Journal of Integrated Science, Technology and Management, 2025; 1(1):76-80

Journal of Integrated Science, Technology and Management

Research Article

Large Language Models in Enterprise Resource Planning: A Comprehensive Review of Applications, Challenges, and Future Directions

A. K. Srivastava

Chief Technology Officer, Light Field Corporation, Philadelphia, PA, USA

ARTICLE INFO

Article history:

Received: 30 July, 2025

Revised: 25 September, 2025

Accepted: 29 October, 2025

Published: 30 December, 2025

Keywords:

Large Language Models, Enterprise Resource Planning, SAP ERP, Natural Language Interfaces, Retrieval-Augmented Generation, Robotic Process Automation, Distributed Systems, Cloud Computing

ABSTRACT

Large language models (LLMs), built on the Transformer architecture and trained on large text corpora, are increasingly proposed as a natural language interface and content-generation layer for enterprise resource planning (ERP) systems, extending capabilities that have historically depended on structured transaction codes and purpose-built reporting tools. This paper presents a comprehensive review of LLM applications within ERP systems, organized around a four-part application taxonomy: natural language query and reporting over enterprise data; automated content and document generation; augmentation of intelligent process automation; and augmentation of embedded analytics and user experience. The review grounds this taxonomy in recent SAP ECC research demonstrating robotic process automation (RPA) combined with machine learning and SAP Fiori-based embedded analytics, both of which represent natural extension points for LLM capability within an existing ERP landscape, and further examines behavioral analytics for fraud detection as an adjacent enterprise analytics capability relevant to LLM-augmented risk narration and explanation. Because LLM-enabled ERP functionality depends on reliable access to live enterprise data across an increasingly distributed application landscape, the review also examines distributed service discovery and communication architecture and intelligent cloud resource scheduling as infrastructural preconditions for deploying LLM capability reliably and cost-effectively at enterprise scale. Drawing on the foundational Transformer, large-scale pre-training, and retrieval-augmented generation literature, the review develops a comparative analysis of the four LLM application types across their grounding requirements, autonomy level, and infrastructural demands, and identifies open challenges in grounding LLM outputs in accurate enterprise data, governance of generated content, and integration with existing SAP business logic, concluding with directions for future research and practice.

Introduction

Enterprise resource planning (ERP) systems have historically presented users with structured, transaction-code-driven interfaces and purpose-built reporting tools, reflecting the structured data models and explicit business rule logic that govern core organizational processes such as finance, procurement, and order management. Large language models (LLMs), built on the Transformer architecture and pre-trained on large text corpora, offer a qualitatively different interaction paradigm, natural language understanding and generation, that is increasingly being explored as a complement to, rather than a replacement for, this structured ERP foundation.

This paper presents a comprehensive review of LLM applications within ERP systems, with a particular focus on SAP ERP environments given the depth of recent SAP ECC modernization research available. The review organizes LLM applications into four categories: natural language query and reporting over enterprise data; automated

***Corresponding Author:** A. K. Srivastava

Address: Chief Technology Officer, Light Field Corporation, Philadelphia, PA, USA

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

© 2025, First author, This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

content and document generation; augmentation of intelligent process automation; and augmentation of embedded analytics and user experience. It grounds this taxonomy in recent SAP ECC research demonstrating RPA-and-machine-learning workflow automation and SAP Fiori-based embedded analytics, both representing natural extension points for LLM capability, and examines behavioral analytics for fraud detection as an adjacent capability relevant to LLM-generated risk explanation. Because reliable LLM-enabled ERP functionality depends on access to live, accurate enterprise data across an increasingly distributed application landscape, the review further examines distributed service discovery and intelligent cloud resource scheduling as infrastructural preconditions for this capability.

The objectives of this review are fourfold: (1) to present a taxonomy of LLM application types within ERP systems; (2) to examine each application type's grounding in existing SAP ERP modernization research; (3) to review the distributed systems and cloud infrastructure capabilities required to support LLM-enabled ERP reliably at scale; and (4) to develop a comparative analysis across application types and identify open challenges and future directions.

The remainder of this paper is organized as follows. Section 2 reviews the foundational LLM technology relevant to enterprise applications. Section 3 presents the four-part application taxonomy. Section 4 examines fraud-detection behavioral analytics as an adjacent, LLM-relevant analytics capability. Section 5 reviews the distributed service infrastructure required to support LLM-enabled ERP. Section 6 reviews cloud resource scheduling as a further infrastructural requirement. Section 7 presents a comparative analysis across application types. Section 8 discusses open challenges, Section 9 outlines future directions, and Section 10 concludes the paper.

