Final Dissertation.

Student Name:	Mohamed Eid
Student Number:	F417057
Programme Title:	Artificial Intelligence and Data Analytics
Dissertation Title:	How can an AI-powered system, built on Retrieval-Augmented Generation (RAG) framework, be applied to organizational operating models such that it is able to detect and optimize process inefficiencies?
Word Count:	10,287
Academic Supervisor:	Thanuja Mallikarachchi
Dissertation route:	Traditional Dissertation

Declaration

By making this submission I confirm that the attached coursework is my own work and that anything taken from or based upon the work of others – or previous work of mine – has its source clearly and explicitly cited; I understand that failure to do so may constitute Academic Misconduct.

I have read the Student Dissertation Handbook and understand the sections on 'Intended Learning Outcomes', 'Academic Integrity', and 'Marking Criteria.'

I have read the Appendix 'Guidance to Students on how to acknowledge, describe and reference the use of Generative AI tools in assessed work' in the Dissertation Handbook. I have acknowledged the use of any Generative AI tools in my dissertation in accordance with this document and understand that a failure to do so will be regarded as academic misconduct by the University.

Tick this box if you do not wish Loughborough University to use your dissertation for administrative, promotional, educational and/or teaching purposes. ☒

Signature: Mohamed Eid

Date: 16/09/2025

Abstract

Efficient operational retention is a struggle faced by organizations of all types and sizes, with 50-70% failure rates demonstrated by the traditional optimization approach, across different industries. This research aims to fixate on applying artificial intelligence to business process optimization, to ratify a new Retrieval-Augmented Generation (RAG) framework which addresses the drawbacks of the existing methodologies resulting from data scarcity and expertise dependency.

A two-phased evaluation approach was applied to this study. The First Phase set a baseline performance by evaluating four different foundation models (Claude Opus 4, Claude Sonnet 4, DeepSeek R1, Llama 3.3 70B) on 231 workflow patterns from the BPI Challenge 2020 dataset. The Second Phase authenticates the implementation of a RAG-enhanced system with proficient validation from Loughborough University London domain specialists.

Results from foundational model evaluation showed convergent classification accuracy of 64-65%, with an 0.4% spread indicating RAG enhancement holds a greater value than model selection. Claude Opus 4 was selected due to its exceptional logical reasoning capabilities, and structured output consistency.

The RAG system combines a knowledge base of 23 different industry and academic documents, with 231 empirically-labeled process patterns, enabling recommendations for evidence-based optimization. Workflow descriptions are transformed into a six-stage processing structured analysis: input processing, evidence retrieval, compliance analysis, prompt generation, AI analysis, and output generation.

Results from implementation showed comprehensive optimization analysis including quantified compliance scores, detailed bottleneck identification, evidence-based recommendations, and visual workflow comparisons. The web interface enables access to insights through five different dashboard components allowing practical deployment.

The practical application for preliminary process assessment was validated by experts, acknowledging the traditional limitations of consulting through cost-effective baseline analysis, all while maintaining the human expertise requirements for implementation decisions.

With the demonstration of successful RAG application to optimizing workflows, the research contributes to business process management, establishing AI-enhanced analysis methodologies, and providing empirical evidence for knowledge-augmented language models in business applications. Research findings showed compelling scaling potential from traditional process-focused optimization to organizational transformation.

Future research encompasses expanding knowledge bases for process domains, improving automated organizational monitoring, and investigating deployment frameworks of enterprises. The study sets the foundation for the next generation of process optimization, through AI-augmented tools with an endorsed practical value.

Keywords: Business Process Management, Retrieval-Augmented Generation, Artificial Intelligence, Workflow Optimization, Large Language Models, Process Mining, Organizational Efficiency

Table of Contents

1. Introduction

1.1 Background

Organizations are constantly struggling to maintain efficiency alongside the rapidly changing business environments, and such struggles have only become more difficult to navigate as times progress. In today's ever changing business settings, efficiency is no longer an edge, rather a necessity. Trying to uphold the standard quality while still adapting inefficient processes tends to cost said businesses an extensive amount of labour, resources and time beyond what is required to do so (Harvard Business Review, 2023). Bottlenecks are an almost immediate result of such an approach, where inefficiencies contribute to significant hold-ups in a business workflow that are easily avoidable. These repercussions are particularly evident in travel and accommodation workflows, where manual processes and layers of approvals have a direct effect on multiple stakeholders. In such cases, all levels of a hierarchy tend to notice the inefficiency, with both workers and CEOs complaining that around 40% of time and effort spent on administrative tasks is unproductive. When poor communication comes into play, competitive disadvantages take effect and subsequently create poor customer relationships and increased business costs (PwC, 2024). Only 7% of organizations have proven effective in all elements of operational excellence, and under 25% are adapting more modern approaches to critical tasks such as workload balancing and resource prioritization through visual tools (McKinsey & Company, 2024).

The financial hits these companies take as a result of operational inefficiencies are substantial. Domestic and international flights require a manual approval rate of 27% and 34% respectively (McKinsey & Company, 2024), and that a mere 56% of travellers make advantage of managed booking channels and end up being repetitive customers of these channels (Deloitte, 2024). Poorly integrated workflows are the perfect depiction of the monetary losses some of these organizations can end up facing, where some companies have gone up to as high as 20-30% of their annual revenue in loss, as a natural outcome of processes inefficiencies, studies have shown. The administrative overhead can increase the cost of a single expense report from $20.65 (ExpensePoint, 2025) to a drastically higher $58 per report (Extend, 2025). The lack of streamlined processes prevents clear visibility of actual versus budgeted expenditures, as well as efficient overall spending management during times of travel (Deloitte, 2024). There are also cases where substantial financial risk directly correlates with manual processes, such as with cases of expense reimbursement fraud. In their Occupational Fraud 2024: A Report to the Nations the Association of Certified Fraud Examiners (ACFE) claims that this category alone represents 13% of all occupational fraud cases globally and has a median loss of $50,000(American Express, 2025).

1.2 Problem Statement

Business process optimization is a field that is well established with myriads of approaches to follow, those of which that are conventional have shown their limitations over time. These limitations have resulted in failure rates that are as high as 50-70% across different industries such as business process re-engineering (BPR), Six Sigma implementations, and management consulting (BCG, 2017; Harvard Business Review, 1993; Emerald Insight, 2015). Due to the nature of these businesses, interventions to improve the current state of organizational processes tend to be more discrete and project based, eventually failing to remain in place past the initial rollout phase, and keep up with changing circumstances across different departments.

This critical gap is what the research aims to address through shifting the focus towards a core question: How can an AI-powered system, built on Retrieval-Augmented Generation (RAG) framework, be applied to organizational operating models such that it is able to detect and optimize process inefficiencies? The answer to this question will encompass an analysis of available data supported by evidence from the literature that is meant to serve as a more efficient successor of methods that have been rendered obsolete or ineffective.

1.3 Research Objectives

The primary objectives of this research are to:

1. **Design and validate an AI-driven framework for evidence-based analysis of business processes.** The Retrieval-Augmented Generation (RAG) system that has been adapted for this framework will be responsible for catering to the challenge of insufficient labelled data through contextualizing Large Language Model (LLM) outputs in a curated knowledge base of established business frameworks.

2. **Perform a systematic empirical analysis of foundational models to be able to compare workflow efficiency and perform a classification.** The study will evaluate the capabilities of multiple LLMS in performing binary classification tasks, determining their effectiveness in predicting workflow efficiency without using the RAG model, through a point of comparison, and subsequently assessing how integration of the RAG influences the quality of the initial recommendations.

3. **Illustrate the scalability and practical deployment of the framework in question via cloud-based infrastructure.** Using services like AWS Bedrock, the research will showcase the usage of different LLMs in terms of usage, deployment and assessment within enterprise contexts, establishing an approach that can be tailored for large-scale organizational adoption.

4. **Validate real-world application possibilities through expert appraisal done by university domain** specialists**.** The framework will be tested on Loughborough University London's travel and accommodation workflows, with experts in the domain providing feedback on accuracy, relevance and quality of the generated recommendations.

1.4 Research Contribution

This thesis aims to be an improvement in the field of business process management, rather than an oversaturation, through the introduction of an innovative, AI-driven methodology, designed to tackle challenges that were previously rendered stubborn in the efficiency of business workflows. The practicality of applying a Retrieval-Augmented Generation (RAG) framework to business process optimization is an unexplored approach, increasing the significance of this research. The methodology also tackles the issue of data scarcity, one that is common in this business domain, by grounding the LLM outputs in the manner that was mentioned as part of the main research goals - an extensive knowledge base of best practices and business frameworks that have already been deemed efficient through the utilization of the RAG system, thereby ensuring the recommendations are essentially provable, while still contextually robust.

Furthermore, the comparative analysis of LLM performance highlights the added value of the RAG enhancements and emphasizes the end-to-end improvement of the quality of the generated recommendations. This, in turn, showcases the potential this solution has of thriving in a real-world setting, through a scalable, cloud-based system capable of adapting to tailored organizational priorities. The data involved also serves as an appropriate test case that when validated by the aforementioned experts, further proves the system's effectiveness.

2. Literature Review

2.1 The Evolution of Business Process Management (BPM)

As the integration of artificial intelligence into almost everything drastically increases, alongside the increasing need for organizational agility, Business Process Management (BPM) has undergone significant transformations (Springer, 2024). The main focus of BPM has been on cutting costs since the introduction of the discipline, but it has now taken a shift to encompass agility, innovation and resilience. This shift is one that takes the discipline away towards AI-enabled practices that supports organizational adaptation in real-time, from what was once mere static documentation of business process requirements and improvements (Springer, 2024). The traditional BPM lifecycle, illustrated in Figure 1, outlines the cyclical phases of process identification, discovery, analysis, redesign, implementation, and monitoring, providing a foundational framework that continues to evolve with these new AI-driven capabilities. This is an essential requirement if BPM is to survive the competitiveness of current business process optimizations.

