Journal Homepage: -www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI: 10.21474/IJAR01/22999
DOI URL: http://dx.doi.org/10.21474/IJAR01/22999

ISSN (O): 2320-5407
ISSN (P): 3107-4928

RESEARCH ARTICLE

LARGE LANGUAGE MODELS AND THE TRANSFORMATION OF PROFESSIONAL ACCOUNTING PRACTICE: FROM DATA PLUMBING TO STRATEGIC INTERPRETATION

Francis Melaragni

1. DHSc, MBA.

Manuscript Info

Manuscript History
Received: December, 2025
Final Accepted: March, 2026
Published: March, 2026

Key words:-
large language models, artificial intelligence, accounting automation, retrieval-augmented generation, financial document analysis, professional judgment

Abstract

Large language models (LLMs) are reshaping the professional landscape of accounting and finance, shifting practitioners from labor-intensive data preparation toward higher-order analytical reasoning. This paper examines how LLMs can be strategically deployed across core accounting workflows — from financial document analysis and regulatory compliance to audit risk assessment — while identifying the governance structures necessary to ensure professional integrity. Drawing on recent empirical benchmarks, industry case studies, and the emerging retrieval-augmented generation (RAG) architecture, we argue that the critical variable is not whether LLMs are used, but whether they are deployed within a framework that preserves professional judgment, data privacy, and regulatory fidelity. Implications for accounting educators, practitioners, and firms undertaking phased AI adoption are discussed.

Introduction:-

The accounting profession has historically absorbed technological disruption with resilience — the ledger yielded to the spreadsheet, the spreadsheet to enterprise resource planning systems, and each transition reconfigured the work without eliminating the professional. The emergence of large language models (LLMs) represents the next such inflection, though its scope may be broader than any previous shift. Unlike prior tools, which primarily accelerated computation, LLMs process language, extract meaning, and generate interpretive commentary across vast unstructured datasets with a speed and coverage that no analyst team could replicate manually.The analogy that best captures the practical dynamic is one of augmentation rather than automation. LLMs function as highly capable but inexperienced assistants — fast, tireless, and fluent — yet lacking the professional judgment, regulatory grounding, and ethical accountability that define competent practice. The accountant who understands this relationship will leverage these tools to amplify their reach and depth; the one who misunderstands it will either underuse the technology or over-trust it at their peril.This paper proceeds as follows: Section 2 surveys the evolving capabilities of LLMs in accounting contexts; Section 3 maps those capabilities to specific functional workflows; Section 4 examines the risk architecture that responsible deployment requires; and Section 5 proposes a phased implementation framework with implications for practitioners and accounting educators.

Corresponding Author:-Francis Melaragni
Adress:-DHSc, MBA.

LLM Capabilities in the Accounting Context:-

The academic literature on AI in accounting has expanded markedly since 2022. A recent systematic review synthesizing 256 peer-reviewed publications across the period 2015–2025 found that the financial analysis cluster — encompassing AI-driven document analysis, forecasting, and anomaly detection — contained the highest volume of empirical work, with publication counts accelerating sharply in 2023 and 2024 (Caballero-Morales et al., 2025). A parallel bibliometric framework proposed by Baber et al. (2025) classified AI-accounting research along two dimensions — accounting-centric versus AI-centric focus, and AI-based versus traditional methodology — identifying significant gaps in research on LLM-specific deployments within auditing and regulatory compliance.

What distinguishes current-generation LLMs from earlier natural language processing tools is the combination of long-context reasoning capacity and generative flexibility. A model such as Claude, with its extended context window, can ingest multi-hundred-page documents — a 10-K filing, a complex vendor contract, or a multi-period audit workpaper — while preserving inferential continuity throughout (IntuitionLabs, 2025). This capability transforms the economics of document-intensive work. Tasks that historically required teams of analysts spending days extracting and cross-referencing data can now be completed in minutes, with the model surfacing relevant clauses, flagging inconsistencies, and producing structured summaries for review.

