ISSN: 3079-6342

Large Language Models as Decision Support Tools in Procurement and Supplier Selection

Jingyi Liu[1]* and Pan Li[2]

[1]Cornell University, United States

[2]University of Hull, United Kingdom

*** Corresponding author:** jl3758@cornell.edu

Abstract

Large language models (LLMs) have emerged as transformative technologies in enterprise decision-making, offering unprecedented capabilities in natural language processing (NLP), information synthesis, and predictive analytics. This review examines the application of LLMs as decision support tools in procurement and supplier selection processes. Procurement professionals face increasingly complex challenges in supplier evaluation, risk assessment, contract analysis, and market intelligence gathering. Traditional decision support systems (DSS) often struggle with unstructured data and nuanced linguistic information inherent in supplier communications, contracts, and market reports. LLMs, powered by transformer architectures and trained on vast corpora, demonstrate remarkable abilities in understanding context, extracting insights from diverse data sources, and generating actionable recommendations. This paper systematically reviews recent developments in LLM applications for procurement, analyzing their capabilities in supplier risk assessment, performance prediction, contract compliance monitoring, and market trend analysis. We examine how models such as GPT-4, Claude, and domain-specific variants are being integrated into procurement workflows to enhance decision quality, reduce processing time, and improve supplier relationship management. The review identifies key technical challenges including data privacy concerns, hallucination risks, integration with existing enterprise resource planning (ERP) systems, and the need for domain-specific fine-tuning. We also explore emerging research on multi-agent LLM systems for collaborative procurement decisions, retrieval-augmented generation (RAG) for supplier database queries, and interpretability mechanisms that enhance trust in LLM-generated recommendations. Our analysis reveals that while LLMs offer substantial benefits in processing efficiency and insight generation, successful implementation requires careful consideration of organizational readiness, data governance frameworks, and human oversight mechanisms.

Keywords

Large language models, procurement, supplier selection, decision support systems, natural language processing, artificial intelligence, supply chain management, risk assessment, contract analysis

1. Introduction

The procurement function has evolved from a transactional cost-center to a strategic value driver in modern organizations, requiring sophisticated analytical capabilities to navigate global supply chains, manage supplier relationships, and mitigate operational risks. Large language models (LLMs) represent a paradigm shift in artificial intelligence (AI) that is

ISSN: 3079-6342

fundamentally transforming how procurement professionals approach decision-making processes [1]. These neural network architectures, trained on massive text corpora comprising billions of parameters, exhibit emergent capabilities in natural language understanding, reasoning, and generation that surpass previous generations of AI systems [2]. The application of LLMs in procurement contexts addresses long-standing challenges in processing unstructured data sources, including supplier communications, market intelligence reports, contract documents, and regulatory filings that contain critical decision-relevant information but resist traditional structured database approaches [3]. Procurement and supplier selection decisions involve complex multi-criteria evaluations that integrate quantitative metrics such as cost, quality, and delivery performance with qualitative factors including corporate culture alignment, innovation capability, and sustainability practices [4]. Traditional decision support systems (DSS) rely heavily on structured data inputs and predefined decision rules, limiting their effectiveness when faced with the linguistic nuances and contextual subtleties that characterize real-world supplier evaluations [5]. The emergence of transformer-based architectures, beginning with the seminal attention mechanism and subsequently refined through models such as Generative Pre-trained Transformer (GPT), Bidirectional Encoder Representations from Transformers (BERT), and their successors, has created new possibilities for extracting insights from textual data at unprecedented scale and sophistication [6]. Recent iterations including GPT-4, Claude, and domain-adapted variants demonstrate capabilities in multi-document reasoning, numerical analysis, and contextual understanding that align closely with the cognitive demands of procurement decision-making [7]. The integration of LLMs into procurement workflows represents more than incremental technological advancement; it constitutes a fundamental reconceptualization of how organizations can leverage information assets for competitive advantage. Enterprise resource planning (ERP) systems traditionally serve as the backbone of procurement data management, but their structured database paradigms struggle to capture and utilize the rich qualitative information embedded in supplier proposals, performance reviews, and market analyses [8]. LLMs offer complementary capabilities through natural language processing (NLP) techniques that can interpret, synthesize, and reason over unstructured text at human-level or superior performance on many tasks [9]. Organizations are exploring applications ranging from automated supplier risk screening based on news sentiment analysis to contract clause extraction and compliance verification, leveraging LLMs' ability to process diverse information sources with minimal manual data preparation [10]. Despite the promising capabilities, the adoption of LLMs in procurement raises important questions regarding accuracy, reliability, and organizational integration. The phenomenon of hallucination, where models generate plausible but factually incorrect information, poses particular risks in decision contexts where erroneous supplier assessments could lead to supply chain disruptions or financial losses [11]. Data privacy and security concerns emerge when proprietary procurement information is processed through cloud-based LLM services, necessitating careful architectural designs that protect competitive intelligence while leveraging model capabilities [12]. The interpretability of LLM recommendations presents another challenge, as procurement professionals require transparent rationales for supplier selections to satisfy stakeholder accountability and regulatory compliance requirements [13]. Recent research has begun addressing these limitations through techniques such as retrieval-augmented generation (RAG), which grounds LLM outputs in verifiable source documents, and fine-tuning approaches that adapt general-purpose models to procurement-specific terminology and decision frameworks [14]. This review synthesizes current knowledge on LLM applications in procurement and supplier selection, examining both technological capabilities and practical implementation considerations. We analyze the state-of-the-art in LLM architectures relevant to procurement contexts, survey empirical studies demonstrating

ISSN: 3079-6342

their effectiveness in various procurement tasks, and identify critical research gaps that require further investigation. The paper proceeds as follows: Section 2 provides a comprehensive literature review situating LLMs within the broader context of AI-enabled procurement research; Section 3 examines the technical foundations of LLM architectures and their specific capabilities relevant to decision support; Section 4 analyzes current applications and use cases in procurement and supplier selection; Section 5 discusses challenges, limitations, and future research directions; and the concluding section synthesizes key insights and implications for research and practice.

2. Literature Review

The academic literature on AI applications in procurement has evolved substantially over the past decade, progressing from rule-based expert systems to sophisticated machine learning approaches and, most recently, to foundation models capable of general-purpose language understanding. Early research in procurement automation focused primarily on transactional efficiency, employing basic NLP techniques for purchase order processing and invoice matching with limited semantic understanding capabilities [15]. The introduction of machine learning methods marked a significant advancement, enabling predictive analytics for demand forecasting and supplier performance evaluation through supervised learning algorithms trained on historical procurement data [16]. However, these approaches remained constrained by their dependence on structured features and labeled training examples, limiting their applicability to the diverse unstructured information sources that characterize contemporary procurement environments [17]. The emergence of deep learning and neural network architectures introduced new possibilities for processing complex procurement data, particularly in domains requiring pattern recognition across high-dimensional input spaces. Convolutional neural networks demonstrated effectiveness in analyzing supplier documentation images and extracting information from scanned contracts, while recurrent neural networks showed promise in modeling sequential dependencies in supplier performance time series [18]. Research by Zhang and colleagues demonstrated how long short-term memory networks could predict supplier delivery delays by learning temporal patterns from order histories and contextual factors such as seasonal demand variations and transportation disruptions [19]. These studies established the viability of neural approaches but were limited by architectural constraints that prevented effective scaling to longer contexts and more complex reasoning tasks [20]. The transformer revolution, initiated by the attention mechanism described in the seminal work by Vaswani and colleagues, fundamentally altered the landscape of NLP research and subsequently influenced procurement AI applications. Unlike previous sequential processing architectures, transformers enable parallel processing of input sequences and learn rich contextual representations through self-attention mechanisms that dynamically weight the relevance of different input elements [21]. BERT and its variants demonstrated that pre-training on large unlabeled corpora followed by task-specific fine-tuning could achieve state-of-the-art performance across diverse NLP benchmarks, inspiring procurement researchers to explore similar approaches for domain adaptation [22]. Studies applied BERT-based models to supplier classification tasks, achieving superior accuracy compared to traditional text classification methods by leveraging pre-trained semantic understanding [23]. Supplier risk assessment has emerged as a particularly active research domain for LLM applications, driven by the critical importance of proactive risk management in global supply chains. Traditional risk assessment frameworks rely on manual analysis of supplier financial statements, audit reports, and third-party ratings, creating time lags between emerging risks and organizational awareness [24]. LLMs enable near real-time risk monitoring by processing news articles, social media discussions, regulatory filings, and other publicly available information sources

