

Received 29 April 2026, accepted 15 May 2026, date of publication 18 May 2026, date of current version 26 May 2026.

Digital Object Identifier 10.1109/ACCESS.2026.3694469

TOPICAL REVIEW

Continuous Business Process Improvement Driven by Large Language Models

ANDERSON S. ARAÚJO^{1,2}, HENRIQUE S. MAMEDE², VÍTOR SANTOS³,
AND VÍTOR FILIPE⁴

¹INESC TEC, University of Trás-os-Montes and Alto Douro, Lisbon, Portugal

²INESC TEC, Portuguese Open University, Lisbon, Portugal

³MagIC, NOVA IMS, Universidade Nova de Lisboa, Lisbon, Portugal

⁴INESC TEC, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal

Corresponding author: Henrique S. Mamede (hsmamede@gmail.com)

This work was supported by the National Funds through the Fundação para a Ciência e a Tecnologia (FCT), I.P., under Grant UID/50014/2025.

ABSTRACT Some of the main challenges faced by organizations when applying Continuous Business Process Improvement are data fragmentation, limited explainability, weak governance, and the isolated use of Artificial Intelligence in Business Process Management. This study initially conducts a Systematic Literature Review on the topic of business process improvement enabled by Large Language Models or Artificial Intelligence in organizations, presenting a comprehensive analysis of prevailing research trends, conceptual frameworks, and persistent limitations, identifying seventeen recurring gaps that affect the effectiveness of integrating the capabilities of Large Language Models and other Artificial Intelligence technologies throughout the entire lifecycle of Continuous Business Process Improvement. As a result, we propose a Framework and its gap-oriented reference architecture that, through modular components, facilitates data integration, reasoning, validation, execution, and monitoring within a closed loop of continuous business process improvement. The framework is operationalized through six phases: Process Understanding, Process Diagnosis, Process Redesign, Process Validation, Process Execution Support, and Continuous Monitoring. The results suggest that designing the framework and architecture directly from the identified gaps creates a coherent foundation for AI-driven process improvement, enabling more reliable, explainable, and easily governed and managed solutions. The study improves the current state of the art by creating a cohesive framework for intelligent, scalable, lifecycle-integrated, and operationally deployable process optimization systems.

INDEX TERMS Artificial intelligence, business process improvement, explainable AI, large language models, process intelligence.

I. INTRODUCTION

Business processes are collections of related structured activities or tasks that are used to generate a product or service [1]. They determine how the work should be done within an organization, becoming the building blocks of the organization. Consequently, business processes are vital to the operation of an organization. Due to their use, resources can be used more efficiently, products and services can be delivered on time, and quality standards are reached. Therefore, it is important for organizations to properly manage their business processes [1].

The associate editor coordinating the review of this manuscript and approving it for publication was Sedat Akleylek¹.

In this context, Business Process Management (BPM) is a systematic methodology to assess, plan, implement, monitor, and optimize business processes. It helps organizations streamline operations, minimize expenses, improve customer satisfaction, and ensure compliance while focusing on governance and strategy [2]. Within the BPM lifecycle, Business Process Improvement (BPI) is a systematic approach to help an organization optimize its underlying processes for more efficient results [1]. Currently, organizations are increasingly dependent on BPI to maintain efficiency, service quality, regulatory compliance, and adaptability in rapidly changing environments. In this context, with the advancement of Artificial Intelligence (AI) technologies, organizations have started to use these technologies, such as Large Language Models

(LLMs), as a boost to promote the continuous improvement of business processes [3].

A Continuous Business Process Improvement (CBPI) approach seeks to maintain and improve organizational performance [2], [4]. In real settings, CBPI depends on the combination empirical execution evidence (e.g., event logs, KPIs, historical baselines) with contextual and normative knowledge (e.g., policies, procedures, regulations, risk controls) to produce diagnostic insights and actionable redesign decisions [5], [6]. Recent advances in AI, especially LLMs when combined with Retrieval-Augmented Generation (RAG), bring new opportunities for BPM and BPI, since these models have the ability to handle unstructured data, create summaries of operational evidence and textual explanations that aid in decision-making processes, particularly when related to business process redesign. [7], [8]. However, LLM-CBPI initiatives frequently struggle with three interrelated challenges:

- 1) integration of heterogeneous structured and unstructured information into a coherent end-to-end improvement workflow [9], [10];
- 2) limited explainability, traceability, and governance of improvement recommendations—especially under compliance and risk constraints [11], [12], [13]; and
- 3) insufficient support for closed-loop operation, where ex ante validation of redesign alternatives and ex-post monitoring evidence systematically trigger subsequent improvement cycles [2], [14].

Moreover, LLMs applications related to process improvement are still fragmented in their focus on specific tasks. Existing studies often focus on isolated use cases, and little attention has been paid to how LLM-based capabilities can be systematically incorporated throughout the CBPI lifecycle. As a result, organizations can experiment with AI-enabled business process improvement tools without a coherent architecture that considers the end-to-end improvement process [15], [16], [17].

In this context, it is necessary to better understand this scenario. To this end, this study initially conducts a Systematic Literature Review (SLR) on research in AI-enabled process improvement. The SLR presents the current state of the art on this topic and identifies seventeen recurring gaps that limit the effective adoption of AI and LLMs in CBPI environments.

Based on the results of the SLR, this study also presents a conceptual framework, the LLM-CBPI Framework and its gap-oriented reference architecture, as a proposal to address the gaps identified in the SLR. The Framework consists of six integrated phases that support explainable, governance-driven, and closed-loop improvement cycles. The reference architecture is based on elements widely used and empirically demonstrated in multiple contexts, such as Human-in-the-Loop, LLM/RAG Pipeline, ETL/ELT Pipeline, and Process Discovery.

Developing the solution directly from the identified gaps creates a more coherent basis for AI-driven process improvement, as aligning improvement needs with architectural

decisions results in a more reliable, explainable, governable, and operationally manageable design. In this context, the main contributions of this article are the presentation of an SLR that shows the state of the art on AI-enabled BPI research and identifies seventeen recurring gaps in this topic; the presentation of a six-phase workflow framework that supports explainable, governance-driven, and closed-loop improvement cycles and its gap-oriented reference architecture to integrate LLMs into the CBPI lifecycle, addressing the gaps identified in the SLR.

II. BUSINESS PROCESS IMPROVEMENT

Business Process Management (BPM) helps organizations streamline operations, minimize expenses, improve customer satisfaction, and ensure compliance while focusing on governance and strategy. According to Dumas et al. [2], the key aspects of BPM are: Identifying and Documenting Existing Business Processes, Sing Technology to Automate Repetitive Tasks, Tracking Process Efficiency Using Key Performance Indicators (KPIs), and Enhancing Processes Through Methodologies such as Six Sigma, Lean, and Process Mining [2], [18].

BPM relies on data-driven optimization, employing Six Sigma to minimize variability [19], Lean to eliminate waste [20], and Process Mining to examine real-time operations [5]. To maintain improvements, enterprises should implement performance monitoring tools such as AI-driven predictive analytics, KPI tracking, and Business Activity Monitoring (BAM) dashboards to guarantee ongoing optimization and compliance [2], [15].

In this context, Business Process Improvement involves analyzing current processes, identifying areas for improvement and implementing changes to improve performance, quality, and reducing costs. BPI is a component of BPM that aims to refine business processes to make them more efficient (doing things right) and practical (doing the right things) [1]. Several methodologies and technologies are commonly used in BPI, each offering a unique approach to process improvement. In many cases, AI models are widely employed [9]. AI algorithms can analyze large datasets to recognize trends, forecast demand, and predict customer behavior [21].

This information is essential to the success of the decision-enabling process in order to anticipate market changes. Moreover, by sifting through large amounts of data, AI can provide actionable insights, helping organizations optimize operations and improve customer experiences [22].

In this sense, techniques such as Machine Learning (ML), Robotic Process Automation (RPA), Natural Language Processing (NLP), Process Mining, and LLMs have been widely used [21]. ML techniques can discern patterns in business processes, optimize workflows, predict customer behavior, forecast demand, and detect anomalies, whereas RPA automates repetitive tasks such as data entry, invoicing, and document processing, excelling with structured data and rule-based workflows [22]. RPA can be augmented with AI (Intelligent RPA) to facilitate decision making [9].

NLP technologies automate text-centric operations such as customer service chatbots and sentiment analysis, extract insights from unstructured data (emails, documents, and social media), and improve document classification and contract analysis [23]. Process Mining approaches employ AI to examine corporate workflows, identify inefficiencies, offer insights into bottle-necks and automation opportunities, and improve BPM [9]. LLMs, such as Generative Pre-trained Transformer 4 (GPT-4), Bidirectional Encoder Representations from Transformers (BERT), and Large Language Model Meta AI (LLaMA), have transformed BPI by improving automation, decision making, and customer engagement [24]. They excel at man-aging unstructured data, automating workflows, and enhancing human productivity [25].

III. LARGE LANGUAGE MODELS

Large Language Models are advanced AI models created to comprehend, produce and modify human language [26]. LLMs such as GPT-3, GPT-4, and Pathways Language Model (PaLM) are based on Deep Learning (DL) architectures, particularly the Transformer Model, and use large volumes of text data to produce coherent text, predict word sequences, and even perform complex reasoning tasks [24]. Among their capabilities are Reinforcement Learning From Human Feedback (RLHF), which enhances performance and aligns responses with human preferences, and in-context learning, where models generate outputs tailored to specific prompts or stimuli [25].

Many aspects of daily life have changed as a result of the use of Large Language Models (LLMs) [24]. Language translation, sentiment analysis, personalized instruction, and virtual assistance are just a few of the many applications where LLMs have demonstrated utility [25]. In the workplace, LLMs help physicians diagnose illnesses and evaluate medical literature [27], while also streamlining corporate procedures and automating administrative duties [23]. In addition, they help software developers with coding and debugging, assist marketers with content creation and consumer trend analysis, and improve education through AI-driven tutoring and adaptive learning [24]. Finally, LLMs support banking services, investment analysis, contract drafting, and legal research in the legal and financial domains [27].

Recently, organizations have started using LLMs to improve business processes [3]. But to do so, it is necessary to have a well-structured process, understand and address the challenges and opportunities this approach presents [16]. These organizations also face several constraints, including the need for high-quality data and data privacy compliance, integration with legacy systems, and significant initial and ongoing costs [3]. Businesses also grapple with skill gaps that require specialized talent and training [28]. Algorithmic bias and the need for transparency are ethical concerns, along with regulatory complexity and uncertainty regarding liability for AI-driven decisions [16].

Consequently, to take advantage of the opportunities offered by these approaches, it is essential to understand their

characteristics and technical requirements [23], [24], as well as the challenges and restrictive aspects involved, including integration, cost, and data governance [29]. Furthermore, it is also important to analyze their advantages and disadvantages, such as the potential for automation vs. constraints in interpretability [30], and to identify opportunities for improvement and refinement, especially in the integration of LLMs with robotic and intelligent automation methodologies [31].

IV. RESEARCH PROBLEM

Taking into account the above, the introduction of LLMs has transformed several fields. Business processes often involve decision making under pressure, unstructured data processing, and repetitive duties. These fields offer a special opportunity for innovation because they closely match the strengths of LLMs [24], [32]. The potential of LLMs is a topic of growing interest among academics. Assessing LLMs is essential because it helps us understand their advantages, disadvantages, and effects on society and business.

Thus, we formulated the Main Research Question (MRQ) based on the “Population and its Problem, Intervention or Issue, Comparative Intervention, Results or Themes and Context” Framework proposed by Booth et al. [33]:

MRQ: In organizational contexts of BPI, how does the integration of LLM/RAG capabilities with process analytics, compared to traditional BPI approaches, ensure explainability, governance, and closed-loop effectiveness of BPI throughout its lifecycle? To answer the RQ, we first conducted a systematic literature review to investigate the current state of the art in BPI using LLMs in order to understand in which business domains and organizational processes LLMs have been used most efficiently to improve business workflows and task performance, the results that have been achieved by adopting LLMs to improve business processes workflows, and the main challenges for business sectors to adopt LLMs to improve business processes. We also identified a set of gaps that hinder the systematic implementation of LLMs in BPI, in order to enhance the explainability, governance, and effectiveness of a continuous closed loop in organizations. After that, we propose a conceptual framework and a reference architecture to address these gaps.

V. RELATED SURVEYS

In order to address the MRQ subject of this study, we first applied the method described in **Section VI** and a set of surveys was recovered from the most pertinent digital libraries: ACM Digital Library, Elsevier, Institute of Electrical and Electronics Engineers (IEEE), Science Direct, Scopus, Springer Link, Web of Science, and Wiley Online Library. In addition, we defined and applied the subsequent Assessment Criteria (AC) to assess surveys.

- AC-01 Business Process Coverage: Does the study encompass a comprehensive array of BPM tasks (e.g. automation, decision making, compliance)?
- AC-02 LLM Implementation Strategies: Does the re-search address optimal implementation strategies,

including fine-tuning, reinforcement learning, or hybrid AI methodologies?

- AC-03 Business-Specific Applications: Does the study examine the use of LLM across various business sectors (e.g. healthcare, banking, retail)?
- AC-04 Methodological Rigor: Does the study employ a systematic literature review or a structured research methodology?
- AC-05 Practical Recommendations: Does the study provide actionable insights for organizations that integrate LLMs in BPM?
- AC-06 Ethical and Regulatory Considerations: Does the study address obstacles related to AI governance, data privacy, and ethical issues?

No single survey was found that addresses the MRQ proposed in this study. In this regard, we found seven surveys that touch on the topic of our investigation in part: Vidgof and Bachhofner [34], Grohs et al. [17], Wang et al. [35], Kourani et al. [36], Zhao et al. [37], Kampik et al. [38], and Ziche and Apruzzese [39]. Therefore, **Tables 1 and 2** list the evaluation results.

TABLE 1. Comparative analysis of similar reviews (AC-01 to AC03).

Review	AC-01	AC-02	AC-03
[34]	✓		
[17]	Limited		✓
[35]	Limited	✓	
[36]	Limited	✓	
[37]	✓	✓	✓
[38]	✓		✓
[39]	Limited	✓	
This Study	✓	Reference Architecture	✓

TABLE 2. Comparative analysis of similar reviews (AC-04 to AC06).

Review	AC-04	AC-05	AC-06
[34]	Theoretical Model		✓
[17]	Experimental Study	✓	
[35]	Survey	Limited	✓
[36]	Structured Framework	✓	
[37]	General Review	✓	✓
[38]	Conceptual Framework		✓
[39]	PoC	✓	
This Study	SLR	Structured Framework	✓

A. OUR CONTRIBUTION

This work provides the most comprehensive cross-sectoral state-of-art of LLM-enabled BPI to date, addressing challenges, governance constraints, across 49 qualified studies, in contrast to previous reviews that focus on isolated tasks, domains, or conceptual discussions. Unlike previous studies that focus predominantly on isolated activities, the solution proposed in this study adopts an end-to-end lifecycle perspective, integrating the phases of understanding, diagnosis, redesign, validation, execution support and continuous monitoring of the business process.

The study delineates seventeen gaps that hinder the systematic implementation of LLMs in BPM and BPI and proposes two novel scientific contributions to address these gaps.

A gap-oriented approach was used both in the development of the framework and in its reference architecture. This approach provides strong traceability between evidence from the literature, identified needs, and design decisions, unlike previous studies that generally introduce AI features without a systematic design justification.

Another striking characteristic of this project is the integration of reasoning, governance, explainability, simulation, operational deployment, and monitoring components into a single modular architecture structure. This differs from most similar projects that present structures focused solely on analytical support. Similarly, opting for a closed-loop logical improvement structure, in which evidence monitoring continuously informs new cycles of diagnosis and redesign, the proposed solution allows for adaptive and iterative process improvement, rather than one-off interventions as identified in most of the analyzed studies.

Finally, this study contributes a practical and research-oriented basis for future empirical implementations of AI-enabled CBPI systems that are reliable, explainable, and scalable.

VI. DATA AND METHODS

An Systematic Literature Review (SLR) is a review process of existing research on a particular topic using a rigorous and comprehensive approach. It involves identifying, evaluating, and synthesizing all relevant studies to answer a specific research question or set of questions. Unlike a traditional literature review, which can rely on the knowledge and biases of the researcher, an SLR follows a predefined protocol that ensures a thorough and unbiased literature search. Conducting an SLR involves several steps, including defining the research question, searching for relevant studies using a predetermined set of criteria, assessing the quality of the studies and synthesizing the results [40], [41].

A. A SYSTEMATIC LITERATURE REVIEW METHODOLOGY

The protocol for conducting systematic reviews proposed by Kitchenham [40] was followed to carry out this study. The first step is to define the research question or topic that guides the entire review process. Next, a comprehensive search strategy is developed to identify all relevant studies. This includes searching multiple databases and using specific keywords and inclusion/exclusion criteria to narrow the results. After identifying all relevant studies, they undergo screening based on their abstracts and titles to determine if they meet the inclusion criteria. The full-text articles are then retrieved for the remaining studies and the quality and relevance are assessed. Finally, data are extracted from the selected studies and synthesized to answer the research question [40]. **Figure 1** depicts the SLR protocol proposed by Kitchenham [40] and followed to perform this review.

