Conversational Systems for AI-Augmented Business Process Management

Angelo Casciani[1], Mario L. Bernardi[2], Marta Cimitile[3], and Andrea Marrella[1]

[1] Sapienza University of Rome, Rome, Italy
{casciani,marrella}@diag.uniroma1.it
[2] University of Sannio, Benevento, Italy
bernardi@unisannio.it
[3] Unitelma Sapienza, Rome, Italy
marta.cimitile@unitelma.it

Abstract. AI-augmented Business Process Management Systems (ABPMSs) are an emerging class of process-aware information systems empowered by AI technology for autonomously unfolding and adapting the execution flow of business processes (BPs). A central characteristic of an ABPMS is the ability to be *conversationally actionable*, i.e., to proactively interact with human users about BP-related actions, goals, and intentions. While today's trend is to support BP automation using reactive conversational agents, an ABPMS is required to create dynamic conversations that not only respond to user queries but even initiate conversations with users to inform them of the BP progression and make recommendations to improve BP performance. In this paper, we explore the extent to which state-of-the-art conversational systems (CSs) can be used to develop such proactive conversation features, and we discuss the research challenges and opportunities within this area.

Keywords: AI-augmented Business Process Management · Conversational Systems · Large Language Models · Process Mining

1 Introduction

In the era of Industry 4.0 (I4.0), the increased availability of event data tracing the execution of Business Processes (BPs), combined with advances in Artificial Intelligence (AI), is laying the ground for a new breed of AI-augmented BPM Systems (ABPMSs), capable of autonomously unfolding and adapting the BP execution flows. A recent research manifesto [19] describes the vision of ABPMSs and delineates the lifecycle of an ABPMS, which expands that of a classical BPMS in two directions. On the one hand, the traditional lifecycle phases (i.e., modeling, analysis, execution, monitoring, etc.) are continuously iterated, and empowered with AI capabilities. On the other hand, the lifecycle includes additional tasks that can only be realised with AI support, namely those of adaptation, explanation, and continuous improvement.

In this transition, one particularly relevant aspect is that BP modelling is lifted to the more general notion of framing, which entail establishing multiple constraints encompassing procedural rules, best practices, and norms that must be considered during BP execution. Within the provided frame, an ABPMS is expected to be: (i)

autonomous to act independently and proactively; (*ii*) *conversationally actionable* to seamlessly interact and cooperate with human users whenever the restrictions imposed by the frame cannot be met; (*iii*) *adaptive* to react to changes in its environment; (*iv*) (*self-*)*improving* to ensure the optimal achievement of its goals; (*v*) *explainable* to provide trust and, hence, foster collaboration with human users.

Among the most significant characteristics of an ABPMS is its ability to be *conversationally actionable*, i.e., being able to seamlessly interact with humans to not only respond to user queries and perform actions on their behalf but also initiate conversations with users to inform them of the BP progression, alert them of relevant BP changes, and make recommendations for interventions for improving the BP concerning relevant performance targets [19]. Indeed, integrating ABPMSs into a human workforce alters the role of human employees and dynamics, fueling a lack of trust, a notorious barrier to the adoption of automated technologies in Information Systems (ISs) [43]. In light of the considerations above, a possible solution to the lack of human trust in these ABPMSs can be the adoption of *Conversational Systems* (CSs).

CSs enable machines to engage with users in human-like dialogues to offer spoken, text-based, or multimodal conversational interactions with humans [51]. Thus, CSs can act as a natural language interface for the ABPMS towards the human and, consequently, boost the explainability of these ISs, since when users understand the reasoning behind the system's actions, they feel in control of the system and are more likely to trust it. In [16], the authors posit that ABPMSs can intensely benefit from the emergence of CSs, as they have the potential to empower the four main data-driven BPM approaches, namely: *Descriptive Process Analytics*, *Predictive Process Analytics*, *Prescriptive Process Optimization*, and *Augmented Process Execution*.