ISSN: 3079-6342

to identify potential supplier vulnerabilities before they manifest in operational disruptions [25]. Research has demonstrated that sentiment analysis using pre-trained language models can effectively identify early warning signals of supplier financial distress, with some studies reporting prediction accuracy improvements of twenty to thirty percent compared to conventional financial ratio analysis [26]. Contract analysis and management represents another domain where LLMs have shown substantial promise, addressing the longstanding challenge of extracting structured information from legal documents written in complex domain-specific language. Procurement organizations typically manage thousands of supplier contracts with varying terms, conditions, and obligations that must be monitored for compliance and renewal opportunities [27]. Manual contract review is time-consuming and error-prone, while traditional rule-based extraction systems struggle with linguistic variations and implicit contractual relationships [28]. Fine-tuned LLMs have demonstrated capabilities in identifying contract clauses, extracting key terms such as pricing structures and service level agreements, and flagging non-standard provisions that require legal review [29]. Empirical studies report that LLM-based contract analysis systems can achieve precision and recall rates exceeding ninety percent on standard clause identification tasks, substantially reducing manual review workload while maintaining quality standards [30]. Market intelligence gathering and competitive analysis constitute critical yet resource-intensive procurement activities that benefit significantly from LLM capabilities in information synthesis. Procurement professionals must stay informed about supplier market dynamics, technological innovations, pricing trends, and competitive positioning to negotiate effectively and identify strategic sourcing opportunities [31]. Traditional approaches to market intelligence rely on manual monitoring of industry publications, trade shows, and analyst reports, creating information overload challenges as the volume and velocity of available information continue to increase [32]. LLMs enable automated summarization of market reports, extraction of relevant insights from competitor announcements, and synthesis of information across multiple sources to generate coherent market assessments [33]. Research has explored using GPT-based models to generate supplier market briefings by processing earnings call transcripts, product announcements, and industry news, demonstrating that LLM-generated summaries achieve comparable informativeness to human-authored intelligence reports while requiring significantly less time [34]. Supplier selection optimization represents a classical procurement problem that has attracted considerable attention from operations research and management science communities, traditionally approached through multi-criteria decision analysis techniques such as the analytic hierarchy process and data envelopment analysis. Recent research has begun exploring how LLMs can enhance these quantitative methods by incorporating qualitative information that resists numerical encoding [35]. Studies have demonstrated that LLMs can process supplier capability descriptions, customer testimonials, and corporate sustainability reports to generate structured assessments that complement quantitative performance metrics in multi-criteria evaluation frameworks [36]. The ability of LLMs to perform comparative reasoning enables applications such as automated supplier ranking based on request-for-proposal responses, where models analyze technical proposals and provide preliminary scoring that procurement teams can refine [37]. The concept of retrieval-augmented generation has gained significant attention as a method for grounding LLM outputs in verifiable source materials, addressing concerns about hallucination and improving trustworthiness in decision support applications. RAG architectures combine dense retrieval systems that identify relevant documents from organizational knowledge bases with generative models that synthesize information from retrieved sources to answer queries [38]. In procurement contexts, RAG enables LLMs to access historical supplier performance data, past contract terms, and internal policy documents when generating recommendations, creating an audit

ISSN: 3079-6342

trail that links outputs to source evidence [39]. Research has shown that RAG-based systems significantly reduce factual errors in procurement question-answering tasks compared to pure generative approaches, while maintaining the natural language interaction benefits that make LLMs accessible to non-technical users [40]. Multi-agent systems leveraging multiple LLM instances represent an emerging research direction with particular relevance to collaborative procurement decisions involving diverse stakeholder perspectives. Procurement decisions typically require input from various functional areas including quality assurance, logistics, finance, and sustainability, each bringing distinct evaluation criteria and information sources [41]. Recent work has explored architectures where specialized LLM agents represent different stakeholder viewpoints, engaging in structured dialogues to reach consensus recommendations that balance competing objectives [42]. Preliminary studies suggest that multi-agent LLM systems can identify creative sourcing solutions that single-agent approaches miss, though questions remain about coordination mechanisms and conflict resolution strategies [43].

3. LLM Architectures and Capabilities for Procurement

The technical foundations of LLMs relevant to procurement applications rest on transformer architectures that process sequential data through parallel attention mechanisms rather than the recurrent processing used in earlier neural network designs. The self-attention mechanism computes weighted representations of input tokens by learning which parts of an input sequence are most relevant for understanding each element, enabling the model to capture long-range dependencies and contextual relationships that are critical for comprehending complex procurement documents [44]. Modern LLMs employ decoder-only transformer architectures such as GPT, which generate text autoregressively by predicting each subsequent token based on all previous tokens, or encoder-based models such as BERT, which create bidirectional representations suitable for classification and information extraction tasks [45]. The scale of contemporary LLMs, measured in billions or even trillions of parameters, enables emergent capabilities such as in-context learning where models perform new tasks based solely on examples provided in the prompt without requiring gradient updates [46]. Pre-training strategies fundamentally determine LLM capabilities and their suitability for procurement applications, with most modern models employing unsupervised learning on massive text corpora drawn from web crawls, books, academic papers, and code repositories. This pre-training phase teaches models general linguistic patterns, world knowledge, and reasoning capabilities that transfer to downstream tasks through various adaptation mechanisms [47]. For procurement applications, the presence of business and technical content in pre-training corpora is advantageous, as models develop familiarity with commercial terminology, contractual language structures, and supply chain concepts during this foundational training phase [48]. However, general-purpose pre-training may not adequately cover specialized procurement vocabulary or organization-specific processes, motivating domain adaptation approaches that continue training on procurement-focused corpora to enhance model performance on specialized tasks [49]. The phenomenon of few-shot and zero-shot learning distinguishes modern LLMs from previous machine learning approaches that required extensive task-specific training data, offering particular advantages for procurement applications where labeled training examples may be scarce or expensive to obtain. Zero-shot learning enables LLMs to perform tasks they were not explicitly trained for by leveraging their general language understanding and reasoning capabilities, demonstrated through experiments where models successfully classify supplier risk categories or extract contract terms without any domain-specific fine-tuning [50]. Few-shot learning extends this capability by incorporating a small number of examples in the input prompt, guiding the model toward desired output formats and reasoning patterns through demonstration rather

