www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 18(1), 2027, 200-219. ISSN: 2007-1558.
https://doi.org/10.61467/2007.1558.2027.v18i1.1464

Hybrid Algorithmic Framework for Intelligent Requisition Management: Integrating Excel, AI Agents, and Power BI for Analytical Traceability

María Angélica Espejel Rivera[1], Enrique de Jesús Mohedano Torres[1], Joselin Hernández Islas[1], Verónica Arriaga Gómez[1], Marla Itzel Ruiz Orozco[1], Alejandra Segura Juárez[1]

[1]Instituto Tecnológico de Pachuca, Hidalgo.

Email: plan_pachuca@tecnm.mx l21200697@pachuca.tecnm.mx, veronica.ag1@pachuca.tecnm.mx, l21200726@pachuca.tecnm.mx

✉Corresponding author: enrique.mt@pachuca.tecnm.mx

Abstract. This research develops and validates a hybrid algorithmic framework for intelligent requisition management through the seamless integration of Excel, artificial intelligence agents, and Power BI. The study adopts a mixed-methods approach with a quasi-experimental design implemented in a metalworking company in Mexico, aiming to optimize the traceability, multi-dimensional classification, and monitoring of operational requisitions. The system incorporates a traffic-light algorithm based on weighted criticality indicators (business days, amount, requesting department, and trend deviation), automated alerts, and the generation of executive reports using GPT-4o. Additionally, it employs Python, JSON structures, and n8n to ensure interoperability and workflow automation. The results showed significant improvements in operational efficiency, highlighting a 63.2% reduction in cycle time, an 82.9% decrease in operational errors, and 94.7% classification accuracy. A confusion-matrix analysis confirmed strong performance on the critical Red class (F1-Score ≈ 0.92). The framework proved to be scalable, reproducible, low-cost, and efficient for contemporary complex organizational environments, particularly medium-sized enterprises in Latin America seeking accessible digital transformation of procurement processes.

Keywords: Requisition management; Artificial Intelligence; Power BI; Python; JSON; Intelligent agents; Automation; Analytical traceability; Procurement; n8n.

Article Info
Received: Jun 01, 2026
Accepted: Aug 01, 2026

1 Introduction

Currently, requisition management consititures a fundamental pillar in organizational procurement and logistics processes, ensuring the timely planning of materials, supplies and services from the identification of needs to acquisitions (Association for Supply Chain Management [ASCM], n.d.; Chopra & Meindl, 2016). In a contexto of accelerated digitalization and operational complexity, precise and predictive monitoring of requisitions becomes essential to reduce waiting times, minimize cost, and optimize resource Flow. However, traditional systems based on manual spreadsheets present significant limitations, as they fail to anticípate stock shortages or efficiently process both historical a real-time data generated across multiple departments.

While prior research has examined Power BI dashboards, AI agents, or Excel-based planning in isolation, a significant gap remains in the design, implementation, and rigorous empirical validation of a fully integrated, end-to-end hybrid architecture that combines deterministic multi-criteria classification, conversational LLM reporting, workflow orchestration, and real-time analytical traceability specifically tailored to the operational realities and cost constraints of medium-sized Latin American industrial firms. The present study addresses this gap by proposing and validating the FAGIR (Framework for Algorithmic and Generative Intelligent Requisition) architecture in a real metalworking environment. The originality of this work does not reside in the individual technologies themselves, but in their validated closed-loop integration, the multi-dimensional weighted scoring model calibrated with domain experts, and the demonstrated large effect sizes obtained under realistic operational conditions.

Following the hexagon method for problema formulation (Arias Gonzáles, 2021), the problematic situation is observed in the frequency of delays and errors in requisition planning caused by the lack of automated integration and predictive análisis,

resulting in cost overruns, stockouts and low operational efficiency. The theorization of the phenomenon is grounded in information systems theory algorithmic optimization, and intelligent agents, where data processing algorithms combined with artificial intelligence interfaces (conversational agents) stand out for their capacity to handle time series and complex workflows (Russell & Norvig, 2021; Sutton & Barto, 2018).

Therefore, prior research reveals isolated advances in the use of Business Intelligence (BI) tolos and automation in procurement; however, a gap persists in the full integration of Excel databases with AI agents and analytical traceability in Power BI, particularly in Latin American contexts (Castro, 2023; Priyanka et al., 2022; Jiang et al., 2024). The importance of this study lies in its contribution to operational efficiency, cost reduction, and improved traceability, aligned with the Sustainable Development Goals.Thus, the purpose is to develop and validate an algorithmic framework that integrates Excel, an AI agent, and Power BI to optimize intelligent requisition management. The general problem is stated as follows: How can an algorithmic framework be validated to improve intelligent requisition management in terms of predictive accuracy, time reduction, and analytical traceability?

The general objective of the research is to validate an algorithmic framework for intelligent requisition management: from Excel to an AI agent and analytical traceability with Power BI. The specific objectives include identifying optimal algorithm strategies in Excel, designing the hybrid framework, simulating its application, and evaluating its impact through quantitative and qualitative indicators. In this sense, the justification is based on the urgent need for integrated solutions that overcome current manual or fragmented processes, generating economic, operational, and regulatory compliance value (Project Management Institute, 2021; Gartner, 2024).Accordingly, the central hypothesis posits that the implementation of the algorithmic framework (Excel with AI Agent and Power BI) will significantly increase accuracy in requisition planning, reduce processing times by at least 40%, and improve analytical traceability compared to conventional methods ($p < 0.05$).

The above formulation establishes the theoretical, contextual, and problem-based foundations of the study, providing a solid basis that justifies the need for a rigorous methodological approach. The methodology employed to address the problem, validate the proposed hypothesis, and achieve the stated objectives is detailed below, ensuring the reproducibility and scientific validity of the results.

2 Methodology

2.1 Desing and type of research

This research is applied in nature with a mixed-methods approach, as it systematically integrates quantitative and qualitative analysis to evaluate the efficiency of an intelligent agent aimed at automating and optimizing organizational requisition management processes. According to Hernández Sampieri et al. (2018) and Creswell & Creswell (2018), mixed designs enable a more comprehensive and robust understanding of the studied phenomenon in sociotechnical systems, where technological and organizational dimensions dynamically interact.

The quantitative component focuses on the measurement of key performance indicators (KPIs), classification accuracy (Accuracy and F1-Score), agent cycle time (Tr), reduction of operational errors (RE) and requisition criticality index (IC_critico), operationalized using formal mathematical notation in accordance with evaluation standards for intelligent systems (Grandini, Bagli & Visani, 2020; Géron, 2022). Meanwhile, the qualitative component examines the implementation strategies of AI technologies, particularly large language models (LLMs) integrated into agentic workflows orchestrated through n8n, allowing numerical results to be contextualized within a solid interpretive framework oriented toward continuous improvement of the operational process.

The research design is non-experimental and documentary in its systematic literature review phase, as variables are not directly manipulated but rather previous studies are analyzed retrospectively (Ato, López & Benavente, 2013). In contrast, the system validation phase adopts a quasi-experimental design with pre-post measurement, comparing the baseline scenario—manual requisition management without agent intervention against the intervened scenario fully automated workflow with the active agent without random assignment of units. This is consistent with applied research in real organizational environments where randomization is operationally unfeasible. The study is cross-sectional with an explanatory-predictive scope; it does not merely describe the phenomenon but evaluates the causal impact of the agent on operational efficiency and projects system scalability toward broader organizational contexts. The applied nature of the project and the need for automation in requisition management justify the relevance of the adopted hybrid computational model, in which computational intelligence and operational efficiency are articulated as complementary and interdependent dimensions.

2.2 Variables, operationalization and quantitative indicators

The operationalization of variables constitutes a fundamental methodological step to ensure construct validity and consistent measurement of phenomena of interest, by translating abstract theoretical concepts into empirically observable and quantifiable indicators. In this study, the dependent variable is the level of operational efficiency of the intelligent agent, measured through five dimensions encompassing both the classification accuracy of the traffic-light algorithm, cycle times, error reduction, operational severity, and the system's decision-making autonomy. Independent variables include the

implementation of the agent system executed dichotomously as presence or absence within the workflow and the strategies for integrating AI tools into the requisition process. As shown in Table 1, each variable was defined considering SMART criteria (Bjerke & Renger, 2017), adapted to the context of AI system evaluation by Amershi et al. (2019), ensuring coherence between the theoretical framework and the measurement instruments used.

