# PROCESS MINING & RPA

## How To Pick Your Automation Battles?

Prof.dr.ir. Wil van der Aalst
Lehrstuhl fur Informatik 9 / Process and Data Science
RWTH Aachen University
www.vdaalst.com

Robotic Process Automation (RPA) has lowered the threshold for process automation. Repetitive tasks done by people are handed over to software robots. For RPA, there is no need to change or replace the pre-existing information systems. Instead, software robots replace users by interacting directly with the Graphical User Interface (GUI). Actually, RPA can be seen as "the poor man's workflow management solution" because it is cheaper than traditional automation. Therefore, it can be used to automate routine work that would normally not be cost-effective. Process mining plays a key role in deciding what to automate and how. Therefore, RPA is closely related to process mining. Before introducing RPA, one needs to analyze the processes to be automated. Process mining can help to identify promising candidates. Moreover, after RPA has been implemented, process mining can be used to monitor processes even if they use a mixture of RPA, workers, and traditional automation.

## INTRODUCTION

This paper aims to relate Robotic Process Automation (RPA) and process mining and put both in a historical context. Workflow Management (WFM) has been around for several decades. In the mid-nineties, the term Straight Through Processing (STP) was used to describe the ultimate goal of WFM: Making operational processes cheaper, faster, and better by avoiding manual intervention [1]. This turned out to be challenging and many WFM projects failed. WFM was subsequently replaced by Business Process Management (BPM), which had a broader scope and put more emphasis on management aspects [2]. However, traditional BPM often relied on modeling, leading to a "disconnect" with reality. We have all seen the idealized process models expressed in languages like BPMN that completely failed to capture the real problems. Moreover, the goal should not be to model, but to improve the process at hand. This often did not happen because it would be too expensive to change the information systems or the actual inefficiencies and compliance problems remained invisible. Some will argue that nothing is new and refer to "screen scraping" (capturing data by reading text from a computer display and transferring it to a new application) and "Taylorism" (i.e., scientific management named after Frederick Winslow Taylor). However, process mining and RPA provide new ways of learning and automating routine processes [3,4,5]. This explains the growing interest in both.

## A BRIEF HISTORY OF WFM AND BPM

Since the industrial revolution, productivity has been increasing because of technical innovations, improvements in the organization of work, and the use of information technology. Adam Smith (1723-1790) showed the advantages of the division of labor. Frederick Taylor (1856-1915) introduced the initial principles of scientific management. In the seventies, people like Skip Ellis and Michael Zisman already worked on so-called office information systems, which were driven by explicit process models [2]. Skip Ellis developed the Officetalk system at Xerox PARC in the late 1970s using Information Control Nets (ICN), a variant of Petri nets, to model processes. Also, the office automation system SCOOP (System for Computerizing of Office Processes) developed by Michael Zisman used Petri nets to represent business processes. These systems can be seen as early Workflow

Management (WFM) systems. However, it took another 15 years until WFM technology was ready to be applied at a large scale. In the mid-nineties, many commercial WFM systems were available and there was the expectation that WFM systems would be an integral part of any information system.

Figure 1 shows the development of information systems over time, explaining the initial great optimism related to WFM technology. Initially, information systems were developed from scratch, i.e., everything had to be programmed, even storing and retrieving data. Soon people realized that many information systems had similar requirements with respect to data management. Therefore, this generic functionality was subcontracted to a database system. Later, generic functionality related to user interaction (forms, buttons, graphs, etc.) was subcontracted to tools that can automatically generate user interfaces. The trend to subcontract recurring functionality to generic tools continued in different areas. Workflow Management (WFM) systems are similar to Database Management (DBM) systems but focus on processes rather than data. In the mid-1990s, many WFM systems became available. These systems focused on automating workflows with little support for process analysis, process flexibility, and process management. Nevertheless, many expected that WFM systems would be as common as DBM systems. However, this did not happen. WFM systems were succeeded by Business Process Management (BPM) systems that were broader in scope. The BPM discipline combines knowledge from information technology and knowledge from management sciences and applies this to operational business processes. BPM systems are generic software systems that are driven by explicit process designs to enact and manage operational business processes. Examples of BPM systems include the software products from Pegasystems, Appian, IBM, Bizagi,