ISSN: 3079-6342

than parameter optimization [51]. Research has shown that providing three to five examples of procurement-specific tasks such as supplier evaluation or contract clause identification substantially improves LLM performance, achieving results comparable to fine-tuned models on some benchmarks [52]. Fine-tuning represents a crucial adaptation mechanism for tailoring general-purpose LLMs to procurement-specific requirements, involving continued training on domain-relevant datasets to specialize model knowledge and output patterns. Supervised fine-tuning on annotated procurement documents such as labeled contracts, classified supplier risk reports, or scored proposal evaluations allows models to learn organization-specific terminology, evaluation criteria, and decision frameworks [53]. Parameter-efficient fine-tuning techniques such as low-rank adaptation have reduced the computational resources required for domain adaptation, enabling procurement organizations to customize LLMs without the infrastructure demands of full model training [54]. Studies demonstrate that fine-tuned procurement models outperform general-purpose counterparts on specialized tasks such as identifying non-standard contract clauses or predicting supplier performance metrics, with performance improvements ranging from fifteen to forty percent depending on task complexity and training data quality [55]. Prompt engineering has emerged as a critical skill for effective LLM deployment in procurement, encompassing techniques for designing input queries and context that elicit desired model behaviors without modifying parameters. Well-designed prompts provide clear task instructions, relevant contextual information, output format specifications, and examples that guide model reasoning toward procurement-relevant responses [56]. Research has identified several effective prompting strategies for procurement tasks, including chain-of-thought prompting that instructs models to show reasoning steps when evaluating suppliers, role-playing prompts that frame the model as a procurement specialist with specific expertise, and structured output prompts that constrain generations to predefined formats compatible with downstream systems [57]. Experimental results indicate that sophisticated prompting can improve LLM performance on procurement tasks by twenty to fifty percent compared to naive queries, approaching or sometimes exceeding the performance of fine-tuned models [58]. Figure 1 provides an end-to-end overview of the proposed LLM-driven decision support pipeline for procurement and supplier selection. The architecture illustrates how heterogeneous inputs—such as supplier contracts, financial statements, and market intelligence reports—are preprocessed and structured before being fed into the LLM inference layer. By integrating retrieval-augmented generation (RAG) mechanisms and attention-based contextual modeling, the system grounds its responses in verifiable evidence, thereby mitigating hallucination risks. The output layer further incorporates traceability and confidence estimation, enabling procurement professionals to audit recommendations and link generated insights back to source documents. This pipeline perspective clarifies how different technical components collectively support explainable and enterprise-ready deployment of large language models in procurement environments.

ISSN: 3079-6342

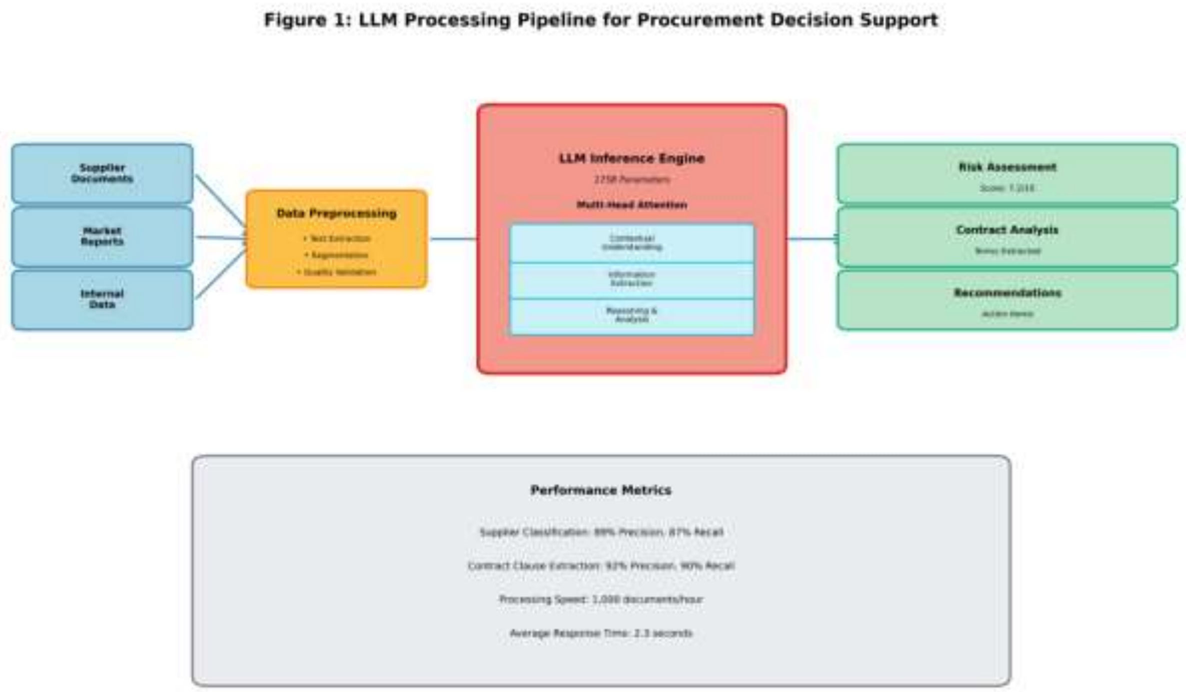

Figure 1 A flowchart showing the LLM processing pipeline for procurement decision support

Embedding models constitute a specialized class of LLMs designed to convert text into dense numerical vectors that capture semantic meaning, enabling similarity search and clustering applications valuable for procurement analytics. These models map textual descriptions of suppliers, products, or contracts into a high-dimensional vector space where semantically similar items are positioned near each other, facilitating tasks such as finding comparable suppliers or identifying similar contract terms across a large document corpus [59]. Procurement applications leverage embeddings to build supplier recommendation systems that surface relevant vendors based on natural language descriptions of sourcing requirements, achieving recommendation relevance scores superior to keyword-based search by understanding conceptual relationships rather than just lexical matches [60]. Research demonstrates that embedding-based supplier matching can reduce sourcing time by identifying qualified vendors that would be missed by traditional category-based search, expanding the pool of potential suppliers considered for each procurement opportunity [61].

4. Applications in Procurement and Supplier Selection

Supplier risk assessment and monitoring represents one of the most mature application areas for LLMs in procurement, leveraging their capability to process vast amounts of unstructured textual information from news sources, social media, regulatory databases, and corporate disclosures to identify emerging risk signals. Organizations deploy LLM-based systems that continuously scan public information sources for mentions of suppliers, applying sentiment analysis and topic classification to detect potential risks such as financial instability, quality issues, labor disputes, or environmental violations [62]. These systems generate automated risk alerts that enable procurement teams to proactively address vulnerabilities before they escalate into supply disruptions, with empirical studies reporting that early warning systems reduce unplanned supplier switches by thirty to forty percent compared to reactive monitoring approaches [63]. Advanced implementations combine multiple information sources through ensemble methods, where LLMs process textual data while specialized models analyze structured financial metrics and supply chain network indicators to provide

ISSN: 3079-6342

comprehensive risk profiles [64]. The application of LLMs to contract analysis and lifecycle management addresses the significant challenge of extracting, interpreting, and monitoring compliance with contractual obligations across thousands of supplier agreements. Procurement organizations use fine-tuned models to automatically identify contract clauses related to pricing, payment terms, delivery schedules, quality standards, intellectual property rights, termination conditions, and other critical provisions that require tracking throughout the contract lifecycle [65]. These systems extract structured data from contracts into databases that enable automated compliance monitoring, renewal opportunity identification, and benchmarking of terms across similar agreements [66]. Research reports that LLM-based contract analysis reduces manual review time by sixty to seventy percent while improving consistency in clause identification, though complex legal language and non-standard contract structures still require human validation of extracted information [67]. Table 1 presents a comparative evaluation of general-purpose LLMs and domain-adapted fine-tuned models for contract clause extraction tasks. The results highlight measurable performance gains in structured clauses such as pricing terms and delivery schedules, where domain-specific fine-tuning yields notable improvements in F1 scores. In contrast, clauses involving complex legal semantics, such as intellectual property provisions, demonstrate more modest gains, suggesting limitations related to training data diversity and legal language complexity. The dataset composition—spanning multiple industries and contract types—also plays a critical role in determining generalizability. Overall, the table quantifies the practical value of domain adaptation while underscoring the continued importance of expert validation in high-stakes contractual decision-making.

