Supplementary Materials

The Role of Large Language Models in Supply Chain Decisions:
A Systematic Review and Research Agenda

Table S1 reports the study-level classifications and TCCM information for all 81 articles. Table S2 describes the evidence from operational applications, and Table S3 reports topic-model stability and parameter checks. Tables S4–S7 provide bibliographic profiles and keyword statistics. The complete database queries and journal ISSNs are provided in Section S1.

Bracketed reference numbers refer to the manuscript bibliography. Groups are A: operational applications (46, including one embedding-only study); B: research measurement (9); C: adoption, analytical or design studies (26). Task and integration coding covers the 55 implemented application and measurement studies. A dash indicates no applicable code; multiple codes can be assigned.

Key to Table S1

Decision area: P = Procurement and suppliers; I = Inventory and production; L = Logistics and network design; R = Risk and resilience; D = Demand and price forecasting; G = Circularity and sustainability; C = Channels and technology contracts; B = Supply chain finance; X = Cross-functional adoption and tools.

Research method: E = Computational and task experiments; A = Econometric and archival studies; S = Survey studies; J = Expert judgement and structural modelling; T = Analytical models; Q = Qualitative cases and content analysis; C = Conceptual frameworks and synthesis.

Model role: A = Generative LLM executed in operational tasks; R = LLM for research measurement or coding; E = Text embeddings only; S = Adoption or organisational effects; no LLM run; T = Technology assumptions in analytical models; C = Proposed technology in conceptual frameworks.

Implemented task: a = Evidence extraction; b = Risk / sentiment scoring; c = Text generation / synthesis; d = Semantic / forecast features; e = Criteria / comparative judgement; f = Model / code / query generation; g = Tool / workflow configuration; h = Direct decisions / policies; i = Negotiation; j = Explanation / interaction.

Integration: p = Prompt / context design; r = Implemented retrieval augmentation; w = Reported weight fine-tuning; o = Connected to OR / simulation; m = Connected to other ML; n = Interacting LLM agents.

Table S1. Study-level classifications (81 studies)

Reference	Group	Decision area	Method	Model role	Tasks	Integration
[6] Jackson et al. (2024)	A	I	E	A	f, j	p, o
[7] Sadeghi et al. (2026)	A	P	E	A	c, e	p, o
[8] Kirshner et al. (2026)	A	P	E	A	h, i	p, r, n
[25] Jackson et al. (2026)	A	R	E	A	a	p, r
[28] Le et al. (2026)	A	L	E	A	f, j	p, r, w, o
[29] Amarasinghe P.T. et al. (2026)	A	I	E	A	f	p, w, o
[30] Sukharamwala et al. (2026)	A	L	E	A	f, j	p, o
[34] Li Q. et al. (2026)	A	I	E	A	a, f	p, r, o
[35] Nie et al. (2026)	A	I	E	A	a	p, r, o
[36] Du et al. (2026)	A	G	E	A	c, e	p, o
[37] Niu et al. (2027)	C	C	T	T	-	-
[38] Lei et al. (2026)	B	R	A	R	a, b	p
[39] Yi et al. (2026)	A	D	E	A	b, d, j	p, m
[40] Herold et al. (2025)	A	P	E	A	h, i	p
[41] Dou et al. (2026)	C	C	T	T	-	-
[49] Wamba et al. (2023)	C	X	S	S	-	-
[50] Wamba et al. (2024)	C	X	S	S	-	-
[51] Hendriksen (2023)	C	X	C	C	-	-
[52] Li et al. (2024)	C	X	S	S	-	-
[55] Aslam et al. (2025)	C	X	S	S	-	-
[57] Sonar et al. (2026)	C	X	J	S	-	-
[58] Chowdhury et al. (2026)	C	R	S	S	-	-
[59] McKinley et al. (2026)	A	P	E	A	e	p, r
[60] Qian et al. (2026)	C	P	A	S	-	-
[61] Aspeteg et al. (2026)	B	R	Q	R	a, c	p
[62] Matsuno et al. (2026)	A	L	E	A	j	p, o
[63] Genetti et al. (2026)	A	I	E	A	j	p, o, m
[64] Du et al. (2026)	C	C	T	T	-	-
[65] Chen et al. (2025)	A	X	E	A	h	p
[66] Wang et al. (2025)	A	B	E	A	a, c, f, j	p
[67] Chen S. et al. (2026)	A	L	E	A	f, j, g	p, o, m
[68] Simchi-Levi et al. (2026)	A	I	E	A	c, f, j	p, o, m
[69] Chen S. (2026)	A	I	E	A	f, g	p, o
[70] Hamdan S. et al. (2026)	A	L	E	A	h	p
[71] Çetinkaya I.O. et al. (2026)	A	I	E	A	f	p, o
[72] Xiang W. et al. (2026)	A	I	E	A	g, h	p, w, o