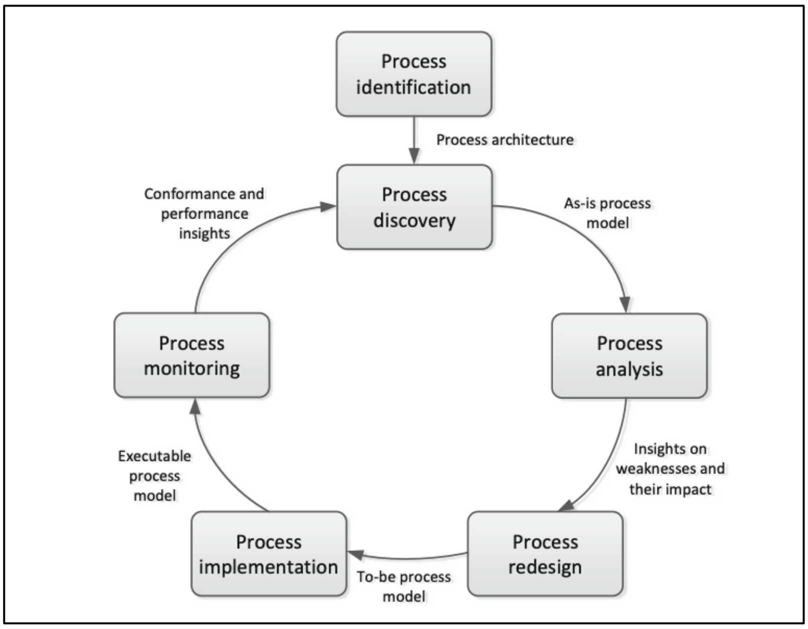

Figure 1: Business Process Management Life Cycle (adapted from Dumas et al., 2013).

2.1.1 The Integration of AI, Agile, and Lean Methodologies

In today's dynamic business environment, organizations are constantly reevaluating their operating models to achieve higher levels of flexibility and responsiveness, leading to the adoption of Agile and Lean methodologies. Both of these are being continuously integrated with artificial intelligence in the production of data driven systems guided by adaptability as their main objective.

The primary focus of agile methodology is iterative development and rapid responses to changing conditions. The approach requires great levels of flexibility across collaborative, cross-functional teams, in order to be able to centre decision making around customer needs over rigid adherence to existing processes (Dikert et al., 2016; Conforto et al., 2014). While traditional BPM assumes an approach built in blocks of sequential processes, Agile BPM follows more case-specific approaches where continuous improvement and rapid adjustments are the key pillars, making them essentially different, and sometimes even opposing one another (Springer, 2024). Return on investment (ROI) rates for companies who follow the latter are leading with up to 2.1x higher than those who stick to traditional BPM, as it gives them a more focused view of calculated use cases rather than broad automation initiatives that are unspecific to their organizational models. This allows such businesses to adapt within timeframes as short as days rather than months or years (BCG, 2025).

Recent studies have shown a synergy between AI and Agile methodologies across fields, as with Almalki (2025), where it has become evident that AI systems built for decision support have the potential to enhance workload management by 25% when put into use within an agile framework, enabling them to identify risks with an accuracy of 94%. This type of integration also speeds up sprint completion by 18% and reduces defect solution time by 35%, saving significant amounts of time for the course of business use cases. Moreover, research has shown that AI has been labelled a top three ranked priority by 75% of business leaders in the market for 2025, indicating that Agile BPM has been recognized as an effective tool to be made advantage of within AI augmented environments, especially where organizational environments shift towards the complex (BCG, 2025).

A close alternative is Lean methodology, a result of the Toyota Production System. It's concerned with waste management, operational streamlining and continuous improvement in the straightforward direction of maximizing customer value through resources benchmarked as only completely essential and nothing more (Tashkinov, 2024).

What stemmed from a place of standardization of manufacturing practices, has now been implemented across the most diverse industries, especially after its integration with AI. Its use alongside the technology has been proven to yield tangible business outcomes, with example use cases in industrial processing plant operators reporting 10-15% increases in production rates and 4-5% increases in Earnings Before Interest, Taxes, Depreciation, and Amortization (EBITA) through the adoption of an AI solution that is essentially Lean-based and more modern team structures (McKinsey & Company, 2023). The ROI data that is delivered by this research is particularly useful as a concrete basis of AI-driven process optimization initiatives in industrial settings (McKinsey & Company, 2023). The integration of this methodology framework with AI allows for a proactive approach to business processes that puts waste levels to a minimum and turns such a standard as a continuous, real-time functionality within modern organizational workflows, while also maintaining the required levels of precision (Davenport et al., 2023).

2.2 Critical Analysis of Traditional and Contemporary Solutions

Businesses have historically relied on a variety of approaches when it comes to streamlining existing processes, from traditional consulting approaches to emerging AI-driven solutions. Nonetheless, systematic analysis of these practices has made it even more evident that failure rates remain high, and some limitations are yet to be bypassed by the adopted approaches.

2.2.1 Management Consulting Approaches

Mckinsey, with its place in the market as one of the leading companies in the industry, has managed to develop a number of intricately effective frameworks meant to accelerate organizational improvement (McKinsey & Company, 2024). Its "Organize to Value" system is a twelve-element approach flexibly built around pillars such as purpose, talent and leadership. The beauty in the said system is that it can be tailored to the specific needs of an organization, in terms of goals, circumstances, and more. As the system requires existing operating models to both predict changes and immediately react to them, it revolutionizes preexisting frameworks that tend to follow a more general approach, using what they know has always worked (McKinsey & Company, 2024). A financial services firm that has significant presence in the global market was able to use the framework to adopt a new operating model, resulting in an extensive transformation of a whole 11 parts of their system, and a subsequent 40% increase in their speed to market. The approach also yielded a 25% increase in their technology ROI (McKinsey & Company, 2024). The unfortunate side to such an approach, is that external expertise is the key factor, with costly approaches being brought into practice by management consulting companies such as BCG (BCG, 2025). Furthermore, change programs are not a guaranteed success, with research showing that 50% of said programs tend to fail to meet objectives, and more complex approaches have even higher failure rates of 75% (BCG, 2017). While some of the aforementioned approaches have been proven successful, the overall outcome of management consulting usage entails that over 67% of large-scale technology programs fail to deliver within the planned scope, budget, or time. This is often a result of costly, project-based engagements as opposed to deep organizational understanding (BCG, 2024). This, in turn, shifts focus away from the underlying causes of inefficiencies which the business is left to resolve on its own after external contractors have served their time on the projects, leading to other spending streams down the line (BCG, 2025). The first thing the proposed model aims to address is the sustainability of its presence as a built-in solution that aims to resolve the dependency on external, costly intervention.

2.2.2 Business Process Re-engineering (BPR)

Business process engineering seeks to "do better things" as opposed to just "doing things better". Shorter wait times in the healthcare industry has been a direct outcome of BPR usage in addition to better patient experiences through the redesign of patient flows. Reimagining core service processes in financial institutions has also proven effective, with the introduction of AI-based customer support solutions through BPR (SixSigma.us, 2024). Operational efficiency can be significantly improved through BPR, with rates of productivity increasing by 20-30%, as shown by use cases in the manufacturing sector (Six Sigma Institute, 2025). However, its failure rate has been documented to be as high as 50-70% (SciELO, 2021), the cause of which is said to be the methodological shortcomings, where it starts reevaluating the business operational workflows from a "clean slate fallacy" point of view. This strategy ignores embedded capabilities and knowledge by assuming that organizations can start from scratch, with 70% of failures attributed to human factors such as institutional memory (Al-Omran et al., 2019).

2.2.3 Six Sigma and Lean Methodologies

Data driven analysis is the key method used by Six Sigma to streamline end to end processes where possible, through using the results to ensure process consistency and reduce defects (Six Sigma Online, 2025). Six Sigma is the adopted methodology for business process re-engineering at J.P. Morgan Chase (iSixSigma, 2025). However, the methodological and contextual limitations of Lean and Six Sigma approaches lead to failure rates of higher than 60%. The service industry provides a wide set of use cases on the matter, where the organizational workflows are often people-intensive and not as systematic as with manufacturing business processes, rendering statistical approaches inefficient. Moreover, the time and budgeting constraints behind getting new employees trained and certified (ranging from $300 to $15,000 per employee) in addition to the actual time it takes for the implementation to take place, commonly 18-24 months, create significant barriers for many organizations (Emerald Insight, 2015).

2.2.4 Process Mining Techniques

Process mining extracts digital traces from preexisting transactions and uses the derived information to showcase, analyse and improve business processes (Celonis, 2024). The generated output acts as a vivid depiction of actual organizational behaviour and works efficiently in respect to showing the discrepancies between the aforementioned and theoretical work that is preplanned and often ends up mismatching circumstances (Celonis, 2024). Organizational cost cutting is often linked with the usage of process mining, where a single company, such as Deutsche Telekom, can save up to €66 million through the optimization of a single business workflow (Procure-to-Pay processes). PepsiCo also attributes some of its successes to process mining, where they have reduced rejected sales orders by a generous 86%, alongside GE Healthcare, who have managed to increase free cash flow to $1.3 billion using the same technology (Celonis, 2024).

Nonetheless, this method does not come without shortcomings, where process mining has been proven to face systematic challenges categorized under scalability and data contextualization and usability. 90% of organizations have rendered their desired business outcomes unreachable through these initiatives alone by 2026, with the leading cause being business process management maturity, as per the research (Gartner, 2024). Data quality in this approach is a pressing challenge, with an end result of most of the efforts being spent on finding data, determining what is considered relevant and transforming it into the output, only so much effort is left for the actual analysis (Celonis, 2024). Even when this entire approach is deemed successful in terms of implementation, there are times where the model can end up being highly intricate, to the point where business users struggle to use them effectively and understand their strong suits (McKinsey & Company, 2024).

2.3 The Architectural Gap

The methodologies covered in the literature provide accumulating evidence of a fundamental architectural limitation, provided by the analysis of approaches in management consulting, BPR, Six Sigma and process mining. It is worth noting that while existing solutions are not short of their effectiveness where relevant, but are rather discrete and use case specific, and are rendered inefficient under the conditions of the dynamic nature of modern organizational workflows (McKinsey & Company, 2024). The optimization lifecycle is a factor that often fails to be taken into account, and current methods end up failing to continuously adapt or continue to be maintainable in the long run and across departments and industries. As per the research, failure rates are as high as 50-70%, a number that speaks for itself in saying that further steps need to be taken (BCG, 2017; Harvard Business Review, 1993; Emerald Insight, 2015). It is quite evident that the approaches in question are unable to grow under the umbrella of modern organizational complexity. This points the direction of the preexisting literature, and this one moving forward, to a gap that needs to be bridged by a new architectural paradigm - one that can provide a more persistent solution that is both scalable and evidence-based.

This research in particular questions the maturity of Artificial Intelligence (AI), more specifically Large Language Models (LLMs) and the extent to which it can be used to overcome the aforementioned challenges. LLMs are now exhibiting substantial progress in advanced reasoning and analysis, as well as the ability to process unstructured data on a larger scale. Open-source models have become increasingly sophisticated and widely available, and with models like DeepSeek opening the door for practical organizational use without the worry of incurring additional costs or legislative limitations, businesses are less worried and more welcoming of the potential of using LLMs to streamline operations and enhance decision making.

