Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.
ISSN 2707-2207 / ISSN 2707-2215 (en línea), julio-agosto 2026,
Volumen 10, Número 4.

https://doi.org/10.37811/cl_rcm.v10i4

RETRIEVAL-AUGMENTED GENERATION FOR COMMERCIAL PROPOSAL MANAGEMENT IN THE ELECTRICAL SECTOR: AN ENGINEERING PROJECT MANAGEMENT EVALUATION

GENERACIÓN POTENCIADA POR LA RECUPERACIÓN PARA LA GESTIÓN DE PROPUESTAS COMERCIALES EN EL SECTOR ELÉCTRICO: UNA EVALUACIÓN DE LA GESTIÓN DE PROYECTOS DE INGENIERÍA

Nefi Alejandro Barron Herrera
Posgrado CIATEQ, A.C

Elsa Marias de LA Calleja Mora
Posgrado CIATEQ, A.C

DOI: https://doi.org/10.37811/cl_rcm.v10i4.25662

Retrieval-Augmented Generation for Commercial Proposal Management in the Electrical Sector: An Engineering Project Management Evaluation

Nefi Alejandro Barron Herrera[1]
nefi.barron@gmail.com
https://orcid.org/0009-0007-2754-8462
Posgrado CIATEQ, A.C

Elsa Marias de LA Calleja Mora
elsa.delacalleja@ciateq.mx
https://orcid.org/0000-0003-0577-3785
Posgrado CIATEQ, A.C

ABSTRACT

This study evaluates the deployment of a Retrieval-Augmented Generation (RAG) LLM assistant for information management in commercial proposals for electrical generation plant retrofit projects. Using an experimental mixed-methods design framed by data-driven, risk-based, and value-driven project management lenses, the performance of the AI assistant was benchmarked against manual execution by proposal engineers. The results revealed that the RAG assistant reduced information processing lead time from a manual range of 2–30 minutes down to a predictable 1–2 minutes (a cycle-time reduction of approximately 90%), while maintaining a zero hallucination rate and generating a projected operating-cost saving of nearly 90% per cycle. In conclusion, RAG architectures provide a reliable and efficient decision-support instrument to enhance information management in pre-project engineering workflows.

Keywords: Retrieval-Augmented Generation (RAG), Commercial proposals, Electrical sector, Engineering project management, Large Language Models (LLM)

[1] Autor principal
Correspondencia: nefi.barron@gmail.com

Generación potenciada por la recuperación para la gestión de propuestas comerciales en el sector eléctrico: una evaluación de la gestión de proyectos de ingeniería

RESUMEN

Este estudio evalúa la implementación de un asistente basado en modelos de lenguaje grande con generación aumentada por recuperación (RAG, por sus siglas en inglés) para la gestión de la información en propuestas comerciales de proyectos de modernización (*retrofit*) en el sector de generación eléctrica. Mediante una metodología experimental y un marco metodológico que combina la gestión de proyectos basada en datos, en riesgos y en valor, se comparó el desempeño del sistema de IA con la ejecución manual realizada por ingenieros. Los resultados demostraron que el asistente RAG redujo el tiempo de búsqueda y procesamiento de información de un rango manual de 2 a 30 minutos a solo 1 o 2 minutos (una reducción del tiempo de ciclo cercana al 90%), logrando además una tasa nula de alucinaciones y un ahorro proyectado del 90% en costos operativos por ciclo. En conclusión, la arquitectura RAG constituye una herramienta eficaz para optimizar la eficiencia, predictibilidad y confiabilidad en la fase previa a la ejecución de proyectos de ingeniería comercial.

Palabras clave: Generación aumentada por recuperación (RAG), Propuestas comerciales, Sector eléctrico, Gestión de proyectos de ingeniería, Modelos de lenguaje grande (LLM)

Artículo recibido 20 mayo 2026
Aceptado para publicación: 20 junio 2026

1. INTRODUCTION

The electrical generation industry operates within long, capital-intensive investment cycles dominated by retrofit, modernization, and repowering projects of existing plants. In Mexico, fossil-based combined-cycle, conventional thermal, hydroelectric, and coal-fired units continue to define the asset base of the National Electric System (SENER, 2023), and operators routinely undertake upgrades whose cost ranges from tens of thousands to several million U.S. dollars and whose execution windows extend up to twenty weeks or more (EPRI, 2021). The business development teams that prepare commercial proposals for these upgrades operate under stringent technical, regulatory, and commercial constraints, requiring fluent interaction with original equipment manufacturer (OEM) datasheets, IEEE, IEC, ASME, and ANSI specifications, regulatory frameworks issued by the *Comisión Federal de Electricidad* (CFE), the *Centro Nacional de Control de Energía* (CENACE), and the *Comisión Reguladora de Energía* (CRE), and historical records of analogous projects.

Empirical evidence from the B2B sales literature indicates that sales representatives spend less than thirty percent of their time on direct selling activities and that information search and proposal preparation consume the bulk of the remaining capacity (Baldino, 2023; Data Axle, 2025; Jahic, 2026a, 2026b). In the specific case of the electrical sector, internal observations within proposal teams suggest that approximately eighty percent of the executed cycle time is absorbed by three bottlenecks: searching historical documentation, classifying it, and analyzing it for transfer to the new offer. The remaining twenty percent supports authoring, internal review, and client interaction. This imbalance directly degrades response times, conversion rates, and risk-adjusted margins. The cost of these inefficiencies is not abstract: a single retrofit proposal can mobilize between forty and two hundred engineering hours, and a proposal team operating in a competitive bidding window of four to twelve weeks must repeatedly retrieve, contrast, and recompose information drawn from disparate sources — OEM technical bulletins, prior similar bids, vendor catalogues, regulatory annexes, and internal cost templates. When this volume of unstructured knowledge is processed through manual search, the probability of omission or inconsistency rises with the cognitive load of the engineer, particularly in high-pressure rounds where multiple proposals run in parallel.

