Department of Industrial Engineering & Maintenance

# Bridging the Gap: A Review of Robotic Process Automation and Process Mining Integration

MESSAI Haythem, BENTAYEB Adel

Master's thesis in Industrial Engineering

**Advisor:**
GHOMARI Leila

**Academic year:**
2023-2024

**Abstract:** This master's thesis reviews the literature on Robotic Process Automation (RPA) and Process Mining (PM), exploring their development, integration, and impact on business process management. The analysis includes a detailed examination of the methodologies employed in RPA and PM, their operational synergies, and the resultant enhancements in process efficiency and data-driven decision-making in various industries. This study categorizes existing research into thematic areas, identifies current knowledge gaps, and suggests future research directions. Significantly, it highlights how the convergence of RPA and PM can provide strategic insights within organizations, augmenting processes that traditionally require intensive manual oversight. The findings indicate that the combined application of RPA and PM enhances operational efficiency and provides strategic insights that can lead to sustainable competitive advantages.

**Key-Words:** Process Mining, Process Discovery, Event Logs, Conformance Checking, Process Automation, Robotic Process Automation, Software Robots, Task Automation, Process Efficiency, Automation Analytics

## Contents

# 1. Introduction

Digital transformation has led to significant changes in how businesses operate, with Robotic Process Automation (RPA) and Process Mining at the forefront of enhancing operational efficiencies and analytical capabilities. This master's thesis explores the interplay between RPA and Process Mining, focusing on their roles in automating and optimizing business processes. RPA and Process Mining, though distinct technologies are increasingly integrated to provide end-to-end business process improvements. RPA automates repetitive and rule-based tasks that were once manually performed by humans, thus reducing error and increasing efficiency. Meanwhile, Process Mining provides deep insights into business process performance, using existing data from IT systems to visualize and analyze how business processes are conducted in reality. This analysis helps identify bottlenecks and inefficiencies but also aids in pinpointing the optimal areas for RPA application, thereby ensuring targeted and effective automation strategies.

This research attempts to bridge the gap between theoretical frameworks and practical RPA and Process Mining implementations. By reviewing the literature on these technologies, this thesis aims to synthesize current knowledge, identify gaps in the existing research, and provide a comprehensive overview of the combined utility of RPA and Process Mining. This approach is particularly relevant given the rapid adoption of these technologies across industries and the need for a nuanced understanding of their impact on organizational performance and strategy.

This thesis is structured into several sections. The first sections review the foundational concepts and technologies underlying RPA and Process Mining. Subsequent sections delve into integrating these technologies, exploring case studies and theoretical models highlighting their synergistic effects. The final section discusses the implications of these findings for future research and practical application in business settings.

# 2. Process mining

Data mining is a well-known concept that involves extracting valuable information from data for various purposes, such as decision-making and prediction. Process mining, on the other hand, is similar to data mining but specifically focused on managing processes.

## 2.1. Process Mining Definition

Process mining is a research domain that develops innovative methods to gather insights from event logs [1]. It involves applying specialized algorithms to event log data to identify trends, patterns, and details of how a process unfolds[2]. This technique combines data science with process analytics to discover, validate, and enhance workflows, providing organizations with valuable insights to optimize their processes and drive better business outcomes.

## 2.2. Process Mining Techniques

Process mining provides various uses for process improvement using event data stored in today's information systems. These techniques encompass aspects such as business process intelligence, business activity monitoring, and business process management (BPM), but process mining is commonly used for three primary purposes:

**a. Process Discovery:** A process mining technique derives process models from event logs devoid of preexisting information. It is a primary technique within process mining to uncover the actual occurrences by scrutinizing the recorded events in an event log. This method proves especially valuable in elucidating the genuine conduct of a process, distinct from its anticipated or optimal trajectory. Process discovery facilitates organizational comprehension of inefficiencies, bottlenecks, and deviations present within their processes, offering significant insights conducive to process enhancement [3]; there are many algorithms for process discovery, such as the heuristic miner [4]. The evaluation of process model quality encompasses diverse perspectives and employs varied assessment methodologies, as underscored in [5]. One such method involves utilizing model-log metrics, which entails comparing the traces present in the event log and those derived from the mined model. Alternatively, another approach compares a pre-existing model with the model generated through mining, necessitating the presence of an apriori model (referred to as model-model metrics)[5]

**b. Conformance Checking:** according to the process mining manifesto [3], this technique involves comparing an existing process model with an event log of the same process. The aim is to determine if the behavior recorded in the event log aligns with the behavior described by the model and vice versa. This comparison can help identify discrepancies, deviations, or commonalities between the modeled and observed process behavior. Different models, including procedural, organizational, declarative process models, business rules/policies, and laws, can be considered for conformance checking. Conventional methods of conformance checking include Token-Based Replay [6], Alignment-Based Techniques[7], and Declarative Conformance Checking, which compares an event log with a declarative process model instead of a procedural [8]. In recent years, researchers have been trying to create more stochastic-aware methods to extract additional information and perform more in-depth analyses such as time and cost instead of control flow only. Conformance-checking techniques face significant challenges when applied to systems characterized by weak supervision, where limited data availability hinders the extraction of meaningful insights such as anomaly detection and process improvement opportunities. Addressing this challenge, [9] proposes an innovative approach utilizing an activity-based Variational Autoencoder (VAE) with a Bidirectional Long Short-Term Memory (Bi-LSTM) architecture. A comparative analysis against established methods, including BiNet, Denoising Autoencoder, Conditional Probabilistic Model, and Anomaly-Free Automaton, demonstrates the superior performance of the Bi-LSTM VAE with Self-Attention mechanism, as measured by precision and recall metrics. This study concludes that the proposed activity and weighted-based classification model effectively leverages the Bi-LSTM VAE with Self-Attention to outperform existing anomaly detection techniques in scenarios with limited data. This finding aligns with the results presented in [10], which demonstrate the efficacy of the proposed approach in surpassing five competing methods by efficiently utilizing scarce anomalous examples.

**c. Process Enhancement:** process enhancement denotes the augmentation or refinement of an extant process model by integrating insights derived from actual process data. This enhancement endeavor seeks to elucidate problematic process pathways, uncover deviations from the expected course, and explain their ramifications on organizational operations. Enriching process models enables enterprises to discern segments ripe for automation, conduct root cause analyses, and initiate process amelioration initiatives. Process enhancement is a pivotal facet of process mining, empowering organizations to refine their operations using empirical data and insights extracted from event logs [1].

## 2.3. Process mining algorithms

Research has shown that the most prominent algorithms in process discovery depend on features of event logs and process characteristics[11]. Discovery algorithms in process mining encounter significant challenges when applied to real-world event logs, particularly those arising from unstructured processes. These challenges include noise, duplicate tasks, hidden tasks, non-free choice constructs, and loops, as identified in [12]. The inherent complexity and variability of real-world processes[5] contribute to these issues and impact the performance of discovery algorithms. Consequently, the effectiveness of such algorithms is contingent upon the specific characteristics of the event log and the underlying process it represents.

**a. Alpha Algorithm:** A foundational approach in process mining introduced to address the discovery of workflow nets from event logs. This algorithm identifies and analyzes the workflow by examining the sequences of activities in a process, distinguishing between parallel and sequential operations. It works by extracting the starting and ending activities, identifying pairs of activities directly followed by each other, and discerning parallelism within the log. Based on these observations, the Alpha Algorithm constructs a Petri net that models the process, capturing the dynamics and concurrency of tasks. This model helps visualize the process flow, identify bottlenecks, and improve efficiency. The Alpha Algorithm's ability to derive a direct representation of a process from event data marks a significant advancement in process mining, providing a systematic approach to uncovering the underlying structures in the process log.[13].