This is further emphasized by how companies might be intimidated by the idea of building customized models from scratch. Such approaches are often associated with high costs and longer scopes and are often deemed unnecessary for a wide range of business process scales, especially with the threat of data scarcity hanging over business' owners' heads in this domain. Upon understanding this, this research brings into light a pragmatic approach, where it proposes and validates a framework that takes advantage of existing models and enhances their capabilities through concatenation. This methodology can be implemented over a short course of time and provides sufficient foundational work for organizations to base their further business model optimization on, easily adding their own internal documentation, thereby allowing the recommendations made to be tailored to their own business processes and maintaining their own compliance requirements.

3. Methodology

3.1 Research Design and Approach

The methodology behind the system in place is composed of two phases that complement one another to provide an analysis of the efficiency increases provided by the use of Retrieval-Augmented Generation (RAG) systems for business process optimization, consisting essentially of foundational model comparison, followed by university expert validation. This approach entails that the criticalness of the standalone capabilities of the foundational model in place is highlighted, while still separately emphasizing the enhancement in performance provided by the RAG integration as experts deem the different steps of the outputs as ones that progressively improve across the phases.

Phase 1: Foundation Model Selection and Comparative Analysis

The first phase aims to perform the aforementioned comparative analysis of the different foundation models to determine which is the most suitable for RAG integration. Since all four models in review are all considered state of the art, establishing baseline performance requires a more detailed assessment of their capabilities in classifying workflow efficiency. The models Claude Opus 4, DeepSeek R1, and Llama 3.3 70B, are all used with patterns derived from the BPI Challenge 2020 dataset. Each model is required to mark a workflow pattern as “efficient” or “inefficient” as per a binary classification framework, which benchmarks its predictions on algorithmically derived ground truth labels, providing a systematic comparison of each of the models’ abilities, while providing quantitative performance metrics for easier selection. The analysis is comprised of data related to 33,000+ process instances from real world implementations dispersed across 231 unique workflow patterns, which renders it statistically reliable.

Phase 2: RAG Enhancement Validation with University Expert Assessment

The subsequent phase entails a collaboration with domain experts from Loughborough University London to examine the practical effectiveness of integration with the RAG to allow for enhanced recommendations as output. The specialists’ knowledge on processes in accommodation requests and travel authorization provides a suitable evaluation tool for academic business process optimization systems.

In this stage, the university experts evaluate the RAG-enhanced recommendations, as well as the baseline performance for the models, allowing for qualitative insights to come into play as well. This provides groundwork for a comparison in terms of practical applicability, implementation feasibility and improvement quality. The role of the experts here also expands to include the assessment of real-world utility, as opposed to a mere audit of algorithmic performance metrics. This ensures that the findings of this research are useful beyond technical benchmarks and can be turned into actionable business outcomes.

The two-part methodology delivers a comprehensive view of the evaluation process, where the first phase establishes proper foundation through intricate analysis of quantitative metrics, and the latter validates the qualitative aspects of RAG enhancement through expert opinions.

3.2 System Architecture and Design

3.2.1 Overall System Architecture with RAG-Enhanced Processing Pipeline

The workflow optimization framework is designed on a robust foundation and aims at tackling the existing limitations presented by the literature through the transformation of raw business process descriptions into academically calculated recommendations, building upon RAG principles and performing rigorous context retrieval.

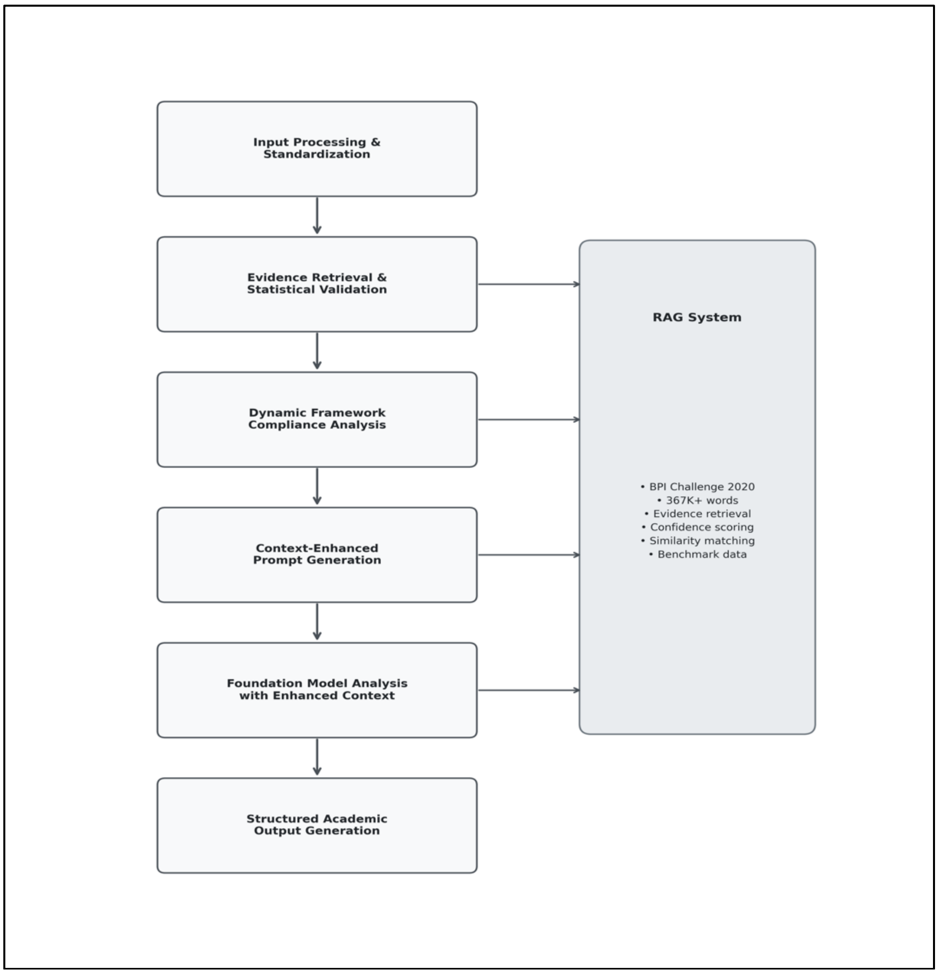

Figure 2: RAG-Enhanced Business Process Optimization Framework Data Processing Pipeline (Source: Author's own work)

The submitted business process descriptions are used as the baseline for the generation of recommendations in accordance with verified evidence. This is done through a streamlined user experience that generously takes into account intelligent context enhancements. As shown in Figure 2, the data processing pipeline is organized into six sequential stages, each of which act as a building block towards arriving at the outcome. The AI-driven framework includes compliance scoring to ensure the system is academically informed as well as practically applicable.

Stage 1: Input Processing and Standardization

- **Input Processing:** After the submission of the business workflow descriptions into the web interface, the framework parses the natural language workflow descriptions through the use of Claude Opus 4.
- **Component Standardization:** The model generates a structured representation of the data, compatible with the BPI dataset, meanwhile extracting the required features of the workflow in analysis.
- **Adaptive Data Handling:** The system does not stop at academic analysis, and entails the use of a mechanism for handling incomplete submissions through constraint identification.

Stage 2: Evidence Retrieval and Statistical Validation

- **Empirical Foundation:** TF-IDF vectorization methodology matches user workflows against the BPI Challenge repository.

- **Confidence Scoring Framework:** Implementation of confidence threshold scoring for the recommendations to meet academic validation requirements, which is rendered through a cosine similarity analysis against the BPI Challenge dataset.

- **Benchmark Assembly:** Combines user workflow data with standard patterns and optimization outcomes from the documentation, and compares it with preexisting statistical performance ranges and similar successful implementations for classification purposes.

Stage 3: Dynamic Framework Compliance Analysis

Performs a systematic evaluation of the identified workflow characteristics in order to be able to assess Agile and Lean methodologies compliance in real-time. This provides the necessary quantitative baselines that, moving forward, complement the qualitative expert validation that is undergone in the second phase of the methodology.

Agile Compliance Scoring (0-100 scale): This metric evaluates the cycle efficiency in several aspects including cross-functional collaboration, adaptability and delivery orientation and is derived through an automated workflow analysis.

Lean Efficiency Scoring (0-100 scale): Lean scoring measures waste reduction, defect detection, handoff efficiency, and digital integration. This combination provides a comparable view of methodological alignment.

Stage 4: Context-Enhanced Prompt Generation for AI Analysis

The system constructs comprehensive analysis prompts following standardized academic format as illustrated in Figure 3.

```python
def _build_enhanced_analysis_prompt(self, workflow_data: Dict[str, Any],
                                    context: Optional[Dict[str, Any]],
                                    rag_context: str) -> str:
    """Build dynamic analysis prompt that performs real workflow analysis."""

    # Base workflow information
    workflow_title = workflow_data.get('title', 'Unnamed Workflow')
    workflow_desc = workflow_data.get('description', 'No description provided')
    processes = workflow_data.get('processes', workflow_data.get('steps', []))
    stakeholders = workflow_data.get('stakeholders', workflow_data.get('roles', []))

    # Current metrics if available
    metrics_section = ""
    if 'current_metrics' in workflow_data:
        metrics = workflow_data['current_metrics']
        metrics_section = f"""
**Current Performance Metrics:**
- Cycle Time: {metrics.get('cycle_time_days', 'N/A')} days
- Annual Volume: {metrics.get('annual_cases', 'N/A')} cases
- Process Steps: {len(processes)} steps
- Stakeholders Involved: {len(stakeholders)} parties
```

Figure 3: Code Implementation for Context-Enhanced Prompt Generation

Stage 5: Foundation Model Analysis with Enhanced Context

Claude Opus 4 processes the enhanced prompts and combines the originally provided workflow data with verified evidence, providing a context for the generated recommendations that is a step further than generalized optimization approaches. The main contributor towards this outcome is the integration of real-world performance data and best practice frameworks.

Stage 6: Structured Academic Output Generation

The system produces thoroughly validated reports containing bottleneck classification and phased implementation roadmaps. The reports follow renowned format requirements and encompass a quantitative compliance analysis generated by executive assessments.

3.2.3 Cloud Infrastructure Integration and Deployment Strategy

The system entails the use of Amazon Web Services (AWS) for its cloud infrastructure via AWS Bedrock. The objective behind this approach is to provide unified access to the foundation models. This also ensures flexibility for deployment across varying environments for any potential future usage. The integration strategy in question creates a balanced implementation that allows for scalable enterprise production as well as local development support, all while satisfying the requirements for the research.