Continued on next page

Reference	Group	Decision area	Method	Model role	Tasks	Integration
[73] Huang J. et al. (2026)	A	I	E	A	f, g	p, w, o
[74] Li R. et al. (2026)	A	I	E	A	f	p, o
[75] Yu F. et al. (2026)	A	I	E	A	f	p, o
[76] Xu M. et al. (2026)	A	I	E	A	c, f	p, o
[77] Zhao F. et al. (2026)	A	I	E	A	f	p, o, m
[80] Zhou et al. (2025)	A	P	E	A	a, b, c, e, j	p
[81] Dang et al. (2026)	A	P	E	A	c, g, j	p, m
[82] Lin et al. (2025)	A	P	E	A	b, d	p, w, o, m
[83] Liu et al. (2026)	C	P	T	T	-	-
[87] Jannelli et al. (2025)	A	I	E	A	h, i, j	p, o, n
[88] Zhou et al. (2026)	A	I	E	A	f, h	p, o
[89] Zouaghi et al. (2026)	A	X	E	A	c, h, j	p
[90] Niu et al. (2025)	B	I	A	R	a	p
[91] Ivanov D. (2026)	C	X	C	C	-	-
[93] Quan et al. (2026)	A	L	E	A	b, j	p, o, n
[94] De A. (2026)	A	L	E	A	a, b, c	p, o
[96] Khojasteh et al. (2026)	A	X	E	A	c, f, g, h	p, r, o, n
[97] Aslam M.W. et al. (2026)	A	L	E	A	d, h, j	p, w, o
[102] AlMahri et al. (2026)	A	R	E	A	a	p
[103] Li et al. (2026)	A	R	E	E	d	m
[104] Joglekar et al. (2025)	A	R	E	A	b, c	p
[105] Lin K.-Y. et al. (2025)	A	R	E	A	a, b	p, o
[106] Fang et al. (2026)	B	R	A	R	a, b, j	p
[107] Huchzermeier and Kuppek (2027)	B	R	Q	R	a, b	p, r
[108] Lo (2027)	C	R	J	S	-	-
[109] Zhu et al. (2026)	A	D	E	A	c, d	p, m
[110] Abolghasemi M. et al. (2025)	A	D	E	A	h	p
[111] Abbaszadeh Nakhost M. et al. (2025)	A	D	E	A	f, d	p
[112] Baharmand et al. (2026)	A	G	E	A	a	p, r, o
[113] Sorooshian et al. (2026)	B	G	Q	R	c, e	p
[114] Tong et al. (2026)	C	G	T	T	-	-
[115] Wang S. et al. (2026)	C	G	S	S	-	-
[116] Hou Y. et al. (2025)	C	G	S	S	-	-
[117] Tan et al. (2026)	C	C	T	T	-	-
[118] Yu et al. (2026)	C	C	T	T	-	-
[119] Li Z. et al. (2026)	C	C	T	T	-	-
[120] Wang et al. (2026)	C	C	T	T	-	-
[122] Zhou et al. (2026)	C	C	T	T	-	-
[123] Ning et al. (2026)	C	C	T	T	-	-

Continued on next page

Reference	Group	Decision area	Method	Model role	Tasks	Integration
[124] Zhang et al. (2026)	B	B	C	R	a	p
[125] Sriraman et al. (2026)	B	B	Q	R	a, c	p
[126] Niu et al. (2026)	B	X	A	R	a	p, w
[127] Maghroor et al. (2025)	C	X	J	S	-	-
[128] Qarqash Z. et al. (2026)	C	X	Q	S	-	-
[129] Trifone et al. (2026)	C	X	Q	S	-	-

Table S1 (continued). Theory and research context

Reference	Theory or conceptual basis	Research context
[6]	A human-AI collaboration framework specifies the responsibilities of domain experts and the model.	Simplified single-product inventory and single-server queue examples implemented in Python.
[7]	Analytic hierarchy process and virtual expert judgement.	Heavy-machinery purchasing, compared with 14 procurement professionals; specific supplier alternatives are not evaluated.
[8]	Anchoring, heuristics and information asymmetry in bilateral supply chain bargaining.	Simulated supplier-retailer bargaining with uncertain demand and supplier inventory risk.
[25]	Supply network science, nexus suppliers and cascading disruption.	Public disclosures around Foxconn, Flex and Jabil in electronics contract manufacturing.
[28]	Organisational information processing, with dynamic-capability and decision-process arguments.	A Vietnamese steel trader's national closed-loop network.
[29]	No explanatory management theory is tested. Supervised code-model adaptation and executable optimisation validation guide the design.	Classical job-shop cases, an embroidery-factory scheduling case and an LP benchmark; some case parameters are generated.
[30]	Solver-grounded explanations and grey-box transparency are design principles.	A stylised network with two factories, six distribution centres and six customers.
[34]	No explanatory management theory is tested. Expert modelling graphs and deterministic template assembly organise the division of work.	Thirty offline workforce-scheduling problems: 15 shift-scheduling and 15 days-off cases; no enterprise deployment.
[35]	Supply chain resilience and inventory buffering.	Critical-mineral stockpiling under shortage and depletion scenarios, using public reports and trade data.
[36]	Organisational information processing, with knowledge-hierarchy and value-chain classifications.	A Chinese textile recycler collaborating with hotel groups.
[37]	Co-opetition, technology spillovers and free riding.	A component supplier that also competes with its manufacturing customer.
[38]	Host-country sentiment, risk amplification and transaction costs motivate the empirical relationships.	1,460 Chinese listed firms and 21,796 overseas relationship-year observations, 2017–2022.
[39]	Exogenous volatility motivates financial and news-sentiment inputs.	US wheat futures and approximately 21,000 English agricultural news items over ten years.
[40]	Distributive and integrative negotiation and purchasing power-dependence.	Simulated desk and customised-power-bank procurement, with students and professional negotiators.
[41]	A manufacturer-led Stackelberg model incorporates consumer utility and double marginalisation.	A stylised manufacturer-retailer chain selling customised products.
[49]	Organisational learning informs interpretation of adoption-experience differences.	A UK-US online panel of 315 operations and supply chain practitioners.
[50]	Diffusion of innovation, including relative advantage and compatibility.	Three hundred managers across 19 industries in the United States and United Kingdom in early 2023.
[51]	Sensemaking and interpretivism; transaction-cost and complex-adaptive-system arguments provide supporting explanations.	A cross-industry conceptual analysis illustrated through inventory, supplier selection and distribution.
[52]	The relational view, with coordination mediation and supply chain dynamism.	A survey of 236 Chinese adopting firms in early 2024, including 158 manufacturers.
[55]	Task-technology fit and stimulus-organism-response.	A cross-industry, multinational sample of practitioners familiar with ChatGPT.
[57]	Technology-organisation-environment and dynamic capabilities.	Digital supply chain experts, primarily from western India.
[58]	Dynamic capabilities link AI-LLM augmentation to improvisation; resistance to digital innovation and decisiveness moderate this relationship.	224 UK supply chain managers from firms using AI-LLM, including manufacturing and transport businesses.