**b. Heuristics Miner:** The Heuristic Miner (HM) algorithm is pivotal in process mining for discovering the control-flow perspective of a process model primarily by analyzing the order of events in a log rather than their timing or correlation across different cases. This algorithm utilizes event logs to identify dependencies between activities, mapping out how one activity may precede another. The methodology of HM comprises three main steps: creating a dependency graph to visualize and assess the relationships between activities; detecting complex structures like AND/XOR splits or joins and non-observable tasks; and mining for long-distance dependencies which are less apparent but significant in understanding the process dynamics. By iteratively processing traces within the log, HM builds a refined model that highlights the most frequent patterns of behavior, providing insights into the process structure that are invaluable for optimizing operations and understanding workflow dynamics[14].

**c. Genetic Process Mining:** The genetic algorithm for process mining represents a robust approach to modeling and optimizing various business and healthcare processes. This algorithm, a type of evolutionary algorithm, simulates natural selection to generate high-quality solutions for complex problems. It iteratively evolves a population of individual solutions toward an optimal solution using operations like selection, crossover, mutation, and termination. In process mining, the genetic algorithm analyzes event logs to discover, enhance, or check the conformance of the process models to real-world processes. It is beneficial in dealing with noisy, incomplete, or unusual data logs. The genetic algorithm stands out for its ability to handle diverse data and to model complex relationships within the process data, thus providing significant insights that can drive process optimization and innovation[**?** ].

**d. Fuzzy miner:** The Fuzzy Miner algorithm, as outlined in the paper by Sarno, Sinaga, and Sungkono [15], is a sophisticated tool used for process mining, particularly in detecting anomalies and fraud within business processes. This algorithm leverages fuzzy logic to handle process data's inherent uncertainty and variability, which is standard in dynamic and complex business environments. By constructing fuzzy models from event logs, the Fuzzy Miner algorithm effectively maps and analyzes deviations from standard operating procedures (SOPs), identifying unusual patterns that may indicate fraudulent activities. The strength of this approach lies in its ability to adapt the level of detail and abstraction based on the fuzziness of the data, providing a flexible and robust framework for uncovering subtle yet critical irregularities that rigid, deterministic methods might overlook [15].

## 2.4. Process mining assessment metrics

Evaluating process discovery methodologies is crucial for assessing the effectiveness and applicability of process models generated from event logs. This chapter discusses the primary dimensions and specific metrics used to evaluate these methodologies, focusing on accuracy and comprehensibility, and introduces additional metrics that have gained prominence in recent research.

**a. Accuracy and Comprehensibility:** process discovery methodologies are evaluated along two main dimensions: accuracy and comprehensibility [14]. Accuracy refers to the degree to which a process discovery technique accurately reflects the behavior recorded in an event log. It challenges the balance between overgeneralization, which can omit critical details, and excessive granularity, which may introduce noise and irrelevant elements into the model [1]. Comprehensibility involves the understandability of the discovered process models, emphasizing their ease of interpretation and simplicity. This metric assesses the ability of stakeholders to effectively grasp and utilize the process models in practical scenarios[1].

**b. Conformance Checking Metrics:** conformance checking is integral to validating the accuracy of process models against actual event logs. Developing a State-Based Deterministic Finite Automaton (SDFA) is a noteworthy method wherein the SDFA is constructed iteratively from an event log. Initially starting with a single state, this automaton expands by adding new states and transitions as it encounters new events in the log, thus forming a probabilistic model through normalized transition probabilities [16].

**c. Precision and Recall:** precision and recall are critical metrics derived from information retrieval and classification. They evaluate the specificity and completeness of the elements within a discovered model, respectively. Precision measures the proportion of accurately identified elements within the model, reflecting its specificity and exclusion of irrelevant details. Recall assesses the extent to which a model captures all relevant process elements, indicating its comprehensiveness. Balancing these metrics is crucial as overemphasis on one can detrimentally affect the utility of the process model[9].

**d. Additional Key Metrics:** recent studies have highlighted several other metrics that are essential for a holistic evaluation of process models:

- Fitness: assesses how well a model can reproduce the behavior seen in the event log using various methods, such as token-based replay or behavioral alignment [17].
- Generalization: Measures the model's ability to predict unseen instances, ensuring it is not overfitted to the training data [18].
- Simplicity: Evaluates the model's ease of understanding based on its structure and complexity [18].
- Overall Accuracy: Encompasses various aspects of model quality, including precision, recall, fitness, and generalization, to provide a comprehensive evaluation of its effectiveness[19] .

The selection of appropriate metrics (see Table 1) depends on the specific goals and context of the process mining project. Considerations include the model's purpose, the complexity of the process, the stability of the process environment, and the quality of the event log. Through careful metric selection, researchers and practitioners can derive significant insights into the capabilities and limitations of discovered process models, thereby enhancing their practical applications in organizational contexts.

Table 1: Model Evaluation Parameters

| **Parameter** | **Focus** | **Usage** |
|---|---|---|
| Precision | Accuracy of positive predictions | Conformance checking, filtering evaluation, algorithm comparison |
| Recall | Completeness in capturing positive cases | Conformance checking, filtering evaluation, algorithm comparison |
| Fitness | Fit to the observed data | Overall model evaluation, model selection |
| Generalization | Ability to handle unseen data | Overall model evaluation, model selection |
| Accuracy | Ease of understanding | Overall model evaluation, model selection, communication |
| Simplicity | Overall correctness and reliability | Overall model evaluation, model selection |

## 2.5. Process mining benefits and challenges

Process mining is an innovative analytical approach that leverages data mining techniques to analyze business processes. It has gained substantial attention due to its ability to provide detailed, data-driven insights and its applicability across various industries, including healthcare[20] , banking, finance[21], and production industries [22] One of the primary benefits of process mining is its reliability in extracting meaningful information from event logs generated by various information systems. This reliability stems from the objectivity of the data-driven approach (see Fig. 1), which minimizes human biases and errors. Furthermore, its wide range of applications shows process mining's versatility. For instance, it aids in conformance checking, identifying the root causes of deviations, pinpointing bottlenecks, and predicting future trends or possible outcomes of process adjustments. These applications demonstrate process mining's critical role in understanding and optimizing business processes.