Foundational Large Language Model Technology

Modern large language models are built on the Transformer architecture, which replaced earlier recurrent neural network approaches to sequence modeling with a self-attention mechanism capable of modeling long-range dependencies in text more effectively and with substantially greater training parallelism. Subsequent large-scale pre-training of Transformer-based models on large text corpora, exemplified by bidirectional encoder representations and large generative language models, has demonstrated strong natural language understanding and, at sufficient model and data scale, substantial few-shot task generalization without task-specific fine-tuning, a capability that underlies much of the current interest in applying LLMs to previously bespoke enterprise natural-language tasks.

A critical limitation of LLMs relevant to enterprise deployment, however, is that their pre-trained knowledge is static and cannot reflect an individual organization's current transactional state; retrieval-augmented generation addresses this limitation by combining a pre-trained language model with a retrieval mechanism that supplies relevant, current documents or data at inference time, allowing the model's output to be grounded in external, up-to-date information rather than relying solely on its pre-trained parameters. This retrieval-augmented pattern is directly relevant to each of the four ERP application types examined in Section 3, since each requires LLM outputs to be grounded in an organization's current SAP data rather than the model's static pre-trained knowledge.

A Taxonomy of LLM Applications in ERP Systems

Natural Language Query and Reporting

The most directly accessible LLM application within ERP systems is natural language query and reporting, in which a user poses a question in ordinary language, such as a request for current inventory status or outstanding invoices, and the LLM translates this request into an appropriate query against underlying SAP data, returning a generated natural language summary of the result. This application requires high grounding fidelity, since the user directly relies on the returned data for operational decisions, but imposes comparatively low autonomy demands, since the system remains read-only and the human user retains full interpretive control over the result.

Content and Document Generation

A second application category uses LLMs to draft enterprise content and documents, such as customer correspondence, exception explanations, or procurement documentation, informed by relevant SAP data context. This application tolerates somewhat lower grounding stakes than direct data query, since generated drafts are typically reviewed by a human before use, but still benefits substantially from grounding in accurate reference data to minimize the burden of human review and correction.

Augmentation of Intelligent Process Automation

A third and structurally more demanding application integrates LLM capability directly into automated process execution, building on recent SAP ECC research demonstrating a framework combining robotic process automation (RPA) with machine learning to automate standard SAP ECC business processes. Within this framework, an LLM component could extend

existing machine learning sub-tasks, such as document classification, by directly interpreting unstructured free-text inputs, such as customer emails or scanned correspondence, and generating structured process actions, a natural language extension of the judgment-based automation already demonstrated in this RPA-and-machine-learning research. This application category carries the highest grounding and autonomy demands of the four, since its outputs feed directly into live transaction execution rather than remaining subject to full human review before action.

Augmentation of Embedded Analytics and User Experience

A fourth application extends the embedded analytics capability documented in SAP ECC modernization research, which surfaces relevant business context and key performance indicators directly within transactional workflows through SAP Fiori expansion, by using an LLM to generate natural language explanations or summaries of the analytical content already being surfaced. Rather than introducing new analytical capability, this application improves the accessibility and interpretability of existing embedded analytics, building directly on the user-experience foundation already established by this research, and consequently carries comparatively low autonomy demands, since it supplements rather than replaces the existing human-reviewed analytics display.

Fraud Detection Analytics as an Adjacent LLM-Relevant Capability

Behavioral analytics research for fraud detection in digital payment systems, in which machine learning models trained on transaction and user behavior data identify anomalous patterns indicative of fraud, represents a capability adjacent to, though distinct from, the four LLM application types reviewed in Section 3. A natural extension point lies in content generation, discussed in Section 3.2: an LLM can be used to translate a fraud model's numerical risk score and the underlying behavioral features driving it into a natural language explanation for a human reviewer, improving the interpretability of an otherwise opaque behavioral analytics output without altering the underlying fraud detection model itself.

This pattern, using an LLM to narrate or explain the output of a separate, specialized machine learning model rather than performing the underlying analytical task itself, is a recurring theme across enterprise LLM applications generally, and is consistent with the content generation and embedded analytics augmentation application types discussed in Sections 3.2 and 3.4, reinforcing the view that LLMs are most effectively deployed within ERP systems as a complementary interaction and explanation layer rather than a replacement for existing specialized analytics capability.

Distributed Service Infrastructure for LLM-Enabled ERP

Each of the four application types reviewed in Section 3 requires an LLM component to reliably access live SAP data, whether to answer a query, inform generated content, feed a process automation step, or explain embedded analytics content. Research on service discovery and communication in distributed enterprise applications addresses the resulting integration requirement, examining how distributed application components locate and reliably communicate with one another across a dynamic enterprise application landscape, a capability directly applicable to the retrieval-augmented generation pattern discussed in Section 2, which depends on reliable, low-latency access to current enterprise data services at inference time.

