, J., Shladover, S.E., 2014. Potential Cyberattacks on Automated Vehicles. IEEE Trans. Intell. Transport. Syst. 1–11. https://doi.org/10.1109/TITS.2014.2342271

Petropoulos, F., Makridakis, S., Assimakopoulos, V., Nikolopoulos, K., 2014. 'Horses for Courses' in demand forecasting. European Journal of Operational Research 237, 152–163. https://doi.org/10.1016/j.ejor.2014.02.036

Pettersson, T., Riveiro, M., Löfström, T., 2024. Multimodal fine-grained grocery product recognition using image and OCR text. Machine Vision and Applications 35, 79. https://doi.org/10.1007/s00138-024-01549-9

Pietrini, R., Paolanti, M., Mancini, A., Frontoni, E., Zingaretti, P., 2024. Shelf Management: A deep learning-based system for shelf visual monitoring. Expert Systems with Applications 255, 124635. https://doi.org/10.1016/j.eswa.2024.124635

Pillac, V., Gendreau, M., Guéret, C., Medaglia, A.L., 2013. A review of dynamic vehicle routing problems. European Journal of Operational Research 225, 1–11. https://doi.org/10.1016/j.ejor.2012.08.015

Pournader, M., Ghaderi, H., Hassanzadegan, A., Fahimnia, B., 2021. Artificial intelligence applications in supply chain management. International Journal of Production Economics 241, 108250. https://doi.org/10.1016/j.ijpe.2021.108250

Qin, Z. (Tony), Tang, X., Jiao, Y., Zhang, F., Xu, Z., Zhu, H., Ye, J., 2020. Ride-Hailing Order Dispatching at DiDi via Reinforcement Learning. INFORMS Journal on Applied Analytics 50, 272–286. https://doi.org/10.1287/inte.2020.1047

R Azmi, P.A., Yusoff, M., Mohd Sallehud-din, M.T., 2024. A Review of Predictive Analytics Models in the Oil and Gas Industries. Sensors 24, 4013. https://doi.org/10.3390/s24124013

Rad, F.F., Oghazi, P., Onur, İ., Kordestani, A., 2025. Adoption of AI-based order picking in warehouse: benefits, challenges, and critical success factors. Rev Manag Sci. https://doi.org/10.1007/s11846-025-00858-1

Radanliev, P., De Roure, D., Page, K., Nurse, J.R.C., Mantilla Montalvo, R., Santos, O., Maddox, L., Burnap, P., 2020. Cyber risk at the edge: current and future trends on cyber risk analytics and artificial intelligence in the industrial internet of things and industry 4.0 supply chains. Cybersecur 3, 13. https://doi.org/10.1186/s42400-020-00052-8

Rafi, T.H., Noor, F.A., Hussain, T., Chae, D.-K., 2024. Fairness and privacy preserving in federated learning: A survey. Information Fusion 105, 102198. https://doi.org/10.1016/j.inffus.2023.102198

Rasheed, F., Yau, K.-L.A., Noor, R.Md., Wu, C., Low, Y.-C., 2020. Deep Reinforcement Learning for Traffic Signal Control: A Review. IEEE Access 8, 208016–208044. https://doi.org/10.1109/ACCESS.2020.3034141

Ren, S., Tomlinson, B., Black, R.W., Torrance, A.W., 2024. Reconciling the contrasting narratives on the environmental impact of large language models. Sci Rep 14, 26310. https://doi.org/10.1038/s41598-024-76682-6

Riachy, C., He, M., Joneidy, S., Qin, S., Payne, T., Boulton, G., Occhipinti, A., Angione, C., 2025. Enhancing deep learning for demand forecasting to address large data gaps. Expert Systems with Applications 268, 126200. https://doi.org/10.1016/j.eswa.2024.126200

Ristić-Durrant, D., Franke, M., Michels, K., 2021. A Review of Vision-Based On-Board Obstacle Detection and Distance Estimation in Railways. Sensors 21, 3452. https://doi.org/10.3390/s21103452

Ristvej, J., Kubás, J., Kollár, B., 2024. METHODOLOGICAL PROCEDURE FOR CASE STUDIES IN CRISIS MANAGEMENT 12, 60–68.