Three structural features of the electrical retrofit market amplify the value of any productivity improvement in the proposal phase. First, the buyer base is concentrated and standards-driven: state-owned utilities, independent power producers (IPPs), and regulated distribution companies dominate procurement, and their requests for quotation reference dense technical specifications anchored on IEEE, IEC, ASME, and ANSI norms, plus jurisdiction-specific permits issued by CRE and CFE. Second, proposals are repeatedly iterated through Request for Information (RFI) and Request for Quotation (RFQ) cycles, generating multiple internal versions whose traceability is critical for negotiation, scope governance, and post-award change-order management. Third, the technical content is heterogeneous: a single proposal can mix narrative scope-of-work statements, normalized cost tables, single-line diagrams, piping and instrumentation diagrams, and warranty terms — each requiring careful classification when reused in subsequent bids. Together, these features create the conditions under which an AI-augmented information layer can compound benefits across many proposals, rather than yielding a one-off saving.

Artificial intelligence (AI), and in particular large language models (LLMs) coupled with retrieval-augmented generation (RAG) architectures, has matured into a viable decision-support technology for knowledge-intensive workflows (Brown et al., 2020; Lewis et al., 2020; Russell and Norvig, 2021). RAG systems have demonstrated reductions in hallucinations and improvements in factual grounding when an LLM is conditioned on a curated corpus rather than relying on parametric memory alone (Béchard and Marquez Ayala, 2024; Zhang and Zhang, 2025). At the same time, the project management discipline has progressively incorporated data-driven, risk-based, and value-driven approaches that demand quantitative evidence of efficiency, control of residual risk, and traceability of value delivery (Aven, 2015; ISO, 2018; Project Management Institute, 2021; Riaz et al., 2022). Yet, the intersection of these two trajectories — RAG-augmented LLMs deployed as a project management instrument in the pre-project phase of electrical-sector proposals — has received limited empirical attention.

This article addresses that gap by posing the following research question: *to what extent does a RAG-augmented LLM assistant, evaluated under data-driven, risk-based, and value-driven project management lenses, improve the efficiency, consistency, and reliability of information management in*

the elaboration of commercial proposals for retrofit projects of electrical generation plants? The study contributes to the engineering business management literature in three ways. First, it develops a triangulated evaluation framework combining data-driven, risk-based, and value-driven project management as an integrated lens for assessing AI-based decision-support tools in the pre-project phase. Second, it provides empirical evidence — both qualitative (proof of concept) and quantitative (controlled comparison) — of the operational and economic impact of a RAG-augmented LLM in a real commercial-proposal workflow. Third, it formalizes two reference prompts for supplier-quotation comparison and proposal-version traceability that practitioners and researchers can replicate, extend, and benchmark.

The remainder of the paper is organized as follows. Section 2 synthesizes the relevant literature on AI in project management, B2B proposal processes, RAG architectures, and hallucination mitigation. Section 3 details the methodology, including the technical stack, the corpus, the experimental design, and the project management lenses. Section 4 reports the proof of concept, the comparative results, and their economic projection. Section 5 discusses the findings against prior literature and managerial implications. Section 6 concludes with limitations and directions for future research.

2. LITERATURE REVIEW

2.1. Artificial Intelligence in Project and Knowledge Management

The conceptual foundations of computational intelligence trace back to the universal Turing machine (Turing, 1936), whose theoretical framing of decidability and algorithmic computation underpins all later AI architectures (Copeland, 2004; Hodges, 2019). Modern AI is understood as a multi-level capability hierarchy that ranges from narrow automation to artificial general intelligence (Boden, 2016; Ringel Morris et al., 2024); within this hierarchy, machine learning and deep learning provide the technical substrate for present-day systems (Goodfellow et al., 2016; LeCun et al., 2015).

Project management research has progressively integrated AI, big data analytics, and automated decision-making into the planning, monitoring, and control of project workflows. Davenport and Harris (2005) anticipated the rise of automated decision-making as a managerial competency. More recently, Obaid and Abu-Naser (2023) and Riaz, Sumayya, and Philbin (2022) argue that big data analytics is now a key determinant of project success, particularly when paired with structured project

management bodies of knowledge such as PMBOK 7 (Project Management Institute, 2021). The wider socio-economic literature has examined the labor-market consequences of AI-induced automation (Acemoglu and Restrepo, 2020; Brynjolfsson and McAfee, 2016; Frey and Osborne, 2017; Goos, Manning, and Salomons, 2014), reinforcing the case for managerial frameworks that explicitly account for value creation rather than mere cost displacement.

2.2. The Pre-Project Phase of B2B Engineering Sales

The B2B sales literature consistently identifies prospecting, qualification, and proposal elaboration as the three early stages with the greatest information intensity and the longest cycle times (Barnes, 2025; Natoli, 2025; Rius, 2026; salesken.ai, 2025; The Dock Team, 2026). Recent empirical work indicates that long sales cycles, fragmented information sources, and multi-stakeholder buying committees characterize industrial B2B markets, with sales teams spending less than a third of their time on direct selling activities (Baldino, 2023; Data Axle, 2025; Jahic, 2026a, 2026b). AI-based lead generation and lead-scoring approaches have shown promising results in increasing conversion rates (González-Flores, Rubiano-Moreno, and Sosa-Gómez, 2025; Kaplan, Seker, and Yoruk, 2025), and conversational AI is reshaping consumer decision-making in adjacent contexts (Lopez-Lopez and Bara Iniesta, 2025). However, the specialized literature on AI-augmented proposal generation in heavy-industry settings — particularly under regulatory frameworks such as those of the Mexican electrical sector — remains thin.