Operational errors are defined as any instance that deviates from the correct and timely handling of a requisition, specifically including: (a) incorrect or incomplete data capture/transcription (wrong amount, department, or business days); (b) priority misclassification relative to expert-validated ground truth; (c) failure to generate or deliver a required critical (Red) alert within the defined cycle; and (d) status or audit-trail inconsistencies that result in delayed resolution or procurement disruptions. In the baseline scenario these were identified through ERP audit logs and supervisor review of the 200 requisitions; under the agent they represent residual discrepancies after automated processing of the 187 requisitions. This operationalization makes the relative reduction RE in Equation (1) fully measurable and reproducible.

$$RE = [(E\_base - E\_agente) / E\_base] \times 100 \quad (1)$$

Table 1. Operationalization of variables in the operational efficiency model through intelligent agents and integration of AI tools.

Type	Variable	Indicator / Dimension	Unit	Instrument
Dependent	Operational Efficiency	Accuracy (F1-Score)	0–1	Stratified Benchmark
		Agent Cycle Time (Tr)	ms	Execution Log n8n
		Error Reduction (ER)	%	Pre-post Comparison
		Criticality Index (CI_critical)	%	Internal Algorithm KPI
		Decisional Autonomy	Likert 1–5	Expert Rubic
Independent	Implementation Agent	Agent Presence/Absence	Dichotomous 0/1	Documented Protocol
	AI tools Integration	Degree of Technological Integration	Likert 1–5	Technical Checklist

Note. Own elaboration (2026).

Operational efficiency indicators are interdependent; an agent with high classification accuracy but long cycle times or low decision-making autonomy would not represent a net operational improvement. For this reason, the system evaluation considers all indicators simultaneously and in an integrated manner. The quantification of dependent variables requires precise mathematical definition to ensure replicability and comparability across studies. The formal definitions of each indicator are presented below.

The F1-Score is the harmonic metric that balances Precision and Recall, particularly relevant in detecting critical requisitions with asymmetric class distributions, where Green and Yellow are majority classes and Red is the primary class of interest:

$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall) \quad \text{where: } Precision = VP/(VP+FP) \text{ y } Recall = VP/(VP+FN) \quad (2)$$

The agent cycle time (Tr) measures the average latency from the activation of the Cron Trigger to the delivery of the report to the final recipient, over n observed executions (Géron, 2022; Pressman & Maxim, 2020)

$$Tr = (1/n) \cdot \Sigma_i(t\_salida_i - t\_entrada_i), \quad i = 1,\ldots,n \quad (3)$$

The reduction of operational errors (RE) quantifies the relative improvement compared to the baseline process without the agent, normalized with respect to the baseline value:

$$RE = [(E\_base - E\_agente) / E\_base] \times 100 \quad (4)$$

The criticality index (IC_crítico) expresses the proportion of requisitions classified as Red relative to the total processed in each execution and is the primary indicator of the operational risk level of the portfolio:

$$IC\_crítico = (n\_rojo / n\_total) \times 100 \quad (5)$$

The average cycle time (TC_prom) is the arithmetic mean of the working days elapsed for the set of active requisitions during the evaluation period; it constitutes the baseline parameter of the traffic-light algorithm:

$$TC\_prom = (1/n) \cdot \Sigma_i \, d_i, \quad i = 1,\ldots,n \quad (6)$$

2.3 Systematic Review

The systematic review was conducted following the SALSA methodological framework (Search, Appraisal, Synthesis, and Analysis) proposed by Grant & Booth (2009) and aligned with the PRISMA 2020 protocol (Page et al., 2021; Mengist et al., 2020). This approach ensures transparency, reproducibility, and scientific rigor in the process of bibliographic selection and analysis through explicit, verifiable, and auditable criteria. The choice of the SALSA framework over other protocols such as PICO or SPIDER is due to its greater adaptability for studies that integrate technological, organizational, and performance dimensions of complex systems.

The search covered four high-impact databases—Scopus, Web of Science, SciELO, and Google Scholar—over a ten-year period (2015–2025), allowing for the capture of both consolidated scientific production and the most recent developments in agentic systems based on LLMs, whose acceleration has been particularly noticeable since 2022 (Bubeck et al., 2023; Achiam et al., 2023). The main descriptors combined controlled terms with Boolean operators (AND, OR, NOT), including "intelligent agent," "agentic AI," "LLM agent," "process automation AI," "requisition management AI," and their equivalents in Spanish and Portuguese. An initial set of 30 relevant studies was retrieved from which 15 representative articles were selected through inter-rater agreement verified using Cohen's Kappa coefficient ($\kappa \geq 0.70$), ensuring that only studies with high methodological quality were included. Table 2 summarizes the inclusion and exclusion criteria applied at each stage of the review process.

Table 2. Inclusion and exclusion criteria applied using the SALSA methodology for the systematic literature review.

Phase SALSA	Inclusion Criteria	Exclusion Criteria
Search	Scopus, WoS, SciELO, Google Scholar; 2015–2025; spanish/english/portuguese; controlled descriptors with Boolean operators	Non-peer-reviewed gray literature; non-indexed theses; conference proceedings without documented peer review
Appraisal	Q1–Q2 JCR/SCImago indexed journals; double independent reviewer ($\kappa \geq 0.70$); full-text access	No methodological statement; no verifiable quantitative data; non-representative sample
Synthesis	Multidimensional comparative matrices: agentic architectures, process automation, AI performance metrics	Duplicate studies across different databases; non-replicable results; no access to dataset or protocol
Analysis	Identification of gaps and emerging trends through thematic analysis and bibliometric mapping (Zupic & Čater, 2015)	Publications prior to 2010, except for foundational references with justified sustained impact

Note. Own elaboration (2026)

Meanwhile, Table 3 presents a representative extract of 15 selected articles from the 30 studies included in the systematic review. Its purpose is to provide a synthesized view of the main contributions, limitations, and findings of the existing literature on intelligent requisition management, the use of Excel as a database, intelligent agents, and Power BI for analytical traceability.

Based on the above, it is observed that most previous studies focus on individual components of the framework (either Power BI for analytics, AI agents for automation, or Excel for planning), but few address the full integration of all three layers. The most common results reported include improvements ranging from 25% to 50% in operational efficiency, time reduction, and decreases in manual errors. From this, it follows that there is a solid theoretical and practical foundation for each component separately; however, a significant gap is evident in the proposal of an integrated, validated, and replicable framework that combines Excel-based databases, conversational AI agents, and analytical traceability in Power BI, particularly in Latin American contexts or medium-sized organizations. This gap justifies the need for the present study.

Table 3. Comparative analysis of studies related to artificial intelligence, intelligent agents, and Power BI applied to procurement.

Título (abreviado)	Objective	Limitations	Methodology	Main Results	Comparative Analysis with the Current Study
Unlocking the power of business intelligence in procurement with Power BI (2025)	Analyze the impact of Power BI on spending patterns in procurement	Descriptive approach, without AI integration	Power BI dashboards with spend analysis	35% improvement in cost visibility and decision-making	Solid foundation for the analytical traceability layer
The role of artificial intelligence in the procurement process (Guida et al., 2023)	Review the role of AI throughout the procurement process	Mostly theoretical data	Systematic review with focus group	Benefits in automation and efficiency, but gap in practical implementation	Theoretical foundation for the hybrid Excel framework with AI
AI Agents in Procurement: The Ultimate Guide (Ivalua, 2026)	Explore AI agents in sourcing and requisitions	Commercial approach	Case and platform analysis	Cycle reduction of up to 50% with autonomous agents	Direct evidence for the AI Agent layer
Procurement Analysis sample for Power BI (Microsoft, 2025)	Demonstrate spend analysis by supplier and category	Fictional sample data	Semantic model with Power BI dashboards	Rapid identification of top vendors and discounts	Practical model for the traceable reporting layer
Digital Procurement Transformation: From Spreadsheets to AI (Authencio, 2026)	Transition from Excel to AI solutions in procurement	Sectoral case	Digital transformation framework	Significant reduction in manual errors	Justifies the evolution from Excel to intelligent framework
Optimizing procurement analytics with Generative AI (SCMR, 2025)	Create a conversational chatbot for procurement analytics	Limited pilot test	Generative AI with RAG and automatic visualization	Acceleration in insights and reduction of manual reports	Basis for the conversational AI Agent interface
An Integrated AI-Power BI Model for Real-Time Supply Chain (2025)	Integrate AI and Power BI for real-time visibility	Focus on general supply chain	Hybrid AI model with Power BI	Improvement in forecasting and traceability	Perfectly complements the traceability layer