The main contribution of this paper is a survey for the analysis of the techniques developed in the field of CSs applied to BPM and the investigation of the related research problems and opportunities in the area. To achieve these objectives, we employed a rigorous search protocol across prominent digital libraries for each of the four identified topics, aiming to discover the state-of-the-art conversational techniques in BPM and to outline the research challenges to make an ABPMS conversationally actionable.

The rest of the paper is structured as follows. Section 2 introduces background knowledge about CSs and their taxonomies. Section 3 describes the adopted search protocol. Sections 4, 5, 6, and 7 explore CSs applied in each BPM area as identified in [16], discussing research challenges for improving ABPMSs conversational features. Section 8 reports the related work in the field. Finally, Section 9 concludes the paper by summarizing its key findings and discussing threats to the study's validity.

2 Background on Conversational Systems

This research area sits at the intersection of Natural Language Processing (NLP), Machine Learning (ML), and Information Retrieval, taking advantage of the techniques developed in these fields to make sense of user queries, provide context-aware responses, and engage, oftentimes, in multi-turns conversations [51].

Given these capabilities, CSs have historically attracted both scientific and industrial interests thanks to their potential for enhancing user interactions in various application domains, from customer support and healthcare to education and enterprises. The roots of the field can be traced back to the early attempts of the 1960s, with chatbots embedding predefined scripts to direct responses [70]. However, only in modern times, the vast availability of data about human conversations freely accessible on the Internet and the late breakthroughs in ML, and more specifically in Deep Learning, have enabled CSs to achieve goals initially deemed unattainable, leading to widespread popularity and generating considerable hype even among individuals without a technical background [33].

Several works have contributed significantly to the understanding and advancement of conversational agents. Indeed, they delve into the historical development and taxonomies, underlying technologies, and practical implementations of CSs. The first categorization that can be drawn for these systems is according to the problem they aim to address [29]:

- *Question answering* CSs respond to the user query in a direct and precise way, exploiting a huge amount of data coming from heterogeneous sources (i.e., Web documents) or local knowledge bases (i.e., business datasets).
- *Task completion* CSs execute a specific task requested by the user, spanning from setting reminders to scheduling meetings. In this case, the task is usually defined beforehand and the system is tailored to it.
- *Social chat* or *open dialogue* CSs engage in fluid and contextually appropriate conversations with users for the sake of entertainment or companionship, resembling human interaction similarly to the Turing test.

Another traditional categorization is based on the nature of the supported conversations, namely *single-turn* or *multi-turns* [71]. The former is relatively straightforward, focusing uniquely on the user query to generate an answer. In contrast, the latter considers the *context*, incorporating utterances from previous turns in the conversation to achieve a major degree of user engagement.

From a technological standpoint, many paradigms have been implemented to realize CSs [51]. In *rule-based* CSs, the dialogue is handled with a modularized architecture, and the conversation flow is defined in advance by designers who meticulously craft dialogue rules forming a set of if-then statements. These rules define the system's understanding and response to user inputs, yet limit its expressiveness. Over the years, *statistical data-driven* approaches have become predominant, learning conversational strategies from data but maintaining the overall modular framework. In particular, it is worth highlighting the possibility of addressing the dialogue formulation as a sequential decision-making process, modeling it as a Markov Decision Process, and employing Reinforcement Learning to find an optimal solution. Recently, researchers in CSs have shifted their focus towards the creation of *end-to-end neural* conversational agents, marking a significant change from traditional modular techniques that involve distinct components for understanding user input and generating responses. Instead, these systems directly map input utterances to output responses by leveraging Deep Neural Networks.

Among end-to-end neural approaches, we identify diverse methodologies [28].