Four capabilities stand out as particularly consequential for professional practice. First, unstructured data extraction: LLMs can ingest SEC filings, earnings call transcripts, and legal contracts, normalizing disparate formats and identifying discrepancies that would otherwise require exhaustive manual review. Second, narrative and sentiment analysis: models detect qualitative signals — shifts in managerial hedging language, changes in disclosure tone, unusual emphasis on contingent liabilities — that frequently precede measurable financial events (Islam et al., 2025). Third, natural language querying: analysts can interrogate large datasets using plain-language questions, receiving structured outputs without requiring SQL expertise or custom report configurations. Fourth, pattern recognition across longitudinal data: LLMs can scan transaction streams across multiple periods simultaneously, surfacing anomalies at a sensitivity that exceeds what a human reviewer is practically able to sustain (Rao et al., 2025).

The practical specialization of leading models matters here. Empirical evaluations suggest that general-purpose models differ meaningfully in their strengths: GPT-based systems with code execution capabilities perform well on structured data modeling tasks; Claude-class models with extended context windows are better suited to deep document reasoning; and Gemini's native integration with productivity ecosystems enables real-time synthesis of internal data with external market context (Lee, 2025). Matching the tool to the task is not merely a preference — it is a prerequisite for reliable outcomes.

Functional Applications: Where LLMs Create Value:-

Several functional areas in accounting present high-value opportunities for LLM deployment that the empirical literature has begun to document.

Financial Report Summarization and MD&A Drafting:-

The Management Discussion and Analysis (MD&A) section of annual reports represents one of the most time-consuming drafting tasks in corporate reporting. It demands synthesis of quantitative results, narrative explanation, and regulatory precision. A Deutsche Bank initiative using a Gemini-powered research assistant demonstrated that structured summarization tasks previously requiring days of analyst time could be completed in minutes, while maintaining the data privacy controls required under institutional governance standards (Training the Street, 2025). The critical discipline in this application is prompt specificity: anchoring the model to a defined reporting period and a set of themes — foreign exchange impacts, segment performance, liquidity trends — substantially reduces the risk of generalized or off-topic outputs.

Revenue Recognition and Regulatory Compliance:-

Complex revenue arrangements under ASC 606 require careful identification of performance obligations, determination of standalone selling prices, and precise timing of revenue recognition across multi-deliverable contracts. LLMs can ingest contract language and surface the relevant clauses, flag terms that implicate variable consideration or significant financing components, and generate preliminary memoranda for accountant review. A similar utility exists for ASC 842 lease accounting, where the categorization of leases across large portfolios — particularly in retail and real estate — involves substantial judgment and interpretive labor. The efficiency gains here are real, but so is the risk: models do not inherently know accounting rules and may generate treatments that are linguistically plausible but technically non-compliant with current FASB or IASB guidance (Baber et al., 2025).

Audit and Anomaly Detection:-

Perhaps the highest-stakes application of LLMs in accounting is in audit support and fraud detection. The Industrial and Commercial Bank of China's deployment of automated document processing reduced loan approval processing time by 42% and generated substantial annual cost savings (Rao et al., 2025). In audit contexts, LLMs can process large transaction populations to flag statistically unusual entries — debit-credit anomalies, round-number concentrations, vendor payment timing irregularities — providing auditors with a risk-ranked population for further investigation rather than requiring them to identify risks from scratch. Industry projections suggest that well-configured AI systems can improve anomaly detection rates substantially over manual methods, though firm-specific figures vary considerably by implementation quality and data architecture.

Risk Architecture: The Guardrails That Professional Practice Requires:-

The efficiency gains catalogued above carry commensurate risks that the accounting profession cannot afford to minimize. Four risk domains demand explicit governance attention.