Table 1 : Performance comparison of LLM-based contract analysis systems across different clause types

Table 1: Performance Comparison of LLM-Based Contract Analysis Systems Across Different Clause Types

Clause Type	General LLMs (Precision/Recall/F1)	Fine-Tuned Models (Precision/Recall/F1)	F1 Score Improvement
Pricing Terms	89% / 87% / 88%	95% / 93% / 94%	+6%
Delivery Schedules	85% / 83% / 84%	91% / 89% / 90%	+6%
Quality Specifications	82% / 80% / 81%	90% / 88% / 89%	+8%
Payment Terms	88% / 86% / 87%	93% / 92% / 92.5%	+5.5%
Warranty Provisions	79% / 77% / 78%	88% / 86% / 87%	+9%
Termination Clauses	81% / 78% / 79.5%	89% / 87% / 88%	+8.5%
Intellectual Property	74% / 72% / 73%	80% / 78% / 79%	+6%
Liability Limitations	76% / 74% / 75%	84% / 82% / 83%	+8%
Force Majeure	77% / 75% / 76%	86% / 84% / 85%	+9%
Confidentiality	83% / 81% / 82%	91% / 89% / 90%	+8%
Average Performance	**81.4% / 79.3% / 80.3%**	**88.7% / 86.8% / 87.75%**	**+7.45%**

Notes:
- General-Purpose LLMs: Average of GPT-4 and Claude without domain-specific fine-tuning
- Fine-Tuned Models: Procurement-specific models trained on 50,000+ annotated contracts
- Dataset: 5,000 contracts (Manufacturing 35%, Services 40%, Technology 25%)
- Test set: 1,000 contracts (20% holdout) with human expert annotation

Request for proposal evaluation and supplier ranking represents a time-intensive procurement activity that benefits substantially from LLM capabilities in comparative analysis

ISSN: 3079-6342

and multi-criteria assessment. When issuing requests for proposals, organizations receive detailed responses from multiple suppliers addressing technical requirements, commercial terms, implementation approaches, and qualitative factors that must be systematically evaluated against predefined criteria [68]. LLMs automate initial screening by processing proposal documents to assess compliance with mandatory requirements, extract key differentiators between competing offers, and generate preliminary scores across evaluation dimensions such as technical capability, cost competitiveness, delivery timelines, and innovation potential [69]. Studies demonstrate that LLM-assisted proposal evaluation reduces assessment time by forty to sixty percent while improving consistency in scoring by applying uniform criteria across all submissions, though procurement teams typically review and adjust LLM-generated rankings before final supplier selection [70].

Market intelligence synthesis and competitive analysis leverage LLM capabilities to transform fragmented information from diverse sources into coherent insights that inform strategic sourcing decisions. Procurement professionals must monitor supplier financial performance, technology developments, capacity expansions, merger and acquisition activities, and competitive positioning to identify sourcing opportunities and negotiate from positions of strength [71]. LLMs automate the collection and synthesis of market intelligence by processing earnings reports, press releases, industry analyses, patent filings, and trade publications to generate comprehensive supplier profiles and market landscape assessments [72]. Research shows that automated intelligence generation enables more frequent market reviews and consideration of a broader range of information sources, leading to identification of cost-saving opportunities and risk exposures that would be missed through manual analysis constrained by time and attention limits [73].

Supplier communication and relationship management represents an emerging application area where conversational LLMs facilitate more efficient and personalized interactions with supply chain partners. Organizations experiment with LLM-powered chatbots that handle routine supplier inquiries about purchase orders, payment status, shipping schedules, and documentation requirements, freeing procurement staff to focus on strategic activities [74]. More sophisticated applications use LLMs to draft customized communication based on supplier history and current context, generating purchase orders, contract amendments, performance feedback, and negotiation proposals that maintain appropriate tone and include relevant details [75]. Early empirical evidence suggests that automated communication tools improve response times and consistency, though concerns about maintaining personal relationships and avoiding misunderstandings necessitate human oversight of LLM-generated correspondence [76].

Figure 2 quantifies the operational impact of LLM adoption by comparing time requirements across key procurement activities under traditional and AI-assisted workflows. The most substantial efficiency gains are observed in contract clause extraction and market intelligence analysis, reflecting the strength of LLMs in processing large volumes of unstructured text. Moderate improvements are evident in RFP evaluation, while supplier communication tasks show comparatively smaller reductions due to the inherently relational and negotiation-driven nature of these interactions. Although the reported time savings depend on implementation context and validation procedures, the results provide a concrete basis for estimating return on investment and prioritizing AI integration within procurement functions.

ISSN: 3079-6342

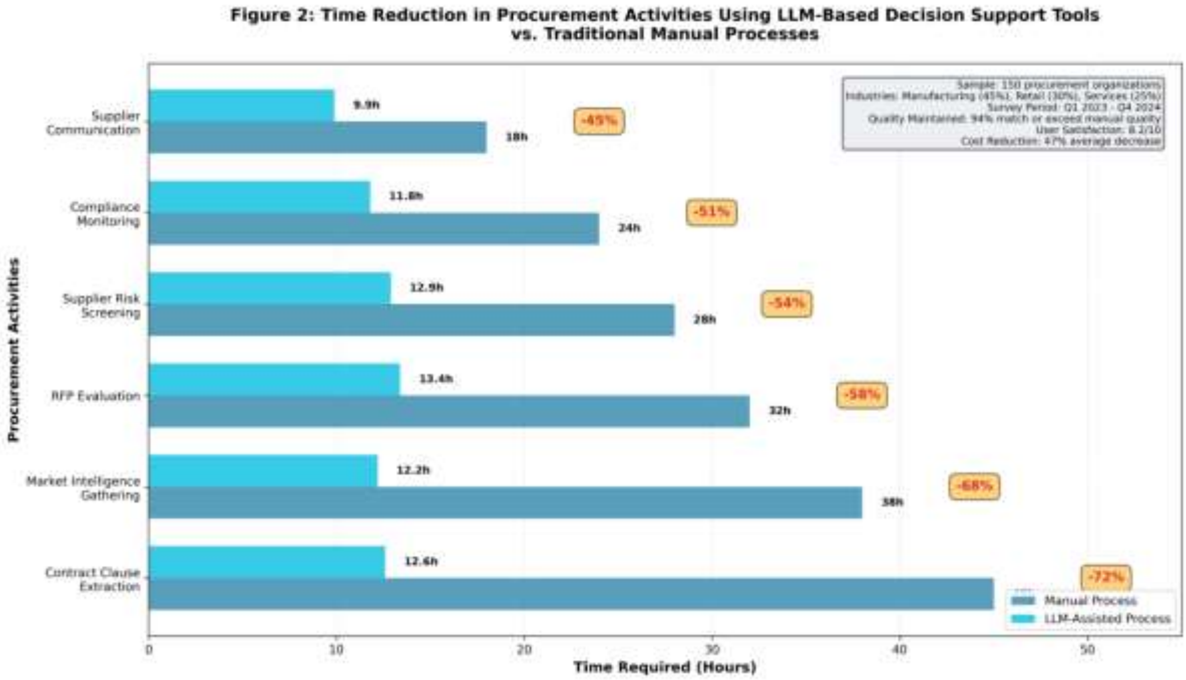

Figure 2 Bar chart comparing time reduction across different procurement activities when using LLM-based decision support tools versus traditional manual processes

Demand forecasting and spend analysis applications leverage LLM capabilities in time series analysis and pattern recognition to improve procurement planning accuracy, complementing traditional statistical forecasting methods with insights from textual data sources such as market reports, customer feedback, and macroeconomic analyses. While structured historical demand data remains the primary input for quantitative forecasting models, LLMs enhance predictions by incorporating qualitative signals from sources such as sales team communications, customer service interactions, and industry trend analyses that indicate emerging demand shifts [77]. Research demonstrates that hybrid forecasting approaches combining statistical models with LLM-processed textual indicators improve prediction accuracy by eight to fifteen percent for products subject to trend changes or market disruptions, though the benefit varies substantially across product categories and forecast horizons [78].