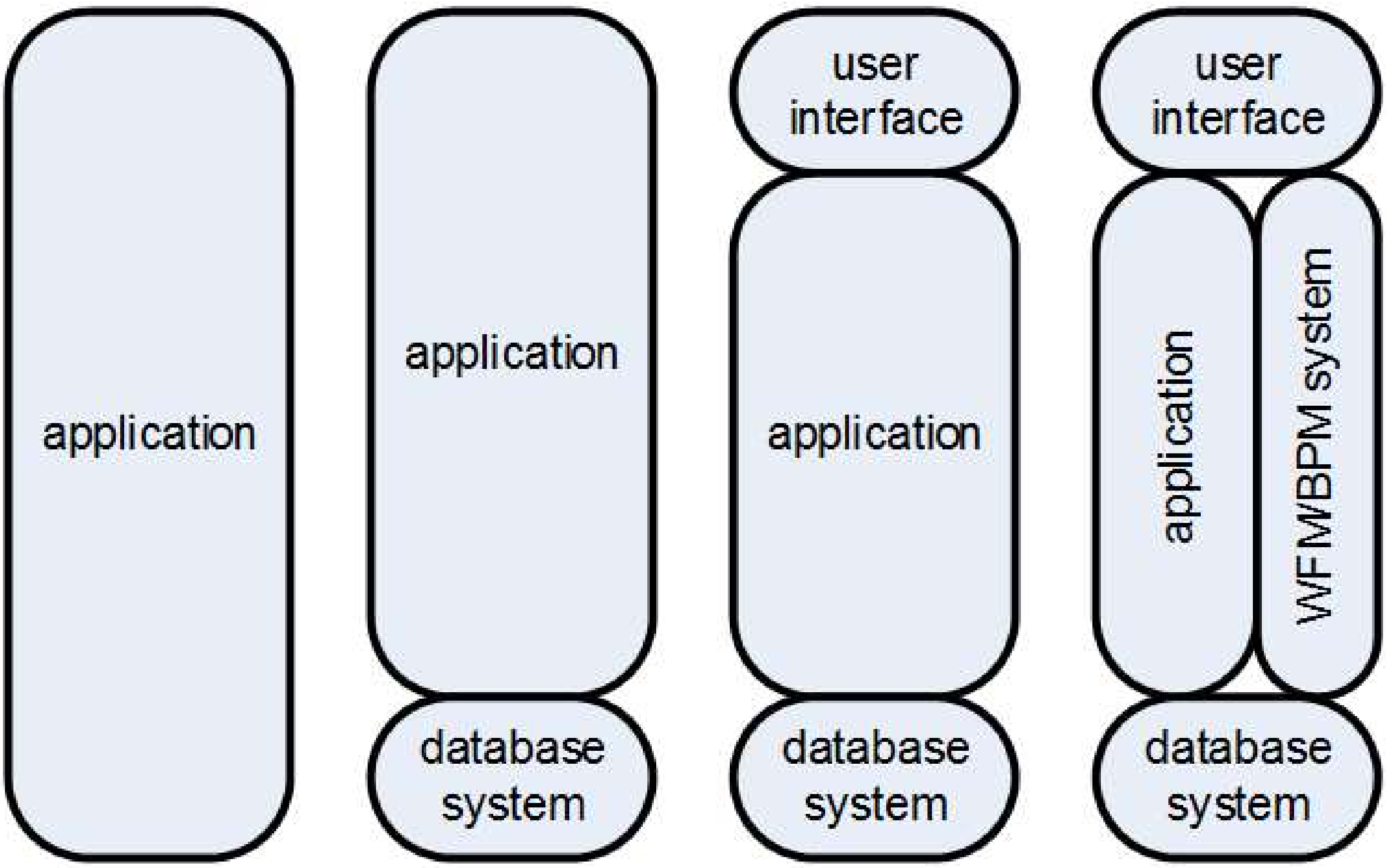



*Figure 1: Positioning of WFM/BPM systems in a historical context (based on [1,2]).*

Oracle, Software AG, TIBCO Software, Bonitasoft, Kofax, and Signavio. However, despite the availability of WFM/BPM systems, process management is not subcontracted to such systems at a scale comparable to DBM systems. The application of "pure" WFM/BPM systems is still limited to specific industries such as banking and insurance. However, WFM/BPM technology is often hidden inside other systems. For example, ERP systems like SAP and Oracle provide workflow engines. Therefore, the landscape is not so clear. Organizations such as Gartner also invent new terms such as "Intelligent Business Process Management Suites" (iBPMS), yet the actual usage of such systems remains limited.