AWS Integration Architecture:

- **AWS Bedrock Integration:** Provides the unified access to Claude Sonnet 4, Claude Opus 4, DeepSeek R1, and Llama 3.3 70B models

- **Credential Management**: Supports flexible authentication mechanisms - environment variables, AWS profiles, and credential files

- **Cost Optimization:** Implements intelligent API usage patterns with throttling safeguards and resource management policies that retain necessary performance levels as well as financial efficiency

- **Hybrid Storage:** Combined with usage of Amazon S3 for scalable, cloud-based deployment with local storage options

3.3 Data Foundation and Processing

3.3.1 BPI Challenge 2020 Dataset Integration and Preprocessing

This section of the methodology analyses the undertaking of the Business Process Intelligence Challenge 2020 dataset, which is used as the benchmark for the workflow classification throughout the research. The dataset is centred on university travel and accommodation processes, laying the domain groundwork for the Loughborough University London collaboration and making the second phase consisting of expert validation directly relevant to the prior. The dataset consists of a number of travel expense claim events spanning over the course of two years from a European university, providing a close representation of how the target implementation environment is meant to operate through the authentic business process instances.

The dataset captures relevant business processes across different organizational levels, where the data from 2017 covered claim events from two departments, moving on to 2018 in an institution-wide scope from all departments, which offers a valuable lens on both the departmental and institutional levels. The process variations provided by this context enables empirical evaluation of the approaches to streamlining workflows across administrative contexts.

The domain alignment between the present data and the academic alliance comprised with the university for the purpose of the research renders the interaction strategic and contributes towards the equinity of the model's depiction of real-world challenges in academic travel authorization. Expert opinions will be particularly relevant as the dataset reflects common approval workflows, including employee submission, travel administration approval and budget owner authorization, as well as similar documentation, with even exceptions requiring additional director approval included.

Dataset Process Characteristics:

- **Process Instances:** Over 33,000 real life cases of travel and accommodation workflows from real university environments

- **Domain Alignment:** Focus on travel authorization and expense reimbursement processes is directly relevant to university implementation

- **Process Flow Structure:** Captures a standardized approval hierarchy, involving travel administration, budget owners, supervisors, and directors

- **Trip Classification:** Domestic trips (post-travel reimbursement) and international trips (pre-approval required through travel permits)

- **Temporal Scope:** Covers a two-year period, spanning departmental (2017) and university-wide (2018) process coverage

Process Type Categories:

- **Travel Permit Processes:** Domestic and international pre-approval workflows

- **Request for Payment Processes**: Cost reimbursement workflows for approved travel

- **Prepaid Travel Costs Processes:** Advance payment workflows for confirmed expenses

- **International Declarations Processes**: Post-travel expense documentation for international trips

- **Domestic Declarations Processes:** Post-travel expense documentation for domestic trips


```xml
<trace>
        <string key="id" value="declaration 86791"/>
        <string key="concept:name" value="declaration 86791"/>
        <string key="BudgetNumber" value="budget 86566"/>
        <string key="DeclarationNumber" value="declaration number 86792"/>
        <float key="Amount" value="26.85120450862128"/>
        <event>
                <string key="id" value="st_step 86794_0"/>
                <string key="org:resource" value="STAFF MEMBER"/>
                <string key="concept:name" value="Declaration SUBMITTED by EMPLOYEE"/>
                <date key="time:timestamp" value="2017-01-09T09:49:50.000+01:00"/>
                <string key="org:role" value="EMPLOYEE"/>
        </event>
        <event>
                <string key="id" value="st_step 86793_0"/>
                <string key="org:resource" value="STAFF MEMBER"/>
                <string key="concept:name" value="Declaration FINAL_APPROVED by SUPERVISOR"/>
                <date key="time:timestamp" value="2017-01-09T11:27:48.000+01:00"/>
                <string key="org:role" value="SUPERVISOR"/>
        </event>
        <event>
                <string key="id" value="dd_declaration 86791_19"/>
                <string key="org:resource" value="SYSTEM"/>
                <string key="concept:name" value="Request Payment"/>
                <date key="time:timestamp" value="2017-01-10T09:34:44.000+01:00"/>
                <string key="org:role" value="UNDEFINED"/>
        </event>
        <event>
                <string key="id" value="dd_declaration 86791_20"/>
                <string key="org:resource" value="SYSTEM"/>
                <string key="concept:name" value="Payment Handled"/>
                <date key="time:timestamp" value="2017-01-12T17:31:22.000+01:00"/>
                <string key="org:role" value="UNDEFINED"/>
        </event>
</trace>
```


Figure 4: Raw Event Structure from DomesticDeclerations.xes File (BPI Challenge 2020 Dataset)

The structured format of the event logs taken from the BPI Challenge 2020 dataset is shown in figure 4. The organizational context required for workflow analysis is provided by the XML structure, which reveal the tracking of process execution from initial employee submission through final supervisor approval.

Event Log Field Structure:

- **trace_id:** Unique identifier for each process instance
- **event_id:** Unique identifier for individual events within a trace
- **activity**: Sequential process step name (e.g., "Declaration SUBMITTED by EMPLOYEE", "Declaration FINAL_APPROVED by SUPERVISOR")
- **timestamp:** Precise temporal occurrence of each activity
- **resource**: Individual system user or automated system executing the activity (e.g., "STAFF MEMBER")
- **role:** Organizational position responsible for activity execution
- **amount:** Monetary value associated with travel expense claim
- **budget_number:** Budget allocation identifier for financial tracking (e.g., "budget 86566")
- **declaration_number:** Unique declaration reference number

3.3.2 Advanced Data Processing and Novel Algorithmic Labeling Methodology

The BPI Challenge 2020 dataset is supplied as raw process event logs in XES format, without pre-existing efficiency classifications, presenting the fundamental limitation of data scarcity discussed in the literature, which hinders the supervised machine learning for the research. Assigning an expert for labeling would be high-priced and liable to subjective bias which necessitated the development of an algorithmic labelling approach that will help with the model's reasoning while maintaining academic rigor.

Multi-Dimensional Scoring Framework Implementation: The labelling system examines each of the 33,000+ process instances covering three main set dimensions, generating scores on validated 1-10 with detailed academic Reasoning. Using Claude Sonnet 4 integrated with RAG enhancement, the Enhanced Framework Analyst system was built ensuring a uniform application of scoring criteria while offering a transparent justification for each evaluation.

Process Efficiency Scoring Algorithm: Peer-reviewed academic publications analyzing the BPI Challenge 2020 were used to derive benchmarks grounding the efficiency scores. The system classification involves assessment of cycle time (process duration relative to BPI median benchmarks), process complexity (stakeholders involved and unique activity counts, and success rates (completion versus failure patterns), ensuring the statistical validity through median-based threshold while incorporating research-backed business process reengineering methodologies.

Agile Methodology Compliance Assessment: The labeling framework performs a quantitative analysis, which turns Agile Manifesto principles into measurable criteria. The analysis takes into account early value delivery ratios, which is a representation of the time to the first approval in relation to the total time required for process completion, process simplicity evaluation, an activity count and complexity patterns reflecting agile principles, role handoffs and stakeholder interaction patterns, and successful process completion assessment. The methodology takes the focus towards administrative workflows, making them compliant with Agile principles beyond simple development.

Lean Waste Elimination Analysis: The evaluation looks into waste and inefficiency in existing administrative business processes. The elements of the evaluation are essentially bottleneck detection through continuous assessment against statistical thresholds, analysis of value contribution for each activity, and the recognition of applicable process standards and optimization patterns. The approach uses the methodology once applied to manufacturing principles and extends it to encompass the evaluation of business processes, which maintains the integrity of the initial research and principles of the Lean methodology.

The developed labeler (bpi_data_labeler.py) creates consistent efficiency classifications through a comprehensive scrutiny of the academic research through batch processing methodology, integrated with quality assurance mechanisms. The literature tackling BPM provides the necessary benchmarks derived from the BPI Challenge academic publications to assess the workflows, while the use of multiple models, enhanced by the Framework Analyst Agent with RAG, ensures consistent, high-quality reasoning is produced by all process instances.

Through this approach, the absence of ground truth labels becomes a strong suit, rather than a discrepancy. The creation of objective, theoretically grounded efficiency measures enables a backed-up appraisal for the foundation models that also upholds academic compliance and sufficient defense for the dissertation.

3.4 Retrieval-Augmented Generation Implementation

Even though large pre-trained language models evidently have a massive ability to store a substantial amount of knowledge, these models undoubtedly encounter a significant share of limitations in terms of acquiring and manipulating knowledge with precision, which often produces information that is factually incorrect, also known as "hallucinations" (Lewis et al., 2020). These large pre-trained language models however still face complications when it comes to making a clear precise decision and updating their knowledge base when new information prevails (Lewis et al., 2020). RAG, also known as Retrieval-Augmented Generation, tackles the aforementioned limitations by merging both parametric memories, provided from pre-trained models, and non-parametric memory, provided through external knowledge retrieval (Lewis et al., 2020). The combined architecture of parametric and non-parametric memory allows models to retrieve relevant documents from external knowledge sources throughout the generation process, which leads to the generation of an increased factually correct output while supporting flexibility to updating domain-specific knowledge without having to retrain the model (Lewis et al., 2020). RAG can also be an extremely useful framework for business process optimization application, even if they have minimal training data, because it can use their proprietary knowledge sources, such as internal process documentation, operational procedures, and organizational guidelines, to take advantage of these sources without any additional external model training.

3.4.1 Dynamic Knowledge Architecture with Automatic Document Discovery

RAG is executed by merging three unambiguous knowledge sources through an intelligent auto-discovery system which allows for a smooth expansion of the knowledge base without requiring any manual involvement in the process. What differentiates this implementation from the traditional static RAG systems is that the system architecture (rag_system.py) performs automatic PDF detection, processing, and integration capabilities while also upholding evaluation consistency through intelligent mode detection.

Knowledge Base Composition:

- **BPI Workflow Patterns**
- **Academic Literature:** Peer-reviewed papers on business process mining, scaled agile frameworks, and critical success factors in large-scale implementations
- **Consulting Research:** Industry frameworks and methodologies from distinguished consulting firms covering transformation and operating model optimization

Current Document Collection (23 Research Documents):

- **Academic Papers:**
  - Business Process Mining Survey 2021
  - SAFe Scaled Agile Framework Lean Engineering 4.0
  - Critical Success Factors in Large Scale Agile Software Development
- **BPI Challenge 2020:**
  - ICPM 2020 Paper 99
- **McKinsey & Company Reports:**
  - McKinsey HR New Operating Model 2024
  - McKinsey Next Generation Operating Model 2024
  - McKinsey Scaling Next Gen Operating Model 2024
- **Boston Consulting Group Reports:**
  - BCG Agile Transformation Management 2019
  - BCG Transformation Methodology 2016
- **Deloitte Reports:**
  - Deloitte Digital Transformation Series 15 2024
- **2024 Agile Transformation Research:**
  - Overcoming the Agile Operating Model Challenges
  - From Agile Experiments to Operating Model Transformation
  - Agile Transformation Survey
  - Transforming for Growth: An Evidence-Based Guide

3.4.2 Automatic Document Discovery and Integration System

The RAG system enforces advanced auto-discovery capabilities which observes the document repository and automatically creates new research without having to go through a system restart or code modification.