Without reliable distributed service discovery of the kind examined in this research, LLM components risk operating on stale, incomplete, or unavailable data, undermining the grounding requirements identified for each application type in Table 1 and Section 3, particularly the high-grounding-stakes query and process automation applications discussed in Sections 3.1 and 3.3.

Table 1: Comparative Characteristics of LLM Application Types Within ERP Systems

LLM Application Type	*Grounding Requirement*	*Autonomy Level*	*Infrastructural Demand*
Natural language query & reporting	High; must reflect current transactional data	Low; read-only, human interprets result	Moderate; live data access via service layer
Content & document generation	Moderate; factual accuracy matters, lower stakes	Low to moderate; human reviews before use	Moderate; access to templates and reference data
Process automation augmentation	High; feeds directly into transaction execution	Moderate to high; acts within defined process	High; tight integration with SAP process logic
Embedded analytics & UX augmentation	Moderate; explains or summarizes existing analytics	Low; supplements human-reviewed dashboards	Moderate; access to analytics and KPI data

Intelligent Cloud Resource Scheduling for LLM Inference Workloads

A second infrastructural requirement concerns the compute resources required to serve LLM inference workloads reliably and cost-effectively alongside existing SAP transactional processing. Research on intelligent resource scheduling frameworks for cloud computing platforms addresses this requirement directly, examining learning-based approaches to allocating compute resources dynamically in response to variable workload demand, a capability well suited to the variable, often bursty inference demand characteristic of natural language query and content generation applications discussed in Sections 3.1 and 3.2.

Applied to LLM-enabled ERP specifically, intelligent resource scheduling of this kind is necessary to sustain acceptable response latency for interactive applications, such as natural language query, discussed in Section 3.1, without over-provisioning compute capacity during periods of low demand, a cost-efficiency consideration that becomes increasingly significant as LLM-enabled applications scale from pilot deployment to broad organizational use across a SAP landscape.

Comparative Analysis Across Application Types

Table 1 synthesizes the four-part taxonomy presented in Section 3 across three dimensions: grounding requirement, autonomy level, and infrastructural demand.

The comparison in Table 1 indicates that process automation augmentation, discussed in Section 3.3, carries the highest combined grounding, autonomy, and infrastructural demands of the four application types, consistent with its direct integration into live transaction execution, while natural language query and embedded analytics augmentation, discussed in Sections 3.1 and 3.4, remain comparatively lower-risk, human-reviewed applications better suited to early adoption.

Open Challenges

Grounding LLM Outputs in Accurate, Current SAP Data: As discussed throughout Section 3 and reinforced in Table 1, every application type depends on reliable grounding in current enterprise data; achieving this reliably, particularly for the high-stakes process automation application discussed in Section 3.3, remains an open technical challenge extending beyond the retrieval-augmented generation techniques discussed in Section 2.

Latency and Cost of LLM Inference at Enterprise Scale: Even with the intelligent resource scheduling capability discussed in Section 6, serving LLM inference reliably within the response-time expectations of interactive ERP use cases, discussed in Section 3.1, at acceptable cost remains a significant practical constraint on broad deployment.

Governance of LLM-Generated Content and Actions: The varying autonomy levels identified in Table 1 imply correspondingly varying governance requirements; organizations risk applying insufficiently rigorous review to higher-autonomy applications, such as process automation augmentation, or excessively burdensome review to lower-autonomy applications, such as query and reporting, if governance is not explicitly calibrated to each application type's autonomy level.

Integration With Existing SAP Business Logic: As discussed in Section 3.3, LLM-augmented process automation must integrate with, rather than bypass, existing SAP business rule logic; ensuring this integration preserves the auditability and compliance guarantees of the existing rule-governed process remains an open architectural challenge.

Reliability of Distributed Service Discovery Under Failure: As discussed in Section 5, LLM-enabled ERP applications depend on reliable distributed service discovery; ensuring graceful degradation, rather than silent grounding failure, when underlying data services are partially unavailable remains an underaddressed reliability challenge.

Future Directions

Deeper Retrieval-Augmented Integration With SAP Data Models: Building on the retrieval-augmented generation pattern discussed in Section 2, future research is likely to explore retrieval mechanisms specifically designed around SAP's structured data models, rather than generic document retrieval, to improve grounding fidelity for the high-stakes application types identified in Table 1.