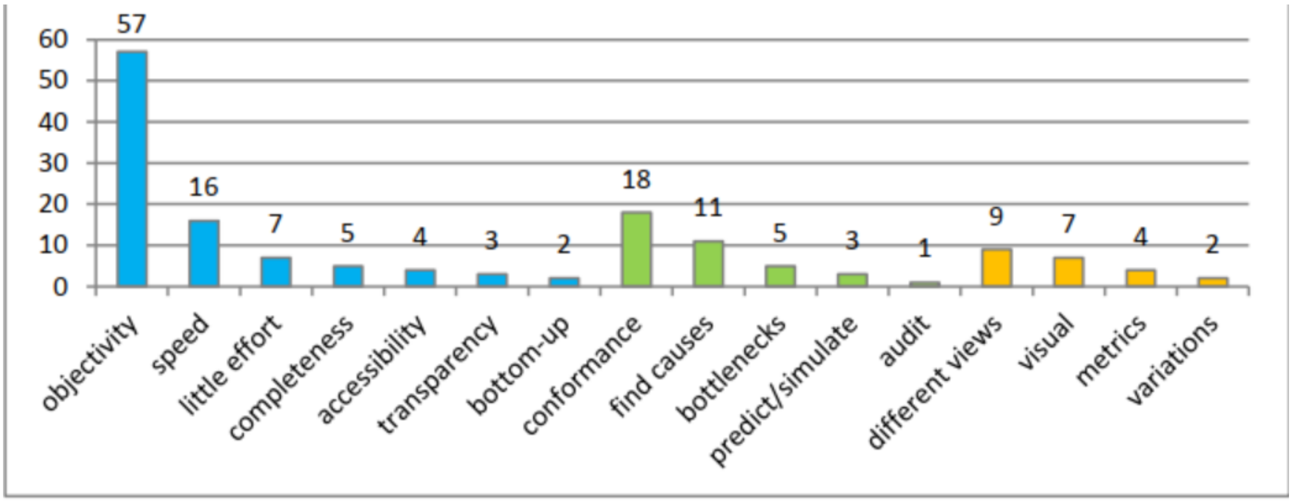


Figure 1: Benefits of process mining techniques (42 questions, 94 respondents) (blue: characteristic, green: application, orange: representation) According to Jean Claes et al

Process mining significantly enhances transparency and process efficiency. It provides a bottom-up analysis approach, ensuring a comprehensive view of the organizational process. This is further supported by the ability to visualize process flows in various representations, which helps stakeholders understand complex data sets and process metrics with little effort. The visual representations and metrics developed through process mining make tracking a process's complete journey possible, promoting a transparent audit trail and facilitating continuous improvement.

The strategic benefits of process mining are realized through its impact on decision-making. Process mining supports strategic decisions that enhance productivity and efficiency by providing detailed insights into process performance and compliance. Simulating process changes before implementation also enables decision-makers to foresee potential impacts and proactively adjust strategies.

The adaptability of process mining extends its benefits across multiple sectors. In healthcare, process mining can improve patient flow and optimize treatment processes. In the banking and finance sectors, it enhances compliance and fraud detection. Moreover, process mining is instrumental in streamlining manufacturing processes and reducing waste in production industries. These examples highlight the broad applicability and significant advantages of process mining in improving operational efficiencies and contributing to cost reductions and enhanced service delivery.

Despite its potential, process mining's adoption and effectiveness are hindered by several challenges (see Fig. 2). A primary obstacle in process mining is accessing the correct data and ensuring its quality. Research shows these are the most significant barriers organizations face when implementing process mining techniques. Poor data quality or incomplete data can lead to inaccurate process models, compromising the results' reliability and usefulness.

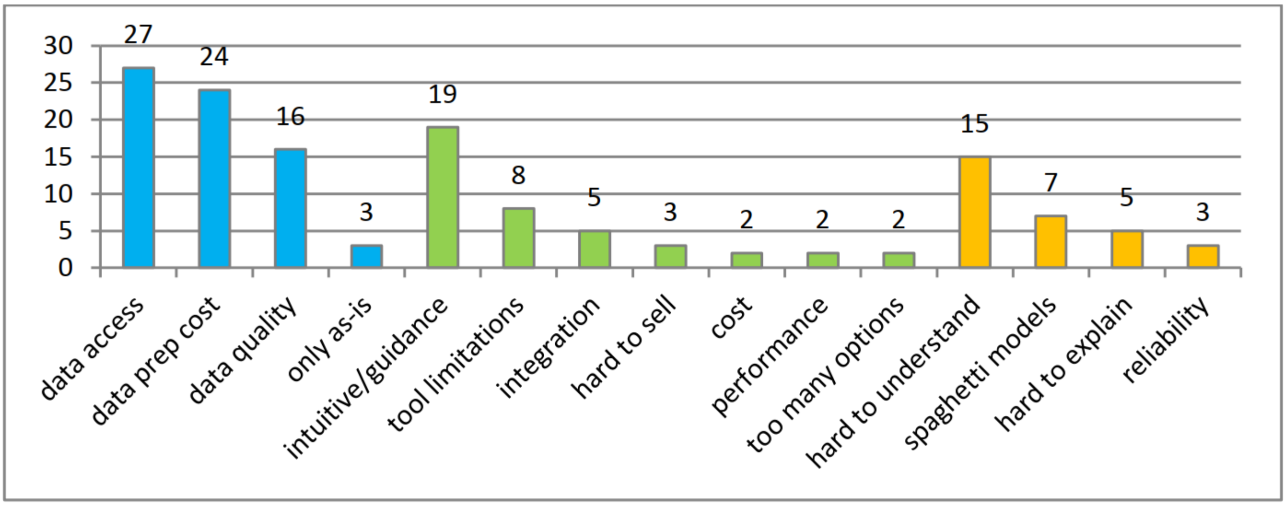


Figure 2: Drawbacks of process mining techniques (question 5 2, 90 respondents) (blue input, green techniques, orange output). According to Jean Claes et al

The complexity of process mining techniques and the usability of related tools also pose considerable challenges. For practitioners, especially those at the managerial level, the process mining tools may seem too complex or unintuitive, making them hard to use and understand. This complexity can discourage adoption, mainly when the benefits are not immediately apparent to decision-makers. In addition, integrating process mining tools with existing IT infrastructures is another significant challenge. This integration often involves substantial costs and requires technical expertise, which may not be readily available. Additionally, the overall cost of implementing process mining solutions, including training and maintenance, can be prohibitive for some organizations. Process mining outputs, such as process models and diagnostic analytics, can be challenging to interpret. For instance, complex models, often called "spaghetti models," are hard to understand and communicate to stakeholders. This lack of clarity can reduce the actionable insights derived from process-mining endeavors.

### 2.6. Future Directions in Process Mining

Looking ahead, the future of process mining lies in addressing these challenges while leveraging advancements in technology and methodology:
**a. Enhanced Data Management Techniques:** It will be crucial to improve the accessibility and quality of data through advanced data management techniques. This includes developing more sophisticated data cleaning tools and methodologies to ensure the integrity and completeness of data used in process mining.
**b. User-friendly Tools:** More intuitive process mining tools that cater to users with varying technical expertise are needed. Simplifying the user interface and providing more explicit guidance on using tools can help make process mining more accessible to a broader audience.
**c. Integration Solutions:**Developing better integration solutions that reduce the cost and complexity of deploying process mining tools will encourage more organizations to adopt these techniques. This could involve creating more modular and scalable tools that can easily fit into different IT environments.
**d. Advanced Analytical Techniques:** Future research should also focus on refining analytical techniques to handle complex data and provide more precise, interpretable models. Artificial intelligence and machine learning could play a significant role in developing these advanced techniques.

## 3. Robotic process automation

### 3.1. Definition

Robotic Process Automation (RPA) is a rapidly growing approach to process automation that uses software robots to mimic human tasks. RPA automates repetitive tasks or workflows previously performed manually, streamlining business processes through technology and software[23]. According to Dr. Choi et al [24]., RPA is a category of software tools that automates repetitive tasks involving structured data, rules, and user interface interactions. The primary objective of RPA is to minimize human effort in labor-intensive processes, thereby increasing the speed and efficiency of high-volume transactional tasks.

### 3.2. RPA vs. Conventional Automation

RPA is a means of automation, similar to conventional script automation and the automation included in standard IT implementations. So, firstly, what is meant by conventional automation and typical IT implementations? **Conventional automation** refers to automating it through conventional programming techniques or other tools. It involves direct integration with backend systems through APIs or other connectivity means. In general, this needs to be developed and maintained by an IT staff who are highly expert in the systems and technologies underneath. Traditional automation, integrated much deeper into system architecture, handles many tasks, such as data processing, system operation, and complex business logic [25].