2.3. Text Mining, Knowledge Discovery, and Information Extraction

Text mining (TM), knowledge discovery (KD), and knowledge management (KM) provide the analytical scaffolding for transforming heterogeneous proposal corpora into structured, reusable knowledge (Aggarwal, 2018; Fayyad et al., 1996; Grobelnik and Mladenic, 2005). Information extraction (IE) and natural language processing (NLP) operationalize this scaffolding, enabling the recovery of named entities, numerical fields, and relations from unstructured text (Singh, 2018). Recent advances in deep-learning-based NLP allow full-text data extraction from technical literature with task-specific pipelines (Du et al., 2025; Hussain, Usman Akram, and Abdul Salam, 2025), with applications ranging from materials science (Dagdelen et al., 2024; Huang and Cole, 2020) to systematic literature reviews. These approaches are directly transferable to commercial-proposal

corpora, where buyer, scope, price, lead time, and warranty fields recur across documents but are seldom uniformly tagged.

2.4. Retrieval-Augmented Generation and Hallucination Mitigation

Retrieval-augmented generation (RAG), formalized by Lewis et al. (2020), couples a generative LLM with a document retriever, so that responses are grounded on a verifiable corpus rather than the model's parametric memory alone. RAG architectures typically comprise a chunking and indexing stage, a retrieval stage based on dense or sparse representations, and a generation stage that consumes retrieved evidence as conditioning context. Recent work has refined evaluation metrics, hybrid pipelines, and multi-evidence designs that further reduce hallucinations (Ko, Gürkan, and Yarman Vural, 2024; Song et al., 2024; Sun et al., 2025; Wołk, 2025; Xu et al., 2025; Zhang and Zhang, 2025). Hallucinations — outputs that are fluent but factually unsupported — are the primary residual risk of LLM-based applications in technical contexts (Massenon et al., 2025; Yang et al., 2025). Detection methods range from internal-state probes (Chen et al., 2024; Sriramanan et al., 2024) and token-probability analyses (Quevedo et al., 2024) to black-box self-checks and semantic-entropy approaches (Ji et al., 2024; Liu et al., 2025; Manakul, Liusie, and Gales, 2023; Wei et al., 2025). Industrial guidance has converged on hybrid pipelines that combine grounded retrieval, structured-output constraints, hallucination-aware tuning, and downstream review (Afolabi et al., 2025; Béchard and Marquez Ayala, 2024). Despite this progress, deployments in the engineering proposal domain remain rare.

2.5. Project Management Lenses for AI Adoption

Three project management approaches frame the evaluation of AI-based information tools. Data-driven project management (DDPM) emphasizes evidence-based decisions through quantitative key performance indicators (KPIs) such as lead time and forecast accuracy (Davenport and Harris, 2005; Riaz, Sumayya, and Philbin, 2022). The risk-based approach (RBA), aligned with ISO 31000:2018, prioritizes the systematic identification, mitigation, and tracking of residual risk (Aven, 2015; Glette-Iversen, Flage, and Aven, 2023; ISO, 2018; Parviainen et al., 2021). Value-driven project management (VDPM) extends success criteria beyond scope, time, and cost to encompass measurable benefits realization, target benefits, and benefits maps (Chih and Zwikael, 2014; Martins Serra and Kunc,

2015; Project Management Institute, 2021; Shenhar et al., 2001). Triangulating these lenses provides a defensible evaluation logic for deploying AI-augmented decision-support tools, since improvements in efficiency must be balanced against residual risks and aligned with strategic value.

2.6. Research Gap

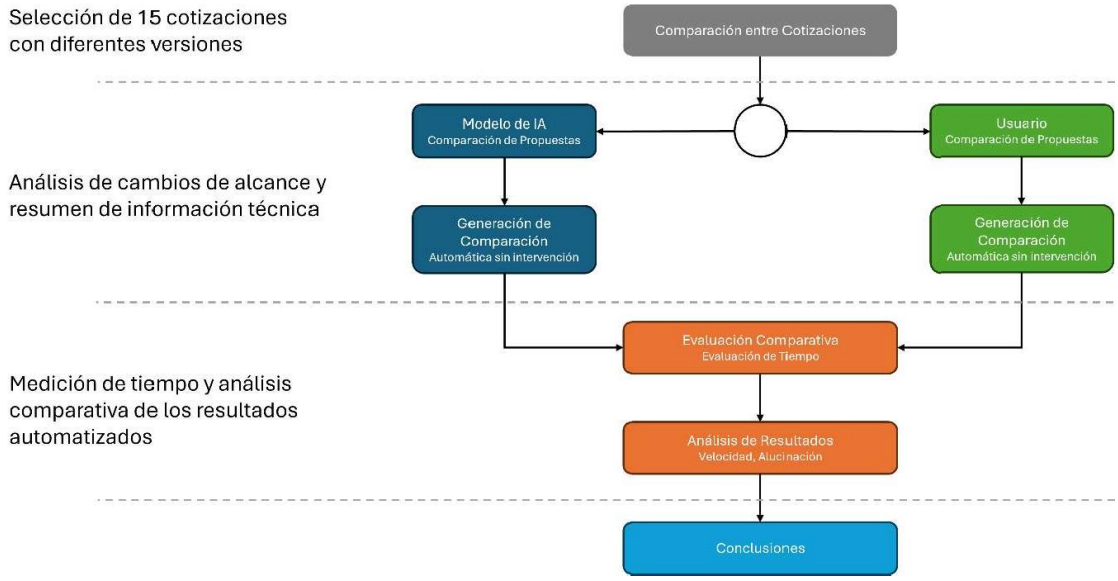
The literature offers substantial evidence on RAG architectures, hallucination mitigation, AI-assisted B2B sales, and project management frameworks, but few studies (i) deploy a RAG-augmented LLM end-to-end in the elaboration of engineering commercial proposals, (ii) measure its effect on cycle time and reliability under controlled conditions, and (iii) interpret the findings simultaneously through DDPM, RBA, and VDPM. The closest analogues are domain-specific RAG deployments in clinical decision support (Wołk, 2025) and public health (Xu et al., 2025), and conversational-AI studies in consumer decision-making (Lopez-Lopez and Bara Iniesta, 2025). None of these settings, however, faces the exact combination of regulated technical specifications, multi-version negotiation, and risk-adjusted margin pressure that characterizes the electrical-sector proposal cycle. This study addresses these three gaps in a unified empirical setting and proposes a transferable evaluation template for engineering business management research on AI adoption.