- *Retrieval-based* techniques generate an answer to the user query by selecting the answer from a large set of candidate responses. Upon receiving a user

utterance, the model encodes it together with the conversational context in a dense representation. Subsequently, it iterates over the whole set of candidate responses, assigning a score to each one depending on its appropriateness through a function that matches the context and the possible response. Eventually, the model produces in output the candidate utterance having the highest score.
- *Generation-based* techniques adopt an opposite approach, synthesizing the answer sequentially, word by word. Preeminent solutions implementing this paradigm primarily rely on the *encoder-decoder* architecture. The encoder translates the context into a hidden state, representing contextual information as a vector. Subsequently, the decoder selects a new word and adjusts consequently the hidden state at each time step in an auto-aggressive fashion. Even if they lag in performance compared to their counterparts, also non-auto-aggressive methods were explored to allow parallel token generation, considering each word conditionally independent in per-step distribution. The aforementioned encoder-decoder architecture can be implemented by means, respectively, of Long Short-Term Memory (LSTM) or Gated recurrent unit (GRU) for recurrent neural networks [62], or a stack of self-attention layers and cross-attention layers for transformer networks [67].
- *Hybrid* techniques overcome the limitations of the above methodologies, combining their strength. Indeed, retrieval-based approaches can provide high-quality responses but offer a limited hypothesis space of candidates while, on the contrary, generation-based methods can produce novel answers but with no guarantees about their quality. For this reason, hybrid techniques first retrieve instances of similar conversations from their dataset and, afterward, exploit them to support the generation of the response in many ways (e.g., Retrieval-Augmented Generation [45]).

Within generation-based approaches, the prominence and versatility of *Language Models* (LMs) have reached unprecedented heights, demanding a dedicated investigation. From the first statistical models, LMs have evolved into transformer models (e.g., BERT) pre-trained over massive textual datasets, demonstrating robust effectiveness in addressing NLP tasks via text generation. Notably, by increasing the size of these models (i.e., the number of parameters) over a specific threshold and moving to *Large Language Models* (LLMs) such as GPT-4 and Llama 2, researchers observed not only an important enhancement in performance but also the manifestation of peculiar capabilities that are not demonstrated by smaller LMs (in-context learning, instruction following, step-by-step reasoning) [75].

3 Search Protocol

To conduct this survey, a reproducible search protocol was employed to ensure a comprehensive exploration of the literature produced in the field of CSs applied to BPM, borrowed by the scientific methodology exposed by Kitchenham [40].

Initially, we formulated the research questions to define the scope of the search and produced a list of search strings. Afterward, we executed the search strings across diverse data sources. Ultimately, we applied inclusion criteria to select the studies acquired through the search. To the end of exploring the aforementioned research area, we identified the following research questions tailored to each BPM family identified in Section 1:

- **RQ1**: *Which conversational techniques are adopted in the BPM field?*
- **RQ2**: *What research challenges need to be addressed in creating actionable conversations for ABPMSs?*

In a nutshell, *RQ1* aims at discovering the most relevant conversational techniques implemented at the moment of writing for each BPM approach and *RQ2* explores future research challenges (RCs) and possible solutions.

Next, we formulated the search strings tailored to each data-driven BPM area as identified in [16]. Through an iterative trial-and-error process, we recognized the need to extend the scope of the search by incorporating broader terms to refer to the fields under examination. The search strings resulting from these considerations were defined as: *(Q1) AND (Q2)*, where *(Q1)* represents the fixed part that remains consistent across all search strings and *(Q2)* represents the variable part that changes based on the specific area of the search. Notably, *(Q1)* corresponds to: *("conversational" OR "dialogue" OR "chatbot" OR "natural language" OR "LLM" OR "language model" OR "ChatGPT")*. It is worth justifying the decision to consider "*ChatGPT*" as opposed to its competitors. We opted for its inclusion due to its widespread diffusion, often misinterpreted as a synonym for general LLM by non-technical users. Conversely, *(Q2)* is:

- *("descriptive process analytics" OR "process discovery" OR "conformance checking" OR "performance mining" OR "variant analysis" OR "process mining" OR "process modeling")* for Descriptive Process Analytics;
- *("predictive process analytics" OR "what-if analysis" OR "digital twin" OR "predictive process monitoring" OR "process analysis")* for Predictive Process Analytics;
- *("prescriptive process optimization" OR "process optimization" OR "prescriptive process monitoring" OR "process redesign")* for Prescriptive Process Optimization;
- *("process execution" OR "robotic process automation" OR "process automation" OR "process implementation")* for Augmented Process Execution.