This process happens when the _discover_and_process_new_documents() function scans the real_documents/ directory structure and authenticates unprocessed PDFs by comparing them with existing processed files and continuously integrating new content into the knowledge base.

Dynamic Knowledge Management Features:

- **Automatic PDF Detection**: Attaining new research papers through recursive scanning of document directories
- **Duplicate Prevention:** In order for re-processing not to occur, intelligent comparison needs to happen against existing processed documents
- **Seamless Integration**: Automatic incorporation of new content into existing knowledge base structure
- **Evaluation Mode Intelligence:** For the integrity of experimental consistency, the auto-discovery is disabled throughout model evaluations (EVALUATION_MODE detection) to provide experimental consistency and the preservation of operational capabilities

3.4.3 Evidence-Based Context Generation with Statistical Validation

The improved context generation method alters the analysis of foundation models from theoretical suggestions to data-driven suggestions by incorporating an extensive amount of evidence-based context. The system produces quantitative confidence metrics by utilising similarity scores, pattern frequency distributions, and collective evidence strength. This permits foundation models to allocate the correct amount of prominence to the recommendations created based on the 231-pattern BPI dataset, using its empirical support, and the 23-document research collection, using its theoretical framework.

Evidence-Based Recommendation Enhancement: The RAG system evolves model recommendations from broad guidelines to precise, feasible recommendations backed by real-world data. The system does not provide theoretical advice, but instead, it demonstrates quantified improvement opportunities through leveraging the best-performing patterns, finding risk factors in cases that failed to pan out, and analysing performance between various identical workflow cases. This technique safeguards that the outcomes of the foundation model are founded on authentic organisational information that is supported by authoritative research frameworks.

Multi-Source Intelligence Integration: Context augmentation incorporates retrieved BPI patterns alongside pertinent research findings and statistical analysis to produce extensive context-based data. The system deploys advanced content selection algorithms to guarantee that models obtain both particular examples of analogous workflows from the BPI dataset and general theoretical frameworks from the research document collection. Such suggestions depend on concrete proof opposed to general best practices and involve quantified confidence assessments.

3.5 Evaluation Framework and Methodology

The previously mentioned framework builds on the dual-phase evaluation design that was covered in Section 3.1. It concentrates on strict protocols and attainable prerequisites for evaluating both technical precision and pragmatic application.

In the classification phase, a blind evaluation protocol is utilised. This implies that the foundation models analyse patterns of workflow efficiency extracted from the dataset, accordingly, this technique guarantees an impartial evaluation of the models' analytical skills, as opposed to mere data retention. The evaluation protocol creates extensive accuracy indicators, classification confidence scores, processing times, and error analysis for all four models, making it simpler to find the ideal foundation model for determining classified productive workflow tasks.

By involving university personnel with valuable backgrounds in process management to conscientiously assess model-generated optimisation outcomes, the expert validation phase bridges the vital gap among algorithmic performance and practical relevance. This method is tailored for research due to its dependence on prompt peer evaluation of advised optimisation strategies, which falls in line with industry practice just like operating model assessments are usually governed post-implementation using KPI metrics and stakeholder surveys (Gilsing et al., 2022).

In order to identify and assess foundation models enhanced with RAG, Retrieval-Augmented Generation, for business process improvement applications, this evaluation framework combines both quantitative and qualitative assessments.

4. Result and Analysis

4.1 Phase 1: Foundation Model Classification Performance

In this phase, the model evaluated all four state-of-the-art language models on their capabilities to identify and class the workflow efficiency patterns without the need of any external additional knowledge. This evaluation signifies that each model inherents analytical capabilities for the business process analysis.

4.1.1 Evaluation Dataset and Ground Truth Distribution

The classification utilized 231 distinctive workflow patterns extracted from the BPI challenge 2020 statistics, showcasing real travel and accommodation travel processes. The Algorithmic labeling methodology created ground truth categorization classifed as follows:

- **Efficient workflows:** 168 patterns (72.7%)
- **Inefficient workflows:** 63 patterns (27.3%)
- **Total evaluations:** 924 (231 patterns × 4 models)

The classification reflects real life scenarios where the majority of the organizational workflows achieve the fundamental functionalities, however there can be significant improvements and opportunities within the inefficient subset.

4.1.2 Foundation Model Performance Results

The comparative evaluation suggests distinguishable performance features across all four foundation models, with insignificant accuracy difference and noticeable variations in processing efficiency and response.

Model	Accuracy (%)	Avg. Processing Time (s)	Confidence	Avg. Agile Score (/100)	Avg. Lean Score	Avg. Overall Score	Avg. Response Length (chars)
Llama 3.3 70B	64.5	2.8	0.79	13.4	15.3	14.4	2,535
Claude Sonnet 4	64.1	8.4	0.80	17.1	13.2	15.1	546
Claude Opus 4	64.1	14.0	0.80	14.8	11.3	13.1	539
DeepSeek R1	64.1	3.6	0.80	28.7	15.3	27.2	443

Table 1: Foundation Model Classification Performance

4.1.3 Performance Analysis and Model Characteristics

Classification Accuracy: The evaluation shows remarkably consistent accuracy across all models, with Llama 3.3 70B achieving the highest performance at 64.5%, followed by the three other models at 64.1%. The 0.4% performance difference suggests that foundation model selection alone can result in a limited differentiation for workflow efficiency classification. Given this analysis, additional evaluation criteria is very essential for optimal model selection. These results also highlight the potential value of Retrieval-Augmented Generation enhancement, as foundation models alone demonstrate moderate classification capabilities that require augmentation for practical organizational deployment.

Processing Efficiency: noticeable variations are shown in computational efficiency, with Llama 3.3 70B demonstrating the fastest processing at 2.8 seconds average per evaluation, while Claude Opus 4 required 14.0 seconds. However, these processing time differences are

primarily due to the AWS Bedrock API limitations rather than inherent model performance characteristics. The AWS platform imposes different rate limiting constraints for various models, requireing mandatory delays between API requests to prevent throttling. Especially, Claude models that required 8-second delays between requests due to stricter AWS rate limits, while Llama and DeepSeek operated with 2-second delays. This represents a deployment constraint rather than computational efficiency, highlighting the importance of considering infrastructure limitations in enterprise AI implementation strategies.
Response Quality and Reasoning Analysis: The models showcased distinct ways to workflow analysis not just simple classification. Each one of the models provided comprehensive reasoning to support their efficiency evaluation, enabling qualitative evaluation of analytical depth and consistency.

Claude models (Sonnet 4 and Opus 4) showed consistently organized responses averaging 540 characters with standardized formatting and focused analytical reasoning. On the other hand, Llama 3.3 70B generated considerably longer outputs at 2,535 characters average, demonstrating a more structured analysis but exhibiting hallucination patterns where responses included excessive elaboration, repetitive content, and tangential information not directly relevant to the core classification task. This inconsistent outcome has resulted in elimination of Llama 3.3 70B for sub RAG integration although it showed higher accuracy marigns.

Scoring Framework Interpretation: DeepSeek R1 exhibited significantly different scoring patterns, assigning higher Agile (28.7/100) and Lean (25.7/100) scores in comparison to other models. This difference suggests divergent interpretation frameworks for business methodology assessment, which indicates calibration differences in how the model evaluates organizational processes against established frameworks.

4.1.4 Response Storage and Analysis Framework

All four model outputs were meticulously classified and stored in production-readiness formats for subsequent analysis, including complete response texts, confidence scores, processing timestamps, and extracted numerical assessments. This detailed data collection entails comprehensive qualitative analysis of the model reasoning patterns and provides the foundation for recognizing optional RAG integration strategies in the second phase.

The stored responses disclose distinct analytical approaches: Claude models demonstrate brief, structured reasoning with consistent formatting, while Llama 3.3 70B provides a detailed analysis but with significant inconsistencies in output structure and occasional content that extends further than the core classification requirements.

4.1.5 Reliability and Error Analysis

The four models were able to achieve 100% API success rates, indicating robust technical reliability across the evaluation framework. No systematic errors or failures occurred during all the 924 evaluations, showcasing the model’s stability under continuous workload constraints which demonstrates the production readiness of the infrastructure model.

The evaluation protocol's blind classification approach states that the noticeable performance differences are a result of genuine analytical capabilities and not memorization or data leakage, providing a credible foundation for subsequent RAG enhancement assessment.
4.1.6 Foundation Model Selection for RAG Integration

As per the evaluation, the foundational models demonstrated similar capabilities in analysing the raw workflow data. The analysis provided an accuracy convergence of 64%, which entailed that additional factors needed to be included in the decision, such as documented analytical strengths, response consistency and reasoning quality.

The Llama 3.3 70B showed inconsistent output structure and frequent hallucinations, resulting in its elimination from the foundational model selection. The model initially showed higher accuracy, but Claude Opus 4 was eventually chosen due to the aforementioned discrepancies. Amongst the remaining Claude models, the said model had substantially superior logical reasoning capabilities. According to Anthropic's system card, external evaluators observed that responses from the Claude Opus 4 were qualitatively preferable, which gives it an edge over the other models in terms of analytical depth (Anthropic, 2025).

Furthermore, Claude Opus 4 is more specialized in handling problem solving scenarios, where its hybrid reasoning architecture allows for the use of alternative cognitive pathways, enabling deeper analytical processing than the other models, confirmed by benchmark evaluations. Claude Opus 4 leads in performance rankings, with 72.5% on SWE-bench, on reasoning-intensive tasks. The rankings establish the model as the “world's best coding model" (Anthropic, 2025).

The superiority provided by the model in the process of inference, combined with the qualitative analysis of its responses being deemed successful, structured and consistent, establishes it as the proper selection for business process streamlining, performed as a result of extensive analytical processing.

The Phase 2 RAG integration therefore makes use of Claude Opus 4, leveraging its capabilities for the analysis of the RAG enhancement effects on the recommendations for workflow efficiency. The model’s proven reliability renders it applicable to the process, making it a value-added concatenation for practical applications.