Robinson, P., Narayanan, A., Sahin, F., 2009. Coordinated deterministic dynamic demand lot-sizing problem: A review of models and algorithms☆. Omega 37, 3–15. https://doi.org/10.1016/j.omega.2006.11.004

Rockafellar, R.T., Royset, J.O., Miranda, S.I., 2014. Superquantile regression with applications to buffered reliability, uncertainty quantification, and conditional value-at-risk. European Journal of Operational Research 234, 140–154. https://doi.org/10.1016/j.ejor.2013.10.046

Rolf, B., Jackson, I., Müller, M., Lang, S., Reggelin, T., Ivanov, D., 2023. A review on reinforcement learning algorithms and applications in supply chain management. International Journal of Production Research 61, 7151–7179. https://doi.org/10.1080/00207543.2022.2140221

Roman, E.-A., Stere, A.-S., Roșca, E., Radu, A.-V., Codroiu, D., Anamaria, I., 2025. State of the Art of Digital Twins in Improving Supply Chain Resilience. Logistics 9, 22. https://doi.org/10.3390/logistics9010022

Roth, E., 2025. OpenAI says ChatGPT users send over 2.5 billion prompts every day.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533–536. https://doi.org/10.1038/323533a0

Rushton, A., Croucher, P., Baker, P., 2014. The handbook of logistics & distribution management, 5th edition. ed. Kogan Page, London ; Philadelphia.

Sadiku, M.N.O., Musa, S.M., 2021. Artificial Intelligence, in: A Primer on Multiple Intelligences. Springer International Publishing, Cham, pp. 133–150. https://doi.org/10.1007/978-3-030-77584-1_11

Sajadieh, S.M.M., Noh, S.D., 2025. From Simulation to Autonomy: Reviews of the Integration of Artificial Intelligence and Digital Twins. Int. J. of Precis. Eng. and Manuf.-Green Tech. https://doi.org/10.1007/s40684-025-00750-z

Salinas, D., Flunkert, V., Gasthaus, J., Januschowski, T., 2020. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. International Journal of Forecasting 36, 1181–1191. https://doi.org/10.1016/j.ijforecast.2019.07.001

Samani, A.R., Talebian, A., Mishra, S., Golias, M., 2025. Evaluating consumer shopping, delivery demands, and last-mile preferences: An integrated MDCEV-HCM approach. Transportation Research Part E: Logistics and Transportation Review 197, 104067. https://doi.org/10.1016/j.tre.2025.104067

Samuels, A., 2025. Examining the integration of artificial intelligence in supply chain management from Industry 4.0 to 6.0: a systematic literature review. Front. Artif. Intell. 7, 1477044. https://doi.org/10.3389/frai.2024.1477044

Sánchez-Ibáñez, J.R., Pérez-del-Pulgar, C.J., García-Cerezo, A., 2021. Path Planning for Autonomous Mobile Robots: A Review. Sensors 21, 7898. https://doi.org/10.3390/s21237898

Sani, S., Zarifnia, A., Salonitis, K., Milisavljevic-Syed, J., 2024. Supply Chain 4.0 and the Digital Twin Approach: A Framework for Improving Supply Chain Visibility. Procedia CIRP 128, 321–326. https://doi.org/10.1016/j.procir.2024.03.014

Santra, B., Mukherjee, D.P., 2019. A comprehensive survey on computer vision based approaches for automatic identification of products in retail store. Image and Vision Computing 86, 45–63. https://doi.org/10.1016/j.imavis.2019.03.005

Sarac, A., Absi, N., Dauzère-Pérès, S., 2010. A literature review on the impact of RFID technologies on supply chain management. International Journal of Production Economics 128, 77–95. https://doi.org/10.1016/j.ijpe.2010.07.039

Sarp, S., Kuzlu, M., Jovanovic, V., Polat, Z., Guler, O., 2024. Digitalization of railway transportation through AI-powered services: digital twin trains. Eur. Transp. Res. Rev. 16, 58. https://doi.org/10.1186/s12544-024-00679-5

Shahsavari, M., Hussain, O.K., Sharma, P., Saberi, M., 2025. Modelling supply chain risk events by considering their contributing events: a systematic literature review. Enterprise Information Systems 19, 2472303. https://doi.org/10.1080/17517575.2025.2472303

Shahzad, M.M., Saeed, Z., Akhtar, A., Munawar, H., Yousaf, M.H., Baloach, N.K., Hussain, F., 2023. A Review of Swarm Robotics in a NutShell. Drones 7, 269. https://doi.org/10.3390/drones7040269