First, each of the four search strings was employed in querying Google Scholar to retrieve studies where the search strings appeared in the *title*, *keywords*, or *abstract* of the paper. Subsequently, we double-checked and complemented the studies discovered in this primary search using widely recognized academic databases such as Scopus, ACM Digital Library, and IEEE Xplore. The search was completed in January 2024.

To maintain the focus on the most pertinent studies, to be considered a study must satisfy all the following inclusion criteria.

- **IN1**: *The study encompasses a technique for CSs in BPM.*
- **IN2**: *The study is peer-reviewed.*
- **IN3**: *The study is electronically available.*
- **IN4**: *The study is written in English.*
- **IN5**: *In case of multiple publications discussing the same technique, only the most comprehensive study is included.*

Table 1. Statistics of the search for each BPM area.

Phase	Descriptive Process Analytics	Predictive Process Analytics	Prescriptive Process Optimization	Augmented Process Execution	Total
Google Scholar	32	73	14	91	210
Scopus	356	236	119	198	909
ACM Digital Library	20	8	1	12	41
IEEE Xplore	25	19	3	34	81
All Publications	433	336	137	335	1241
Duplicates	60	61	7	63	191
All (-duplicates)	373	275	130	272	1050
Excluded	346	266	127	257	996
Final	27	9	3	15	54

Hence, the applied exclusion criterion was to eliminate studies that violated at least one of the aforementioned inclusion criteria. The results of our search protocol were documented in four distinct spreadsheets, each corresponding to a specific category. Comprehensive statistical data are presented in Table 1, showing that 54 studies were finally selected to conduct the survey analysis.

In the following sections, we present the outcomes of applying our search protocol for each BPM area considered, addressing the above research questions by introducing the conversational techniques developed for BPM, and identifying the research challenges (RCs), outlined in the concluding Table 2, along with potential solutions. The outcomes of the papers' extraction and selection for the survey are available at: `https://zenodo.org/doi/10.5281/zenodo.10827054`. Figure 1 illustrates the distribution of selected publications per year, categorized by BPM area.

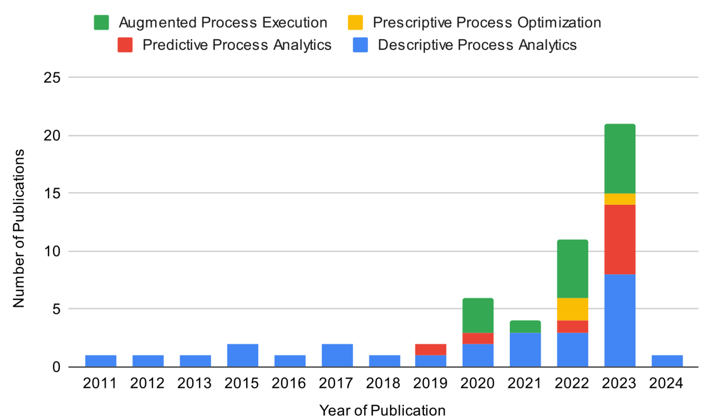

Fig. 1. Year distribution of the selected publications, grouped by BPM area.

4 Descriptive Process Analytics

The first data-driven BPM area we tackle in our analysis is *Descriptive Process Analytics* (DPA). This family of methodologies deals with the *as-is* description of BPs, i.e., their current state, supporting domain experts in identifying problems and potential improvements [16]. Its scope ranges from BP modeling to BP monitoring, including advanced process mining (PM) applications for analyzing the performance and conformance of BPs based on event logs produced during their execution [20].