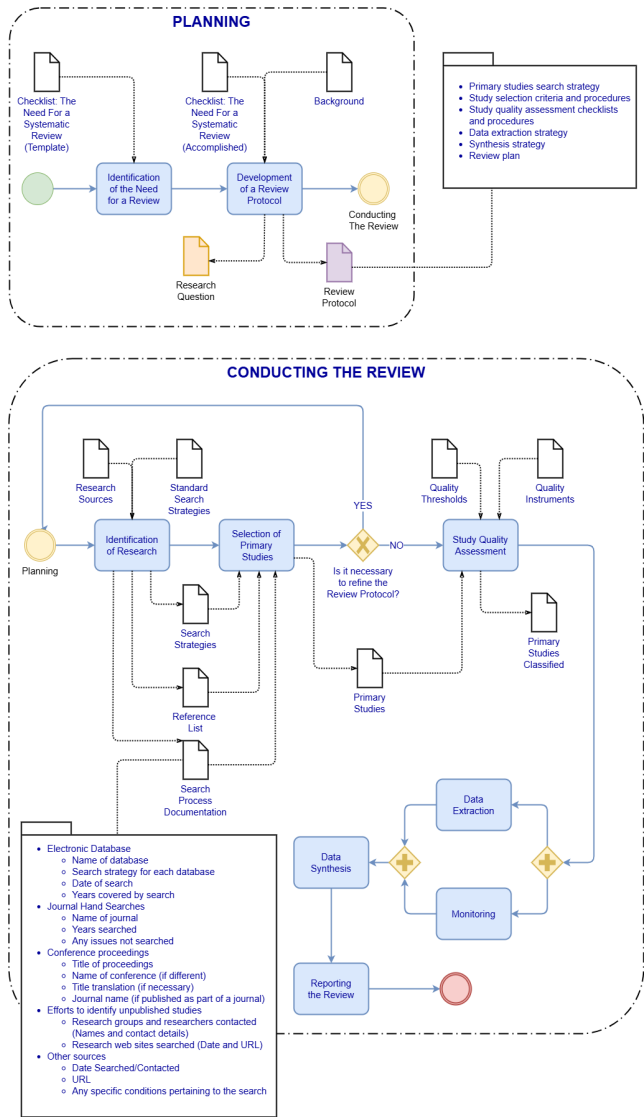


FIGURE 1. SRL protocol based on Kitchenham [40].

In order to do that, a search process was carried out by using Harzing’s Publish or Perish (Windows GUI Edition) 8.12.4612.8838 software [42], [43] to search the main available data sources such as ACM Digital Library, Elsevier, IEEE, Science Direct, Scopus, Springer Link, Web of Science, and Wiley Online Library, as well as the platforms of specific conference proceedings and journals. Figure 2 depicts the workflow of the search process.

In order to identify the publications to be analyzed, a search string was defined and a single search string algorithm was created using conventional Boolean operators to make the following query: (“Business Process Improvement” OR “Business Process Management” OR “Business Process Reengineering”) AND (“Large Language Models” OR “Generative AI” OR “Generative Artificial Intelligence”). When running the algorithm, we identified that the documents retrieved that directly reference LLMs date back

only to 2020. Consequently, we define a search period from 2020 to 2025.

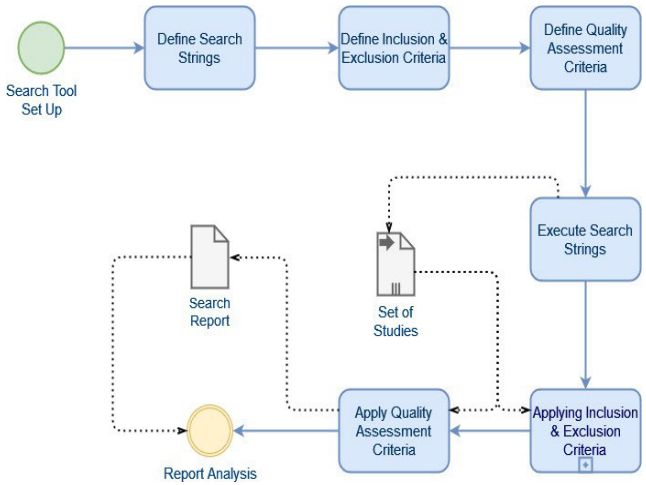


FIGURE 2. Workflow of the search process.

B. SCREENING AND ELIGIBILITY

Continuing to execute the workflows, the inclusion/exclusion criteria were used to filter documents, followed by the execution of the search string algorithm. The objective is to ensure that the selection process is systematic, unbiased, and transparent. Table 3 lists the Inclusion / Exclusion Criteria (IEC). Then each document was reviewed using these criteria in the order in which they are listed in Table 3. The objective is to ensure relevance and quality.

TABLE 3. Inclusion and exclusion criteria.

Id	Criteria	Description	Action
IEC-01	Type of study	Corrections and retractions	Exclude
IEC-02	Language	English	Include
IEC-03	Type of study	Articles published in conferences, proceedings, scientific journals, e-Book, bachelor thesis, master thesis, Ph.D dissertations or doctoral thesis	Include
IEC-04	Publisher	Articles whose publisher is a Potential Predatory Journal and Publishers according to Bell’s criteria for determining predatory open-access publishers [44], [45]	Exclude
IEC-05	Availability	Studies available for access by the public or through an academic account	Include
IEC-06	Topic correspondence	Studies that are not related to the topic under study	Exclude
IEC-07	Quality assessment	Studies that failed qualitative assessment	Exclude

As shown in Table 3, studies that met criteria IEC-01 to IEC-06 undergo a quality assessment (IEC-07) and the results are compiled into a search report. Studies that scored less than 6 out of 12 in the Quality Assessment Criteria (QAC) listed in Table 4 were rejected.

In order to proceed with the analysis, a three-step approach was used to read the research articles based on the method proposed by Keshav [46].

As a result, 534 documents were retrieved. After removing 34 duplicates, 28 studies did not meet IEC-01 (corrections and retractions). Consequently, 472 records were screened. After that, 13 studies were identified to not meet IEC-02 (non-English publications), 57 studies did not meet IEC-03 (non-academic sources), 87 studies did not meet IEC-04 (studies published by potential predatory journals or publishers [44], [45]) and 45 studies did not meet IEC-05 (studies not accessible by public or academic account). Finally, 128 studies did not meet IEC-06 (studies that do not correspond to the topic under study) and 93 studies did not meet IEC-07 (studies that did not meet qualitative assessment).

TABLE 4. Qualitative assessment criteria.

Id	Criteria	Score
QAC-01	Does the study have a clear and well-defined research question or hypothesis?	0 = Not clear; 1 = Partially clear; 2 = Well-defined
QAC-02	Is the study relevant to the research questions about the SLR?	0 = Not relevant; 1 = Somewhat relevant; 2 = Highly relevant
QAC-03	Were the study's findings adequately supported by the data?	0 = Not supported; 1 = Partially supported; 2 = Well supported
QAC-04	Were the limitations of the study acknowledged and discussed?	0 = Not mentioned; 1 = Briefly mentioned; 2 = Well discussed
QAC-05	Does the study provide sufficient detail to allow others to reproduce the research?	0 = Insufficient; 1 = Some details; 2 = Sufficient detail
QAC-06	Are the study procedures and analyses transparent and detailed?	0 = Not transparent; 1 = Somewhat transparent; 2 = Fully transparent

It important to note that due to the rapid evolution of research in the field of LLMs, the returned results include pre-publication sources, such as arXiv. Peer-reviewed sources were prioritized throughout the manuscript to ensure academic rigor and relevance. However, pre-publication sources were used selectively and only in emerging and specific topics where peer-reviewed evidence is still limited or where recent technical developments have not yet been fully reflected in indexed journal publications. It is noteworthy that whenever suitable peer-reviewed alternatives were available, references were re-evaluated and updated accordingly. **Figure 4** depicts the workflow of the systematic review process.

C. INCLUDED STUDIES

Since we want to understand how the integration of LLM/RAG capabilities with process analytics affects the explainability, governance, and closed-loop effectiveness of BPI, the set of Research Questions (RQ) listed in **Table 5** was formulated. The purpose of this study is to be able to answer these research questions.

Consequently, according to the research problem statement described in Section IV, the 49 selected studies underwent

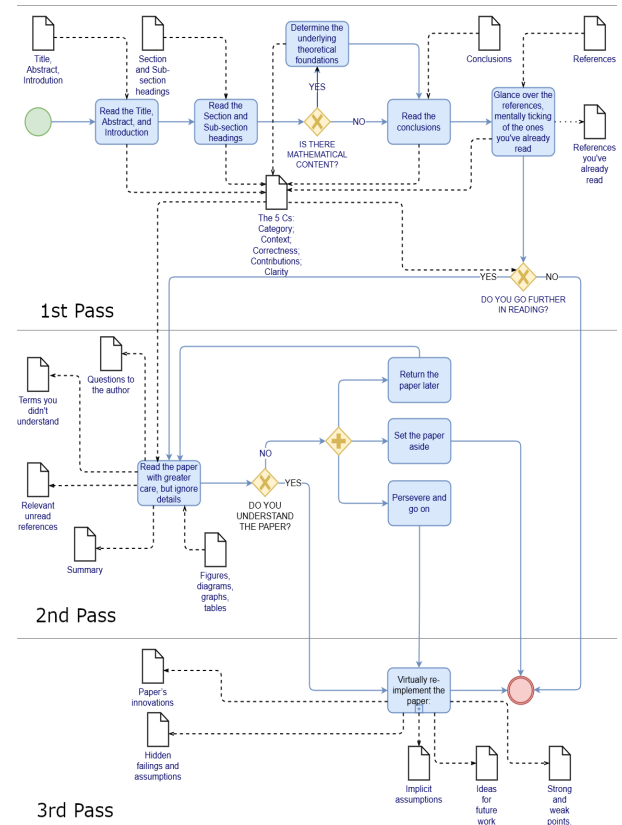


FIGURE 3. Three-step approach to reading research articles based on the method proposed by Keshav [46].

TABLE 5. Research questions.

Id	Research Question
RQ-01	In which business domains and organizational processes have LLMs been used most efficiently to improve business workflows and task performance?
RQ-02	What results have been achieved by adopting LLMs to improve business processes workflows?
RQ-03	What are the main challenges for business sectors to adopt LLMs to improve business processes?
RQ-04	What are the gaps in the effective integration LLM/RAG capabilities with Business Process Improvement, in order to enhance the explainability, governance, and effectiveness of the closed loop in organizations?

an assessment to respond to these RQ. Some selected studies address adjacent or peripheral themes related to AI-enabled process improvement. It is important to note that these studies were identified by the search strategy and predefined selection criteria, and not through manual reference selection. Although not all focus directly on the topic of AI-enabled continuous process improvement or LLM, they help to illustrate the current scenario more broadly where these AI-enabled continuous process improvement or LLM approaches are being adopted, which does not alter the main analytical focus of this study.

To obtain reliable and reproducible findings from the final set of 49 included studies, we used a dual analytical approach,

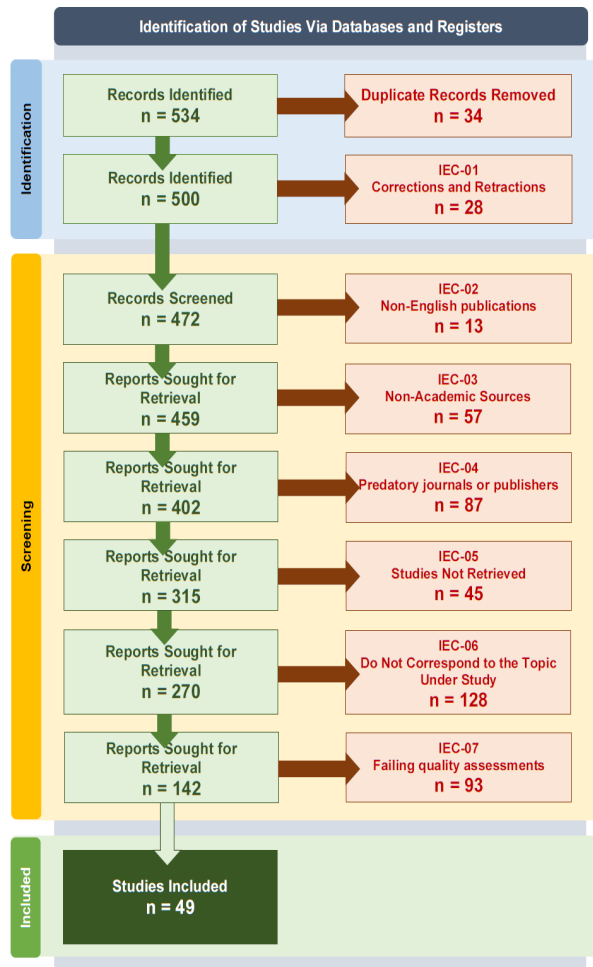


FIGURE 4. Workflow for the identification of studies.

integrating qualitative content analysis with bibliometric and scientific mapping methodologies. This hybrid methodology facilitated both conceptual integration and quantitative assessment of the literature environment.

D. QUALITATIVE CONTENT ANALYSIS

In order to analyze the text of the sources in the final sample to obtain the results, we compiled a systematic content analysis methodology from Mayring [47], Bengtsson [48], and Kuckartz [49] to rigorously analyze the complete text of each study.

This procedure encompassed three fundamental phases:

- 1) **Data Extraction:** Each study was assessed using a standardized extraction template corresponding to the four research questions (RQ01 - RQ04). Essential information was recorded, encompassing business domains, addressed processes, AI approaches (such as rapid engineering and RLHF), deployment strategies, empirical results, and recognized obstacles.
- 2) **Open Coding and Thematic Classification:** A manual open coding procedure was performed on pertinent content. The codes were developed iteratively and

classified into subject areas such as ‘Prompt Engineering for Semantic Enhancement’, ‘Hybrid AI Integration’, and ‘RLHF’. This inductive-deductive methodology guaranteed coherence with both emerging discoveries and established analytical frameworks.

- 3) **Cross-Case Synthesis:** Coded data were amalgamated to juxtapose findings across studies, discern prevailing trends, and differentiate outcomes within sectors or methodologies. Frequency counts were used, if applicable, to measure patterns (e.g., the number of studies utilizing LLMs in healthcare or implementing process mining).

This qualitative process facilitated a comprehensive analysis of the implementation of LLMs to enhance business processes while maintaining transparency and consistency.

VII. BIBLIOMETRIC ANALYSIS

Bibliometric analysis evaluates research output using quantitative methods to measure the impact of research in a particular field or discipline. It can evaluate the performance of individual researchers, departments, and institutions. The goal of bibliometric analysis is to provide information on trends and patterns of research output and to identify influential authors and publications [50], [51].

To assess and analyze the scientific domains of the included studies, the following indicators provided by SCImago [52] were used: Quartile, H-Index, and Scimago Journal Rank (SJR). These indicators aim to provide a more accurate and meaningful measure of the scientific prestige of the journal by considering various factors beyond simple citation counts [52]. Table 6 lists the overall dataset with the respective quartiles and H-index scores retrieved from SCImago [52] and Google Scholar [53], [54].

Donthu et al. [97] state that the main techniques for a bibliometric analysis are Performance Analysis and Science Mapping. According to the authors, while Performance Analysis considers the contributions of each research constituent, Scientific Mapping concentrates on the connections between research constituents. It involves visualizing and understanding the structure and dynamics of scientific knowledge [98]. In order to do that, the VOSviewer tool [99], [100], Google Scholar [53], [54], and Google Looker Studio [101] were used to create visual analytics dashboards based on the data recovered by the query algorithm presented in Section VI-A. Accordingly, we present below a series of indicators suggested by Donthu et al. [97], Guerrero-Bote and Moya-

Anegón [102], and SCImago [52], in order to perform the bibliometric analysis with the included studies.

A. PERFORMANCE ANALYSIS

Performance analysis aims to quantify the contributions and impact of these various entities in the research landscape. This helps identify the leading researchers, institutions and journals, as well as trends in research productivity and influence [98].

TABLE 6. Bibliometric data for the general dataset ordered alphabetically by article title.

Article	Quartile	H-index
Abbasi <i>et al.</i> [3]	Q1	96
Beheshti <i>et al.</i> [16]	Not available	29
Mendling <i>et al.</i> [15]	Q1	77
Vidgof <i>et al.</i> [34]	Not available	Not available
Grohs <i>et al.</i> [17]	Not available	Not available
Wang <i>et al.</i> [35]	Q1	78
Kampik <i>et al.</i> [38]	Q3	70
Naveed <i>et al.</i> [55]	Q1	38
Lee <i>et al.</i> [56]	Q1	77
Afrin <i>et al.</i> [57]	Q1	242
Du <i>et al.</i> [58]	Not available	Not available
Siderska <i>et al.</i> [59]	Not available	Not available
Casciani <i>et al.</i> [60]	Not available	24
Jameaba [61]	Not available	Not available
Nica <i>et al.</i> [62]	Q2	18
Novak <i>et al.</i> [63]	Q1	95
Bell and Bell [64]	Not available	Not available
Ilagan <i>et al.</i> [65]	Not available	132
Zeng <i>et al.</i> [66]	Not available	Not available
Feuerriegel <i>et al.</i> [67]	Q1	63
Banh and Strobel [68]	Q1	57
Redek and Godnov [69]	Q1	53
Grupac <i>et al.</i> [70]	Q1	17
Devgov <i>et al.</i> [71]	Q1	17
Nica <i>et al.</i> [72]	Q1	17
Makarov <i>et al.</i> [73]	Q1	125
Ke and Ng [74]	Q1	38
Poliak <i>et al.</i> [75]	Q2	18
Haase <i>et al.</i> [76]	Q2	59
Demir and Demir [77]	Q1	104
Kampik <i>et al.</i> [78]	Not available	Not available
Weinzierl <i>et al.</i> [79]	Q1	271
Gura <i>et al.</i> [80]	Q1	17
Schreiber <i>et al.</i> [81]	Q1	96
Frajt <i>et al.</i> [82]	Q2	18
Ayad and Alsayoud [83]	Q1	95
Ackermann <i>et al.</i> [84]	Not available	Not available
Nankya <i>et al.</i> [85]	Q2	243
Jain <i>et al.</i> [86]	Not available	Not available
Baier <i>et al.</i> [87]	Q1	95
Beerepoot <i>et al.</i> [88]	Not available	Not available
Atalan [89]	Q1	27
Sliz' [90]	Q1	27
Sharma <i>et al.</i> [91]	Q2	59
Ibrahim <i>et al.</i> [92]	Q1	50
Sharma [93]	Q3	20
Ceravolo <i>et al.</i> [94]	Q1	242
Norton <i>et al.</i> [95]	Q3	53
Bader <i>et al.</i> [96]	Q1	59

In order to do that, we analyzed the number of studies published yearly. Initially with little research activity in 2020 and 2022, when only one article was published each year, the data show a notable rise trend over time. Following this, there was a notable increase in 2023, with a peak of 21 studies, suggesting that the value of LLMs in improving business processes in organizations was becoming increasingly acknowledged. In 2024 there was consistent and increasing research activity in this area, as evidenced by the number of publications, which increased to 24 in 2024. This ongoing interest shows how relevant this topic is.