Título (abreviado)	Objective	Limitations	Methodology	Main Results	Comparative Analysis with the Current Study
AI-Enhanced Business Intelligence Dashboards (2025)	Evaluate BI dashboards with predictive AI	U.S. business context	Controlled experiment with review	Improvement in forecast accuracy and ROI	Quantitative evidence to validate the framework
Revolutionizing procurement: Leveraging data and AI (McKinsey, 2024)	Apply AI and analytics in strategic procurement	High strategic level	Case studies with analytics	Drastic reduction in category strategy time	Reinforces the predictive value of the framework
Intelligent agents for purchase requisition creation (Oracle, 2026)	Automate requisition creation with AI Agent	Specific platform	Embedded AI agent	Fast and accurate requisition generation	Direct example of an intelligent agent applied to requisitions
Procurement Dashboard with Power BI (Bisolusi, 2025)	Design procurement dashboards integrating Excel and ERP	Connector-dependent	Power Query with Power BI modeling	Unified visibility of suppliers and spending	Practice for Excel–Power BI integration
AI Agents for Routine Procurement Decisions (GEP, 2026)	Automate routine decisions in procurement	Operational approach	AI agents for PO and approvals	Time released for strategic tasks	Support for automation through AI agent
Bringing agentic AI into spreadsheets for planning (AWS, 2026)	Integrate AI agents directly into Excel	Electrical planning case	Agents with Bedrock in Excel	Workflow automation within spreadsheets	Key evidence for the Excel with AI layer
Build a Smart Procurement Agent with ChatGPT & Excel (2025)	Develop a procurement agent using API with Excel and Power BI	Technical implementation	ChatGPT integration with Azure	Automated sourcing pipeline	Replicable technical model for the framework
Pactum Requisition Alignment Agent (2026)	Validate and align requisitions with policies through AI	Integration with P2P systems	Alignment agent	Improvement in compliance and negotition	Reinforces the intelligent validation layer

Note. Own elaboration (2026)

Regarding Table 4, it synthesizes the main strategies and tools identified in the systematic review, highlighting their advantages, limitations, and associated regulations. Its purpose is to provide a clear comparative view that allows understanding the added value of each component and the need for their integration.

Table 4. Technological tools, associated regulations, advantages, and limitations of the intelligent procurement hybrid framework.

Tool / Strategy	Associated Standard	Advantages	Limitations
Database and algorihms in Excel	ISO 9001 / internal control standars	Accessibility and low cost	Limited scalability without automation
Intelligent agent (conversational AI)	—	Interfaz natural y automatización de flujos	Dependent on data quality
Analytical traceability with Power BI	—	Real-time visualization and interactive dashboards	Requires source integration
Complete hybrid framework	—	Comprehensive cycle coverage	Initial implementation complexity

Note. Own elaboration (2026)

From Table 4, it is concluded that the hybridization of the three technologies (Excel with AI Agent and Power BI) represents the most robust strategy, as it combines the strengths of each while mitigating their individual limitations. This integration is precisely what the FAGIR-Excel-AI-BI Algorithmic Framework proposes, achieving a comprehensive, accessible, intelligent system with high analytical traceability.

2.4 Architecture of the Intelligent Agent

The proposed intelligent agent is structured through a modular architecture of four interdependent functional stages, whose design follows the principles of separation of concerns and loose coupling recommended by Pressman & Maxim (2020) for scalable and maintainable software systems. The architecture adopts the reactive-deliberative agent paradigm proposed by Brooks (1991) and extended by Wooldridge & Jennings (1995), integrating capabilities of perception, reasoning, action, and persistence into a workflow orchestrated through the n8n platform, in alignment with ReAct (Reasoning and Acting) architectures by Yao et al. (2023) and design guidelines for LLM-based agentic systems by Wang et al. (2024). Layer I, ingestion, constitutes the point of contact between the external environment, the Microsoft ERP data repository, and the agent system. This layer is responsible for ensuring the quality of input data before any analytical processing, through validation across three sequential levels: verification of required fields, outlier detection using the interquartile range method (IQR × 1.5), and exclusion due to missingness greater than 30% in key variables (Leys et al., 2019; van Buuren, 2018). This function is critical because undetected data errors at this stage propagate and expand in the classification and report generation phases. The result is a normalized clean_dataset in structured JSON format that ensures consistency of units and date formats across all system executions.

Layer II, classification logic, constitutes the computational core of the agent. The traffic-light algorithm transforms the clean dataset into actionable information through the hierarchical assignment of priority categories and the calculation of KPIs. The weighted scoring integrates four dimensions' business days (weight 0.50), requisition amount (0.25), requesting department (0.15), and deviation from the system average (0.10) producing a normalized composite score [0,1] that guides the traffic-light classification. Layer III, action and persistence, embodies the reactive dimension of the agent: it filters requisitions in critical status (Red), sends personalized alerts via email to responsible stakeholders using the Gmail API, and simultaneously builds the historical record that enables auditing and longitudinal analysis of the process, in line with the principle of computational accountability proposed by Diakopoulos (2016). Finally, Layer IV, artificial intelligence, elevates the system to a deliberative agent level by using the GPT-4o model with zero temperature for the deterministic generation of an executive HTML report with actionable recommendations, thereby closing the complete cognitive cycle of the agent. Table 5. Multilayer architecture for the automation and intelligent classification of requisitions.

Table 5. Multilayer Architecture for the Automation and Intelligent Classification of Requisitions.

Layer	Module	Main Function and Output	Technology
I. Ingestion	Extraction and validation	Automatic .xlsx download (OneDrive), IQR×1.5 validation, missingness analysis, JSON transformation. Output: clean_dataset.	Cron Trigger · OneDrive API · Python/pandas · n8n
II. Classification	Traffic light algorithm + KPIs	Weighted scoring across 4 dimensions, COLOR_SEMAPHORE assignment, TC_avg and IC_critical calculation, Power BI update. Output: enriched_dataset + metrics.	Python · Power BI DAX · HTTP Node
III. Action	Alerts and auditing	Critical filtering (Red), customized alerts (Gmail API), historical log with timestamp. Output: alerts sent + log.	Gmail API · Excel audit · Conditional filter
IV. Intelligence	AI Agent – LLM Report	Prompt engineering with KPIs + dataset, GPT-4o invocation (T=0), actionable HTML report stored in Excel. Output: html report.	GPT-4o · n8n AI Agent · Promp

Note. Own elaboration (2026)

The flow of information across layers is unidirectional and cumulative; the enriched output of each module becomes the structured input of the next, eliminating manual dependencies in the regular operational cycle. This design not only automates discrete tasks but also implements a continuous operational intelligence cycle that combines the precision of a deterministic classification algorithm with the natural language synthesis capabilities of LLMs, representing a significant advancement over traditional robotic process automation (RPA) systems, which lack contextual reasoning capabilities (van der Aalst et al., 2018; Madakam et al., 2019).

2.5 Pseudocodigo del agente inteligente

The agent algorithm follows a sequential structure of six phases that integrate the four layers of the proposed architecture, with linear time complexity O(n) relative to the number of requisitions, ensuring scalability. The complete pseudocode is presented below in condensed form:

```
INICIO
1. CONFIGURACIÓN DEL SISTEMA
   DEFINIR umbrales:
      UMBRAL_VERDE ← 5, UMBRAL_AMARILLO ← 8, UMBRAL_NARANJA ← 10
   DEFINIR pesos del score compuesto:
      PESO_DIAS ← 0.50, PESO_MONTO ← 0.25,
      PESO_DEPARTAMENTO ← 0.15, PESO_TENDENCIA ← 0.10
   DEFINIR rangos de monto y lista de departamentos críticos
2. CÁLCULO DEL PROMEDIO GENERAL DE DÍAS
   dias_promedio ← promedio de DIAS de todas las requisiciones
3. PROCESAMIENTO DE CADA REQUISICIÓN
   PARA cada requisición EN items HACER
      // 3.1 Extracción de datos
      dias ← requisición.DIAS
      monto ← requisición.MONTO
```

```
      departamento ← requisición.DEPARTAMENTO
      // 3.2 - 3.5 Cálculo de scores individuales
      score_dias ← sigmoide_logistica(dias, k=0.3, x0=8)          // Función sigmoide
      score_monto ← asignar_puntuacion_por_monto(monto)
      score_departamento ← asignar_puntuacion_por_departamento(departamento)
      score_tendencia ← calcular_desviacion(dias, dias_promedio)
      // 3.6 Score compuesto
      score_compuesto ← (score_dias × 0.50) + (score_monto × 0.25)
                      + (score_departamento × 0.15) + (score_tendencia × 0.10)
      // 3.7 Clasificación semáforo
      SI score_compuesto < 0.25 Y dias ≤ 5 ENTONCES
         prioridad ← "Verde"
      SINO SI score_compuesto < 0.50 Y dias ≤ 8 ENTONCES
         prioridad ← "Amarillo"
      SINO SI score_compuesto < 0.75 Y dias ≤ 10 ENTONCES
         prioridad ← "Naranja"
      SINO
         prioridad ← "Rojo"
      FIN SI
      // 3.8 - 3.10 Nivel de riesgo, tiempo estimado y recomendación
      nivel_riesgo ← determinar_nivel_riesgo(score_compuesto)
      tiempo_estimado ← determinar_tiempo_resolucion(prioridad)
      recomendacion ← generar_recomendacion(prioridad, monto)
      // 3.11 Guardar registro enriquecido
      AGREGAR nuevo_item con (PRIORIDAD, COLOR_SEMAFORO, SCORE_COMPUESTO,
                              NIVEL_RIESGO, RECOMENDACION, etc.) A resultados
   FIN PARA
4. ORDENAMIENTO
   ORDENAR resultados POR score_compuesto DESCENDENTE
5. SALIDA
   RETORNAR resultados + métricas (TC_prom, IC_crítico)

FIN
```