3. METHODOLOGY

3.1. Research Design

The study adopts a mixed-methods, applied, and experimental design (Figure 1). Qualitative analysis identifies the factors that drive cycle time in commercial-proposal preparation, while quantitative analysis measures the effect of an AI-augmented information workflow on those factors. The unit of analysis is a single information-search task drawn from the pre-project phase of an electrical-sector commercial proposal — specifically (a) supplier-quotation comparison and (b) traceability of successive proposal versions. The treatment is the use of a RAG-based assistant; the control is the same task executed manually by a qualified user.

Figure 1. Methodological schema for the comparative evaluation of AI- and user-generated information processing in commercial proposals. Adapted from the source thesis.



3.2. Technical Stack and System Architecture

The experimental environment was implemented as a conversational assistant built on four components organized in three functional layers. The presentation layer relies on React 19 and Material-UI v7 (MUI), which guarantee a stable, reproducible interface and a homogeneous interaction surface across trials. The typing layer uses TypeScript to enforce static contracts between the user interface, the application state, and the inference service, reducing run-time errors and documenting the data shapes that flow between modules. The reasoning layer is anchored on Claude Opus 4.6, accessed via an authenticated application programming interface (API), and is responsible for prompt comprehension, structured information extraction, and response generation.

The retrieval layer underlying the LLM follows a standard RAG architecture in three steps. First, ingestion: each historical proposal in the corpus is parsed into its component blocks (header, scope, prices, warranties, exclusions), normalized, and chunked into overlapping segments sized to the model's context window. Second, indexing: each chunk is encoded into an embedding — a dense numerical representation of meaning — and stored in a vector index alongside the original metadata. Third, query-time retrieval and generation: an incoming user prompt is embedded with the same encoder, the top-k most similar chunks are returned by similarity search, and the LLM receives them as conditioning evidence together with the prompt. The model is instructed to ground each answer on retrieved fragments and to cite them inline, an established design pattern for reducing hallucinations in

technical domains (Béchard and Marquez Ayala, 2024; Lewis et al., 2020). This separation of concerns isolates the model's behavior from interface variability, supports reproducibility under the data-driven evaluation logic, and allows the pipeline to be benchmarked independently from the user-facing layer.

3.3. Corpus and User Profile

The reference corpus comprises fifty historical commercial-proposal records targeting seven buyers in the electrical generation sector. Each record contains a proposal identifier, customer code, upgrade type, technical description, price (in U.S. dollars), reported benefits, and buyer responsibilities. Eight upgrade categories appear in the corpus: power-output increase, cooling-system upgrade, exhaust-system improvement, remote monitoring system, noise-reduction kit, control-panel upgrade, fuel-efficiency enhancement, and automatic transfer switch. Reported prices range from approximately USD 6,000 to USD 50,000.

Three engineer-users participated in the comparative experiment. Each user holds a graduate or undergraduate degree in electrical, mechanical, mechatronic, electromechanical, or control engineering, with a minimum of four years of experience in retrofit projects of combined-cycle, gas-turbine, or steam-turbine plants. Their working languages are Spanish and English, and they routinely operate under ISO 9001 and ISO 55000 traceability requirements. The profile mirrors that of a proposal engineer, application engineer, or technical-commercial lead in a pre-sales team. The three early B2B stages — prospecting, qualification, and proposal elaboration — bracket the experiment.

3.4. Reference Prompts

Two reference prompts formalize the assistant's behavior under the *role–context–inputs–task–constraints–output* structure recommended in the prompt-engineering literature. The first prompt supports supplier-quotation comparison and produces four artifacts: a normalized comparative table, a 250-word executive note, a supplier ranking with strengths and weaknesses, and a flag list of risk signals classified as high, medium, or low. The second prompt addresses traceability of successive versions of an outgoing proposal and produces a comparative version table, a 300-word evolution note, a ranking of versions against internal success criteria (margin, residual risk, lead time, technical compliance), and a list of traceability risk flags. Both prompts incorporate explicit anti-hallucination

constraints: the model is instructed not to invent fields, to mark inferences distinctly, and to cite each datum to its source document.

3.5. Experimental Protocol

The experiment proceeded in two stages. The first stage was a qualitative proof of concept that verified end-to-end feasibility on the corpus through two extraction prompts: (i) “Make a list of customers and how many proposals each has”, and (ii) “Identify which are the most common upgrades and benefits”. The proof of concept tested whether the RAG pipeline could correctly traverse the corpus, recognize repeated entities across heterogeneous documents, and emit a structured response without external scaffolding.

The second stage was a quantitative comparative test in which each of the three users executed fifteen independent search tasks under each of two scenarios — supplier-quotation comparison and proposal-version comparison — first manually and then with the RAG-based assistant. The fifteen tasks per user per scenario were drawn from a pool designed to vary along three dimensions: information density (number of input documents per task), entity complexity (heterogeneity of fields to extract), and temporal scope (single-round vs multi-round comparisons). Tasks were administered in randomized order to mitigate learning effects.