Continuing the performance analysis, the values of the indicators listed in **Table 6** were analyzed. This led us to deliver two analyzes: **Journal Data Analysis** and **Author H-Index**. These analyzes and indexes are important for understanding the landscape of academic publishing and for guiding research-related decisions.

1) JOURNAL DATA ANALYSIS

Journal Data Analysis refers to the examination and interpretation of data collected from journals. The quartile metric is used to categorize journals according to their relative position within a specific field. It means raking journals by impact using quartiles (Q1–Q4). The Q1 journals are the most prestigious, highly cited, and influential. The Q2 journals have strong impact and above-average citations. The Q3 journals have moderate influence, receiving fewer citations than Q1 and Q2. The Q4 journals are newer or highly specialized and have the lowest impact. Quartile rankings use metrics such as Impact Factor and SJR [103], [104].

As we can see in **Table 6**, most of the articles (35 articles - 71%) were published in a journal that belongs to a quartile and most of them (26 articles - 53%) were published in Q1 journals. 6 articles (12,2%) were published in Q2 journals and 3 articles were published in Q3 journals. It is important to note that the set of 14 studies that were published in unavailable quartile journals refers to articles presented at renowned international conferences whose proceedings were also published by renowned publishers such as Springer Nature, IEEE, Elsevier, ACM Digital Library, and arXiv.org, as stated in **Section VI-B**.

2) AUTHOR H-INDEX

The H-index is a metric used to measure productivity and impact of citation [104], [105]. Hirsch [105] states that the H-index evaluates the weight, significance, and overall influence of the total research contributions of a scientist. According to Hirsch [105], an H-index of 20 characterizes a successful scientist. An H-index of 40 characterizes outstanding scientists, and an H-index of 60 characterizes truly unique individuals.

As we can see in **Table 6**, in most of the analyzed studies (30 studies – 61.22%) at least one of the authors has an H-index greater than 20. In 23 studies (47%) at least one of the authors has an H-index greater than 40, and in 16 studies (32,7%), at least one of the authors has an H index greater than 60.

3) ANALYSIS OF AUTHOR AFFILIATION DATA

The analysis of the author affiliation involves examining information related to the affiliations of authors who have contributed to academic publications. This type of analysis can provide information on various aspects of research and academic collaboration [97], [106].

Consequently, analyzing the set of 49 articles that are the subject of this study, we can identify that the affiliations of

the majority of authors are from Germany (15), the United States of America (10), Romania (5), Slovakia (5), Australia (4) and the Netherlands (4).

4) CITATION ANALYSIS

A citation analysis examines how often other works have cited the article to determine its impact and influence. A higher citation count indicates a more significant influence and recognition within the field [97]. Data was recovered from the reliable and widely used Google Scholar search engine [53], [54]. The included articles have been referenced, ranging from 1930 to 0 times. Table 7 lists the ten most referenced articles.

TABLE 7. Top ten articles per citation count.

Article	#Citations
Wang <i>et al.</i> [35]	1930
Naveed <i>et al.</i> [55]	1637
Feuerriegel <i>et al.</i> [67]	531
Mendling <i>et al.</i> [15]	236
Banh and Strobel [68]	180
Beerepoot <i>et al.</i> [88]	145
Vidgof <i>et al.</i> [34]	77
Du <i>et al.</i> [58]	75
Grohs <i>et al.</i> [17]	57
Demir and Demir [77]	53

5) CONCLUSION

Performance analysis reveals that the research landscape is marked by rapid expansion and significant scholarly credibility. The significant increase in publication volume after 2022 indicates that LLM-driven process improvement has evolved from an emerging topic to an established research focus. The prevalence of publications in Q1 journals and the significant proportion of contributions by authors with high H-index values offer strong evidence of intellectual maturity and academic credibility in the discipline.

The geographical distribution of institutional affiliations indicates a concentration of research effort in technologically sophisticated countries, suggesting that the field is now influenced by a limited number of innovation ecosystems. Citation patterns further underscore the field’s prominence, with numerous works exhibiting remarkably high effect. Nonetheless, the broad range of citations indicates uneven dissemination of knowledge and disparate levels of consolidation within subtopics.

In summary, these identified robust performance metrics provide substantial evidence that the investigated research is credible and of high academic quality.

B. SCIENCE MAPPING

According to Donthu et al. [97], scientific mapping visualizes and analyzes the relationships and interactions between these entities to understand the structure and evolution of scientific knowledge. This helps in understanding the intellectual structure of a field and identifying core topics, emerging trends,

and potential areas for future research. Therefore, we present the analysis of the relationships and interactions between the studies included below: Keyword Co-occurrence Analysis.

1) KEYWORDS CO-OCCURRENCE ANALYSIS

In research and bibliometric analysis, a keyword is a term or phrase that captures the essential topics, concepts, or themes of a research paper, article, or document [97]. Using VOSviewer [99], [100], a visual and analytical perspective was created on the interrelationships of various concepts and Keywords addressed in the included studies. Figure 5 depicts the visualization of the co-keyword network based on occurrences, and Figure 6 depicts the visualization of the co-keyword network based on occurrences and average publication per year scores between 2022 and 2025.

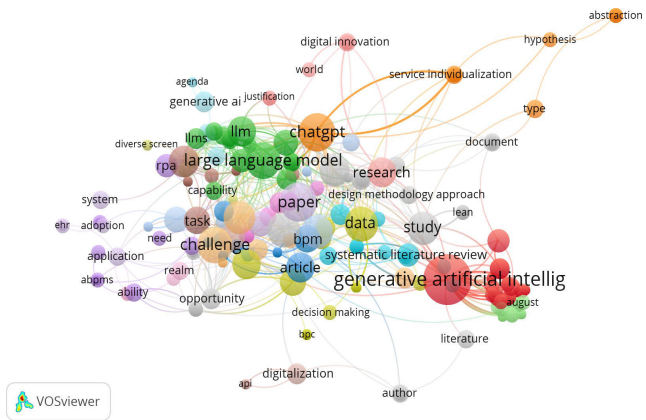


FIGURE 5. Co-keyword network visualization, based on occurrences.

As depicted in Figure 5, there are six keyword clusters based on their patterns of co-occurrence. Taking into account this image, it is evident that the centrality of Generative AI and LLMs (Green & Red Clusters) suggests that these subjects predominate in the dataset. Strong ties between AI and digitalization (Yellow Cluster) and BPI (Purple Cluster) illustrate the use of AI in automation and BPI. The prominent positions of “Generative AI” and “LLMs” demonstrate how these subjects are propelling innovation and research. The links to “decision making”, “BPM”, and “service individualization” suggest that AI’s potential effects on digital transformation and corporate efficiency are being investigated.

In contrast to Figure 5, Figure 6 shows the average year of keyword occurrence with a color-gradient bar in the bottom right corner. Green Nodes correspond to terms that were popular throughout 2023, Blue or Purple Nodes reflect themes that were more popular in early 2023 but have since declined in prominence, and Yellow Nodes represent the most recent topics, mostly occurring in 2024. “Generative AI”, “LLMs”, and “ChatGPT” continue to be the most related and important subjects. Nevertheless, more recent conversations on “abstraction”, “hypothesis”, “service individualization”, and “application” shown by Yellow Nodes are beginning to emerge, pointing to a move toward more

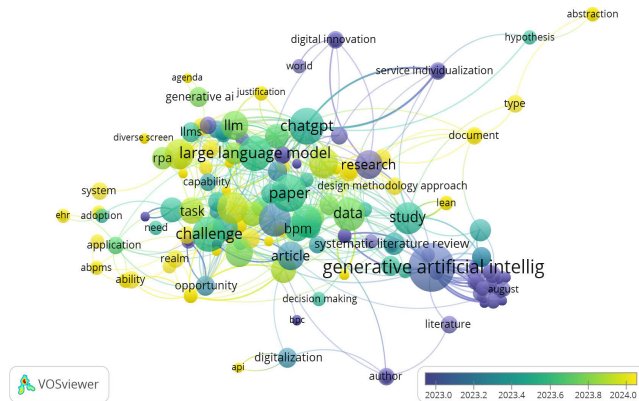


FIGURE 6. Co-keyword network visualization, based on the occurrences and average publication year scores between 2020 and 2025.

theoretical and applied research. Meanwhile, expressions such as “RPA”, “BPM”, and “decision-making” appear in Green and Yellow Nodes, suggesting that more people are paying attention to the role of AI in business process automation and BPI. Furthermore, the associations found between the expressions “digitalization”, “API”, and “automation” imply that the incorporation of AI into software systems and business procedures is becoming increasingly significant.

2) CONCLUSION

The co-keyword analysis reveals that Generative AI and LLM serve as the conceptual core of the research landscape, organizing subject clusters and inter-topic connections. Their prominence indicates a domain propelled more by technological advancement than by completely established theoretical underpinnings. The temporal overlay indicates a shift from an initial experimental topic to more contemporary applications and abstraction-focused subjects, reflecting a progressive development of the field. The continued relevance of fundamental concepts like BPM, RPA and decision-making in recent developments indicates that the incorporation of AI into established process frameworks is still an unresolved research challenge rather than a definitive paradigm change. The evidence indicates a rapidly consolidating, yet structurally developing, knowledge space, marked by robust technological influences, growing multidisciplinary convergence, and a nascent focus on operationalization and empirical validation.

C. OVERALL CONCLUSIONS

The information synthesized from the performance analysis and science mapping indicates that the research domain has advanced beyond initial emergence and is now transitioning into a phase of systematic scientific consolidation. The observed increase in publication output, along with the prevalence of contributions in high-impact venues authored by highly cited scholars from prestigious institutions, suggests that the field has attained significant academic legitimacy and intellectual authority. Simultaneously, the conceptual

frameworks revealed by keyword co-occurrence analysis indicate that Generative AI and Large Language Models have emerged as crucial organizational constructs inside the knowledge network, influencing thematic clustering and inter-topic connectedness. This pattern indicates a shift from exploratory technological discussion to a more established and conceptually robust research environment. It is evident that this is a rapidly expanding field, however, with limited structural consolidation. The rapid diffusion of generative AI tools from 2023 onward spurred a significant increase in publications on the subject in that same year. However, although much research exists on the topic, the predominant research focuses on task-specific applications, but the development of integrative frameworks is not evolving at the same pace.

The alignment of quantitative maturity indicators with conceptual stabilization is a significant hallmark of disciplinary creation. The findings reveal a cohesive intellectual trajectory marked by the growing integration of AI capabilities with BPI frameworks, rather than only showcasing isolated research endeavors. Residual asymmetries, including geographic concentration and unequal citation dispersion, are evident. However, these traits align with the rapid evolution of scientific fields and indicate structural growth dynamics rather than conceptual fragmentation.

The fact that we identify a concentration of studies in the areas of finance, health, and supply chains indicates that initial adoption is stronger in highly regulated and data-intensive sectors. This is because the speed of decision-making and the complexity of the process create opportunities for immediate value. On the other hand, the recurring gaps identified related to explainability, governance, and closed-loop monitoring suggest that organizational control mechanisms have not evolved at the same pace as AI capabilities. Finally, a very limited amount of evidence is identified that accompanies a solution developed over time in real-world environments, indicating that the field remains largely exploratory, with a great need for real-world validation.

VIII. RESULTS

This study examined 49 pertinent publications chosen from an initial set of 534 based on their rigor, relevance, and contribution. The report provides a detailed examination of the impact of LLMs on BPI across various sectors and methodological frameworks. Although most of the studies demonstrate a rigorous methodology, some reveal constraints on sample size, generalizability, or the absence of longitudinal data.

Numerous sectors have made extensive use of traditional BPM models, including Business Process Model and Notation (BPMN), Six Sigma, and Process Reengineering. On the other hand, the advent of LLMs presents new approaches to business process automation, analysis, and improvement.

This study aims to understand these new approaches and understand how the integration of LLM/RAG capabilities with process analytics improves the explainability, governance, and closed-loop effectiveness of BPI across its

lifecycle. Furthermore, we aim to understand the characteristics of these approaches used and discover their unresolved issues. Consequently, taking this into account, we can now answer the RQs that are the subject of this study.

RQ-01: In what business domains and organizational processes have LLMs been used most efficiently to improve business workflows and task performance?

LLMs have improved corporate processes and productivity in several business domains. The findings indicate that LLMs applications are most prevalent in financial services (17% of total citations), particularly in the organizational processes of risk assessment, fraud detection, and compliance automation. The healthcare sector ranks second (15% of total citations), using LLMs for workflow automation, medical diagnostics, and patient care optimization. Supply chain management (13% of total citations) employs LLMs for demand forecasting and inventory management, while Customer Service (13% of total citations) improves service automation, inquiry management, and user interactions.

Other business domains include manufacturing (11% of total citations), Retail & E-commerce (11% of total citations), public sector (9% of total citations) and education (6% of total citations) using LLMs for document processing, regulatory automation, and service optimization.

RQ-02: What results have been achieved by adopting LLMs to improve business processes workflows? Considering the set of studies analyzed, the findings indicate that LLMs markedly improve efficiency, reduce costs, and facilitate decision making in BPI. Businesses indicate a 30–50% improvement in automation efficiency, leading to a reduction in process cycle durations and bottlenecks, coupled with a 40% reduction in human processing time, particularly in document-intensive workflows. AI-driven automation reduces operating expenses by 20–30%, decreasing manual labor and improving resource efficiency.

The precision of the forecast has improved by 30–40%, improving supply chain management, market analysis, and financial planning. AI-assisted compliance improves regulatory conformance, while chatbot-driven customer service has improved resolution accuracy by 25%, hence increasing customer satisfaction and decreasing dependence on human agents.

In addition to efficiency, LLMs improve agility and adaptability, facilitating dynamic AI-driven workflows that adapt to market fluctuations. Enhanced user-friendliness and workforce participation further facilitate uptake.

RQ-03: What are the main challenges for business sectors to adopt LLMs to improve business processes?

Considering the set of studies analyzed, the findings indicate that LLMs adoption is complicated by data protection restrictions (e.g., General Data Protection Regulation - GDPR), high implementation costs, and information technology integration issues, especially with outdated systems. As employees fear job loss, AI adoption is hindered by work-force opposition and upskilling. Bias in training data

raises ethical and legal problems about AI-driven conclusions.

In the same way, the calculation of the Return on Investment (ROI) is complicated and varies by business and use case. High-performance computers raise operational expenses, and LLMs management requires AI competence. In regulated industries such as healthcare and banking, clear AI choices are essential. The adoption of LLMs sometimes requires redesigning workflows, which complicates operations. Although LLMs improve corporate operations, its application requires combining efficiency with technical, financial, and ethical constraints.

These findings repeatedly underscore the substantial advantages that LLMs confer on organizational operations, such as increased automation efficiency, cost savings, higher customer satisfaction, compliance, and improved decision making. The advantages are especially evident in business sectors such as banking, healthcare, and supply chain management, where activities such as risk assessment, fraud detection, medical diagnostics, and customer service are data-intensive and depend on rapid information processing.

Notwithstanding its benefits, the implementation of LLMs frequently exposes enduring problems. Challenges include compliance with data privacy, model interpretability, integration with legacy systems, elevated computing expenses, and labor opposition are frequently noted in the literature. These limits highlight the need to create governance structures and technical techniques to mitigate risks while optimizing benefits.

RQ-04: What are the gaps in the effective integration of LLM/RAG capabilities with Business Process Improvement, in order to enhance the explainability, governance, and effectiveness of the closed loop in organizations?

Taking into account the findings described above aligned with existing specialized literature, we extracted 17 gaps that organizations must address to successfully integrate LLM/RAG capabilities with process analytics, in order to enhance the explainability, governance, and closed-loop effectiveness of BPI across its lifecycle.

To identify gaps, reports and information on capability deficiencies, limitations, and contextual problems related to LLM-enabled BPM approaches were investigated in the 49 selected studies. After extracting this information, recurring patterns were identified in the studies analyzed. Finally, these patterns were consolidated into the 17 gaps presented below. This ensures the use of multiple sources of empirical or conceptual evidence to provide adequate support for architectural decisions.

It is important to note that this mapping process is based much more on representative evidence than exhaustive evidence. This means that the gaps are not identical to those presented in the 49 studies analyzed, nor do they have the same probative force. While some gaps are directly supported by a subset of the analyzed studies, others are only partially derived from those

studies and result from refinements of design abstractions informed by the literature rather than directly specified by it.

Gap-01: End-to-end BPI Lifecycle Coverage

Process improvement constitutes a cyclical and dynamic lifecycle encompassing discovery, diagnosis, redesign, implementation, and monitoring. The BPM literature regularly highlights lifecycle integration as essential for enduring organizational learning and performance improvements [2], [4]. Disjoint or phase-isolated enhancements typically yield localized optimizations instead of a comprehensive advancement. Addressing the entire BPI lifecycle ensures methodological consistency and facilitates closed-loop learning, a fundamental characteristic of adaptive companies [3], [15], [88].

Gap-02: Heterogeneous Data Ingestion and Access

Digital processes produce data from various sources, such as event logs, structured records, papers, and policy repositories. Studies in analytics-driven businesses indicate that the integration of diverse data enhances decision-making quality and contextual understanding [6]. Process mining illustrates that the integration of operational logs with contextual knowledge yields enhanced diagnostic insights [5]. Consequently, extensive data input is essential for evidence-based BPI [3], [69].

Gap-03: Separation of Concerns and Modular Decomposition

Complex socio-technical systems are enhanced by modular architectures that allow analytics, reasoning, governance, and execution components to develop independently. Fundamental software engineering theory posits that modular breakdown enhances maintainability, scalability, and robustness [107]. A distinct separation and modular decomposition diminish the connection, enhance maintainability, and facilitate the independent development of data pipelines and AI components [16], [60].