There seem to be three main reasons why the adoption of WFM/BPM technology is low.

- Applying WFM/BPM technology is rather expensive. Processes are hardcoded in application software or not supported at all. Many processes also involve software from different vendors, making integration difficult and time-consuming.
- Although the "M" in WFM and BPM refers to "Management", the focus is on modeling and automation rather than management. Traditional WFM/BPM systems fail to learn from the event data they collect.
- Real-life processes are more complex than people like to believe. The well-know 80-20 rule applies to processes, i.e., 80% of all cases are rather simple, but explain only 20% of the complexity of the process. The remaining 20% of cases tend to be neglected by software and management, but consume 80% of the resources of an organization.

The above three obstacles for WFM/BPM explain the current interest in Robotic Process Automation (RPA) and process mining.

## RPA: THE POOR MAN'S WFM

Robotic Process Automation (RPA) is a form of automation using software robots (bots) replacing humans. The three main RPA vendors are UIPath (founded in 2005), Automation Anywhere (founded in 2003), and Blue Prism (founded in 2001). The key difference between RPA and traditional WFM/BPM is that RPA does not aim to replace existing (back-end) information systems. Instead, software robots interact with the existing information systems in the same way as humans do. In traditional WFM/BPM systems, the process is specified precisely and the WFM/BPM system orchestrates the modeled process by implementing simple activities and calling pre-existing applications through Application Programming Interfaces (APIs). In contrast, RPA software interacts with the pre-existing applications through Graphical User Interfaces (GUIs) directly replacing humans, i.e., automation is realized by taking over tasks from workers directly though the GUI. A typical RPA scenario is a sequence of copy-and-paste actions normally performed by a human. Because there is no need to replace the existing information systems, RPA can be seen as "The Poor Man's WFM".

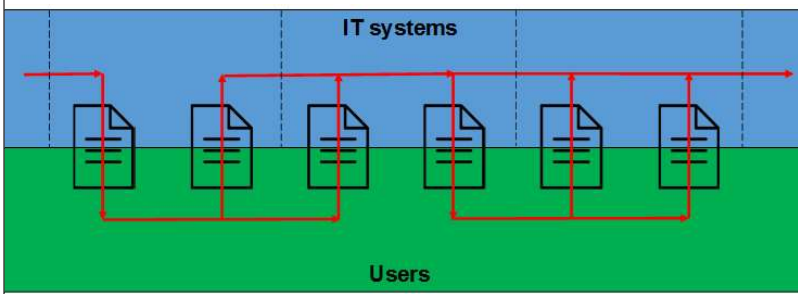



*Figure 2: Interaction between IT systems and users. RPA assumes that pre-existing information systems cannot be changed. Therefore, the RPA software accesses these systems in the same way as users do: Through the Graphical User Interfaces (GUIs) provided by these systems.*

To understand RPA, it is important to realize that workers and information are "dancing" together (see Figure 3). An information system may trigger its users and provide information. Similarly, people start applications and enter information. Consider, for example, the usage of forms. Most forms are partly prefilled with information and users complete the missing information thereby possibly triggering new actions. Sometimes the user takes the initiative and sometimes the system. When there are multiple information systems, people are often the "glue" between the different parts. See, for example, the scenario where a user copies address information from one information system to another one.

Figure 3 shows the positioning of RPA with respect to the traditional setting and the situation where WFM/BPM software is used. Both RPA and WFM/BPM automate simple tasks and provide the glue between existing information systems. WFM/BPM connects to these systems via the "backend" using APIs. RPA connects to these systems via the "frontend" using GUIs. In [5] the terms "inside-out" and "outside-in" are used for respectively the backend WFM/BPM approach and the frontend RPA approach.