For every trial, the recorded variables were execution time (in minutes, measured from prompt issuance to delivery of a complete answer), presence or absence of hallucinations in the AI output (validated against the source corpus), and successful completion. Hallucinations were defined operationally as fluent assertions in the AI response that could not be traced to any retrieved source fragment, following the typology used in recent industrial RAG evaluations (Massenon et al., 2025; Yang et al., 2025). They were assessed by an expert panel of three reviewers, each with more than twenty years of experience in project engineering, commercial management, and electrical generation; the panel applied a binary “detected/not detected” judgement at the trial level and a structured fact-by-fact check on each AI response. Internal validity was supported by (i) a stable model and prompt template across trials, (ii) the same corpus and retrieval index for all users, and (iii) timekeeping by a non-participant observer. External validity is bounded by the size of the user panel and the corpus, both of which are addressed in Section 6.

3.6. Evaluation Lenses and KPIs

Three KPIs operationalize the evaluation. Under DDPM, lead time is measured per task; the target threshold is a reduction of at least twenty-five percent against the manual baseline, complemented by a forecast-accuracy proxy expressed as the variance of AI-task durations. Under RBA, residual risk is operationalized as the proportion of trials with detectable hallucinations; the target threshold is below fifteen percent. Under VDPM, the target benefits are an increase in the win rate of proposals and an improvement in the risk-adjusted margin, both proxied by the freed engineering capacity that AI assistance redirects to higher-value tasks.

4. RESULTS

4.1. Comparative Time and Reliability Results

The comparative experiment produced thirty paired observations distributed across the supplier-quotation and proposal-version scenarios. Table 3 reports the supplier-quotation results; Table 4 reports the proposal-version results. Across both scenarios, AI-assisted tasks completed in a stable one-to-two-minute range (mean ≈ 1.0 min; near-zero variance), while manual execution times varied widely, from two to thirty minutes. A single AI execution returned an error in supplier-quotation Trial 3, which was flagged but did not propagate further; no hallucinations were detected by the expert panel in any of the thirty trials.

Table 3. Execution times for supplier-quotation comparison tasks (Scenario 1).

User	Trial	User time (min)	AI time (min)	Hallucination detected
User 1	1	25	1	No
User 2	2	15	2	No
User 3	3	3	Error	No
User 1	4	10	1	No
User 2	5	7	1	No
User 3	6	2	1	No
User 1	7	3	1	No
User 2	8	8	1	No
User 3	9	5	1	No

User 1	10	6	1	No
User 2	11	20	1	No
User 3	12	16	1	No
User 1	13	30	1	No
User 2	14	9	1	No
User 3	15	8	1	No

Table 4. Execution times for proposal-version comparison tasks (Scenario 2).

User	Trial	User time (min)	AI time (min)	Hallucination detected
User 1	1	10	1	No
User 2	2	12	2	No
User 3	3	5	1	No
User 1	4	7	1	No
User 2	5	20	2	No
User 3	6	3	1	No
User 1	7	8	1	No
User 2	8	5	1	No
User 3	9	15	1	No
User 1	10	10	1	No
User 2	11	7	1	No
User 3	12	3	1	No
User 1	13	4	1	No
User 2	14	13	1	No
User 3	15	16	1	No

Figures 3 and 4 visualize the per-trial comparison for each scenario. The two charts confirm the consistency of AI-assisted execution and the heterogeneity of human execution. The largest absolute gaps occur in trials with high information density (Trials 1, 11, and 13 of Scenario 1), where the

manual approach exceeds twenty minutes while the AI-assisted approach remains at one minute. Across both scenarios, the manual mean approaches eleven minutes per task with a standard deviation greater than seven minutes, whereas the AI-assisted mean stays close to one minute with negligible dispersion. Read together, the descriptive statistics characterize two qualitatively different process regimes: a high-variance, individually-driven regime under manual execution and a low-variance, near-deterministic regime under AI assistance. From a planning standpoint, this consistency is operationally meaningful — it makes capacity forecasts at the proposal-team level much tighter and reduces the buffer time that managers must allocate to absorb individual variability.

Figure 3. Per-trial execution times for the supplier-quotation comparison scenario; AI assistance compared with manual execution.