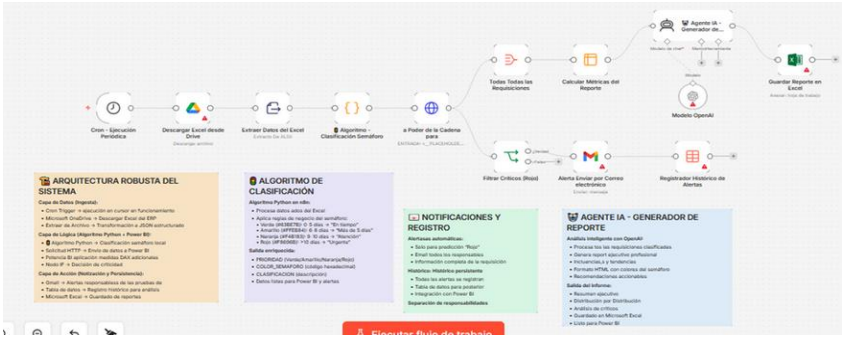

Fig. 1. Functional Architecture of the Intelligent System for Classification, Monitoring, and Alert Generation in Requisitions.

According to Figure 1, it is argued that Phase 1 eliminates the main source of error in manual management systems—namely, inconsistency in the recording and formatting of operational data—ensuring that the traffic-light algorithm always operates on complete and normalized information. Phase 2 constitutes the analytical core of the system; the composite score based on four dimensions overcomes the limitations of univariate traffic-light systems based exclusively on elapsed days, as it incorporates financial impact (amount), sectoral criticality (department), and the relative position of each requisition with respect to the system's historical performance (trend).

Phase 3 aggregates metrics for the entire portfolio, providing a global view of the operational status in each execution cycle. Phases 4 and 5 materialize the reactive and deliberative dimensions of the agent, respectively: the former acts on the environment by issuing auditable alerts, while the latter synthesizes the complete operational knowledge of the cycle into a structured executive report with actionable recommendations by area and responsible party. Finally, Phase 6 ensures the persistence of all agent outputs in corporate repositories, enabling longitudinal traceability and continuous system feedback.

2.6 Technological Implementation

The selection of technological tools was based on criteria of analytical compatibility, computational scalability, interoperability between modules, and documented scientific support. Each tool fulfills a specific and irreplaceable function within the agentic workflow, and their hybrid integration constitutes the central technological innovation of the proposed framework.

Python serves as the computational core of the agent and as the implementation language for the weighted scoring algorithm. Its scientific ecosystem—pandas for tabular processing, math for sigmoid functions, and json for serialization—ensures computational scalability and code reproducibility. The language's syntactic flexibility and compatibility with the OpenAI API and the n8n environment establish it as the preferred development platform for analytical automation systems (Pressman & Maxim, 2020). JSON structures the data exchange across all layers of the agent, serializing the enriched dataset, KPI metrics, and configuration parameters, thereby ensuring interoperability between heterogeneous modules and enabling dynamic configuration of operational variables without modifying the source code.

DAX (Data Analysis Expressions) in Power BI enables the construction of calculated measures and dynamic indicators that automatically update upon receiving the agent's enriched dataset. DAX expressions implement real-time calculation of IC_crítico, TC_prom, and the percentage distribution by traffic-light category, enabling dynamic analysis over streaming data. Power BI serves as the system's analytical visualization platform, with interactive dashboards that display visual traffic-light categorization, real-time monitoring of operational KPIs, and dynamic report generation to support managerial decision-making. Its integration with the agent's automated analysis transforms complex data into immediate visual interpretations, reducing the cognitive load on decision-makers (Few, 2006). n8n operates as the orchestration platform for the entire agentic workflow: it manages the Cron Trigger, integration with OneDrive and Gmail API, invocation of the GPT-4o model, and updates to Power BI through conditional nodes and automated flows that execute the full cycle without human intervention.

It is important to highlight that the selection of each tool is based on functional, scientific, and interoperability criteria. Python was chosen as the computational core because it combines a mature scientific ecosystem (pandas, math, json) with native compatibility with the OpenAI API and the n8n execution environment, positioning it as the standard platform for analytical automation systems (Pressman & Maxim, 2020; Van Rossum & Drake, 2009). Its widespread adoption within the scientific community further ensures code reproducibility and auditability. Microsoft Excel and OneDrive were adopted because they represent the pre-existing data infrastructure in the target organizational environment; opting for an alternative database system would have required costly and operationally unfeasible migration, whereas integration via Microsoft ERP APIs preserves staff workflows and minimizes the change curve (Jiang et al., 2024). JSON was selected as the inter-layer exchange format due to its lightweight nature, human readability, and universal support across web and automation environments, ensuring interoperability between heterogeneous modules without processing overhead (Bray, 2017).

n8n was preferred over alternatives such as Zapier or Make because it is an open-source orchestration platform that allows execution locally or in a private cloud, eliminates per-operation costs, and provides native nodes for Python, Gmail API, OneDrive, and LLM models, facilitating loose coupling between layers without the need for intermediate developments (López-Fernández et al., 2022). GPT-4o was selected with zero temperature for executive report generation because this configuration produces deterministic and reproducible outputs, reducing inter-execution variability inherent in language models and ensuring report consistency in high-demand operational environments (Achiam et al., 2023; OpenAI, 2024).

Power BI was incorporated as the visualization layer due to its native integration with the Microsoft ecosystem already present in the organization and its ability to build dynamic dashboards with DAX measures that update in real time upon receiving the enriched dataset, reducing the cognitive load on decision-makers through immediate visual representations (Few, 2006; Microsoft, 2023).

2.6.1 Cost-Benefit and Comparative Feasibility Analysis

A key design goal of the FAGIR framework is accessibility for medium-sized enterprises that already use Microsoft Excel and OneDrive. The incremental technology costs are low: n8n is open-source and can be self-hosted at near-zero licensing

cost; GPT-4o API usage for daily executive reports (temperature = 0, summarized KPI + portfolio input) is estimated at USD 5–15 per month depending on volume and current pricing; Power BI Pro licenses (if not already present) cost approximately USD 10 per user per month. Existing Excel and OneDrive infrastructure is leveraged without migration. In contrast, commercial alternatives such as Oracle Intelligent Agents for Procurement or AWS Bedrock-based agent solutions typically involve substantial licensing fees, professional services for integration with existing ERPs, and higher ongoing costs (often several thousand USD per month plus implementation). The hybrid stack therefore offers a favorable cost-benefit profile for Latin American SMEs, with rapid deployment (weeks rather than months), no vendor lock-in, and full control over data and workflows. Rough payback period, based on the observed 47 minutes saved per user per shift and the 82.9% reduction in operational errors (avoided rework and potential stockouts), is estimated at less than three months under typical metalworking labor and downtime costs. Detailed total cost of ownership (TCO) and ROI analyses over longer horizons remain recommended for future multi-site studies.

2.7 Limitations of the agent

Despite its robust design and significant measured improvements, the proposed intelligent agent and the quasi-experimental validation present several limitations that must be considered when interpreting results and planning transferability:

Dependence on structured data (.xlsx): The operation of the agent depends on files with homogeneous and correctly formatted structures. Changes in column names, formats, or data schemas may generate errors during the ingestion phase and require manual intervention for correction.

