New Data Sources for Process Mining

Dissertation
zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaft
der Rechts- und Wirtschaftswissenschaftlichen Fakultät
der Universität Bayreuth

vorgelegt
von
Andreas Egger
aus
Illertissen

Dekan:	Prof. Dr. Claas Christian Germelmann
Erstberichterstatter:	Prof. Dr. Maximilian Röglinger
Zweitberichterstatterin:	Prof. Dr. Moe Thandar Wynn
Tag der mündlichen Prüfung:	24.10.2024

If you can't describe what you are doing as a process, you don't know what you're doing.

William Edwards Deming (1900 – 1993)

Diese Dissertation widme ich allen, die mich auf meinem Weg begleitet, unterstützt und geprägt haben!

Ich möchte meiner gesamten Familie danken, insbesondere meinen Eltern Armin und Brigitte. Ihr habt mich zu dem Menschen gemacht, der ich heute bin und habt mich bei all meinen Vorhaben immer unterstützt!
Danke an alle Kolleginnen und Kollegen, von denen viele zu Freunden fürs Leben geworden sind. Ebenso danke ich allen Freunden aus der Heimat, die mich treu begleitet haben. Eine solch langjährige, große und beständige Freundesgruppe ist etwas Besonderes.
Ich danke auch Dir, Max, für Deine Vorbildfunktion als Doktorvater, Dein unermüdliches Engagement und die positive persönliche Entwicklung, die ich mit Dir erfahren durfte.
Liebe Amelie, vielen Dank für Deine bedingungslose Liebe und dafür, dass Du auch in anstrengenden Zeiten immer an mich glaubst! Deine Unterstützung und Dein Verständnis haben diese Promotion erst möglich gemacht. Ich bin unendlich dankbar, dass Du an meiner Seite bist und ohne Dich wäre ich nicht so weit gekommen.
Lieber Jakob, auch wenn die Abgabe dieser Dissertation einen großen Meilenstein in meinem Leben darstellt, wird alles überschattet durch das Wunder Deiner Geburt. Du hast mir so viel Motivation und eine neue Perspektive aufs Leben gegeben.

Copyright Statement

The following sections are partly comprised of content from the research articles included in this thesis. To improve the readability of the text, I have omitted the standard labeling of these citations.

Abstract

Process mining is situated between process science and data science, with the objective of discovering, monitoring, and improving processes. As a technique in business process management, process mining uses event logs as input and enables building a view of reality by making process executions tangible, facilitating the identification of bottlenecks, providing new insights, and anticipating problems in business processes. The process mining discipline has undergone a notable evolution over recent decades, both in academic research and in practice. The development of innovative algorithms and algorithmic extensions is an ongoing process, with previously unused data sources being explored to enhance the existing body of knowledge. Traditionally, the predominant data sources for process mining are information systems, such as enterprise resource planning systems. However, process information can also be extracted from alternative sources and utilized for process mining. The majority of available data is unstructured, thereby offering a substantial opportunity to provide valuable contextual insights into business processes. This can facilitate a more comprehensive representation and analysis of real-world processes, as well as the reduction of blind spots, i.e., parts of processes that were previously unable to be captured in event logs.

The overarching objective of this dissertation is to contribute to the advancement of process mining by enabling the use of new data sources. This objective is pursued by facilitating the integration of sensor, video, bot, and text data, as well as providing a systematic overview of approaches to unstructured data in process mining. In line with design science research principles, multiple artifacts are developed that contribute to process mining in both research and practice. This dissertation comprises five research articles addressing three opportunities that aim to expand the scope of process mining analysis.

First, the integration of unstructured data into process mining is addressed by initial research. However, a systematic overview that presents a comprehensive summary of the approaches employed has yet to be provided. Research Article 1 addresses this opportunity by providing a systematic literature review of the current state of research on the use of unstructured data in process mining. In light of the findings, a research agenda is put forth that identifies both open challenges and potential avenues for future research.

Second, data derived from video and sensor data could facilitate the detection of previously hidden but relevant process activities, thus enabling a more transparent process picture. Accordingly, Research Article 2 presents a reference architecture that offers guidance on the utilization of unstructured data sources and traditional event logs for

object-centric process mining. Moreover, an instantiation of the proposed architecture is provided, demonstrating the specific use of video and sensor data for object-centric process mining. Additionally, Research Article 3 proposes a reference architecture for the unsupervised exploration of video data. This architecture enables the extraction of actual process activities from video data without the need for predefined activities, serving as a starting point for process discovery.

Third, the conjunction of several emerging technologies with process mining has facilitated the integration of additional data sources. As the use of Robotic Process Automation (RPA) bots becomes more prevalent in business processes, there is a need to integrate the steps performed by bots into process mining analysis. Accordingly, Research Article 4 presents an approach that makes bot logs from RPA software usable for process mining and develops process mining measures that analyze bot logs and process event logs in an integrated manner. Chatbots represent a further type of bot that is deployed in scenarios where it is essential to align with the underlying business processes and comply with regulatory requirements. Consequently, there is a need to integrate textual conversation data from chatbots to investigate whether chatbots achieve process-compliant behavior. Research Article 5 addresses this opportunity by providing an approach that converts textual conversation data from chatbots to event logs for process mining and quantifies chatbots' ability to learn and adhere to organizations' business processes.

Overall, this dissertation, comprising the five research articles, contributes to the advancement of process mining by enabling the use of new data sources. The conducted literature review provides new insights to systematically advance research at the intersection of unstructured data and process mining. The dissertation presents two artifacts that provide guidance for incorporating video and sensor data as new data sources for process mining, completed by a generic architecture on the use of unstructured data for object-centric process mining. Finally, two artifacts that integrate data from emerging technologies are presented. Both artifacts facilitate a more holistic process view by incorporating data from chatbots and RPA bots as new data sources for process mining.

Table of Contents

List of Figures

List of Tables

Acronyms

BPM Business Process Management

CRM Customer Relationship Management

DO Design Objective

DSR Design Science Research

ERP Enterprise Resource Planning

NLI Natural Language Inference

NLP Natural Language Processing

OCRAUD Object-Centric Reference Architecture for Unstructured Data in process mining

ORM 2 Object-Role Modeling 2

OVaSA Object-centric Video and Sensor Architecture

RAVEE Reference Architecture for Video Event Extraction

RPA Robotic Process Automation

Seq2Seq Sequence-to-Sequence

UML Unified Modeling Language

I Introduction

I.1 Motivation

Process mining is situated between process science and data science, with the objective of discovering, monitoring, and improving processes (van der Aalst et al., 2012; van der Aalst, 2016a). As a technique in Business Process Management (BPM), process mining uses event logs as input and enables building a view of reality by making process executions tangible, facilitating the identification of bottlenecks, providing new insights, and anticipating problems in business processes (Dumas et al., 2018; van der Aalst et al., 2012). The process mining discipline has undergone a notable evolution over the past few decades. The discipline's foundational principles were established in the late 1990s with the objective of utilizing example traces to automatically construct a Petri net (van der Aalst, 2022). Nevertheless, the discipline has seen a significant expansion over time, encompassing a much broader range of techniques and applications. These include organizational perspectives on the adoption and effective use of the technology (Martin et al., 2021; Mamudu et al., 2022; Milani et al., 2022), predicting process executions (Kratsch et al., 2021; Di Francescomarino and Ghidini, 2022), and recommending appropriate next best actions (Leoni et al., 2020).

In practice, process mining has facilitated process improvements and cost savings. For example, it has been used to reduce over 10 million manual activities associated with order-to-cash processes at Siemens and to enhance transparency in manufacturing processes at the BMW Group (Reinkemeyer, 2020b). The economic potential is also reflected in the dynamic development of the market for process mining software. Gartner currently monitors 39 process mining platforms, including IBM, UiPath, and the market leader Celonis (Gartner, Inc., 2024). It is projected that the global process mining market will experience a substantial growth trajectory, increasing from $2.46 billion in 2024 to $46.39 billion by 2032 (Fortune Business Insights, 2024). This represents a compound annual growth rate of 44.3%.

Notwithstanding these promising prospects in practice, new and established process mining techniques indicate areas for enhancement (van der Aalst, 2020). These are continuously being addressed through the development of innovative algorithms (Boltenhagen et al., 2021; Li et al., 2022) and algorithmic extensions (Li and Zelst, 2022), as demonstrated by ongoing research in the field. For example, object-centric process mining (van der Aalst, 2019; Esser and Fahland, 2021; Berti and van der Aalst, 2023) has been introduced, offering a revised notation for the extraction of more realistic event logs, which represents one of several promising avenues for the enhancement of process mining.

An ongoing challenge in the field of process mining is the acquisition of the required high-quality data from various potential sources (van der Aalst, 2016b). Traditionally, the predominant data sources for process mining are information systems, such as Customer Relationship Management (CRM) systems or Enterprise Resource Planning (ERP) systems (Diba et al., 2020; van der Aalst, 2016b). These systems provide data that is either already captured in a process-centric format or can be made process mining-ready through preprocessing. Furthermore, process information can be extracted from alternative sources and utilized for process mining (Reinkemeyer, 2020a; van der Aalst, 2016b).

The majority of available data is unstructured (Gandomi and Haider, 2015), and it is anticipated that this proportion will continue to grow (Balducci and Marinova, 2018). Nevertheless, despite the availability of these sources, they are often not used for process mining analysis, and current studies mainly focus on traditional structured data sources (Diba et al., 2020; Janiesch et al., 2020). Using unstructured data sources in process mining, however, presents a significant opportunity to provide valuable contextual insights into business processes (Beverungen et al., 2021; Koschmider et al., 2023). Furthermore, unstructured data can assist in more comprehensive representation and analysis of real-world processes (Grisold et al., 2021), as well as in the reduction of blind spots, i.e., parts of processes that previously were not captured in event logs (Kratsch et al., 2022). It is, therefore, unsurprising that in a recent Delphi study, all academic experts voted that the exploration of non-process-related and unstructured data should be designated as a priority within the field of BPM (Kerpedzhiev et al., 2021).

As a result, preliminary research has commenced utilizing unstructured data for process mining, such as textual data (Banziger et al., 2018; Nakayama et al., 2018; Jlailaty et al., 2017), sensor data (van Eck et al., 2016; Leotta et al., 2020), or image and video data (Kratsch et al., 2022; Knoch et al., 2020). Other research has sought to enhance process mining with emerging technologies, such as Robotic Process Automation (RPA) (El-Gharib and Amyot, 2023; Leno et al., 2021), which aims to automate simple and repetitive tasks through the use of bots (van der Aalst et al., 2018; Aguirre and Rodriguez, 2017). These trends also facilitate the potential exploitation of hitherto untapped data sources for process mining. Figure 1 illustrates how traditional data sources in process mining, such as CRM, BPM, or ERP systems, can be complemented with new data sources that have not previously been utilized for process mining analysis. These include, but are not limited to, sensor, video, bot, or text data. While there are numerous untapped opportunities associated with acquiring the necessary data for process mining, this dissertation will address three primary opportunities, which are outlined in the following.

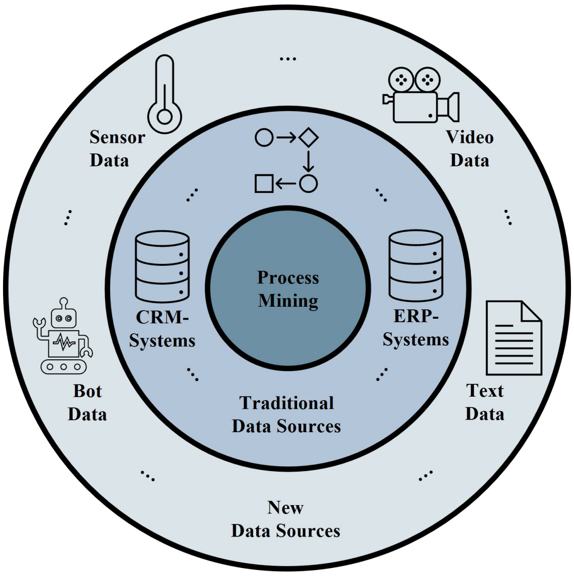

Figure 1: Overview of new data sources for process mining in this dissertation

The extant studies that have employed unstructured data in process mining have demonstrated the potential value of such data for process mining applications. These studies are frequently tailored to a specific application scenario, utilize only particular data types, and focus on specific process mining activities (e.g., Khowaja et al. (2020), Gupta et al. (2020), and Knoch et al. (2020)). Currently, there is a lack of an overview that summarizes the approaches that integrate unstructured data into process mining, which hinders systematic progress in this field. A structured review of the existing literature on unstructured data in process mining would assist researchers and practitioners in identifying research gaps and enabling the design of new studies. This opportunity could reveal, for example, the extent to which specific data types, such as sensor, video, or text data, are already utilized in the existing literature or which process mining activities, such as discovery, prediction, and recommendation, have already been covered. Furthermore, a comprehensive summary of the open challenges and potential avenues for advancing future research would provide valuable guidance for other researchers in formulating relevant research questions.

The integration of a particular data type, such as textual, sensory, or video data, into a process mining analysis is contingent upon the specific process scenario. In some cases, the inclusion of one or more of these data types may offer more substantial advantages than others. For instance, the extraction of meaningful process information from textual data may prove advantageous in contexts where a considerable volume of emails are utilized (Elleuch et al., 2020; Jlailaty et al., 2017). Nevertheless, this approach may be less effective in scenarios involving manual process activities. In particular, data derived from video and sensors could prove beneficial in such scenarios. Sensors can be installed, incorporating the gathered data into process mining analysis (Hemmer et al., 2021), or, in compliance with privacy aspects (Elkoumy et al., 2022), video cameras could be used in areas where textual data is not available, extracting relevant context information. This would enable the detection of previously hidden but relevant process activities, thus facilitating transparent process discovery. In the processing of video data, techniques for object tracking and activity recognition may be employed for the identification and tracking of specific objects and activities (Wu et al., 2013; Aggarwal and Ryoo, 2011). This offers opportunities as a potential fit for object-centric process mining, which enables a more flexible representation of process-related data using objects and events (van der Aalst, 2019). Notwithstanding the potential advantages of utilizing sensor and video data, research in the field of process mining that addresses these data types is relatively limited, mainly focusing on specific use cases (e.g., Kratsch et al. (2022), Knoch et al. (2018), and Hemmer et al. (2021)). This offers opportunities for research that generically leverages video and sensor data for process mining, enabling researchers and practitioners to systematically use these data types to enhance process understanding.

In addition to incorporating unstructured data sources, such as video or sensor data, into process mining analysis, the advent of several technologies has facilitated the integration of additional data sources. RPA is one such technology using software bots to automate processes (El-Gharib and Amyot, 2023; Leno et al., 2021). As the Gartner report shows (Gartner, Inc., 2024), the RPA vendor UiPath has entered the process mining market, and process mining vendors, such as Celonis, have incorporated RPA functionality into their software (Geyer-Klingeberg et al., 2018; Celonis, 2024b). Moreover, research on RPA and process mining are becoming increasingly integrated (Leno et al., 2021; El-Gharib and Amyot, 2023), with a primary focus on the initial stages of RPA projects. This integration presents a promising avenue for further research. As bots become more prevalent in business processes, there is a need to integrate the actions performed by bots into process mining analysis to gain a comprehensive understanding of the underlying processes. The advent of new technologies is giving rise to a new category of bots, enabled by advances in Natural Language Processing (NLP). Chatbots, exemplified by ChatGPT (OpenAI, 2022), represent one such emerging category. Chatbots can be deployed in the business-to-consumer domain, for example, in customer service (Poddar et al., 2009), and in such scenarios, they must adhere to organizational or regulatory requirements (Gunson et al., 2011). Consequently, the implementation of chatbots within an organizational context is contingent upon the underlying model's capacity to learn and align with the specific business processes. To investigate whether chatbots achieve process-compliant behavior, approaches are needed that enable the use of textual conversation data from chatbots for process mining. Incorporating steps executed by chatbots at the interface between customers and business processes (Pallotta and Delmonte, 2013) into process mining could facilitate a more holistic view of business processes, ultimately enhancing process analysis and enhancement. Accordingly, integrating chatbot data presents novel opportunities for process mining.

I.2 Research Objectives

In light of the identified opportunities and research needs, this dissertation seeks to make contributions in three key areas. First, initial research has been conducted on unstructured data in process mining. However, a systematic overview that presents a comprehensive summary of the approaches employed has yet to be provided. Such an overview would facilitate progress in the systematic exploration of this promising area. To address this opportunity, this dissertation provides a systematic literature review of the current state of research on the use of unstructured data in process mining. In light of the findings, a research agenda is put forth that identifies both open challenges and potential avenues for future research.

Second, data derived from video and sensor data could facilitate the detection of previously hidden but relevant process activities, thus enabling a more transparent process picture. The paucity of research in this area presents opportunities for further investigation, such as leveraging video and sensor data for object-centric process mining. Therefore, this dissertation presents a reference architecture that provides guidance on the utilization of unstructured data sources and traditional event logs for object-centric process mining. Specifically contributing to the integration of video and sensor data into process mining, an instantiation of the proposed architecture is provided that demonstrates the use of these two data types for object-centric process mining. As a further contribution, this dissertation proposes a reference architecture for the unsupervised exploration of video data. This architecture enables the extraction of actual process activities from video data without the need for predefined activities and serves as a starting point for process discovery. In summary, the two contributions provide a means of utilizing video and sensor data as new data sources for process mining, facilitating a more comprehensive process picture.

Third, the conjunction of several emerging technologies with process mining has facilitated the integration of additional data sources. As the use of RPA bots becomes more prevalent in business processes, there is a need to integrate the steps performed by bots into process mining analysis. Accordingly, this dissertation presents an approach that makes bot logs from RPA software usable for process mining and develops process mining measures that analyze bot logs and process event logs in an integrated manner. Chatbots represent a further type of bots that are deployed in scenarios where it is essential to align with the underlying business processes and to comply with regulatory requirements. Consequently, there is a need to integrate textual conversation data from chatbots to investigate whether chatbots achieve process-compliant behavior. This dissertation addresses this opportunity by providing an approach that converts textual conversation data from chatbots to event logs for process mining and quantifies chatbots' ability to learn and adhere to organizations' business processes. In conclusion, the two contributions leverage the use of bot and text data as new data sources for process mining.

The overarching objective of this dissertation is to contribute to the advancement of process mining by enabling the use of new data sources. This objective is pursued by facilitating the integration of sensor, video, bot, and text data, as well as providing a systematic overview of approaches to unstructured data in process mining. Accordingly, multiple artifacts are developed in line with the Design Science Research (DSR) principles as set out by Gregor and Hevner (2013) and Peffers et al. (2007). In sum, the findings of this dissertation contribute to the field of process mining in both research and practice.

I.3 Structure of the Thesis and Embedding of the Research Articles

This dissertation is comprised of five research articles that address the identified opportunities and research objectives. Figure 2 illustrates the focus of the individual contributions. Research Article 1 provides an overview of current research on unstructured data in process mining. Research Article 2 facilitates the use of sensor and video data for object-centric process mining, whereas Research Article 3 concentrates on the unsupervised exploration of video data. Research Articles 4 and 5 present two contributions to the incorporation of data from emerging technologies into process mining. While Research Article 4 integrates data from RPA bots, Research Article 5 enables the use of textual conversation data from chatbots for process mining.