Shan, J., Miccinesi, L., Beni, A., Pagnini, L., Cioncolini, A., Pieraccini, M., 2025. A Review of Foreign Object Debris Detection on Airport Runways: Sensors and Algorithms. Remote Sensing 17, 225. https://doi.org/10.3390/rs17020225

Shao, Y., Fan, Z., Zhu, B., Lu, J., Lang, Y., 2023. A Point Cloud Data-Driven Pallet Pose Estimation Method Using an Active Binocular Vision Sensor. Sensors 23, 1217. https://doi.org/10.3390/s23031217

Singh, R., Gill, S.S., 2023. Edge AI: A survey. Internet of Things and Cyber-Physical Systems 3, 71–92. https://doi.org/10.1016/j.iotcps.2023.02.004

Soeffker, N., Ulmer, M.W., Mattfeld, D.C., 2022. Stochastic dynamic vehicle routing in the light of prescriptive analytics: A review. European Journal of Operational Research 298, 801–820. https://doi.org/10.1016/j.ejor.2021.07.014

Soori, M., Jough, F.K.G., Dastres, R., Arezoo, B., 2024. AI-Based Decision Support Systems in Industry 4.0, A Review. Journal of Economy and Technology S2949948824000374. https://doi.org/10.1016/j.ject.2024.08.005

Southern, R.N., 2011. Historical Perspective of the Logistics and Supply Chain Management Discipline. Transportation Journal 50, 53–64. https://doi.org/10.5325/transportationj.50.1.0053

Stephens, S., 2001. Supply Chain Operations Reference Model Version 5.0: A New Tool to Improve Supply Chain Efficiency and Achieve Best Practice. Information Systems Frontiers 3, 471–476. https://doi.org/10.1023/A:1012881006783

Stern, R.E., Cui, S., Delle Monache, M.L., Bhadani, R., Bunting, M., Churchill, M., Hamilton, N., Haulcy, R., Pohlmann, H., Wu, F., Piccoli, B., Seibold, B., Sprinkle, J., Work, D.B., 2018. Dissipation of stop-and-go waves via control of autonomous vehicles: Field experiments. Transportation Research Part C: Emerging Technologies 89, 205–221. https://doi.org/10.1016/j.trc.2018.02.005

Stüve, D., Van Der Meer, R., Ali Agha, M.S., Lütke Entrup, M., 2025. Supply chain planning in the food industry: mixed methods research on the adoption of advanced planning systems. Production Planning & Control 1–30. https://doi.org/10.1080/09537287.2025.2508719

Sun, J., Dong, Q., 2023. A Survey on Open-Set Image Recognition. https://doi.org/10.48550/ARXIV.2312.15571

Sun, X., Wu, H., Abdolmaleki, M., Yin, Y., Zou, B., 2021. Investigating the Potential of Truck Platooning for Energy Savings: Empirical Study of the U.S. National Highway Freight Network. Transportation Research Record: Journal of the Transportation Research Board 2675, 784–796. https://doi.org/10.1177/03611981211031231

Sunday, A.A., Omolayo, M.I., Samuel, A.U., Samuel, O.O., Abdulkareem, A., Moses, E.E., Olamma, U.I., 2021. The Role of Production Planning in Enhancing an Efficient Manufacturing System – An Overview. E3S Web Conf. 309, 01002. https://doi.org/10.1051/e3sconf/202130901002

Swapno, S.M.M.R., Nobel, S.N., Meena, P., Meena, V.P., Azar, A.T., Haider, Z., Tounsi, M., 2024. A reinforcement learning approach for reducing traffic congestion using deep Q learning. Sci Rep 14, 30452. https://doi.org/10.1038/s41598-024-75638-0

Talebpour, A., Mahmassani, H.S., 2016. Influence of connected and autonomous vehicles on traffic flow stability and throughput. Transportation Research Part C: Emerging Technologies 71, 143–163. https://doi.org/10.1016/j.trc.2016.07.007

Tan, W.C., Sidhu, M.S., 2022. Review of RFID and IoT integration in supply chain management. Operations Research Perspectives 9, 100229. https://doi.org/10.1016/j.orp.2022.100229

Tang, H., Wu, K., Zhou, J., 2025. Smarter supply chains, stronger resilience? The impact of AI on preparation, response, and recovery. Economics Letters 254, 112488. https://doi.org/10.1016/j.econlet.2025.112488