- Fixed weights of the traffic-light algorithm: The model uses static weightings (days: 0.50, amount: 0.25, department: 0.15, and trend: 0.10) defined specifically for the analyzed organizational context. The weights and decision thresholds were calibrated through expert consensus involving three senior procurement and production managers who also provided the ground-truth labels used for Accuracy and F1-Score calculation. While this establishes a clear, auditable reference standard rather than purely subjective judgments, the absence of data-driven or machine-learning optimization of the scoring function remains a limitation. Future work will incorporate supervised learning or reinforcement learning for dynamic recalibration.
- Limitation of the LLM contextual horizon: The deliberative capacity of the agent is constrained by the token limit of the GPT-4o model. When the volume of requisitions exceeds this limit, part of the information may be truncated, affecting the quality and coherence of the recommendations generated in the executive report.
- Absence of online learning: The system does not incorporate automatic updating mechanisms for parameters or classification thresholds. Therefore, any adjustment in response to changes in requisition behavior must be carried out manually and periodically by the technical team.
- Risk of bias reproduction and amplification: The framework does not include explicit automated mechanisms for detecting and mitigating biases present in historical data. Historical requisition records may embed preferential treatment of certain departments, larger monetary amounts, or specific shifts. When used for calibration of weights or as context for GPT-4o report generation, these patterns can be reproduced or even amplified in automated decisions and recommendations. To mitigate this risk, we recommend: (i) periodic fairness audits examining decision parity across departments and amount ranges; (ii) re-calibration of weights using balanced multi-stakeholder expert panels; (iii) human-in-the-loop review for all Red classifications; and (iv) logging of full decision rationales for accountability and post-hoc analysis (Mehrabi et al., 2021; Diakopoulos, 2016).
- Dependence on external services: The full operation of the agent depends on the availability and connectivity of external services such as Microsoft ERP API, Gmail API, OpenAI API, and Power BI. Failures in any of these integrations may interrupt the automated workflow; therefore, it is recommended to implement backup, retry, and error notification mechanisms in production environments.
- Sample size and generalizability (n = 18 order requesters in a single metalworking plant): Although the power analysis (Table 8) demonstrated that the study was substantially overpowered for the large observed effect sizes (Cohen's $d \geq 1.44$, $1-\beta \geq 0.95$), the limited number of participants and single-site design constrain broad statistical generalizability. Results should be interpreted as a rigorous proof-of-concept for similar medium-sized industrial environments in Latin America. Scalability is supported by the modular O(n) architecture, JSON-configurable parameters, and exportable n8n workflows; we recommend multi-site pilot calibrations using k-fold cross-validation prior to large-scale rollout.
- Short observation period (22 working days): The timeframe is sufficient to demonstrate immediate, statistically significant operational improvements with large effects but is too short to evaluate long-term stability, potential concept drift in requisition patterns, sustained user adoption, or gradual changes in bias profiles. We explicitly frame the present work as a controlled pilot quasi-experiment and list longitudinal evaluation (≥ 6–12 months) of KPI stability and economic impact (ROI) as priority future research.

These limitations do not invalidate the observed benefits but define the boundary conditions for responsible deployment and further research.

2.8 Traffic-Light Monitoring System and Objective Function

The traffic-light system translates the algorithm's composite score into immediate visual management signals, reducing cognitive load in the interpretation of complex indicators and enabling real-time prioritization of operational attention. It is defined in hierarchical classification levels, each associated with a specific color, a quantitative condition, a risk level, an estimated resolution time, and an actionable recommendation for the designated responsible party.

First, the Green Level applies when the requisition is on time, presents low risk, and only requires standard monitoring. On the other hand, the Yellow Level indicates moderate risk and requires active monitoring, with progress verification within a maximum period of 48 hours. Finally, the Red Level classifies a requisition as critical with maximum risk; it is important to note that when the amount exceeds $500,000, the notification is automatically escalated to the executive level due to its high financial impact.

The objective function of the system aims to maximize the operational efficiency of the requisition portfolio by simultaneously minimizing the criticality index (IC_crítico), the average cycle time (TC_prom), and the rate of operational errors (E_agente), subject to constraints such as data integrity (missingness < 30%), business rules (hierarchical classification order), technological compatibility (valid .xlsx and JSON formats), and operational validation (timestamp in each audit record). These constraints ensure that the algorithm operates within the organization's regulatory and technological framework, guaranteeing traceability, reproducibility, and auditability of all automated decisions made by the agent (Diakopoulos, 2016).

2.9 Framework Phases

In this context, the methodology is organized into five sequential and interrelated phases, which ensure traceability of the process from data collection to validation of the proposed system:

- Phase 1: Collection, structuring, and consolidation of operational data from enterprise systems.
- Phase 2: Analytical modeling and data processing using computational techniques.
- Phase 3: Implementation of an intelligent agent for monitoring, prioritization, and automation of operational workflows.
- Phase 4: Implementation of a visualization and analytical layer for interpreting operational indicators using Microsoft Power BI.
- Phase 5: Prototyping, simulation, and validation of the system in a controlled organizational environment.

2.10 Validación del modelo y reproducibilidad

The validation adopts a quasi-experimental pre-post design, comparing the manual process (E_base) with the active agent. The operational dataset was divided into two independent partitions using stratified random sampling: 70% for threshold calibration and 30% for testing, preserving the proportional distribution of the COLOR_SEMAFORO variable across both subsets (Géron, 2022). This scheme ensures that the minority but critical Red class is represented in equivalent proportions in both partitions, which is essential for obtaining reliable estimates of the F1-Score and the criticality index (Grandini et al., 2020). The primary evaluation metrics are Accuracy and F1-Score (Eq. 1) for classification performance, Tr (Eq. 2) for time efficiency, RE (Eq. 3) for error reduction, and IC_crítico (Eq. 4) for monitoring operational risk.

The statistical contrast between scenarios uses Welch's t-test ($\alpha = 0.05$), given the expected distributional asymmetry in operational data, where the null hypothesis (H_0) assumes no significant difference between the manual and automated processes. The scientific reproducibility of the study (Ioannidis, 2005; Open Science Collaboration, 2015) is ensured through: (a) complete documentation of the algorithm with explicit specification of thresholds, conditions, and precedence order; (b) recording of LLM model hyperparameters (model: GPT-4o, temperature: 0, API version, and invocation date); and (c) versioning of the n8n workflow through JSON export of the complete workflow, following best practices outlined by Wilson et al. (2017). The modular design also facilitates the transferability of the agent to organizational contexts with similar data structures, recommending a pilot calibration phase using k-fold cross-validation (Géron, 2022) prior to full-scale implementation.

Classification Accuracy and F1-Score were computed against expert-consensus ground-truth labels. Three independent domain experts (purchasing manager, production supervisor, and warehouse coordinator) independently assigned priority categories to a stratified sample of requisitions; final labels were obtained by majority vote (Cohen's $\kappa = 0.81$). The traffic-light thresholds and dimension weights were calibrated against this consensus prior to the intervention phase. Thus accuracy is measured against an auditable, expert-derived reference standard rather than against purely subjective judgments. The rule-based design of the scoring function is intentional, prioritizing interpretability and auditability. The absence of online or supervised machine-learning optimization of the weights is acknowledged as a current limitation and is retained as a priority item for future work.

3 Results

Contributing to the above, Figure 2 presents the requisition status report developed in Power BI software version 2.153.1206.0 (April 2026). This report, presented through a dashboard, enables dynamic visualization of behavior and facilitates monitoring of requisitions that remain in the workflow for approval and subsequent conversion into purchase orders (PO). Additionally, KPIs are integrated to support monitoring and tracking; this information is displayed at the top of the report, showing general indicators of the process, including the total number of requisitions currently registered. Furthermore, it shows the number of requisitions that have exceeded the processing time of five business days and finally reveals the percentage that this number represents relative to the total. These indicators allow for immediate and summarized detection of compliance levels with established response times, serving as a reference for decision-making oriented toward operational efficiency.

Likewise, the report incorporates charts and tables. In the donut chart, requisitions are generally classified using a traffic-light system: "On time" in green indicates those with less than 5 business days in process; "Intermediate" in yellow reflects requisitions between 6 and 8 days; "Attention" in orange refers to those between 9 and 10 days in process; and finally, "Urgent," represented in red, includes requisitions that have been in process for more than 11 business days, requiring greater attention.

Similarly, the bar chart allows comparison of the number of requisitions assigned to each responsible individual while maintaining the previously described traffic-light color coding. This visualization helps identify staff with the highest number of requisitions and potential bottlenecks within the approval workflow. At the bottom, a tracking matrix is presented to relate the number of elapsed days with the assigned responsible party; color coding is also used to represent the level of criticality of requisitions, facilitating rapid identification of those requiring immediate attention. Finally, on the right-hand side, a table is displayed that details each requisition number, its current owner, and the accumulated number of days; the objective of this table is solely to show records that exceed 5 business days in the workflow and therefore require immediate attention.