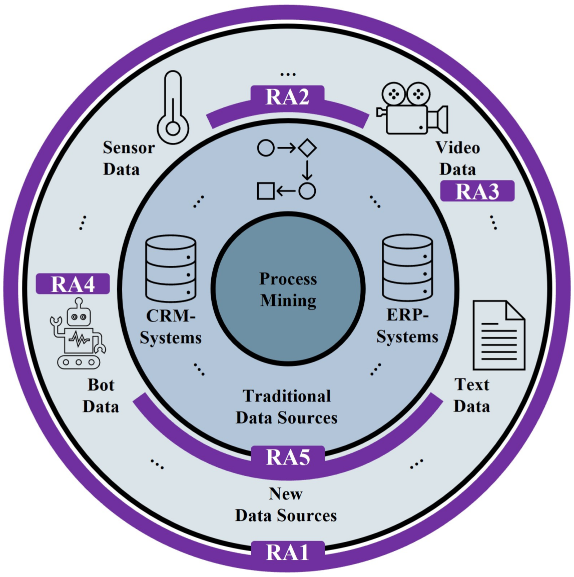

Figure 2: Focus and embedding of the five Research Articles (RA1 – RA5) in this dissertation

The structure of this dissertation and the embedding of the five research articles are presented in Table 1. Section I motivated the research and defined the research objectives of this dissertation. Section II (including Research Article 1) presents an overview of unstructured data in process mining and identifies research opportunities. Thereby, Research Article 1 provides a systematic literature review with a focus on technical process mining artifacts that enable (semi-)automated consideration of unstructured data. Out of 1,379 identified research items, 24 primary studies were selected and analyzed. Based on the results, a research agenda is presented that summarizes challenges and opportunities for future research.

Section III (including Research Articles 2 and 3) is concerned with the integration of video and sensor data as new data sources for process mining. Research Article 2 proposes a reference architecture that provides guidance on the integrated use of unstructured data sources and traditional event logs for object-centric process mining. To demonstrate the application of the reference architecture, a particular instantiation is presented that is specifically tailored to the utilization of video and sensor data. Multiple evaluation episodes were conducted to validate the research, including a competing artifacts analysis, two rounds of expert interviews, the implementation of a software prototype with a graphical user interface, and the use of the prototype on a real-world dataset. Research Article 3 introduces a reference architecture that facilitates the extraction of actual process information from video data through the integration of computer vision and clustering capabilities. A multitude of evaluation activities have been conducted, including a prototypical instantiation of the reference architecture tested against a real-world dataset, a competing artifacts analysis, and interviews with experts from industry and research. The two articles provide guidance for incorporating video and sensor data as new data sources for process mining, completed by a generic architecture on the use of unstructured data for object-centric process mining.

Table 1: Structure of this dissertation and embedding of the five research articles (RA1 – RA5)

I	**Introduction**
II	**Overview and Opportunities of Unstructured Data in Process Mining**
RA1	**Unstructured Data in Process Mining: A Systematic Literature Review** König F, Egger A, Kratsch W, Röglinger M, Wördehoff N
III	**Video and Sensor Data**
RA2	**Refining the Process Picture: Unstructured Data in Object-Centric Process Mining** Egger A, Fehrer T, Kratsch W, Wördehoff N, König F, Röglinger M
RA3	**Beyond Assumptions: A Reference Architecture to Enable Unsupervised Process Discovery from Video Data** Wördehoff N, Egger A, Kratsch W, König F, Röglinger M
IV	**Bot and Text Data**
RA4	**Bot Log Mining: An Approach to the Integrated Analysis of Robotic Process Automation and Process Mining** Egger A, ter Hofstede AHM, Kratsch W, Leemans SJJ, Röglinger M, Wynn MT
RA5	**Quantifying Chatbots' Ability to Learn Business Processes** Kecht C, Egger A, Kratsch W, Röglinger M
V	**Conclusion**
VI	**References**
VII	**Appendix**

Section IV (including Research Articles 4 and 5) recognizes the importance of integrating emerging technologies into process mining, thus enabling the inclusion of bot and text data as new data sources for process mining. Research Article 4 proposes an approach that integrates RPA with process mining. Thereby, a conceptual data model is developed that describes the relations between bots and business processes. Moreover, an approach is presented that makes bot logs usable for process mining, and 12 integrated process mining measures are developed that consider bots as integral components of business processes. Various evaluation activities were conducted, including two rounds of interviews with experts and the implementation and testing of a software prototype with both real-world and artificial data. Research Article 5 introduces an approach that derives topics and process activities from textual customer service conversations with the goal of quantifying chatbots' ability to learn and adhere to organizations' business processes. To evaluate the approach, several evaluation activities were conducted, including a competing artifact analysis, the development of a prototype, and an assessment of its applicability to real-world data.

Section V summarizes the dissertation, provides limitations, and suggests avenues for future research. Section VI comprises the list of references. Section VII includes an index of the research articles, lists my contributions to the research articles, and provides the abstracts of the research articles.

II Overview and Opportunities of Unstructured Data in Process Mining

As outlined in Section I, integrating unstructured data into process mining holds significant potential to generate valuable context-related insights for BPM (Beverungen et al., 2021; Koschmider et al., 2023) and can help to analyze real-world processes more comprehensively (Grisold et al., 2021). The benefits of unstructured data for specific process mining use cases are demonstrated by initial research on text data (Teinemaa et al., 2016), sensor data (Leotta et al., 2020), and image and video data (Knoch et al., 2018; Kratsch et al., 2022; Lepsien et al., 2023). As these studies are often tailored to a specific application scenario, generic approaches to integrate unstructured data into process mining are still needed. However, the lack of an overview of the usage of unstructured data in process mining hinders systematic progress in this research area. Against this backdrop, Research Article 1 seeks to ascertain the extent to which unstructured data are incorporated into process mining, to structure the current literature, and to present a research agenda. Therefore, the following research questions are addressed:

RQ1 *Which types of unstructured data are used in process mining?*

RQ2 *How are unstructured data leveraged for different process mining use cases?*

RQ3 *What are the open challenges and areas for improvement?*

To answer these research questions, in Research Article 1, a systematic literature review (Kitchenham and Charters, 2007) was conducted. Thereby, the focus lies on technical process mining artifacts that enable (semi-)automated consideration of unstructured data. From 1,379 research items, 24 primary studies have been selected and reviewed. In light of these findings, a research agenda is proposed that highlights open challenges and potential opportunities for future research.

Following the steps recommended by Kitchenham and Charters (2007), first, a review protocol was developed for the literature review. In order to ensure the identification of as many potentially relevant research items as possible (Hiebl, 2023), a purely database-driven approach was selected, encompassing seven prominent databases within the domains of information systems and computer science. Additionally, in an iterative approach, the final search phrase was developed, encompassing combinations and different spellings of the terms "process mining", "unstructured data", "heterogeneous data", as well as the term "information" instead of "data". The initial search resulted in 1,379 research items. In the next step, multiple study selection criteria were defined, which two co-authors applied to identify relevant research items. Following a deductive data extraction approach (Bandara et al., 2015) and using a developed coding scheme (Kitchenham and Charters, 2007), the final 24 studies were analyzed by the same two co-authors. Table 2 presents for each of the final studies the respective goal and the types of unstructured data used.

Table 2: Summary of the 24 research items included in the final sample. Note: unstructured data can be text (T), audio (A), image (I), video (V), simple sensor (SiS), and complex sensor (CoS)

No.	**Paper**	**Goal**	**Type of data**					
			T	**A**	**I**	**V**	**SiS**	**CoS**
1	Li et al. (2015)	Improve task identification and event log extraction by leveraging text documents	x					
2	Banziger et al. (2018)	Discover processes from CRM data by incorporating related text data	x					
3	Weerdt et al. (2012)	Discover incident management processes by combining text data with trace clustering	x					
4	Gupta et al. (2020)	Discover incident management processes through the extraction of key phrases from textual attributes to enrich event logs	x					
5	Zhu et al. (2019)	Identify activities in software development logs	x					
6	Jlailaty et al. (2017)	Identify process activities in emails	x					
7	Elleuch et al. (2020)	Identify multiple process activities in emails	x					
8	Chambers et al. (2020)	Discover processes from textual data	x					
9	Holstrup et al. (2020)	Discover knowledge-sharing processes based on text-based conversations	x					
10	Tang et al. (2021)	Identify and extract activities from text data to uncover bottlenecks	x					
11	Kecht et al. (2021)	Discover processes based on text conversations	x					

Table 2: Summary of the 24 Research Items Included in the Final Sample. Note: unstructured data can be text (T), audio (A), image (I), video (V), simple sensor (SiS), and complex sensor (CoS)

No.	Paper	Goal	Type of data					
			T	A	I	V	SiS	CoS
12	Holz et al. (2021)	Discover processes from case worker reports in the personal services domain	x					
13	Epure et al. (2015)	Discover processes from text reports	x					
14	Chiudinelli et al. (2020)	Discover patient care flows by extracting frequent event patterns for breast cancer patients	x					
15	Kecht et al. (2023)	Quantify the ability of chatbots to learn and comply with rules in business processes	x					
16	Folino et al. (2015)	Predict the fix-time for incident tickets in incident management systems	x					
17	Teinemaa et al. (2016)	Leverage text data and structured data for predictive process mining	x					
18	Ronzani et al. (2022)	Predict the success of home hospitalization by mapping textual diagnoses to ICD-9-CM codes	x					
19	Jimenez-Ramirez et al. (2019)	Discover process models by leveraging event extraction on screen-mouse-key-logging to support the initial phases of RPA projects			x		x	
20	Kratsch et al. (2022)	Extract structured process information from unstructured video data				x		
21	Knoch et al. (2018)	Extract worker activities in manufacturing scenarios				x		x
22	Khowaja et al. (2020)	Discover high-level activities from low-level sensor measurements building on behavior patterns of other users						x
23	Leotta et al. (2020)	Discover processes by transforming sensor measurements from smart environments into higher-level activities					x	
24	Hemmer et al. (2021)	Detect misbehaviors in the context of the internet of things					x	x
			18	0	1	2	3	3

Regarding the publication timeline of relevant research at the intersection of unstructured data and process mining, until 2013, publications increased slowly but steadily. Nevertheless, from 2013 to 2022, the annual research output increased by more than fourfold, indicating a more rapid growth than that observed in the field of process mining research in general (van der Aalst, 2020). This is possibly due to the increasing maturity and adoption of process mining research (see, e.g., the process mining manifesto (van der Aalst et al., 2012)) as well as the enhanced processing capabilities for unstructured data, made possible by advances in data science and artificial intelligence.

As Table 2 shows and addressing RQ1, 75% of the 24 studies deal with unstructured text data, three studies with data from simple sensors (such as passive infrared sensors, temperature sensors, and mouse and keyboard), three with complex sensors (such as engine rotation sensors, ultrasonic sensors, and gyroscope sensors), two with video data, and only one with image data. The literature search revealed no research items with a focus on audio data, suggesting untapped potential. Furthermore, no other types of unstructured data emerged that have not already been considered in the literature. While half of the studies exclusively process unstructured data, the other half presents artifacts that combine structured and unstructured data. Two studies combine different unstructured data types, while the other 22 studies have a unidimensional focus on one type of unstructured data.

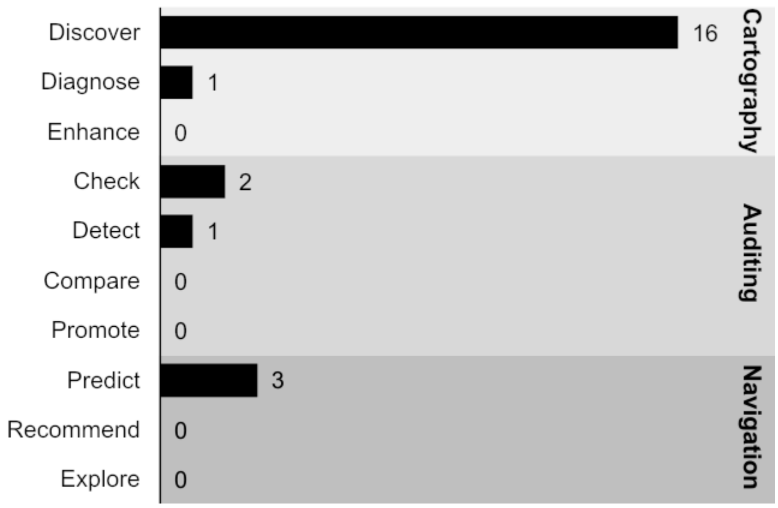

Figure 3: Overview of process mining activities. Note: none or multiple activities can apply per study

To address RQ2, the 24 selected studies were categorized along the ten process mining activities of the refined process mining framework by van der Aalst (2016d). Figure 3 shows the distribution of the ten process mining activities. A majority of 17 studies are associated with the cartography category, with process discovery as the leading activity (67%).

The number of studies in the auditing and navigation categories is equal, with three studies (13%) in each category. One study from the auditing category and three studies from the navigation category provide forward-looking operational support for process mining. The exploration of running process instances for human stakeholders and the recommendation of the next best actions are not mentioned in any study, highlighting the purely technical nature of the artifacts presented. Moreover, several studies at the intersection of unstructured data and process mining do not target a specific process mining activity but instead aim to identify relevant process activities and extract appropriate event logs for subsequent analysis.

To answer RQ3, a research agenda is proposed based on the previous results, featuring seven opportunities that help advance research at the intersection of process mining and unstructured data. As discussed, most identified studies are associated with process discovery and deal with unstructured text data. Therefore, Opportunity 1 (*Extending the scope beyond event log extraction for process discovery*) and Opportunity 2 (*Extending the scope beyond text data*) call for research that goes beyond these scopes. Since most studies focus on a single type of unstructured data in an isolated manner, Opportunity 3 (*Combining structured and different types of unstructured data*) expresses the need to devise approaches that extract relevant contextual attributes from multiple unstructured data sources and combine them with structured data. Opportunity 4 (*Consolidating existing research to build integrative artifacts*) highlights that generic artifacts that provide holistic solutions to a specific process mining challenge independent of a particular use case or application scenario are needed (e.g., reference architectures). Analyzing the identified studies in the literature review showed that there is a lack of publicly available process-oriented benchmark datasets. This is one reason why progress at the intersection of unstructured data and process mining cannot be reliably determined. Thus, Opportunity 5 (*Providing process-oriented and open-access benchmark datasets*) calls for providing such datasets to increase the transparency of research results and facilitate the extension of previous research. The analysis also revealed a research gap for process mining activities that typically require human judgment (e.g., enhance, promote, and explore). This need is expressed in Opportunity 6 (*Accounting for the human in the loop*). Since unstructured data has multifaceted meanings, it may also contain personal information that is protected by regulations. Therefore, Opportunity 7 (*Considering privacy aspects*) emphasizes the need to address privacy concerns to enable the integration of novel solutions into everyday business applications.

In summary, Research Article 1 presents a systematic literature review in which 24 primary studies were selected and analyzed from a total of 1,379 research items addressing the intersection of unstructured data and process mining. One of the main findings is that current research predominantly deals with textual data and concentrates on extracting event logs for process discovery. The study thus contributes to existing process mining knowledge and the use of unstructured data in process mining. Furthermore, Research Article 1 proposes a research agenda that includes seven opportunities to address the research gaps identified in the analysis of the selected studies. This paves the way for future process mining research incorporating new data sources.

III Video and Sensor Data

As outlined in Section II there are several research opportunities when dealing with unstructured data in process mining, such as extending the scope beyond text data and accounting for the human in the loop. The literature review revealed that the identified studies focus primarily on text data and only a few on video, image, and sensor data. Seizing these opportunities and following the overall goal of this dissertation, i.e., enabling the use of new data sources for process mining, the following two studies integrate video and sensor data into process mining. Research Article 2 (Section III.1) presents a reference architecture that gives guidance in using unstructured data sources and traditional event logs for object-centric process mining. Thereby, an instantiation of the architecture is provided that enables the use of video and sensor data for object-centric process mining. Research Article 3 (Section III.2) proposes a reference architecture that enables the extraction of actual process activities from video data without having predefined these activities. This provides an initial basis for unsupervised exploration of video data and serves as a starting point for process discovery.

III.1 Video and Sensor Data in Object-Centric Process Mining

Traditional process mining techniques are based on flat event data, where each event is associated with one case (i.e., with one process instance) (van der Aalst et al., 2012; van der Aalst, 2016d). However, in real-life processes, objects interact with one another, which can be interpreted as events involving different objects that depend on each other. To address this, a recent development in process mining research has introduced object-centric process mining (van der Aalst, 2019; Esser and Fahland, 2021), which allows events to be related to one or many objects, enabling a more comprehensive view of processes. Object-centric process mining and the use of unstructured data in process mining both aim to provide a more holistic picture of processes, and a highly structured yet restricted view of objects and their process relations can be well enriched with unstructured data. Consequently, combining the two research streams has the potential to reveal hitherto unidentified process information and thus create new insights. In the literature, there are approaches that incorporate the object-centric perspective, enrich traditional process data, or use unstructured data in process mining. Nevertheless, to my knowledge, no single approach is capable of addressing all points simultaneously, and no guidance is provided for incorporating the object-centric perspective to enrich structured process data with multiple unstructured data sources. Therefore, Research Article 2 investigates the following research question: *How can unstructured data be combined with structured event logs for object-centric process mining?*

To address this research question, Research Article 2 proposes the Object-Centric Reference Architecture for Unstructured Data in process mining (OCRAUD) that guides the integrated use of unstructured data sources and traditional event logs for object-centric process mining. In addition, to demonstrate the use of the reference architecture, the Object-centric Video and Sensor Architecture (OVaSA) is provided, a specific instantiation of the OCRAUD that is tailored to using video and sensor data with event logs. The research is structured along the DSR paradigm by Gregor and Hevner (2013) and DSR reference process by Peffers et al. (2007). Furthermore, the methodology outlined by Galster and Avgeriou (2011) is integrated as a specialized method for the development of reference architectures. The artifact builds on four Design Objectives (DOs) derived from the literature, and multiple evaluation episodes are carried out (Venable et al., 2016).

The OCRAUD is shown as Unified Modeling Language (UML) diagram in Figure 4. It serves as a blueprint and can be instantiated for different use cases with unstructured data in process mining. Therefore, instantiation variants are represented by optional components (dotted frames) and modular unstructured data processor subsystems for different data sources. The OCRAUD includes several main subsystems, i.e., the *Process Log Processor*, an *Unstructured Data Processor* per unstructured data source, and the *OCEL Generator*. The OCRAUD can be summarized as follows: A traditional structured process log can be loaded, preprocessed, and, if necessary, converted to the OCEL 2.0 format (Berti et al., 2024) before events and objects are extracted from it. Additionally, different unstructured data sources can be loaded and preprocessed before objects are extracted. Thereby, already known objects, e.g., from the process log, can be correlated. After objects are labeled, events can be extracted. At various points, regarding both the structured process log and the unstructured data, domain knowledge can be used as input if necessary. After all data sources are processed, the collected events can be correlated, and finally, the collection of objects and events is exported. The output of the OCRAUD is a valid OCEL 2.0 log that can be used for various process mining applications.