**Standard IT implementation** encompasses the comprehensive deployment of IT solutions through a systematic process that includes planning, system requirements analysis, system design, development, integration, testing, and maintenance. This sequence aims to ensure the IT solutions meet business requirements effectively and adhere to predefined budgets and timelines. This process highlights the critical integration of new technologies within an organization's IT framework to improve or replace existing functionalities [26]. According to Rajagopal, G. & Ramamoorthy [27], IT implementation, such as CRM, is highly intricate. Implementation requires an expert as it involves complex integration at the data or application layers and concerns the whole business process. On the other hand, RPA implementation consists of training the bot directly on the software (e.g., UiPath) and affects only the application layer.

RPA operates on the front end, mimicking a human user's behavior. On the contrary, conventional automation requires access to the backend as it involves controlling machines to conduct certain operations in certain phases (see Table 2).

Table 2: Comparison of Conventional Automation and RPA

| Criteria | Conventional Automation | RPA |
|---|---|---|
| IT infrastructure adjustments | Necessary | Unnecessary |
| Human behavior emulation | Incapable | Capable |
| Coding knowledge | Necessary | Recommended but not necessary |
| Customization flexibility | High | Low |
| Speed | Fast | Slower compared to CA, but still much faster than manual |

## 3.3. RPA components

According to [24], RPA can be divided into three components (see Fig. 3):

- **Robots:** Virtual software bots that perform mundane, repetitive tasks instead of human resources. They can be "attended" type bots, which work alongside their human counterparts, or "unattended" bots, which work independently and require little to no human involvement. .
- **Orchestrator:** an RPA orchestrator is a management server that schedules, monitors, manages, and audits robots. It is used in the development, testing, and production [28]. As a highly scalable platform that connects the studio to the robots, the orchestrator also bridges the development environment (studio) and the robots, enabling efficient and centralized control of automated processes.
- **Studio:** the RPA studio is a user-friendly, intuitive tool for designing and automating robotic processes. Additionally, it allows users to create and automate workflows for robots.

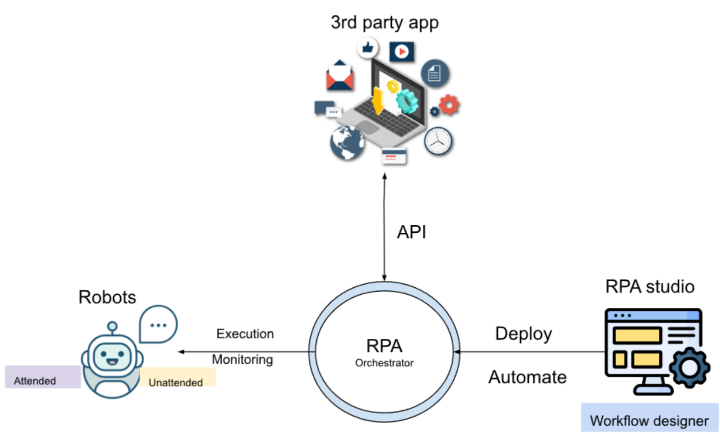


Figure 3: RPA Components According to Dr Choi et al

## 3.4. RPA tools

In [28], Amira Khan conducted a comparative study on the three most common RPA tools: Ui Path, Automation Anywhere, and BluePrism. Other tools include Windows Power Automate, Taskt, RoboCorp, and many more. She specified that tools can have two types of architectures. It is either a client-server architecture, meaning that every node can be a client or a server. Or a web-based orchestrator that links automated tasks to create a unified workflow; a web-based architecture can be like the .Net Framework. We will now conduct a comparative study on some popular RPA tools: UiPath, Automation Anywhere, Blue Prism, and TASKT, a popular open-source solution.
Then We will briefly overview each tool, its components, and its advantages and disadvantages. Then, we shall compare these tools according to determined criteria (see Table 3).
**a. Automation Anywhere (AA)** is a software platform that enables businesses to automate their entire business processes using Robotic Process Automation (RPA). Automation Anywhere provides all the features needed for a company in RPA through its Control Room, serving bot development, configuration, and monitoring in one single and central environment. These bots can be used for many tasks, including data entry, validation, and complex calculations, mainly using AI and ML technologies[29]. AA supports three types of bot creation: Task Bots for rule-based tasks, Meta Bots for reusable building blocks, and IQ Bots for processing unstructured data. It also provides three types of recorders to automate functions by replicating user actions. It offers features such as BOT INSIGHTS for data visualization and business insights, BOT FARM for usage-based RPA tool purchases, and BOT STORE for plug-and-play bots.
**b. Uipath** is one of the top platforms for Robotic Process Automation (RPA), and it offers automation functionality combined with process discovery and analytics. UiPath platform facilitates the software robots (SRs) development, deployment, and management designed to perform automated repetitive and rules-based business tasks. Other vital components are the orchestrator of task management, workflow designers, and analytic tools[30]. Components of UiPath are [28]: (1) Core RPA Capabilities: Allows accessible building and deployment. (2) Process Discovery and Analytics Tools: These are business-oriented ideas whereby the impact of the process on automation is provided. (3) Orchestrator: It shall be a central control that manages task assignments and performance appraisal. (4) Workflow Designer allows you to design processes with a drag-and-drop surface. These mimic human operations on digital systems and carry out robotic process automation—software Robots (SRs). Advantages of UiPath include improved efficiency, ease of scaling the company, and performance analytics[30]. Disadvantages include the constant updating of the software, its complexity in set-up and management, potential high costs, and dependence on the current IT system. The UiPath software has been actively improved to integrate newer innovations like machine learning and AI as part of its advancing functions.
**c. Blue Prism** is a robotic automation software used to automate the business processing system through the integration of presentation. This approach, formerly known as "screen scraping," has been remodeled to permit efficient interaction with various applications, simplifying business process automation. It empowers business analysts with the ease of low-technical skills to create and modify automation through direct interaction with application user interfaces. Blue Prism provides functionality that allows the automation of interfaces from contemporary web interfaces to the most mature mainframe applications, including interface automation [31]. Blue Prism mainly includes several components. (1) Visual Business Objects (VBOs) are application interface adapters that graphically create and execute specific tasks, such as logging in or entering data, without using coding through Object Studio. (2) Process Studio is a graphical tool for defining and sequencing the steps in a business process, using VBOs for application interactions. (3) Control Room: This room oversees the execution of Blue Prism processes and handles process control, monitoring, and scheduling. (4) System Manager: Administer users, manage user settings, administer processes, deploy processes, and manage the overall system for successful, efficient, and secure operation. (5) SQL Server Database that stores the details about the processes and VBOs for management and audit purposes. One of the Blue Prism advantages is efficiency: Easily and fast, you can automate business processes easily and quickly through user interfaces without changing the applications. It is cost-effective: It is cheaper than the traditional way of doing things and can be applied in low-value processes. Also, adaptability: Easily variable to changing business changes. It has broad compatibility and can interface virtually any application with a user interface. It ensures robust security, such as safeguarding encrypted credentials and role-based access controls. One of the disadvantages of Blue Prism is performance issues: Complex multi-screen processes and extensive data retrieval can be a struggle. Also, UI Limitation: Only automates tasks that can be managed through the user interface, lacking direct backend access. In addition to some maintenance issues due to updates on the significant changes in application interfaces, Skill Dependency requires a sound fundamental understanding of cutting across the business processes and the Blue Prism tools for effective implementation.
**d. TASKT** (formerly known as sharpRPA) represents the pioneering instance of a genuinely free, user-friendly, and open-source process automation tool developed within the .NET Framework using C#. TASKT empowers users to create and customize process automation workflows without coding application logic [32]. It offers an extensive suite of task management features, including subtasks, alerts and notifications, task visualization tools, and comprehen-