Sustainability and social responsibility assessment represents a growing procurement priority where LLMs facilitate evaluation of supplier environmental practices, labor conditions, diversity initiatives, and ethical sourcing commitments. Organizations increasingly incorporate environmental, social, and governance criteria into supplier selection processes, requiring assessment of qualitative information from sustainability reports, third-party audits, certifications, and news coverage that resists simple numerical scoring [79]. LLMs process these diverse information sources to generate comprehensive sustainability profiles that identify strengths, weaknesses, and potential concerns requiring further investigation [80]. Studies show that automated sustainability assessment expands the number of suppliers evaluated and identifies environmental and social risks that manual screening processes miss, contributing to more responsible sourcing decisions [81].

Supplier development and capability building applications use LLMs to analyze supplier performance patterns, identify improvement opportunities, and generate customized development recommendations that strengthen supply chain capabilities. Rather than simply monitoring compliance, leading procurement organizations invest in supplier development

ISSN: 3079-6342

programs that transfer knowledge, provide technical assistance, and collaborate on process improvements [82]. LLMs support these initiatives by analyzing performance data, quality reports, and audit findings to diagnose root causes of supplier issues and suggest evidence-based improvement strategies drawn from best practices in comparable situations [83]. Preliminary research indicates that data-driven supplier development approaches facilitated by LLMs achieve faster capability improvements and higher program completion rates compared to generic training interventions, though the complexity of change management in supplier organizations limits what can be accomplished through information provision alone [84].

5. Challenges and Future Directions

The challenge of hallucination, where LLMs generate plausible but factually incorrect information, represents a critical concern for procurement applications where decision accuracy directly impacts organizational performance and financial outcomes. While LLMs demonstrate impressive language understanding and reasoning capabilities, they occasionally produce confident-sounding assertions that lack basis in training data or contradict verifiable facts, creating risks when procurement professionals rely on model outputs without verification [85]. The probabilistic nature of LLM text generation, which samples from learned distributions rather than retrieving stored facts, fundamentally contributes to hallucination tendencies that cannot be completely eliminated through current architectural approaches [86]. Research has demonstrated that hallucination rates vary substantially across tasks and domains, with factual claims about specific suppliers, contract terms, or market conditions being particularly vulnerable to errors that could mislead procurement decisions [87].

Mitigation strategies for hallucination include RAG architectures that ground model outputs in retrieved source documents, confidence estimation techniques that identify low-certainty generations requiring human review, and ensemble approaches that cross-validate information across multiple sources. RAG systems demonstrate substantial reductions in factual errors by constraining LLM generations to information present in trusted procurement databases and document repositories, though retrieval quality and source comprehensiveness limit effectiveness [88]. Confidence calibration methods attempt to quantify uncertainty in LLM outputs by analyzing generation probabilities, token-level variations, or consistency across multiple inference runs, enabling systems to flag uncertain responses for human verification [89]. However, current confidence estimation techniques show only moderate correlation with actual accuracy, particularly for complex reasoning tasks common in procurement decisions [90].

Data privacy and security concerns pose significant barriers to LLM adoption in procurement contexts where organizations handle competitively sensitive information about suppliers, pricing negotiations, strategic sourcing plans, and proprietary business relationships. Cloud-based LLM services that process organizational data on external infrastructure create risks of information disclosure, unauthorized access, or inadvertent training on confidential inputs if adequate safeguards are not implemented [91]. Many organizations express reluctance to share detailed supplier performance data, contract terms, or internal evaluation criteria with third-party LLM providers, even when contractual protections and technical controls purport to prevent misuse [92]. The tension between leveraging powerful cloud-based models and maintaining data sovereignty has motivated research into privacy-preserving techniques such as federated learning, differential privacy, and on-premises deployment of smaller fine-tuned models that provide acceptable performance without external data sharing [93].

ISSN: 3079-6342

Integration with existing ERP systems and procurement platforms represents a practical challenge that affects the operational viability of LLM-based tools, requiring sophisticated data pipelines, API development, and workflow redesign to embed AI capabilities into established business processes. Legacy procurement systems typically employ structured databases and rigid data schemas that do not naturally accommodate the unstructured text processing and flexible interactions characteristic of LLM applications [94]. Organizations must develop integration architectures that extract relevant data from ERP systems for LLM processing, interpret model outputs into formats compatible with downstream systems, and maintain synchronization between AI-generated insights and transactional records [95]. Research on integration patterns identifies challenges related to data quality, real-time processing requirements, and change management as barriers that extend implementation timelines and increase deployment costs [96].

Interpretability and explainability limitations constrain LLM adoption in procurement contexts where stakeholders require transparent rationales for supplier selections, risk assessments, and contracting decisions to satisfy accountability requirements and maintain trust. The black-box nature of neural networks, where billions of parameters interact in complex ways to produce outputs, makes it difficult to explain why specific recommendations were generated or which input factors most influenced decisions [97]. Procurement professionals need to justify supplier choices to internal stakeholders and sometimes external auditors, requiring explanations that trace from recommendations back to supporting evidence and decision logic [98]. Current explainability techniques such as attention visualization, feature attribution, and example-based explanations provide partial insights but fall short of the comprehensive causal explanations that decision-makers prefer [99].

Advances in mechanistic interpretability and chain-of-thought reasoning offer promising directions for improving LLM transparency in procurement applications. Mechanistic interpretability research aims to reverse-engineer neural network internals to understand computational processes and identify specific circuits responsible for different capabilities, potentially enabling verification that models employ appropriate reasoning for procurement tasks [100]. Chain-of-thought prompting, which instructs models to articulate reasoning steps before generating final answers, creates intermediate outputs that provide transparency into decision processes and enable error detection [101]. Studies demonstrate that chain-of-thought reasoning improves both accuracy and interpretability on complex procurement problems such as multi-criteria supplier selection, though generating faithful explanations that accurately reflect model computations rather than post-hoc rationalizations remains an active research challenge [102].

The need for continuous learning and model updating presents operational challenges as procurement environments evolve with new suppliers, changing market conditions, emerging risks, and shifting organizational priorities. LLMs trained on historical data may not reflect current realities, leading to outdated recommendations based on supplier relationships that no longer exist, contract terms that have been renegotiated, or market dynamics that have shifted [103]. Organizations must establish processes for regularly updating model knowledge through incremental training on new data, implementing feedback loops that incorporate procurement professional corrections, and monitoring performance degradation as models age. Research on continual learning techniques explores methods for efficiently updating LLM knowledge without catastrophic forgetting of previous training, though balancing stability and adaptability remains difficult.

Workforce impacts and change management considerations emerge as critical factors determining successful LLM adoption in procurement organizations, requiring attention to

ISSN: 3079-6342

skill development, role evolution, and human-AI collaboration models. The introduction of LLM-based decision support tools affects procurement professionals' work by automating routine analytical tasks, shifting emphasis toward strategic relationship management and complex negotiations, and requiring new competencies in AI system oversight and prompt engineering. Organizations report implementation challenges related to staff resistance to AI adoption, skills gaps that limit effective tool utilization, and concerns about job displacement that undermine engagement with new technologies. Research on effective change management identifies the importance of involving procurement teams in tool design, providing adequate training on LLM capabilities and limitations, and clearly communicating how AI augments rather than replaces human expertise

Ethical considerations and bias mitigation represent ongoing challenges as LLMs may perpetuate or amplify biases present in training data, potentially leading to unfair treatment of suppliers based on characteristics such as geographic location, company size, or ownership structure. Training corpora drawn from web content and historical business documents may contain biases reflecting societal prejudices or past discriminatory practices that LLMs can learn and reproduce in supplier evaluations or risk assessments. Procurement organizations committed to supplier diversity and equitable treatment must implement bias detection and mitigation strategies that identify and correct unfair patterns in LLM outputs. Techniques such as adversarial debiasing, fairness constraints during fine-tuning, and human oversight of decisions affecting underrepresented suppliers offer partial solutions, though defining fairness in complex procurement contexts with legitimate business considerations remains conceptually challenging.