Continued on next page

Reference	Theory or conceptual basis	Research context
[59]	Signalling and interpretation of qualification versus differentiation signals.	A total of 123 bids for 31 Ohio IT procurements in 2023-2024.
[60]	Signalling and upstream demand, productivity and innovation spillovers.	North American listed-company announcements from January 2023 to May 2024.
[61]	Resource dependence distinguishes external bridging from internal buffering.	Annual reports of seven defence firms covering 2014–2024.
[62]	Queueing fairness and human-centred decision support motivate scheduling and interaction.	A simulated Japanese produce wholesale market; the main experiment includes 294 trucks and 10,000 pallets.
[63]	Human-involved digital twins and interpretable decision support motivate the framework.	Six months of action research at Italy's Adige Spa, including 345 real orders from 2021–2024.
[64]	Consumer valuation biases, information asymmetry and noncooperative channel games.	A stylised digital-product dual channel, illustrated by digitally manufactured products.
[65]	Behavioural operations; prospect theory, dual-process reasoning and reference dependence.	English decision vignettes, with standard psychological problems reframed as inventory and other operations tasks.
[66]	Information asymmetry motivates the workflow.	Chinese new-energy bus procurement and production, with feedback from financing practitioners.
[67]	No explanatory management theory is tested. Quality-diversity evolution and interpretable heuristic archives provide the methodological basis.	Offline routing and scheduling benchmarks under static, perturbed, stochastic and dynamic conditions; no enterprise deployment.
[68]	The 4I design framework comprises Insight, Interpretability, Interactivity and Improvisation; it is a design contribution.	Microsoft cloud supply-chain fulfilment. A production MILP already operated before the GenAI assistant; pilot use in March–October 2023 preceded wider deployment in November 2023.
[69]	No explanatory management theory is tested. Coupling-aware destruction and multiobjective search provide the methodological basis.	Offline blocking hybrid-flow-shop scheduling with time-of-use electricity prices; 100 synthetic instances and 15 SCC cases.
[70]	No explanatory management theory is tested. External verification of feasibility and objective values structures the evaluation.	Nine routing benchmark families, additional large instances and a multistage inventory-routing problem; three model-generation snapshots.
[71]	No explanatory management theory is tested. Automatic heuristic discovery through evolutionary program search is the methodological basis.	Offline single-machine scheduling; 10,000 training instances and tests spanning 20 to 500 jobs.
[72]	No explanatory management theory is tested. Multiagent reinforcement learning and expert-guided model adaptation provide the design basis.	Semiconductor wafer batch-processing simulation, including a factory-scale SimPy model based on an enterprise prototype.
[73]	No supply-chain explanatory theory is tested. Textual gradients motivate feedback aggregation; information-theoretic arguments are presented as intuition.	One thousand synthetic dynamic flexible-job-shop instances and 100 scenarios based on a Guangzhou MINO production line.
[74]	No explanatory management theory is tested. Memetic search and cooperative automatic heuristic design form the methodological basis.	Lot-streaming hybrid job-shop benchmarks and an aerospace tooling-shop case.
[75]	No explanatory management theory is tested. Evolutionary selection evaluates LLM-generated heuristic programs.	Offline flow-shop, job-shop and open-shop benchmarks, using public and synthetic instances.
[76]	No explanatory management theory is tested. A neuro-symbolic division combines LLM knowledge extraction with genetic-programming search.	Dynamic flexible-job-shop simulation with six single-objective and four multiobjective scenarios.
[77]	No explanatory management theory is tested. A hybrid algorithm combines heuristic initialisation, evolutionary expansion and multiobjective policy refinement.	Fifteen synthetic hybrid-flow-shop instances motivated by aluminium casting; peak-power constraints are imposed.
[80]	Green supply chain management and the GSCMEIS evaluation framework.	A sample of 215 Chinese listed firms and an illustrative shipbuilding case.
[81]	Complex adaptive supply networks and task decomposition motivate the workflow.	A global aerospace manufacturer's 2021-2022 purchasing data.
[82]	Bayesian risk updating, multicriteria weighting and Markov decision processes.	Historical data for eighteen suppliers of an anonymous German semiconductor manufacturer.
[83]	Channel competition and sequential or simultaneous quantity games.	A retailer, a reliable national-brand manufacturer and an unreliable overseas source.