Tang, R., De Donato, L., Bešinović, N., Flammini, F., Goverde, R.M.P., Lin, Z., Liu, R., Tang, T., Vittorini, V., Wang, Z., 2022. A literature review of Artificial Intelligence applications in railway systems. Transportation Research Part C: Emerging Technologies 140, 103679. https://doi.org/10.1016/j.trc.2022.103679

Tang, S., Zhang, Z., Zhang, Y., Zhou, J., Guo, Y., Liu, S., Guo, S., Li, Y.-F., Ma, L., Xue, Y., Liu, Y., 2023. A Survey on Automated Driving System Testing: Landscapes and Trends. ACM Trans. Softw. Eng. Methodol. 32, 1–62. https://doi.org/10.1145/3579642

Tang, Y., He, H., Wang, Y., Mao, Z., Wang, H., 2023. Multi-modality 3D object detection in autonomous driving: A review. Neurocomputing 553, 126587. https://doi.org/10.1016/j.neucom.2023.126587

Tang, Y., Zakaria, M.A., Younas, M., 2025. Path Planning Trends for Autonomous Mobile Robot Navigation: A Review. Sensors 25, 1206. https://doi.org/10.3390/s25041206

Taylor, S.J., Letham, B., 2018. Forecasting at Scale. The American Statistician 72, 37–45. https://doi.org/10.1080/00031305.2017.1380080

The Rise Of Autonomous Supply Chains: Powered By Cognitive Digital Twins, 2025.

Thomas, D.M., Kleinberg, S., Brown, A.W., Crow, M., Bastian, N.D., Reisweber, N., Lasater, R., Kendall, T., Shafto, P., Blaine, R., Smith, S., Ruiz, D., Morrell, C., Clark, N., 2022. Machine learning modeling practices to support the principles of AI and ethics in nutrition research. Nutr. Diabetes 12, 48. https://doi.org/10.1038/s41387-022-00226-y

Tian, S., Li, L., Li, W., Ran, H., Ning, X., Tiwari, P., 2024. A survey on few-shot class-incremental learning. Neural Networks 169, 307–324. https://doi.org/10.1016/j.neunet.2023.10.039

Toorajipour, R., Sohrabpour, V., Nazarpour, A., Oghazi, P., Fischl, M., 2021. Artificial intelligence in supply chain management: A systematic literature review. Journal of Business Research 122, 502–517. https://doi.org/10.1016/j.jbusres.2020.09.009

Torra, V., Karlsson, A., Steinhauer, H.J., Berglund, S., 2019. Artificial Intelligence, in: Said, A., Torra, V. (Eds.), Data Science in Practice, Studies in Big Data. Springer International Publishing, Cham, pp. 9–26. https://doi.org/10.1007/978-3-319-97556-6_2

Trend report, 2025. TREND REPORT: AI DRIVEN COMPUTER VISION. URL https://www.dhl.com/us-en/microsites/csi/computer-vision/understanding-computer-vision.html

Tseng, Y., Yue, W.L., Taylor, M., 2005. THE ROLE OF TRANSPORTATION IN LOGISTICS CHAIN, in: Proceedings of the Eastern Asia Society for Transportation Studies. pp. 1657–1672.

Turing, A.M., 1950. I.—COMPUTING MACHINERY AND INTELLIGENCE. Mind LIX, 433–460. https://doi.org/10.1093/mind/LIX.236.433

Ulmer, M.W., Goodson, J.C., Mattfeld, D.C., Hennig, M., 2019. Offline–Online Approximate Dynamic Programming for Dynamic Vehicle Routing with Stochastic Requests. Transportation Science 53, 185–202. https://doi.org/10.1287/trsc.2017.0767

Ulmer, M.W., Mattfeld, D.C., Köster, F., 2018. Budgeting Time for Dynamic Vehicle Routing with Stochastic Customer Requests. Transportation Science 52, 20–37. https://doi.org/10.1287/trsc.2016.0719

Vakalopoulou, M., Christodoulidis, S., Burgos, N., Colliot, O., Lepetit, V., 2023. Deep Learning: Basics and Convolutional Neural Networks (CNNs), in: Colliot, O. (Ed.), Machine Learning for Brain Disorders, Neuromethods. Springer US, New York, NY, pp. 77–115. https://doi.org/10.1007/978-1-0716-3195-9_3