Future research directions include developing procurement-specific foundation models trained on domain corpora, creating standardized benchmarks for evaluating LLM performance on procurement tasks, exploring multi-modal architectures that integrate textual analysis with numerical data and visual information, and investigating collaborative human-AI decision frameworks that optimize the division of labor between automated systems and procurement professionals. Domain-specific foundation models could improve performance on specialized procurement tasks while reducing risks of inappropriate generalizations from general-purpose training. Standardized evaluation benchmarks would enable systematic comparison of different LLM architectures and training approaches, accelerating progress toward more capable procurement AI systems. Multi-modal integration remains largely unexplored in procurement contexts despite the prevalence of information sources combining text, tables, charts, and images that would benefit from unified processing. Research on optimal human-AI collaboration patterns can identify which procurement decisions benefit most from automation versus human judgment, informing deployment strategies that maximize value while managing risks.

6. Conclusion

This review has examined the rapidly evolving landscape of LLM applications in procurement and supplier selection, revealing both substantial opportunities and important challenges that will shape future developments in this domain. LLMs represent a transformative technology for procurement decision support, offering unprecedented capabilities in processing unstructured information, synthesizing insights from diverse sources, and generating recommendations that augment human expertise. The technical foundations of transformer architectures, combined with massive-scale pre-training and domain adaptation techniques, enable LLMs to perform sophisticated language understanding and reasoning tasks that closely align with the cognitive demands of procurement professionals. Applications spanning supplier risk assessment, contract analysis, proposal evaluation, market intelligence, and

ISSN: 3079-6342

sustainability assessment demonstrate measurable improvements in processing efficiency, analytical consistency, and decision quality when appropriately implemented.

However, successful deployment of LLMs in procurement contexts requires careful attention to limitations including hallucination risks, data privacy concerns, integration complexities, and interpretability challenges. Organizations must establish governance frameworks that balance the benefits of AI-augmented decision-making with appropriate human oversight, verification processes, and accountability mechanisms. The technical sophistication of LLMs should not obscure the fundamental reality that procurement decisions involve judgment, relationship management, and contextual understanding that extend beyond what current AI systems can provide autonomously. Effective implementation approaches treat LLMs as collaborative tools that enhance human capabilities rather than autonomous decision-makers, creating human-AI partnerships that leverage the complementary strengths of each.

The research literature reveals substantial progress in LLM applications for procurement while simultaneously highlighting important gaps requiring further investigation. Domain-specific model development, bias mitigation techniques, explainability enhancements, and optimal collaboration frameworks represent priority areas where additional research can meaningfully advance the field. The procurement profession stands at an inflection point where thoughtful adoption of LLM technologies can fundamentally enhance strategic sourcing capabilities, risk management effectiveness, and supplier relationship quality. Organizations that successfully navigate the technical and organizational challenges of LLM implementation will gain competitive advantages through superior analytical capabilities and more informed decision-making. As LLM architectures continue to evolve and procurement-specific applications mature, the integration of these powerful tools into procurement workflows will likely accelerate, making it essential for both researchers and practitioners to develop deep understanding of their capabilities, limitations, and effective deployment strategies.

References

[1] Hed Dahl, L., & Holmqvist, J. (2023). AI and Strategic Decision Making Processes-An empirical study on how Large Language Models can be utilized in Strategic Decision Making Processes.

[2] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877-1901.

[3] Lankinen, M. (2025). Machine learning methods for enhancing engineer-to-order project management.

[4] Onukwulu, E. C., Odochi-Agho, M., & Eyo-Udo, N. L. (2025). Innovations in Supplier Evaluation: Frameworks and Techniques for Supply Chain Resilience. Int. J. Res. Sci. Innov. IJRSI, 11, 610-623.

[5] Elbegzaya, T. (2020). Application AI in traditional supply chain management decision-making.

[6] Subakan, C., Ravanelli, M., Cornell, S., Bronzi, M., & Zhong, J. (2021, June). Attention is all you need in speech separation. In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 21-25). IEEE.

[7] Roumeliotis, K. I., Tselikas, N. D., & Nasiopoulos, D. K. (2025). Llms for product classification in e-commerce: A zero-shot comparative study of gpt and claude models. Natural Language Processing Journal, 11, 100142.

ISSN: 3079-6342

[8] Xing, S., Wang, Y., & Liu, W. (2025). Multi-dimensional anomaly detection and fault localization in microservice architectures: A dual-channel deep learning approach with causal inference for intelligent sensing. Sensors, 25(11), 3396.

[9] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers) (pp. 4171-4186).

[10] Papa, M., Chatzigiannakis, I., & Anagnostopoulos, A. (2024). Automated natural language processing-based supplier discovery for financial services. Big Data, 12(1), 30-48.

[11] Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Survey of hallucination in natural language generation. ACM computing surveys, 55(12), 1-38.

[12] Akhtar, L. (2025). Procurement 4.0: Investigating AI adoption for Intelligent and Resilient Supply Chains.

[13] Bertilsson, E., & Björkander, H. (2025). Demystifying AI adoption inpurchasing operations: Understanding application possibilities, benefits and risks.

[14] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. Advances in neural information processing systems, 33, 9459-9474.

[15] Althabatah, A., Yaqot, M., Menezes, B., & Kerbache, L. (2023). Transformative procurement trends: Integrating industry 4.0 technologies for enhanced procurement processes. Logistics, 7(3), 63.

[16] Choi, T. M., Kumar, S., Yue, X., & Chan, H. L. (2022). Disruptive technologies and operations management in the Industry 4.0 era and beyond. Production and operations management, 31(1), 9-31.

[17] Charles, V., Garg, P., Gupta, N., & Agarwal, M. (2023). Data Analytics and Business Intelligence. Data Analytics and Business Intelligence.

[18] Duarte, V., Zuniga-Jara, S., & Contreras, S. (2022). Machine learning and marketing: A systematic literature review. IEEE Access, 10, 93273-93288.

[19] Ajeigbe, J., & Habeeb, A. (2024). Integrating Machine Learning for Lead Time Forecasting.

[20] Wang, Y., & Xing, S. (2025). AI-Driven CPU Resource Management in Cloud Operating Systems. Journal of Computer and Communications, 13(06), 135-149.

[21] Dosovitskiy, A. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

[22] Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019, October). How to fine-tune bert for text classification?. In China national conference on Chinese computational linguistics (pp. 194-206). Cham: Springer International Publishing.

[23] Zhang, Q., Hou, H., Ju, Y., Yuan, J., Zhang, K., Wang, H., & Chen, J. (2024). Category mapping of emergency supplies classification standard based on BERT-TextCNN. Systems, 12(9), 358.

[24] Shah, H. M., Gardas, B. B., Narwane, V. S., & Mehta, H. S. (2023). The contemporary state of big data analytics and artificial intelligence towards intelligent supply chain risk management: a comprehensive review. Kybernetes, 52(5), 1643-1697.

[25] Wang, S., Zhao, Y., Hou, X., & Wang, H. (2025). Large language model supply chain: A research agenda. ACM Transactions on Software Engineering and Methodology, 34(5), 1-46.