Hallucination and Output Unreliability:-

The most documented and potentially most consequential risk is model hallucination — the generation of outputs that are linguistically coherent but factually fabricated. A 2024 benchmark study using the FailSafeQA evaluation framework found that LLMs produced hallucinated responses in up to 41% of finance-related queries under real-world conditions (Prabhakar, 2024). The problem is not random noise; it is plausible-sounding error. A model may fabricate a specific financial figure, construct a reference to a non-existent regulatory provision, or misrepresent the terms of a contract clause — all while presenting the output with apparent authority. One documented case involved an AI tool falsely certifying tax compliance during an acquisition review by citing a non-existent source document; the error was not discovered until post-close, resulting in a seven-figure liability (Development Corporate, 2026).

The primary technical mitigation is retrieval-augmented generation (RAG), which anchors model outputs to a verified corpus of source documents rather than relying on parametric knowledge. Under a RAG architecture, the model does not generate financial data from training memory; it retrieves and synthesizes it from the actual financial statements, audit workpapers, or regulatory texts provided. This structural constraint substantially reduces — though does not eliminate — the risk of fabricated outputs (CFA Institute, 2025).

Data Privacy and Confidentiality:-

Public-facing LLM interfaces do not provide the data confidentiality controls required for client financial information. Inputting proprietary financial data, personally identifiable information, or materially non-public data into a non-enterprise AI system creates legal and ethical exposure that cannot be remediated after the fact. The appropriate response is institutional: enterprise-grade deployments with contractual zero-retention guarantees, or anonymization protocols that prevent the transmission of client-identifiable information to any external model. This is not an area where informal workarounds are acceptable; the GLBA, GDPR, and state-level privacy frameworks create enforceable obligations that apply to AI-mediated data handling (BizTech Magazine, 2025).

Regulatory Non-Compliance:-

An LLM trained on data through a particular cutoff date may apply accounting guidance that has since been superseded, propose tax treatments that do not reflect current IRS positions, or summarize IFRS standards in ways that diverge from recent amendments. The model presents its outputs with no inherent signal that the underlying guidance is current or applicable. Every AI-generated regulatory interpretation requires verification against the authoritative source — the FASB codification, IASB standards, or applicable regulatory bulletins. The professional, not the model, remains the final arbiter.

Over-Reliance and Atrophy of Professional Judgment:-

The Journal of Accountancy (2026) has noted the risk of "shadow AI" — practitioners using unapproved tools in ways that circumvent firm governance — and the complementary risk of "AI slop," where unvetted outputs are forwarded without substantive review, creating an illusion of productivity without its substance. A related concern, particularly relevant for accounting educators, involves the potential atrophy of foundational skills among early-career professionals who lack the diagnostic competence to identify errors in AI-generated outputs. Conceptualizing LLMs as highly capable interns — requiring oversight from a senior professional who reviews, interrogates, and takes final responsibility for every work product — provides a practical frame that reinforces rather than displaces professional accountability.

Implementation Framework and Implications:-

A phased approach to LLM adoption allows firms to build institutional competence, establish governance infrastructure, and demonstrate demonstrable returns before extending AI tools to higher-stakes workflows.Phase one focuses on low-risk, high-volume tasks where errors are easily detected and the consequences of occasional inaccuracies are limited: summarizing public earnings reports, drafting internal communications and client briefing templates, and synthesizing publicly available market research. These applications build practitioner familiarity with model behavior and prompt construction without creating material professional risk.Phase two formalizes governance: a documented policy specifying which data tiers are permissible for AI input, a mandatory verification protocol for AI-generated figures, and an explicit mapping of which workflows require human review before any output is relied upon.

This phase should also address the "shadow AI" problem directly, acknowledging that staff will use whatever tools are available and that governance through blocking is neither sustainable nor productive.Phase three involves continuous professional development oriented not toward technology adoption per se but toward the cultivation of what might be called AI-critical judgment — the capacity to interrogate model outputs, identify the conditions under which outputs are most likely to be unreliable, and integrate LLM assistance into workflows that strengthen rather than substitute for professional reasoning.For accounting educators, these phases have direct pedagogical implications. Students entering the profession will work in environments where LLMs handle much of the mechanical labor that previously constituted the early-career apprenticeship. Curricula will need to cultivate both the technical fluency to use these tools productively and the professional skepticism to treat their outputs with appropriate care — the same skepticism one would apply to any delegated work product before signing off.