Gap-04: Event-log Readiness and Empirical Process Evidence

Process mining research identifies event logs as objective, empirical depictions of process activity, facilitating impartial diagnostics and compliance analysis. Evidence-based enhancement rooted in actual operational data diminishes dependence on subjective perceptions and increases the precision of intervention [5]. It depends on the event logs as the primary source of behavioral evidence for observed executions [63], [94].

Gap-05: Process Model Management and Traceability

Process models serve as the formal foundation for BPM. Robust model governance ensures version control, evidence traceability, and compliance with execution environments. Traceable models allow organizations to understand the relationship between decisions, previous baselines, and prospective alterations [2]. BPI needs clear representations of the process and the regulated evolution of the model over time. [17], [36], [83].

Gap-06: Diagnostic Insight Generation and Interpretability

Diagnostic analytics must be comprehensible to facilitate actionable insights. Studies on explainable analytics and decision-support indicate that stakeholders are more inclined to believe and act on findings they comprehend [2], [5], [12]. Interpretability connects technical evidence with management logic, making AI-supported diagnostics operationally effective. This facilitates the prioritization of initiatives [3], [63], [89].

Gap-07: Improvement Reasoning and Alternative Generation

Decision theory posits that superior decisions result from the examination of several alternatives. In BPM, redesign is fundamentally innovative and advantages from methodical examination of enhancement scenarios [6]. AI-assisted reasoning increases this exploratory capability [2]. An effective CBPI needs prescriptive reasoning instead of purely descriptive reporting [28], [66].

Gap-08: Human-in-the-Loop Accountability and Approval Gates

Socio-technical research emphasizes that human supervision mitigates automation bias and ensures ethical and organizational accountability [108], [109]. Research on Responsible AI emphasizes the need for explicit human accountability in automated decision-making systems [74], [76].

Gap-09: Ex-ante Simulation and Validation of Redesigns

Simulation is a long-established method in BPM for assessing the potential impact of process changes prior to deployment [2], [110]. Ex-ante evaluation mitigates implementation risk and facilitates evidence-based selection of redesign alternatives [36], [39].

Gap-10: Operational Implementation and Integration with Enterprise Systems

Operational implementation of process improvement is essential for value creation. Research on BPM and information systems integration indicates that the alignment of models and enterprise systems is essential for enduring digital transformation [2], [111]. Improvement is only effective when it can be implemented and assimilated into operational systems [57], [59].

Gap-11: Continuous Monitoring and Closed-loop Feedback

The theory of continuous improvement, specifically the PDCA cycle, identifies feedback as the fundamental mechanism for ongoing learning and improvement of performance [14]. BPM monitoring strategies apply this notion to digital contexts [2], [5]. CBPI needs a closed loop that juxtaposes observed outcomes with anticipated behavior and baselines [3], [87].

Gap-12: Historical Baselines and Comparative Evaluation

Performance assessment needs historical benchmarks to differentiate genuine improvement from inherent variability.

Research in evidence-based management underscores the importance of baseline comparison in analytics [2], [5], [6]. In this sense, baselines are essential to illustrate improvement effects and mitigate confounding variables [63], [94].

Gap-13: Explainability, Traceability, and Auditable Reasoning (LLM-grounded)

Research on explainable AI indicates that transparent and traceable reasoning enhances accountability and trust in automated systems [11], [112]. Auditable reasoning is crucial in organizational and regulatory frameworks. [34], [93].

Gap-14: Knowledge Freshness and Adaptability Without Retraining

Organizational knowledge evolves rapidly, and a retrieval-based grounding facilitates quick updates while maintaining constant model parameters. Retrieval-augmented reasoning enables AI systems to integrate current knowledge dynamically without the need for retraining [8]. Adaptability enhances relevance in dynamic organizational contexts. [16], [60].

Gap-15: Compliance and Risk Constraints Enforcement

The literature on risk governance emphasizes that innovation should function within established limitations to prevent unmanaged uncertainty [113]. Incorporating compliance checks ensures adherence to institutional regulations [114], [115]. CBPI in regulated settings must maintain compliance and risk awareness, ensuring that improvements do not create intolerable exposures [85], [93].

Gap-16: Security, Privacy, and Access Control

AI-driven CBPI systems handle critical operational data, becoming essential security features. Data processing frequently involves personal and confidential information; governance mandates the principle of least privilege and regulated exposure [116], [117], [118], [119]. Research in information security underscores that data protection and regulated access are essential for reliable information systems [85].

Gap-17: Reproducibility and Operational Reliability

Scientific integrity is contingent on reproducibility. Systems research posits that reproducible pipelines enhance transparency and institutional confidence [120]. Dependable operation is crucial for the long-term sustainability of CBPI [109], [121]. Operational discipline is essential for managing technical debt and maintaining quality in production ML systems over time [73].

So, in order to ensure the traceability and interpretability of the information presented here, **Table 8** presents the mapping between the identified gaps and their respective literature support.

IX. DISCUSSION

The incorporation of LLMs signifies a transformative shift from conventional BPM models, providing AI-driven flexibility and scalability that exceed the static or manual characteristics of existing methods. Once we are interested in comparing traditional BPI approaches with approaches using

TABLE 8. Literature support for the identification of research gaps.

Gap	Foundational Support	SLR Support (49 Studies)
Gap-01	Dumas <i>et al.</i> [2]; vom Brocke <i>et al.</i> [4]	Mendling <i>et al.</i> [15]; Beerepoot <i>et al.</i> [88]; Abbasi <i>et al.</i> [3]
Gap-02	Davenport and Harris [6]; van der Aalst [5]	Redek and Godnov [69]; Abbasi <i>et al.</i> [3]
Gap-03	Bass <i>et al.</i> [107]	Beheshti <i>et al.</i> [16]; Casciani <i>et al.</i> [60]
Gap-04	van der Aalst [5]	Novak <i>et al.</i> [63]; Ceravolo <i>et al.</i> [94]
Gap-05	Dumas <i>et al.</i> [2]	Ayad and Alsayoud [83]; Kourani <i>et al.</i> [36]; Grohs <i>et al.</i> [17]
Gap-06	Miller [12]; van der Aalst [5]; Dumas <i>et al.</i> [2]	Abbasi <i>et al.</i> [3]; Novak <i>et al.</i> [63]; Atalan [89]
Gap-07	Davenport and Harris [6]; Dumas <i>et al.</i> [2]	van Dun <i>et al.</i> [28]; Zeng <i>et al.</i> [66]
Gap-08	Shneiderman [108]; Amershi <i>et al.</i> [109]	Ke and Ng [74]; Haase <i>et al.</i> [76]
Gap-09	Banks <i>et al.</i> [110]; Dumas <i>et al.</i> [2]	Kourani <i>et al.</i> [36]; Ziche and Apruzzese [39]
Gap-10	Weske [111]; Dumas <i>et al.</i> [2]	Afrin <i>et al.</i> [57]; Siderska <i>et al.</i> [59]
Gap-11	Deming [14]; Dumas <i>et al.</i> [2]; van der Aalst [5]	Abbasi <i>et al.</i> [3]; Baier <i>et al.</i> [87]
Gap-12	Davenport and Harris [6]; Dumas <i>et al.</i> [2]; van der Aalst [5]	Novak <i>et al.</i> [63]; Ceravolo <i>et al.</i> [94]
Gap-13	Doshi-Velez and Kim [11]; Rudin [112]	Vidgof <i>et al.</i> [34]; Sharma [93]
Gap-14	Lewis <i>et al.</i> [8]	Beheshti <i>et al.</i> [16]; Casciani <i>et al.</i> [60]
Gap-15	Power [113]; ISO 31000 [114]; COSO ERM [115]	Sharma [93]; Nankya <i>et al.</i> [85]
Gap-16	Anderson [116]; ISO 27001 [122]; ISO 27701 [119]	Nankya <i>et al.</i> [85]
Gap-17	Saltelli <i>et al.</i> [120]; Breck <i>et al.</i> [121]; Amershi <i>et al.</i> [109]	Makarov <i>et al.</i> [73]

LLMs, it is important to understand the scope, methods, advantages, and disadvantages of using traditional BPM models with the approaches presented above using LLMs.

A. COMPARISON OF TRADITIONAL APPROACHES TO USING BPM WITH APPROACHES USING LLMs

Starting with **BPMN**, it is a standardized notation that visually represents workflows, facilitating process standardization and effective communication between stakeholders [2], [123]. Nevertheless, it requires manual updates and is not adaptable in real-time [2]. Conversely, LLMs systematically evaluate and improve workflows, pinpointing inefficiencies within unstructured data, such as emails and documents [23], [24]. Although LLMs automate and improve processes, they lack BPMN’s structured visual representation, making them less suitable for standardized documentation [25].

Six Sigma employs a structured, data-driven DMAIC (De-fine, Measure, Analyze, Improve, and Control) framework to reduce errors and enhance efficiency, with a focus on statistical analysis and specialized teams [124]. However, it involves in-depth study, iterative improvements, and challenges with unstructured data [16]. In contrast, LLMs assess massive datasets instantaneously, adjust dynamically,

and anticipate optimal process parameters without requiring human interaction [24], [26]. Whereas Six Sigma ensures precision and control, LLMs provide automation, speed, and flexibility, making them more suitable for dynamic business environments [25]. Consequently, highly regulated sectors benefit greatly from the greater structure and rigor of Six Sigma [124]. However, LLMs provide more automation, speed, and flexibility, especially in dynamic commercial settings [24].

Finally, **Business Process Reengineering (BPR)** fundamentally restructures workflows to achieve significant performance improvements, enhance customer experience, and match operations to strategic objectives [20]. Nevertheless, it faces prolonged adoption, substantial expenses, and institutional resistance [125]. Conversely, LLMs improve processes incrementally, modifying workflows in real time without significant reorganization [23], [24]. BPR facilitates strategic transformation, whereas LLMs automates and improves current processes [24]. In this sense, LLMs offer continuous AI-driven optimization, while BPR is strategic and disruptive [25]. Businesses can combine the two strategies, using LLMs to support and maintain BPR improvements [126]. **Table 9** presents a comparative summary of LLMs vs. traditional BPM models.

TABLE 9. Comparative summary of LLMs vs. traditional BPM models.

Aspect	BPM	Six Sigma	BPR	LLMs
Approach	Visual workflow	Data-driven optimization	Radical redesign	AI-Driven process automation
Speed	Slow (manual updates)	Moderate (data analysis required)	Slow (strategic transformation)	Fast (real-time AI processing)
Automation	Low	Moderate (statistical modeling)	Low	High
Flexibility	Low	Moderate	Low	High
Real-Time Adaptation	No	Limited	No	Yes
Complexity	High (requires process expertise)	High (requires data expertise)	Very high (requires business transformation)	Moderate (requires AI governance)
Best Use Case	Process documentation	Process quality control	Business transformation	Process automation and continuous optimization

The findings indicate that LLMs can significantly enhance operational efficiency, especially in business sectors that are dependent on document-intensive procedures, consumer engagement, or data-driven compliance.

Nonetheless, real application requires addressing many challenges:

- **Technical Readiness:** organizations must allocate re-sources towards high-performance computing, secure data pipelines, and AI governance structures.

- **Workforce Transformation:** successful implementation relies on reskilling personnel, promoting AI knowledge, and alleviating concerns about job displacement.
- **AI Governance:** addressing bias, ensuring explainability, transparency, and adhering to standards such as GDPR are imperative prerequisites for sustainable AI integration, as well as for ensuring the ethics and compliance of BPI.
- **Integration:** achieving an adequate level of compatibility with current BPM tools and enterprise systems presents a significant challenge that requires both technological solutions and change management strategies.

In summary, LLMs do not replace standard BPM models. Instead, they enhance them with flexibility, adaptability, and automation. A hybrid methodology that uses the advantages of BPMN for visualization, Six Sigma for quality assurance, BPR for strategic transformation, and LLMs for AI-enhanced optimization presents the most promising path forward. As organizations increasingly pursue agility in unpredictable marketplaces, the astute integration of various techniques will characterize the forthcoming evolution of process management.

B. EVIDENCE STRENGTH AND LIMITATIONS OF THE REVIEWED LITERATURE

Although the analyzed studies report promising results regarding the use of LLMs and AI techniques to improve business processes, the strength of the available evidence is heterogeneous and should be interpreted with caution. The analysis combines conceptual articles, prototype demonstrations, specific case implementations, research, simulation-based studies, and empirical evaluations with varying levels of methodological rigor. As a result, not all reported benefits have the same degree of empirical support.

While some studies present controlled experiments or use cases, others propose conceptual frameworks, proof-of-concept systems, or exploratory discussions. Furthermore, the analyzed studies are distributed across multiple sectoral contexts, which demonstrates broad applicability but also limits generalization, since the results observed in one do-main may therefore not be directly transferable to another.

It is also important to note that some inconsistencies were observed regarding the sample size and scope of the evaluation. While several studies rely on restricted datasets, limited process instances, pilot implementations or short-term evaluations; long-term evaluations, large-scale deployments, and comparative studies are still scarce. This restricts the ability to infer sustained performance effects over time.

Finally, the level of empirical validation is uneven. While some studies provide measurable operational results, others infer expected benefits from technical feasibility or expert judgment.

In summary, the literature can be considered promising but is still in the process of maturing.

C. ENSURING EXPLAINABILITY, GOVERNANCE, AND EFFECTIVENESS IN A CLOSED-LOOP BPI THROUGHOUT ITS ENTIRE LIFECYCLE

Considering the above, we can now answer the Main Re-search Question that is the object of this study: **In organizational contexts of BPI, how does the integration of LLM/RAG capabilities with process analytics, compared to traditional BPI approaches, ensure explainability, governance, and closed-loop effectiveness of BPI throughout its lifecycle?**

Workflows can be effectively automated, optimized, and improved with the help of LLMs, which increase the intelligence, adaptability, and data-drivenness of business processes. Therefore, in order to ensure explainability, governance, and effectiveness in a closed-loop BPI throughout its entire lifecycle, organizations must address the 17 gaps described in **Section VIII**. In order to do that, we propose a modular and extendable six-phase conceptual LLM-driven framework for Continuous Business Process Improvement (LLM-CBPI Framework) for application in organizations. **Figure 7** depicts the LLM-CBPI Framework.



FIGURE 7. LLM-CBPI framework.

1) LLM-CBPI FRAMEWORK

As can be seen in **Figure 6**, this framework has six phases:

- 1) Process Understanding;
- 2) Process Diagnosis;
- 3) Process Redesign;
- 4) Process Validation;
- 5) Process Execution Support; and
- 6) Continuous Monitoring.

1. Process Understanding

This phase emphasizes the development of a detailed and organized depiction of the existing (“As-Is”) business process. LLMs facilitate the extraction of procedural information from various sources, including papers, procedures, commu-

nications, narratives, guidelines, and event logs. They consolidate this information into coherent descriptions, de-lineate tasks, roles, dependencies, limitations, and are capable of producing preliminary BPMN models or textual process maps. This allows companies to quickly cultivate a collective understanding of the real functioning of the process, encompassing informal practices and undocumented deviations.

2. Process Diagnosis

During this phase, the LLM evaluates the acquired process model and its related data to detect inefficiencies, bottlenecks, deviations, non-conformities, hazards, and resource misalignments. The LLM employs contextual reasoning to elucidate the core causes, identify discrepancies between the intended and actual implementation, and propose avenues for improvement. When logs or measurements are accessible, the model can identify trends that indicate performance deterioration, compliance breaches, or cycles of rework. The result is a systematic diagnostic evaluation that guides the redesign initiative.

3. Process Redesign

This phase emphasizes the collaborative development of enhanced process alternatives (“To-Be” models). LLMs suggest redesigning strategies that meet organizational goals, such as cost reduction, efficiency, quality, or compliance. The model can assess trade-offs, compare various redesign alternatives, and produce BPMN diagrams or textual process representations that integrate automation, simplification, standardization, task elimination, or role reallocation. A HITL methodology ensures that domain specialists enhance, verify, and supplement the options produced to archive feasible and contextually relevant solutions.

4. Process Simulation and Validation

This phase involves evaluating the modified (“To-Be”) process models to ascertain their anticipated performance, practicality, and alignment with corporate objectives prior to deployment. LLMs facilitate the analysis of the ramifications of proposed modifications, identify potential hazards, and assess if the restructured processes adhere to compliance, capacity, and resource limitations. Human specialists evaluate and contrast various scenarios produced by the model, enhancing the redesign until the To-Be version exhibits an acceptable equilibrium of efficiency, reliability, and operational realism. The phase culminates with a validated process design prepared for implementation.

5. Process Execution Support

In this phase, LLMs operate as an intelligent assistant integrated into routine operational processes. The model can offer task advice, interpret policies or Standard Operating Procedures (SOPs) as needed, assist decision-making in extraordinary or unclear circumstances, and recommend corrective measures when discrepancies arise. When combined with BPM or RPA systems, LLM can oversee execution traces, identify anomalies in real time, and suggest alternate paths or human interventions. The purpose of this phase is not to automate execution, but to enhance the workforce and digital

systems with contextual, adaptive, and cognitively sophisticated process knowledge.

6. Continuous Monitoring

In this phase, the LLM methodically collects and evaluates feedback from event logs, performance metrics, user reports, key performance indicators, and compliance assessments. It recognizes indicators of deterioration, such as increased delays, novel bottlenecks, developing dangers, role overload, or departures from the sanctioned To-Be model. By contrasting the actual execution with the planned design, the model can identify the drift of the process, elucidate root causes, and propose incremental modifications or novel improvement opportunities.

2) LLM-CBPI FRAMEWORK REFERENCE ARCHITECTURE

In order to deploy the 6 phases of the LLM-CBPI Framework, we designed a reference architecture that incorporates diverse data sources, process mining functionalities, LLM-driven reasoning, human-centered redesign processes, and ongoing monitoring systems.

This is a modular structure that incorporates process analytics, LLMs, human governance, and workflow execution functionalities to support the 6 phases of the LLM-CBPI Framework. The proposed architecture was designed using a notation inspired by the Unified Modeling Language (UML)

[18] with a set of visual elements that represent a distinct set of functions in the architecture. Below is a description of the visual elements that make up the proposed architecture.