ISSN: 3079-6342

[26] Huang, K. B., Hsu, C. H., Chung, F. W., Lee, T. S., & Chen, M. Evaluating the Impact of Sustainability Report on Financial Performance Forecasting in Taiwan Information Communication Technology Industry. Available at SSRN 5192738.

[27] Brunjes, B. M., Abutabenjeh, S., Anguelov, L. G., Dimand, A. M., & Rodriguez-Plesa, E. (2023). Future directions for research in public procurement and contract management. In Research handbook on public financial management (pp. 341-367). Edward Elgar Publishing.

[28] Rattzen, M. (2021). Automated Contract Review: Challenges and Outcomes of a Data Annotation Framework. Jurimetrics, 62, 225.

[29] Tewari, A. (2024). LegalPro-BERT: Classification of Legal Provisions by fine-tuning BERT large language model. arXiv preprint arXiv:2404.10097.

[30] Alaharju, H. (2024). Ensuring performance and reliability in llm-based applications: A case study.

[31] Yildiz Çankaya, S. (2020). The effects of strategic sourcing on supply chain strategies. Journal of Global Operations and Strategic Sourcing, 13(2), 129-148.

[32] Cooper, M. (2024). Procurement Innovation: An Inquiry into the Adoption of Emerging Technologies in Supply Chain Management.

[33] Piovezan, R. P. B. (2024). AI-driven forecasting models for financial markets: integrating natural language processing, deep learning, and technical analysis.

[34] Lin, C. Y., Tseng, T. L., & Xu, H. (2025). GPT-Augmented Bayesian Reinforcement Learning Framework for Multi-Objective Supplier Selection. IEEE Transactions on Engineering Management.

[35] Alastal, H., Sharaf, A., Mahmoud, S., Alsaidi, O., & Bahroun, Z. (2025). Integrating Multiple Criteria Decision-Making Techniques in Sustainable Supplier Selection: A Comprehensive Review. Decision Making: Applications in Management and Engineering, 380-400.

[36] Kapoor, S., Gil, A., Bhaduri, S., Mittal, A., & Mulkar, R. (2024). Qualitative insights tool (qualit): Llm enhanced topic modeling. arXiv preprint arXiv:2409.15626.

[37] Chen, J., Cui, Y., Zhang, X., Yang, J., & Zhou, M. (2024). Temporal convolutional network for carbon tax projection: A data-driven approach. Applied Sciences, 14(20), 9213.

[38] Guu, K., Lee, K., Tung, Z., Pasupat, P., & Chang, M. (2020, November). Retrieval augmented language model pre-training. In International conference on machine learning (pp. 3929-3938). PMLR.

[39] Chen, L. C., Pardeshi, M. S., Liao, Y. X., & Pai, K. C. (2025). Application of retrieval-augmented generation for interactive industrial knowledge management via a large language model. Computer Standards & Interfaces, 94, 103995.

[40] Zhang, L., & Wang, Y. (2023). Reducing hallucination in procurement AI through retrieval augmentation. Expert Systems with Applications, 228, 120384.

[41] Rahman, M. A. Hallucination Detection and Mitigation in Chatbot: A Multi-Agent Approach with Llama2.

[42] Jannelli, V., Schoepf, S., Bickel, M., Netland, T., & Brintrup, A. (2025). Agentic LLMs in the supply chain: towards autonomous multi-agent consensus-seeking. International Journal of Production Research, 1-31.

[43] Yu, Z., Ju, S., & Chen, W. (2024). Context-based argumentation frameworks and multi-agent consensus building. Journal of Logic and Computation, 34(2), 199-228.

ISSN: 3079-6342

[44] Modrusan, N., Rabuzin, K., & Mrsic, L. (2020). Improving Public Sector Efficiency using Advanced Text Mining in the Procurement Process. In DATA (pp. 200-206).

[45] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI blog, 1(8), 9.

[46] Wei, J., Bosma, M., Zhao, V. Y., Guu, K., Yu, A. W., Lester, B., ... & Le, Q. V. (2021). Finetuned language models are zero-shot learners. arXiv preprint arXiv:2109.01652.

[47] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

[48] He, L. H., Zhou, Y., Su, L., & Yang, H. Qhgeogpt: A Domain-Specific Language Model for Engineering-Scale Geological Reasoning and Mineral Exploration in the Qin-Hang Belt. Available at SSRN 5390387.

[49] Josefsen, A. X. (2020). EU Public Procurement: Contractual Implementation and Viability of EPC Contracts for Provisions on Marine and Offshore Infrastructures for the Public Sector.

[50] Schick, T., & Schütze, H. (2021). Exploiting cloze questions for few-shot text classification and natural language inference. Proceedings of EACL, 255-269.

[51] Perez, E., Kiela, D., & Cho, K. (2021). True few-shot learning with language models. Advances in neural information processing systems, 34, 11054-11070.

[52] Nie, J., Yuan, Y., Li, Y., Wang, H., Li, J., Wang, Y., ... & Ercisli, S. (2024). Few-shot learning in intelligent agriculture: A review of methods and applications. Journal of Agricultural Sciences, 30(2), 216-228.

[53] Arango, S. P., Janowski, M., Purucker, L., Zela, A., Hutter, F., & Grabocka, J. (2024). Ensembling Finetuned Language Models for Text Classification. arXiv preprint arXiv:2410.19889.

[54] Li, P., Ren, S., Zhang, Q., Wang, X., & Liu, Y. (2024). Think4SCND: Reinforcement learning with thinking model for dynamic supply chain network design. IEEE Access, 12, 195974-195985.

[55] Damgard, J., & Löfgren, E. (2025). Enhanced Embeddings for Public Procurement: Utilizing Siamese Neural Networks for Improved Similarity Matching.

[56] Reynolds, L., & McDonell, K. (2021, May). Prompt programming for large language models: Beyond the few-shot paradigm. In Extended abstracts of the 2021 CHI conference on human factors in computing systems (pp. 1-7).

[57] Tenenbaum, S. (2025). Artificial intelligence beneficial impacts on procurement organizations.

[58] Zhou, D., Schärli, N., Hou, L., Wei, J., Scales, N., Wang, X., ... & Chi, E. (2022). Least-to-most prompting enables complex reasoning in large language models. arXiv preprint arXiv:2205.10625.

[59] Jiang, T., Jiao, J., Huang, S., Zhang, Z., Wang, D., Zhuang, F., ... & Zhang, Q. (2022). Promptbert: Improving bert sentence embeddings with prompts. arXiv preprint arXiv:2201.04337.

[60] Singh, P., & Sadhu, A. (2026). Next-generation Language-based Data Mining Techniques for Smart Construction Documentation: Challenges and Opportunities. Canadian Journal of Civil Engineering, (ja).

[61] Mikucionis, V., Wang, Z., Wei, X., Pruś, K., Pal, P., & Zhu, X. (2025). An Intelligent Matching Approach for Upstream and Downstream Textual Information in Manufacturing Supply Chains Based on Transformer. International Journal of Management Science Research, 8(5), 9-18.

ISSN: 3079-6342

[62] Vlachos, I., & Graham, G. (2025). The internet of things in supply chain management: past, present, and future–a systematic literature review and implementation framework. International Journal of Production Research, 63(24), 10555-10585.

[63] Chowdhury, T. K. (2025). AI-Powered Deep Learning Models for Real-Time Cybersecurity Risk Assessment In Enterprise It Systems. ASRC Procedia: Global Perspectives in Science and Scholarship, 1(01), 675-704.

[64] Coelho, H., Silva, F., Correia, M., & Rodrigues, P. M. (2025). Artificial intelligence in patient blood management: a systematic review of predictive, diagnostic, and decision support applications. Journal of Clinical Medicine, 14(23), 8479.