sive reporting and analytics capabilities. Additionally, its integrations seamlessly connect with other applications, enhancing workflow productivity and efficiency [33]. One of its advantages is that it is free and open-source, making it reachable even for small businesses and individuals. It provides an intuitive interface and accessible commands to automate tasks, making it easy for users to come up to speed quickly. It supports web and desktop applications, thus making it very flexible and allowing for different automation scenarios. One of its disadvantages is that, being a smaller project, it may not provide the same level of support or community activity as some of the more extensive commercial RPA tools. It is limited to the Windows environment, implying that this software would be ineffective when implementing across-platform functionality. It has the smaller scale of the project can mean less frequency and scope with which updates or new features are made.

**e. Robocorp** According to the Robocorp and Rpabotsoworld [34] website, Robocorp is an RPA open-source tool based on Python that is used for automation across different platforms. This software, purposefully made for non-code and code user interfaces, targets developers and non-developers. That rests on the cloud-native architecture foundation, allowing it to handle data and execute tasks with fortitude, whether on-cloud or on-premise. Its components are the Robocode Lab, An IDE that supports the development of automation scripts and the Control Room, a central dashboard for deploying, managing, and scaling bots and automation. The Robocorp Cloud offers cloud services for bot execution, making it easier to manage and deploy bots remotely. One of its advantages [35] is the flexibility and Open Source; Robocorp is perfectly poised to give users the ideal flexibility to connect an extensive range of Python libraries and APIs inside their automation workflow, making it functional and flexible. It is cost-effective and supports a consumption-based pricing model since its features are affordable for the user and what is only utilized. It supports scalability; the system will expand operations excellently and take in those of small and large businesses; it will do so without the need for colossal infrastructures. It provides community and documentation; the community is robust, and the guides are well-documented. New learners, hence, find the tool accessible for learning and troubleshooting. One of its disadvantages[35] is the complexity of setup; setting up the environment for Robocorp can be time-consuming and challenging for users who need it for quick deployment. Its interface and usability of the tool might not be straightforward for users who do not possess coding skills, therefore increasing the learning curve. Some users added that the tool could use huge memory and space, requiring robust systems specifications for better functionality.

| **Criteria** | **UiPath** | **Taskt** | **Robocorp** | **Automation Anywhere** | **BluePrism** |
|---|---|---|---|---|---|
| Architecture | .Net Framework | .Net Framework | Robot Framework and Jupyter Notebook | Client Server | Client Server |
| Availability | Community Edition (Bots cannot be distributed), 60-day free trial (UI Path Pro) | Open Source | Consumption-based pricing with free trial; packages for various usage levels from personal to enterprise, costs vary based on usage. | One month free trial (Industry edition), community edition (BotCreator rights only) | One month free trial, Learning edition (1 digital worker, 15 processes) |
| Usability | Simple | Simple (according to Github) | Simple (for the paid version) and offers some complexity for the free version. | Complex | Simple |
| Automatable Processes | back/front office | Back/front office | Back/front office | back/front office | Back-office |

| Criteria | UiPath | Taskt | Robocorp | Automation Anywhere | BluePrism |
|---|---|---|---|---|---|
| Recorders | Innovative, screen, and web (Desktop and web applications) | recorder available | It does not have a recorder | Primary, web, desktop, image, and Citrix | No recorders |
| Cognitive ability | Medium | Medium | High Cognitive Ability due to AI integration | Medium | Low |

Table 3: RPA tools comparison according to some of the criteria mentioned by [28] and [29]

### 3.5. RPA Project Lifecycle

Implementing Robotic Process Automation (RPA) involves a structured six-phase lifecycle (see Fig. 4). The process starts with the Discovery Phase, where suitable processes for automation are identified. The Analysis Phase then assesses the feasibility of automating these processes. The specifications for the automated processes are outlined in the design phase. The Development Phase transforms these designs into actionable components. The Deployment Phase follows, where robots are executed in operational environments. Control and Monitoring oversee the robots' performance, while the Evaluation Phase evaluates their effectiveness, facilitating continuous improvement [27].

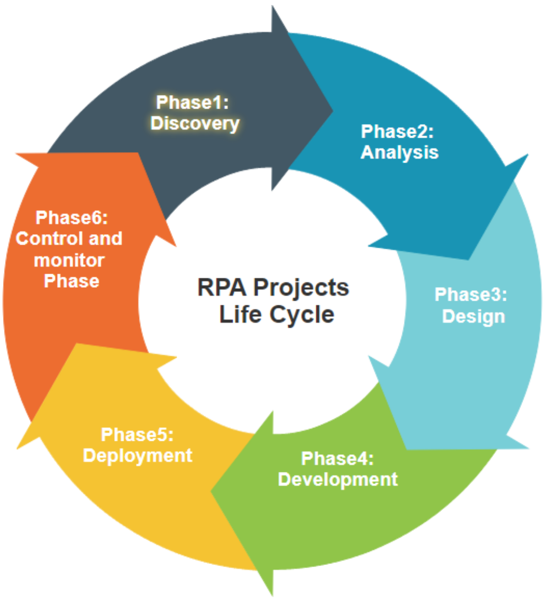


Figure 4: RPA Project Lifecycle According to Gowri Rajagopal et al

### 3.6. Robotic Process Automation (RPA) Benefits and Challenges

Robotic Process Automation (RPA) presents numerous benefits that can significantly enhance business operations. Firstly, RPA enables rapid efficiency gains and cost savings, often within weeks or months of implementation (source). The initial investment and return on investment (ROI) are manageable and predictable, making RPA an attractive option for cost-conscious businesses. Furthermore, RPA provides a solution that requires minimal changes to existing applications and business processes, facilitating incremental improvements without substantial disruption[36]. RPA operates 24/7, ensuring continuous productivity and operational availability. This capability enhances efficiency and maintains consistent compliance with regulatory requirements. Additionally, RPA is scalable; as the system expands, it achieves more significant cost advantages and can support extensive data generation necessary for Lean Six Sigma programs, thereby improving process repeatability and reducing human error [37].

Despite its advantages, RPA faces significant challenges. The perception of RPA is polarized; some view it as a revolutionary advancement akin to artificial intelligence, while others dismiss it as overhyped by marketing efforts. One of the primary technical challenges is the maintenance required when underlying software updates occur. These updates can disrupt RPA by altering critical elements that the bots interact with, necessitating frequent adjustments to maintain functionality. Furthermore, RPA is particularly effective in environments with legacy systems that lack API or database access, functioning as a "glue" between disparate software applications. However, as more modern systems with better integration capabilities become prevalent, the utility of RPA may diminish, highlighting its suitability mainly for outdated systems [37] [36],.

There are alternative views on RPA's efficacy. Critics argue that RPA merely accelerates existing processes without addressing underlying inefficiencies. This perspective suggests that rather than relying on RPA, organizations should focus on reducing software fragmentation and improving process efficiency through more traditional automation techniques. This approach would streamline operations and mitigate the accumulating software burdens that could lead to future operational issues. In summary, while RPA offers substantial benefits in terms of efficiency, scalability, and compliance, it also faces challenges related to maintenance and relevance in modern IT environments. The debate continues on whether RPA represents a technological advancement or a temporary solution to deeper systemic issues.