Progressive Autonomy Expansion Within Process Automation: Extending the RPA-and-machine-learning automation framework discussed in Section 3.3, future research is likely to explore progressive expansion of LLM autonomy within process automation, beginning with narrow, well-grounded sub-tasks and expanding only as grounding reliability, discussed in Section 8, is demonstrated.

Autonomy-Calibrated Governance Frameworks: Development of ERP-specific governance frameworks explicitly calibrated to the autonomy levels identified in Table 1, addressing the governance challenge identified in Section 8, represents an important direction distinct from applying uniform governance across all LLM application types.

Resilient Distributed Infrastructure for LLM-Enabled ERP: Continued research into resilient distributed service discovery and intelligent resource scheduling, building on the infrastructure capabilities discussed in Sections 5 and 6, is likely to remain a necessary enabling direction as LLM-enabled ERP applications scale from pilot to

production deployment.

Extension of LLM-Based Explanation to Broader Analytics: The fraud-detection explanation pattern discussed in Section 4 is likely to extend to other specialized analytics capabilities across the SAP landscape, applying the content generation and embedded analytics augmentation application types discussed in Sections 3.2 and 3.4 more broadly across enterprise risk and operational analytics.

Conclusion

This paper has presented a comprehensive review of large language model applications within ERP systems, organized around a four-part taxonomy spanning natural language query and reporting, content and document generation, process automation augmentation, and embedded analytics augmentation. The review grounded this taxonomy in recent SAP ECC research on RPA-and-machine-learning process automation and Fiori-based embedded analytics, examined behavioral analytics for fraud detection as an adjacent, LLM-relevant explanation capability, and reviewed distributed service discovery and intelligent cloud resource scheduling as infrastructural preconditions for reliable, cost-effective LLM deployment at enterprise scale. A comparative analysis across the four application types showed a consistent relationship between grounding requirement, autonomy level, and infrastructural demand, with process automation augmentation carrying the highest combined demands and natural language query and embedded analytics augmentation representing comparatively lower-risk entry points for adoption. While substantial challenges remain in grounding, inference cost, governance, integration with existing business logic, and infrastructure reliability, the literature reviewed in this paper indicates that LLMs are most effectively positioned within ERP systems as a complementary natural language interaction and explanation layer built atop existing structured business logic and specialized analytics capability, rather than as a wholesale replacement for either.

References

1. Sannidhanam, A.H., Alladi, S. (2017). Behavioral Analytics for Fraud Detection in Digital Payment Systems. International Journal of Engineering and Science Research, 8(3), 01-10. https://ijesr.org/index.php/ijesr/article/view/1849
2. Sannidhanam, A. H., & Alladi, S. (2017). Service discovery and communication in distributed enterprise applications. International Journal of Engineering and Science Research, 7(4), 1–12.
3. Sannidhanam, A. H., & Alladi, S. (2019). An intelligent resource scheduling framework for cloud computing platforms. International Journal of Information Technology and Computer Engineering, 7(3), 114–127. https://ijitce.org/index.php/ijitce/article/view/1909
4. Asuri, R. (2020). An Intelligent Enterprise Workflow Automation Framework for SAP ECC Standard Business Processes through the Integration of Robotic Process Automation and Machine Learning Techniques. SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology, 12(02), 181-195. https://doi.org/10.18090/samriddhi.v12i02.17
5. Asuri, R. (2019). Transforming Legacy Enterprise Resource Planning Systems through SAP Fiori Expansion and Embedded Analytics to Enhance User Experience and Operational Performance in SAP ECC. SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology, 11(02), 176-190. https://doi.org/10.18090/samriddhi.v11i02.13
6. A. Vaswani et al., "Attention Is All You Need," in Advances in Neural Information Processing Systems (NeurIPS 2017), vol. 30, pp. 5998-6008, 2017.
7. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL-HLT 2019, Minneapolis, MN, USA, 2019, pp. 4171-4186.
8. T. B. Brown et al., "Language Models are Few-Shot Learners," in Advances in Neural Information Processing Systems (NeurIPS 2020), vol. 33, pp. 1877-1901, 2020.
9. P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in Advances in Neural Information Processing Systems (NeurIPS 2020), vol. 33, pp. 9459-9474, 2020.
10. T. H. Davenport and R. Ronanki, "Artificial Intelligence for the Real World," Harvard Business Review, vol. 96, pp. 108-116, 2018.

HOW TO CITE THIS ARTICLE: Srivastava, A. K. (2025). Large Language Models in Enterprise Resource Planning: A Comprehensive Review of Applications, Challenges, and Future Directions. *Journal of Integrated Science, Technology and Management*, 1(1), 76-80.