The evaluation strategy applied in Research Article 2 adheres to Venable et al. (2016), comprising multiple evaluation episodes to validate the research across various criteria, including completeness, feasibility, applicability, and usefulness. The evaluation

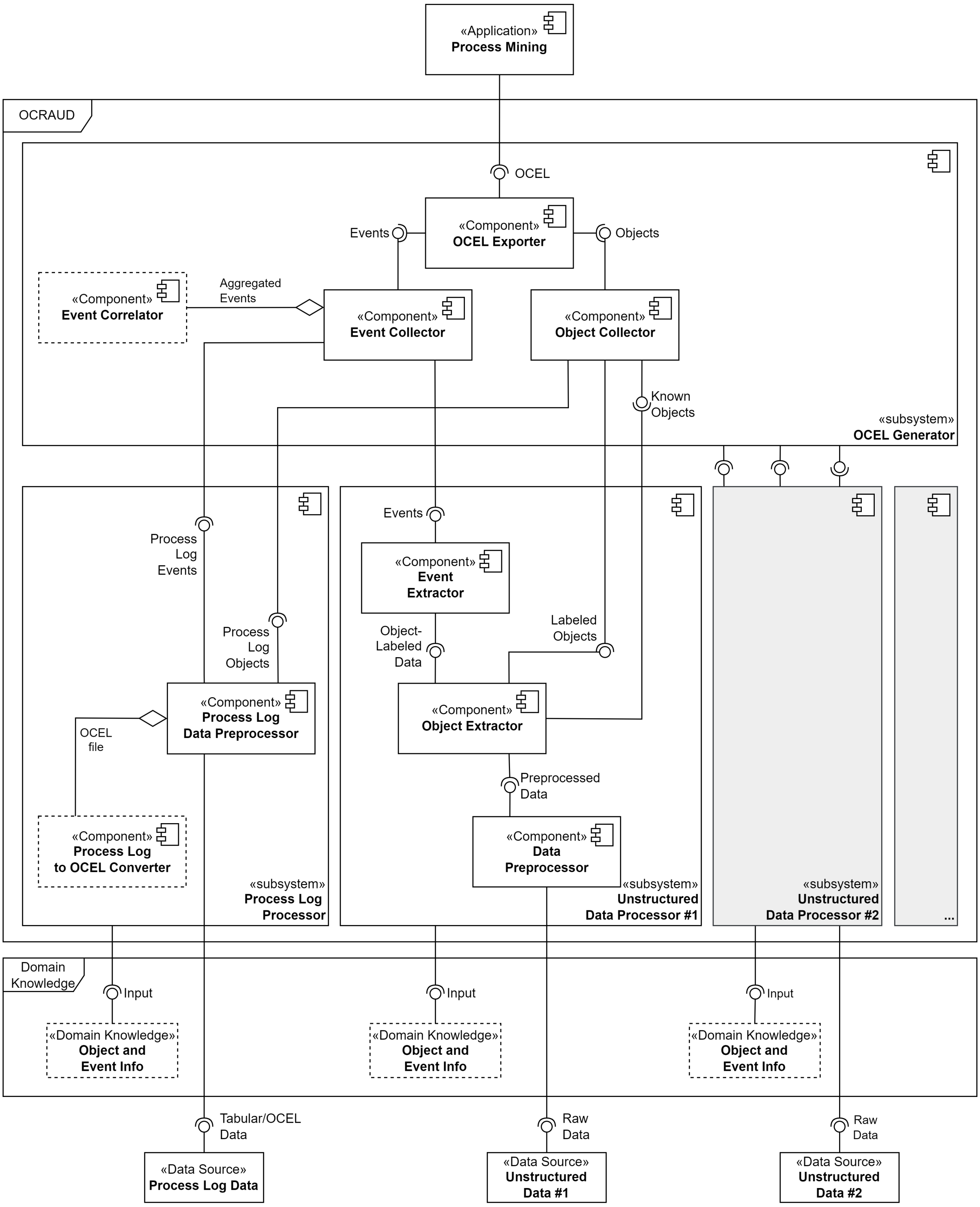

Figure 4: The OCRAUD (optional components have dotted frames)

episodes include comparing the OCRAUD to competing artifacts and conducting 10 interviews with industry and research experts, all with a background in process mining. To demonstrate the use of the artifact, the OVaSA is provided, a specific instantiation of the OCRAUD that is tailored to using video and sensor data with event logs. Moreover, a software prototype [1], including a graphical user interface, has been implemented based on it, facilitating the utilization and further development of the implemented algorithms. Figure 5 shows the object labeling step in the software prototype. In addition, a second round of interviews was conducted with the 10 experts, and the prototype has been applied to a real-world dataset (Fehrer et al., 2024; Chvirova et al., 2024).

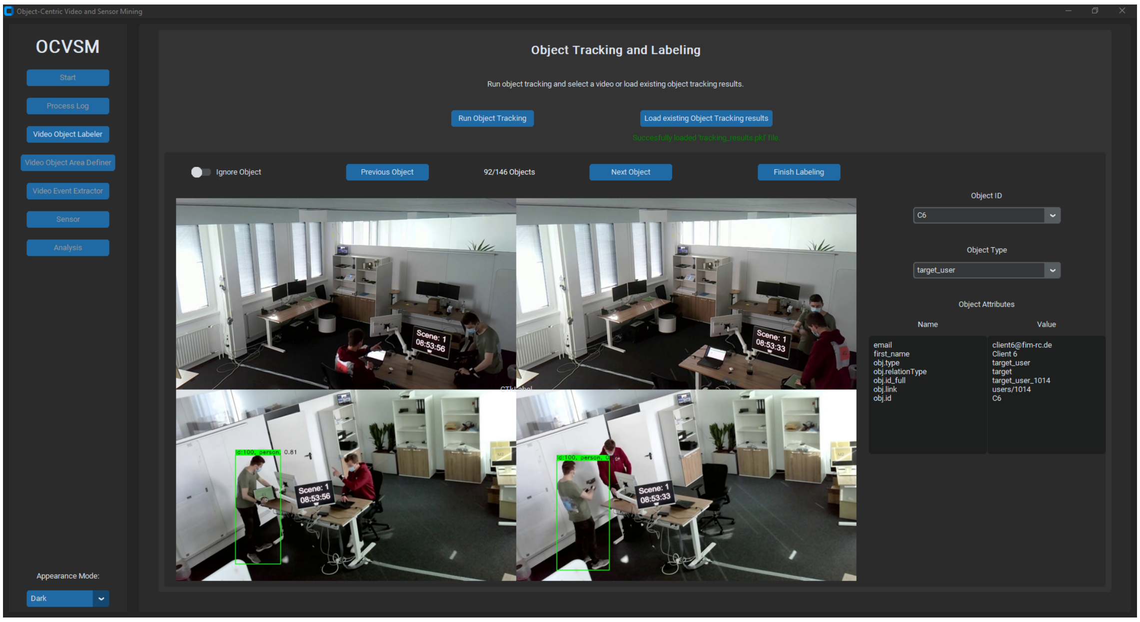

Figure 5: Screenshot of the software prototype for object-centric video and sensor mining

Research Article 2 contributes to prescriptive knowledge in process mining by giving guidance in incorporating additional unstructured data sources into event logs from information systems for object-centric process mining. It extends the existing knowledge in process mining by closing the identified research gaps, i.e., it incorporates the object-centric representation of event data, combines different unstructured data sources, and combines the unstructured data sources with event logs. The OVaSA demonstrates how to use the OCRAUD for incorporating video and sensor data as new data sources for process mining. Researchers and practitioners can extend this by instantiating new architectures for other data types or specific use cases. The code of the software prototype is made publicly available, facilitating the utilization and further development of the implemented algorithms.

[1] The source code is publicly available here: `https://github.com/550e8400e29b41d4a716446655440000/object-centric-video-sensor-mining`

III.2 Process Discovery from Video Data

While Research Article 2 in Section III.1 gives general guidance in using unstructured data sources for object-centric process mining and provides an instantiation for video and sensor data, Research Article 3 focuses specifically on video data and how to extract actual process activities from it for process discovery.

The use of video data for process mining can be particularly helpful in obtaining information about blind spots, i.e., process parts that cannot be fully captured in information systems (Kratsch et al., 2022). Despite recent developments in computer vision, approaches that use video data for process mining are sparse and tend to be tailored for specific use cases (Abbad Andaloussi et al., 2021; Knoch et al., 2019; Knoch et al., 2020; Zhou et al., 2024). A major challenge in extracting process mining activities from video data is the need to define potentially relevant activities in advance before they can be extracted into event logs. Only assumed process behavior can be identified from the video data instead of the actual process behavior since the set of potentially relevant activities is typically defined based on process models or by process experts. Therefore, deviations in process behavior, which may occur due to changes in the process environment, cannot be detected. This remains a challenge, as the identification of process deviations and new process steps is embedded within the core of process mining (van der Aalst et al., 2012). To my knowledge, there are no video-based approaches that manage to derive BPM relevant events from video data without defining a set of potentially relevant activities. This shows a need for extraction methods that can process video data in a completely unsupervised manner to identify actual processes instead of assumed ones based solely on the information contained in the video data. This would enable the detection of previously hidden but relevant process activities and thus enable transparent process discovery. Hence, Research Article 3 poses the following research question: *How can actual process activities be derived from video data as part of process discovery?*

To answer this research question, Research Article 3 introduces the Reference Architecture for Video Event Extraction (RAVEE), a reference architecture that supports extracting actual process information from unstructured video data. The RAVEE facilitates the unsupervised extraction of process steps by integrating various computer vision and clustering capabilities. The research approach is structured following the DSR paradigm and reference process (Gregor and Hevner, 2013; Peffers et al., 2007) and the six steps for the construction of empirically grounded reference architectures, proposed by Galster and

Avgeriou (2011). In developing the reference architecture, five DOs are defined based on the literature, serving as requirements for the artifact. Moreover, several evaluation activities (Sonnenberg and vom Brocke, 2012a) are completed, including a prototypical instantiation of the reference architecture.

Figure 6 illustrates the RAVEE as UML diagram, which enables the identification of actual process activities as opposed to assumed ones by performing the identification of individual process steps in an unsupervised manner. Building on Kratsch et al. (2022), the RAVEE comprises the subsystem layers *Data Preprocessor*, *Feature Extractor*, *Information Extractor*, and *Event Processor*. The RAVEE includes different instantiation variants (Galster and Avgeriou, 2011) to enable explorative, data-driven process mining using computer vision in various application areas. To ensure it can be adapted to specific process mining use cases, it comprises a set of optional components represented by the dotted frames.

Initially, the *Data Preprocessor* layer creates an interface between raw video data and the following subsystems. The *Feature Extractor* receives the preprocessed frame batch and extracts the required video features to gain a representation of the video reduced to the relevant information. Thereby, the *Feature Generator* comprises up to five sub-components, which can be combined for specific use cases in such a way that the best possible video frame representation for the selected approach is generated. The need for different video frame representations is based on the fact that the action segmentation step performed in the *Information Extractor* subsystem can be supported by various algorithms with different input needs. The *Information Extractor* layer obtains the previously computed video frame representations and generates a frame-based clustering of the individual process steps. Subsequently, activity labels are assigned to the individual process steps, realized by the two processing loops in the subsystem, both of which receive relevant process information from a human-in-the-loop instance. In the blue loop, a dictionary with activity labels is created based on the individual process steps and contextual process information. The green loop, in turn, uses this dictionary to train a supervised approach for automated recognition of the process steps. The output of this layer can be fed into the *Event Processor* subsystem layer. This subsystem converts the received information into a format suitable for an event log. It creates the respective event log that can be used for further process mining applications.

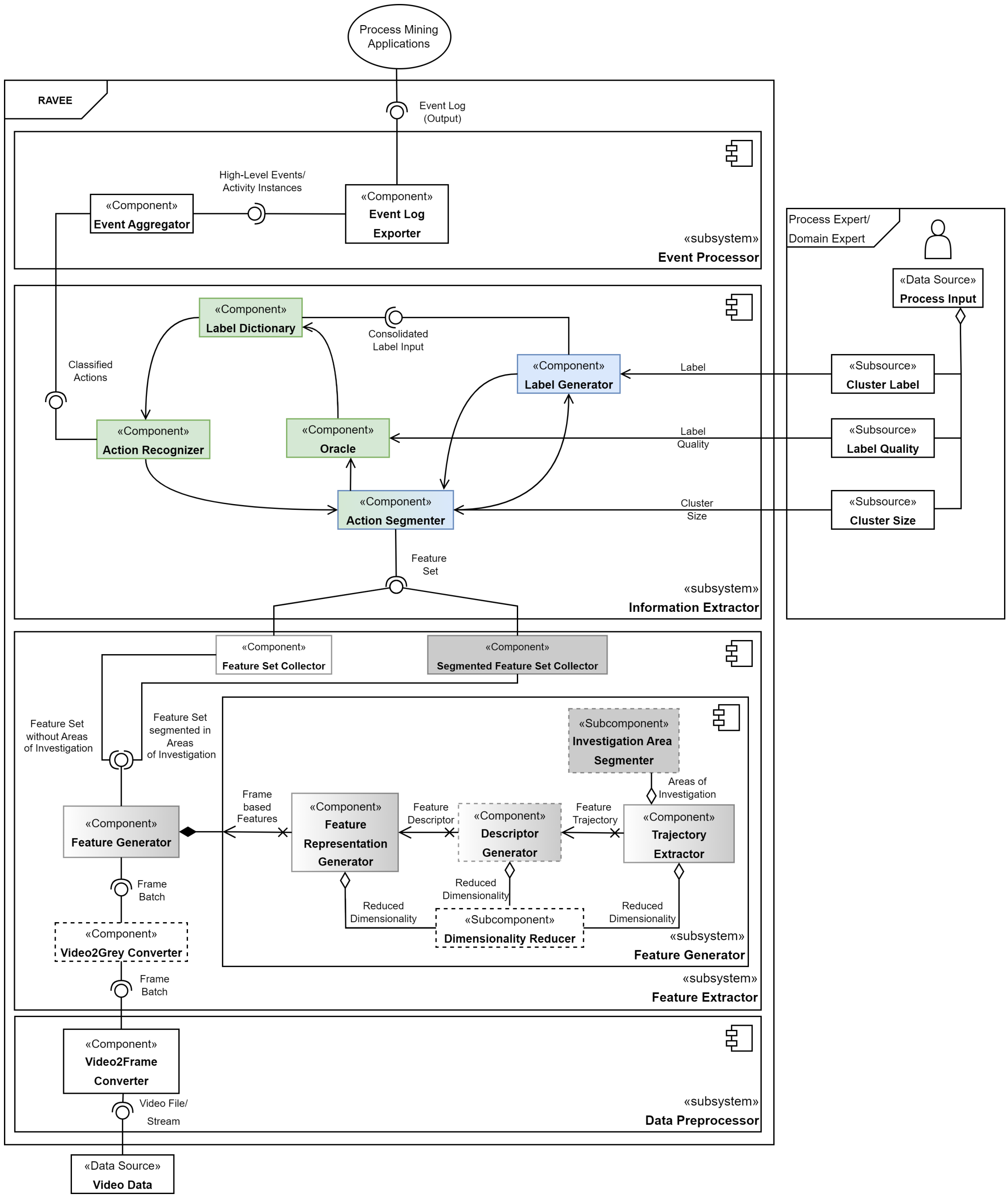

Figure 6: The RAVEE as UML diagram

The research process involved several evaluation activities in line with Sonnenberg and vom Brocke (2012a) to ensure that the artifact addresses the research problem. A competing artifacts analysis compared the RAVEE against related work and assessed their compliance with the defined DOs. Furthermore, 10 semi-structured interviews with experts from industry and research were conducted, addressing several evaluation criteria, such as understandability, completeness, and real-world fidelity. In a subsequent ex-post evaluation, parts of the reference architecture were instantiated in the form of a software prototype[2] and tested against a publicly available real-world dataset (Ben-Shabat et al., 2021). Thereby, the prototype demonstrated the real-world fidelity of the artifact by generating an event log from unstructured video data containing an assembly process performed by different actors in front of various backgrounds. The results indicate that the RAVEE can extract a high degree of process-relevant activities.

Research Article 3 extends existing knowledge by providing a reference architecture that enables the data-based, unsupervised extraction of actual process activities from unstructured video data for process mining. This enables evidence-based decision support using video data without having to rely on assumed activities and, therefore, provides holistic insights into underlying processes. The reference architecture includes several variation points, serving as a basis for further approaches and application-specific design variants. Moreover, by providing a software prototype, which has been tested as part of the evaluation activities, an end-to-end example is provided on how to instantiate the reference architecture, thus reducing the barrier of entry for creating further artifacts.

To summarize, Section III presented two contributions to video and sensor data in process mining. Research Article 2 first introduced a reference architecture that guides the use of unstructured data sources with traditional event logs for object-centric process mining. Moreover, an instantiation was provided that enables the use of video and sensor data for object-centric process mining. Research Article 3 focused specifically on video data, proposing a reference architecture that enables the unsupervised extraction of actual process activities from video data as a starting point for process discovery. Combining both contributions, guidance is provided for incorporating video and sensor data as new data sources for process mining, completed by a generic architecture on the use of unstructured data for object-centric process mining.

[2]The source code is publicly available here: `https://github.com/RAVEE-PM/RAVEE`

IV Bot and Text Data

The literature review and analysis in Section II revealed opportunities to integrate hitherto unused unstructured data sources into process mining analysis. Section III seizes these opportunities and presents two contributions that enable the use of video and sensor data for process mining and give guidance in using unstructured data sources with traditional event logs for object-centric process mining. Nevertheless, in addition to the identified unstructured data sources, there are several emerging technologies that provide further opportunities for new data sources in process mining. RPA is one such technology aiming to automate simple and repetitive tasks (van der Aalst et al., 2018; Aguirre and Rodriguez, 2017), and RPA bots are increasingly becoming part of business processes. Therefore, recorded logs of steps executed by RPA bots could be a valuable new data source for process mining to get a comprehensive overview of RPA-enabled business processes. Research Article 4 (Section IV.1) thus presents an approach that makes bot logs from RPA software usable for process mining and develops process mining measures that analyze bot logs and process event logs in an integrated way.

Another type of bots emerging due to advances in NLP are chatbots, such as ChatGPT (OpenAI, 2022). Chatbots can be applied in digital communication channels in customer service (Poddar et al., 2009) and are already in use by the public sector (Androutsopoulou et al., 2019) and various well-known companies (Følstad and Brandtzæg, 2017; Simon, 2019; Ukpabi et al., 2019). To successfully deploy chatbots in the business-to-consumer domain, they need to comply with organizational or regulatory requirements (Gunson et al., 2011), thus, the adoption of chatbots depends on the underlying model's capability to learn and comply with organizations' business processes. Therefore, Research Article 5 (Section IV.2) introduces an approach that converts textual customer conversations to event logs for process mining and quantifies chatbots' ability to learn and adhere to organizations' business processes.