Literature Analysis and Research Challenges. During interactions with business stakeholders, the ability to extract BP models from natural language descriptions proves to be highly time-efficient and effective. Over the years, researchers have developed many NLP techniques to this end, such as [27] for BPMN models. In [65], semantic unification is embedded to deal with partial and potentially contradictory information, while [14] generates BP models from descriptions in controlled natural language, enabling the BP discovery through user interactions. Other NLP methodologies to extract BP models from unstructured organizational documentation are exposed in [63,23]. In [56], the authors introduce an encoder-decoder translator to represent BP models in a middle representation and decode them into natural language descriptions, to mitigate information loss and semantic errors. With [26], a platform architecture enables the export of BP models into natural language and vice versa through a Web service interface, whereas, [37] presents a modeling environment facilitating the rapid generation of visual BPMN models from constrained natural language input during interviews and design workshops. [66] proposes a Machine Translation-inspired approach to simplify the BP modeling phase by generating BPMN diagrams from textual descriptions in natural language. [21] presents a method that combines BPMN with NLP to generate BPMN diagrams from natural language BP descriptions, employing Probabilistic Latent Semantic Analysis to cope with ambiguities in natural language. [54] introduces an automated approach utilizing NLP and Prolog to extract BP models in UML from user stories. [55] extends a data-driven pipeline for the automated generation of BP models from natural language text. Approaches to allow the natural language description of BP were also investigated for Declare, notably in [48] and in [1], whereby constraints can be expressed verbally and are converted via speech recognition into the closest set of Declare constraints. Prospective improvements in these methodologies may address the implicit ambiguity of natural language, which can result in unpredictable and varied representations of the same BP (**RC1**) [2]. A potential solution to this challenge involves refining the algorithms for semantic analysis and context-aware processing, ensuring a more consistent and standardized interpretation of natural language inputs.

The reverse approach was also explored, wherein BP models are expressed using natural language to enhance human comprehension. Relevant studies on this topic can be identified in [44] for prescriptive BP models and in [4] for declarative ones. Furthermore, [49] introduces an approach to automatically generate a natural language representation of BPMN models leveraging business rules in SBVR as an intermediate representation, and [24] utilizes PM, fuzzy linguistic protoforms, and natural language generation (NLG) to generate specialized textual explanations of BPs automatically. The challenges primarily revolve around comprehending model labels and explaining parallel behaviors in natural language without compromising the reader's grasp of the underlying BP semantics (**RC2**) [2]. To overcome the former, NLP techniques incorporating contextual analysis and domain-specific knowledge can be applied for precise word categorization. Meanwhile, addressing the latter challenge may involve adopting a structured and standardized template to articulate parallel behaviors, thereby enhancing clarity and maintaining human comprehensibility.

Moving to methodologies based on chatbots, it is worth mentioning [60], whereby chatbots are employed to enable adaptive learning of BPs in a multi-actor environment, and [25], which presents a conversational agent designed for tasks such as consistency, conformance, and model checking for declarative BP models. In tackling the challenge of bidirectional interaction between natural language descriptions and BP models (**RC3**), a promising direction can be embodied by multimodal LLMs. These models can comprehend BP models and respond to user queries in a grounded manner, offering visual aids such as charts for improved human interpretation. Moreover, their generative abilities enable them to transform inputs in natural language into the corresponding BP model, relying on advanced NLP techniques for semantic interpretation. Furthermore, the exploration of explainable AI (XAI) methodologies can enhance the transparency of the chatbot's decision-making process, fostering user trust and usability in dynamic process environments.

Another important trend is the development of conversational interfaces to enhance the understanding of PM findings and make them accessible to non-technical users. Notably, [8] addresses this challenge by developing a natural language querying interface in combination with Everflow and proposing a taxonomy of PM questions. [72] suggests a methodology to facilitate data extraction from PM using a natural language interface, eliminating the need for programming in a PM query language. [42] enhances the querying experience for domain analysts with limited technological expertise by introducing a natural language interface that utilizes graph-based storage techniques, i.e., labeled property graphs, and executes queries through the Cypher language. In [31], a solution is presented to automatically discover BPs models from textual documentation using a neural network with the Ordered Neurons LSTM architecture and a process-level language model objective. [22] suggests enhancing BP discovery with causal BP discovery and XAI for improving the interpretability of BP execution outcomes through LLMs. [57] reports a method facilitating advanced PM by automatically extracting BP information, including resources and business objects, from event data through semantic role labeling, employing an attribute classification technique. [58] introduces an approach for defining measurable Process Performance Indicators by combining NLP techniques and tailored matching strategies to integrate textual descriptions with event logs. The challenges in this domain involve the limited generalization of rule-based semantic parsing, necessitating new rules for novel questions, prompting the need for hybrid approaches mixing AI and rule-based techniques for NLP in PM, extending systems capabilities to support complex queries, multimodal conversational interfaces, and specialized evaluation frameworks (**RC4**). Moreover, in the analysis of event logs, the integration of LLMs may enable the automated execution of entity-extraction tasks and the computation of semantic similarity.