Component. A component is a self-contained functional element that incorporates a series of duties in the LLM-CBPI Framework. Components contain well-defined interfaces and interact with other elements of the architecture via a *dependency relationship* that allows that component to make requests to other elements of the architecture. They are more system capabilities than software implementations. These components may contain subcomponents or modules that, when combined, address a set of related issues [18], [107]. Within the LLM-CBPI Framework reference architecture, we have the following components:

- Human-in-the-Loop Redesign Module
- Improvement Reasoning Engine
- LLM-Based Process Intelligent Layer
- Monitoring & Analytics Layer
- Process Discovery
- Simulation / Validation Module
- Workflow Execution Engine

Packages. A package is a logical grouping of data from the same issue domain. They are not executable entities, but feed these entities with data or information, acting as a facilitating mechanism for communication between the other elements of the architecture [18], [107]. Within the LLM-CBPI Framework reference architecture, we have the following packages:

- AS-IS
- Compliance & Risk Constraints Layer
- Diagnostic Insights

- Input Data Layer
- LLM Use Cases
- Model Layer
- Process Models Repository
- Simulation / Validation Layer
- Structured Data Sources
- TO-BE

Data Source. A data source is a persistent storage or an origin of a data set used by the elements of the architecture. These are structured and semi-structured data sources, external or passive entities that are not executable but provide or retain information used or produced by other elements of the architecture [18], [107]. Within the LLM-CBPI Framework reference architecture, we have the following data sources:

- Event Records
- Vector Database

Pipeline. A pipeline is a sequential data-processing system that converts inputs into outputs via organized steps. They are flow-oriented semantic elements [8], [127]. Within the LLM-CBPI Framework reference architecture, we have the following pipelines:

- ETL/ELT Pipeline
- LLM/RAG Pipeline

External System. An external system is an application existing in the organization beyond the LLM-CBPI Framework reference architecture. These systems interact with elements of the proposed architecture through well-defined interfaces. They are not part of the CBPI itself, but produce or use data relevant to the CBPI process [18], [107]. Within the LLM-CBPI Framework reference architecture, we have the following external system:

- Business Application Layer

Dependency Relationships. For a better understanding of the proposed architecture, only one type of relationship between elements was drawn in the diagram: dependency. This is a unidirectional relationship in which one element (the client) depends on another element (the provider) for its accurate description, specification, or functionality. This relationship is conventionally represented by a dashed arrow that points from the dependent element to the element on which it is based [18], [107]. In the dependency relationships of the proposed architecture, the flow of requests occurs in the direction of the arrow, whereas the flow of information returned from the requests occurs in the opposite direction. The design maintains a loose link between components by using dependency relationships instead of more rigid structural ties, which is very relevant to ensure the scalability, modularity and autonomous development of individual architectural pieces.

Figure 8 depicts this reference architecture.

As we can see in **Figure 8**, this architecture is based on elements widely used and empirically demonstrated in multiple contexts, such as *Human-in-the-Loop*, *LLM/RAG Pipeline*, *ETL/ELT Pipeline*, and *Process Discovery*. It is designed to facilitate all six phases of the CBPI lifecycle by coordinating

a suite of interoperable components that collectively constitute the cognitive and analytical foundation of LLM-assisted continuous improvement. Every subsystem inside the architecture provides distinct functionalities to the CBPI lifecycle, creating a cohesive ecosystem that connects data, models, processes, and human expertise. In the following, we present how each element of the reference architecture contributes to the workflow execution of each phase of the LLM-CBPI Framework.

Phase 1 (Process Understanding)

This phase combines human expertise with the LLMs capabilities to analyze information about organizational processes and transform it into workflows that reflect how processes actually unfold within the organization. The generated representations need to be validated by stakeholders to serve as a diagnostic basis for subsequent business process redesign activities.

This phase begins with the definition of scope, stakeholders, and information sources by human actors. For this, the elements of the *Input Data Layer*, *Structured Data Sources*, and the *Process Models Repository* are used to access organizational records, legacy workflows, governance documents, and operational systems. Here, too, the *Compliance & Risk Constraints Layer* provides information on access, ownership, or regulatory boundaries.

After these activities are completed, the LLM comes into action when the *LLM/RAG Pipeline* is fed with this information, while the *ETL/ELT Pipeline* prepares the structured operational data. In this context, the *Vector Database* stores embeddings that enable semantic retrieval of contextual evidence during later interactions, and the *LLM-Based Process Intelligence Layer* extracts activities, roles, rules, exceptions, and dependencies from these data sources, converting this extracted information into *AS-IS* process descriptions.

In this way, the architecture supports the generation of preliminary process models, such as a workflow description or even a draft of a BPMN workflow, if there is support for integration with a BPMN tool, including task sequences, gateways, handoffs, and control-flow assumptions. These artifacts are stored in the *Process Models Repository*, which performs version control in the refinement cycles.

It is important to note that not all input data mentioned in the phase descriptions will be available within the organization. Even so, LLMs should be able to generate the expected results. However, the accuracy of these output data will be greater if all suggested input data are available, which will also depend on the quality of that data. If the organization does not possess the set of input data suggested in the description of each phase or if the quality of the available data is poor, the process of analyzing and evaluating the results of each phase will require greater effort from stakeholders.

These artifacts are subjected to validation by human actors through interviews, walkthrough sessions, and stakeholder reviews. These tasks are coordinated by the *Human-in-the-*

Loop Redesign Module. At this stage, experts analyze the generated information and may suggest further refinements. If inconsistencies persist, the workflow iterates until an acceptable representation is achieved. Finally, the approved *AS-IS* baseline is formally stored in the repository and released as input to the next phase.

Phase 2 (Process Diagnosis)

This phase combines managerial judgment, analytical capabilities, and LLM-enabled reasoning in order to transform raw execution evidence into structured information. This information is essential to guide future redesign decisions.

In this context, the workflow begins with human actors, who need to define the scope of the diagnosis and develop a set of questions that are within that scope and are properly aligned with business objectives, compliance requirements, and target KPIs. In the architecture, these activities are supported by the *Process Models Repository* and the *Compliance & Risk Constraints Layer*. The *Process Models Repository* provides the query base for *AS-IS* processes, and the *Compliance & Risk Constraints Layer* contains the policies, regulations, standards, and tolerance thresholds that should be considered in this process.

Next, the *ETL/ELT Pipeline* accesses the execution data of the *AS-IS* processes in the *Structured Data Sources* and performs cleansing, standardization, and event-trace structuring. These results are then processed by the *Monitoring & Analytics Layer*, where the previously defined target KPIs are compared with the current values stored in the *Historical Baselines*. The results are stored in *Diagnostic Insights* as empirical evidence.

This information is then interpreted by the *LLM/RAG Pipeline* and the *LLM-Based Process Intelligence Layer* to generate an interpretive diagnosis using process evidence, business rules, constraints, and contextual documentation retrieved from the *Vector Database*. This provides support for identifying root causes, operational risks, and cause-effect relationships that may not be immediately visible through quantitative metrics alone. The outputs of this process include narrative explanations, hypothesis generation, and consolidated diagnostic findings.

Subsequently, this information undergoes a review process by stakeholders, managers, and domain experts. If inconsistencies or areas for improvement are identified, the workflow iterates through further analysis and clarification cycles. Once validated, it may be necessary to prioritize the issues, considering aspects such as impact, feasibility, organizational constraints, and governance thresholds. This is done with the help of the *Improvement Reasoning Engine*, which helps to rank opportunities based on performance gains, risk exposure, and strategic relevance.

In the next step, human actors must produce and approve the diagnostic report that includes confirmed bottlenecks, validated root causes, prioritized issues, and evidence-based conclusions. This artifact is stored for traceability and formally transferred as input to the next phase.

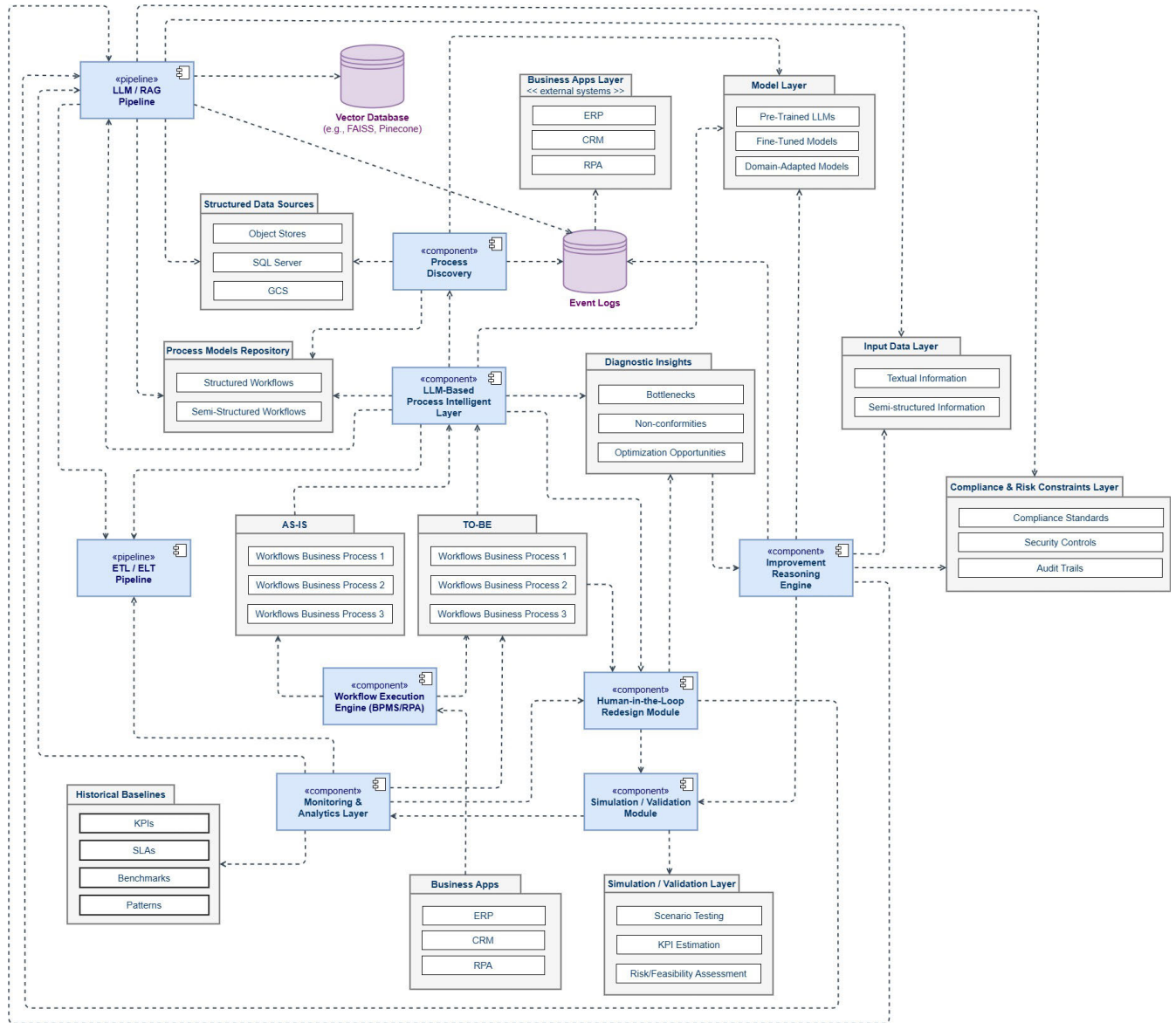


FIGURE 8. LLM-CBPI framework reference architecture.

Phase 3 (Process Redesign)

This phase combines the judgment capabilities of human actors with LLM-enabled generative reasoning. The objective is to create, compare, refine, and formally approve re-design alternatives for later implementation.

Here, the workflow begins by defining the objectives and establishing the constraints that will guide the process re-design strategy. To do this, human actors must perform these activities considering strategic priorities, policies, legal requirements, budget limitations, and technological conditions. These activities are supported by the **Compliance & Risk Constraints Layer**.

After performing these activities, human actors use the **AS-IS** process model stored in the **Process Models Repository** and the prioritized issues identified during diagnosis to define

redesign principles such as automation, control strengthening, cycle-time reduction, workload balancing, etc. [128].

Once these steps are completed, the **LLM/RAG Pipeline** and the **LLM-Based Process Intelligence Layer** support the proposal of process changes and the creation of multiple **TO-BE** alternatives. To achieve this, these components utilize contextual knowledge retrieved from the **Vector Database**, prior process models, business rules, and redesign principles. Expected outputs include recommended modifications such as task elimination, activity consolidation, decision-rule simplification, role reassignment, exception handling improvements, sequencing changes, or automation opportunities. These results are stored in the **TO-BE** repository for version control and traceability.

Another important point is the generation of a summary with the expected changes and trade-offs associated with each alternative, for example, expected efficiency gains, implementation complexity, compliance implications, operational risks, required resources, or organizational disruption, etc. The **Improvement Reasoning Engine** supports these activities.

The next step is the evaluation of these proposed alternatives and the selection of the preferred redesign option. This can be done through workshops and stakeholder feedback, supported by the **Human-in-the-Loop Redesign Module**.

Finally, the selected **TO-BE** process model is submitted for governance approval. Once approved, the redesigned process is formally stored in the repository and released as an authorized input to the next phase. It is important to note that, as stated in **Phase 1** description, these models (AS-IS and TO-BE) can be stored as a workflow description or even a draft of a BPMN workflow if there is support for integration with a BPMN tool.

Phase 4 (Process Validation).

Before the operational deployment of the business process, it is necessary to validate the new workflow model proposed in the previous phase. In this context, this phase aims to verify whether the redesigned process is feasible, compliant, and capable of achieving the desired performance improvements under realistic conditions, thus reducing the risk of implementation and strengthening evidence-based decision-making.

The workflow begins with human actors defining the validation objectives and acceptance criteria for the approved redesign proposal. To do this, they must specify target KPIs, expected service levels, stakeholder expectations, and acceptable risk thresholds [128]. These activities can be carried out with the support of the **Compliance & Risk Constraints Layer**, in order to ensure the inclusion of regulatory obligations, governance criteria, and tolerance limits in the validation process.

After these activities are carried out, there must be a preparation stage for the validation data and simulation scenarios. This is done using historical records, prior execution evidence, workload assumptions, and alternative operating conditions obtained from the **Historical Baselines**, **Structured Data Sources**, and **Process Models Repository**. At this stage, the **ETL/ELT Pipeline** supports data cleaning and scenario preparation.

Therefore, the approved **TO-BE** model is then processed by the **Simulation / Validation Module**. The objective is to simulate the redesigned workflow under different conditions, such as different demand profiles, resource constraints, or policy assumptions. The expected outputs of this stage include projected cycle times, throughput levels, queue formation, resource utilization, Service Level Agreement (SLA) compliance, and expected cost implications. Subsequently, these outputs are assessed by the **Monitoring & Analytics Layer**, which compares the simulated results with acceptance criteria and historical benchmarks.

At the same time, the **LLM/RAG Pipeline** and the **LLM-Based Process Intelligence Layer** support a higher-level interpretation of outcomes validation. Using simulation evidence, organizational rules, and contextual documentation retrieved from the **Vector Database**. The goal is to identify potential implementation risks, hidden bottlenecks, policy

conflicts, unrealistic assumptions, or operational dependencies that may not be evident through numerical metrics alone. At this stage, it is also possible to recommend refinements such as control adjustments, resource reallocations, sequencing changes, or exception-handling improvements.

All these findings must be submitted to review by human actors for verification, by managers and domain specialists, whether acceptance criteria have been met and whether identified issues can be resolved through incremental refinements. If necessary, it is possible to return for adjustment through iterative redesign cycles until the validation requirements and objectives are met and formally proven. This approved artifact is stored for traceability and is released as authorized input to the next phase.

Phase 5 (Process Execution Support) The objective of this phase is to apply the approved re-design business process to the execution of the organization's daily operations. To do this, it is necessary to develop execution guidance and obtain real-time evidence, in order to prepare the basis for continuous improvement of business processes and for the next monitoring phase.

This phase begins with the development of the deployment plan by human actors. For this, they used the validated **TO-BE** process approved in **Phase 4**. In this context, stakeholders must define rollout priorities, resource allocation, transition schedules, and organizational readiness actions. These activities are supported by the **Compliance & Risk Constraints Layer**. This layer ensures that deployment decisions are aligned with governance requirements, segregation-of-duty rules, security controls, and operational risk thresholds.

After this step, human actors must configure (integration parameters, user permissions, interfaces, system dependencies, etc.) the execution environment of the proposed workflow through the **Workflow Execution Engine (BPMS/RPA)** and connected enterprise systems represented in the **Business Apps Layer**, such as Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), and automation platforms. The goal is to make the proposed workflow operate under realistic conditions that reflect the daily operations of the organization.

Once the execution environment is configured, the re-designed process is deployed as an executable workflow instance. At this stage, the **Workflow Execution Engine** is responsible for orchestrating tasks, decision logic, handoffs, escalations, and automated activities. All of this is done according to the approved **TO-BE** model stored in the **Process Models Repository**.

In this context, the **LLM/RAG Pipeline** and the **LLM-Based Process Intelligence Layer** provide contextual execution support to operational users. For this, policies, SOPs,

business rules, role definitions, and organizational knowledge retrieved from the **Vector Database** are used. These procedures reduce ambiguities and improve consistency in workflow execution.

Here, we also have the role of the **Monitoring & Analytics Layer** through the continuous collection of evidence related to the execution of the business process, such as executed tasks, decision logs, timestamps, resource utilization records, event streams and system audit trails. This involves creating an empirical basis for future evaluations of performance evaluation and process learning.

In this step, it is also important to collect structured feed-back from stakeholders, such as execution difficulties, unexpected constraints, usability concerns, or suggestions regarding AI-generated guidance. **Human-in-the-Loop Redesign Module** can support these activities in order to complement the system-generated evidence captured by the **Monitoring & Analytics Layer**.