IV.1 Integrated Analysis of Robotic Process Automation and Process Mining

RPA refers to tools that mimic human actions on computer systems and aim to automate simple and repetitive tasks (van der Aalst et al., 2018; Aguirre and Rodriguez, 2017). The technology is already used by many organizations, and the RPA market is expected to rise (Gartner, 2022; Deloitte, 2017). In practice, the two technologies of process mining and RPA are already brought together, e.g., by process mining vendors adding RPA functionality to their software (Geyer-Klingeberg et al., 2018; Celonis, 2024b) or vice versa (UiPath, 2024). There exist various RPA solutions with different functionalities (Dilmegani, 2023), and often, the software can be configured to record bot logs, i.e., logs of steps executed by bots. RPA bots are increasingly becoming part of business processes; thus, bot logs could be a valuable new data source for process mining analysis, and the research streams of both fields are already growing together (El-Gharib and Amyot, 2023; Leno et al., 2021). However, to my knowledge, there are no process mining approaches yet that use bot logs and focus on the post-implementation phase, i.e., when bots are already running in production. Combining bot logs and business process event logs enables insights into bot behavior and effects on the underlying business process. Therefore, an integrated view of steps performed by bots in the context of business processes is beneficial to get a comprehensive overview of RPA-enabled business processes. Research Article 4 thus investigates the following first research question: *How can bot logs be used for process mining?*

The main goal here is to find a solution for integrating bot logs with business process event logs, thus enabling the use of bot logs for process mining. A further challenge is to analyze these merged logs then. Once event logs are available, process mining offers various techniques and measures to analyze underlying processes (van der Aalst, 2016c). On the RPA side, most RPA software also includes some basic bot-related measures to monitor bot performance (Enríquez et al., 2020). However, measures specifically developed for merged process mining and RPA logs would enable integrated analyses of bots in business processes. Some approaches in the literature bring the performance measure perspectives of process mining and RPA together (Geyer-Klingeberg et al., 2018; Wewerka and Reichert, 2020; Di Bisceglie et al., 2019). However, to my knowledge, no specifically tailored process mining measures exist for analyzing merged bot and business process event logs. This could give insights into the interaction of bots with other process participants and enable an end-to-end analysis in RPA-enabled business processes. Therefore, an integrated process mining analysis with measures tailored for this purpose is needed to analyze the effects of bots on business processes. Research Article 4 investigates the following second research question: *How can bot logs and business process event logs be analyzed in an integrated way?*

To answer the two research questions, Research Article 4 first develops an integrated conceptual data model in Object-Role Modeling 2 (ORM 2) notation (Halpin and Morgan, 2008) describing the relations between bot and business processes. Furthermore, an approach is introduced that makes bot logs usable for process mining, including the ‘bot log parser’ and the ‘log merger’. Additionally, 12 integrated process mining measures are developed that consider bots as part of business processes. The measures focus on different process mining perspectives (van der Aalst et al., 2007) and on bot exceptions (Leno et al., 2018; Wewerka et al., 2021; Wewerka and Reichert, 2020). The research is structured using the DSR paradigm and reference process (Gregor and Hevner, 2013; Peffers et al., 2007) as well as evaluation patterns (Sonnenberg and vom Brocke, 2012a; Sonnenberg and vom Brocke, 2012b). The primary artifacts build on five DOs derived from the literature. Various evaluation activities (Sonnenberg and vom Brocke, 2012a; Sonnenberg and vom Brocke, 2012b) are carried out, including two interview rounds with experts and implementing and testing a software prototype with real-world and artificial data.

In line with the two research questions, an approach is developed that enables the use of bot logs for process mining and the integrated analysis of the behavior of bots in RPA-enabled business processes. The approach consists of several components that build on each other to achieve this. Figure 7 gives an overview of these components.

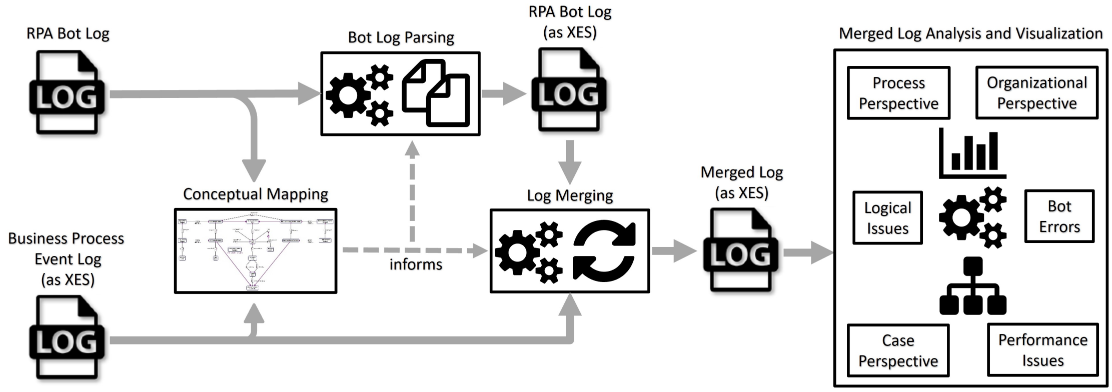

Figure 7: Conceptual overview of the approach (Egger et al., 2020; van der Aalst et al., 2007)

First, a data model is introduced that details the structure and relation of bot processes and business processes, including the attributes required for using bot logs for process mining. The data model in ORM 2 notation (Halpin and Morgan, 2008) is shown in Figure 8. This first step helps to develop the other components of the approach and ultimately supports answering the research questions.

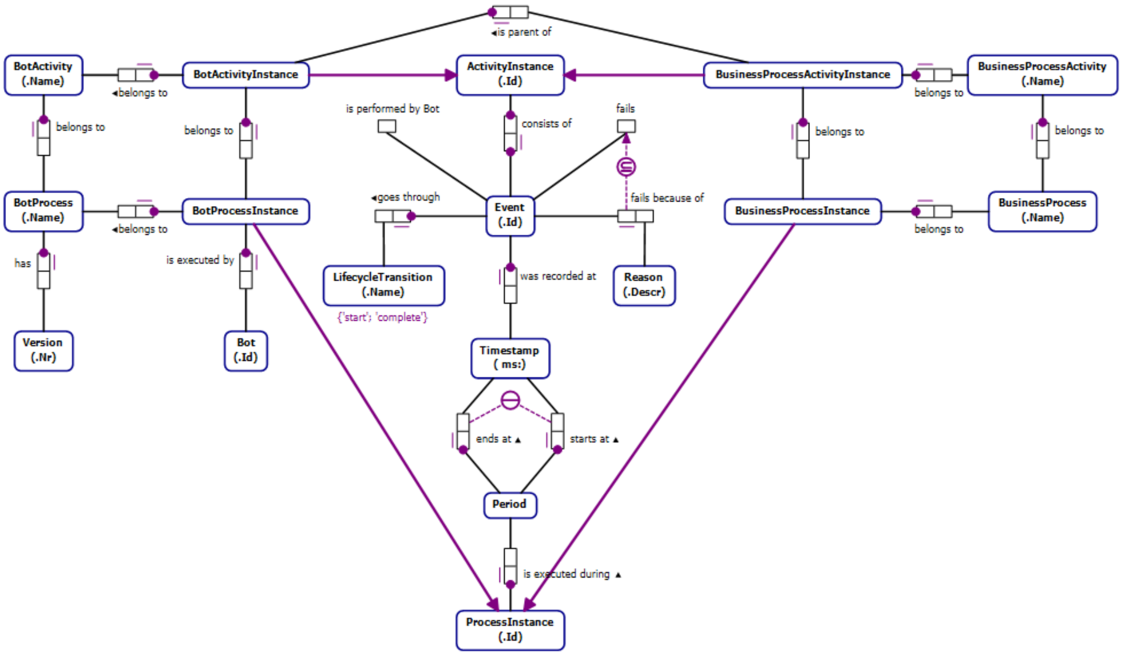

Figure 8: Structure of bot processes and business processes, including relevant attributes (Egger et al., 2020)

Following this, the bot log parser is presented. This component parses bot logs of three RPA vendors into the XES format (Acampora et al., 2017) that enables the further use of these logs in other process mining approaches. The logging formats of the different RPA tools differ; e.g., UiPath provides bot logs in a JSON-like format with many different attributes. Figure 9 shows the parsing of an exemplary UiPath bot log. Furthermore, the log merger is specified, which combines XES-parsed bot logs with business process event logs to one aggregated 'merged log'. Figure 10 shows the merging of the parsed UiPath bot log with an exemplary business process event log.

The resulting merged log can be used to gain new insights regarding the role of bots in business processes. For this purpose and specifically targeting RQ2, measures are developed that are specifically tailored to analyze the merged bot and process event logs to enable an end-to-end analysis of RPA-enabled business processes. The 12 developed measures are named: *Exception Time Impact, Exception Time Variance, Relative Fails, Relative Case Fails, Case Activities Execution Time, Case Activities Execution Time Variance, Relative Execution Time, Execution Time Variance, Automation Rate, Bot Human Handover Count, Bot Human Handover Impact,* and *Bot Human Handover Variance*. The measures are based on the literature, including different process mining perspectives (process perspective, organizational perspective, case perspective) and dimensions (logical is-

Figure 9: Exemplary parsing of a UiPath bot log to XES

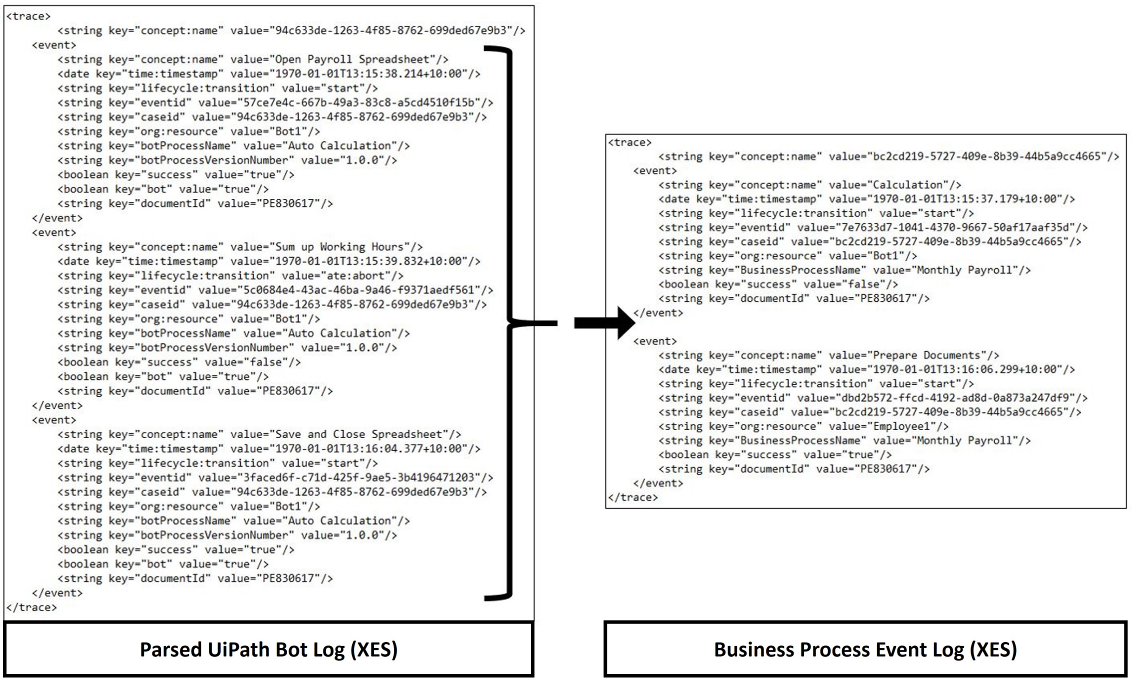

Figure 10: Exemplary merging of a parsed UiPath bot log with a business process event log

sues and performance issues) (van der Aalst et al., 2007) as well as challenges (Hindel et al., 2020) and trends (Ruiz et al., 2022) in RPA. In developing the measures, existing measures in RPA software (Enríquez et al., 2020) and process mining tools (Berti et al., 2023) have also been observed.

For evaluating the research several evaluation activities in line with Sonnenberg and vom Brocke (2012a) were carried out. This includes conducting 11 semi-structured interviews with experts from industry and research and using the feedback to develop the approach. Moreover, a software prototype[3] was implemented and instantiated. The prototype was first tested on artificial data (van Dongen and Borchert, 2018) and then using anonymized real-world data collected from the company where three interviewed experts work. Additionally, follow-up interviews were conducted with these three experts to discuss the results and assess the approach's applicability and usefulness in practice.

RPA is an emerging technology and associated bots are increasingly becoming part of business processes. The streams of RPA and process mining are already brought together in research and practice. Research Article 4 addresses these developments and contributes to prescriptive knowledge by providing a concept of using bot logs as a new data source for process mining. The study also contributes to descriptive knowledge by describing the structure of bot and business processes in an integrated data model. This facilitates a better understanding of the behavior of bots in business processes. The developed measures enable integrated analyses of bot logs and business process event logs. Furthermore, the implemented algorithms are published and thus provide a basis for further developments by practitioners and researchers.

IV.2 Chatbots' Conformance to Business Processes

While Section IV.1 focuses on enabling the use of RPA bot logs for process mining, Research Article 5 recognizes the emergence of chatbots and the need for integrating textual conversation data from chatbots into process mining. Proactive customer orientation is considered a main driver of customer value (Blocker et al., 2011) and, thus, a crucial success factor of established business models. Thereby, digital communication channels enable new opportunities to reduce customer service costs without decreasing customer satisfaction. In particular, the application of chatbots facilitates appropriate responses to repetitive customer requests while preserving the customer orientation that is essential to uphold purchase intentions (Poddar et al., 2009). Due to advances in NLP, chatbots provide a feasible, affordable, and scalable solution and are already in use by the public sector (Androutsopoulou et al., 2019) and various well-known companies, such as Amazon (Følstad and Brandtzæg, 2017), Telekom (Simon, 2019), and Lufthansa (Ukpabi

[3] The source code is publicly available here: `https://github.com/pandyke/bot-log-mining`

et al., 2019). To successfully deploy chatbots in the business-to-consumer domain, they need to comply with organizational or regulatory requirements, such as authenticating a customer, before providing sensitive details (Gunson et al., 2011). Thus, the organizational adoption of chatbots depends on the underlying model's capability to learn and comply with organizations' business processes. Consequently, Research Article 5 formulates the following research question: *How can a chatbot's ability to learn and adhere to organizations' business processes be quantified?*

To answer the research question and investigate whether chatbots achieve process-compliant behavior, Research Article 5 develops an approach that applies conformance checking, providing a viable solution to investigate whether a collection of events, represented as an event log, conforms to a process model (van der Aalst et al., 2012) or another event log (Rogge-Solti et al., 2016). For synthesizing event logs from the chatbots' training data, Natural Language Inference (NLI) is used to derive topics and process activities from customer service conversations and to represent them in an XES event log. The solution incorporates four DOs, and the research is structured along DSR methodologies (Gregor and Hevner, 2013; Peffers et al., 2007). For evaluating the approach, several evaluation activities are carried out (Sonnenberg and vom Brocke, 2012a), including a competing artifact analysis, developing a software prototype, and assessing its applicability to real-world data.

Figure 11 provides an overview of how the approach applies conformance checking to quantify a chatbot's ability to learn and adhere to organizations' business processes. Thereby, chatbots are trained on an existing set of conversations between humans and humans, such as customers and customer service agents. Apart from noisy process instances, these conversations adhere to a normative process model. If no such model exists, it can be derived from event logs synthesized from the conversations. By reporting the trace alignment (Rogge-Solti et al., 2016) between an event log constructed from conversations the chatbot has not seen during the training process and an event log from the same conversations in which the customer service agents' responses were replaced with chatbot-generated responses, the approach can assess the chatbot's overall ability to learn and adhere to business processes. To enable a breakdown on a particular process variant and the comparison against normative process models, four metrics are calculated: *Fitness* (Rozinat and van der Aalst, 2008) (measures to which extent the chatbot does not introduce new process variants), *Precision* (Muñoz-Gama and Carmona, 2010) (describes whether the process model disallows the creation of new process variants by the chatbot), *Generalization* (Buijs et al., 2012) (ensures that the process model does not overfit by capturing each trace of the event log as a separate path), and *Simplicity* (Buijs et al., 2012) (denotes whether the process model has low complexity).

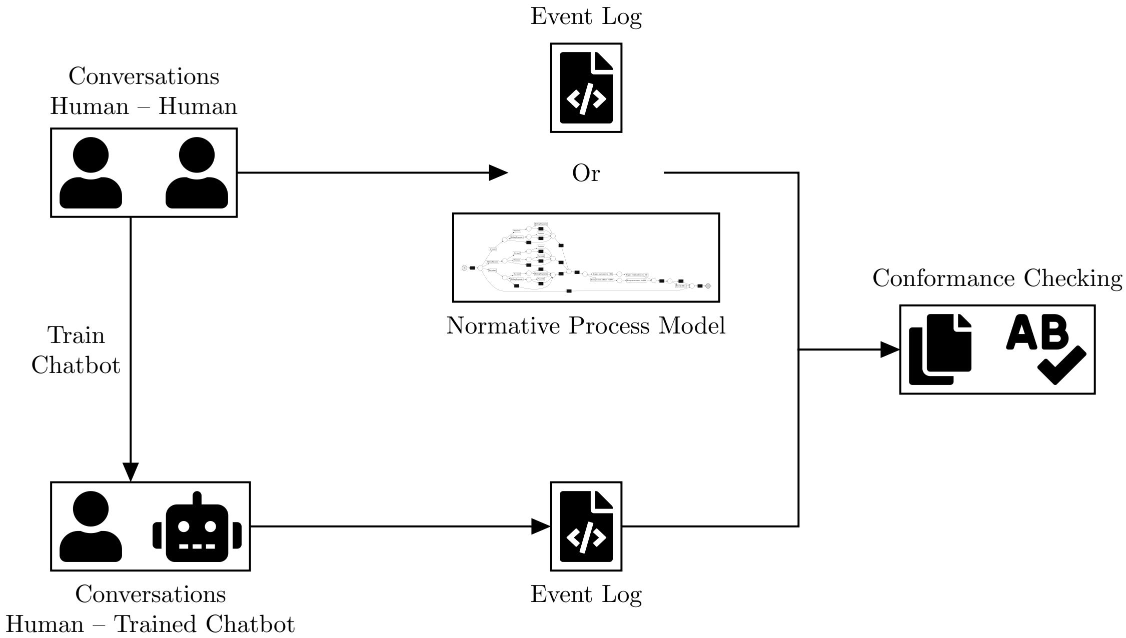

Figure 11: Overview of the approach to quantify chatbots' ability to learn and adhere to organizations' business processes

The developed approach builds upon the classical supervised machine learning evaluation workflow and complements it with a business process perspective. Figure 12 visualizes the proposed workflow. Thereby, in step 1, the chatbot is trained on the training data, followed by converting the training data to an XES event log in step 2. In step 3, the conversations are replayed in the training data by replacing the agents' responses with a chatbot-generated response, and the resulting dataset is converted to an XES event log. If a quantification of the chatbot's ability to adhere to a normative process model is desired, the normative process model can optionally be specified in step 4. However, in case there is no normative process model yet, a discovered process model using the discovery algorithms of PM4Py (Berti et al., 2023) (e.g., the Alpha Miner (van der Aalst et al., 2004), the Inductive Miner (Leemans et al., 2013), or the Heuristics Miner (Weijters et al., 2006)) can either serve as a suitable proxy or supports the specification of the normative process model. In step 5, depending on whether a quantification of the ability to learn business processes from the training data or to adhere to a normative process model is desired, the event logs resulting from step 3 are compared either to the proxy models discovered in step 4 or to the specified normative process model using the four metrics fitness, precision, generalization, and simplicity. The test data is then converted to an XES event log in step 6. In step 7, the conversations are replayed in the test data by replacing the agents' responses with a chatbot-generated response, and the resulting dataset is converted to an XES event log. In step 8, depending on whether quantification of the ability to learn business processes from the training data or to adhere to a normative process model is desired, the event logs resulting from step 7 are compared either to the proxy models discovered in step 4 or to the specified normative process model and trace alignment is applied on the event logs resulting from steps 6 and 7.