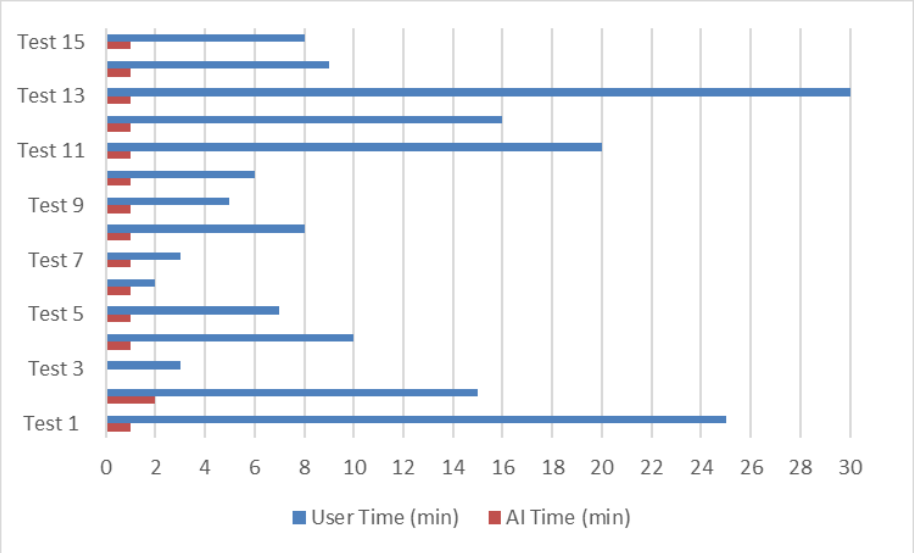
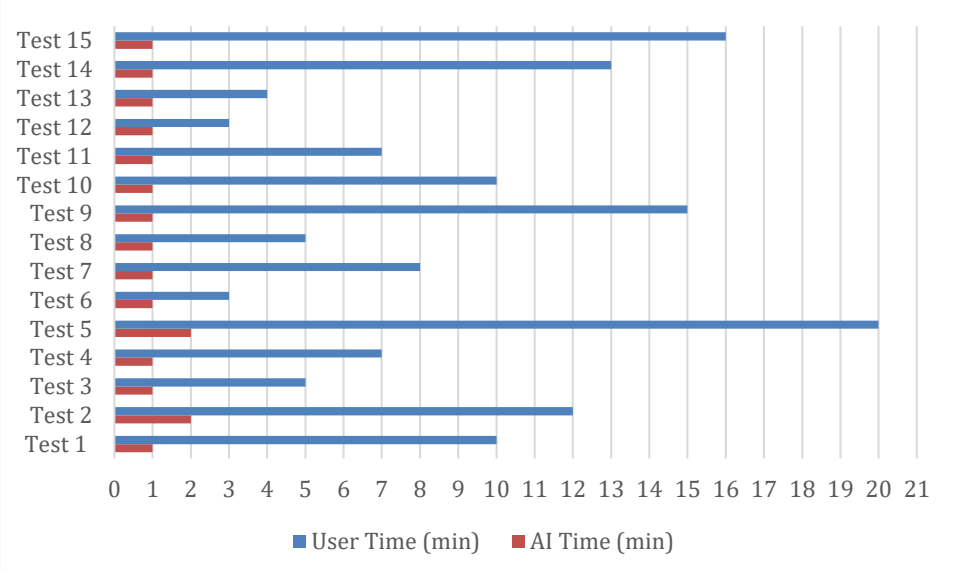


Figure 4. Per-trial execution times for the proposal-version comparison scenario; AI assistance compared with manual execution.



4.2. Economic Projection

A first-order economic projection translates the time savings into operating-cost terms. Assuming a reference rate of USD 150 per engineering hour — representative of the loaded cost of a senior proposal engineer in retrofit projects — and an average manual execution time of five hours per cycle (combining historical-proposal search, quotation comparison, and version verification), the manual cost is approximately USD 750 per cycle (≈ MXN 13,500 at the exchange rate observed during the test period). Under AI assistance, the total cycle time drops to approximately thirty minutes (including human review), yielding a projected cost of USD 75 per cycle (≈ MXN 1,350) — roughly ten percent of the manual cost and consistent with the ninety percent lead-time reduction observed in the comparative trials. These figures are indicative; full validation requires instrumentation across larger volumes of proposals and inclusion of additional costs such as inference subscription, corpus maintenance, and user training.

5. DISCUSSION

5.1. Triangulated Interpretation

Under the data-driven lens, the observed reduction of cycle time from a two-to-thirty-minute range to a one-to-two-minute range corresponds to a lead-time reduction of approximately ninety percent, comfortably exceeding the −25% target threshold defined ex ante for the pre-project phase. The near-zero variance of AI-task durations supports a favorable forecast-accuracy reading: AI execution is not only faster but more predictable, which improves capacity planning for the pre-sales team. This finding is consistent with prior big-data analytics studies in project management (Obaid and Abu-Naser, 2023; Riaz, Sumayya, and Philbin, 2022) and aligns with reported productivity gains in adjacent B2B AI applications (Kaplan, Seker, and Yoruk, 2025; González-Flores, Rubiano-Moreno, and Sosa-Gómez, 2025).

Under the risk-based lens, the absence of hallucinations across thirty trials suggests a low residual information risk under the experimental conditions, in line with the core promise of grounded RAG architectures (Lewis et al., 2020; Zhang and Zhang, 2025). The result is also consistent with industrial guidance favoring grounded retrieval and structured output (Afolabi et al., 2025; Béchard and Marquez Ayala, 2024). However, the thirty-trial sample is insufficient to claim that the residual-risk

KPI of less than fifteen percent generalizes; broader corpora and adversarial prompts would be required for a strong claim.

Under the value-driven lens, the freed engineering capacity per proposal cycle — from approximately five hours of manual search to about thirty minutes of supervised AI-assisted execution — can be redirected to higher-value activities such as risk-of-retrofit assessment, OEM negotiation, and strategic scope decisions. This redirection is the operational mechanism through which the target benefit (an increase in the win rate and an improvement in the risk-adjusted margin) is plausibly realized. Yet, as observed in the benefits-realization literature (Chih and Zwikael, 2014; Martins Serra and Kunc, 2015; Shenhar et al., 2001), conclusive evidence on win rate and margin requires longitudinal measurement across complete proposal cycles — a path beyond the present experimental scope.

5.2. Comparison with Prior Work

The findings extend prior RAG and LLM evaluations, typically conducted on academic or general-purpose corpora (Lewis et al., 2020; Wołk, 2025; Xu et al., 2025), to the specialized engineering-proposal domain, where domain knowledge and traceability requirements are higher. The reported reduction in cycle time is at the upper bound of efficiency gains documented in B2B sales productivity studies (Baldino, 2023; Data Axle, 2025), suggesting that the heavy reliance on document search in proposal preparation amplifies the marginal value of RAG. The absence of hallucinations contrasts favorably with the persistent risk highlighted in user-reported field studies of LLM applications (Massenon et al., 2025), reinforcing the case for RAG-grounded designs over pure generative deployments. Compared with structured-output reduction studies (Béchard and Marquez Ayala, 2024) and hybrid retrieval architectures (Ko, Gürkan, and Yarman Vural, 2024; Song et al., 2024), the present setup achieves comparable factual reliability with a simpler implementation footprint, indicating that for narrow corpora and well-scoped tasks, additional architectural complexity may yield diminishing returns. The triangulated framework introduced here — DDPM × RBA × VDPM — complements prior single-lens evaluations and provides a transferable template for engineering business management research on AI adoption. By embedding the technical evaluation inside an explicit project management logic, the framework also offers a vocabulary that engineering and

managerial audiences can share, which is itself a contribution to bridging the gap between technical and managerial AI research.