4.2 Phase 2: RAG-Enhanced System Results and Expert Validation

Based on the foundation model assessment in Phase 1, Phase 2 shows the practicality of the RAG-enhanced business process optimization system along with expert validation from Loughborough University London’s Business Services HR and Finance department. As mentioned, Claude Opus 4 is adopted as the foundation model adjoining the comprehensive knowledge base to generate optimized workflow recommendations

4.2.1 University Expert Input Interface and Workflow Submission

The validation process began with university domain experts providing their real-world procurement workflows on the web interface. The expert decided to test on a Purchase Order Request with the following features:

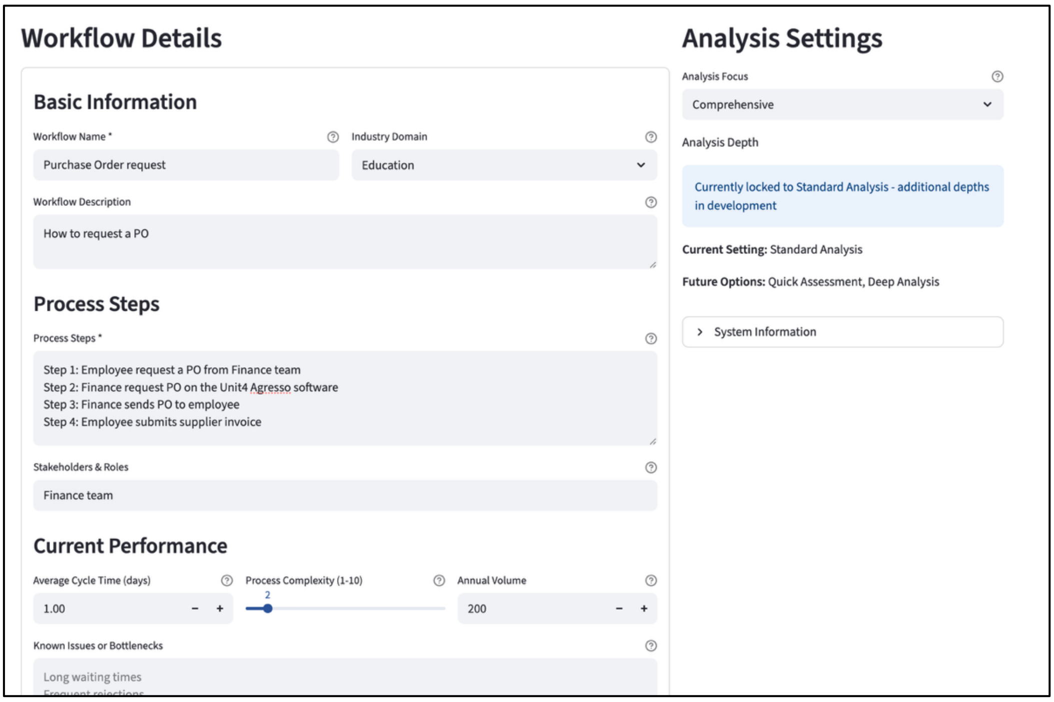

Figure 4: Input Interface Screenshot - Workflow submission form showing Purchase Order request specifications

- **Workflow Name:** Purchase Order request
- **Description**: "How to request a PO"
- **Process Steps:** 4 sequential steps
  - Employee request a PO from Finance team
  - Finance request PO on the Unit4 Agresso software
  - Finance sends PO to employee
  - Employee submits supplier invoice
- **Stakeholders:** Finance team (single stakeholder)
- **Performance Metrics:**
  - Cycle time: 1 day
  - Complexity score: 2/10
  - Annual cases: 200
  - Domain: Education sector

4.2.2 Executive Dashboard - Framework Compliance Scores

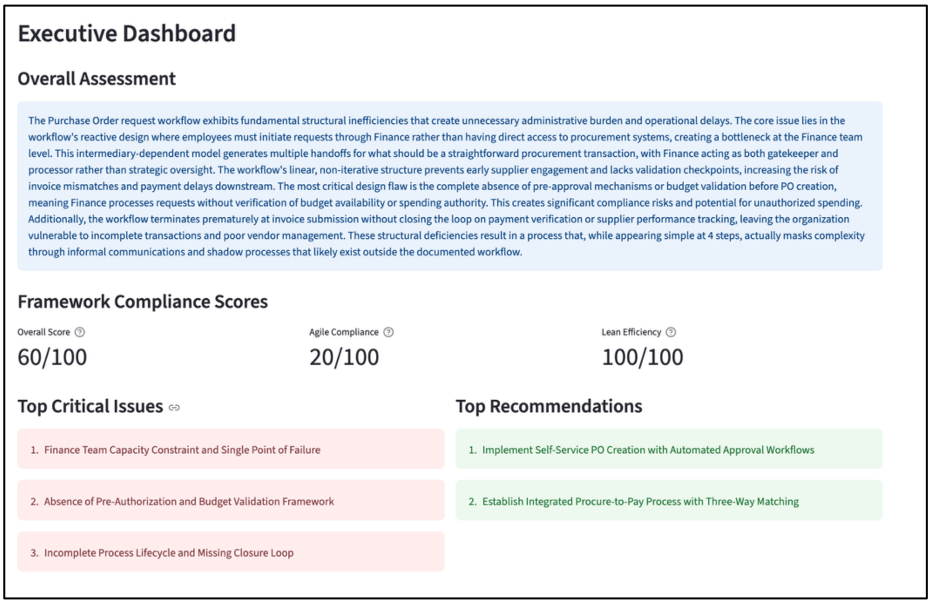

Figure 5: Executive Dashboard Screenshot - Author's simulation of Purchase Order workflow analysis

Figure 5 shows the model's executive summary on the expert's workflow. The model indicates low agile adaptation for the workflow giving it a score of 20/100, but excellent waste/ Lean efficiency with a perfect score of 100/100, and a fairly moderate overall performance.

***Screenshot Methodology Note:** Figure 5 represents a simulation analysis done by the author using the same workflow parameters to demonstrate the interface functionality. The analysis content is still on expert's actual workflow assessment results.

4.2.3 RAG-Enhanced System Analysis Output

Note: The following analysis is a direct output from the RAG-enhanced system based on the expert's workflow submission. This content is generated entirely by the AI system and demonstrates the automated analysis capabilities of the framework.

Primary Recommendations Generated by the System:

1. Implement Self-Service Procurement Portal with Automated Approval Routing - Deploy Unit4 Agresso self-service module with spending thresholds (under £1,000 auto-approved, £1,000-5,000 manager approval, over £5,000 Finance review)

2. Establish Integrated Procure-to-Pay Process with Three-Way Matching - Create seamless integration from PO request through invoice payment with systematic matching controls

Critical Bottlenecks Identified by the System:

1. Finance Team as Single Point of Failure: Positioned as sole gateway for all PO requests, creating processing bottleneck affecting 200 annual cases with constant interruptions diverting from strategic activities

2. Disconnected Invoice Processing: Step 4 operates in isolation without systematic PO-to-invoice matching, creating compliance gaps and control weaknesses

3. Absence of Approval Hierarchies: Lacks spending controls or authorization mechanisms regardless of purchase value or risk level

Primary System Recommendation:

- Implement Self-Service Procurement Portal with Automated Approval Routing
- Deploy Unit4 Agresso self-service module enabling direct employee access
- Establish spending thresholds (under £1,000 auto-approved, £1,000-5,000 manager approval, over £5,000 Finance review)
- Expected Benefits: 60-70% reduction in Finance processing time, hours instead of 1-day cycle time
- Implementation: Medium complexity, 6-12 weeks timeline

4.2.4 Visual Workflow Optimization

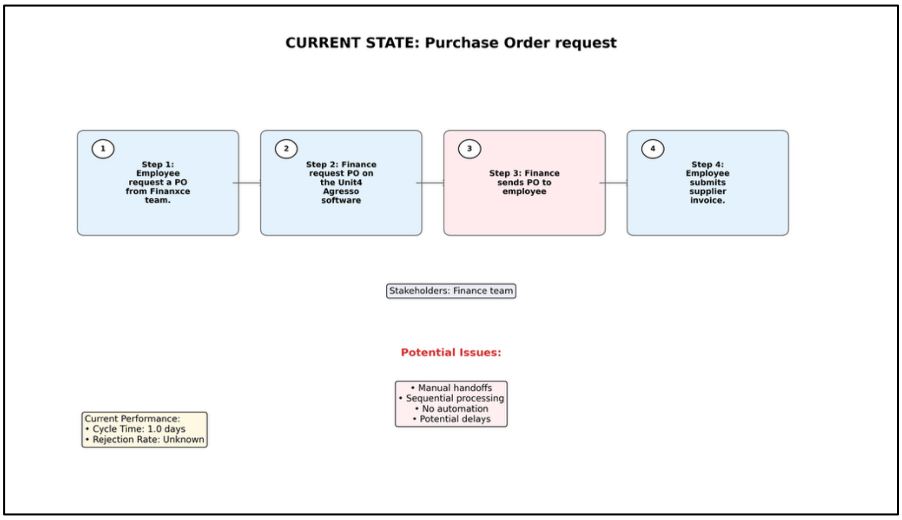

Figure 5: Current State Workflow Diagram

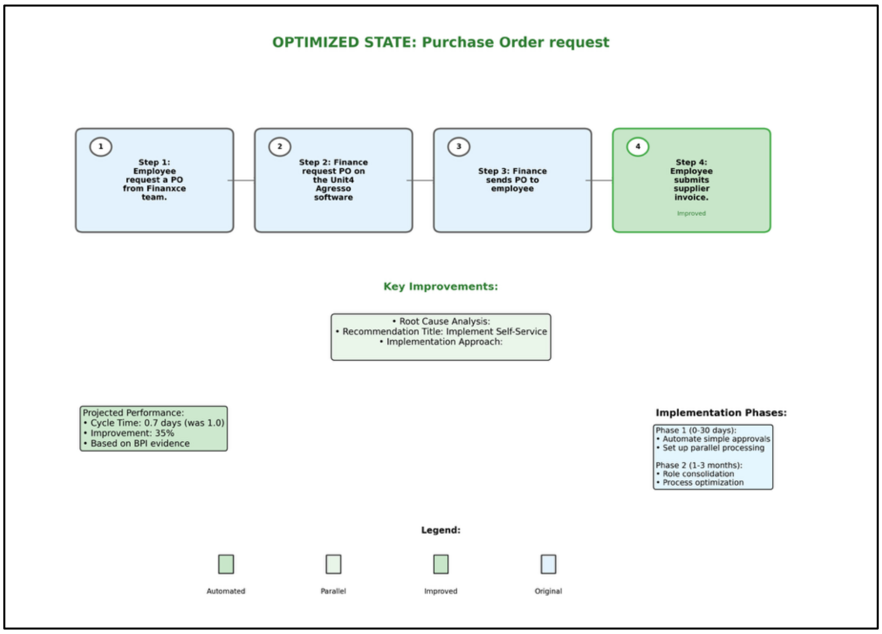

Figure 6: Optimized State Workflow Diagram

Figures 5 and 6 depicts user's workflow before and after the model assessment and recommendation. In the current workflow diagram, a bottleneck is highlighted in red, at Step 3, showing where exactly the inefficiency occurred and what could be the cause for such an issue. As this particular workflow will require the same steps but with different implementation style, more automation as mentioned, the diagram is similar to the previous. In other cases, the diagram adds or removes steps according to the recommendations while showing how this can affect the workflow positively.