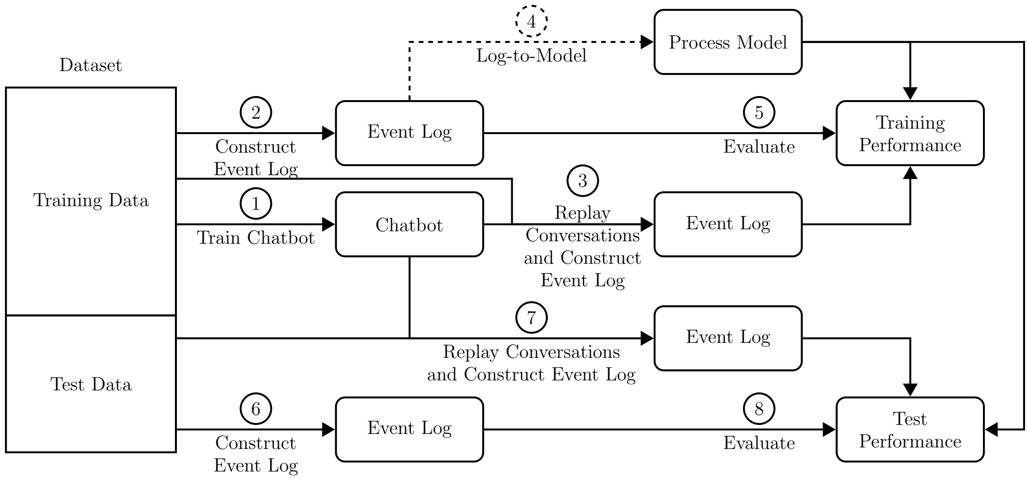

Figure 12: Proposed chatbot evaluation workflow

For evaluating the proposed approach, several evaluation activities were conducted in alignment with the framework by Sonnenberg and vom Brocke (2012a). The activities included discussing the developed solution against competing approaches, developing a software prototype, and assessing its applicability to a real-world dataset. To this end, three chatbots were trained using a Sequence-to-Sequence (Seq2Seq) model (Sutskever et al., 2014) on a publicly available corpus of Twitter conversations (Thought Vector, 2017) of AmazonHelp, AppleSupport, and SpotifyCares. The dataset comprises over 500,000 Tweets in total. Afterward, it is demonstrated how the approach quantifies a chatbot's overall ability to learn business processes from the training data and how it quantifies a chatbot's ability to learn a particular variant of the underlying process. Following this, it is shown how to compare the chatbot's executed steps against a given normative process model. The implementation of the software prototype was separated into the three modules ‘chatbot’, ‘event log construction’, and ‘chatbot process mining’[4]. Each module can serve its use case independently and can be exchanged by a different implementation if necessary.

Due to advances in NLP, chatbots are emerging, and Research Article 5 recognized the need for integrating chatbot conversations into process mining. A new chatbot evaluation workflow was developed that quantifies a chatbot’s ability to adhere to organizations’ business processes. Thereby, textual conversations were enabled as a new data source for process mining. The approach contributes to a more holistic view of business processes by incorporating steps executed by chatbots. Consequently, mining interaction steps at the interface between customers and business processes (Pallotta and Delmonte, 2013) enables filling blind spots (Kratsch et al., 2022) with information. The implementation of the approach is publicly available and, thus, can be adopted and extended by the community. Further, the approach can complement previously available business process event logs with chatbot-related event logs.

In summary, Section IV presented two contributions to integrating data from emerging technologies into process mining. Research Article 4 introduced an approach that enables using bot logs from RPA software for process mining. Research Article 5 proposed a workflow that quantifies a chatbot’s ability to adhere to organizations’ business processes, thereby enabling the use of textual conversation data from chatbots for process mining. Both contributions facilitate a more holistic process view by incorporating data from chatbots and RPA bots as new data sources for process mining.

[4]The source code of the three modules is publicly available here: `https://github.com/kechtel/customer-support-chatbot`, `https://github.com/kechtel/ericsson`, `https://github.com/kechtel/chatbot-process-mining`

V Conclusion

V.1 Summary

The field of process mining has experienced significant growth and development over the past few decades, introducing innovative techniques and applications in both research (van der Aalst, 2022; Boltenhagen et al., 2021; Li et al., 2022; Li and Zelst, 2022) and practice (Reinkemeyer, 2020b; Gartner, Inc., 2024). In process mining, information systems, such as CRM systems and ERP systems, have historically constituted the primary data sources (Diba et al., 2020; van der Aalst, 2016b). Nevertheless, process information can also be derived from alternative sources and employed in process mining (Reinkemeyer, 2020a; van der Aalst, 2016b). Notwithstanding the availability (Gandomi and Haider, 2015) and potential advantages (Beverungen et al., 2021; Koschmider et al., 2023; Grisold et al., 2021; Kerpedzhiev et al., 2021) of unstructured data sources, they are frequently not used for process mining analysis and current studies mainly focus on traditional structured data sources (Diba et al., 2020; Janiesch et al., 2020). This dissertation, comprising five research articles, contributes to the advancement of process mining by enabling the use of new data sources. First, an overview of the current state of research on the use of unstructured data in process mining and a research agenda are presented, providing new insights to systematically advance research in this field. Second, this dissertation presents two artifacts that provide guidance for incorporating video and sensor data as new data sources for process mining, completed by a generic architecture on the use of unstructured data for object-centric process mining. Third, two artifacts that integrate data from emerging technologies are presented. Both artifacts facilitate a more holistic process view by incorporating data from chatbots and RPA bots as new data sources for process mining.

Although initial research has been conducted on unstructured data in process mining, a systematic overview that summarizes existing approaches is needed to accelerate progress in this promising area. Section II addresses this opportunity, presenting Research Article 1, which provides a systematic literature review with a focus on technical process mining artifacts that enable (semi-)automated consideration of unstructured data. From the 1,379 identified research items, 24 primary studies were selected for analysis. Based on the results, Research Article 1 proposes a research agenda comprising seven opportunities to address the identified research gaps. This paves the way for future process mining research incorporating new data sources.

The integration of a specific data type, such as textual, sensory, or video data, into process mining analysis is contingent upon the particular process scenario under consideration. In many scenarios, the use of data derived from video and sensor data has the potential to facilitate the detection of previously hidden but relevant process activities, thereby enabling a more transparent process representation. Consequently, a variety of opportunities for process mining emerge, including leveraging video and sensor data for object-centric process mining. Accordingly, Section III presents two artifacts developed based on DSR principles that concentrate on the incorporation of video and sensor data as new data sources for process mining. Research Article 2 proposes the OCRAUD, a reference architecture that offers guidance on the integrated use of unstructured data sources and traditional event logs for object-centric process mining. To demonstrate the application, the OVaSA is introduced, an instantiation that focuses specifically on video and sensor data. A series of evaluation episodes were conducted, including a competing artifacts analysis, two rounds of expert interviews, the implementation of a software prototype, and its use on a real-world dataset. Research Article 3 focuses specifically on video data, introducing the RAVEE, a reference architecture designed to facilitate the extraction of actual process information from unstructured video data. This enables the data-driven, unsupervised identification of real process steps, thereby facilitating evidence-based decision support using video data. It is not contingent on assumptions regarding activities and provides a comprehensive understanding of the underlying processes. A multitude of evaluation activities were conducted, including a prototypical instantiation tested against real-world data, a competing artifacts analysis, and expert interviews.

The combination of process mining with several emerging technologies has enabled the integration of further data, extending beyond the previously considered unstructured data sources. As the prevalence of RPA bots in business processes continues to grow, there is an increasing need to incorporate the actions performed by bots into process mining analysis. Furthermore, chatbots are deployed in contexts where alignment with the underlying business processes is essential. Consequently, it is neccessary to integrate textual conversation data from chatbots to ascertain whether chatbots are achieving process-compliant behavior. Section IV recognizes these needs and presents two artifacts developed based on DSR principles that enable the inclusion of bot and textual conversation data as new data sources for process mining. Research Article 4 proposes an approach that integrates RPA with process mining. A conceptual data model is developed, which describes the relations between bots and business processes, and an approach is presented that makes bot logs usable for process mining. Moreover, 12 integrated process mining measures are developed that consider bots as integral components of business processes. The approach facilitates a more comprehensive understanding of the behavior of bots in business processes and enables the integrated analysis of bot logs and business process event logs. A series of evaluation activities were conducted, comprising two rounds of expert interviews and the implementation and testing of a software prototype with both real-world and artificial data. Research Article 5 presents an approach that enables textual customer service conversations from chatbots to be utilized as a new data source for process mining to quantify chatbots' ability to learn and adhere to organizations' business processes. The approach can be used to supplement previously available business process event logs with chatbot-related event logs, thereby providing a more comprehensive view of business processes. To validate the approach, a number of evaluation activities were conducted, including a competing artifact analysis, the development of a prototype, and an assessment of its applicability to real-world data.

V.2 Limitations and Future Research

The findings of this dissertation, including those of the embedded research articles, are constrained by limitations that call for further research. Sections VII.3 to VII.7 present the individual limitations of each research article, while this section discusses limitations and avenues for future research in a broader view and in light of the used research methods.

As primary research methods, the five research articles in this dissertation used DSR principles (Gregor and Hevner, 2013; Peffers et al., 2007), guidance on the development of reference architectures (Galster and Avgeriou, 2011), and design steps for systematic literature reviews (Kitchenham and Charters, 2007). Research Articles 2, 3, 4, and 5 build on DSR principles to construct artifacts. March and Smith (1995) highlight that feasibility is shown by building an artifact, whereas the evaluation of it provides insight into its efficacy. Accordingly, for evaluating the presented artifacts, evaluation methods proposed by Sonnenberg and vom Brocke (2012a) and by Venable et al. (2016) were used. The evaluation activities included competing artifacts analyses, interviews with experts from both research and practice, the implementation of software prototypes, and the use of the prototypes on artificial and real-world datasets. Nevertheless, future research might benefit from more comprehensive evaluation activities, such as conducting additional interview rounds throughout the artifact development process and interviews with a greater number of experts. Moreover, the implemented software prototypes could be utilized on additional real-world datasets derived from organizations, facilitating a more naturalistic evaluation, i.e., with real tasks, systems, and users (Sonnenberg and vom Brocke, 2012a). A more comprehensive evaluation process would further substantiate the efficacy and usefulness of the proposed artifacts.

Research Articles 2 and 3 develop reference architectures as artifacts (Galster and Avgeriou, 2011), which are designed to serve as blueprints for software architectures and corresponding instantiations. Illustrating the utilization of a reference architecture, for instance, by providing and describing instantiations of it, as demonstrated in Research Article 2, offers guidance on the use of the artifact. Nevertheless, future work would benefit from the inclusion of comprehensive user guides and a greater number of exemplary instantiations. Furthermore, when designing reference architectures, a substantial challenge is the selection of an appropriate level of abstraction that strikes a balance between generality and context-specificity (Galster and Avgeriou, 2011). From one perspective, reference architectures should be sufficiently generic to address the general research problem. From another perspective, a more detailed architectural level can be tailored to specific use cases, thereby facilitating efficient use in these particular contexts. The objective of the reference architectures introduced in the research articles was to present a balanced solution. However, future research could develop reference architectures concerning new data sources for process mining tailored to specific contexts or organizational types. For example, architectures could be developed for only the manufacturing context or only for small or medium-sized organizations, facilitating a more efficient use in these contexts.

In Research Article 1, a systematic literature review was conducted in accordance with the guidelines set forth by Kitchenham and Charters (2007). Despite adherence to established methods, literature reviews can be subject to several limitations. The literature search was conducted using a database-driven search approach and a generic search phrase that did not include any specific search terms for the different types of unstructured data. This approach was employed to avoid bias towards pre-identified types of unstructured data. Moreover, the search phrase was refined in several iterations to ensure comprehensive coverage of relevant research. However, it is possible that some potentially relevant studies may not have been identified. Furthermore, the literature search was confined to peer-reviewed scientific literature. As part of future research, the findings could be enhanced by integrating non-scientific sources of information, such as white papers, tools and reports from leading software vendors, as well as practitioner-oriented books (Silva et al., 2021). Moreover, a literature review is constrained by the temporal boundaries of the search period. Given the rapid pace of research, it is essential to ensure that the results of the literature search are updated with regularity, thus providing a current foundation for other researchers and practitioners. This requires future work to observe and summarize the developments in the field of unstructured data in process mining.

As this dissertation aims to enable new data sources for process mining, the embedded research articles construct artifacts that bring previously unused data into a format that can be used by process mining applications. Consequently, Research Articles 3, 4, and 5 facilitate the transformation of data into the well-established XES format (Acampora et al., 2017). However, the recent introduction of object-centric process mining (van der Aalst, 2019; Esser and Fahland, 2021) has resulted in the creation of new log standards (Berti et al., 2024; Berti and van der Aalst, 2023), as employed in Research Article 2. These new developments have already been incorporated by vendors, paving the way for the future of process mining in practice (Celonis, 2024a). It is thus essential that future research builds upon and updates the created artifacts in light of new developments in process mining. Furthermore, the research articles included in this dissertation employ techniques and algorithms that introduce uncertainty into the data. For example, Research Article 2 uses object tracking (Wu et al., 2013), whereby objects are identified with a certain probability, leading to uncertainty in the resulting data used for process mining. Similarly, Research Article 5 applies NLI, resulting in inaccuracies in the created event logs. It is thus imperative that future research in process mining considers the inherent uncertainty associated with such event data (van der Aalst, 2020). Finally, this dissertation does not purport to be exhaustive in its coverage of potential data sources for process mining. While sensor, video, bot, and text data are addressed, there are further possibilities regarding these four data types or completely other data sources. Future work could explore the use of audio data for process mining or concentrate on further emerging technologies that may also contain valuable insights, contributing to a more comprehensive process picture.

Overall, the rapid growth in accessible data presents considerable opportunities for process mining. I am confident that this dissertation will contribute to a more comprehensive understanding of processes and provide valuable assistance to researchers and practitioners in their pursuit of exploring new data sources for process mining.

Use of Writing Assistance Please note that I have utilized DeepL and Grammarly to enhance the language and readability of this dissertation. However, I take full responsibility for its content and have thoroughly reviewed and edited the material as necessary.

VI References

Abbad Andaloussi, A., F. Zerbato, A. Burattin, T. Slaats, T. T. Hildebrandt, and B. Weber (2021). "Exploring how users engage with hybrid process artifacts based on declarative process models: a behavioral analysis based on eye-tracking and think-aloud". In: *Software and Systems Modeling* 20.5, pp. 1437–1464. DOI: 10.1007/s10270-020-00811-8.

Acampora, G., A. Vitiello, B. N. D. Stefano, W. M. P. van der Aalst, C. W. Günther, and E. Verbeek (2017). "IEEE 1849: The XES Standard: The Second IEEE Standard Sponsored by IEEE Computational Intelligence Society [Society Briefs]". In: *IEEE Computational Intelligence Magazine* 12.2, pp. 4–8. DOI: 10.1109/MCI.2017.2670420.

Aggarwal, J. K. and M. S. Ryoo (2011). "Human activity analysis: A review". In: *ACM Computing Surveys* 43.3, 16:1–16:43. DOI: 10.1145/1922649.1922653.

Aguirre, S. and A. Rodriguez (2017). "Automation of a business process using robotic process automation (RPA): A case study". In: *Applied Computer Sciences in Engineering: 4th Workshop on Engineering Applications, WEA 2017, Cartagena, Colombia, September 27-29, 2017, Proceedings 4*. Springer, pp. 65–71. DOI: 10.1007/978-3-319-66963-2_7.

Androutsopoulou, A., N. Karacapilidis, E. Loukis, and Y. Charalabidis (2019). "Transforming the communication between citizens and government through AI-guided chatbots". In: *Government Information Quarterly* 36.2, pp. 358–367. DOI: 10.1016/j.giq.2018.10.001.

Balducci, B. and D. Marinova (2018). "Unstructured data in marketing". In: *Journal of the Academy of Marketing Science* 46, pp. 557–590. DOI: 10.1007/s11747-018-0581-x.

Bandara, W., E. Furtmueller, E. Gorbacheva, S. Miskon, and J. Beekhuyzen (2015). "Achieving rigor in literature reviews: Insights from qualitative data analysis and tool-support". In: *Communications of the Association for Information systems* 37.1, pp. 154–204. DOI: 10.17705/1CAIS.03708.

Banziger, R. B., A. Basukoski, and T. Chaussalet (2018). "Discovering Business Processes in CRM Systems by Leveraging Unstructured Text Data". In: *2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Confer-*

ence on Data Science and Systems (HPCC/SmartCity/DSS), pp. 1571–1577. DOI: 10.1109/HPCC/SmartCity/DSS.2018.00257.

Ben-Shabat, Y., X. Yu, F. S. Saleh, D. Campbell, C. R. Opazo, H. Li, and S. Gould (2021). "The IKEA ASM Dataset: Understanding People Assembling Furniture through Actions, Objects and Pose". In: *IEEE Winter Conference on Applications of Computer Vision, WACV 2021, Waikoloa, HI, USA, January 3-8, 2021*. IEEE, pp. 846–858. DOI: 10.1109/WACV48630.2021.00089.

Berti, A., I. Koren, J. N. Adams, G. Park, B. Knopp, N. Graves, M. Rafiei, L. Liß, L. T. genannt Unterberg, Y. Zhang, C. T. Schwanen, M. Pegoraro, and W. M. P. van der Aalst (2024). "OCEL (Object-Centric Event Log) 2.0 Specification". In: *CoRR* abs/2403.01975. DOI: 10.48550/ARXIV.2403.01975.

Berti, A. and W. M. P. van der Aalst (2023). "OC-PM: analyzing object-centric event logs and process models". In: *International Journal on Software Tools for Technology Transfer* 25.1, pp. 1–17. DOI: 10.1007/s10009-022-00668-w.

Berti, A., S. van Zelst, and D. Schuster (2023). "PM4Py: A process mining library for Python". In: *Software Impacts* 17, p. 100556. DOI: 10.1016/j.simpa.2023.100556.