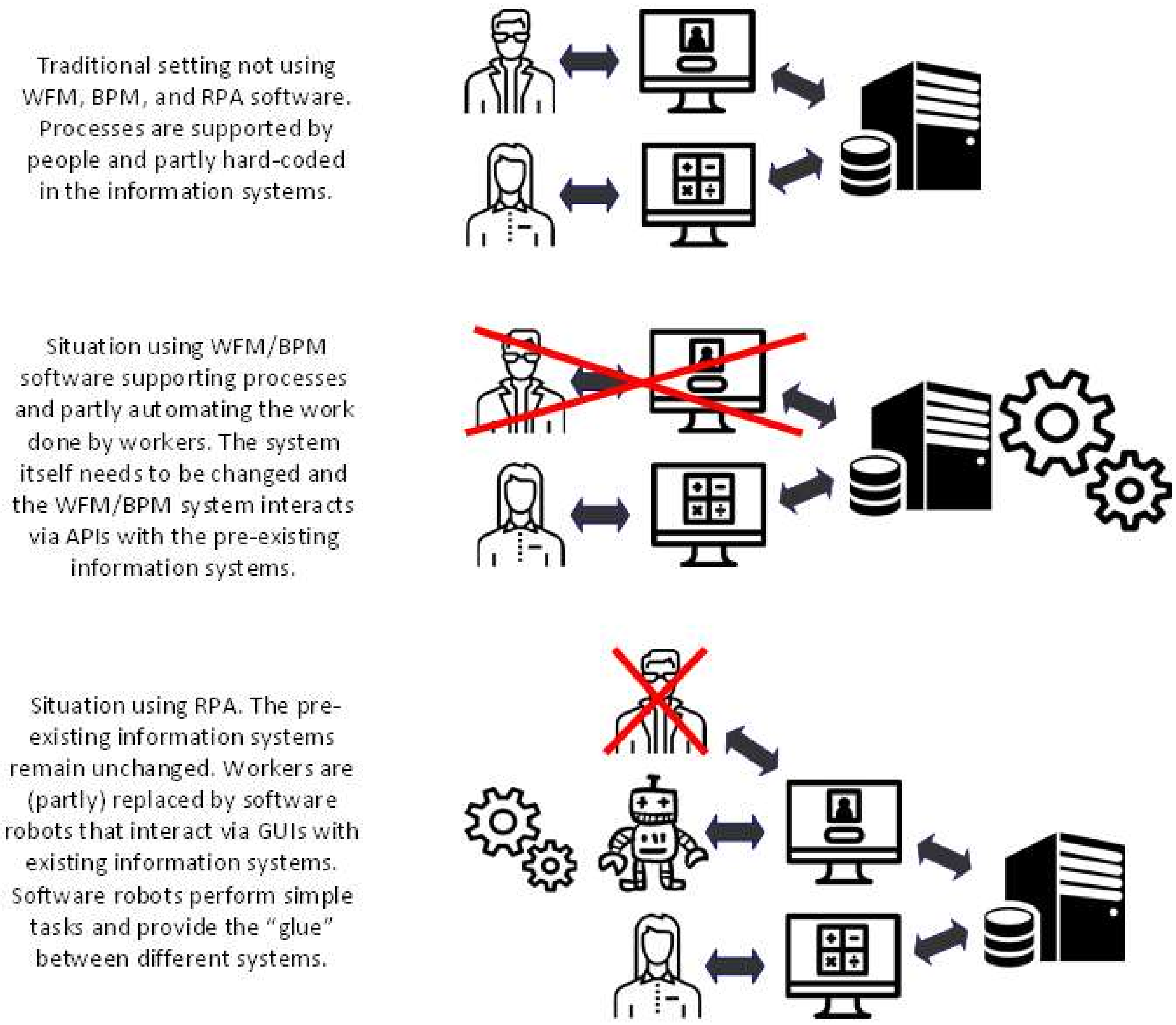



*Figure 3: Three situations: (a) traditional setting, (b) WFM/BPM setting, and (c) RPA setting.*

RPA can be much more cost-effective than traditional automation because the information systems do not need to be changed. RPA can automate various mundane and routine tasks in the workplace. At the same time, there are some risks. RPA can handle processes and tasks that are repetitive and deterministic. However, these should require little to no judgment and have few exceptions. Technical glitches, exceptions, changing user interfaces, or changing contextual factors provide problems for software robots. There are also obvious security risks, and the lack of communication may conceal important issues (e.g., recurring problems are detected too late).