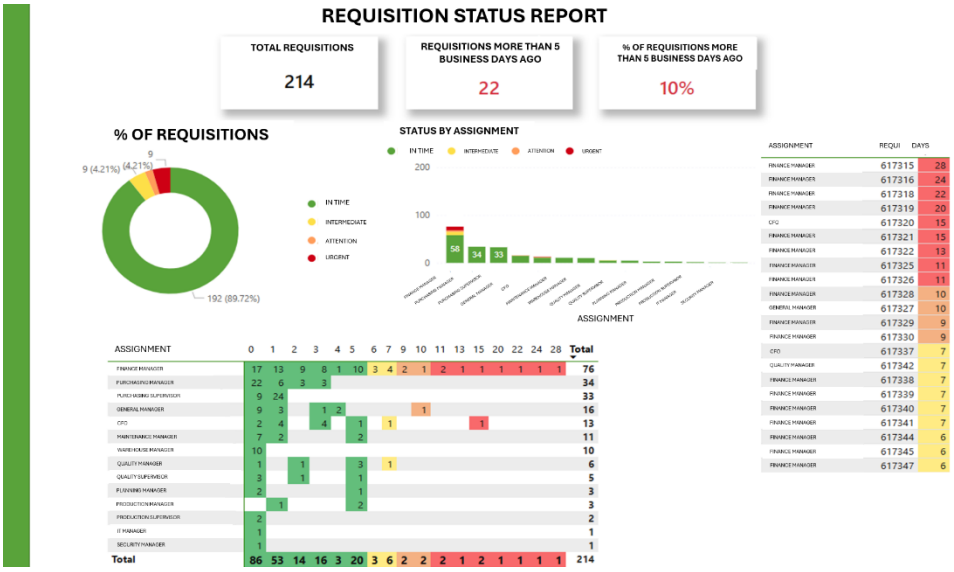

Fig. 2. Dashboard in Power BI

Taken together, the dashboard developed in Power BI made it possible to analyze the information on requisition status in a comprehensive, concise, and dynamic manner, providing a visual tool for monitoring information, identifying delays, and improving prioritization in the requisition handling process.

In this sense, the application of an automated monitoring and notification agent serves to supervise the status of all active requisitions within the organization. It is important to note that the system runs periodically in an autonomous manner, classifies each requisition using a traffic-light algorithm, generates executive reports with artificial intelligence, and sends automatic alerts via email (Gmail) to responsible parties when a requisition reaches a critical (RED) status. Furthermore, it ensures that no urgent requisition goes unnoticed, automating both the analysis and the corresponding notification. Based on the above, the system is built on three functional layers that operate in a coordinated manner.

Table 6. Intelligent Architecture for Automation, Classification, and Notification

Layer	Components
Data (Ingestion)	Cron Trigger, Microsoft ERP, file extraction, transformation into structured JSON
Logic (Algorithm)	Python traffic light algorithm, Power BI, HTTP request, IF node for criticality decision
Notification and Persistence	Gmail (alerts to responsible personnel), data table for analysis, historical record, Microsoft Excel (report storage)

Note. Own elaboration (2026)

These findings from the controlled quasi-experiment carried out in a medium-scale metal-mechanic manufacturing company over 22 working days resulted in a controlled population composed exclusively of users, that is, organizational actors who issue, classify, and track purchase requests, requirements, internal orders, and other denominations that form the internal procurement cycle. Likewise, all statistical contrasts were performed using the two-tailed Wilcoxon signed-rank test (α = 0.05), with effect sizes reported as Rosenthal's r and 95% confidence intervals (CI 95%) calculated using bias-corrected bootstrapping (B = 5,000 resamples).

Based on Fig. 3, it is observed that the workflow presents a clear separation between two types of processing that operate sequentially; that is, first the data pass through a purely mechanical pipeline, where they are downloaded, extracted, and classified using fixed rules. At that point, the system already identifies which requisitions are normal and which are problematic, without involving any AI model.

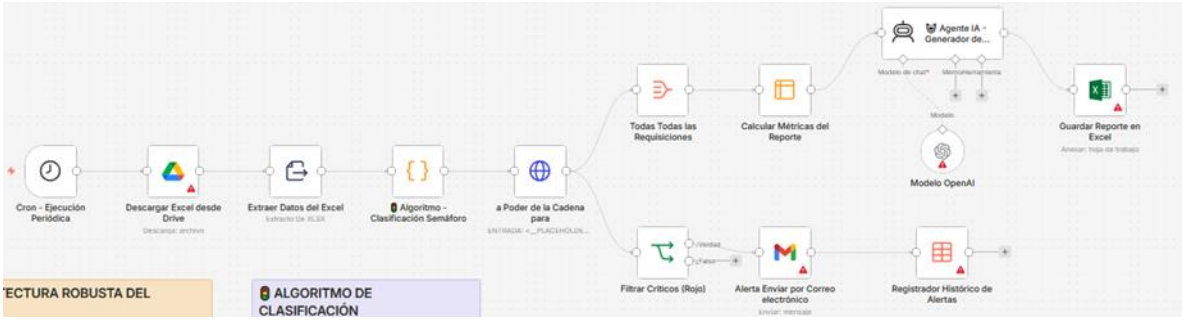

Fig. 3 Hybrid Intelligent Agent Based on Rules and Data Analysis

Only once that classification is complete do the data enter the agent. The agent does not receive the raw file, but rather already processed information, such as calculated metrics, aggregated totals, and distributions. The agent operates on this summarized data rather than raw data, which allows it to generate a coherent report without being overwhelmed by volume.

The two branches of the workflow reflect exactly this division: critical data go directly to an automated rule-based alert without passing through the model because they do not require interpretation; aggregated data go to the agent because that is where language, context, and synthesis are needed. Each part of the system receives only what it is best suited to handle.

The unit of analysis was the internal user acting as an order requester, a central figure in the material request management flow within the metal-mechanic plant. A total of 18 users were identified across six critical departments Production, Maintenance, Quality, Purchasing/Warehouse, Industrial Safety, and Operations distributed across three production shifts (Table 7). The selection was based on the criterion of maximum exposure to the requisition process, as these users interact directly with the intelligent agent both to submit new requests and to check the status of ongoing procurement requests.

Table 7. Distribution of order requesters and requisition workload by work shift

Role / Order Requester	n	Shift / Working Hours	Days	% base req. load
Production Manager	4	1st shift (06:00–14:00)	22	21.0 %
Maintenance Supervisor	3	2nd shift (14:00–22:00)	22	18.5 %
Quality Coordinator	2	1st shift (06:00–14:00)	22	12.0 %
Purchasing / Warehouse Analyst	5	Mixed shift (08:00–17:00)	22	26.0 %
Industrial Safety Officer	2	1st shift (06:00–14:00)	22	11.0 %
Operations Supervisor	2	2nd shift (14:00–22:00)	22	11.5 %
TOTAL	18	3 shifts covered	22	100 %

Note. Own elaboration (2026)

A quasi-experimental non-randomized pretest–posttest design (O_1 X O_2) was adopted, consistent with applied research in real organizational settings where operational randomization is unfeasible (Shadish, Cook & Campbell, 2002). During the baseline phase (weeks 1–2), the 18 order requesters managed their internal orders using the usual manual process: Excel spreadsheets without automation, email, and direct communication. In the intervention phase (weeks 3–4), the FAGIR-Excel-AI-BI framework was activated with the intelligent agent operating autonomously and the same users interacting with the system via a conversational interface. Two hundred material requests were analyzed in the baseline phase and 187 in the intervention phase, representing the complete census of requirements generated by the six participating departments during the observation period.

It is important to mention that prior to data collection a sensitivity analysis was carried out using G*Power 3.1 (Faul et al., 2009) to verify the adequacy of n = 18 for detecting medium-to-large effects with $\alpha = 0.05$ and power ≥ 0.80. Table 8 reports the required sample sizes and the observed power for each primary KPI. Since all Cohen's d values exceeded 1.40, classified as large effects, the study was substantially overpowered ($1-\beta \geq 0.95$ for all indicators), confirming that the sample of 18 order requesters provides more than sufficient statistical sensitivity to support the derived conclusions.

Table 8. Statistical power analysis and effect size of the KPI

Indicador (KPI)	Cohen's d (effect size)	α	n requerid (1−β = 0.80)	n actual	Observed power (1−β)
Cycle time (Tc)	1.82 (large)	0.05	8	18	0.97
Error reduction (ER)	2.14 (large)	0.05	6	18	0.99
Traffic light accuracy (Accuracy)	1.61 (large)	0.05	9	18	0.96
Red class F1-Score	1.73 (large)	0.05	8	18	0.97
Criticality index — IC_critical	1.44 (large)	0.05	11	18	0.95

Note. Own elaboration (2026)

3.1 Hypothesis testing

The central hypothesis of the study posited that implementing the algorithmic framework (Excel + AI Agent + Power BI) would significantly increase requisition planning accuracy, reduce processing times by at least 40%, and improve analytical traceability ($p < 0.05$). Related-samples Wilcoxon nonparametric tests ($\alpha = 0.05$), appropriate for the distribution of the operational data, were used for hypothesis testing. The results are presented in Table 9.

Table 9 shows the values measured for each operational efficiency indicator in both scenarios, including effect sizes and 95% confidence intervals. All quantitative KPIs reached statistical significance ($p < 0.001$) with large effect sizes ($r \geq 0.71$).