Valenzuela, A.M., Noguera, F.B., 2025. Task Allocation in Mobile Robot Fleets: A review. https://doi.org/10.48550/ARXIV.2501.08726

Vargas, V., 2009. An optimal solution for the stochastic version of the Wagner–Whitin dynamic lot-size model. European Journal of Operational Research 198, 447–451. https://doi.org/10.1016/j.ejor.2008.09.003

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention Is All You Need. https://doi.org/10.48550/ARXIV.1706.03762

Villegas-Ch, W., Navarro, A.M., Sanchez-Viteri, S., 2024. Optimization of inventory management through computer vision and machine learning technologies. Intelligent Systems with Applications 24, 200438. https://doi.org/10.1016/j.iswa.2024.200438

Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C., 2014. Short-term traffic forecasting: Where we are and where we're going. Transportation Research Part C: Emerging Technologies 43, 3–19. https://doi.org/10.1016/j.trc.2014.01.005

Wadud, Z., MacKenzie, D., Leiby, P., 2016. Help or hindrance? The travel, energy and carbon impacts of highly automated vehicles. Transportation Research Part A: Policy and Practice 86, 1–18. https://doi.org/10.1016/j.tra.2015.12.001

Walker, S.E., 2003. Active learning strategies to promote critical thinking. J Athl Train 38, 263–267.

Wang, H., Vaze, S., Han, K., 2025. Dissecting Out-of-Distribution Detection and Open-Set Recognition: A Critical Analysis of Methods and Benchmarks. Int J Comput Vis 133, 1326–1351. https://doi.org/10.1007/s11263-024-02222-4

Wang, L., Cao, C., Zou, B., Ye, J., Zhang, J., 2023. License Plate Recognition via Attention Mechanism. Computers, Materials & Continua 75, 1801–1814. https://doi.org/10.32604/cmc.2023.032785

Warwick, K., Shah, H., 2016. Passing the Turing Test Does Not Mean the End of Humanity. Cogn Comput 8, 409–419. https://doi.org/10.1007/s12559-015-9372-6

Wasi, A.T., Islam, M.S., Akib, A.R., Bappy, M.M., 2025. Graph Neural Networks in Supply Chain Analytics and Optimization: Concepts, Perspectives, Dataset and Benchmarks. https://doi.org/10.48550/arXiv.2411.08550

Weber, F., 2023. Artificial Intelligence, in: Artificial Intelligence for Business Analytics. Springer Fachmedien Wiesbaden, Wiesbaden, pp. 33–64. https://doi.org/10.1007/978-3-658-37599-7_2

Wei, G., Zhu, S., Wang, Y., Chen, W., Lu, S., Sha, S., Li, L., 2022. Energy-efficient automatic train operation for high-speed railways: Considering discrete notches and neutral sections. Transportation Research Part C: Emerging Technologies 145, 103884. https://doi.org/10.1016/j.trc.2022.103884

Wei, X.-S., Cui, Q., Yang, L., Wang, P., Liu, L., 2019. RPC: A Large-Scale Retail Product Checkout Dataset. https://doi.org/10.48550/ARXIV.1901.07249

Wei, Y., Tran, S., Xu, S., Kang, B., Springer, M., 2020. Deep Learning for Retail Product Recognition: Challenges and Techniques. Computational Intelligence and Neuroscience 2020, 1–23. https://doi.org/10.1155/2020/8875910

Weizenbaum, J., 1966. ELIZA—a computer program for the study of natural language communication between man and machine. Commun. ACM 9, 36–45. https://doi.org/10.1145/365153.365168

Wilson, M., 2015. Automation System Components, in: Implementation of Robot Systems. Elsevier, pp. 39–73. https://doi.org/10.1016/B978-0-12-404733-4.00003-5

Woschank, M., Rauch, E., Zsifkovits, H., 2020. A Review of Further Directions for Artificial Intelligence, Machine Learning, and Deep Learning in Smart Logistics. Sustainability 12, 3760. https://doi.org/10.3390/su12093760

Wu, G., De Carvalho Servia, M.Á., Mowbray, M., 2023. Reinforcement Learning for inventory management in multi-echelon supply chains, in: Computer Aided Chemical Engineering. Elsevier, pp. 795–800. https://doi.org/10.1016/B978-0-443-15274-0.50127-X