## PROCESS MINING

Process mining techniques use event data to show what people, machines, and organizations are really doing. Process mining provides novel insights that can be used to identify and address performance and compliance problems [3]. Just like spreadsheets can do anything with numbers, process mining can do anything with event data, i.e., it is a generic, domain independent, technology to improve processes [4].

The application of process mining is much broader than RPA. However, let us first relate both using Figure 4. The diagram sketches the typical Pareto distribution found in event logs. Often, a small percentage of activities account for most of the events and a small percentage of traces variants account for most of the traces [5]. For example, 20% of the activities may account for 80% of the events. Similarly, the 20% most frequent process variants may explain 80% of the cases. Traditional process automation focuses on the most frequent activities and process variants. Only for high-frequent activities and process variants, it may be cost-effective to automate tasks and introduce WFM/BPM. Less frequent activities and process variants need to be handled by workers that exploit human flexibility and creativity. RPA focuses on the middle part, i.e., routine work that is not frequent enough to be automated in the traditional sense. Process mining is a key technology to identify routine work that can be supported using RPA.

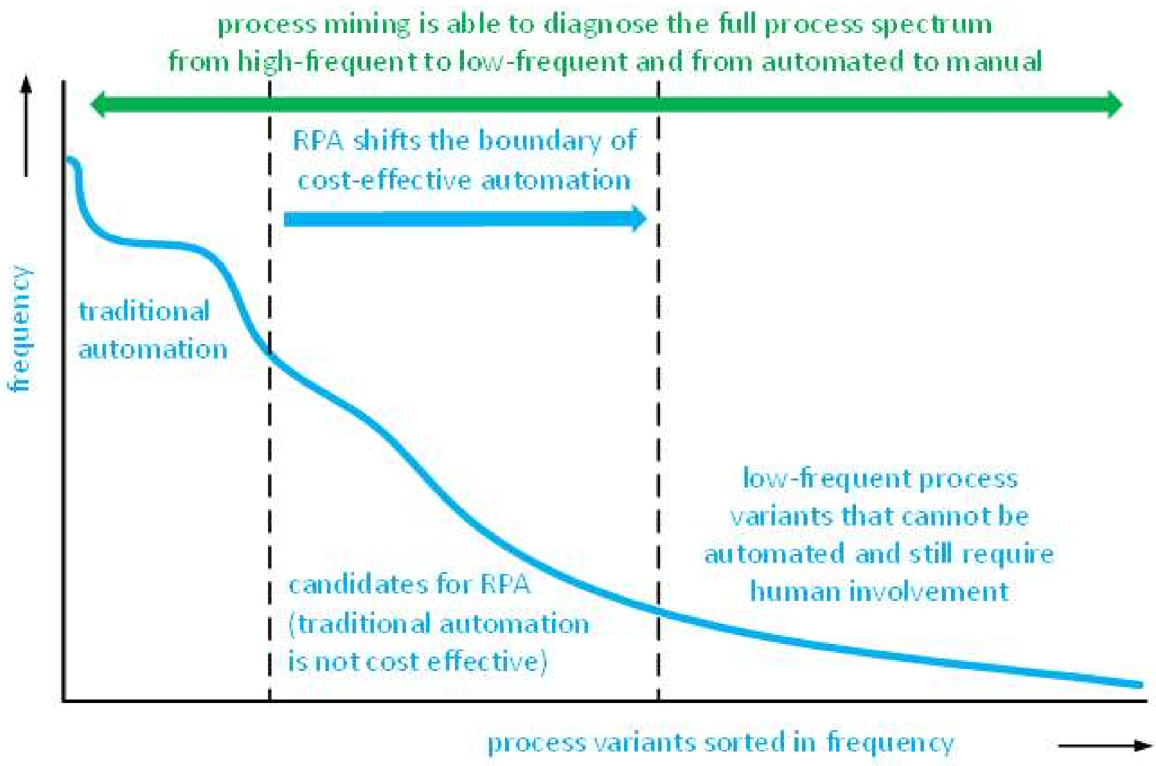



*Figure 4: Relating RPA and process mining (based on [5]).*

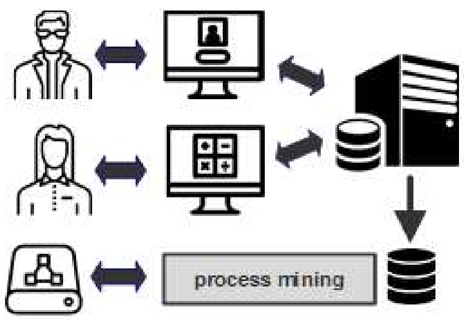
Process mining starts from event data, typically stored in an event log. An event log views a process from a particular angle. Each event in the log refers to (1) a particular process instance (called a case), (2) an activity, and (3) a timestamp. There may be additional event attributes referring to resources, people, costs, etc., but these are optional. With some effort, such data can be extracted from the information systems used by the organization. For example, an SAP system may hold thousands of tables with information about hundreds of processes. Process mining uses these event data to answer a variety of process-related questions. Process mining techniques such as process discovery, conformance checking, model enhancement, and operational support can be used to improve performance and compliance [3].