## 3.7. The RPA implementation Frameworks

Implementing robotic Process Automation (RPA) necessitates a well-defined framework, as the complexity of these projects demands structured guidance to ensure efficacy and scalability. However as RPA is a new technology and it being relatively to other IT technologies immature; it does not have many well structured frameworks. As such we have nitpicked the latest and most popular frameworks to discuss in our article

### 3.7.1 Process mining-based RPA frameworks

Using process mining-based framework for an RPA project: Unlike conventional frameworks, a PM-based framework makes less room for guesswork and depends directly on data logs generated by process behavior, thus giving a better understanding of the process and detecting RPA opportunities in a better way. Moreover, it helps during and after the implementation of the RPA bot (as described in the previous section). In what follows, we describe some frameworks that rely on process mining to implement RPA in organizations.
**a. PLOST framework:** Hilde Jongling [38] created the "Prioritized List Of Suitable Tasks" Framework. It utilizes process mining and consists of eight qualitative and quantitative steps that must be performed chronologically (see Fig. 5).

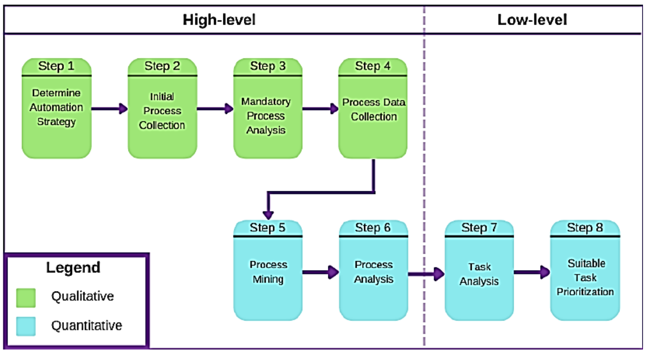


Figure 5: The PLOST framework steps

First, the automation strategy should be determined to customize the framework to the organization's needs. The automation strategy consists of two parts: the prioritization of the business values and the determination of the risk Level. Next, processes are gathered from the organization through semi-structured interviews with domain experts, including various roles from managers to system administrators. In the third step, processes selected in the previous phase undergo assessment based on six mandatory qualitative criteria. These criteria are digital and structured input, easy data access, few variations, repetitiveness, clear rules, and maturity. In the fourth step, the process data is collected for the processes in the revised process selection from the previous step. With the event logs of the processes in the revised process selection, the next step is to apply process mining. The framework user can choose which process mining tool is used for this step. The Process Analysis step assesses the remaining processes from the revised process selection against different quantitative criteria. This happens at a high level. It is done with the help of the output of the previous step. The requirements are Cycle Time, Case Frequency, Activity Frequency, Standardization, Length, Automation rate, Human Error Prone. The Task Analysis step involves assessing individual tasks within the identified processes using specific quantitative criteria, focusing on the low-level details. The requirements are task-specific, and their values are obtained through visualizations in the fifth step, ensuring a detailed analysis of process components. The requirements are Activity Frequency, Case Frequency, Duration, Automation Rate, Human Error Prone, and Irregular Labor. The final step of the framework produces a prioritized list of tasks suitable for RPA automation. This output relies on two key components: the automation strategy established in the initial step and the task analysis conducted in the seventh step. The six criteria analyzed in the previous step align with various business values outlined in the automation strategy, facilitating the final task prioritization process.

**Comments on this framework:**
While this framework provides a step-by-step detailed guide to selecting suitable tasks for automation and focuses on repeatable tasks that match the criteria for automation, it misses the ROI (return on investment) in the process selection phase. Moreover, this framework is specialized for IT-related processes, and although it is still applicable in other functions, it may pose a challenge.

**b. A framework for implementing Process mining and RPA in Organizations** in 2023, a paper[39] proposed a framework that depends on PM to assess tasks' suitability for automation; in said framework, the process mining helps implement RPA by:

- Discovering Automation Opportunities: Process mining techniques are used to analyze the collected event logs and identify candidate routines that can be automated. By analyzing the process execution data, process mining can reveal patterns, bottlenecks, and inefficiencies. This analysis helps organizations identify which processes can benefit the most from RPA automation.
- Assessing Feasibility: Process mining helps assess the feasibility of automating identified processes using RPA. It provides insights into the execution frequency, the number of process variants, and the number of exceptions encountered. This information helps determine whether a process suits automation and provides a basis for decision-making.
- Process Understanding and Improvement: Process mining enables organizations to understand their current processes deeply. It helps uncover hidden process variations, deviations, and inefficiencies that may not be apparent through traditional documentation or manual observations. This understanding is crucial for e actively implementing RPA, allowing organizations to optimize and streamline processes before automation.
- Continuous Monitoring: After implementing RPA, process mining continues to monitor the performance of the automated processes. By comparing the actual execution of the automated processes with the expected process models, process mining can identify any deviations, errors, or unexpected behaviors in the RPA implementation. This monitoring capability helps organizations ensure their automated processes' accuracy, consistency, and efficiency.

**Comments on this framework:** This framework uses process mining to identify and assess opportunities for automation so that a suitable selection can be made for RPA. It also provides continuous monitoring to maintain the RPA-implemented process efficiency. However, it might lack the same treatment regarding the scale of RPA solutions and the cultural and change management aspects required for adoption success.

### 3.7.2 Frameworks that do not primarily rely on process mining

Frameworks that are not based on process mining require domain expertise and might depend on a lot of guesswork, which, as a result, may lead to a wrong task choice and the failure of the RPA project as a whole. They are necessary in enterprises that do not have well-structured data.

**a. A framework for implementing robotic process automation projects:** a paper [40] proposed a robust and adaptable framework, offering a significant tool for organizations to approach RPA implementations systematically and effectively. The framework comprises four phases: Initialization, Implementation, Scaling, and Rollout. The initialization phase identifies areas in the enterprise that can benefit from automation using RPA technology. The implementation phase involves selecting processes and RPA software, creating a pilot to test project feasibility, and evaluating the business case to determine the feasibility of full-scale implementation. The scaling phase involves rolling out the RPA project to cover other processes, increasing automation scale and percentage within tasks, and implementing RPA support processes to ensure reliable operation and maintenance. A Center of Excellence is set up to oversee the organization's RPA-related activities, development, and enhancement. The framework ensures that RPA projects align with business objectives and are supported by mechanisms to ensure reliable operation and maintenance.

**Comments on this framework:** This framework is comprehensive and methodically structured to cover the breadth of considerations necessary for successfully adopting and scaling RPA technologies. It could be improved by incorporating specific identification methodologies, such as process mining and employee workshops, to discover automation opportunities systematically. This would help identify and prioritize suitable processes based on potential return on investment and ease of implementation.