Continued on next page

Reference	Theory or conceptual basis	Research context
[87]	Complex adaptive supply networks, local incentives and the bullwhip effect.	A simulated three-echelon sequential chain with two-step lead times and 200 operating steps.
[88]	A partially observable stochastic game captures decentralised inventory competition and channel substitution.	Simulated two-echelon, dual-channel markets, including Rossmann demand histories.
[89]	An efficiency-innovation framework motivates task-dependent temperature settings.	Nine simulated supply chain tasks covering forecasting, planning and circular operations.
[90]	Servitisation stages, information sharing and production-efficiency mechanisms in bullwhip research.	US listed manufacturers, using SEC reports and Compustat data for 1997-2018.
[91]	An extension of the digital-twin hierarchy adds an agentic layer; ICARUS organises proposed capabilities.	Conceptual analysis with secondary Microsoft and Ford examples; no original field implementation or model experiment.
[93]	Supply chain risk management and Physical Internet hub deployment.	A total of 376 properties in seven southeastern US states, with NOAA weather narratives and transport, crime and rental data.
[94]	Resilience and viability motivate risk-informed network planning.	An Asian network initially containing 247 manufacturing nodes and 32 ports.
[96]	Design science and SCOR organise a control-tower architecture with governance functions; no explanatory management theory is tested.	Python/CrewAI proof of concept with a four-echelon beer game, simplified ERP and simulated external services.
[97]	No explanatory management theory is tested. Spatial reasoning and constrained layout search provide the design basis.	An eight-machine layout benchmark with obstacles; larger 15-, 25- and 35-machine cases; a 16-department drinking-water-machine factory case.
[102]	Knowledge-graph representation of entities and relationships.	Public automotive, battery and mineral-source information, supplemented by sentence-level extraction tests.
[103]	The SCRAM capability-vulnerability framework and recovery-time concepts.	An anonymous automotive OEM supply network with 1031 suppliers, 1321 directed links and six tiers.
[104]	Contextual information processing motivates attention to regional differences in supply chain information.	5,168 English news articles from the Democratic Republic of Congo, China and North America around a November 2023 truck strike.
[105]	Informational agility motivates a sensing-prediction-response architecture; no management theory is tested independently.	Three simulated supply regions with different digital maturity and reliability; a ten-period planning horizon.
[106]	Macro and micro supply chain risk categories, extended to digital risks; inventory buffering.	Shenzhen-listed nonfinancial firms' investor-relations records for 2018-2024 and a stylised inventory simulation.
[107]	Existing resilience dimensions and the Triple-P complexity framework organise interpretation; geopolitical risk management is distinguished from operational risk management.	2024 annual reports of 20 European multinational firms across five industries and four countries.
[108]	Extended FMEA is the assessment framework; dynamic-capability and absorptive-capacity arguments interpret the results.	Taiwanese machinery SMEs; senior expert judgements and an anonymised retrospective case.
[109]	A multimodal forecasting framework combines market narratives with numerical price features.	A Chinese battery manufacturer; copper, ternary precursor and lithium carbonate prices over 1,081 trading days each.
[110]	Judgemental forecasting provides the behavioural comparison; no independent supply-chain explanatory theory is tested.	Controlled forecasting experiment with 123 graduate participants and five LLMs, using 24 sales series from an Australian food manufacturer.
[111]	No explanatory management theory is tested. Intermittent-demand forecasting and self-reflective code generation provide the methodological basis.	Four simulated spare-parts datasets and two real datasets, using training/test splits; a complete self-reflective process is repeated 30 times.
[112]	Circular economy informs a multiobjective decision framework.	Pakistan flood relief in 2022, with five suppliers, three hubs and five camps.
[113]	DPSIR organises causal relationships; natural-resource and social-ecological perspectives inform application.	Cross-industry sustainable supply chains, with prospective palm-oil sourcing examples.
[114]	A downstream-led Stackelberg model combines operational hedging with cost sharing.	Stylised upstream and downstream manufacturers serving an environmentally sensitive market.

Continued on next page

Reference	Theory or conceptual basis	Research context
[115]	Dynamic capabilities and TOEH motivate absorptive and transformative capacity mechanisms and AI-turbulence moderation.	334 Chinese cross-border e-commerce firms, followed by ten management interviews.
[116]	Resource orchestration theory motivates resource-orchestration and decarbonisation capability mechanisms.	Firm-level survey of Chinese high-technology manufacturers; no LLM execution experiment.
[117]	A supplier-manufacturer Stackelberg game with consumer time utility and government welfare.	A stylised two-member chain without empirical calibration.
[118]	Stackelberg channel competition.	A stylised manufacturer-supplier-platform system with a direct website channel.
[119]	Sequential games and consumer expected utility represent platform adoption, retailer pricing and consumer trust.	Analytical e-commerce platform models with monopoly/competition and numerical extensions; no firm-level intervention data.
[120]	Stackelberg and supplier Nash games with heterogeneous consumer utility.	A stylised retailer with branded and ordinary suppliers and trusting or sceptical consumers.
[122]	Tripartite evolutionary game theory incorporates bounded rationality and industry development.	A stylised Chinese fresh-food chain comprising a supplier, platform and community group leader.
[123]	Stackelberg investment and two-part charging link platform incentives to manufacturer production.	A stylised platform-manufacturer channel, extended to competing manufacturers and two periods.
[124]	Resource mobilisation and incentive alignment, with financial frictions and information asymmetry.	US corporate 10-K disclosures from January 2022 to June 2025.
[125]	Theory elaboration links financing frictions to blockchain mechanisms.	11 global financing platforms, including seven continuing and four failed or reduced operations.
[126]	Self-attribution, impression management and signalling; reflection and feedback learning.	US listed-company earnings calls from late 2005 to late 2023.
[127]	TOEH combines technology-organisation-environment, HOT-Fit and force-field perspectives.	Eleven North American and European academic or practitioner experts.
[128]	Abductive extension of an existing AI decision framework develops middle-range propositions about human-AI collaboration.	Seven expert interviews in Germany, Jordan, Canada and the United States, supplemented by secondary information.
[129]	The technology-organisation-environment framework organises drivers and constraints; inductive qualitative analysis develops propositions.	Nineteen experienced supply-chain and AI practitioners interviewed across organisational roles and industries.