Finally, the consolidated data from this phase serves as input for the next phase, allowing longitudinal performance tracking and future optimization cycles.

Phase 6 (Continuous Monitoring)

This is the phase that completes a cycle of the CBPI closed-loop. In this phase, the objectives are to measure sustained performance, verify compliance, detect emerging risks or process drift, and trigger new improvement cycles when-ever operational conditions or strategic objectives change.

The workflow begins with the computation of the performance metrics (predefined KPIs, SLA thresholds, throughput indicators, etc.) by the **Monitoring & Analytics Layer**. These values provide an up-to-date view of the workflow's operational performance over time.

At the same time, the **Compliance & Risk Constraints Layer** monitors regulatory and constraint indicators such as business rules, standards, policies, and governance requirements. The expected values are compared with the values obtained from the current execution of the workflow. The goal is to identify deviations, control violations, segregation-of-duty conflicts, or growing risk exposure. These two layers ensure efficiency and conformity.

This information is compared with historical references stored in the **Historical Baselines**. In this way, it is possible to detect anomalies, gradual performance deterioration, demand shifts, resource imbalances, or behavioral drift in the deployed process.

Next, the **LLM/RAG Pipeline** and the **LLM-Based Process Intelligence Layer** interpret the monitoring signals at a higher level. To do this, these components use contextual process knowledge from the **Vector Database**, previous diagnostic findings, and organizational objectives. In this way, it is possible to translate the raw indicators into understandable managerial insights. Outputs from this stage include narrative explanations of anomalies, likely causes of decreased performance, identification of recurring bottlenecks, and contextualized interpretations of compliance alerts. Potential recommendations are also generated, such as process

recalibration, workload redistribution, control reinforcement, parameter adjustment, etc.

Finally, human actors must analyze the monitoring in-sights. This can be done through governance forums, dash-boards, reports, or data visualization, provided there is support for integration with Business Intelligence and data visualization tools. Based on these analyzes, new improvement cycles are initiated, and the monitoring baselines are updated.

3) DESIGN RATIONALE

The LLM-CBPI reference architecture is designed to address three sometimes competing demands in modern process improvement efforts: analytical rigor, explainable intelligence, and organizational governance. This subsection elucidates the reasoning behind the principal architectural selections and delineates how each selection aligns with the CBPI objectives.

A key design choice is the clear distinction between deterministic analytics and generative AI reasoning. Quantitative duties, including KPI computation, conformance verification, anomaly detection, and simulation, are assigned to the **Monitoring & Analytics Layer**, while interpretative and explanatory tasks are managed by the **LLM-Based Process Intelligent Layer**.

This distinction prevents LLMs from being erroneously regarded as numerical or statistical engines and mitigates hazards related to fabricated measurements or unverifiable outcomes. The design maintains methodological rigor by anchoring all reasoning in verifiable analytical outputs while also using the capacity of LLMs to synthesize, contextualize, and elucidate intricate discoveries. This design decision directly facilitates **Phases 2, 4, and 6** of the LLM-CBPI Framework, where analytical evidence must precede interpretation and decision-making.

Another essential justification is the establishment of human oversight within the framework. The architecture clearly incorporates governance checkpoints through the **Human-in-the-Loop Redesign Module** and decision gateways at each level of the CBPI, rather than considering human involvement as an exception. Humans maintain dominion over:

- Determination of scope and diagnostic framework (Phases 1–2);
- Selection and approved redesign options (Phase 3);
- Decisions about acceptance and deployment (Phases 4–5); and
- Initiation of fresh improvement cycles (Phase 6).

This design ensures accountability, regulatory adherence, and institutional confidence, particularly in scenarios where automated process modifications may entail legal, ethical, or operational consequences.

The architecture is deliberately constructed as a closed-loop system, rather than a linear BPM pipeline. The results of execution and monitoring (Phases 5–6) are directly sent

back to the preceding phases, especially **Phase 1 - Process Understanding** and **Phase 2 - Process Diagnosis**.

This design, informed by feedback, facilitates the ongoing acquisition of knowledge from operational data, adaptive

baseline modifications, and premature identification of deviations and nascent threats. The *Monitoring & Analytics Layer* is important in operationalizing this loop, converting runtime evidence into organized improvement signals instead of static reports.

Another striking feature of the proposed architecture is modularization for scalability and adaptation. Every architectural capability—data input, reasoning, validation, execution, and monitoring—is executed as a loosely connected module. This modularity enables the architecture to progress over time without necessitating a complete system redesign. For instance: LLM models may be substituted or refined autonomously within the Model Layer. New analytical engines may be integrated into the *Monitoring & Analytics Layer*. Furthermore, additional governance constraints can be integrated into the *Compliance & Risk Constraints Layer*. This architecture facilitates long-term maintainability and scalability within actual organizational contexts.

In the same way, the suggested solution, unlike conventional BPM designs that emphasize either discovery or execution, retains explicit representations of both As-Is and To-Be process models in the *Process Models Repository*. This dual representation facilitates traceability of enhancement decisions, contrast between the present and desired situations, simulation, and validation preceding deployment. Additionally, it facilitates *LLM-Based Process Intelligent Layer* and *Improvement Reasoning Engine* by providing structured artifacts that ground semantic explanations in formal process models.

Furthermore, explainability is regarded as a fundamental architectural consideration rather than a secondary matter. The architecture ensures that each AI-assisted out-put—diagnostic insight, redesign recommendation, or monitoring alert—can be traced to fundamental data sources, analytical findings, and defined restrictions or objectives. By designating LLMs as elucidators and synthesizers instead of independent decision-makers, the framework improves transparency and fosters appropriate AI utilization in BPM scenarios.

Finally, the architecture is designed to seamlessly interact with existing business applications, including ERP, CRM, and RPA platforms. This prevents the formation of independent AI systems and ensures that CBPI functions within established organizational frameworks. The *Compliance & Risk Constraints Layer* ensures that all process modifications adhere to organizational policies, regulatory mandates, and auditability standards, thus improving the appropriateness of the architecture for practical implementation.

In summary, proposed LLM-CBPI Framework reference architecture is deliberately structured as a modular, regulated, and elucidative socio-technical system that harmonizes analytical precision, AI-enhanced reasoning, and human oversight. Every architectural decision immediately facilitates the

ongoing, evidence-driven, and accountable enhancement of business processes.

4) FRAMEWORK AND ARCHITECTURAL RESPONSE TO IDENTIFIED GAPS

In this section, we summarize the correlation between the gaps identified in the SLR and presented in **Section VIII** and the elements of the LLM-CBPI Framework that address these gaps. This approach underscores the function of LLMs as facilitators of CBPI, focusing on automation, optimization, and decision support. By applying these elements, organizations can optimize the advantages of LLM-driven continuous process improvement.

Gap-01: End-to-end CBPI Lifecycle Coverage Through its closed-loop approach, the LLM-CBPI framework provides clear assistance for the entire CBPI lifecycle, encompassing its six phases: process comprehension, diagnosis, redesign, simulation/validation, execution support, and ongoing monitoring. In the same way, its architecture incorporates elements for analysis, reasoning, validation, enactment, and monitoring, while delineating the data and information flows between them. Moreover, each phase is characterized by specific inputs and outputs, at least one architectural component operationally supports each phase, and feedback from monitoring can initiate a renewed diagnosis and redesign.

Gap-02: Heterogeneous Data Ingestion and Access The proposed framework assimilates and offers regulated access to diverse data sources, encompassing structured (e.g., relational records, KPIs, event logs), semi-structured (e.g., JSON/XML), and unstructured data (e.g., policies, manuals, tickets, emails). The *Input Data Layer* provides standardized inputs for structured and semi-structured data. *Structured Data Sources* and the *ETL/ELT Pipeline* perform the extraction, cleansing, transformation, and integration of structured operational data. Simultaneously, the *LLM/RAG Pipeline* and *Vector Database* provide semantic ingestion, indexing, and retrieval of unstructured knowledge sources. Furthermore, *Event Logs* offer dedicated support for event-driven evidence. In this way, this strategy provides a suitable approach to separating responsibilities between ingestion, storage, retrieval, higher-level reasoning, and analytical functions.

Gap-03: Separation of Concerns and Modular Decomposition As evidenced in the previous paragraph, the LLM-CBPI reference architecture establishes a modular division of duties, segregating data access, transformation, modeling, retrieval, reasoning, validation, execution, and monitoring. It also features a layered structure with explicit interfaces and defined responsibilities for each layer. Each element serves a singular core function, cross-cutting issues (such as security and governance) are uniformly enforced, and components can be substituted without compromising the integrity of the entire architecture.

Gap-04: Event-log Readiness and Empirical Process Evidence In order to support event-log readiness and ensure an empirical basis for process discovery, conformance ver-

ification, and performance evaluation, *ETL/ELT Pipeline* extracts, cleanses, transforms, and enriches execution data from enterprise systems and operational repositories. *Event Logs* was designed as an architectural element dedicated to storing and maintaining the resulting logs that serve as input to the *Process Discovery* and *Monitoring & Analytics Layer* for downstream analysis. This approach provides quality checks, completeness and consistency assessments, standardized trace generation, and continuous availability of empirical evidence for diagnosis and monitoring activities.

Gap-05: Process Model Management and Traceability The LLM-CBPI framework reference architecture provides support for centralized management, versioning, comparison, and controlled evolution across successive redesign iterations through the *Process Models Repository*, *AS-IS*, and *TO-BE* storage repositories. In addition, these architectural elements also provide traceability links between process models and supporting evidence, including event logs, KPIs, compliance constraints, simulation results, and redesign decisions. Thus, it is possible to track changes, compare alternative versions, diagnoses, findings, and governance and compliance requirements. These elements also offer support for integration with redesign, validation, and execution components, ensuring continuity between modeling activities and downstream operational stages.

Gap-06: Diagnostic Insight Generation and Interpretability The proposed solution supports the generation of diagnostic insights and interpretability. To this end, the reference architecture combines the use of the *Monitoring & Analytics Layer*, the *Model Layer*, and the *LLM-Based Process Intelligence Layer*. These architectural elements are responsible for transforming the execution evidence into structured diagnostic outputs. These outputs are subsequently consumed by downstream decision-support components, including the *Improvement Reasoning Engine* and the *Human-in-the-Loop Redesign Module*. In this way, they provide redesign prioritization and stakeholder evaluation. In this sense, diagnostic findings have an adequate level of explainability about what happened, where issues occur, and why they are likely emerging, while preserving traceability to underlying evidence sources (event logs, KPIs, compliance constraints, etc.).

Gap-07: Improvement Reasoning and Alternative Generation The framework produces improvement hypotheses and re-design alternatives, encompassing logic, assumptions, anticipated impacts, and clear trade-offs (cost–time–quality–risk). An *Improvement Reasoning Engine* analyzes diagnostics, restrictions, and contextual knowledge to generate structured redesign alternatives. In this way, alternatives are generated as organized change sets, each alternative encompasses justification, and trade-offs and limits are clearly delineated. This architectural element works in conjunction with the *LLM-based Process Intelligence Layer* and the *Compliance and Risk Constraints Layer*. In this way, it is possible to ensure that the proposed alternatives remain contextually relevant and aligned with governance. This

approach ensures that the redesign options are always accompanied by a clear justification, expected benefits, implementation limitations, and comparable compensation profiles for subsequent human evaluation.

Gap-08: Human-in-the-Loop Accountability and Approval Gates

The LLM-CBPI framework incorporates human-in-the-loop methods for the review, refinement, and approval of diagnostic interpretations and redesign ideas, to ensure accountability, contextual judgment, and integration of tacit knowledge. *Human-in-the-Loop Redesign Module* offers interaction points, explanatory views, and decision documentation. As a result, deployment necessitates prior approval, decisions are documented with rationale, and expert comment informs revisions of design artifacts. Furthermore, in cooperation with the *Improvement Reasoning Engine* and the *Process Models Repository*, these three elements of the proposed architecture allow stakeholders to evaluate proposed changes in a structured way, request adjustments, compare alternatives, and formally approve the selected design artifacts. In this context, critical redesign decisions are subject to explicit approval gates, including managerial rationale and expert feedback that can iteratively refine the results before validation or downstream deployment.

Gap-09: Ex-ante Simulation and Validation of Re-designs

The proposed framework provides ex-ante simulation and validation of redesigned process alternatives before operational deployment. The objective is to mitigate the implementation risk and improve evidence-based decision-making. In this context, the *Simulation / Validation Layer* evaluates the *TO-BE* models under alternative demand conditions, resource allocations, policy constraints, and operational assumptions. The *Simulation / Validation Layer* works together with the *Process Models Repository*, *Historical Baselines*, and the *Monitoring & Analytics Layer*, allowing redesigned processes to be compared against current performance levels, target KPIs, and prior execution patterns. As a result, it is possible to check projected cycle times, throughput, queue levels, resource utilization, SLA achievement, and expected cost implications. In parallel, the *LLM-Based Process Intelligence Layer* interprets the results of the validation process by identifying hidden bottlenecks, unrealistic assumptions, policy conflicts, or additional improvement opportunities. Thus, redesign alternatives can be screened

before implementation, high-risk proposals can be rejected or refined, and only validated process models are promoted to the downstream execution stages.

Gap-10: Operational Implementation and Integration with Enterprise Systems

In the LLM-CBPI Framework reference architecture, the *Workflow Execution Engine* orchestrates task routing, decision logic, transfers, escalations, and automated activities according to the redesigned process approved in the *TO-BE* models. This work is done in conjunction with the *Business Applications Layer*, as this layer connects the elements of

the designed architecture with business systems such as ERP, CRM, databases, legacy platforms, and RPA. In this way, it is possible to deploy workflows, exchange data, process transactions, and promote synchronization with existing operational infrastructures. Simultaneously, the evidence generated during execution is captured and forwarded to the **Monitoring & Analytics Layer**. This ensures real-time traceability of operations and continuous visibility of process performance.

Gap-11: Continuous Monitoring and Closed-loop Feedback

The LLM-CBPI Framework provides a phase of continuous monitoring of implemented processes. The objective is to obtain feedback that allows for continuous performance evaluation as well as to detect deviations and promote iterative improvement. In this context, the architectural element **Continuous Monitoring & Feedback Layer** is responsible for obtaining real-time evidence from operational environments and continuously evaluating process behavior in relation to predefined KPIs, SLA targets, and governance boundaries. The **Monitoring & Analysis Layer** and **Historical Baselines** allow current performance to be compared with previous execution standards, expected service levels, and historical trends. Furthermore, relevant signals can be forwarded to the elements of the architecture **Diagnostic Insights** and **Improvement Reasoning Engine** for reanalysis and generation of corrective recommendations.

Gap-12: Historical Baselines and Comparative Evaluation

The proposed reference architecture maintains **Historical Baselines** that contain previous operational behavior, process performance records, and previously observed KPI levels. This approach allows for before-and-after comparisons, boundary setting, and evidence-based evaluation of improvement initiatives. These baselines provide a reference point for determining whether observed changes represent genuine improvement or normal operational variation. Also noteworthy here are the architectural elements: the **Simulation/Validation Module**, the **Simulation/Validation Layer**, the **Monitoring & Analytics Layer**, and the **Continuous Monitoring Layer**. These elements allow redesigned processes and real-time executions to be evaluated against historical performance standards. Thus, the proposed solution allows for the calibration of simulation parameters, supports post-implementation performance evaluation, enables anomaly detection, and strengthens the longitudinal monitoring of continuous improvement results.

Gap-13: Explainability, Traceability, and Auditable Reasoning (LLM-grounded)

To support explainability, traceability, and auditable reasoning, the proposed solution links diagnostic and prescriptive outputs to the underlying evidence used during the analysis and generation of recommendations. In the reference architecture, the **LLM/RAG Pipeline** works with organizational context information and evidence to support the answers, pre-serving references related to supporting artifacts. To this end, the **LLM/RAG Pipeline** works in

conjunction with the **Vector Database**, **Process Model Repository**, **Diagnostic Insights**, and the **Compliance & Risk Constraints Layer** in order to provide recommendations linked to process models, KPI evidence, business rules, historical records, and governance constraints. This approach provides greater transparency to stakeholders as it allows the **Improvement Reasoning Engine** to present the assumptions, trade-offs, and decision criteria used to propose process redesign alternatives.

Gap-14: Knowledge Freshness and Adaptability Without Retraining

In order to support knowledge freshness and adaptability without retraining, the LLM-CBPI Framework reference architecture provides continuous updates of organizational knowledge sources such as policies, manuals, constraints, and process documentation. To do that, the **Input Data Layer**, in conjunction with the **Structured Data Sources**, the **LLM/RAG Pipeline**, and the **Vector Database**, allow newly available content to be ingested, indexed, and retrieved incrementally at the time of inference. In this way, updated artifacts are immediately available for subsequent reasoning tasks, while obsolete or superseded documents can be deactivated through repository maintenance procedures. Furthermore, the **Compliance & Risk Restrictions Layer** offers metadata filters, document provenance attributes, and access control policies to provide greater control over retrieval operations.

Gap-15: Compliance and Risk Constraints Enforcement

The proposed Framework incorporates and applies compliance requirements, control obligations, and risk limits throughout the entire improvement lifecycle, including diagnosis, redesign, validation, execution, and continuous monitoring. This capability is primarily implemented through the dedicated **Compliance & Risk Constraints Layer**, which centralizes business rules, regulatory requirements, security controls, approval policies, and organizational risk tolerances. This element of the LLM-CBPI Framework reference architecture works in a coordinated manner with the

Improvement Reasoning Engine, the **Simulation/Validation Layer**, the **Workflow Execution Engine**, and the **Monitoring & Analytics Layer**. The goal is to ensure that analytical output, redesign alternatives, executable workflows, and runtime operations remain aligned with governance.