Wu, H., Li, M., Yu, C., Ouyang, Z., Lai, K., Zhao, Z., Pan, S., Wang, S., Zhong, R.Y., Kuo, Y.-H., Zhang, F., Huang, W., Shen, Z.-J.M., Ballot, E., Huang, G.Q., 2025. Towards cyber-physical internet: A systematic review, fundamental model and future perspectives. Transportation Research Part E: Logistics and Transportation Review 197, 104051. https://doi.org/10.1016/j.tre.2025.104051

Wyrembek, M., Baryannis, G., Brintrup, A., 2025. Causal machine learning for supply chain risk prediction and intervention planning. International Journal of Production Research 63, 5629–5648. https://doi.org/10.1080/00207543.2025.2458121

Xu, D., Li, T., Li, Y., Su, X., Tarkoma, S., Jiang, T., Crowcroft, J., Hui, P., 2020. Edge Intelligence: Architectures, Challenges, and Applications. https://doi.org/10.48550/ARXIV.2003.12172

Xu, M., Wang, Y., Xing, D., 2025. Applying RFID and NLP for efficient warehouse picking. International Journal of RF Technologies: Research and Applications 15, 49–64. https://doi.org/10.3233/RFT-230055

Yan, Y., Chow, A.H.F., Ho, C.P., Kuo, Y.-H., Wu, Q., Ying, C., 2022. Reinforcement learning for logistics and supply chain management: Methodologies, state of the art, and future opportunities. Transportation Research Part E: Logistics and Transportation Review 162, 102712. https://doi.org/10.1016/j.tre.2022.102712

Yang, M., Lim, M.K., Qu, Y., Ni, D., Xiao, Z., 2023. Supply chain risk management with machine learning technology: A literature review and future research directions. Computers & Industrial Engineering 175, 108859. https://doi.org/10.1016/j.cie.2022.108859

Yang, X., Hu, H., Cheng, C., Wang, Y., 2023. Automated Guided Vehicle (AGV) Scheduling in Automated Container Terminals (ACTs) Focusing on Battery Swapping and Speed Control. JMSE 11, 1852. https://doi.org/10.3390/jmse11101852

Yang, Y., Liu, Yang, Li, G., Zhang, Z., Liu, Yanbin, 2024. Harnessing the power of Machine learning for AIS Data-Driven maritime Research: A comprehensive review. Transportation Research Part E: Logistics and Transportation Review 183, 103426. https://doi.org/10.1016/j.tre.2024.103426

Yatani, K., Sramek, Z., Yang, C.-L., 2024. AI as Extraherics: Fostering Higher-order Thinking Skills in Human-AI Interaction. https://doi.org/10.48550/ARXIV.2409.09218

Yazeedi, B.M.A., Shakman, L.M.W., Sunderraj, S.E.J., Prabhakaran, H., Arulappan, J., Roach, E.J., Al Hashmi, A., Al Azri, Z., 2024. Perceived efficacy of case analysis as an assessment method for clinical competencies in nursing education: a mixed methods study. BMC Nurs 23, 441. https://doi.org/10.1186/s12912-024-02102-9

Ye, T., Hijazi, A., Van Hentenryck, P., 2025. Conformal Predictive Distributions for Order Fulfillment Time Forecasting. https://doi.org/10.48550/ARXIV.2505.17340

Yeong, D.J., Velasco-Hernandez, G., Barry, J., Walsh, J., 2021. Sensor and Sensor Fusion Technology in Autonomous Vehicles: A Review. Sensors 21, 2140. https://doi.org/10.3390/s21062140

Yin, S., Fu, C., Zhao, S., Li, K., Sun, X., Xu, T., Chen, E., 2024. A survey on multimodal large language models. National Science Review 11, nwae403. https://doi.org/10.1093/nsr/nwae403

Yılmaz, Ö., Son, Y., Shang, G., Arslan, H.A., 2024. Causal inference under selection on observables in operations management research: Matching methods and synthetic controls. J of Ops Management 70, 831–859. https://doi.org/10.1002/joom.1318

Yoon, Y., Ban, K.-D., Yoon, H., Kim, J., 2016. Automatic Container Code Recognition From Multiple Views. ETRI J. https://doi.org/10.4218/etrij.16.0014.0069

Yu, L., Zin, Z.M., 2023. The critical thinking-oriented adaptations of problem-based learning models: a systematic review. Front. Educ. 8, 1139987. https://doi.org/10.3389/feduc.2023.1139987