Table S2. Evidence appraisal of operational-application studies (46 studies)

Ref.	Data	Comparator	Reference standard	Model	C/M/A/R/P	Outcome
[6]	Inventory and queueing examples	Alternative descriptions and starting code	Author inspection of executable simulations	GPT-3 Codex	0/1/0/0/1	T
[7]	Heavy-machinery procurement criteria	Fourteen procurement professionals	Professional weights and AHP consistency	ChatGPT-4	1/1/0/0/1	D
[8]	Simulated supplier-retailer bargaining	Human experiments; information and retrieval variants	Contract profits and theoretical efficiency	GPT-4o	1/1/1/1/0	D
[25]	Public electronics-company disclosures	S&P Capital IQ company records	Company-name matches to external records	GPT-3.5 Turbo	1/1/0/0/0	T
[28]	Steel-company data; query benchmarks	Model, retrieval and fine-tuning variants	Reference queries, answers and explanations	GPT-5.2; GPT-4.1; Llama variants	1/1/1/1/1	T
[29]	Scheduling cases and LP benchmark	Prompt baselines; training configurations	Solver execution and reference solutions	Fine-tuned CodeRL; GPT-3.5	1/1/1/0/1	D
[30]	Thirty network-planning scenarios	Prompt-only LLM and model variants	Expert scenario answers and Gurobi solutions	GPT-4 Turbo; other OpenAI models	1/1/1/0/1	D
[34]	Thirty workforce-scheduling instances	Zero-shot prompts; OptiMUS	Expert model checks and reference optimum	GPT-4o-2024-08-06	1/1/0/1/0	D
[35]	Public documents and mineral-trade data	Stockpiling policies and shortage scenarios	Checked extractions; historical trade validation	Qwen-3-32B	1/1/1/1/1	D
[36]	Hotel-linen interviews and barrier judgements	Experts; prompt and model variants	Human interviews and expert judgements	GPT-4o; GPT-5; Gemini; Qwen	1/1/1/1/1	D
[39]	Wheat prices and agricultural news	Forecast baselines; sentiment/feature variants	Observed prices in temporal test sets	DeepSeek	1/1/1/0/1	D
[40]	Three human-AI negotiation experiments	Cooperative versus competitive instructions	Agreed terms and participant responses	GPT-3.5-turbo	1/1/1/0/1	D
[59]	Actual procurement bids	Human evaluators; three reasoning models	Human procurement assessment records	o3; Grok-3-Mini; DeepSeek-R1	1/1/0/1/0	D
[62]	Wholesale-market scheduling simulation	FCFS; heuristic; genetic scheduling	Simulator metrics; LLM dialogue examples	Gemma 3 27B	1/1/1/0/0	D
[63]	Factory orders and synthetic scheduling data	Interactive versus conventional optimisation	Simulated schedules and expert feedback	GPT-4o; Llama; Mistral; Phi-3	1/1/1/1/1	D
[65]	Decision vignettes across three contexts	Human literature; unbiased benchmarks; model variants	Established bias tests and calculated answers	GPT-3.5-turbo; GPT-4	1/1/1/1/0	D
[66]	Bus procurement pages and financial files	Manual and four LLM-assisted workflows	Manual output and evaluator quality scores	GPT-4; GPT-3.5; Gemini; Ernie	1/1/0/0/1	T
[67]	Routing and scheduling benchmarks	Classical/neural solvers; module removals	Benchmark solutions and external evaluators	GPT-5.4; Claude; Qwen variants	1/1/1/1/1	D
[68]	Microsoft production-deployment account	Same workflow before/after assistant	Operational records and team time estimates	Deployed model unspecified	1/0/0/0/0	D

Continued on next page

Ref.	Data	Comparator	Reference standard	Model	C/M/A/R/P	Outcome
[69]	Synthetic and industrial scheduling cases	Nine algorithms; non-LLM controls	Benchmark objective values and feasibility	LLM unspecified	1/0/1/1/0	D
[70]	Routing and inventory-routing instances	Gurobi and specialised heuristics	External feasibility and objective checks	Eight OpenAI/Gemini/Grok/DeepSeek snapshots	1/1/0/0/0	D
[71]	Single-machine scheduling instances	Classical rules and exact references	Reference total-tardiness values	Mixtral 8x7B Instruct	1/1/0/0/1	D
[72]	Wafer-processing simulation and factory prototype	Scheduling/RL baselines; module removals	Simulator flow-time outcomes	Lightweight LLM; Gemma/Phi/Llama checks	1/0/1/1/0	D
[73]	Synthetic and production-line scheduling cases	Dispatching rules; GP/GEP; LLM variants	Simulator makespan and tardiness	ModouGPT; DeepSeek; GPT; Qwen	1/1/1/0/1	D
[74]	Lot-streaming benchmarks and tooling-shop case	GA; RL; memetic and component variants	Benchmark scheduling objectives	GPT-4o; GPT-4o-mini; Claude-3.5	1/1/1/1/1	D
[75]	Flow/job/open-shop scheduling benchmarks	Classical rules and module removals	Benchmark makespan values	GPT-4o-mini	1/1/1/0/0	D
[76]	Dynamic flexible-job-shop simulation	GP; GPLS; EGP	Simulator scheduling objectives	ChatGPT 4.0; Gemini 2.5; Claude Sonnet 4.5	1/1/1/1/1	D
[77]	Fifteen hybrid-flow-shop instances	Six algorithms and module removals	Simulator objectives and power constraints	GPT-4o; GPT-5	1/1/1/1/1	D
[80]	Listed-company disclosures; shipbuilding case	Human labels; OLS; model versions	Human action labels and evaluator ratings	GPT-3.5-turbo; GPT-4	1/1/1/1/1	D
[81]	Aerospace purchase-line records	Manual pipeline; AutoML; alternative LLMs	Recorded quality-event labels	Llama 3.1 8B	1/1/1/0/0	T
[82]	Supplier records and stress simulations	MCDM; hybrid systems; module removals	Supplier outcomes and simulated costs	GPT (version unspecified)	1/0/1/1/0	D
[87]	Three-echelon supply-chain simulation	Independent agents; information sharing; ordering policies	Simulated costs and order fluctuations	Gemini 1.5 Flash/Pro	1/1/1/0/1	D
[88]	Demand histories and simulated markets	Specialised multiagent RL and other policies	Simulated manufacturer and retailer profits	ChatGPT-4-Turbo; DeepSeek-V3	1/1/0/1/1	D
[89]	Nine simulated supply-chain tasks	Temperature settings; ARIMA	Task references and expert ratings	GPT-4o; GPT-3.5 checks	1/1/1/0/0	D
[93]	Weather narratives and logistics properties	Rules; model and revision variants	Two hundred events rated by study authors	GPT-4o; alternative models	1/1/1/1/1	D
[94]	Geographic, conflict and semiconductor data	Scenario/clustering checks; no LLM ablation	Network and simulation outputs	Flan-T5	1/1/1/0/0	D
[96]	Agent proof of concept; beer-game simulation	Local delayed/immediate and shared-signal policies	Simulator costs; policy test separate from agent runs	LLM agents; backbone unspecified	1/0/1/1/0	D
[97]	Layout benchmarks and factory case	GA; SA; PSO; Gurobi; module removals	Material-handling costs and constraint checks	Fine-tuned VLLM; Gemma-3 variants	1/1/1/1/1	D
[102]	Public reports and annotated sentences	String and TF-IDF matching	Labelled entities and relations	GPT-4o	1/1/0/1/1	T