**b. A UiPath Academy proposed framework:** As mentioned above in 3.4 Uipath is the among the most popular RPA tools, and it is the only one to provide a complete framework for RPA implementation included in its academy training. [41]

The Uipath academy framework outlines the different steps, deliverables, and team members included primarily in each step, (see fig 6)

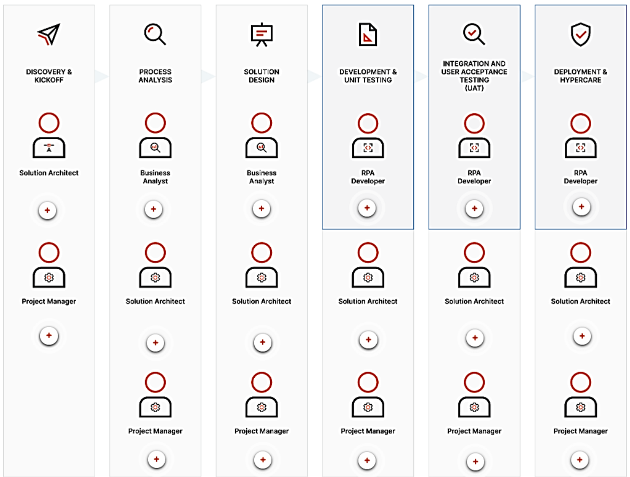


Figure 6: Phases of Uipath Framework and stakeholders concerned by each phase

The automation process involves several steps, including discovery and kickoff, process analysis, solution design, development and unit testing, integration and user acceptance testing (UAT), deployment and hypercare, and deployment and hypercare. The setup team evaluates potential automation projects based on their complexity and intricacy, establishing schedules and resources for successful completion. Next, process analysis involves assessing the customer's process requirements and determining the degree of automation based on the study and complexity of the process. The technical team designs a future state flow and maps out various modules to complete the automation. The development and unit testing involves the creation of modules from the design using PDD and SDD papers, with each module tested individually in set situations before moving forward. The next step is testing and combining modules. The user acceptance testing (UAT) is conducted by users with oversight from the implementing team. Users coordinate with business groups to draft a test plan covering all expected and exceptional use cases. The end of UAT is marked by signoff. Finally, the deployment of robots and hypercare oversee the bots' running. The team reviews automation cases in daily meetings, ensuring errors or issues are fixed quickly.

**Comments on this framework:** This framework provides in detail the deliverables and responsibilities of each stakeholder in an RPA project and focuses primarily on BOT development. However, it is lacking (compared to other frameworks) in process selection and automation potential identification.

# 4. Integrating Process Mining and Robotic Process Automation

Process Mining and Robotic Process Automation (RPA) are two transformative technologies that synergize to enhance business process optimization. Process mining delves into transaction logs to unearth real-time insights into business operations, revealing the actual workflow, inefficiencies, bottlenecks, and compliance deviations. This technology maps out the 'as-is' state of processes, paving the way for informed decision-making.

Conversely, RPA employs software bots to automate mundane, rule-based tasks that are typically manual and repetitive. This automation is not just about replacing human effort but enhancing process efficiency and reliability [23]

The synergy between process mining and RPA stems from their mutual goal of optimizing business processes. Here's how they complement each other (see fig 7) [42]:

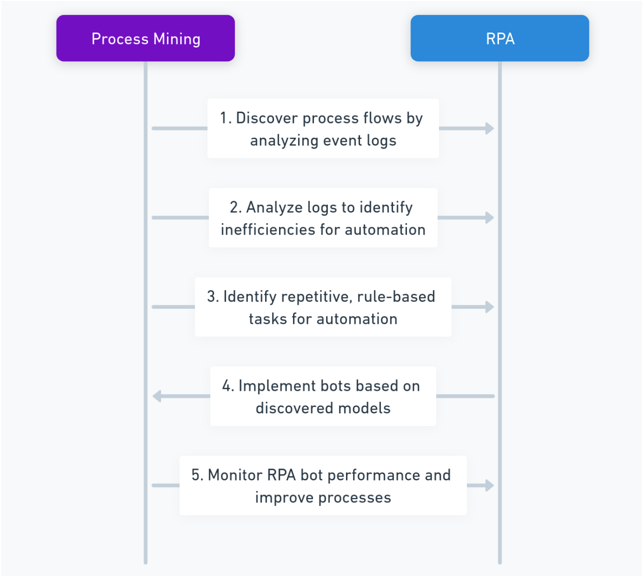


Figure 7: RPA and Process mining working together to automate processes efficiently

- **Process Discovery:** By analyzing event logs during RPA bot activities, process mining uncovers true process flows. This visibility allows organizations to pinpoint inefficiencies and prime areas for automation, ensuring that RPA efforts are targeted and effective.
- **Process Enhancement** Process mining evaluates these logs to identify bottlenecks and deviations. The insights garnered here help in choosing which processes to automate first, thus ensuring that RPA deployment has the most beneficial impact on organizational workflow.
- **Bot Discovery:** Through detailed analysis of event data, process mining identifies tasks that are repetitive, rule-based, and voluminous. These tasks are ideal candidates for RPA, helping organizations automate the right processes.
- **Bot Implementation:** RPA takes the baton by automating the tasks identified through process mining. The bots are guided by detailed process models developed through mining, ensuring that the automation is aligned with actual process needs.
- **Continuous Optimization:** Post-implementation, process mining offers ongoing monitoring and analysis of RPA bot performance by scrutinizing the event logs they generate. This continuous evaluation helps in fine-tuning the bots, addressing any emerging issues, and quantitatively assessing the impact of RPA on process efficiency.

  In essence, the relationship between process mining and RPA is a continuum of improvement and automation, where each technology feeds into and enhances the other, creating a cycle of perpetual refinement and efficiency in business processes.

## 5. Conclusions

This thesis has explored the synergies between Robotic Process Automation (RPA) and Process Mining within a business process management context. Through a literature review, several key findings have emerged, such as that RPA and Process Mining are synergistic technologies that, when combined, can optimize operations and enhance process efficiency across many industries—additionally, combining Process Mining's analytical skills and RPA's execution proficiency results in a robust framework for optimizing processes. The literature shows that implementing Process Mining before RPA installation has significantly improved process comprehension, enabling more focused and efficient automation solutions. Nevertheless, significant obstacles to successfully combining RPA with Process Mining are related to data quality, process complexity, and change management. To achieve the intended objectives, these obstacles must be addressed by implementing strategic planning and continuously improving. The convergence of RPA and Process Mining represents a significant advancement in process management, offering substantial operational efficiency and decision-making support. Future research should aim to validate these findings further through more extensive empirical studies and explore the potential of emerging technologies to enhance the capabilities of RPA and Process Mining further.

## References


[1] W.M.P. van der Aalst. *Introduction*. Springer, Berlin, Heidelberg, 2011.

[2] W. van der Aalst. *Process Mining: The Missing Link*. Springer, Berlin, Heidelberg, 2016.

[3] W. et al. van der Aalst. Process mining manifesto.

In Barkaoui K. Dustdar S. Daniel, F., editor, *Business Process Management Workshops. BPM 2011. Lecture Notes in Business Information Processing*, volume 99. Springer, Berlin, Heidelberg, 2012.

[4] Nutchar Senewong Na Ayutaya, Prajin Palungsuntikul, and Wichian Premchaiswadi. Heuristic mining: Adaptive process simplification in education. In *2012 tenth international conference on ict and knowledge engineering*. IEEE, 2012.

[5] De Backer M. Vanthienen J. Baesens B. De Weerdt, J. A critical evaluation study of model-log metrics in process discovery. In Su J. zur Muehlen, M., editor, *Business Process Management Workshops. BPM 2010. Lecture Notes in Business Information Processing*, volume 66. Springer, Berlin, Heidelberg, 2011.

