# A Hybrid AI Framework for Automating High-Stakes Procurement Workflows


## Abstract

**Problem Statement:** The automation of complex, human-centric workflows remains a critical challenge in enterprise AI. The manual *Purchase Requisition (PR)* approval process exemplifies this, presenting a significant operational bottleneck where each review consumes between 15 minutes to 2 hours. This process is a fundamental pillar of financial governance. As the primary mechanism for initiating the procurement cycle , each PR serves as a critical control point to ensure that proposed expenditures align with internal budgets , adhere to accounting policies by validating fields like GL codes and cost centers , and comply with vendor and VAT requirements. This manual process leads to considerable delays and inefficiencies. The core limitation is the absence of an integrated system capable of programmatically performing over a dozen distinct checks, ranging from budget alignment to content validation within attachments. To address this, we present an automated **PR Review Assistant**. Our primary contributions are: 1) a novel multi-agent architecture featuring specialized agents for each stage of the workflow, including semantic data interpretation, contextual validation, and the generation of **explainable, natural-language outputs**; and 2) a comprehensive real-world benchmark of leading LLMs on a financially critical task.

**Proposed Solution:** Our framework operates as a sophisticated multi-agent system that orchestrates the end-to-end review through three specialized LLM agents in a cascading workflow. The process begins with the ***LLM-EXTRACT*** agent, which ingests raw PR attachments and performs advanced semantic entity recognition to transform unstructured content into a standardized, machine-readable format. This structured data is then passed to the ***LLM-VAL*** agent, the core of our system's intelligence, which performs nuanced contextual validation and cross-references information against complex business logic to identify subtle anomalies that deterministic rules would miss. Finally, all findings are synthesized by the ***LLM-REPORT-GEN*** agent, which translates the technical outputs into a clear, natural-language summary, providing actionable rationale for any failures. This modular, orchestrated process creates a highly efficient and auditable workflow that separates interpretation, validation, and communication into discrete, optimized tasks, resulting in a system that is simultaneously intelligent, trustworthy, and transparent. The final validation status is governed by a **Hybrid Confidence Score ($H_s$)**, which models this cascading logic:

$$H_s(\text{PR}) = \left( \sum_{i=1}^{n} c_{di} \right) + C_L(\text{PR})$$

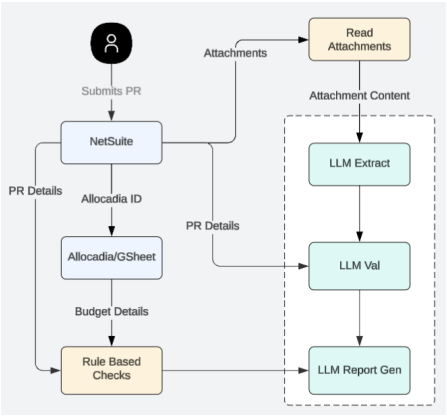
where deterministic checks ($c_{di} \in \{0, 1\}$) act as a gate for the LLM's confidence score ($C_L$).


Table 1: Methodology and Performance on Test Set of 100 PRs

| Model | Extraction Accuracy | Mapping Precision |
|---|---|---|
| **gpt-4o** | 96.2% | 94.8% |
| **gpt-oss-120b** | 93.7% | 91.2% |
| **llama3-1-405b-instruct** | 91.1% | 93.5% |
| **mistral-large** | 91.5% | 90.5% |
| **claude-3-5-sonnet** | 90.8% | 88.2% |

Since $c_{di}$ are constant for all LLMs, $H_s$ is a direct function of $C_L$. Accordingly, **gpt-4o** achieves the highest $H_s$ due to its superior extraction accuracy and mapping precision.

Figure 1: Validation Pipeline




**Experimental Results:** Now in **production** for our Marketing and G&A departments, the assistant processes over 4,000 PRs annually. The benchmark revealed distinct performance characteristics, with `gpt-4o` showing the highest consistency. Implementation yielded transformative business impact, driving a **62% reduction in manual review time** and a **37% reduction in the total end-to-end procurement cycle time** (from 14.7 to 9.3 days). Furthermore, the system achieved a **66% reduction in initial submission error rates** by providing immediate, actionable feedback. User feedback has been overwhelmingly positive, citing faster approvals and improved compliance. By automating routine checks, the framework allows finance and procurement teams to focus on higher-value strategic assessment, establishing a generalizable model for human-in-the-loop automation in other corporate functions.