Continued on next page

Ref.	Data	Comparator	Reference standard	Model	C/M/A/R/P	Outcome
[103]	Automotive supplier-network data	Conventional predictors; feature ablations	Expert and recovery-based resilience targets	text-embedding-ada-002	1/1/1/1/0	T
[104]	Cobalt supply-chain news	Context and aggregation variants	Expert sentiment checks on six articles	Gemini 1.5 Flash; GPT-4 evaluator	1/1/1/1/1	T
[105]	Thirty disruption scenarios across regions	Risk-averse optimisation; digital-twin simulation	Simulated service, costs and recovery	Transformer language model unspecified	1/0/1/1/0	D
[109]	Raw-material prices and generated narratives	Text sources; predictors; module removals	Observed prices in temporal test sets	Qwen; ERNIE Bot; SparkDesk	1/1/1/1/1	D
[110]	Food-sales forecasting experiment	Human forecasters; five LLMs	Observed sales and forecasting errors	ChatGPT-3.5/4; Bard; Bing; Llama2	1/1/1/0/0	D
[111]	Simulated and real spare-parts demand	Seven forecasting baselines; prompt variants	Held-out demand observations	OpenAI o1-mini	1/1/1/1/1	D
[112]	Humanitarian reports and flood-relief case	Optimisation scenarios and sensitivity	Human-checked bounds; no extraction benchmark	PrivateGPT; backbone unspecified	1/0/1/0/1	D

Note. C = reported comparison; M = identifiable model family or version; A = component or parameter checks; R = repeated evaluation; P = prompt examples or a reusable protocol. Values are 1 (confirmed in the examined study materials) or 0 (not confirmed). Outcomes: D = decision-support or operational result; T = task-level result.

Table S3. Topic-model stability and parameter checks

All ten random-seed fits recovered four topics. The 45 pairwise adjusted Rand indices averaged 0.83 (minimum 0.71). Mean Jaccard similarities between each reported theme and its matches in the other nine runs were 0.858, 0.841, 0.848 and 1.000. Of the 81 articles, 68 retained their theme in all ten runs. Jaccard similarity is the number of articles shared by two matched themes divided by the number in their union. Themes were matched one-to-one by maximising total Jaccard similarity; the seed-42 comparisons use the other nine runs.

Random-seed runs

Seed	Topics	Unassigned	Topic sizes
42	4	0	29, 22, 15, 15
7	4	0	28, 24, 15, 14
11	4	4	24, 20, 18, 15
13	4	0	29, 22, 15, 15
17	4	10	24, 17, 15, 15
23	4	1	24, 22, 19, 15
29	4	5	24, 19, 18, 15
37	4	5	24, 19, 18, 15
51	4	7	27, 18, 15, 14
101	4	6	24, 18, 18, 15

Jaccard similarity by theme

		Seed 42 vs other nine		All 45 pairs		Same in
Theme	**Articles**	**Mean**	**Min.**	**Mean**	**Min.**	**all ten**
Risk, models and decision frameworks	29	0.858	0.828	0.905	0.821	23
Generative AI and organisational adoption	22	0.841	0.739	0.841	0.607	17
Supplier/platform interaction and collaboration	15	0.848	0.700	0.825	0.650	13
Scheduling, optimisation and heuristics	15	1.000	1.000	1.000	1.000	15

Note. Topic sizes are listed from largest to smallest. Unassigned articles count as absent from a theme. "Same in all ten" counts articles retaining their reported theme in every run.

The parameter sweep combined four UMAP neighbourhood sizes (10, 15, 20 and 25), two output dimensions (3 and 5), and five HDBSCAN minimum cluster sizes (4, 5, 6, 8 and 10) across the same ten seeds, giving 400 fits. Topic counts ranged from two to nine: four topics occurred in 192 fits (48%), five in 101, six in 49, three in 25, seven in 16, two in 12, eight in four and nine in one. Under an alternative parameter setting with 20 neighbours, five dimensions and a minimum cluster size of eight, all ten seeds produced four topics.

Table S4. Publication profile and first-author affiliation

Issue year / publication outlet	Articles
2023	2
2024	3
2025	14
2026	59
2027	3
Publication outlets	
International Journal Of Production Economics	23
International Journal Of Production Research	19
Ieee Transactions On Engineering Management	8
Ieee Transactions On Evolutionary Computation	5
Computers & Operations Research	3
Journal Of The Operational Research Society	3
Production And Operations Management	3
Supply Chain Management-An International Journal	3
Annals Of Operations Research	2
M&Som-Manufacturing & Service Operations Management	2
Production Planning & Control	2
Decision Sciences	1
International Journal Of Forecasting	1
International Journal Of Operations & Production Management	1
Journal Of Business Logistics	1
Journal Of Purchasing And Supply Management	1
Journal Of Supply Chain Management	1
Management Science	1
Operations Research	1

Note. Counts use publication issue years.