Conclusion:-

The transition the accounting profession is navigating is less about technology than about professional identity. The work that defines accounting's value — the exercise of judgment about complex economic events, the application of regulatory standards to specific circumstances, the acceptance of accountability for conclusions drawn — is not automatable in any meaningful sense. What is changing is the labor surrounding that judgment. LLMs absorb the friction of data collection, document processing, and first-draft generation, creating the conditions under which professional intelligence can operate at higher leverage.The accountant who understands this is not threatened by the technology; they are freed by it. The profession's competitive imperative is clear, and it is not a technical one: it is the ongoing cultivation of the interpretive judgment, ethical grounding, and professional skepticism that no language model can replicate.

References:-

1. Baber, H., Thoufiqulla, M., & Bhati, D. (2025). Artificial intelligence and accounting research: A framework and agenda. International Journal of Accounting Information Systems, 56, 100–124. https://doi.org/10.1016/j.accinf.2025.100
2. Caballero-Morales, S. O., García-Leyva, G., & Castellanos-Bolaños, M. (2025). Recent evolution and growth of AI and advanced technologies in accounting and finance: Systematic review and bibliometric analysis. Spanish Journal of Finance and Accounting. https://doi.org/10.1080/02102412.2025.2582120
3. CFA Institute. (2025). Agentic AI for finance: Workflows, tips, and case studies. https://rpc.cfainstitute.org/research/the-automation-ahead-content-series/agentic-ai-for-finance
4. Development Corporate. (2026, March). AI hallucination rates are a due diligence crisis. https://developmentcorporate.com/corporate-development/ai-hallucination-rates-are-a-due-diligence-crisis/
5. IntuitionLabs. (2025). LLMs for financial document analysis: SEC filings and decks. https://intuitionlabs.ai/articles/llm-financial-document-analysis
6. Islam, M., Hossain, M., & Cai, C. (2025). Financial document analysis and large language models: Applications and benchmarks. Journal of Financial Data Science, 7(2), 45–62.
7. Journal of Accountancy. (2026, February). AI risks CPAs should know. https://www.journalofaccountancy.com/issues/2026/feb/ai-risks-cpas-should-know/
8. Lee, D. H. (2025, October). AI model recommendations for accounting professionals. Medium. https://medium.com/@danny_54172/ai-model-recommendations-b075d0525ca2

9. Prabhakar, A. V. (2024, September). FailSafeQA: Evaluating AI hallucinations, robustness, and compliance in financial LLMs. https://ajithp.com/2025/02/15/failsafeqa-evaluating-ai-hallucinations-robustness-and-compliance-in-financial-llms/
10. Rao, W., Zhang, Y., & Li, X. (2025). A study on the influence of large language models on financial systems: Taking DeepSeek as an example. Proceedings of the 2025 International Conference on Big Data, Artificial Intelligence and Digital Economy. ACM. https://doi.org/10.1145/3767052.3767053
11. Training the Street. (2025, October). The state of AI in finance: 2025 global outlook. https://trainingthestreet.com/the-state-of-ai-in-finance-2025-global-outlook/
12. BizTech Magazine. (2025, August). LLM hallucinations: What are the implications for financial institutions? https://biztechmagazine.com/article/2025/08/llm-hallucinations-what-are-implications-financial-institutions
13. Zheng, X. L., Li, J. Y., Lu, M. Y., & et al. (2024). New paradigm for economic and financial research with generative AI: Impact and perspective. IEEE Transactions on Computational Social Systems, 1–11. https://doi.org/10.1109/TCSS.2024.3390123