Beverungen, D., J. C. A. M. Buijs, J. Becker, C. Di Ciccio, W. M. P. van der Aalst, and et al. (2021). "Seven Paradoxes of Business Process Management in a Hyper-Connected World". In: *Business & Information Systems Engineering* 63.2, pp. 145–156. DOI: 10.1007/s12599-020-00646-z.

Blocker, C. P., D. J. Flint, M. B. Myers, and S. F. Slater (2011). "Proactive customer orientation and its role for creating customer value in global markets". In: *Journal of the Academy of Marketing Science* 39.2, pp. 216–233. DOI: 10.1007/s11747-010-0202-9.

Boltenhagen, M., T. Chatain, and J. Carmona (2021). "An A*-Algorithm for Computing Discounted Anti-Alignments in Process Mining". In: *2021 3rd International Conference on Process Mining (ICPM)*. IEEE, pp. 25–31. DOI: 10.1109/ICPM53251.2021.9576887.

Buijs, J. C. A. M., B. F. van Dongen, and W. M. P. van der Aalst (2012). "On the Role of Fitness, Precision, Generalization and Simplicity in Process Discovery". In: *On the Move to Meaningful Internet Systems: OTM 2012*. Berlin, Heidelberg: Springer, pp. 305–322. DOI: 10.1007/978-3-642-33606-5_19.

Celonis (2024a). *What is object-centric process mining?* https://www.celonis.com/blog/what-is-object-centric-process-mining-ocpm/. Accessed: 2024-09-08.

Celonis (2024b). *What is RPA and how does combining it with process intelligence help drive process excellence?* https://www.celonis.com/what-is-rpa/#5-orchestrate-rpas-with-celonis. Accessed: 2024-09-08.

Chambers, A. J., A. M. Stringfellow, B. B. Luo, S. J. Underwood, T. G. Allard, I. A. Johnston, S. Brockman, L. Shing, A. Wollaber, and C. VanDam (2020). “Automated Business Process Discovery from Unstructured Natural-Language Documents”. In: *Business Process Management Workshops*. Cham: Springer International Publishing, pp. 232–243. DOI: 10.1007/978-3-030-66498-5_18.

Chiudinelli, L., A. Dagliati, V. Tibollo, S. Albasini, N. Geifman, N. Peek, J. H. Holmes, F. Corsi, R. Bellazzi, and L. Sacchi (2020). “Mining post-surgical care processes in breast cancer patients”. In: *Artificial Intelligence in Medicine* 105, p. 101855. DOI: 10.1016/j.artmed.2020.101855.

Chvirova, D., A. Egger, T. Fehrer, W. Kratsch, M. Röglinger, J. Wittmann, and N. Wördehoff (2024). “A multimedia dataset for object-centric business process mining in IT asset management”. In: *Data in Brief* 55, p. 110716. DOI: 10.1016/j.dib.2024.110716.

Deloitte (2017). *The robots are ready. Are you? Untapped advantage in your digital workforce*. https://www2.deloitte.com/content/dam/Deloitte/tr/Documents/technology/deloitte-robots-are-ready.pdf. Accessed: 2024-09-08.

Di Bisceglie, C, E Ramezani Taghiabadi, and H Aklecha (2019). *Data-driven insights to Robotic Process Automation with Process Mining*. https://www.compact.nl/pdf/C-2019-3-Bisceglie.pdf. Accessed: 2024-09-08.

Di Francescomarino, C. and C. Ghidini (2022). “Predictive Process Monitoring”. In: *Process Mining Handbook*. Ed. by W. M. P. van der Aalst and J. Carmona. Springer International Publishing, pp. 320–346. DOI: 10.1007/978-3-031-08848-3_10.

Diba, K., K. Batoulis, M. Weidlich, and M. Weske (2020). “Extraction, correlation, and abstraction of event data for process mining”. In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 10.3. DOI: 10.1002/WIDM.1346.

Dilmegani, C. (2023). *Top 53 RPA Tools / Vendors & Their Features in 2023*. https://research.aimultiple.com/rpa-tools/#what-is-the-best-rpa-software. Accessed: 2023-12-03.

Dumas, M., M. L. Rosa, J. Mendling, and H. A. Reijers (2018). *Fundamentals of Business Process Management, Second Edition*. Springer. DOI: 10.1007/978-3-662-56509-4.

Egger, A., A. H. M. ter Hofstede, W. Kratsch, S. J. J. Leemans, M. Röglinger, and M. T. Wynn (2020). "Bot Log Mining: Using Logs from Robotic Process Automation for Process Mining". In: *Conceptual Modeling - 39th International Conference, ER 2020, Vienna, Austria, November 3-6, 2020, Proceedings*. Vol. 12400. Springer, pp. 51–61. DOI: 10.1007/978-3-030-62522-1_4.

El-Gharib, N. M. and D. Amyot (2023). "Robotic process automation using process mining — A systematic literature review". In: *Data & Knowledge Engineering* 148, p. 102229. DOI: 10.1016/j.datak.2023.102229.

Elkoumy, G., S. A. Fahrenkrog-Petersen, M. F. Sani, A. Koschmider, F. Mannhardt, S. N. von Voigt, M. Rafiei, and L. von Waldthausen (2022). "Privacy and Confidentiality in Process Mining: Threats and Research Challenges". In: *ACM Transactions on Management Information Systems* 13.1, pp. 1–17. DOI: 10.1145/3468877.

Elleuch, M., O. A. Ismaili, N. Laga, W. Gaaloul, and B. Benatallah (2020). "Discovering Activities from Emails Based on Pattern Discovery Approach". In: *Business Process Management Forum*. Springer International Publishing, pp. 88–104. DOI: 10.1007/978-3-030-58638-6_6.

Enríquez, J. G., A. J. Ramirez, F. J. D. Mayo, and J. A. García-García (2020). "Robotic Process Automation: A Scientific and Industrial Systematic Mapping Study". In: *IEEE Access* 8, pp. 39113–39129. DOI: 10.1109/ACCESS.2020.2974934.

Epure, E. V., P. Martín-Rodilla, C. Hug, R. Deneckère, and C. Salinesi (2015). "Automatic process model discovery from textual methodologies". In: *2015 IEEE 9th International Conference on Research Challenges in Information Science (RCIS)*, pp. 19–30. DOI: 10.1109/RCIS.2015.7128860.

Esser, S. and D. Fahland (2021). "Multi-Dimensional Event Data in Graph Databases". In: *Journal on Data Semantics* 10.1-2, pp. 109–141. DOI: 10.1007/S13740-021-00122-1.

Fehrer, T., A. Egger, D. Chvirova, J. Wittmann, N. Wördehoff, W. Kratsch, and M. Röglinger (2024). *Business Processes in IT Asset Management Multimedia Event Log*. DOI: 10.6084/m9.figshare.25246291.v1.

Folino, F., M. Guarascio, and L. Pontieri (2015). "An Approach to the Discovery of Accurate and Expressive Fix-Time Prediction Models". In: *Enterprise Information Systems*. Ed. by J. Cordeiro, S. Hammoudi, L. Maciaszek, O. Camp, and J. Filipe. Springer International Publishing, pp. 108–128. DOI: 10.1007/978-3-319-22348-3_7.

Følstad, A. and P. B. Brandtzæg (2017). "Chatbots and the New World of HCI". In: *Interactions* 24.4, pp. 38–42. DOI: 10.1145/3085558.

Fortune Business Insights (2024). *Process Mining Software Market Size, Share & Industry Analysis*. https://www.fortunebusinessinsights.com/process-mining-software-market-104792. Accessed: 2024-09-08.

Galster, M. and P. Avgeriou (2011). "Empirically-grounded reference architectures: a proposal". In: *7th International Conference on the Quality of Software Architectures, QoSA 2011 and 2nd International Symposium on Architecting Critical Systems, ISARCS 2011. Boulder, CO, USA, June 20-24, 2011, Proceedings*. ACM, pp. 153–158. DOI: 10.1145/2000259.2000285.

Gandomi, A. and M. Haider (2015). "Beyond the hype: Big data concepts, methods, and analytics". In: *International Journal of Information Management* 35.2, pp. 137–144. DOI: 10.1016/J.IJINFOMGT.2014.10.007.

Gartner (2022). *Gartner Says Worldwide RPA Software Spending to Reach $2.9 Billion in 2022*. https://www.gartner.com/en/newsroom/press-releases/2022-08-1-rpa-forecast-2022-2q22-press-release. Accessed: 2023-12-03.

Gartner, Inc. (2024). *Process Mining Platforms Reviews and Ratings*. https://www.gartner.com/reviews/market/process-mining-tools. Accessed: 2024-09-08.

Geyer-Klingeberg, J., J. Nakladal, F. Baldauf, and F. Veit (2018). "Process Mining and Robotic Process Automation: A Perfect Match". In: *Proceedings of the Dissertation Award, Demonstration, and Industrial Track at BPM 2018, Sydney, Australia, September 9-14, 2018*. Vol. 2196. CEUR Workshop Proceedings. https://ceur-ws.org/Vol-2196/BPM_2018_paper_28.pdf. CEUR-WS.org, pp. 124–131.

Gregor, S. and A. R. Hevner (2013). "Positioning and Presenting Design Science Research for Maximum Impact". In: *MIS quarterly* 37.2, pp. 337–355. DOI: 10.25300/MISQ/2013/37.2.01.

Grisold, T., J. Mendling, M. Otto, and J. vom Brocke (2021). "Adoption, use and management of process mining in practice". In: *Business Process Management Journal* 27.2, pp. 369–387. DOI: 10.1108/BPMJ-03-2020-0112.

Gunson, N., D. Marshall, H. Morton, and M. Jack (2011). "User perceptions of security and usability of single-factor and two-factor authentication in automated telephone banking". In: *Computers & Security* 30.4, pp. 208–220. DOI: https://doi.org/10.1016/j.cose.2010.12.001.

Gupta, M., P. Agarwal, T. Tater, S. Dechu, and A. Serebrenik (2020). "Analyzing Comments in Ticket Resolution to Capture Underlying Process Interactions". In: *Business Process Management Workshops*. Vol. 397. Springer International Publishing, pp. 219–231. DOI: 10.1007/978-3-030-66498-5_17.

Halpin, T. A. and T. Morgan (2008). *Information modeling and relational databases (2. ed.)* Morgan Kaufmann. DOI: 10.1016/B978-1-55860-672-2.X5000-3.

Hemmer, A., M. Abderrahim, R. Badonnel, J. François, and I. Chrisment (2021). "Comparative Assessment of Process Mining for Supporting IoT Predictive Security". In: *IEEE Transactions on Network and Service Management* 18.1, pp. 1092–1103. DOI: 10.1109/TNSM.2020.3038172.

Hiebl, M. R. (2023). "Sample selection in systematic literature reviews of management research". In: *Organizational research methods* 26.2, pp. 229–261. DOI: 10.1177/1094428120986851.

Hindel, J., L. M. Cabrera, and M. Stierle (2020). "Robotic Process Automation: Hype or Hope?" In: *Entwicklungen, Chancen und Herausforderungen der Digitalisierung: Proceedings der 15. Internationalen Tagung Wirtschaftsinformatik, WI 2020, Potsdam, Germany, March 9-11, 2020. Zentrale Tracks*. GITO Verlag, pp. 1750–1762. DOI: 10.30844/WI_2020_R6-HINDEL.

Holstrup, A., L. Starklit, and A. Burattin (2020). "Analysis of Information-Seeking Conversations with Process Mining". In: *2020 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. DOI: 10.1109/IJCNN48605.2020.9207187.

Holz, F., B. Lantow, and M. Fellmann (2021). "Towards a Content-Based Process Mining Approach in Personal Services". In: *Enterprise, Business-Process and Information Systems Modeling*. Ed. by A. Augusto, A. Gill, S. Nurcan, I. Reinhartz-Berger, R.

Schmidt, and J. Zdravkovic. Springer International Publishing, pp. 62–77. ISBN: 978-3-030-79186-5. DOI: 10.1007/978-3-030-79186-5_5.

Janiesch, C., A. Koschmider, M. Mecella, B. Weber, A. Burattin, C. Di Ciccio, G. Fortino, A. Gal, U. Kannengiesser, F. Leotta, et al. (2020). "The internet of things meets business process management: a manifesto". In: *IEEE Systems, Man, and Cybernetics Magazine* 6.4, pp. 34–44. DOI: 10.1109/MSMC.2020.3003135.

Jimenez-Ramirez, A., H. A. Reijers, I. Barba, and C. Del Valle (2019). "A Method to Improve the Early Stages of the Robotic Process Automation Lifecycle". In: *Advanced Information Systems Engineering*. Springer International Publishing, pp. 446–461. DOI: 10.1007/978-3-030-21290-2_28.

Jlailaty, D., D. Grigori, and K. Belhajjame (2017). "Mining Business Process Activities from Email Logs". In: *2017 IEEE International Conference on Cognitive Computing (ICCC)*, pp. 112–119. DOI: 10.1109/IEEE.ICCC.2017.28.

Kecht, C., A. Egger, W. Kratsch, and M. Röglinger (2021). "Event Log Construction from Customer Service Conversations Using Natural Language Inference". In: *2021 3rd International Conference on Process Mining (ICPM)*, pp. 144–151. DOI: 10.1109/ICPM53251.2021.9576869.

Kecht, C., A. Egger, W. Kratsch, and M. Röglinger (2023). "Quantifying chatbots' ability to learn business processes". In: *Information Systems* 113, p. 102176. DOI: 10.1016/j.is.2023.102176.

Kerpedzhiev, G. D., U. M. König, M. Röglinger, and M. Rosemann (2021). "An Exploration into Future Business Process Management Capabilities in View of Digitalization". In: *Business & Information Systems Engineering* 63.2, pp. 83–96. DOI: 10.1007/s12599-020-00637-0.

Khowaja, S. A., B. N. Yahya, and S.-L. Lee (2020). "CAPHAR: context-aware personalized human activity recognition using associative learning in smart environments". In: *Human-centric Computing and Information Sciences* 10.1, p. 35. DOI: 10.1186/s13673-020-00240-y.

Kitchenham, B. A. and S. Charters (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering*. EBSE Technical Report, Keele University and University of Durham.

Knoch, S., N. Herbig, S. Ponpathirkoottam, F. Kosmalla, P. Staudt, P. Fettke, and P. Loos (2019). "Enhancing Process Data in Manual Assembly Workflows". In: *Business Process Management Workshops: BPM 2018 International Workshops, Sydney,* *NSW, Australia*. Springer, Cham, pp. 269–280. DOI: 10.1007/978-3-030-11641-5_21.

Knoch, S., N. Herbig, S. Ponpathirkoottam, F. Kosmalla, P. Staudt, D. Porta, P. Fettke, and P. Loos (2020). "Sensor-based Human–Process Interaction in Discrete Manufacturing". In: *Journal on Data Semantics*. Lecture Notes in Business Information Processing. Ed. by Daniel and Gerstner. DOI: 10.1007/S13740-019-00109-Z.

Knoch, S., S. Ponpathirkoottam, P. Fettke, and P. Loos (2018). "Technology-Enhanced Process Elicitation of Worker Activities in Manufacturing". In: *Business Process Management Workshops*. Ed. by E. Teniente and M. Weidlich. Springer International Publishing, pp. 273–284. DOI: 10.1007/978-3-319-74030-0_20.

Koschmider, A., M. Aleknonytė-Resch, F. Fonger, C. Imenkamp, A. Lepsien, K. Apaydin, M. Harms, D. Janssen, D. Langhammer, T. Ziolkowski, et al. (2023). "Process Mining for Unstructured Data: Challenges and Research Directions". In: *arXiv preprint arXiv:2401.13677*. DOI: 10.48550/arXiv.2401.13677.

Kratsch, W., F. König, and M. Röglinger (2022). "Shedding light on blind spots - Developing a reference architecture to leverage video data for process mining". In: *Decision Support Systems* 158, p. 113794. DOI: 10.1016/J.DSS.2022.113794.

Kratsch, W., J. Manderscheid, M. Röglinger, and J. Seyfried (2021). "Machine Learning in Business Process Monitoring: A Comparison of Deep Learning and Classical Approaches Used for Outcome Prediction". In: *Business & Information Systems Engineering* 63.3, pp. 261–276. DOI: 10.1007/s12599-020-00645-0.

Leemans, S. J. J., D. Fahland, and W. M. P. van der Aalst (2013). "Discovering Block-Structured Process Models from Event Logs – A Constructive Approach". In: *Application and Theory of Petri Nets and Concurrency*. Ed. by J.-M. Colom and J. Desel. Berlin, Heidelberg: Springer, pp. 311–329. DOI: 10.1007/978-3-642-38697-8_17.

Leno, V., M. Dumas, F. M. Maggi, and M. La Rosa (2018). "Multi-perspective process model discovery for robotic process automation". In: *Proceedings of the Doctoral Consortium Papers Presented at the 30th International Conference on Advanced Information Systems Engineering (CAiSE 2018)*. Vol. 2114. CEUR-WS, pp. 37–45.

Leno, V., A. Polyvyanyy, M. Dumas, M. L. Rosa, and F. M. Maggi (2021). "Robotic Process Mining: Vision and Challenges". In: *Business & Information Systems Engineering* 63.3, pp. 301–314. DOI: 10.1007/S12599-020-00641-4.

Leoni, M. de, M. Dees, and L. Reulink (2020). “Design and Evaluation of a Process-aware Recommender System based on Prescriptive Analytics”. In: *2020 2nd International Conference on Process Mining (ICPM)*, pp. 9–16. DOI: 10.1109/ICPM49681.2020.00013.

Leotta, F., M. Mecella, and D. Sora (2020). “Visual process maps: a visualization tool for discovering habits in smart homes”. In: *Journal of Ambient Intelligence and Humanized Computing* 11.5, pp. 1997–2025. DOI: 10.1007/s12652-019-01211-7.

Lepsien, A., A. Koschmider, and W. Kratsch (2023). “Analytics Pipeline for Process Mining on Video Data”. In: *Business Process Management Forum*. Ed. by C. Di Francescomarino, A. Burattin, C. Janiesch, and S. Sadiq. Cham: Springer Nature Switzerland, pp. 196–213. DOI: 10.1007/978-3-031-41623-1_12.

Li, G.-L., S.-Y. Jing, Y. Shen, and B. Guo (2022). “Efficient discrete particle swarm optimization algorithm for process mining from event logs”. In: *International Journal of Computational Intelligence Systems* 15.1, p. 21. DOI: 10.1007/S44196-022-00074-9.

Li, J., H. J. Wang, and X. Bai (2015). “An intelligent approach to data extraction and task identification for process mining”. In: *Information Systems Frontiers* 17.6, pp. 1195–1208. DOI: 10.1007/s10796-015-9564-3.

Li, T. and S. J. van Zelst (2022). “Cache Enhanced Split-Point-Based Alignment Calculation”. In: *2022 4th International Conference on Process Mining (ICPM)*. IEEE, pp. 120–127. DOI: 10.1109/ICPM57379.2022.9980633.

Mamudu, A., W. Bandara, M. T. Wynn, and S. J. Leemans (2022). “A process mining success factors model”. In: *International Conference on Business Process Management*. Springer, pp. 143–160. DOI: 10.1007/978-3-031-16103-2_12.