4.2.5 University Expert Validation Results

The expert then provided a comprehensive personal assessment on the model's output through a structured questionnaire following the hands-on run on the Purchase Order Request workflow

Quantitative Expert Assessment Results:

- **Technical Quality and Professional Standards:** 8/10 - confirming high sophistication and accuracy of AI-generated workflow analysis
- **Bottleneck Identification Accuracy:** 7/10 - validating effective identification of Finance team bottleneck and process inefficiencies
- **Recommendation Quality and Actionability:** 8/10 - endorsing practicality of self-service procurement portal recommendation
- **Workflow Diagram Clarity:** 8/10 - confirming effective visual communication of current vs. optimized states
- **Workflow Model Accuracy:** 10/10 - perfect validation of system's capture of key process steps and decision points

Main System Strengths Identified:

- **Faster Processing Time:** Confirmed significant efficiency gains over manual analysis
- **Actionable Recommendation Generation**: Confirmed practical value of self-service portal recommendation

System Limitations Identified:

- **Limited Contextual Understanding:** Recognition that system may not capture organizational nuances affecting real-world implementation

Expert Validation Summary: The validation shows solid positive assessment with an ideal 10/10 score for workflow model accuracy, which asserts the system's capability to accurately capture university procurement processes. The key strengths' identification provides empirical evidence for the added value by the RAG-enhanced system as a cost-effective preliminary tool for analysis. The expert's acknowledgement of the tool's limitations aligns with the realistic expectations of it being a complementary tool rather than being human-replacing, asserting the readiness of the system for practical deployment in organizations.

Expert Improvement Suggestion: A "more standard language to present" the analysis, was the expert's suggestion when they were asked about specific amendments or additions to enhance the workflow model presentation. This implies the presence of opportunities for improved professional terminology and presentation formatting.

The Purchase Order request case study effectively shows the value of RAG enhancement, through the integration of BPI Challenge empirical benchmarks with structured business framework analysis, to generate specific, and quantifiable suggestions that are evidence-backed from similar 381 actual workflow cases.

5. Discussion

5.1 Analysis of Research Findings

The research findings demonstrate the potential and limitations when it comes to applying RAG-enhanced AI systems to business process optimization. The Phase 1 foundation model evaluation indicated significantly consistent classification accuracy across all tested models, with performance converging around 64-65%. This very convergence suggests that the current state-of-the-art language models have apparently attained near-equal analytical ability to assess workflow efficiencies, rendering a RAG enhancement more crucial than selecting the foundation model for any real-world business application.

The minimal performance differences between models (0.4% spread) reveal that the choice of foundation model may, in some instances, be less important than initially anticipated. Nonetheless, the qualitative analysis showed important distinctions in output consistency and reasoning quality, particularly Claude Opus 4's superior structured reasoning capabilities compared to other models' tendency toward hallucination and inconsistent analysis depth. Hence, this finding reinforces the premise of this research-the enhancement of RAG rather than raw model performance.

The classification accuracy rates lying between 64-65% reflect the inherent complexity of workflow efficiency assessment and cannot be attributed to any methodological shortcomings. Existing business process datasets that lack descriptive information, forced the labelling of business processes to be done algorithmically. Although this has been supported by published benchmarks for business process data, particularly for the BPI Challenge 2020 dataset, it remains a primary stumbling block in business process improvement research. The RAG enhancement filled this gap by providing associative contexts that enabled the newest process analyses to become more subtle than positive-negative dichotomy analysis.

5.2 RAG Enhancement Value and Evidence

Although direct quantitative comparison between RAG-enabled and baseline performance was not systematically quantified, qualitative evidence have proven the value of the RAG enhancement applied in providing a structured analytical framework. Even foundation models could hardly give specific actionable optimization guidance without a RAG system that integrated through Lean and Agile methodologies. Therefore, RAG allows the system to recommend evidence-based approaches from established business framework and empirical benchmarks from similar process patterns.

The enhanced system's capability to perform detailed bottleneck analysis, quantify expected improvement, and output a structured implementation road map is, in fact, far advanced from what generic AI advice can deliver. Expert validation is, thus, needed to provide evidence of the practical value of practical implementation, but those types of validations are realized as subjective, so that validity should be sought among many domain experts and within more organizational settings.

5.3 Research Limitations and Challenges

Several substantial limitations affected the scope and findings of the study. The primary constraint lay in the lack of labelled business process data, which could be used for rigorous verification. A number of datasets, including the BPI Challenge 2020 repository, provide event logs but no efficiency classification at all, nor any descriptive process documentation. This limitation required the researchers to create a labelling methodology of their own, which while based on previous studies and benchmarked on empirical data, constitutes a potential source of bias and circularity of validation.

The academic collaboration, although in favour of adding domain expertise, unearthed yet another limitation tied to data availability. Many organizational processes go undocumented while the execution of workflows depends upon many contextual considerations, lateral or informal communication channels, and decisions made outside of any formal procedural guidance. Reality underscores that there is a great deal of sophistication involved in accurately modelling real-world business processes and that static workflow analysis falls short.

Technical contraptions such as token limitations on AWS have been imposed on this analysis. Limited to given token counts, requests had to be optimized, and responses truncated-all of which could have affected the comprehensiveness of the offered recommendations. In application, this inquiry highlights the needs of this kind of system at the enterprise scale.

While aligning with industry practice, the expert validation approach introduces subjective assessment bias and limited sample size constraints. The research acknowledges a wider set of validation across multiple experts and organizational contexts to enhance the generalizability of the findings.

5.4 Methodological Considerations

Firstly, process optimization in travel and accommodation was strategically chosen due to the timeframe and resource constraints of a master’s thesis. Although limiting immediate generalizability, this allowed a deeper dive into a specific context and provided a basis for broader future application. Starting with process-specific optimization before expanding to department or organization levels is thus a pragmatic research progression.

Since RAG enhancement was firstly developed and employed for business process optimization, the work had to develop new evaluation approaches. This innovative direction, although enriching the research, meant there were neither best practices nor validation frameworks to rely upon. In this way, the research ventured into uncharted territory and sought to alleviate such by tackling the challenge through systematically integrating recognized business frameworks as well as empirical benchmarking grounds, thus serving as theoretical support for actual implementations.

5.5 Implications for Practice and Research

The research indicates that AI-based systems for business process optimization can be a cost-effective alternative for preliminary analysis when compared with consulting. An organization can use such a system for an initial assessment of the workflow before contracting an expensive external consultant, or else to verify the consultant's recommendations against the data-driven analysis of the systems. This practical application should be of great assistance, especially to organizations with tight budgets who require solutions to others' process improvements.

The findings thus indicate that the value of RAG-based systems resides largely in their ability to bolster the human-originated analysis with a structured framework and data support. Through generating detailed implementation roadmaps, quantitative forecasts, and recommendations supported by data, this enhancement system creates a basis for making informed decisions while leaving the actual decision of implementation to human discretion.

6. Conclusion

6.1 Research Summary and Contributions

The research successfully showcased the application of Retrieval-Augmented Generation (RAG) frameworks to business process optimization, addressing a critical gap in automated workflow analysis capabilities. The study's primary contribution lies in developing and validating a novel approach that combines advanced language models with domain-specific knowledge bases to generate evidence-based optimization recommendations for organizational workflows.

The two-phase evaluation methodology established foundation model capabilities before demonstrating RAG enhancement value through practical implementation and expert validation. The research revealed that while current state-of-the-art language models show convergent analytical capabilities, the integration of structured business frameworks and empirical benchmarks through RAG enhancement enables significantly more actionable and evidence-based optimization guidance.

The developed system architecture acts as a scalable system for the enterprise deployment, integrating cloud infrastructure, automatic knowledge discovery, and structured output generation. The comprehensive web interface is a practical example of how intricate AI analysis can be translated into business insights through executive dashboards, bottleneck analysis, and implementation roadmaps.

6.2 Theoretical and Practical Implications

In theory, the research, by showing how retrieval-augmented generation can overcome age-old shortcomings faced by using pure language model approaches in domain-specific applications, contributes to the emergent AI-augmented business process management. The blend of established business frameworks (Lean and Agile methodologies) with empirical process data creates a rigorously defendable system for automated process analysis.

In practice, the research ensures organizations access a cheap means of conducting rudimentary process assessments and receiving guidance for optimization. Theism of the approach stands, in fact, in opposition to replacing contemporary avenues of consulting, while providing data-driven baseline analyses together with evidence-based recommendations upon which to more confidently engage in the optimization initiatives. This happens to be a particular value for companies really trying to answer the right questions concerning process improvement on their own before making such heavy investment to hire consultants.

6.3 Future Research Directions

The research lays several promising future research directions. Urgently, there lies expansion of knowledge from being oriented only on travel and accommodation processes to include more expansive organizational workflows, and furthermore, to create more sophisticated ground truth generation techniques to address the shortage of labelled data in business process research.

These technical augmentations must develop to include industry-specific frameworks into the RAG knowledge base, multimodal input capabilities to ingest data from various media formats, and real-time organizational data streams for active process monitoring. Due to cloud-based architecture implementation, it is possible to envision automatic data collection, trend analysis, and proactive optimization recommendations.

Long-term longitudinal research could view the long-term effect of AI-driven recommendations for process optimizations, adaptive learning, so that these are improved based on the results of implementation, and integration with the so-called existing ERP and business intelligence platforms.

6.4 Scalability and Enterprise Deployment

It is evident from the research that pathways exist for general enterprise deployment through a cloud infrastructure and a modular system architecture. Other developmental aspects should encompass data privacy requirements for proprietary organizational information, integration with extant business systems, and a customizable framework for industry-specific criteria of optimization.

The potential for automated organizational health monitoring through continuous process assessment represents a significant opportunity for future research. Such systems could proactively identify emerging inefficiencies, recommend preventive optimizations, and provide ongoing performance monitoring to ensure sustained process effectiveness.

6.5 Research Limitations and Future Validation

While the research provides strong preliminary evidence for RAG-enhanced business process optimization, broader validation across multiple organizational contexts, process domains, and expert evaluations remains necessary. Future research should address the subjective nature of process efficiency assessment through the development of more objective measurement frameworks and larger-scale validation studies.