Yu, M., Zhu, S., Lu, B., Chen, Q., Wang, T., 2024. A Two-Stage Automatic Container Code Recognition Method Considering Environmental Interference. Applied Sciences 14, 4779. https://doi.org/10.3390/app14114779

Zaidi, S.A.H., Khan, S.A., Chaabane, A., 2024. Unlocking the potential of digital twins in supply chains: A systematic review. Supply Chain Analytics 7, 100075. https://doi.org/10.1016/j.sca.2024.100075

Zaidi, S.S.A., Ansari, M.S., Aslam, A., Kanwal, N., Asghar, M., Lee, B., 2022. A survey of modern deep learning based object detection models. Digital Signal Processing 126, 103514. https://doi.org/10.1016/j.dsp.2022.103514

Zhai, C., Wibowo, S., Li, L.D., 2024. The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. Smart Learn. Environ. 11, 28. https://doi.org/10.1186/s40561-024-00316-7

Zhang, H., Lu, G., Zhang, Y., D'Ariano, A., Wu, Y., 2025. Railcar itinerary optimization in railway marshalling yards: A graph neural network based deep reinforcement learning method. Transportation Research Part C: Emerging Technologies 171, 104970. https://doi.org/10.1016/j.trc.2024.104970

Zhang, L., Cai, Z., Yan, Y., Yang, C., Hu, Y., 2024. Multi-agent policy learning-based path planning for autonomous mobile robots. Engineering Applications of Artificial Intelligence 129, 107631. https://doi.org/10.1016/j.engappai.2023.107631

Zhang, L., Chen, F., Ma, X., Pan, X., 2020. Fuel Economy in Truck Platooning: A Literature Overview and Directions for Future Research. Journal of Advanced Transportation 2020, 1–10. https://doi.org/10.1155/2020/2604012

Zhang, X., Mo, T., Zhang, Y., 2023. Optimization of Storage Location Assignment for Non-Traditional Layout Warehouses Based on the Firework Algorithm. Sustainability 15, 10242. https://doi.org/10.3390/su151310242

Zhang, Yunfeng, Li, S., Yuan, Y., Yang, L., 2025. Multi-step look ahead deep reinforcement learning approach for automatic train regulation of urban rail transit lines with energy-saving. Engineering Applications of Artificial Intelligence 145, 110181. https://doi.org/10.1016/j.engappai.2025.110181

Zhang, Yuxing, Lin, Y., Esfahbodi, A., 2025. Digital Transformations of Supply Chain Management via RFID Technology: A Systematic Literature Review. Journal of Digital Economy S2773067025000196. https://doi.org/10.1016/j.jdec.2025.06.001

Zhang, Z., Ding, Y., Li, R., Chen, K., 2024. Enhancing OCR with line segmentation mask for container text recognition in container terminal. Engineering Applications of Artificial Intelligence 133, 108667. https://doi.org/10.1016/j.engappai.2024.108667

Zhao, Y., 2013. R and data mining: examples and case studies, First edition. ed. Academic Press, an imprint of Elsevier, Amsterdam ; Boston.

Zhen, L., Tan, Z., De Koster, R., He, X., Wang, S., Wang, H., 2025. Optimizing Warehouse Operations with Autonomous Mobile Robots. Transportation Science trsc.2024.0800. https://doi.org/10.1287/trsc.2024.0800

Zheng, X., Liang, C., Wang, Y., Shi, J., Lim, G., 2022. Multi-AGV Dynamic Scheduling in an Automated Container Terminal: A Deep Reinforcement Learning Approach. Mathematics 10, 4575. https://doi.org/10.3390/math10234575

Zhu, Z., Wang, S., Wang, T., 2024. Optimizing Robotic Mobile Fulfillment Systems for Order Picking Based on Deep Reinforcement Learning. Sensors 24, 4713. https://doi.org/10.3390/s24144713

Ziegner, G., Choi, M., Le, H.M.C., Sakhuja, S., Sarmadi, A., 2025. Iterative Multi-Agent Reinforcement Learning: A Novel Approach Toward Real-World Multi-Echelon Inventory Optimization. https://doi.org/10.48550/arXiv.2503.18201

Zoting, S., 2025. Artificial Intelligence (AI) Market Size, Share, and Trends 2025 to 2034.