[6] van der Aalst W.M.P. Berti, A. A novel token-based replay technique to speed up conformance checking and process enhancement. In Kordon F. Pomello L. Koutny, M., editor, *Transactions on Petri Nets and Other Models of Concurrency XV. Lecture Notes in Computer Science*, volume 12530. Springer, Berlin, Heidelberg, 2021.

[7] Zsuzsanna Nagy and Agnes Werner-Stark. An alignment-based multi-perspective online conformance checking technique. In *University of Pannonia, Faculty of Information Technology, Department of Electrical Engineering and Information Systems*, 2022.

[8] Montali M. Peñaloza R. Maggi, F.M. Probabilistic conformance checking based on declarative process models. In La Rosa M. Herbaut, N., editor, *Advanced Information Systems Engineering. CAiSE 2020. Lecture Notes in Business Information Processing*, volume 386. Springer, Cham, 2020.

[9] Philippe Krajsic and Bogdan Franczyk. Semi-supervised anomaly detection in business process event data using self-attention based classification. *Procedia Computer Science*, 192:39–48, 2021.

[10] E.A. Elaziz, R. Fathalla, and M. Shaheen. Deep reinforcement learning for data-efficient weakly supervised business process anomaly detection. *J Big Data*, 10(33), 2023.

[11] Fundora-Ramírez O. Lazo-Cortés M.S. Roche-Escobar R. Pérez-Alfonso, D. Recommendation of process discovery algorithms through event log classification. In Martínez-Trinidad J. Sossa-Azuela J. Olvera López J. Famili F. Carrasco-Ochoa, J., editor, *Pattern Recognition. MCPR 2015. Lecture Notes in Computer Science*, volume 9116. Springer, Cham, 2015.

[12] Jochen De Weerdt, Manu De Backer, Jan Vanthienen, and Bart Baesens. A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs. *Information Systems*, 37(7):654–676, 2012.

[13] W. van der Aalst, T. Weijters, and L. Maruster. Workflow mining: discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering*, 16(9):1128–1142, 2004.

[14] Alireza Bakhshi, Erfan Hassannayebi, and Amir Hossein Sadeghi. Optimizing sepsis care through heuristics methods in process mining: A trajectory analysis. *Healthcare Analytics*, 3(100187), 2023.

[15] Sinaga F. Sarno, R. and K.R. Sungkono. Anomaly detection in business processes using process mining and fuzzy association rule learning. *J Big Data*, 7(5), 2020.

[16] Sander J.J. Leemans and Artem Polyvyanyy. Stochastic-aware precision and recall measures for conformance checking in process mining. *Information Systems*, 115(102197), 2023.

[17] Alves de Medeiros A.K.-Wen L. van Dongen, B.F. Process mining: Overview and outlook of petri net discovery algorithms. In van der Aalst W.M.P. Jensen, K., editor, *Transactions on Petri Nets and Other Models of Concurrency II. Lecture Notes in Computer Science*, volume 5460. Springer, Berlin, Heidelberg, 2009.

[18] van Dongen B.F. van der Aalst W.M.P. Buijs, J.C.A.M. On the role of fitness, precision, generalization and simplicity in process discovery. In et al. Meersman, R., editor, *On the Move to Meaningful Internet Systems: OTM 2012. Lecture Notes in Computer Science*, volume 7565. Springer, Berlin, Heidelberg, 2012.

[19] Akhil Kumar Zan Huang. A study of quality and accuracy trade-offs in process mining. *INFORMS Journal on Computing*, 24(2), 2011.

[20] Alvin C. Lin Anna M. Lin, David Baumgartner and andJosef Küng. Towards the use of standardized terms in clinical case studies for process mining in healthcare †. *Enivironmental Search and Public Health*, 17(22), 2020.

[21] M. Werner and N. Gehrke. Multilevel process mining for financial audits. *IEEE Transactions on Services Computing*, 8(6), 2015.

[22] Wilfried Sihn Julian Senoner and Torbjorn Netland. Using process mining to improve productivity in make-to-stock manufacturing. *International Journal of Production Research*, 59(16), 2021.

[23] Jerome Geyer-Klingeberg, Janina Nakladal, Fabian Baldauf, and Fabian Veit. Process mining and robotic process automation: A perfect match, 2018.

[24] Daehyoun Choi, Hind R'bigui, and Chiwoon Cho. Robotic process automation implementation challenges. In *International conference on smart computing and cyber security: strategic foresight, security challenges and innovation*, pages 297–304. Springer, 2020.

[25] Alan Richardson. *Automating and Testing a REST API*. 2017.

[26] Jag Sodhi and Prince Sodhi. *IT Project Management Handbook*. 2001.

[27] Ramamoorthy R. Rajagopal, G. Robotic process automation: The key to reviving the supply chain processes. In Banerjee J.S. De D. Bhattacharyya, S., editor, *Confluence of Artificial Intelligence and Robotic Process Automation. Smart Innovation, Systems and Technologies*, volume 335. Springer, Singapore, 2023.

[28] Sameera Khan. Comparative analysis of rpa tools-uipath, automation anywhere and blueprism. *International Journal of Computer Science and Mobile Applications*, 8(11):1–6, 2020.

[29] David Andrade. Challenges of automated software testing with robotic process automation rpa - a comparative analysis of uipath and automation anywhere. *International Journal of Intelligent Computing Research (IJICR)*, 11(1), 2020.

[30] Liliana Dobrica. Robotic process automation platform uipath. *Communications of the ACM*, 65(4):42–43, 2022.

[31] David Chappell. Introducing blue prism automating business processes with presentation integration, 2010.

[32] Taskt tool, https://github.com/saucepleez/taskt, 2024. Consulted: 20 March 2024.

[33] 6 task management software features your company needs, https://technologyadvice.com/blog/information-technology/task-management-features/, 2024. Consulted: 22 March 2024.

[34] Open source rpa tool: Robocorp , https://rpabotsworld.com/open-source-rpa-tool-robocorp/, 2024. Consulted: 20 March 2024.

[35] Robocorp reviews, https://www.trustradius.com/products/robocorp/reviews?qs=pros-and-consreviews, 2024. Consulted: 20 March 2024.

[36] Rehan Syed, Suriadi Suriadi, Michael Adams, Wasana Bandara, Sander J.J. Leemans, Chun Ouyang, Arthur H.M. ter Hofstede, Inge van de Weerd, Moe Thandar Wynn, and Hajo A. Reijers. Robotic process automation: Contemporary themes and challenges. *Computers in Industry*, 115(103162), 2020.

[37] Pereira R. Santos, F. and J.B. Vasconcelos. Toward robotic process automation implementation: an end-to-end perspective. *Business Process Management Journal*, 26(2):405–420, 2020.

[38] Hilde Jongeling. Identifying and prioritizing suitable rpa candidates in itsm using process mining techniques: Developing the plost framework. Master's thesis, MS thesis, 2022.

[39] Najah Mary El-Gharib. A framework for implementing process mining and robotic process automation in organizations. Master's thesis, School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Canada, 2022.

[40] LV. Herm, C. Janiesch, and A. et al. Helm. A framework for implementing robotic process automation projects. *Information Systems and E-Business Management*, 21:1–35, 2023.

[41] Uipath academy learning path viewer, 2024. Consulted: 15 March 2024.

[42] Volodymyr Leno, Artem Polyvyanyy, Marcello La Rosa, Marlon Dumas, and Fabrizio Maria Maggi. Action logger: Enabling process mining for robotic process automation. In *The University of Melbourne, Parkville, VIC, 3010, Australia, University of Tartu, Estonia*. 2019.