March, S. T. and G. F. Smith (1995). “Design and natural science research on information technology”. In: *Decision Support Systems* 15.4, pp. 251–266. DOI: 10.1016/0167-9236(94)00041-2.

Martin, N., D. A. Fischer, G. D. Kerpedzhiev, K. Goel, S. J. J. Leemans, M. Röglinger, W. M. P. van der Aalst, M. Dumas, M. La Rosa, and M. T. Wynn (2021). “Opportunities and Challenges for Process Mining in Organizations: Results of a Delphi Study”. In: *Business & Information Systems Engineering* 63.5, pp. 511–527. DOI: 10.1007/s12599-021-00720-0.

Milani, F., K. Lashkevich, F. M. Maggi, and C. Di Francescomarino (2022). “Process mining: a guide for practitioners”. In: *International conference on research challenges* *in information science*. Springer, pp. 265–282. DOI: 10.1007/978-3-031-05760-1_16.

Muñoz-Gama, J. and J. Carmona (2010). “A Fresh Look at Precision in Process Conformance”. In: *Business Process Management*. Ed. by R. Hull, J. Mendling, and S. Tai. Berlin, Heidelberg: Springer, pp. 211–226. DOI: 10.1007/978-3-642-15618-2_16.

Nakayama, Y., M. Mori, Y. Naruse, and H. Morikawa (2018). “The Process Discovery Approaches for Decision Making in Sales Activities”. In: *2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS)*, pp. 1394–1399. DOI: 10.1109/SCIS-ISIS.2018.00217.

OpenAI (2022). *ChatGPT: Optimizing language models for dialogue*. Ed. by OpenAI. https://openai.com/blog/chatgpt/. Accessed: 2023-01-03.

Pallotta, V. and R. Delmonte (2013). “Interaction Mining: The New Frontier of Customer Interaction Analytics”. In: *New Challenges in Distributed Information Filtering and Retrieval: DART 2011: Revised and Invited Papers*. Ed. by C. Lai, G. Semeraro, and E. Vargiu. Berlin, Heidelberg: Springer, pp. 91–111. DOI: 10.1007/978-3-642-31546-6_6.

Peffers, K., T. Tuunanen, M. A. Rothenberger, and S. Chatterjee (2007). “A Design Science Research Methodology for Information Systems Research”. In: *Journal of Management Information Systems* 24.3, pp. 45–77. DOI: 10.2753/MIS0742-1222240302.

Poddar, A., N. Donthu, and Y. Wei (2009). “Web site customer orientations, Web site quality, and purchase intentions: The role of Web site personality”. In: *Journal of Business Research* 62.4, pp. 441–450. DOI: https://doi.org/10.1016/j.jbusres.2008.01.036.

Reinkemeyer, L. (2020a). “How to Get Started”. In: *Process mining in action*. Cham: Springer Nature, pp. 11–14. DOI: 10.1007/978-3-030-40172-6_2.

Reinkemeyer, L. (2020b). *Process Mining in Action: Principles, Use Cases and Outlook*. Cham: Springer Cham. DOI: 10.1007/978-3-030-40172-6.

Rogge-Solti, A., A. Senderovich, M. Weidlich, J. Mendling, and A. Gal (2016). “In Log and Model We Trust? A Generalized Conformance Checking Framework”. In: *Business Process Management*. Cham: Springer International Publishing, pp. 179–196. DOI: 10.1007/978-3-319-45348-4_11.

Ronzani, M., R. Ferrod, C. Di Francescomarino, E. Sulis, R. Aringhieri, G. Boella, E. Brunetti, L. Di Caro, M. Dragoni, C. Ghidini, and R. Marinello (2022). "Unstructured Data in Predictive Process Monitoring: Lexicographic and Semantic Mapping to ICD-9-CM Codes for the Home Hospitalization Service". In: *AIxIA 2021 – Advances in Artificial Intelligence*. Ed. by S. Bandini, F. Gasparini, V. Mascardi, M. Palmonari, and G. Vizzari. Cham: Springer International Publishing, pp. 700–715. DOI: 10.1007/978-3-031-08421-8_48.

Rozinat, A. and W. M. P. van der Aalst (2008). "Conformance checking of processes based on monitoring real behavior". In: *Information Systems* 33.1, pp. 64–95. DOI: 10.1016/j.is.2007.07.001.

Ruiz, R. C., A. Jiménez-Ramírez, M. J. E. Cuaresma, and J. G. Enríquez (2022). "Hybridizing humans and robots: An RPA horizon envisaged from the trenches". In: *Computers in Industry* 138, p. 103615. DOI: 10.1016/J.COMPIND.2022.103615.

Silva, N., P. Sousa, and M. Mira da Silva (2021). "Maintenance of Enterprise Architecture Models: A Systematic Review of the Scientific Literature". In: *Business & Information Systems Engineering* 63, pp. 157–180. DOI: 10.1007/s12599-020-00636-1.

Simon, J. P. (2019). "Artificial intelligence: scope, players, markets and geography". In: *Digital Policy, Regulation and Governance* 21.3, pp. 208–237. DOI: 10.1108/DPRG-08-2018-0039.

Sonnenberg, C. and J. vom Brocke (2012a). "Evaluation Patterns for Design Science Research Artefacts". In: *Practical Aspects of Design Science: European Design Science Symposium*. Springer, Berlin, Heidelberg, pp. 71–83. DOI: 10.1007/978-3-642-33681-2_7.

Sonnenberg, C. and J. vom Brocke (2012b). "Evaluations in the Science of the Artificial - Reconsidering the Build-Evaluate Pattern in Design Science Research". In: *Design Science Research in Information Systems. Advances in Theory and Practice - 7th International Conference, DESRIST 2012, Las Vegas, NV, USA, May 14-15, 2012. Proceedings*. Vol. 7286. Lecture Notes in Computer Science. Springer, pp. 381–397. DOI: 10.1007/978-3-642-29863-9_28.

Sutskever, I., O. Vinyals, and Q. V. Le (2014). "Sequence to Sequence Learning with Neural Networks". In: *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*. NIPS'14. Montreal, Canada, pp. 3104–3112.

Tang, J., L. Li, Y. Liu, and K.-Y. Lin (2021). “Automatic Identification of Bottleneck Tasks for Business Process Management using Fusion-based Text Clustering”. In: *IFAC-PapersOnLine* 54.1, pp. 1200–1205. ISSN: 2405-8963. DOI: 10.1016/j.ifacol.2021.08.142.

Teinemaa, I., M. Dumas, F. M. Maggi, and C. Di Francescomarino (2016). “Predictive Business Process Monitoring with Structured and Unstructured Data”. In: *Business Process Management*. Cham: Springer International Publishing, pp. 401–417. DOI: 10.1007/978-3-319-45348-4_23.

Thought Vector (2017). *Customer Support on Twitter*. https://www.kaggle.com/thoughtvector/customer-support-on-twitter. Accessed: 2021-04-03.

UiPath (2024). *Evolve your processes, discover better outcomes*. https://www.uipath.com/product/process-mining. Accessed: 2024-09-08.

Ukpabi, D. C., B. Aslam, and H. Karjaluoto (2019). “Chatbot Adoption in Tourism Services: A Conceptual Exploration”. In: *Robots, Artificial Intelligence, and Service Automation in Travel, Tourism and Hospitality*. Ed. by S. Ivanov and C. Webster. Emerald Publishing Limited, pp. 105–121. DOI: 10.1108/978-1-78756-687-320191006.

van der Aalst, W., T. Weijters, and L. Maruster (2004). “Workflow mining: discovering process models from event logs”. In: *IEEE Transactions on Knowledge and Data Engineering* 16.9, pp. 1128–1142. DOI: 10.1109/TKDE.2004.47.

van der Aalst, W. (2016a). “Data Science in Action”. In: *Process Mining: Data Science in Action*. Springer, Berlin, Heidelberg, pp. 3–23. DOI: 10.1007/978-3-662-49851-4_1.

van der Aalst, W. (2016b). “Getting the Data”. In: *Process Mining: Data Science in Action*. Springer Berlin Heidelberg, pp. 125–162. DOI: 10.1007/978-3-662-49851-4_5.

van der Aalst, W. (2016c). “Mining Additional Perspectives”. In: *Process Mining: Data Science in Action*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 275–300. DOI: 10.1007/978-3-662-49851-4_9.

van der Aalst, W. (2016d). “Operational Support”. In: *Process Mining: Data Science in Action*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 301–321. DOI: 10.1007/978-3-662-49851-4_10.

van der Aalst, W. M. P. (2019). “Object-Centric Process Mining: Dealing with Divergence and Convergence in Event Data”. In: *Software Engineering and Formal Methods - 17th International Conference, SEFM 2019, Oslo, Norway, September 18-20, 2019, Proceedings*. Vol. 11724. Lecture Notes in Computer Science. Springer, pp. 3–25. DOI: 10.1007/978-3-030-30446-1_1.

van der Aalst, W. M. P. (2020). “Academic View: Development of the Process Mining Discipline”. In: *Process Mining in Action: Principles, Use Cases and Outlook*. Ed. by L. Reinkemeyer. Cham: Springer Cham, pp. 181–196. DOI: 10.1007/978-3-030-40172-6_21.

van der Aalst, W. M. P. (2022). “Process Mining: A 360 Degree Overview”. In: *Process Mining Handbook*. Ed. by W. M. P. van der Aalst and J. Carmona. Cham: Springer International Publishing, pp. 3–34. DOI: 10.1007/978-3-031-08848-3_1.

van der Aalst, W. M. P., A. Adriansyah, A. K. A. de Medeiros, F. Arcieri, T. Baier, and et al. (2012). “Process Mining Manifesto”. In: *Business Process Management Workshops*. Ed. by F. Daniel, K. Barkaoui, and S. Dustdar. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 169–194. DOI: 10.1007/978-3-642-28108-2_19.

van der Aalst, W. M. P., M. Bichler, and A. Heinzl (2018). “Robotic Process Automation”. In: *Business & Information Systems Engineering* 60.4, pp. 269–272. DOI: 10.1007/S12599-018-0542-4.

van der Aalst, W. M. P., H. A. Reijers, A. J. M. M. Weijters, B. F. van Dongen, A. K. A. de Medeiros, M. Song, and H. M. W. Verbeek (2007). “Business process mining: An industrial application”. In: *Information Systems* 32.5, pp. 713–732. DOI: 10.1016/J.IS.2006.05.003.

van Dongen, B. and F. F. Borchert (2018). *BPI Challenge 2018*. DOI: 10.4121/uuid:3301445f-95e8-4ff0-98a4-901f1f204972.

van Eck, M. L., N. Sidorova, and W. van der Aalst (2016). “Enabling process mining on sensor data from smart products”. In: *2016 IEEE Tenth International Conference on Research Challenges in Information Science (RCIS)*, pp. 1–12. DOI: 10.1109/RCIS.2016.7549355.

Venable, J. R., J. Pries-Heje, and R. L. Baskerville (2016). “FEDS: a Framework for Evaluation in Design Science Research Open”. In: *European Journal of Information Systems* 25.1, pp. 77–89. DOI: 10.1057/EJIS.2014.36.

Weerdt, J. de, S. K. vanden Broucke, J. Vanthienen, and B. Baesens (2012). “Leveraging process discovery with trace clustering and text mining for intelligent analysis of incident management processes". In: *2012 IEEE Congress on Evolutionary Computation*, pp. 1–8. DOI: 10.1109/CEC.2012.6256459.

Weijters, A. J., W. M. van der Aalst, and A. A. De Medeiros (2006). *Process mining with the HeuristicsMiner algorithm*. https://pure.tue.nl/ws/portalfiles/portal/2388011/615595.pdf.

Wewerka, J., C. Micus, and M. Reichert (2021). "Seven Guidelines for Designing the User Interface in Robotic Process Automation". In: *25th International Enterprise Distributed Object Computing Workshop, EDOC Workshop 2021, Gold Coast, Australia, October 25-29, 2021*. IEEE, pp. 157–165. DOI: 10.1109/EDOCW52865.2021.00045.

Wewerka, J. and M. Reichert (2020). "Towards Quantifying the Effects of Robotic Process Automation". In: *24th IEEE International Enterprise Distributed Object Computing Workshop, EDOC Workshops 2020, Eindhoven, The Netherlands, October 5, 2020*. IEEE, pp. 11–19. DOI: 10.1109/EDOCW49879.2020.00015.

Wu, Y., J. Lim, and M. Yang (2013). "Online Object Tracking: A Benchmark". In: *2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, June 23-28, 2013*. IEEE Computer Society, pp. 2411–2418. DOI: 10.1109/CVPR.2013.312.

Zhou, Q., W. Suraworachet, and M. Cukurova (2024). "Detecting non-verbal speech and gaze behaviours with multimodal data and computer vision to interpret effective collaborative learning interactions". In: *Education and Information Technologies* 29, pp. 1071–1098. DOI: 10.1007/s10639-023-12315-1.

Zhu, R., Y. Dai, T. Li, Z. Ma, M. Zheng, Y. Tang, J. Yuan, and Y. Huang (2019). "Automatic Real-Time Mining Software Process Activities From SVN Logs Using a Naive Bayes Classifier". In: *IEEE Access* 7, pp. 146403–146415. DOI: 10.1109/ACCESS.2019.2945608.

VII Appendix

VII.1 Index of Research Articles

Research Article 1: Unstructured Data in Process Mining: A Systematic Literature Review

König, F.; Egger, A.; Kratsch, W.; Röglinger, M.; Wördehoff, N. (2024). Unstructured Data in Process Mining: A Systematic Literature Review. Submitted to: *ACM Transactions on Management Information Systems* (in first round of revision).
(VHB-JQ3 [5]: B, VHB-PMR [6]: B, SJR [7]: Q1, IF [8]: 2.5)

Research Article 2: Refining the Process Picture: Unstructured Data in Object-Centric Process Mining

Egger, A.; Fehrer, T.; Kratsch, W.; Wördehoff, N.; König, F.; Röglinger, M. (2024). Refining the Process Picture: Unstructured Data in Object-Centric Process Mining. Submitted to: *Information Systems*.
(VHB-JQ3: B, VHB-PMR: B, SJR: Q1, IF: 3.0)

Research Article 3: Beyond Assumptions: A Reference Architecture to Enable Unsupervised Process Discovery from Video Data

Wördehoff, N.; Egger, A.; Kratsch, W.; König, F.; Röglinger, M.; (2024). Beyond Assumptions: A Reference Architecture to Enable Unsupervised Process Discovery from Video Data. Submitted to: *Outlet hidden due to the double-blind review process of the journal* (in first round of revision).

[5] VHB-JQ3: VHB-Jourqual 3
[6] VHB-PMR: VHB-Publikationsmedienrating
[7] SJR: Scimago Journal & Country Rank
[8] IF: Impact Factor

Research Article 4: Bot Log Mining: An Approach to the Integrated Analysis of Robotic Process Automation and Process Mining

Egger, A.; ter Hofstede, A.H.M.; Kratsch, W.; Leemans, S.J.J.; Röglinger, M.; Wynn, M.T. (2024). Bot Log Mining: An Approach to the Integrated Analysis of Robotic Process Automation and Process Mining. In: *Information Systems*.
DOI: 10.1016/j.is.2024.102431.
(VHB-JQ3: B, VHB-PMR: B, SJR: Q1, IF: 3.0)
Earlier version published as a short paper in *ER 2020 – International Conference on Conceptual Modeling*, 2020. DOI: 10.1007/978-3-030-62522-1_4.

Research Article 5: Quantifying Chatbots' Ability to Learn Business Processes

Kecht, C.; Egger, A.; Kratsch, W.; Röglinger, M. (2023). Quantifying Chatbots' Ability to Learn Business Processes. In: *Information Systems*.
DOI: 10.1016/j.is.2023.102176.
(VHB-JQ3: B, VHB-PMR: B, SJR: Q1, IF: 3.0)
Earlier version published in *3rd International Conference on Process Mining (ICPM)*, 2021. DOI: 10.1109/ICPM53251.2021.9576869.

Over the course of the dissertation, I also co-authored the following research articles and book chapter. These are not part of this dissertation.

Dreyer, S.; Egger, A.; Püschel, L.; Röglinger, M. (2020). Prioritising smart factory investments – A project portfolio selection approach. In: *International Journal of Production Research*. DOI: 10.1080/00207543.2020.1849845.

Egger, A.; Püschel, L.; Röglinger, M. (2021). Ausgangslage von und ökonomische Erwartungen an das Internet der Dinge. In: *Beck Verlag*.

Chvirova, D.; Egger, A.; Fehrer, T.; Kratsch, W.; Röglinger, M.; Wittmann, J.; Wördehoff, N. (2024). A multi-media dataset for object-centric business process mining in IT asset management. In: *Data in Brief*. DOI: 10.1016/j.dib.2024.110716.

Kratsch, W.; Stengel, G.; Egger, A.; Fehrer, T. (2024). Unstrukturierte Daten aus der Serienproduktion mit Process Mining zur Fehleranalyse und Prozessoptimierung nutzen. In: *VDI Mechatroniktagung 2024*. URL: https://eref.uni-bayreuth.de/id/eprint/88991/.

VII.2 Individual Contribution to the Research Articles

This dissertation is cumulative, comprising five research articles, each developed in teams with multiple co-authors. This section describes the research settings and my contribution to each of the five research articles.

Research Article 1, titled "*Unstructured Data in Process Mining: A Systematic Literature Review*" (König et al. 2024; Section VII.3), was written by a team of five authors. I contributed to the conceptualization and the formulation of the research goal. Additionally, I was responsible for drafting parts of the initial manuscript and had a key role in reviewing and editing the entire manuscript. One co-author acted as the lead author, while the other co-authors and I acted as subordinate authors.

Research Article 2, titled "*Refining the Process Picture: Unstructured Data in Object-Centric Process Mining*" (Egger et al. 2024; Section VII.4), was written by a team of six authors. As the first author, I held a crucial role in all parts and administered the research. I contributed significantly to the design of the research methodology, conceptualization, investigation, and data curation. Furthermore, I was solely responsible for writing the original draft of the manuscript and was involved in reviewing and editing. I led the development of the software prototype and was responsible for data analysis and evaluation of the approach. I acted as the lead author, while the other five co-authors acted as subordinate authors.

Research Article 3, titled "*Beyond Assumptions: A Reference Architecture to Enable Unsupervised Process Discovery from Video Data*" (Wördehoff et al. 2024; Section VII.5), was written by a team of five authors. I had a key role in most parts of the research project. I contributed to the formulation of the overarching research objectives and the research question, as well as to the design of the research methodology. In terms of writing, I was involved in drafting parts of the initial manuscript and contributed significantly to reviewing and editing the entire manuscript. All authors contributed equally to the research article.

Research Article 4, titled "*Bot Log Mining: An Approach to the Integrated Analysis of Robotic Process Automation and Process Mining*" (Egger et al. 2024; Section VII.6), was written by a team of six authors. I contributed significantly to all parts and administered the research project. Specifically, I was involved in the conceptualization of overall research goals, as well as the design of the research methodology. Additionally, I contributed to software development, validation, data curation, and formal analysis. I led the writing of the original manuscript draft and was involved in reviewing and editing the entire paper, as well as in revising the manuscript for re-submission. As a team, we agreed that I held a crucial role in contributing to this research article, while the other co-authors contributed equally.