5 Predictive Process Analytics

We follow in our discussion with *Predictive Process Analytics* (PPA). This area is aimed at building predictive models to estimate the future state of the BP, enabling the prediction of its performance [16]. Thus, we can view this sub-field of BPM as framed within the activities of the *Process Analysis* step of the usual BPM lifecycle [20]. In particular, there are two preeminent techniques for this BPM approach. The first is *what-if digital process twins* that, constructing a simulation model representing the BP, tries to forecast the impact of changes to the BP concerning relevant KPIs. The second is *predictive process monitoring*, which leverages ML algorithms to learn a predictive model from historical data and uses it to generate predictions both at the case and BP levels.

Literature Analysis and Research Challenges. In [9], the authors introduce a chatbot for BP simulation that allows to conversationally specify what-if scenarios, simplifying the procedure for users without technical knowledge and enabling the comparison of the BP performance under these scenarios against a standard reference. In [46], the authors present a conversational system utilizing GPT-4 to automate the creation of digital twins in the data centers domain. Furthermore, [50] explores the application of ChatGPT in conjunction with Digital Twins in the construction industry, and [64] introduces neuro-symbolic reasoning for interacting with 3D digital twins using natural language in aircraft maintenance. Research opportunities in conversational what-if analysis concern the support for more extensive customization of the digital twin (**RC5**), e.g., allowing domain-specific modifications and tailoring the KPIs on that particular BP. Additionally, LLMs could play a crucial role by integrating with simulation engines, facilitating multimodal interaction, and allowing users to conversationally customize views over the digital twins.

Within the realm of predictive process monitoring, [12] introduces a text-aware approach combining ML and NLP to monitor knowledge-intensive BPs and incorporating structured features and unstructured textual information over the control flow, whereas [13] presents a text-aware technique to predict the next activity and timestamp in BP instances considering semantic information by including contextualized word embeddings. Moreover, [34] proposes the use of Attention-Based LSTM with Multi-Task Learning to improve business behavior prediction accuracy from historical event logs. In contrast to LSTM which relies solely on the last hidden state for predictions, [38] leverages the notion of attention but considers all hidden states to predict future behavior accurately. In [69], the authors investigate the integration of textual data into predictive process monitoring techniques to improve accuracy, while also prioritizing explainability, which boosts transparency and interpretability for black-box ML models at the cost of increased computation time. Potential research directions in predictive process monitoring could focus on enhancing explainability (**RC6**). Specifically, future endeavors may involve integrating explainability analysis with subsequent causality analysis to distinguish whether a variable is correlated or causally related to the outcome. Furthermore, investigating the connection between explainability analysis results and practical interventions could be explored [69].

6 Prescriptive Process Optimization

Prescriptive Process Optimization (PPO) is mainly concerned with the optimization of the process, especially through the translation of the findings from Predictive Process Analytics into actual actions to undertake for enhancing the execution of the process [16]. Drawing the parallelism with the traditional BPM lifecycle, this body of methodologies falls primarily into the *Process Redesign* phase, for the production of a *to-be* version of the process model [20]. In this case, these techniques refer to the training of a predictive model utilizing historical data, later employed for making predictions regarding the underlying process. Such predictions serve as input for a recommender system, which, based on this information, generates recommendations for subsequent courses of action. These recommendations may undergo automatic execution, be presented to domain experts, or be integrated into the prescriptive system to refine the model for subsequent recommendation generation. PPO methodologies can be divided into two broad classes. *Automated process optimization* is directed at proposing alterations to the BPs in order to achieve a balance between competing KPIs, such as reducing costs while simultaneously maximizing the quality. On the other hand, *prescriptive process monitoring* proposes recommendations about actions to perform for the process optimization with respect to the selected KPIs in real-time or near-real-time and, oftentimes, at the case level.