Currently, there are over 30 commercial offerings of process mining software (e.g., Celonis, Disco, ProcessGold, myInvenio, PAFnow, Minit, QPR, Mehrwerk, Puzzledata, LanaLabs, StereoLogic, Everflow, TimelinePI, Signavio, and Logpickr). They all can discover so-called Directly-Follows Graphs (DFGs) showing frequencies and bottlenecks. DFGs can be seamlessly simplified by removing nodes and edges based on frequency thresholds. DFGs are simple and provide interesting insights, but only provide a starting point. More advanced discovery algorithms like the inductive miner discover better process models, also showing concurrency (e.g., Petri nets, BPMN diagrams, and UML activity diagrams) [3]. Typically, four types of process mining are identified [3].

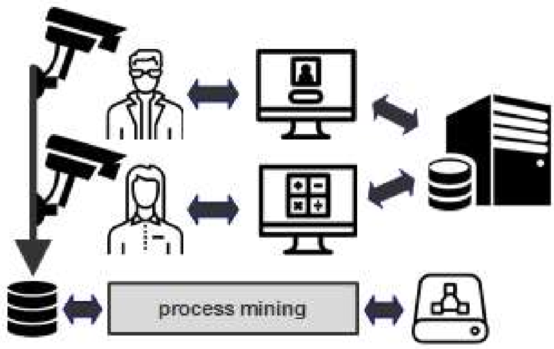
- Process discovery: learning process models from event data. A discovery technique takes an event log and produces a process model without using additional information. An example is the well-known Alpha-algorithm, which takes an event log and produces a Petri net explaining the behavior recorded in the log. Most of the commercial process mining tools first discover DFGs before conducting further analysis.
- Conformance checking: detecting and diagnosing both differences and commonalities between an event log and a process model. Conformance checking can be used to check if reality, as recorded in the log, conforms to the model and vice versa. The process model used as input may be descriptive or normative. Moreover, the process model may have been made by hand or learned using process discovery.
- Process reengineering: improving or extending the model based on event data. Like for conformance checking, both an event log and a process model are used as input. However, now, the goal is not to diagnose differences. The goal is to change the process model. For example, it is possible to repair the model to better reflect reality. It is also possible to enrich an existing process model with additional perspectives. For example, replay techniques can be used to show bottlenecks or resource usage. Process reengineering yields updated models. These models can be used to improve the actual processes.
- Operational support: directly influencing the process by providing warnings, predictions, or recommendations. Conformance checking can be done 'on-the-fly' allowing people to act the moment things deviate. Based on the model and event data related to running process instances, one can predict the remaining flow time, the likelihood of meeting the legal deadline, the associated costs, the probability that a case will be rejected, etc. The process is not improved by changing the model, but by directly providing data-driven support in the form of warnings, predictions, and/or recommendations.

All techniques start from the so-called control-flow perspective, which focuses on the ordering of activities. Then the time perspective (bottlenecks, delays, and frequencies), the data perspective (understanding decisions), and the resource and organization perspective (social networks, roles, and authorizations) are added.