Gap-16: Security, Privacy, and Access Control

The proposed LLM-CBPI framework incorporates security, privacy, and access control mechanisms across all layers of its reference architecture. The objective is to ensure that sensitive process information and organizational data are handled in accordance with internal policies and regulatory requirements. In this context, the **Input Data Layer**, **Structured Data Sources**, the **LLM/RAG Pipeline**, and the **Vector Database** apply authentication, authorization, metadata-based filtering, and least-privilege retrieval policies that restrict the exposure of protected content during the ingestion, indexing, retrieval, and generation stages. Data minimization

principles can also be applied so that only relevant information is processed for each task. In parallel, the **Compliance & Risk Constraints Layer** supports the application of privacy and security rules, while the logging mechanisms and **Event Logs** provide auditable records of data access, model interactions, and operational actions.

Gap-17: Reproducibility and Operational Reliability

In the LLM-CBPI Framework reference architecture, reproducibility is primarily enabled by the **Model Layer**, the **ETL/ELT Pipeline**, the **LLM/RAG Pipeline**, and the **Process Model Repository**, where versions of models, data transformations, recovery configurations, prompts, and generated artifacts can be tracked and historically recorded. This allows previous runs to be executed again with equivalent inputs and configurations whenever auditing, validation, or troubleshooting activities are required. At the same time, the **Monitoring & Analytics Layer** supports operational reliability by detecting model deviations, data deviations, output degradation, abnormal behavior, and service instability. This approach promotes the traceability, repeatability, and sustained performance of AI-enabled continuous improvement operations.

X. CONCLUSION

The results of this study demonstrate how the incorporation of LLMs has revolutionized BPI and BPM. In order to increase productivity, save costs, and ensure compliance, traditional BPM approaches use structured rule-based process optimization techniques such as Six Sigma, Lean, and Process Mining. The use of LLMs, however, brings with it fresh perspectives that contradict and enhance these conventional BPM models. In the following, we highlight key points that show how LLMs challenge traditional BPM models and some applications where the use of traditional BPM models is still relevant.

Process Flexibility and Adaptability: conventional BPM models are less flexible in dynamic business environments

because they place an emphasis on strict procedures and pre-determined decision trees. The processing powers of LLMs facilitate real-time adaptability and decision making, giving organizations greater flexibility in responding to changing circumstances.

Handling Unstructured Data: emails, customer reviews, and legal documents are examples of unstructured data sources that are difficult for traditional BPM systems to integrate. LLMs can efficiently process and analyze such data, offering useful insights and automation possibilities that are not possible with conventional BPM tools.

Enhancing Human-Centric Workflows: BPM frameworks typically call for human participation in data processing and decision making. By automating intricate processes such as document classification, sentiment analysis, and compliance checks, LLMs reduce manual workload and change the human role in BPM from execution to strategic decision-making and monitoring.

Predictive and Prescriptive Capabilities: traditional BPM models lack strong predictive analytics, even though

they optimize processes using past data. When combined with AI-driven techniques like RLHF, LLMs provide predictive insights that assist organizations in foreseeing bottlenecks and recommending changes before inefficiencies occur.

Confirming Traditional BPM Models: the research validates that standardized BPM frameworks are very important for maintaining process integrity and compliance, with LLMs augmenting rather than supplanting governance structures. LLMs improve BPM automation methodologies such as RPA and Process Mining, increasing precision, decreasing variability, and facilitating automated decision-making.

In AI-driven BPM, Lean and Six Sigma continue to be relevant, and LLMs facilitate continuous improvement through automated data analysis and real-time insights. Although LLMs improve workflows, human supervision is essential for fine-tuning, mitigating bias, and ensuring accountability.

Ultimately, LLMs enhance rather than replace traditional BPM by automating processes, improving decision making, and managing unstructured data, while still adhering to essential principles of BPM such as compliance, standardization, and continuous improvement. To optimize advantages, organizations must integrate LLMs with current BPM systems, harmonizing AI-driven automation with human oversight.

In this context, this study enhances the field of AI-driven process improvement by introducing the LLM-CBPI Framework and its reference architecture as a systematic solution to ongoing structural deficiencies identified in current methodologies. Although previous studies have shown the capability of AI to improve business processes, the majority of current solutions are disjointed, focusing on singular analytical or automation functions rather than offering a cohesive, lifecycle-oriented and governable framework. The findings of this study suggest that to overcome these restrictions one needs a systematic architectural approach based on clearly specified requirements that together tackle reasoning, empirical evidence, governance, and operational implementation, rather than mere incremental methodological improvements.

The proposed framework addresses these needs by delineating seventeen essential gaps that must be addressed to effectively operationalize key dimensions for reliable AI-enabled CBPI, encompassing explainability, traceability, compliance alignment, human accountability, lifecycle coverage, and validation mechanisms. By addressing these gaps as essential architectural design requirements instead of optional system features, the framework creates a stringent basis for the development of AI-driven process intelligence solutions that are reliable, auditable, and applicable in actual organizational contexts. The provided reference architecture enhances this contribution by converting conceptual needs into modular components and interaction frameworks, thus closing the enduring gap between theoretical models and implementable systems.

This research establishes a cohesive architectural and requirement-driven foundation that converts AI-enabled process improvement from a disparate array of promising techniques into a unified, governable, and operationally feasible

paradigm, thereby enhancing both the theoretical and practical dimensions of the field.

XI. LIMITATIONS OF THIS STUDY

Regarding the limitations of this study, we can state that because LLMs are evolving so rapidly, some of the findings can become out of date very quickly, requiring ongoing updates and further research to take into account new developments in technology. Another drawback is that, even if the research finds patterns in several business sectors, the precise influence of LLMs may differ depending on the sector, organization size, and state of the digital infrastructure. Moreover, a lot of research focused on case studies and theoretical advantages rather than thorough empirical evaluation of LLM-driven organization improvements.

XII. RECOMMENDATIONS FOR FUTURE RESEARCH

The findings indicate that the future development of AI-driven process enhancement will rely more on the availability of integrative architectures that can coordinate diverse data sources, reasoning modules, validation processes, and monitoring systems within defined governance and risk parameters, rather than solely on improvements in model performance. The LLM-CBPI Framework not only enhances current methodologies, but also creates a design-focused paradigm that redefines the systematic integration of AI inside process improvement lifecycles.

In this context, although the proposed framework and architecture are conceptually grounded in the literature and their elements have already been tested and validated in previous research, the current arrangement proposed in the LLM-CBPI Framework and its reference architecture represents an innovation that needs empirical validation in real organizational environments. In this sense, future research should focus on confirming its practical effectiveness through case studies, pilot implementations, action research, and longitudinal evaluations in different sectors and process contexts.

Another possibility for future research is the refinement and operational validation of the workflows for each of the phases of the LLM-CBPI Framework described in **Section IX-B-2**. The proposed workflows for Phases 1 to 6 should be tested with professionals in the field to assess usability, completeness, sequencing logic, decision points, and feasibility of integration with corporate systems. An interesting approach is to subject each workflow of each phase of the LLM-CBPI Framework to the same continuous improvement logic advocated in the framework itself (**Figure 7**). This involves iteratively evolving the LLM-CBPI Framework over time, transforming it into a self-improvement reference model, based on both theoretical advances and accumulated implementation evidence.

Finally, controlled comparisons with traditional BPM approaches and long-term studies are also needed to evaluate the dynamics of closed-loop learning and the lasting effects of LLMs-enabled improvements on organizational performance, as well as interdisciplinary collaboration that

incorporates insights from information systems, software design, governance, and ethics in AI, in order to enhance the theoretical foundations and ensure responsible implementation.

REFERENCES

- [1] H. Kashfi and F. S. Aliee, "Business process improvement challenges: A systematic literature review," in *Proc. 11th Int. Conf. Inf. Knowl. Technol. (IKT)*, Dec. 2020, pp. 122–126.
- [2] M. Dumas, M. La Rosa, J. Mendling, and H. A. Reijers, *Fundamentals of Business Process Management*, 2nd ed., Cham, Switzerland: Springer, 2018.
- [3] M. Abbasi, R. I. Nishat, C. Bond, J. B. Graham-Knight, P. Lasserre, Y. Lucet, and H. Najjaran, "A review of AI and machine learning contribution in predictive business process management (process enhancement and process improvement approaches)," 2024, *arXiv:2407.11043*.
- [4] J. vom Brocke, T. Schmiedel, J. Recker, P. Trkman, W. Mertens, and S. Viaene, "Ten principles of good business process management," *Bus. Process Manage. J.*, vol. 20, no. 4, pp. 530–548, Jul. 2014.
- [5] W. van der Aalst, *Process Mining: Data Science in Action*. Cham, Switzerland: Springer, 2016.
- [6] T. Davenport and J. Harris, *Competing on Analytics: Updated, With a New Introduction: The New Science of Winning*. Harvard Bus. Press, 2017.
- [7] T. Brown et al., "Language models are few-shot learners," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 1877–1901.
- [8] P. Lewis et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 9459–9474.
- [9] J. Mendling, G. Decker, R. Hull, H. A. Reijers, and I. Weber, "How do machine learning, robotic process automation, and blockchains affect the human factor in business process management?" *Commun. Assoc. Inf. Syst.*, vol. 43, no. 1, p. 19, 2018.
- [10] T. H. Davenport, "From analytics to artificial intelligence," *J. Bus. Analytics*, vol. 1, no. 2, pp. 73–80, Jul. 2018.
- [11] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," 2017, *arXiv:1702.08608*.
- [12] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artif. Intell.*, vol. 267, pp. 1–38, Feb. 2019.
- [13] F. Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press, 2015.
- [14] W. E. Deming, *Out of the Crisis*. Cambridge, MA, USA: MIT Press, 1986.
- [15] J. Mendling, B. T. Pentland, and J. Recker, "Building a complementary agenda for business process management and digital innovation," *Eur. J. Inf. Syst.*, vol. 29, no. 3, pp. 208–219, May 2020.
- [16] A. Beheshti et al., "ProcessGPT: Transforming business process management with generative AI," 2023, *arXiv:2305.01837*.
- [17] M. Grohs, L. Abb, N. Elsayed, and J.-R. Rehse, "Large language models can accomplish business process management tasks," in *Proc. Int. Conf. Bus. Process Manag.* Cham, Switzerland: Springer, 2023, pp. 453–465.
- [18] *OMG Unified Modeling Language (UML) Specification, OMG Std. Version 2.5.1*, Object Management Group, 2017.
- [19] B. Baker, "Lean six sigma: Combining six sigma quality with lean speed," *Quality Progress*, vol. 36, no. 10, p. 96, 2003.
- [20] M. Hammer and J. Champy, "Reengineering the corporation: Manifesto for business revolution," Zondervan, Tech. Rep., 2009.
- [21] P. Bornet, I. Barkin, and J. Wirtz, *Intelligent Automation: Welcome to the World of Hyperautomation: Learn How to Harness Artificial Intelligence to Boost Business & Make Our World More Human*. Singapore: World Scientific, 2021.
- [22] A. Haleem, M. Javaid, R. P. Singh, S. Rab, and R. Suman, "Hyperautomation for the enhancement of automation in industries," *Sensors Int.*, vol. 2, Apr. 2021, Art. no. 100124.
- [23] M. Chen et al., "Evaluating large language models trained on code," 2021, *arXiv:2107.03374*.
- [24] Y. Chang et al., "A survey on evaluation of large language models," 2023, *arXiv:2307.03109*.
- [25] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, E. H. Chi, T. Hashimoto, O. Vinyals, P. Liang, J. Dean, and W. Fedus, "Emergent abilities of large language models," 2022, *arXiv:2206.07682*.

- [26] R. Bommasani et al., "On the opportunities and risks of foundation models," 2021, *arXiv:2108.07258*.
- [27] Y. Wu, F. Jia, S. Zhang, H. Li, E. Zhu, Y. Wang, Y. T. Lee, R. Peng, Q. Wu, and C. Wang, "An empirical study on challenging math problem solving with GPT-4," 2023, *arXiv:2306.01337*.
- [28] C. van Dun, L. Moder, W. Kratsch, and M. Röglinger, "ProcessGAN: Supporting the creation of business process improvement ideas through generative machine learning," *Decis. Support Syst.*, vol. 165, Feb. 2023, Art. no. 113880.
- [29] T. H. Chen, "An artificial intelligence based approach to automate document processing in business area," Ph.D. dissertation, Massachusetts Institute of Technology, 2021.
- [30] J. Ribeiro, R. Lima, T. Eckhardt, and S. Paiva, "Robotic process automation and artificial intelligence in Industry 4.0—A literature review," *Proc. Comput. Sci.*, vol. 181, pp. 51–58, May 2021.
- [31] N. Jha, D. Prashar, and A. Nagpal, "Combining artificial intelligence with robotic process automation—An intelligent automation approach," in *Deep Learning and Big Data for Intelligent Transportation: Enabling Technologies and Future Trends*. Cham, Switzerland: Springer, 2021, pp. 245–264.
- [32] G. Gopal and E. Pilkuskaite, "Implementing process innovation by integrating continuous improvement and business process re-engineering," in *Innovation Management*. Ann Arbor, MI, USA: Edwards, 2020, pp. 93–113.
- [33] A. Booth, M. M.-S. James, M. Clowes, and A. Sutton, *Systematic Approaches to a Successful Literature Review*, 2021.
- [34] M. Vidgof, S. Bachhofner, and J. Mendling, "Large language models for business process management: Opportunities and challenges," in *Business Process Management Forum: BPM 2023 Forum*. Cham, Switzerland: Springer, 2023, pp. 107–123.
- [35] L. Wang, C. Ma, X. Feng, Z. Zhang, H. Yang, J. Zhang, Z. Chen, J. Tang, X. Chen, Y. Lin, W. X. Zhao, Z. Wei, and J. Wen, "A survey on large language model based autonomous agents," *Frontiers Comput. Sci.*, vol. 18, no. 6, Dec. 2024, Art. no. 186345.
- [36] H. Kourani, A. Berti, D. Schuster, and W. M. van der Aalst, "Process modeling with large language models," in *Proc. Int. Conf. Bus. Process Model., Develop. Support*. Cham, Switzerland: Springer, 2024, pp. 229–244.
- [37] H. Zhao, L. Wu, Y. Shan, Z. Jin, Y. Sui, Z. Liu, N. Feng, M. Li, and W. Zhang, "A comprehensive survey of large language models in management: Applications, challenges, and opportunities," *J. Latex Class Files*, vol. 14, no. 8, 2015.
- [38] X. Huang, G. Mei, and J. Zhang, "Cross-source point cloud registration: Challenges, progress and prospects," 2023, *arXiv:2305.13570*.
- [39] C. Ziche and G. Apruzzese, "LLM4PM: A case study on using large language models for process modeling in enterprise organizations," in *Proc. Int. Conf. Bus. Process Management*. Cham, Switzerland: Springer, 2024, pp. 472–483.
- [40] B. Kitchenham, "Procedures for performing systematic reviews," *Keele Univ., Tech. Rep.*, 2004, vol. 33, no. 2004, pp. 1–26.
- [41] A. Nightingale, "A guide to systematic literature reviews," *Surg., Oxford*, vol. 27, no. 9, pp. 381–384, Sep. 2009.
- [42] S. J. Bensman, "Anne-wil harzing: The publish or perish book: Your guide to effective and responsible citation analysis: Tarma software research pty ltd, melbourne, australia, c2010, ISBN 978-0-9808485-1-9, U.S.\$49.95 (paperback, black and white), U.S.\$74.95 (paperback, color)," *Scientometrics*, vol. 88, no. 1, pp. 339–342, Jul. 2011.
- [43] A.-W. Harzing, (2023). *Harzing's Publish or Perish*. Accessed: Jul. 8, 2024. [Online]. Available: <https://harzing.com>
- [44] J. Beall, "Criteria for determining predatory open-access publishers," *Tech. Rep.*, 2015.
- [45] J. Beall, "What i learned from predatory publishers," *Biochemia Medica*, vol. 27, no. 2, pp. 273–278, Jun. 2017.
- [46] S. Keshav, "How to read a paper," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 37, no. 3, pp. 83–84, 2007.
- [47] P. Mayring, "Qualitative content analysis: Theoretical foundation, basic procedures and software solution," *Tech. Rep.*, 2014.
- [48] M. Bengtsson, "How to plan and perform a qualitative study using content analysis," *NursingPlus Open*, vol. 2, pp. 8–14, 2016.
- [49] U. Kuckartz, *Qualitative Text Analysis: A Guide to Methods, Practice and Using Software*. Newbury Park, CA, USA: Sage, 2014.
- [50] J. M. Merigó, W. Pedrycz, R. Weber, and C. De la Sotta, "Fifty years of information sciences: A bibliometric overview," *Inf. Sci.*, vol. 432, pp. 245–268, May 2018.
- [51] A. A. Pimenta, A. Portela, C. D. Oliveira, and R. M. Ribeiro, "A bibliometria nas pesquisas acadêmicas," *Scientia*, vol. 4, no. 7, pp. 1–13, 2017.
- [52] SCImago Research Group. (2025). *SJR—SCImago Journal and Country Rank*. Accessed: Sep. 11, 2025. [Online]. Available: <http://www.scimagojr.com>
- [53] A. Harzing and R. van der Wal, "Google scholar as a new source for citation analysis," *Ethics Sci. Environ. Politics*, vol. 8, pp. 61–73, Jun. 2008.
- [54] Google LLC. (2025). *Google Scholar*. [Online]. Available: <https://scholar.google.com>
- [55] H. Naveed, A. U. Khan, S. Qiu, M. Saqib, S. Anwar, M. Usman, N. Akhtar, N. Barnes, and A. Mian, "A comprehensive overview of large language models," *ACM Trans. Intell. Syst. Technol.*, vol. 16, no. 5, pp. 1–72, 2025.
- [56] M. C. M. Lee, H. Scheepers, A. K. H. Lui, and E. W. T. Ngai, "AI orientation, capabilities, and business value: Case study," *J. Comput. Inf. Syst.*, pp. 1–20, Nov. 2024.
- [57] S. Afrin, S. Roksana, and R. Akram, "AI-enhanced robotic process automation: A review of intelligent automation innovations," *IEEE Access*, vol. 13, pp. 173–197, 2025.
- [58] H. Du et al., "Beyond deep reinforcement learning: A tutorial on generative diffusion models in network optimization," 2023, *arXiv:2308.05384*.
- [59] J. Siderska, S. N. B. M. Aini, and D. Kedziora, "Complementing robotic process automation with generative artificial intelligence (ChatGPT), case of robocorp," in *Proc. Future Inf. Commun. Conf*. Cham, Switzerland: Springer, vol. 2024, pp. 37–53.
- [60] A. Casciani, M. L. Bernardi, M. Cimitile, and A. Marrella, "Conversational systems for ai-augmented business process management," in *Proc. Int. Conf. Res. Challenges Inf. Sci.*, 2024, pp. 183–200.
- [61] M.-S. Jameaba, "Digitalization, emerging technologies, and financial stability: Challenges and opportunities for the indonesian banking sector and beyond, emerging technologies, and financial stability: Challenges and opportunities for the Indonesian banking sector and beyond," *Tech. Rep.*, 2024.
- [62] E. Nica, C. Alpopi, A.-L. D. Constantin, and M. Melenciu, "Employee engagement analytics, generative artificial intelligence and task automation tools, and computerbased teamwork simulations in immersive multi-sensory virtual spaces," *Contemp. Readings Law Social Justice*, vol. 15, no. 2, 2023.
- [63] C. Novak, L. Pfahlsberger, S. Bala, K. Revoredo, and J. Mendling, "Enhancing decision-making of IT demand management with process mining," *Bus. Process Manage. J.*, vol. 29, no. 8, pp. 230–259, Dec. 2023.
- [64] R. Bell and H. Bell, "Entrepreneurship education in the era of generative artificial intelligence," *Entrepreneurship Educ.*, vol. 6, no. 3, pp. 229–244, Sep. 2023.
- [65] J. R. Ilagan, J. B. Ilagan, C. L. Basallo, and Z. M. Alabastro, "Exploratory prompting of large language models to act as co-pilots for augmenting business process work in document classification," *Proc. Comput. Sci.*, vol. 237, pp. 420–425, 2024.
- [66] Z. Zeng, M. Watson, N. Cho, S. Rahimi, S. Reynolds, T. Balch, and M. Veloso, "FlowMind: Automatic workflow generation with LLMs," in *Proc. 4th ACM Int. Conf. AI Finance*, Nov. 2023, pp. 73–81.
- [67] S. Feuerriegel, J. Hartmann, C. Janiesch, and P. Zschech, "Generative AI," *Bus. Inf. Syst. Eng.*, vol. 66, no. 1, pp. 111–126, 2024.
- [68] L. Banh and G. Strobel, "Generative artificial intelligence," *Electron. Markets*, vol. 33, no. 1, p. 63, 2023.
- [69] T. Redek and U. Godnov, "From data to decision: Distilling decision intelligence from user-generated content," *Kybernetes*, vol. 53, no. 13, pp. 1–23, Dec. 2024.
- [70] M. Grupac, R. Miklencicova, M. Barker, and L. C. Voumik, "Generative artificial intelligence and haptic and biometric sensor technologies, spatial data mining and image processing computational algorithms, and virtual team performance and behavior tracking tools in immersive extended reality environments," *Anal. Metaphysics*, vol. 22, pp. 81–100, Jul. 2023.
- [71] V. V. Dengov, K. Zvarikova, and R.-S. Balica, "Generative artificial intelligence and movement and behavior tracking tools, remote sensing and cognitive computing systems, and immersive audiovisual content in virtually simulated workspace environments," *Anal. Metaphysics*, vol. 22, pp. 274–293, May 2023.