Given that PPO is in the early stages of development, a limited number of studies have explored its intersection with CSs.

Literature Analysis and Research Challenges. In [6], ChatGPT is employed for BP optimization within the domain of Additive Manufacturing (AM). The integration of ChatGPT into the manufacturing workflow is reported to result in improved efficiency, cost reduction, and increased accessibility in the field. In [52], the authors present a NLP approach for BP Redesign, aimed at the extraction of the redesign suggestions from end-user feedback in natural language and tested in a real-world use case with an extensive experiment program. Research opportunities include developing NLP-based approaches to automatically extract change proposals for BP Redesign, performing sentiment analysis on the suggestions, prioritizing them, and employing clustering to group suggestions based on similarity and frequency. In this domain, LLMs can be embedded to generate recommendations for BP improvement, working in conjunction with the predictive layer to assess and validate these suggestions (**RC7**).

In [73], crowd-wisdom and goal-driven methods from prescriptive process monitoring are applied to AI-powered BPs, blending classical BPM, goal-driven chatbots, and conversational recommendation systems, introducing a synthesized dataset derived from a real use case. Future research challenges in CSs for prescriptive process monitoring involve considering Reinforcement Learning to leverage implicit and explicit user feedback, handling complex utterances using AI planning for dynamic orchestration of automated tasks, and employing DL and NLP methodologies to map user utterances to activities, followed by the translation of the results back into natural language recommendations (**RC8**).

7 Augmented Process Execution

Augmented Process Execution (APE) brings a paradigm shift from the reactive execution that lies on the human, aided by system suggestions, to the inverse execution model, where the system proactively carries out the BP execution, supported by human operators. [16]. Now we are in the *Process Execution* stage of the BPM lifecycle, where the *to-be* BP is executed within the IS of the organization [20]. Furthermore, this is where the notion of ABPMS introduced in Section 1 comes in, with the domain experts that intervene in the execution only when the system needs it to disambiguate its behavior in a specific situation [19]. Within this sub-field, we can identify two categories of systems, depending on their interaction with the human operator. An *autonomic process execution system* operates within its predefined frame and resorts to human intervention and decision-making whenever it encounters an uncertain scenario. Conversely, an *autonomous process execution system* not only exercises full control over the BP within the frame, but it can even operate modifications on it to achieve specific business goals. In this context, the human assumes the supervisor role, intervening solely to avoid undesired consequences. Notably, Robotic Process Automation (RPA) can be considered a specific instance of AI-augmented BPM, aiming to realize more complex automation than a traditional BPMS. RPA operates on applications' user interface (UI) by creating software robots that automate mouse and keyboard interactions. This enables the automated execution of repetitive tasks on the UI, mitigating human errors stemming from mental lapses induced by boredom or exhaustion.

Literature Analysis and Research Challenges. In [15], the authors introduce a tool leveraging declarative design and AI planning to optimize complex BPs by composing with conversational agents and services. [47] implements a conversational agent using Rasa and Camunda Engine, designed to integrate with BPMSs and to simplify BP execution. [30] relies on LLMs to tackle various BPM tasks, and to assess the suitability of BP tasks for RPA. The paper [10] proposes leveraging GPT technology to generate new BP models, enhance decision-making in data-centric BPs, and improve overall BPM efficiency through task automation, insights provision, and operational enhancements. [7] introduces a no-code conversational interface facilitating collaboration between human users and bots in knowledge-intensive BPs, by enabling bots to identify user intents, orchestrate automation tasks, and include insights from conversations mining for performance monitoring and service quality improvement. [53] introduces an approach for verifying resource compliance requirements in BPs, leveraging GPT-4 for NLP and a customized compliance verification component. The study in [61] investigates the transformative role of AI and chatbots in procurement processes, and develops a chatbot to improve efficiency, reduce costs, and enhance supplier relations. In [32], a user-friendly natural language interface for querying runtime event data in BPMSs is introduced, enabling real-time insights without the need for backend knowledge and including a bootstrapping pipeline for the automatic instantiation of the natural language interface. In the domain of conversational agents for system-driven management, challenges arise in achieving seamless interaction, requiring advanced interfaces capable of understanding nuanced user requests. Ensuring efficient correction and optimization based on user-specified natural language instructions poses a significant challenge, emphasizing the need for automated solutions. Additionally, addressing challenges related to intent classification accuracy and building user trust through explicit approval requests for automation are critical aspects of future research (**RC9**). Furthermore, considerable variation in LLM responses suggests the need for further research into their behavior and reactions to diverse inputs for consistent and reliable performance in APE.