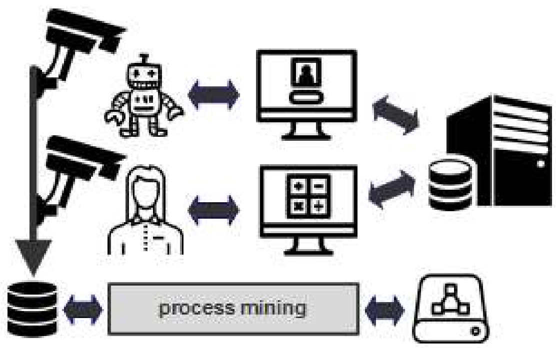
As mentioned, the scope of process mining extends far beyond RPA. However, to conclude the paper, we discuss the relationship between process mining and RPA in more detail using Figure 5. In Figure 5(a), the traditional usage of process mining is described. In this scenario, event data are extracted from the information systems supporting the process. Workers are not observed directly. In Figure 5(b), process mining is applied to event data collected directly from the GUI, i.e., the interactions between workers and information systems are directly recorded. This scenario is particularly useful in the phase before RPA is introduced. Process mining can be used to detect routine work that can be automated by mimicking the behavior of workers. Rather than manually programming robots, process discovery can be used to configure the robots correctly. In Figure 5(c), process mining is used after introducing RPA. Part of the work formerly done by workers is now done by software robots. In this scenario, process mining is used to check whether the processes run as planned. If a software robot malfunctions due to technical glitches, exceptions, changing user interfaces, or changing contextual factors, then this can be detected using conformance checking techniques. Note that a lack of human oversight of the work produced by robots constitutes a real risk of catastrophic outcomes. Figure 5(d) describes the most advanced scenario. In this scenario, the work is flexibly distributed over workers and software robots. For example, tasks are initially performed by robots and are escalated to workers the moment there is a complication of exception. Similarly, workers can hand off work to robots using an "auto-complete" option. Moreover, the RPA solution may adapt due to changes in the underlying process (e.g., concept drift).



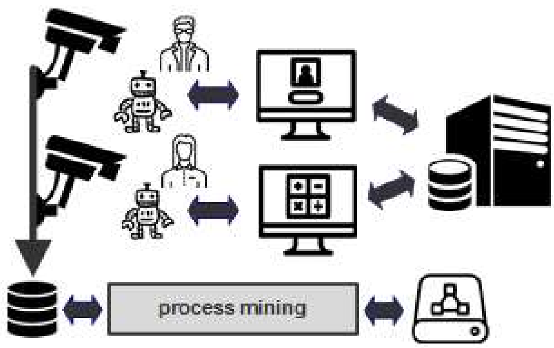

(a) Traditional process mining using event data extracted from the information systems supporting the process.




(b) Process mining using event data obtained by observing the workers using the information systems.




(c) Process mining after introducing RPA. Event data are obtained by observing workers and robots.




(d) Process mining after introducing RPA using an adaptive distribution of work over workers and robots.

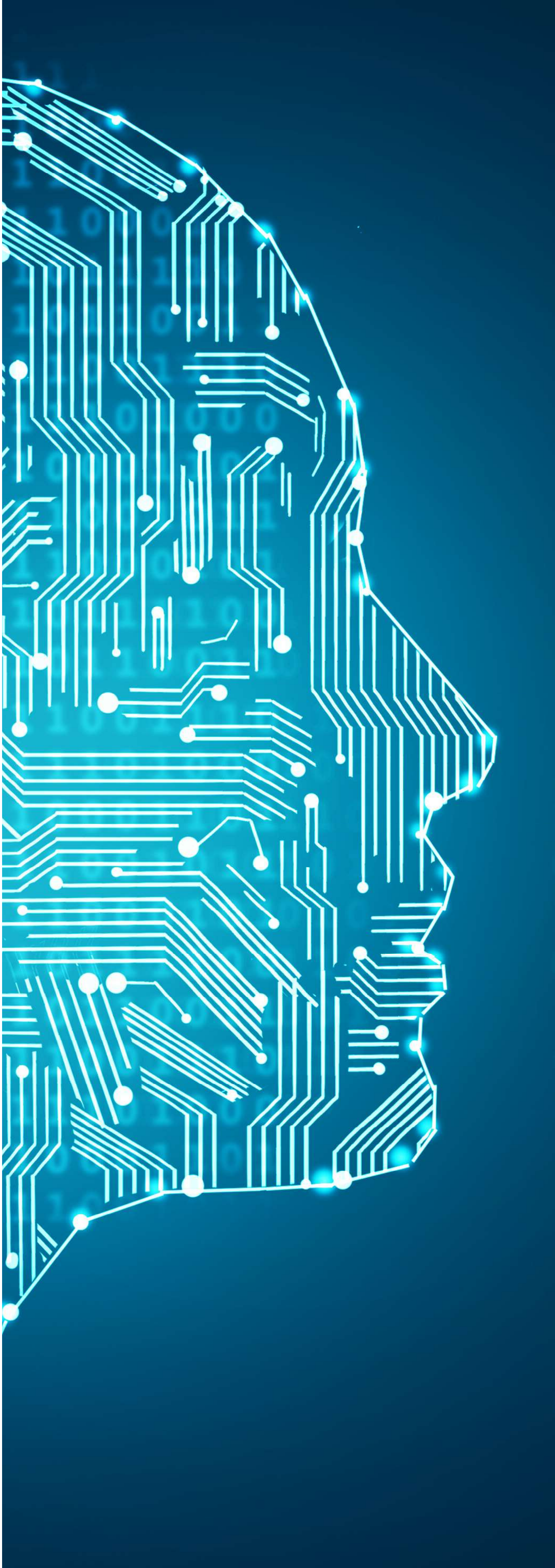
*Figure 5: Process mining can be used before and after the introduction of RPA. Robots and workers use the same GUI and the role-distribution may be flexible and change over time. Fortunately, process mining provides a holistic view of the processes at hand and interplay between robots and workers.*