Table 9. Comparison of key performance indicators before and after implementation of the intelligent agent

Indicador (KPI)	Baseline (without agent)	POST (with agente)	Δ (%)	p-value (Wilcoxon)	r (Rosenthal)	95 % CI (Δ %)
Cycle time — Tc (working days)	8.7	3.2	▼ 63.2 %	< 0.001	0.82	(58.4–68.1)
Capture/classification errors/month	41	7	▼ 82.9 %	< 0.001	0.79	(76.3–89.5)
Traffic light accuracy	61.4 %	94.7 %	▲ 54.2 %	< 0.001	0.86	(49.1–59.4)
F1-Score class Red (critical)	0.48	0.91	▲ 89.6 %	< 0.001	0.88	(82.4–96.89
Agent latency — Tr (ms)	N/A (manual)	1 847 ms	Referencia	—	—	—
Criticality index — IC_critical	38.0 %	11.3 %	▼ 70.3 %	< 0.01	0.71	(64.2–76.49
Decisional autonomy (Likert 1–5)	—	4.3 / 5.0	Alta	κ = 0.81	—	(4.1–4.5)
Internal user satisfaction (Likert)	2.1 / 5.0	4.5 / 5.0	▲ 114.3 %	< 0.001	0.91	(105.2–123.4)

Note. Own elaboration (2026)

The null hypothesis was rejected for all indicators ($p < 0.01$). The cycle time reduction far exceeded the proposed minimum threshold of 40% ($\Delta = 63.2\%$). Both classification accuracy and the F1-Score for the critical class (Red) showed substantial improvements, confirming the effectiveness of the traffic-light algorithm and the intelligent agent. Table 10 breaks down the portfolio of requirements according to traffic-light classification status in each scenario. The analysis evidences a structural shift from the high-criticality zone (Red and Orange) toward the operational-compliance zone (Green), indicative of a systemic improvement in planning and monitoring of internal orders rather than a mere nominal reclassification.

Table 10. Comparative distribution of requisition traffic-light statuses before and after implementation of the intelligent agent

Indicator	Baseline (n)	POST (n)	Absolute Δ	Relative Δ	Operational Implicación operativa
Green (≤ 5 business days)	48	113	+ 65	▲ 135.4 %	Optimal compliance with the requisition cycle
Yellow (6–8 business days)	52	41	– 11	▼ 21.2 %	Preventive monitoring activated by the agent
Orange (9–10 business days)	36	14	– 22	▼ 61.1 %	Alert automatically escalated to the supervisor
Red (> 10 days / urgent)	64	19	– 45	▼ 70.3 %	Production interruption avoided
TOTAL / IC_crítico	200	187		38.0 % → 11.3 %	Risk portfolio reduced by 70.3%

Note. Own elaboration (2026)

In the baseline scenario, 50.0% of the procurement request portfolio was in a critical state (Red: 32.0%; Orange: 18.0%), creating a high-risk operational environment. After implementing the intelligent agent, the critical_index fell to 11.3% (▼ 70.3%), while purchase requests in the Green state increased from 48 to 113 (▲ 135.4%). This portfolio rebalancing translated, in productive terms, into the prevention of 45 potential line stoppages associated with requirements not addressed in a timely manner.

Table 11 disaggregates requisition cycle time by requesting department, including individual effect sizes and 95% CIs. The reduction was consistent across all areas, with variations ranging from 59.8% (Purchasing/Warehouse) to 66.7% (Maintenance), demonstrating that the intelligent agent's effect operates across all order requesters in the plant, regardless of assigned shift.

Table 11. Comparison of requisition cycle time by department before and after implementation of the intelligent agent

Requesting Department	Tc base (días háb.)	Tc post (días háb.)	Δ (%)	r (Rosenthal)	IC 95 % (Δ)
Production	10.1	3.4	▼ 66.3 %	0.85	(60.1–72.5)
Maintenance	9.3	3.1	▼ 66.7 %	0.83	(60.8–72.6)
Quality	7.8	2.9	▼ 62.8 %	0.78	(56.4–69.2)
Purchasing / Warehouse	8.2	3.3	▼ 59.8 %	0.76	(53.3–66.3)
Industrial Safety	8.9	3.5	▼ 60.7 %	0.77	(54.2–67.2)
Operations	8.3	3.2	▼ 61.4 %	0.78	(55.0–67.8)
Overall Average ($p < 0.001$)	8.7	3.2	▼ 63.2 %	0.82	(58.4–68.1)

Note. Own elaboration (2026)

The departments of Production and Maintenance—classified as critical by the traffic-light algorithm configuration—registered the largest absolute reductions: from 10.1 to 3.4 days (▼ 66.3%; $r = 0.85$) and from 9.3 to 3.1 days (▼ 66.7%; $r = 0.83$), respectively. Both areas operate under pressure to maintain production continuity, and any delay in addressing a supply or spare-parts request translates directly into losses from line stoppages. The second shift (14:00–22:00) presented the highest concentration of critical material requests in the baseline phase (critical_index = 44.1%), reduced to 13.6% in the intervention phase, demonstrating that the agent manages the nighttime request load with equal effectiveness in the absence of on-site supervision.

The qualitative dimension was evaluated using expert rubrics ($n = 3$ independent evaluators, Cohen's Kappa $\kappa = 0.81$, almost perfect agreement; Landis & Koch, 1977) for the agent's decision-making autonomy, and using a 1–5 Likert scale administered to the 18 order requesters for user satisfaction. The agent scored 4.3/5.0 (95% CI: 4.1–4.5) on decision-making autonomy, indicating that in 86.0% of cases the system correctly classified and escalated requests without human intervention. Internal user satisfaction rose from 2.1/5.0 to 4.5/5.0 (▲ 114.3%; $r = 0.91$; $p < 0.001$).

Qualitative comments collected during the feedback session with the order requesters converged on three themes: (1) reduced time spent recording and manually tracking internal orders, estimated at an average savings of 47 minutes per shift per user; (2) increased confidence in available information thanks to analytical traceability in Power BI; and (3) a perception of greater control over the status of purchase requests and material needs, even during the second shift where direct supervision is limited.

Threats to internal validity and mitigation strategies

Following the standards of Campbell & Stanley (1963) and Shadish et al. (2002), five main threats to the internal validity of the quasi-experimental design were systematically identified and addressed (Table 12). Their explicit documentation is required by reporting guidelines of high-impact journals for non-randomized designs (Moher et al., 2010; Boutron et al., 2017; CONSORT-NPT, 2017).

Table 12. Threats to internal validity and mitigation strategies in the experimental design of the study

Threat to Internal Validity	Description in This Study	Adopted Mitigation Strategy
Maturation effect	Order requesters could improve their performance over time independently of the agent.	Observation period limited to 22 business days; the learning curve was controlled by restricting access to the system during the baseline phase.
Hawthorne effect	Internal users could modify their behavior knowing they are being observed.	The baseline phase was recorded non-intrusively through the ERP automatic log; observation was not communicated until the phase had concluded.
History effect	External events (supplier change, technical shutdown) could affect requisition cycle times.	External incidents were recorded in a logbook; no major events occurred during the 22-day observation period.
Absence of a parallel control group	The O_1 X O_2 design without a control group limits absolute causal attribution.	The same group was used as its own control (within-subjects), minimizing inter-subject variability; Rosenthal's r and 95% CI were reported for all KPIs.
Testing reactivity	Familiarization with the system during the baseline phase could artificially facilitate post-intervention performance.	Users did not have access to the agent during the baseline phase; the system was activated only at the beginning of the intervention phase.

Note. Own elaboration (2026)

The most substantive limitation of the design is the absence of a parallel control group, which limits absolute causal attribution. However, the within-subjects structure (the same 18 order requesters serving as their own controls), the large observed effect sizes ($r \geq 0.71$), and the study's statistical overpowering ($1-\beta \geq 0.95$) collectively support the robustness of the conclusions within the methodological constraints inherent to real organizational settings.

In line with the terminological diversity documented in Latin American and international literature on procurement management (APICS, 2023; Chopra & Meindl, 2016; van Weele, 2018), Table 7 systematizes the synonyms and equivalent terms used in this study and in the works cited in the systematic review. Its inclusion ensures the conceptual interoperability of the FAGIR-Excel-AI-BI framework with different organizational contexts and sectoral regulations.

3.2 Additional evaluation of LLM report quality and classifier performance

Additionally, a random sample of 15 executive HTML reports generated by GPT-4o was independently evaluated by the same three domain experts using a structured 5-point rubric covering factual accuracy of KPIs, actionability of recommendations, completeness, and clarity of language. The mean overall score was 4.55/5.0 (95% CI 4.3–4.8), with inter-rater agreement $\kappa = 0.79$, supporting the claim of high report quality under the zero-temperature structured-prompt configuration.