5.3. Managerial and Operational Implications

For engineering services firms, the results indicate a tractable path toward measurable productivity and quality gains in the pre-project phase. First, the deployment investment is bounded: a RAG architecture built on commercially available LLM APIs and standard front-end components can be operationalized without custom model training, as long as a curated proposal corpus exists. The principal operational pre-requisite is therefore the curation of that corpus — a task that is well-aligned with knowledge management practices (Aggarwal, 2018; Grobelnik and Mladenic, 2005) and that yields collateral benefits, such as standardization of templates, retirement of obsolete clauses, and explicit identification of the boundary between non-confidential and proprietary content.

Second, the quantitative KPIs (lead time, forecast accuracy, residual hallucination rate, freed engineering hours) are auditable and align with project management best practices (Project Management Institute, 2021), enabling defensible business cases. Engineering managers can stage deployment as a benefits map: phase one targets lead-time reduction in the most repetitive tasks (quotation comparison, version traceability), phase two extends scope to scope-of-work drafting and risk-register generation, and phase three integrates the assistant with downstream proposal-management systems and customer relationship management (CRM) platforms. Each phase is associated with concrete KPIs and target benefits, providing a defensible accountability structure under VDPM.

Third, the reference prompts formalize quality controls — explicit citations to source documents, distinction between observed data and inferences, and structured output formats — that mitigate the operational risks of LLM use under regulated conditions such as ISO 9001 and ISO 55000. Beyond the prompts themselves, organizations should institutionalize three further controls: (i) periodic re-indexing of the corpus to incorporate new awarded proposals while removing superseded ones; (ii) human-in-the-loop review of every AI-generated artifact before client release, with the proposal engineer accepting or rejecting each cited fragment; and (iii) periodic adversarial evaluation, where the panel intentionally challenges the assistant with edge cases (missing fields, conflicting data points,

ambiguous requirements) to surface latent failure modes. These controls operationalize the residual-risk discipline central to RBA (Aven, 2015; ISO, 2018; Parviainen et al., 2021).

For policy and standardization audiences, the study underscores the relevance of regulatory frameworks for data protection (CDHCU, 2025), neutrality (IFT, 2021), and international AI governance (G20, 2024) when AI-augmented assistants are deployed in industrial settings, since proposal corpora typically contain commercially sensitive information. Practical compliance pathways include on-premise or private-cloud deployment of the inference layer, tokenization or redaction of personally identifiable information before indexing, and explicit retention policies linked to contract lifecycles. Where such guarantees cannot be met, the scope of automation should be restricted to non-confidential portions of the corpus until the appropriate safeguards are in place.

6. Conclusions, Limitations, and Future Research

This article reports the design, implementation, and empirical evaluation of a RAG-augmented LLM assistant for managing information in commercial proposals for retrofit projects of electrical generation plants. The proof of concept confirmed end-to-end technical feasibility on a fifty-record proposal corpus, and the comparative experiment showed a stable one-to-two-minute AI execution time against a two-to-thirty-minute manual range, with no hallucinations detected and a projected operating-cost reduction near ninety percent per cycle. Interpreted through the triangulated lens of data-driven, risk-based, and value-driven project management, the assistant performed beyond the lead-time threshold of −25%, met the residual-risk target under experimental conditions, and credibly enabled a redirection of engineering capacity toward higher-value tasks consistent with the target benefits.

Three limitations qualify these conclusions. First, the sample of three engineer-users and thirty trials is suitable for first-stage validation but insufficient for population-level inference; a larger panel and a longer measurement window are needed to confirm the residual-risk KPI and to detect rare-event hallucinations. Second, the study evaluates information-search tasks of moderate inferential demand; tasks that require interpretation of single-line diagrams, piping and instrumentation diagrams, or photographic evidence remain beyond the scope of pure text-RAG and require multimodal extensions.

Third, the economic projection is indicative; longitudinal tracking across full proposal cycles is required to confirm impacts on win rate and risk-adjusted margin.

Three research directions follow. First, a multi-firm replication with anonymized corpora would test external validity across electrical-sector market segments and regulatory regimes; comparative cross-country studies (for example, Mexican CRE-regulated projects versus North American IPP procurement) would shed light on how regulatory heterogeneity shapes the marginal value of an AI-augmented proposal layer. Second, a multimodal extension that incorporates diagrammatic and tabular evidence — building on advances in domain-specific extraction (Dagdelen et al., 2024; Du et al., 2025; Hussain, Usman Akram, and Abdul Salam, 2025) — would broaden applicability to the full engineering proposal pipeline, addressing the current limitation around single-line-diagram and P&ID interpretation. Third, integration of richer hallucination-detection signals (Chen et al., 2024; Liu et al., 2025; Manakul, Liusie, and Gales, 2023; Sun et al., 2025) into the deployment loop would tighten the residual-risk control and provide automated early-warning indicators that reduce reliance on full panel review. A complementary line of work would examine the organizational dynamics of adoption — change management, role redefinition, and skill profiles — through which the productivity gains observed in this controlled setting translate into sustained business performance. Together, these directions consolidate RAG-augmented LLMs as a measurable, defensible, and value-aligned instrument of engineering project management in the pre-project phase.