First-author affiliation	Articles	%
China	35	43.2
United States	12	14.8
Australia	5	6.2
Germany	5	6.2
United Kingdom	4	4.9
France	3	3.7
India	2	2.5
Italy	2	2.5
Japan	2	2.5
Netherlands	2	2.5
Sweden	2	2.5
Taiwan	2	2.5
Algeria	1	1.2
Canada	1	1.2
Denmark	1	1.2
Norway	1	1.2
Switzerland	1	1.2

Table S5. Authors contributing at least two included articles

Author	Included articles	WoS citations to those articles	Articles with WoS coverage	Within-corpus citations
Ivanov, Dmitry	4	144	4	12
Cheng, T. C. Edwin	3	103	3	16
Wamba, Samuel Fosso	3	469	3	58
Wu, Jing	3	47	3	3
Brintrup, Alexandra	2	41	2	7
Jackson, Ilya	2	120	2	12
Kirshner, Samuel N.	2	54	2	8
Kumar, Mukesh	2	9	2	1
Li, Lixu	2	107	2	15
Li, Xinyu	2	0	2	0
Matsuno, Kotomichi	2	16	2	1
Minner, Stefan	2	233	2	32
Nie, Wei	2	9	2	1
Niu, Yimeng	2	45	2	3
Queiroz, Maciel M.	2	464	2	58
Saenz, Maria Jesus	2	120	2	12
Tsolakis, Naoum	2	9	2	1
Zhang, Hao	2	0	2	0

Note. The corpus includes 263 authors and 286 authorships. Counts use full author credit; coauthor citation totals overlap.

Table S6. Citations to the included articles

Web of Science covers 73 articles with 1,232 citations; Scopus covers 72 with 1,514.

Reference	WoS	Scopus	Within corpus	Reference	WoS	Scopus	Within corpus	Reference	WoS	Scopus	Within corpus
[6]	109	159	10	[64]	0	0	0	[102]	21	21	2
[7]	4	3	0	[65]	47	49	6	[103]	0	0	0
[8]	7	6	2	[66]	10	13	3	[104]	5	7	1
[25]	11	9	2	[67]	–	0	0	[105]	15	16	1
[28]	0	0	0	[68]	–	–	0	[106]	4	6	2
[29]	5	6	2	[69]	0	0	0	[107]	0	–	0
[30]	1	1	0	[70]	0	0	0	[108]	0	–	0
[34]	1	1	0	[71]	13	16	1	[109]	0	0	0
[35]	4	4	0	[72]	3	2	1	[110]	13	16	0
[36]	2	3	0	[73]	0	0	0	[111]	9	9	3
[37]	0	–	0	[74]	24	24	1	[112]	1	1	0
[38]	0	0	0	[75]	–	12	0	[113]	1	1	0
[39]	2	3	0	[76]	–	1	0	[114]	3	3	1
[40]	18	21	0	[77]	–	17	1	[115]	0	0	0
[41]	5	7	1	[80]	11	13	1	[116]	8	8	0
[49]	240	344	29	[81]	0	0	0	[117]	7	7	1
[50]	224	279	29	[82]	13	7	0	[118]	0	0	0
[51]	148	190	11	[83]	0	0	0	[119]	5	6	0
[52]	99	114	15	[87]	20	22	5	[120]	3	3	1
[55]	10	14	2	[88]	1	1	0	[122]	4	7	0
[57]	2	1	1	[89]	5	5	0	[123]	6	–	1
[58]	0	0	0	[90]	45	–	3	[124]	2	0	0
[59]	0	0	1	[91]	23	29	0	[125]	0	1	0
[60]	2	3	0	[93]	2	5	1	[126]	0	–	0
[61]	0	0	0	[94]	–	0	0	[127]	10	11	0
[62]	1	2	0	[96]	–	–	0	[128]	1	1	0
[63]	2	4	1	[97]	0	0	0	[129]	–	–	0

Note. A dash indicates unavailable coverage. Within-corpus citations use DOI links in Web of Science cited-reference lists.

Table S7. Author-keyword frequency and occurrence by issue year

Keyword	Total	2023	2024	2025	2026	2027
large language models	38	0	0	6	32	0
generative ai	29	1	3	4	20	1
supply chain	12	0	2	3	6	1
artificial intelligence	11	1	1	4	5	0
supply chain management	9	0	0	1	8	0
decision making	8	0	0	1	6	1
chatgpt	6	1	1	3	1	0
optimization	6	0	0	1	5	0
digital twin	4	0	0	0	4	0
sustainable development	4	0	0	2	2	0
agentic ai	3	0	0	0	3	0
chatbots	3	0	0	2	1	0
costs	3	0	0	1	2	0
decision support system	3	0	0	0	3	0
evolutionary computation	3	0	0	0	3	0
game theory	3	0	0	0	3	0
natural language processing	3	0	1	0	2	0
resilience	3	0	0	1	1	1
supplier selection	3	0	0	1	2	0
supply chain resilience	3	0	0	1	2	0
training	3	0	0	3	0	0
artificial	2	0	0	0	2	0
capabilities	2	0	0	1	1	0
critical mineral supply chains	2	0	0	1	1	0
digital manufacturing	2	0	0	0	2	0
e-commerce platform	2	0	0	0	2	0
explainability	2	0	0	0	2	0
formulation	2	0	0	0	2	0
generative ai (gen-ai)	2	0	0	1	1	0
human-ai collaboration	2	0	0	0	2	0
intelligence	2	0	0	0	2	0
interpretability	2	0	0	0	2	0
machine learning	2	0	0	0	2	0
manufacturing	2	0	0	1	0	1
multi-agent system	2	0	0	1	1	0
negotiation	2	0	0	1	1	0
production scheduling	2	0	0	0	2	0
prompt engineering	2	0	0	1	1	0
real-time systems	2	0	0	2	0	0
risk assessment	2	0	0	0	1	1
simulation	2	0	1	0	1	0
supply chain management (scm)	2	0	0	1	1	0
supply chain mapping	2	0	0	0	2	0
supply chain visibility	2	0	0	0	2	0
surveys	2	0	0	1	1	0
textual analysis	2	0	0	1	1	0