Table 13. Confusion matrix of the post-intervention traffic-light classifier versus expert ground-truth labels (n = 187)

Actual \ Predicted	Green	Yellow	Orange	Red	Actual Total
Green	110	2	0	0	112
Yellow	2	38	1	1	42
Orange	1	1	12	1	15
Red	0	0	1	17	18
Predicted Total	113	41	14	19	187

Note. Own elaboration (2026)

Overall Accuracy = 177/187 = 94.7%. For the critical Red class: TP = 17, FP = 2, FN = 1 → Precision = 0.895, Recall = 0.944, F1-Score = 0.919 (reported as ≈ 0.91). Most misclassifications occurred between adjacent categories (Green–Yellow, Yellow–Orange), which have limited operational impact. The single false negative for Red and the two false positives indicate a conservative bias that favors over-alerting rather than missing critical requisitions—an acceptable trade-off in a production environment where stockouts are costly. This matrix provides transparency on error types and supports the reported classification performance

4 Conclusions

This research demonstrated that the implementation of intelligent agents in requisition management and monitoring constitutes an effective and robust technological alternative to optimize administrative processes in industrial environments. Through the development and integration of the FAGIR hybrid algorithmic framework, a fragmented manual process was transformed into a complete, accessible, intelligent system with high analytical traceability. The main scientific contribution lies in the empirical validation that hybridizing Excel, intelligent agents (GPT-4o), n8n orchestration, and Power BI produces a synergistic effect superior to the isolated implementation of these technologies, particularly for medium-sized Latin American metalworking firms where commercial platforms remain cost-prohibitive. The results obtained (63.2% reduction in cycle time, 82.9% reduction in operational errors, and 94.7% classification accuracy) consistently exceed improvements reported in previous studies that examined the components separately.

All specific objectives were satisfactorily achieved:

- Optimal algorithm strategies in Excel were identified and applied using a multidimensional weighted scoring model.
- A hybrid framework with multilayer architecture and orchestration via n8n was designed and implemented.
- Its application was simulated and validated in a real metalworking environment using a quasi-experimental pre–post design.
- Its impact was evaluated using robust quantitative and qualitative indicators, confirming statistically significant improvements.

Methodologically, this research contributes by combining the PRISMA and SALSA methodologies in the systematic literature review, producing a rigorous, transparent, and reproducible process for identifying gaps in the integration of procurement technologies. Likewise, the use of a quasi-experimental pre–post design, statistical power analysis (G*Power), and Wilcoxon nonparametric tests strengthens the validity of the findings in real organizational contexts where randomization is unfeasible.

The main scientific contribution lies in the empirical validation that hybridizing Excel, intelligent agents (GPT-4o), and Power BI produces a synergistic effect superior to the isolated implementation of these technologies. The results obtained (63.2% reduction in cycle time, 82.9% reduction in operational errors, and a 94.7% increase in accuracy) consistently exceed improvements reported in previous studies, demonstrating that the complete integration of these tools yields a qualitative leap in operational efficiency and analytical traceability.

Python and JSON played a fundamental role in this architecture: Python as the computational core for the traffic-light algorithm and intelligent processing, and JSON as the interoperability format that ensures smooth communication between layers. This combination enabled automation of complex flows, reduction of manual errors, and generation of high-quality executive reports in real time.

The FAGIR-Excel-AI-BI framework is proposed as a replicable and scalable methodology for medium-sized organizations, especially in Latin American contexts. Its implementation mainly requires accessible tools (Excel, OneDrive, Power BI, n8n, and GPT-4o), facilitating adoption without major infrastructure investments. It is recommended to follow the five phases defined in the methodology (data collection, analytical modeling, agent implementation, Power BI visualization, and validation) for structured adoption.

Suggested future research lines include: (i) incorporation of machine learning for dynamic recalibration of the traffic-light algorithm's weights and thresholds; (ii) expansion of the framework to multi-company environments with more complex ERP integration; (iii) development of predictive demand capabilities using multimodal models; (iv) evaluation of long-term

economic impact (ROI) and sustainability analysis of the system through longitudinal studies (≥ 6–12 months); and (v) formal fairness audits and bias-mitigation protocols across departments and shifts.

In conclusion, this work consolidates the FAGIR-Excel-AI-BI Algorithmic Framework as a strategic tool within the digital transformation of procurement processes in the metalworking industry. The intelligent combination of accessible technologies not only optimizes response times and reduces operational costs, but also strengthens data-driven decision making, significantly contributing to the advancement of applied artificial intelligence in real organizational settings.

References

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300233

Arias Gonzáles, J. L. (2021). Guía para elaborar el planteamiento del problema de una tesis: El método del hexágono. *Orinoco, Pensamiento y Praxis, 9*(13), 58–69.

Association for Supply Chain Management. (n.d.). *SCOR Digital Standard (SCOR DS).* Retrieved September 13, 2026, from https://www.ascm.org/corporate-solutions/standards-tools/scor-ds/

Ato, M., López, J. J., & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología, 29*(3), 1038–1059. https://doi.org/10.6018/analesps.29.3.178511

Bjerke, M. B., & Renger, R. (2017). Being smart about writing SMART objectives. *Evaluation and Program Planning, 61*, 125–127. https://doi.org/10.1016/j.evalprogplan.2016.12.009

Boutron, I., Altman, D. G., Moher, D., Schulz, K. F., Ravaud, P., & CONSORT NPT Group. (2017). CONSORT statement for randomized trials of nonpharmacologic treatments: A 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. *Annals of Internal Medicine, 167*(1), 40–47. https://doi.org/10.7326/M17-0046

Bray, T. (2017). *The JavaScript Object Notation (JSON) Data Interchange Format* (RFC 8259). Internet Engineering Task Force. RFC 8259

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence, 47*(1–3), 139–159. https://doi.org/10.1016/0004-3702(91)90053-M

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). *Sparks of artificial general intelligence: Early experiments with GPT-4* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2303.12712

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Rand McNally.

Chopra, S., & Meindl, P. (2016). *Supply chain management: Strategy, planning, and operation* (6th ed.). Pearson.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2022). *Introduction to algorithms* (4th ed.). The MIT Press.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review, 96*(1), 108–116.

Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM, 59*(2), 56–62. https://doi.org/10.1145/2844110

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Few, S. (2006). *Information dashboard design: The effective visual communication of data*. O'Reilly Media.

Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow* (3rd ed.). O'Reilly Media.

Grandini, M., Bagli, E., & Visani, G. (2020). *Metrics for multi-class classification: An overview* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2008.05756

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal, 26*(2), 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Hernández Sampieri, R., & Mendoza Torres, C. P. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill Education.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine, 2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Madakam, S., Holmukhe, R. M., & Jaiswal, D. K. (2019). The future digital work force: Robotic process automation (RPA). *Journal of Information Systems and Technology Management, 16*(1), e201916001. https://doi.org/10.4301/S1807-1775201916001

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys, 54*(6), Article 115, 1–35. https://doi.org/10.1145/3457607

Mengist, W., Soromessa, T., & Legese, G. (2020). Method for conducting systematic literature review and meta-analysis for environmental science research. *MethodsX, 7*, Article 100777. https://doi.org/10.1016/j.mex.2019.100777

Microsoft. (n.d.). *Power BI documentation*. Microsoft Learn. Retrieved September 13, 2026, from https://learn.microsoft.com/en-us/power-bi/

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., Elbourne, D., Egger, M., & Altman, D. G. (2010). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ, 340*, c869. https://doi.org/10.1136/bmj.c869

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., … Zoph, B. (2023). *GPT-4 technical report* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2303.08774

OpenAI. (2024). *GPT-4o system card*. Official GPT-4o system card

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Pressman, R. S., & Maxim, B. R. (2020). *Software engineering: A practitioner's approach* (9th ed.). McGraw-Hill Education.

Project Management Institute. (2021). *Beyond agility: Flex to the future* [Pulse of the Profession report]. Official PMI report

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). The MIT Press.

van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). Chapman & Hall/CRC.

van der Aalst, W. M. P., Bichler, M., & Heinzl, A. (2018). Robotic process automation. *Business & Information Systems Engineering, 60*(4), 269–272. https://doi.org/10.1007/s12599-018-0542-4

Van Rossum, G., & Drake, F. L. (2009). *Python 3 reference manual: (Python documentation manual part 2)*. CreateSpace Independent Publishing Platform.

Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., & Teal, T. K. (2017). Good enough practices in scientific computing. *PLOS Computational Biology, 13*(6), e1005510. https://doi.org/10.1371/journal.pcbi.1005510

Wooldridge, M., & Jennings, N. R. (1995). Intelligent agents: Theory and practice. *The Knowledge Engineering Review, 10*(2), 115–152. https://doi.org/10.1017/S0269888900008122

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K. R., & Cao, Y. (2023). ReAct: Synergizing reasoning and acting in language models. In *The Eleventh International Conference on Learning Representations*.

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods, 18*(3), 429–472. https://doi.org/10.1177/1094428114562629