The current focus on single-process optimization, while appropriate for this research scope, it does represent a constraint that future work should address through the development of integrated departmental and organizational optimization capabilities. The expansion from process-specific to enterprise-wide optimization represents both a significant technical challenge and a substantial opportunity for organizational impact.

6.6 Final Recommendations

Based on such research outcomes, one could recommend the adoption of RAG-enhanced business process optimization as a first business analysis tool, while keeping realistic expectations about what it can offer at the moment in terms of technology. While it provides a structured analysis and an evidence-based recommendation, decisions on implementation and adaptation to context require human experts.

From the point of view of researchers, this work allows the observation of the applicability of advanced AI techniques to the more traditional problems of business, emphasizing the need for integration of domain knowledge and strict validation methods. In essence, the merger between AI and BPM is a promising avenue for further exploration with real-world potential implications.

The research establishes a foundation for the next generation of AI-augmented organizational optimization tools, providing both theoretical frameworks and practical implementation guidance for future development in this emerging field. The successful demonstration of RAG enhancement for business process optimization opens new possibilities for intelligent organizational analysis and evidence-based improvement recommendations.

References

Al-Omran, K., AlZayer, J. and Arnout, M. (2019) 'Causes of business process reengineering failure in the Kingdom of Bahrain and Saudi Arabia', *SSRN*. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3533688 (Accessed: 29 July 2025).

Almalki, S.S. (2025) 'AI-driven decision support systems in agile software project management: enhancing risk mitigation and resource allocation', *Systems*, 13(3), 208.

American Express (2025) *2024 fraud landscape report*. American Express Insights. Available at: https://www.americanexpress.com/en-us/business/trends-and-insights/articles/fraud-landscape-report (Accessed: 29 July 2025).

Anthropic (2025) *Introducing Claude 4*. Available at: https://www.anthropic.com/news/claude-4 (Accessed: 28 August 2025).

Anthropic (2025) *System card: Claude Opus 4 & Claude Sonnet 4*. Available at: https://www-cdn.anthropic.com/4263b940cabb546aa0e3283f35b686f4f3b2ff47.pdf (Accessed: 28 August 2025).

Boston Consulting Group (2017) *Getting smart about change management*. Available at: https://www.bcg.com/publications/2017/change-management-getting-smart-about-change-management (Accessed: 29 July 2025).

Boston Consulting Group (2024) *Most large-scale tech programs fail—here's how to succeed*. Available at: https://www.bcg.com/publications/2024/most-large-scale-tech-programs-fail-how-to-succeed (Accessed: 29 July 2025).

Boston Consulting Group (2025) *The four biggest organizational cost challenges—and how to solve them*. Available at: https://www.bcg.com/publications/2025/four-biggest-organizational-cost-challenges (Accessed: 29 July 2025).

Celonis (2024) *What is process mining?* Available at: https://www.celonis.com/process-mining/what-is-process-mining/ (Accessed: 29 July 2025).

Celonis Academic Alliance (2024) *Process mining research*. Available at: https://www.celonis.com/acal-researchers/ (Accessed: 29 July 2025).

Conforto, E.C., Salum, F., Amaral, D.C., da Silva, S.L. and de Almeida, L.F.M. (2014) 'Can agile project management be adopted by industries other than software development?', *Project Management Journal*, 45(3), pp. 21–34.

Davenport, T.H., Holweg, M. and Snyder, K. (2023) 'How AI fits into Lean Six Sigma', *Harvard Business Review*, 9 November.

Deloitte (2024) *Upward climb with uphill struggles: 2024 Deloitte corporate travel study*. Deloitte Insights.

Dikert, K., Paasivaara, M. and Lassenius, C. (2016) ‘Challenges and success factors for large-scale agile transformations: a systematic literature review’, *Journal of Systems and Software*, 119, pp. 87–108.

Dumas, M., La Rosa, M., Mendling, J. and Reijers, H.A. (2013) *Fundamentals of business process management*. Heidelberg: Springer.

Emburse (2023) *2023 spend management trends report*. Emburse Insights. Available at: https://www.emburse.com/insights/2023-spend-management-trends-report (Accessed: 29 July 2025).

Emerald Insight (2015) ‘Critical failure factors of Lean Six Sigma: a systematic literature review’, *International Journal of Lean Six Sigma*. Available at: https://www.emerald.com/insight/content/doi/10.1108/ijqrm-09-2013-0147/full/html (Accessed: 29 July 2025).

ExpensePoint (2025) *The true cost of a manual expense report*. Available at: https://www.expensepoint.com/blog/breaking-down-costs-associated-with-expense-reports/ (Accessed: 29 July 2025).

Gartner (2024) *Technology overview for travel expense management applications*. Gartner Research.

Gilsing, R., Turetken, O., Grefen, P., Ozkan, B. and Adali, O.E. (2022) ‘Business model evaluation: a systematic review of methods’, *Pacific Asia Journal of the Association for Information Systems*, 14(4), Article 2. doi:10.17705/1pais.14402.

Harvard Business Review (1993) ‘How to make reengineering really work’, *Harvard Business Review*, November. Available at: https://hbr.org/1993/11/how-to-make-reengineering-really-work (Accessed: 29 July 2025).

Harvard Business Review (2013) ‘Consulting on the cusp of disruption’, *Harvard Business Review*, October. Available at: https://hbr.org/2013/10/consulting-on-the-cusp-of-disruption (Accessed: 29 July 2025).

Harvard Business Review (2023) ‘How AI is helping companies redesign processes’, *Harvard Business Review*, March. Available at: https://hbr.org/2023/03/how-ai-is-helping-companies-redesign-processes (Accessed: 29 July 2025).

iSixSigma (2025) *Business process reengineering in a Six Sigma world*. Available at: https://www.isixsigma.com/financial-services/business-process-reengineering-six-sigma-world/ (Accessed: 29 July 2025).

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S. and Kiela, D. (2020) ‘Retrieval-augmented generation for knowledge-intensive NLP tasks’, *Advances in Neural Information Processing Systems*, 33, pp. 9459–9474. Available at: https://arxiv.org/abs/2005.11401 (Accessed: 16 September 2025).

McKinsey & Company (2023) ‘AI: the next frontier of performance in industrial processing plants’, *McKinsey Global Institute*. Available at: https://www.mckinsey.com/industries/metals-and-mining/our-insights/ai-the-next-frontier-of-performance-in-industrial-processing-plants (Accessed: 29 July 2025).

McKinsey & Company (2024) ‘A new operating model for a new world’, *McKinsey Quarterly*. Available at: https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/a-new-operating-model-for-a-new-world (Accessed: 29 July 2025).

PwC (2024) *Global workforce hopes and fears survey 2024*. Available at: https://www.pwc.com/gx/en/issues/workforce/hopes-and-fears.html (Accessed: 29 July 2025).

SciELO (2021) ‘Improving the success rate of business process re-engineering projects: a business process re-engineering framework’, *SciELO*. Available at: http://www.scielo.org.za/scielo.php?script=sci_arttext&pid=S1560-683X2021000100003 (Accessed: 29 July 2025).

Six Sigma Institute (2025) *Business process reengineering with examples [2025/26]*. Available at: https://www.sixsigma-institute.org/Six_SigmaVsBusinessProcessReengineering_Comparison.php (Accessed: 29 July 2025).

Six Sigma Online (2025) ‘Combining Six Sigma with business process reengineering’, *Six Sigma Online*. Available at: https://www.sixsigmaonline.org/six-sigma-business-preocess-reengineering/ (Accessed: 29 July 2025).

SixSigma.us (2024) ‘Business process reengineering (BPR): a strategic approach to organizational transformation’, *SixSigma.us*. Available at: https://www.6sigma.us/process-improvement/business-process-reengineering-bpr/ (Accessed: 29 July 2025).

Springer (2024) ‘Business process management in the age of AI – three essential drifts’, *Information Systems and e-Business Management*, 22, pp. 1–26.

Tashkinov, A. (2024) ‘The implementation of lean and digital management techniques using artificial intelligence in industrial settings’, *Discover Artificial Intelligence*, 4, 94.

Appendix

Appendix A: Digital Resources and Code Repository

A.1 Source Code Repository

The complete source code for the RAG-enhanced business process optimization system is available in the public GitHub repository:

Repository URL: https://github.com/Mando789/diss-data

The repository contains:

- RAG system implementation (rag_system.py)
- Foundation model evaluation scripts (fixed_model_evaluation.py)
- Web interface implementation (web_interface.py)
- Data processing and labeling modules (bpi_data_labeler.py)
- Analysis result generation tools (generate_final_comparison.py)
- Sample workflow analysis outputs (JSON format)
- Documentation and setup instructions

A.2 Expert Validation Questionnaire

The structured questionnaire used for university expert validation is accessible online:

Questionnaire URL: https://tally.so/r/3EpYpl

The questionnaire evaluates:

- Technical quality and professional standards of AI-generated analyses
- Alignment of framework scoring with expert judgment
- Bottleneck identification accuracy and recommendation quality
- Workflow diagram clarity and model accuracy
- System strengths, weaknesses, and improvement suggestions
- Overall quality and applicability assessment

A.3 Dataset Information

BPI Challenge 2020 Dataset

- **Source**: IEEE Task Force on Process Mining
- **Repository**: 4TU.ResearchData
- **URL**: https://data.4tu.nl/articles/dataset/BPI_Challenge_2020_Domestic_Declarations/12692543
- **Content**: 33,000+ process instances from European university travel workflows
- **Coverage**: Travel permits, payment requests, expense declarations (2017-2018)
- **Format**: XES event logs with temporal and organizational metadata

- **Specific Dataset Used**: Domestic Declarations subset containing university travel and accommodation processes

Additional BPI Challenge 2020 Components:

- International Declarations: https://data.4tu.nl/articles/dataset/BPI_Challenge_2020_International_Declarations/12696884
- Travel Permits: https://data.4tu.nl/articles/dataset/BPI_Challenge_2020_Travel_Permits/12688207
- Request for Payment: https://data.4tu.nl/articles/dataset/BPI_Challenge_2020_Request_for_Payment/12688534
- Prepaid Travel Cost: https://data.4tu.nl/articles/dataset/BPI_Challenge_2020_Prepaid_Travel_Cost/12693014

Algorithmic Labeling Dataset

- **Generated**: 231 unique workflow patterns with efficiency classifications
- **Method**: Multi-dimensional scoring across efficiency, Agile, and Lean frameworks
- **Validation**: Grounded in published BPI Challenge research and empirical benchmarks
- **Availability**: Included in GitHub repository as processed JSON files