Note. Keywords are normalised and counted once per article. The table includes the 20 most frequent terms and all 40 network terms; 80 of the 81 articles provide keywords.

Section S1. Search strategies

Search date: 5 November 2026.

Scope: 40 journals rated 3, 4 or 4* in the OR&MANSCI and OPS&TECH categories of AJG 2024; 2019-2026 publication window; Article filter.

Web of Science Core Collection returned 228 records and Scopus returned 195, giving 423 database records. Removing 174 cross-database duplicates left 249 unique records for screening.

Web of Science uses TS (including Keywords Plus), IS, PY and DT. Scopus uses TITLE-ABS-KEY, ISSN, PUBYEAR and the Article document-type filter.

Web of Science Core Collection

TS=("large language model" OR "large language models" OR LLM OR LLMs OR "LLM-based" OR "LLM-driven" OR "LLM-powered" OR "LLM-assisted" OR ChatGPT OR "GPT-4*" OR "GPT-3*" OR "generative pre-trained transformer" OR "generative pretrained transformer" OR "foundation model" OR "foundation models" OR "generative AI" OR "generative artificial intelligence" OR "retrieval-augmented generation" OR "retrieval augmented generation" OR "prompt engineering" OR "LLM agent" OR "LLM agents" OR "language model*" OR GPT OR LLaMA OR Gemini OR DeepSeek OR Qwen OR Claude OR "agentic AI" OR "agentic artificial intelligence" OR "AI agent*" OR chatbot* OR "conversational agent*" OR "conversational AI" OR GenAI OR "Gen-AI") AND TS=("supply chain" OR "supply chains" OR "supply network" OR "supply networks" OR logistic OR logistics OR procurement OR sourcing OR supplier OR suppliers OR inventory OR production OR manufactur* OR scheduling OR transport* OR forecast* OR warehous* OR "operations management" OR purchas*) AND IS=(0025-1909 OR 0030-364X OR 0377-2217 OR 1089-778X OR 0025-5610 OR 1049-3301 OR 0254-5330 OR 0926-6003 OR 0305-0548 OR 0011-7315 OR 1063-6560 OR 1568-4539 OR 2168-2267 OR 2168-2216 OR 2472-5854 OR 1091-9856 OR 0169-2070 OR 1381-1231 OR 0022-3239 OR 0160-5682 OR 0364-765X OR 0894-069X OR 0305-0483 OR 0171-6468 OR 0951-8320 OR 1052-6234 OR 0041-1655 OR 0272-6963 OR 0144-3577 OR 1523-2409 OR 1059-1478 OR 0166-3615 OR 0018-9391 OR 0925-5273 OR 0020-7543 OR 0735-3766 OR 1478-4092 OR 1523-4614 OR 0953-7287 OR 1359-8546) AND PY=(2019-2026) AND DT=(Article)

Scopus

TITLE-ABS-KEY("large language model" OR "large language models" OR llm OR llms OR "llm-based" OR "llm-driven" OR "llm-powered" OR "llm-assisted" OR chatgpt OR "gpt-4*" OR "gpt-3*" OR "generative pre-trained transformer" OR "generative pretrained transformer" OR "foundation model" OR "foundation models" OR "generative ai" OR "generative artificial intelligence" OR "retrieval-augmented generation" OR "retrieval augmented generation" OR "prompt engineering" OR "llm agent" OR "llm agents" OR "language model*" OR gpt OR llama OR gemini OR deepseek OR qwen OR claude OR "agentic ai" OR "agentic artificial intelligence" OR "ai agent*" OR chatbot* OR "conversational agent*" OR "conversational ai" OR genai OR "gen-ai") AND TITLE-ABS-KEY("supply chain" OR "supply chains" OR "supply network" OR "supply networks" OR logistic OR logistics OR procurement OR sourcing OR supplier OR suppliers OR inventory OR production OR manufactur* OR scheduling OR transport* OR forecast* OR warehous* OR "operations management" OR purchas*) AND ISSN("0025-1909" OR "0030-364X" OR "0377-2217" OR "1089-778X" OR "0025-5610" OR "1049-3301" OR "0254-5330" OR "0926-6003" OR "0305-0548" OR "0011-7315" OR "1063-6560" OR "1568-4539" OR "2168-2267" OR "2168-2216" OR "2472-5854" OR "1091-9856" OR "0169-2070" OR "1381-1231" OR "0022-3239" OR "0160-5682" OR "0364-765X" OR "0894-069X" OR "0305-0483" OR "0171-6468" OR "0951-8320" OR "1052-6234" OR "0041-1655" OR "0272-6963" OR "0144-3577" OR "1523-2409" OR "1059-1478" OR "0166-3615" OR "0018-9391" OR "0925-5273" OR "0020-7543" OR "0735-3766" OR "1478-4092" OR "1523-4614" OR "0953-7287" OR "1359-8546") AND PUBYEAR > 2018 AND PUBYEAR < 2027 AND (LIMIT-TO (DOCTYPE , "ar"))