Research Article 5, titled "*Quantifying Chatbots' Ability to Learn Business Processes*" (Kecht et al. 2023; Section VII.7), was written by a team of four authors. I was involved in the conceptualization of the research, contributing to the formulation of the research question and the overall research objectives. In writing the research article, I was responsible for drafting several parts and had a key role in reviewing the entire manuscript. Additionally, I contributed significantly to the revision of the manuscript for re-submission. One co-author acted as a subordinate author, while the other two co-authors and I contributed equally to the research article.

VII.3 Research Article 1: Unstructured Data in Process Mining: A Systematic Literature Review

Authors:
Fabian König, Andreas Egger, Wolfgang Kratsch, Maximilian Röglinger, Niklas Wördehoff

Submitted to:
ACM Transactions on Management Information Systems (in first round of revision)

Extended Abstract:
The majority of available data is in unstructured form (Davis, 2019; Gandomi and Haider, 2015) and the proportion is expected to increase (Balducci and Marinova, 2018). In process mining, studies have traditionally focused on structured data sources (e.g., relational databases) (Diba et al., 2020), but as data analysis methods continue to improve, solutions that integrate unstructured data into process mining are receiving increasing attention. This includes approaches that use text (Teinemaa et al., 2016), sensor (Leotta et al., 2020), and image and video data (Knoch et al., 2018). The integration of unstructured data into process mining holds significant potential to generate valuable context-related insights (Beverungen et al., 2021) and facilitate a more comprehensive representation and analysis of real-world processes (Grisold et al., 2021). However, systematic progress in process mining research in this area is currently hampered by the lack of an overview of the use of unstructured data. This lack of insight also makes it difficult to identify the most promising avenues for advancing process mining research. Therefore, the research article poses the following three research questions:

- *Which types of unstructured data are used in process mining?*
- *How are unstructured data leveraged for different process mining use cases?*
- *What are the open challenges and areas for improvement?*

To answer these research questions, a systematic literature review (Kitchenham and Charters, 2007) is conducted. The focus is on technical process mining artifacts that enable (semi-)automated consideration of unstructured data. A generic search phrase was derived and multiple study selection criteria were employed. A total of 1,379 research items were extracted from seven information systems and computer science databases. Following a rigorous selection process, 24 primary studies were identified for analysis in which two researchers independently applied a deductive coding scheme. The analysis shows

that current research is primarily concerned with textual data and concentrates on extracting event logs for process discovery. To guide future process mining research, a research agenda is proposed that includes seven opportunities to address the identified research gaps. Since the umbrella term unstructured data is used ambiguously in the literature, the research article also proposes a generic conceptualization and definition for unstructured data by summarizing and consolidating existing definitions. The research article provides new insights to systematically advance research at the intersection of unstructured data and process mining. Given the potential of unstructured data to fundamentally improve business process management, many process mining vendors and users will need to develop solutions for managing unstructured data to remain competitive. Therefore, the results and the proposed research agenda are of interest to academics and practitioners alike.

Keywords:
Unstructured data, Process mining, Systematic literature review

References

Balducci, B. and D. Marinova (2018). "Unstructured data in marketing". In: *Journal of the Academy of Marketing Science* 46, pp. 557–590. DOI: 10.1007/s11747-018-0581-x.

Beverungen, D., J. C. A. M. Buijs, J. Becker, C. Di Ciccio, W. M. P. van der Aalst, and et al. (2021). "Seven Paradoxes of Business Process Management in a Hyper-Connected World". In: *Business & Information Systems Engineering* 63.2, pp. 145–156. DOI: 10.1007/s12599-020-00646-z.

Davis, D. (2019). *AI Unleashes the Power of Unstructured Data*. https://www.cio.com/article/3406806/ai-unleashes-the-power-of-unstructured-data.html. Accessed: 2023-11-14.

Diba, K., K. Batoulis, M. Weidlich, and M. Weske (2020). "Extraction, correlation, and abstraction of event data for process mining". In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 10.3. DOI: 10.1002/WIDM.1346.

Gandomi, A. and M. Haider (2015). "Beyond the hype: Big data concepts, methods, and analytics". In: *International Journal of Information Management* 35.2, pp. 137–144. ISSN: 0268-4012. DOI: 10.1016/j.ijinfomgt.2014.10.007.

Grisold, T., J. Mendling, M. Otto, and J. vom Brocke (2021). "Adoption, use and management of process mining in practice". In: *Business Process Management Journal* 27.2, pp. 369–387. DOI: 10.1108/BPMJ-03-2020-0112.

Kitchenham, B. A. and S. Charters (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering*. EBSE Technical Report, Keele University and University of Durham.

Knoch, S., S. Ponpathirkoottam, P. Fettke, and P. Loos (2018). "Technology-Enhanced Process Elicitation of Worker Activities in Manufacturing". In: *Business Process Management Workshops*. Ed. by E. Teniente and M. Weidlich. Springer International Publishing, pp. 273–284. DOI: 10.1007/978-3-319-74030-0_20.

Leotta, F., M. Mecella, and D. Sora (2020). "Visual process maps: a visualization tool for discovering habits in smart homes". In: *Journal of Ambient Intelligence and Humanized Computing* 11.5, pp. 1997–2025. DOI: 10.1007/s12652-019-01211-7.

Teinemaa, I., M. Dumas, F. M. Maggi, and C. Di Francescomarino (2016). "Predictive Business Process Monitoring with Structured and Unstructured Data". In: *Business Process Management*. Cham: Springer International Publishing, pp. 401–417. DOI: 10.1007/978-3-319-45348-4_23.

VII.4 Research Article 2: Refining the Process Picture: Unstructured Data in Object-Centric Process Mining

Authors:
Andreas Egger, Tobias Fehrer, Wolfgang Kratsch, Niklas Wördehoff, Fabian König, Maximilian Röglinger

Submitted to:
Information Systems

Extended Abstract:
Process mining aims to discover, monitor, and improve processes (van der Aalst et al., 2012). To this end, process mining techniques use event data, typically extracted from information systems and organized along process instances (Diba et al., 2020; van der Aalst, 2016; van der Aalst et al., 2012). The inherent complexity of real-world processes has driven the recent introduction of object-centric process mining (van der Aalst, 2019; Esser and Fahland, 2021) that encompasses objects' interaction with one another. Thereby, events can be related to one or many objects, allowing for a more comprehensive view of processes. Another avenue of research contributing to more complete process analyses is integrating unstructured data, which can enhance traditional event logs by extracting hitherto unidentified process information (Beverungen et al., 2021; Koschmider et al., 2023; Reinkemeyer, 2020; van der Aalst, 2016). Object-centric process mining and the use of unstructured data in process mining both aim to provide a more holistic picture of processes. Although combining the object-centric perspective with event log enrichment from unstructured data sources holds promising potential, such investigation remains in its infancy. Thus, this research article investigates the following research question: *How can unstructured data be combined with structured event logs for object-centric process mining?*

To answer this research question, the OCRAUD is developed, a reference architecture that provides guidance on using unstructured data sources and traditional event logs for object-centric process mining. The reference architecture includes three main subsystems and serves as a blueprint that can be instantiated for different use cases. Several optional components and modular unstructured data processor subsystems for different data sources allow for different instantiation variants. Design science research principles are used to design (Gregor and Hevner, 2013; Peffers et al., 2007) and evaluate (Venable et al., 2016)

the artifact. Based on the literature on object-centric process mining and unstructured data, four design objectives are derived to guide the design of the artifact. The evaluation episodes involve conducting a total of 20 interviews with experts from industry and research over two rounds. Moreover, the OCRAUD is compared to competing artifacts to show that it is new and addresses the identified research gap. Furthermore, the OVaSA is presented as a specific instantiation of the OCRAUD that is tailored to the use of video and sensor data. On this basis, a software prototype is developed and applied to real-world data, and the results are analyzed. The research article contributes to the field of process mining by guiding the combination of unstructured data sources with traditional event logs, incorporating an object-centric representation of event data. The instantiation targets video and sensor data, thereby demonstrating the use of the artifact. This enables researchers and practitioners to instantiate the reference architecture for other data types or specific use cases. The published code of the software prototype allows for further development of the implemented algorithms.

Keywords:
Object-centric process mining, Unstructured data, Reference architecture, Design science research, Business process management

References

Beverungen, D., J. C. A. M. Buijs, J. Becker, C. Di Ciccio, W. M. P. van der Aalst, and et al. (2021). "Seven Paradoxes of Business Process Management in a Hyper-Connected World". In: *Business & Information Systems Engineering* 63.2, pp. 145–156. DOI: 10.1007/s12599-020-00646-z.

Diba, K., K. Batoulis, M. Weidlich, and M. Weske (2020). "Extraction, correlation, and abstraction of event data for process mining". In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 10.3. DOI: 10.1002/WIDM.1346.

Esser, S. and D. Fahland (2021). "Multi-Dimensional Event Data in Graph Databases". In: *Journal on Data Semantics* 10.1-2, pp. 109–141. DOI: 10.1007/S13740-021-00122-1.

Gregor, S. and A. R. Hevner (2013). "Positioning and Presenting Design Science Research for Maximum Impact". In: *MIS quarterly* 37.2, pp. 337–355. DOI: 10.25300/MISQ/2013/37.2.01.

Koschmider, A., M. Aleknonytė-Resch, F. Fonger, C. Imenkamp, A. Lepsien, K. Apaydin, M. Harms, D. Janssen, D. Langhammer, T. Ziolkowski, et al. (2023). "Process Mining for Unstructured Data: Challenges and Research Directions". In: *arXiv preprint arXiv:2401.13677*. DOI: 10.48550/arXiv.2401.13677.

Peffers, K., T. Tuunanen, M. A. Rothenberger, and S. Chatterjee (2007). "A Design Science Research Methodology for Information Systems Research". In: *Journal of Management Information Systems* 24.3, pp. 45–77. DOI: 10.2753/MIS0742-1222240302.

Reinkemeyer, L. (2020). "How to Get Started". In: *Process mining in action*. Cham: Springer Nature, pp. 11–14. DOI: 10.1007/978-3-030-40172-6_2.

van der Aalst, W. (2016). "Getting the Data". In: *Process Mining: Data Science in Action*. Springer Berlin Heidelberg, pp. 125–162. DOI: 10.1007/978-3-662-49851-4_5.

van der Aalst, W. M. P. (2019). "Object-Centric Process Mining: Dealing with Divergence and Convergence in Event Data". In: *Software Engineering and Formal Methods - 17th International Conference, SEFM 2019, Oslo, Norway, September 18-20, 2019, Proceedings*. Vol. 11724. Lecture Notes in Computer Science. Springer, pp. 3–25. DOI: 10.1007/978-3-030-30446-1_1.

van der Aalst, W. M. P., A. Adriansyah, A. K. A. de Medeiros, F. Arcieri, T. Baier, and et al. (2012). "Process Mining Manifesto". In: *Business Process Management Workshops*. Ed. by F. Daniel, K. Barkaoui, and S. Dustdar. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 169–194. DOI: 10.1007/978-3-642-28108-2_19.

Venable, J. R., J. Pries-Heje, and R. L. Baskerville (2016). "FEDS: a Framework for Evaluation in Design Science Research Open". In: *European Journal of Information Systems* 25.1, pp. 77–89. DOI: 10.1057/EJIS.2014.36.

VII.5 Research Article 3: Beyond Assumptions: A Reference Architecture to Enable Unsupervised Process Discovery from Video Data

Authors:
Niklas Wördehoff, Andreas Egger, Wolfgang Kratsch, Fabian König, Maximilian Röglinger

Submitted to:
Outlet hidden due to the double-blind review process of the journal (in first round of revision)

Extended Abstract:
Process mining has developed into one of the most important research streams in business process management (van der Aalst, 2020; van der Aalst, 2019). Despite its successful application to improve process performance in industry, there are still various potentials to be realized in the coming years. One of them is the use of unstructured data, which can contain additional contextual process information (Beverungen et al., 2021). In particular, the use of unstructured video data has the potential to facilitate the analysis of previously non-observable parts of processes (Kratsch et al., 2022; Knoch et al., 2020). Existing approaches dealing with video data in process mining derive event logs from video data by extracting a pre-defined set of potentially relevant activities. As this set is typically determined using a process model or input from process experts rather than the available video data, current solutions are unable to identify activities that extend beyond the presumed process behavior, which would be essential to ensure transparency in process analysis. Hence, this research article deals with the following research question: *How can actual process activities be derived from video data as part of process discovery?*

The research question is addressed by developing the RAVEE, a reference architecture that supports the extraction of actual process information from unstructured video data. The RAVEE facilitates the unsupervised extraction of process steps by integrating various computer vision and clustering capabilities. Hence, the reference architecture enables exploratory process analysis in a wide range of process mining application areas by extracting the actual process activities based solely on the available video data. To achieve this objective and to structure the research, design science research principles are followed (Gregor and Hevner, 2013; Peffers et al., 2007). A semi-structured litera-

ture search first enables gaining a comprehensive understanding of the current related approaches. Building on the insights derived from literature, design objectives are derived that outline the requirements for the RAVEE. Several evaluation activities are performed to ensure the completeness and applicability of the reference architecture (Sonnenberg and vom Brocke, 2012). This includes conducting semi-structured interviews to ensure that the design objectives and the RAVEE address the identified research problem. Furthermore, a prototypical instantiation of the RAVEE demonstrates its ability to extract process-relevant activities from video data on a challenging dataset. The results extend existing knowledge by providing an initial basis for unsupervised exploration of video data for process mining purposes. In the future, this can enable evidence-based decision support using video data without having to rely on assumed activities and, therefore, provide holistic insights into underlying processes.

Keywords:
Process mining, Video data, Unstructured data, Computer vision, Unsupervised learning, Reference architecture

References

Beverungen, D., J. C. A. M. Buijs, J. Becker, C. Di Ciccio, W. M. P. van der Aalst, and et al. (2021). "Seven Paradoxes of Business Process Management in a Hyper-Connected World". In: *Business & Information Systems Engineering* 63.2, pp. 145–156. DOI: 10.1007/s12599-020-00646-z.

Gregor, S. and A. R. Hevner (2013). "Positioning and Presenting Design Science Research for Maximum Impact". In: *MIS quarterly* 37.2, pp. 337–355. DOI: 10.25300/MISQ/2013/37.2.01.

Knoch, S., N. Herbig, S. Ponpathirkoottam, F. Kosmalla, P. Staudt, D. Porta, P. Fettke, and P. Loos (2020). "Sensor-based Human–Process Interaction in Discrete Manufacturing". In: *Journal on Data Semantics*. Lecture Notes in Business Information Processing. Ed. by Daniel and Gerstner. DOI: 10.1007/S13740-019-00109-Z.

Kratsch, W., F. König, and M. Röglinger (2022). "Shedding light on blind spots - Developing a reference architecture to leverage video data for process mining". In: *Decision Support Systems* 158, p. 113794. DOI: 10.1016/J.DSS.2022.113794.

Peffers, K., T. Tuunanen, M. A. Rothenberger, and S. Chatterjee (2007). "A Design Science Research Methodology for Information Systems Research". In: *Journal of Man-*

agement Information Systems 24.3, pp. 45–77. DOI: 10.2753/MIS0742-1222240302.

Sonnenberg, C. and J. vom Brocke (2012). “Evaluation Patterns for Design Science Research Artefacts”. In: *Practical Aspects of Design Science: European Design Science Symposium*. Springer, Berlin, Heidelberg, pp. 71–83. DOI: 10.1007/978-3-642-33681-2_7.

van der Aalst, W. (2019). *ICPM Keynote on 20 Years of Process Mining Research: Accompishments, Challenges, and Open Problems*. Ed. by ICPM 2019. DOI: 10.1109/ICPM.2019.00011.

van der Aalst, W. M. P. (2020). “Academic View: Development of the Process Mining Discipline”. In: *Process Mining in Action: Principles, Use Cases and Outlook*. Ed. by L. Reinkemeyer. Cham: Springer Cham, pp. 181–196. DOI: 10.1007/978-3-030-40172-6_21.

VII.6 Research Article 4: Bot Log Mining: An Approach to the Integrated Analysis of Robotic Process Automation and Process Mining

Authors:
Andreas Egger, Arthur H.M. ter Hofstede, Wolfgang Kratsch, Sander J.J. Leemans, Maximilian Röglinger, Moe T. Wynn

Published in:
Information Systems 126, 102431 (2024). DOI: 10.1016/j.is.2024.102431.

Abstract:
Process mining and Robotic Process Automation (RPA) are two technologies of great interest in research and practice. Process mining uses event logs as input, but much of the information available about processes is not yet considered since the data is outside the scope of ordinary event logs. RPA technology can automate tasks by using bots, and the executed steps can be recorded, which could be a valuable data source for process mining. With the use of RPA technology expected to grow, an integrated view of steps performed by bots in business processes is needed. In process mining, various techniques to analyze processes have already been developed. Most RPA software also includes basic measures to monitor bot performance. However, the isolated use of bot-related or process mining measures does not provide an end-to-end view of bot-enabled business processes. To address these issues, we develop an approach that enables using RPA logs for process mining and propose tailored measures to analyze merged bot and process logs. We use the design science research process to structure our work and evaluate the approach by conducting a total of 14 interviews with experts from industry and research. We also implement a software prototype and test it on real-world and artificial data. This approach contributes to prescriptive knowledge by providing a concept on how to use bot logs for process mining and brings the research streams of RPA and process mining further together. It provides new data that expands the possibilities of existing process mining techniques in research and practice, and it enables new analyses that can observe bot-human interaction and show the effects of bots on business processes.

Keywords:
Robotic Process Automation, Process mining, Business Process Management, Design Science Research

VII.7 Research Article 5: Quantifying Chatbots' Ability to Learn Business Processes

Authors:

Christoph Kecht, Andreas Egger, Wolfgang Kratsch, Maximilian Röglinger

Published in:

Information Systems 113, 102176 (2023). DOI: 10.1016/j.is.2023.102176.

Abstract:

Chatbots enable organizations in the business-to-customer domain to respond to repetitive requests efficiently. Extant approaches in Natural Language Processing (NLP) already address the essential requirement of understanding user input and synthesizing a response as close as possible to a response a human interlocutor would give. However, we argue that the organizational adoption of chatbots further depends on the underlying model's capability to learn and comply with organizations' business processes, for example, authenticating a customer before providing sensitive details. To address this issue, we develop an approach that quantifies chatbots' ability to learn business processes using standardized process mining metrics. We demonstrate our approach by training chatbots on a dataset of more than 500,000 customer service conversations from three companies on Twitter and show how our approach supports the quantification of a chatbot's overall ability to learn business processes from the training data. Furthermore, we quantify a chatbot's ability to learn a particular variant of the underlying process and we show how to compare the chatbot's executed steps against a given normative process model. Our approach that seamlessly integrates with existing approaches to evaluate NLP-based chatbots mitigates the current hurdles that practitioners face and, therefore, strives to foster the adoption of chatbots in practice.

Keywords:

Chatbots, Process mining, Natural language processing, Conformance checking