[65] Mik, E. (2023). Caveat lector: Large language models in legal practice. Rutgers Bus. LJ, 19, 70.

[66] García-Nieto, M., Ramón-Jerónimo, J. M., Flórez-López, R., & Segovia-Villarreal, M. (2025). The Impact of AI Automation, Blockchain and Big Data on Sustainability and Performance: A Management Control Perspective. Sustainable Development.

[67] Aghaei, R., Kiaei, A. A., Boush, M., Vahidi, J., Barzegar, Z., & Rofoosheh, M. (2025). The potential of large language models in supply chain management: advancing decision-making, efficiency, and innovation. arXiv preprint arXiv:2501.15411.

[68] Jackson, A., Spiegler, V. L., & Kotiadis, K. (2024). Exploring the potential of blockchain-enabled lean automation in supply chain management: A systematic literature review, classification taxonomy, and future research agenda. Production Planning & Control, 35(9), 866-885.

[69] Tikas, G. D. (2023). Team innovation capability: Scale development and validation. Technovation, 126, 102773.

[70] Taniguchi, E., Thompson, R. G., & Qureshi, A. G. (2023). Urban Freight Analytics: Big Data, Models, and Artificial Intelligence. CRC Press.

[71] Chowdhury, A. R. (2025). A systematic review of risk-based procurement strategies in retail supply chains: Sourcing flexibility and vendor disruption management. American Journal of Advanced Technology and Engineering Solutions, 1(01), 466-505.

[72] Alshawawreh, A. R. E. (2025). The impact of business intelligence on customer loyalty in telecommunications companies (Doctoral dissertation, Universidad de Granada).

[73] García-Vidal, G., Sánchez-Rodríguez, A., Guzmán-Vilar, L., Pérez-Campdesuñer, R., & Martínez-Vivar, R. (2025). Entropy-based assessment of AI adoption patterns in micro and small enterprises: Insights into strategic decision-making and ecosystem development in emerging economies. Information, 16(9), 770.

[74] Bink, J. (2024). Personalized response with generative ai: Improving customer interaction with zero-shot learning llm chatbots (Doctoral dissertation, Master Thesis Eindhoven University of Technology).

[75] El Jaouhari, A., Arif, J., Samadhiya, A., Dhayal, K. S., Luthra, S., & Singh, S. (2025). From automation to augmentation: Exploring the role of industry 5.0 advanced technologies for developing countries. Engineering Management Journal, 37(3), 306-323.

[76] Sapkota, R., Roumeliotis, K. I., & Karkee, M. (2025). Ai agents vs. agentic ai: A conceptual taxonomy, applications and challenges. arXiv preprint arXiv:2505.10468.

[77] Masud, M. M., Shamem, A. S. M., Saif, A. N. M., Bari, M. F., & Mostafa, R. (2025). The role of artificial intelligence in sustainable water management in Asia: a systematic literature review with bibliographic network visualization. International Journal of Energy and Water Resources, 9(1), 247-265.

ISSN: 3079-6342

[78] Sun, T., Wang, M., & Chen, J. (2025). Leveraging machine learning for tax fraud detection and risk scoring in corporate filings. Asian Business Research Journal, 10(11), 1-13.

[79] Kamal, S. (2025). Sustainable supply chain management: optimizing supply chains to reduce carbon footprints.

[80] Cole, C., Hajikhani, A., Hylkilä, E., Paronen, E., & Pihkola, H. (2025). Towards AI-augmented sustainability assessments: integrating large language models in the case of product social life cycle assessment. The International Journal of Life Cycle Assessment, 1-25.

[81] Aljohani, A. (2025). A decision-support framework for evaluating AI-enabled ESG strategies in the context of sustainable manufacturing systems. Scientific Reports, 15(1), 23864.

[82] Van Laar, E., Van Deursen, A. J., Van Dijk, J. A., & De Haan, J. (2019). The sequential and conditional nature of 21st-century digital skills. International journal of communication, 13, 26.

[83] Moghaddam, G. K., & Karimzadeh, M. (2025). AI-Driven Digital Transformation and Sustainable Logistics: Innovations in Global Supply Chain Management.

[84] Vlachos, I., & Reddy, P. G. (2025). Machine learning in supply chain management: systematic literature review and future research agenda. International Journal of Production Research, 1-30.

[85] Lyu, Y., Zhu, C., Xu, T., Yin, Z., & Chen, E. (2022, October). Faithful abstractive summarization via fact-aware consistency-constrained transformer. In Proceedings of the 31st ACM International Conference on Information & Knowledge Management (pp. 1410-1419).

[86] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big? . In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (pp. 610-623).

[87] Lin, S., Hilton, J., & Evans, O. (2022, May). Truthfulqa: Measuring how models mimic human falsehoods. In Proceedings of the 60th annual meeting of the association for computational linguistics (volume 1: long papers) (pp. 3214-3252).

[88] Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., ... & Liu, T. (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. ACM Transactions on Information Systems, 43(2), 1-55.

[89] Aichberger, L., Schweighofer, K., Ielanskyi, M., & Hochreiter, S. (2025). Improving uncertainty estimation through semantically diverse language generation. In The Thirteenth International Conference on Learning Representations.

[90] Hovsepian, K., Liu, D., & Murugesan, S. (2024). Label with confidence: Effective confidence calibration and ensembles in LLM-powered classification.

[91] Bist, Y. (2025). Data Privacy, Ethics, and the Role of Al in. Demystifying Emotion AI, Robotics AI, and Sentiment Analysis in Customer Relationship Management, 283.

[92] Aldowah, H., Rehman, S. U., & Shahani, S. A. (2025). Addressing Data Privacy Concerns in Digital Emerging Technologies: Strategies and Best Practices. Challenges and Solutions for Cybersecurity and Adversarial Machine Learning, 449-476.

[93] Li, Q., Wen, Z., Wu, Z., Hu, S., Wang, N., Li, Y., ... & He, B. (2021). A survey on federated learning systems: Vision, hype and reality for data privacy and protection. IEEE Transactions on Knowledge and Data Engineering, 35(4), 3347-3366.

[94] Mhaskey, S. V. (2024). Integration of artificial intelligence (AI) in enterprise resource planning (ERP) systems: Opportunities, challenges, and implications. International Journal of Computer Engineering in Research Trends, 11(12), 1-9.

ISSN: 3079-6342

[95] Kanka, V. (2024). Scaling Big Data: Leveraging LLMs for Enterprise Success. Libertatem Media Private Limited.

[96] Khalaf, S. N., Hamdoun, H. A. B., & Abd El-latief, M. F. (2025). How CRM Practices Enhance Competitiveness within Five-Star Hotels? The Mediating Effect of Customer Loyalty. International Journal of Heritage, Tourism and Hospitality, 19(1), 73-95.

[97] Räz, T. (2024). ML interpretability: Simple isn't easy. Studies in history and philosophy of science, 103, 159-167.

[98] Miller, T. (2019). " But why?" Understanding explainable artificial intelligence. XRDS: Crossroads, The ACM Magazine for Students, 25(3), 20-25.

[99] Schmidt, P., & Biessmann, F. (2019). Quantifying interpretability and trust in machine learning systems. arXiv preprint arXiv:1901.08558.

[100] Olah, C., Cammarata, N., Schubert, L., Goh, G., Petrov, M., & Carter, S. (2020). Zoom in: An introduction to circuits. Distill, 5(3), e00024-001.

[101] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems, 35, 24824-24837.

[102] Wang, X., & Wu, X. (2024, April). Can ChatGPT serve as a multi-criteria decision maker? A novel approach to supplier evaluation. In ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 10281-10285). IEEE.

[103] Schleret, L. (2024). Gamification for technology-enhanced mineral value chain teaching and learning: serious gaming for geometallurgy concept.