REFERENCES

Acemoglu, D. and Restrepo, P. (2020) 'Robots and jobs: evidence from US labor markets', *Journal of Political Economy*, 128(6), pp. 2188–2244. https://doi.org/10.1086/705716.

Afolabi, Z., Taleb, A., Kozodoi, N. and Zinovyeva, E. (2025) *Detect hallucinations for RAG-based systems*. Available at: https://aws.amazon.com/blogs/machine-learning/detect-hallucinations-for-rag-based-systems/ (Accessed: 16 May 2025).

Aggarwal, C.C. (2018) *Machine learning for text*. New York: Springer. https://doi.org/10.1007/978-3-319-73531-3.

Aven, T. (2015) ‘Risk assessment and risk management: review of recent advances on their foundation’, *European Journal of Operational Research*, 253(1), pp. 1–13. https://doi.org/10.1016/j.ejor.2015.12.023.

Baldino, A. (2023) *B2B sales statistics*. Available at: https://www.thinkific.com/blog/b2b-sales-statistics/ (Accessed: 6 May 2026).

Barnes, O. (2025) *B2B long sales cycles*. Equinet Media. Available at: https://www.equinetmedia.com/blog/b2b-long-sales-cycles (Accessed: 25 March 2025).

Béchard, P. and Marquez Ayala, O. (2024) ‘Reducing hallucination in structured outputs via retrieval-augmented generation’, in *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. https://doi.org/10.18653/v1/2024.naacl-industry.19.

Boden, M.A. (2016) *AI: its nature and future*. Oxford: Oxford University Press.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P. *et al.* (2020) ‘Language models are few-shot learners’, *Advances in Neural Information Processing Systems*, 33, pp. 1877–1901. https://doi.org/10.48550/arXiv.2005.14165.

Brynjolfsson, E. and McAfee, A. (2016) *The second machine age: work, progress, and prosperity in a time of brilliant technologies*. New York: W.W. Norton & Company.

Cámara de Diputados del H. Congreso de la Unión (CDHCU) (2025) *Ley Federal de Protección de Datos Personales en Posesión de los Particulares*. Mexico City: CDHCU.

Chen, C., Liu, K., Gu, Y., We, Y., Tao, M., Fu, Z. and Ye, J. (2024) ‘INSIDE: LLMs’ internal states retain the power of hallucination detection’, in *International Conference on Learning Representations 2024*. arXiv:2402.03744.

Chih, Y.-Y. and Zwikael, O. (2014) ‘Project benefit management: a conceptual framework of target benefit formulation’, *International Journal of Project Management*, 33(2), pp. 352–362. https://doi.org/10.1016/j.ijproman.2014.06.002.

Copeland, B.J. (2004) *The essential Turing: seminal writings in computing, logic, philosophy, artificial intelligence, and artificial life*. Oxford: Oxford University Press.

Dagdelen, J., Dunn, A., Lee, S., Walker, N., Rosen, A.S. and Ceder, G. (2024) 'Structured information extraction from scientific text with large language models', *Nature Communications*, 15, 1418. https://doi.org/10.1038/s41467-024-45563-x.

Data Axle (2025) *Sales productivity statistics*. Available at: https://www.salesgenie.com/blog/sales-productivity-statistics/ (Accessed: 6 May 2026).

Davenport, T.H. and Harris, J.G. (2005) 'Automated decision making comes of age', *MIT Sloan Management Review*. Available at: https://sloanreview.mit.edu/article/automated-decision-making-comes-of-age/ (Accessed: 15 July 2005).

Du, J., Wang, D., Lin, B., He, L., Huang, L.-C. and Wang, J. (2025) 'Use of deep learning-based NLP models for full-text data elements extraction for systematic literature review tasks', *Scientific Reports*, 15. https://doi.org/10.1038/s41598-025-03979-5.

EPRI (2021) *O&M cost estimates for gas turbine and combined-cycle plants*. Palo Alto: Electric Power Research Institute.

Fayyad, U., Piatetski-Shapiro, G., Smith, P. and Uthurusamy, R. (1996) *Advances in knowledge discovery and data mining*. Cambridge: MIT Press.

Frey, C.B. and Osborne, M.A. (2017) 'The future of employment: how susceptible are jobs to computerisation?', *Technological Forecasting & Social Change*, 114, pp. 254–280. https://doi.org/10.1016/j.techfore.2016.08.019.

G20 (2024) *G20 Brazil Sherpa Track Digital Economy Ministers Maceió Ministerial Declaration*. Available at: https://g7g20-documents.org/database/document/2024-g20-brazil-sherpa-track-digital-economy-ministers-ministers-language-g20-dewg-maceio-ministerial-declaration (Accessed: 6 May 2026).

Glette-Iversen, I., Flage, R. and Aven, T. (2023) 'On the foundational issues of risk and risk analysis', *Safety Science*. https://doi.org/10.1016/j.ssci.2023.106317.

González-Flores, L., Rubiano-Moreno, J. and Sosa-Gómez, G. (2025) 'The relevance of lead prioritization: a B2B lead scoring model based on machine learning', *Frontiers in Artificial Intelligence*, 8. https://doi.org/10.3389/frai.2025.1554325.

Goodfellow, I., Bengio, Y. and Courville, A. (2016) *Deep learning*. Cambridge: MIT Press.

Goos, M., Manning, A. and Salomons, A. (2014) ‘Explaining job polarization: routine-biased technological change and offshoring’, *American Economic Review*, 104(8), pp. 2509–2526. https://doi.org/10.1257/aer.104.8.2509.

Grobelnik, M. and Mladenic, D. (2005) ‘Automated knowledge discovery in advanced knowledge management’, *Journal of Knowledge Management*, 9(5), pp. 132–146. https://doi.org/10.1108/13673270510622500.

Hodges, A. (2019) ‘Alan Turing’, in *Stanford