- [72] E. Nica, A. Androniceanu, O.-M. Sabie, and A. Mis, "Generative artificial intelligence and workplace tracking systems, synthetic training data, and computer-based teamwork simulations in 3D immersive spaces," *Anal. Metaphysics*, vol. 22, pp. 24–42, Jul. 2023.
- [73] V. Makarov, C. Chabbert, E. Koletou, F. Psomopoulos, N. Kurbatova, S. Ramirez, C. Nelson, P. Natarajan, and B. Neupane, "Good machine learning practices: Learnings from the modern pharmaceutical discovery enterprise," *Comput. Biol. Med.*, vol. 177, Jul. 2024, Art. no. 108632.
- [74] P. F. Ke and K. C. Ng, "Human-AI synergy in survey development: Implications from large language models in business and research," *ACM Trans. Manage. Inf. Syst.*, vol. 16, no. 1, pp. 1–39, Mar. 2025.
- [75] M. Poliak, A. Poliakova, and G. Lăzăroi, "Image processing computational algorithms, multimodal behavioral analytics, and generative artificial intelligence and ambient scene detection tools in immersive interconnected 3D worlds," *Contemp. Readings Law Social Justice*, vol. 15, no. 2, pp. 103–121, 2023.
- [76] J. Haase, W. Kremser, H. Leopold, J. Mendling, L. Onnasch, and R. Plattfaut, "Interdisciplinary directions for researching the effects of robotic process automation and large language models on business processes," *Commun. Assoc. Inf. Syst.*, vol. 54, no. 1, pp. 579–604, 2024.
- [77] M. Demir and Ş. Ş. Demir, "Is ChatGPT the right technology for service individualization and value co-creation? Evidence from the travel industry," *J. Travel Tourism Marketing*, vol. 40, no. 5, pp. 383–398, Jun. 2023.
- [78] T. Kampik, C. Warmuth, A. Rebmann, R. Agam, L. N. P. Egger, A. Gerber, J. Hoffart, J. Kolk, P. Herzig, G. Decker, H. van der Aa, A. Polyvyanyy, S. Rinderle-Ma, I. Weber, and M. Weidlich, "Large process models: A vision for business process management in the age of generative AI," 2023, *arXiv:2309.00900*.
- [79] S. Weinzierl, S. Zilker, S. Dunzer, and M. Matzner, "Machine learning in business process management: A systematic literature review," 2024, *arXiv:2405.16396*.
- [80] K. Gura, A. AlAkoum, M. Melenciu, and S. Henley, "Mobile biometric and sentiment data, generative artificial intelligence and behavior tracking tools, and wearable sensor-based and connected monitoring devices in immersive interconnected 3D worlds," *Anal. Metaphysics*, vol. 22, pp. 255–273, Apr. 2023.
- [81] C. Schreiber, A. Abbad-Andaloussi, and B. Weber, "On the cognitive and behavioral effects of abstraction and fragmentation in modularized process models," *Inf. Syst.*, vol. 125, Nov. 2024, Art. no. 102424.
- [82] T. Frajt, K. F. Michalikova, and R. Balica, "Productivity software tools, wearable augmented reality devices, and generative artificial intelligence and immersive remote collaboration systems in virtually simulated workspace environments," *Contemp. Readings Law Social Justice*, vol. 15, no. 2, 2023.
- [83] S. Ayad and F. Alsayoud, "Prompt engineering techniques for semantic enhancement in business process models," *Bus. Process Manage. J.*, vol. 30, no. 7, pp. 2611–2641, Nov. 2024.
- [84] L. Ackermann et al., "Recent advances in data-driven business process management," 2024, *arXiv:2406.01786*.
- [85] M. Nankya, A. Mugisa, Y. Usman, A. Upadhyay, and R. Chataut, "Security and privacy in E-health systems: A review of ai and machine learning techniques," *IEEE Access*, vol. 12, pp. 148796–148816, 2024.
- [86] A. Jain, S. Paliwal, M. Sharma, L. Vig, and G. Shroff, "SmartFlow: Robotic process automation using LLMs," 2024, *arXiv:2405.12842*.
- [87] M.-S. Baier, J. Lockl, M. Röglinger, and R. Weidlich, "Success factors of process digitalization projects—Insights from an exploratory study," *Bus. Process Manage. J.*, vol. 28, no. 2, pp. 325–347, Mar. 2022.
- [88] I. Beerepoot et al., "The biggest business process management problems to solve before we die," *Comput. Ind.*, vol. 146, Jun. 2023, Art. no. 103837.
- [89] A. Atalan, "The ChatGPT application on quality management: A comprehensive review," *J. Manage. Analytics*, vol. 12, no. 2, pp. 229–259, Apr. 2025.
- [90] P. Sliž, "The role of ChatGPT in elevating customer experience and efficiency in automotive after-sales business processes," *Appl. Syst. Innov.*, vol. 7, no. 2, p. 29, Mar. 2024.
- [91] R. K. Sharma, G. Bharathy, F. Karimi, A. V. Mishra, and M. Prasad, "Thematic analysis of big data in financial institutions using NLP techniques with a cloud computing perspective: A systematic literature review," *Information*, vol. 14, no. 10, p. 577, Oct. 2023.
- [92] M. Syed Ibrahim, A. Hanif, F. Q. Jamal, and A. Ahsan, "Towards successful business process improvement—An extension of change acceleration process model," *PLoS ONE*, vol. 14, no. 11, Nov. 2019, Art. no. e0225669.
- [93] S. Sharma, "Trustworthy artificial intelligence: Design of AI governance framework," *Strategic Anal.*, vol. 47, no. 5, pp. 443–464, Sep. 2023.
- [94] P. Ceravolo, S. B. Junior, E. Damiani, and W. Van Der Aalst, "Tuning machine learning to address process mining requirements," *IEEE Access*, vol. 12, pp. 24583–24555, 2024.
- [95] L. Norton, A. Howell, J. DiGirolamo, and T. Hayes, "Using artificial intelligence in consulting psychology," *Consulting Psychol. J.*, vol. 76, no. 2, pp. 137–162, 2024.
- [96] M. Bader, J. Antony, R. Jayaraman, V. Swarnakar, R. S. Goonetilleke, M. Maalouf, J. A. Garza-Reyes, and K. Linderman, "Why do process improvement projects fail in organizations? A review and future research agenda," *Int. J. Lean Six Sigma*, vol. 15, no. 3, pp. 664–690, Apr. 2024.
- [97] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," *J. Bus. Res.*, vol. 133, pp. 285–296, Sep. 2021.
- [98] A. H. Alsharif, N. Salleh, and R. Baharun, "Bibliometric analysis," *J. Theor. Appl. Inf. Technol.*, vol. 98, no. 15, pp. 2948–2962, 2020.
- [99] N. J. van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, pp. 523–538, Aug. 2010.
- [100] Centre for Science and Technology Studies (CWTS), Leiden University. (2024). *VOSviewer*. Accessed: Jul. 8, 2024. [Online]. Available: <https://www.vosviewer.com>
- [101] Google LLC. (2024). *Google Looker Studio*. Accessed: Jul. 8, 2024. [Online]. Available: <https://lookerstudio.google.com>
- [102] V. P. Guerrero-Bote and F. Moya-Anegón, "A further step forward in measuring journals' scientific prestige: The SJR2 indicator," *J. Informetrics*, vol. 6, no. 4, pp. 674–688, Oct. 2012.
- [103] H. Taş, A. Mikhaylova, H. Ödemiş, E. Ç. Uçar, and M. Söğüt, "Characteristics of articles published in different quartiles: A bibliometric analysis of sport and exercise psychology journals," *Baltic J. Sport Health Sci.*, vol. 4, no. 123, pp. 35–44, Feb. 2022.
- [104] L. Bornmann and H.-D. Daniel, "Does the h-index for ranking of scientists really work?" *Scientometrics*, vol. 65, no. 3, pp. 391–392, Dec. 2005.
- [105] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 46, pp. 16569–16572, Nov. 2005.
- [106] W. Lou and J. He, "Does author affiliation reputation affect uncitedness?" *Proc. Assoc. Inf. Sci. Technol.*, vol. 52, no. 1, pp. 1–4, Jan. 2015.
- [107] L. Bass, P. Clements, and R. Kazman, *Software Architecture in Practice*, 4th ed., Reading, MA, USA: Addison-Wesley, 2021.
- [108] B. Shneiderman, "Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered AI systems," *ACM Trans. Interact. Intell. Syst.*, vol. 10, no. 4, pp. 1–31, Dec. 2020.
- [109] S. Amershi et al., "Guidelines for human-AI interaction," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2019.
- [110] J. Banks, J. S. Carson, B. L. Nelson, and D. M. Nicol, *Discrete-Event System Simulation*, 5th ed., Upper Saddle River, NJ, USA: Pearson, 2010.
- [111] M. Weske, "Business process management architectures," in *Business Process Management: Concepts, Languages, Architectures*. Cham, Switzerland: Springer, 2012, pp. 333–371.
- [112] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Mach. Intell.*, vol. 1, pp. 206–215, May 2019.
- [113] M. Power, *Organized Uncertainty: Designing a World of Risk Management*. London, U.K.: Oxford Univ. Press, 2007.
- [114] *Risk Management—Guidelines*, Standard ISO 31000:2018, 2018.
- [115] *Enterprise Risk Management: Integrating With Strategy and Performance*, Committee of Sponsoring Organizations of the Treadway Commission (COSO), New York, NY, USA, 2017.
- [116] R. Anderson, *Security Engineering: A Guide to Building Dependable Distributed Systems*. Hoboken, NJ, USA: Wiley, 2010.
- [117] *Information Security, Cybersecurity and Privacy Protection—Information Security Controls*, Standard ISO/IEC 27002:2022, 2022.
- [118] *Information Technology—Security Techniques—Information Security Risk Management*, Standard ISO/IEC 27005:2018, 2018.
- [119] *Security Techniques—Extension to ISO/IEC 27001 and ISO/IEC 27002 for Privacy Information Management*, Standard ISO/IEC 27701:2019, 2019.

- [120] A. Saltelli et al., “Five ways to ensure that models serve society: A manifesto,” *Nature*, vol. 582, no. 7813, pp. 482–484, Jun. 2020.
- [121] E. Breck, S. Cai, E. Nielsen, M. Salib, and D. Sculley, “The ML test score: A rubric for ML production readiness and technical debt reduction,” in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2017, pp. 1123–1132.
- [122] *Information Security, Cybersecurity and Privacy Protection—Information Security Management Systems—Requirements*, Standard ISO/IEC 27001:2022, 2022.
- [123] Object Management Group, (2011) *Business Process Model and Notation (BPMN) Version 2.0*. [Online]. Available: <https://www.omg.org/spec/BPMN/2.0>
- [124] M. L. George, *Lean Six Sigma: Combining Six Sigma Quality With Lean Production Speed*. New York, NY, USA: McGraw-Hill, 2002.
- [125] T. Kaufmann, P. Weng, V. Bengs, and E. Hüllermeier, “A survey of reinforcement learning from human feedback,” *Trans. Mach. Learn. Res.*, 2024.
- [126] N. Prakash and K. Mathewson, “Conceptualization and framework of hybrid intelligence systems,” *Conceptualization and framework of hybrid intelligence systems*, Cornell University, Tech. Rep., 2012.
- [127] D. Sculley, G. Holt, D. Golovin, E. Davydov, T. Phillips, D. Ebner, V. Chaudhary, M. Young, J.-F. Crespo, and D. Dennison, “Hidden technical debt in machine learning systems,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015.
- [128] L. Sánchez-González, F. García, F. Ruiz, and J. Mendling, “Quality indicators for business process models from a gateway complexity perspective,” *Inf. Softw. Technol.*, vol. 54, no. 11, pp. 1159–1174, Nov. 2012.



in Public Services of Brazil. He is also an Associate Researcher with INESC Technology and Science—Associate Laboratory, Portugal. He also has a Cybersecurity Executive Certification from the Hebrew University of Jerusalem.

ANDERSON S. ARAÚJO received the B.Sc. degree in computer science from the Federal University of Rio Grande do Norte, Brazil, and the M.Sc. degree in systems engineering from the Federal University of Rio de Janeiro (COPPE/UFRJ). He is currently pursuing the Ph.D. degree in web science and technology with the University of Trás-os-Montes and Alto Douro, Vila Real, Portugal. He is a Chief Privacy Officer’s Advisor with the Ministry of Management and Innovation



Center. He is a Specialist in the digital transformation of organizations and information systems. He has more than 30 years of experience in academia and consulting, having worked in management within the corporate sector. He has published over 100 scientific papers in conferences and journals.

HENRIQUE S. MAMEDE received the B.Sc. degree in computer science and engineering, the M.Sc. degree in informatics, and the Ph.D. degree in information systems and technologies. He is an Associate Professor (with Habilitation) with the Department of Sciences and Technology, Portuguese Open University. He is an invited Professor with the Information Management School, New University of Lisbon. He is a Senior Researcher with INESC TEC and the HumanISE



Specialist title from polytechnic institutes Guarda, Castelo Branco, and Viseu. He is an Associate Professor with the NOVA Information Management School (NOVA IMS), NOVA University of Lisbon, teaching artificial intelligence, information systems, and digital systems courses in data science and informatics systems degrees. Before that, he was an invited Professor with the University of Trás-os-Montes and Alto Douro (UTAD) and Minho University (UM). He integrates several national and international conference scientific committees and has authored several academic publications. He was an Elected Member of the Order of Engineers and the APDSI Board. He was the Microsoft Portugal Academic Computer Science Program Manager. Before that, he occupied senior management positions with Santander Bank Group Companies and has developed computer engineering activities for about 15 years.

VÍTOR SANTOS received the D.E.A. degree from the University of Minho, the B.Sc. degree in informatics engineering from Cocite, the Postgraduate degree in computer science from the Faculty of Science, Lisbon University, and the M.Sc. degree in information systems science and the Ph.D. degree in science and information and technology systems from the University of Minho. He is currently pursuing the second Ph.D. degree in culture and communication. He received the Computer



with UTAD. He is a Senior Researcher with the Institute for Systems and Computer Engineering, Technology and Science (INESC TEC). He was a President of the School of Science and Technology, UTAD, which has over 2 000 students, from 2022 to 2025. He has co-authored more than 160 Scopus-indexed articles, including over 60 in journals. He has participated as a Researcher in more than 25 projects, with financial support from scientific agencies and industrial partners. His research interests include medical image analysis, computer vision, and machine learning.

VÍTOR FILIPE received the bachelor’s degree in electrical and computer engineering and the master’s degree in informatics from the University of Minho in 1994 and 1998, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Trás-os-Montes and Alto Douro (UTAD), Portugal, in 2003. He is an Associate Professor (with Habilitation) in electrical and computing engineering, specializing in digital signal processing and biomedical image

...