# CONCLUSION

Process automation has a long history. WFM and BPM systems have been around for decades, but their application is limited to high-volume structured processes. RPA has lowered the threshold for automation. The phrase "RPA is the Poor Man's WFM" (coined in this paper) illustrates this. Due to RPA, it is possible to automate many mundane repetitive routines in an economically viable manner. Process mining helps to identify process fragments that can be supported using RPA. This is the reason that process mining and RPA vendors have joined forces. For example, in October 2019, process mining vendor ProcessGold was acquired by RPA vendor UiPath. Similarly, vendors like Celonis started to support "task mining" and "action automation" (using the action engine) to boost RPA-related capabilities.

According to Deloitte and EY, up to 30 to 50% of RPA projects fail, and most are more expensive and time-consuming than planned [6,7]. Process mining can be used to avoid such failures. As Figure 4 shows, the scope of process mining includes everything from routine activities and processes automated using WFM, BPM, and RPA to one-of-a-kind activities and processes that require human interventions and creativity. Moreover, process mining helps to support the different phases of RPA as highlighted in Figure 5.

The uptake of RPA triggers many interesting research questions [5]. What characteristics make processes suitable to be supported by RPA? How to control software robots and avoid security, compliance, and economic risks? How can software robots and people seamlessly work together? Process mining plays a key role in answering these questions. **In short: Process mining can be used to pick your automation battles!**



# REFERENCES


1. W.M.P. van der Aalst. The Application of Petri Nets to Workflow Management. The Journal of Circuits, Systems and Computers, 8(1):21-66, 1998. https://doi.org/10.1142/S0218126698000043
2. W.M.P. van der Aalst. Business Process Management: A Comprehensive Survey. ISRN Software Engineering, pages 1-37, 2013. https://doi.org/10.1155/2013/507984
3. W.M.P. van der Aalst. Process Mining: Data Science in Action. Springer-Verlag, Berlin, 2016. http://www.springer.com/978-3-662-49850-7
4. W.M.P. van der Aalst. Spreadsheets for Business Process Management: Using Process Mining To Deal With "Events" Rather Than "Numbers". Business Process Management Journal, 24(1):105-127, 2018. https://doi.org/10.1108/BPMJ-10-2016-0190
5. W.M.P. van der Aalst, M. Bichler, and A. Heinzl. Robotic Process Automation. Business and Information Systems Engineering, 60(4):269-272, 2018. https://doi.org/10.1007/s12599-018-0542-4
6. EY. Get Ready for robots: Why planning makes the difference between success and disappointment. 2016. https://www.ey.com/Publication/vwLUAssets/Get_ready_for_robots/$FILE/ey-get-ready-for-robots.pdf
7. Deloitte. The Robots Are Ready. Are You? Untapped advantage in your digital workforce. 2017. https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/consultancy/deloitte-uk-consulting-robots-are-ready.pdf


# ACKNOWLEDGMENTS:


We thank the Alexander von Humboldt (AvH) Stiftung for supporting our research.