Shifting the focus on how CSs can empower RPA, a solution combining RPA and chatbots to automate iterative BPs is presented in [35]. [59] presents a conversational digital assistant framework for interactive automation that addresses the accessibility issues of RPA for business users through natural language interaction and a multi-agent orchestration model. [17] develops a multi-channel chatbot integrated with RPA, automating the end-to-end BP of product exploration, purchase, and transaction, for consistent operation across channels with minimal human intervention. [18] integrates chatbots with RPA in manufacturing to efficiently process and present data, offering a solution for rapid access to information in manufacturing sites through an architecture capable of managing complex queries and processes. Another relevant problem in RPA is the identification of suitable routines to automate, which is a time and cost-intensive manual task and is tackled in [3] by exploring its automation through NLP techniques. In [36], the authors address the challenge of making APIs, which provide access to RPA bots, accessible to non-technical users by introducing a data augmentation approach using LLMs for intent recognition. [74] presents a technique leveraging LLMs to enhance RPA capabilities through automatic workflow generation, ensuring reliable reasoning and maintaining data integrity. Given this plethora of works, we can indicate possible future improvement opportunities. Notably, a unified conversational interface could simplify the integration of various RPA automation solutions, fostering exploration of sophisticated orchestration models and autonomous agent composition through natural language (**RC10**). Moreover, employing NLP for identifying automation candidates in BPs highlights potential future improvements. Indeed, NLP techniques play a crucial role in supporting the design and execution of automation routines and advancing the automation of non-trivial tasks through cognitive automation (**RC11**).

8 Related Work

While the field is still undergoing significant development, numerous surveys have already delved into the applications and challenges of conversational techniques applied in the context of BPM. One of the first examples we can find in the literature is [2], a position paper that explores the potential of NLP in enhancing the advantages of BPM activities across various organizational levels. The authors report the principal research directions for a successful implementation of NLP in automating specific tasks that, otherwise, would require compelling effort to be performed. The paper also describes possible concrete applications, considering both the process perspective and its improvement through NLP.

The exploration of NLP in BPM is further addressed in two notable studies, namely, [5] and [11]. The former conducts a systematic literature review (SLR) to investigate the utilization of NLP techniques in extracting BPs and ensuring BP quality from unstructured text throughout the BPM lifecycle. The latter complements this by performing a qualitative analysis of state-of-the-art tools for BP extraction from unstructured documents, in the direction of uncovering existing limitations and challenges within the field.

Transitioning the focus from NLP to CSs, their realm has undergone significant transformation recently, particularly with the emergence of LLMs, which introduced

Table 2. Research challenges for CSs in BPM.

BPM Area	Selected Papers	Identifier	Research Challenge
Descriptive Process Analytics	[27]; [65]; [14]; [63]; [23]; [56]; [26]; [37]; [66]; [21]; [54]; [55]; [48]; [1]; [44]; [4]; [49]; [24]; [60]; [25]; [8]; [72]; [42]; [31];[22]; [57]; [58]	RC1 RC2 RC3 RC4	*Unambiguous BP Discovery* *BP model semantics explanations* *From natural language to BP and vice versa* *Conversational interfaces for PM*
Predictive Process Analytics	[9]; [46]; [50]; [64]; [12]; [13]; [34]; [38]; [69]	RC5 RC6	*Conversational what-if analysis* *Explainable predictive process monitoring*
Prescriptive Process Optimization	[6]; [52]; [73]	RC7 RC8	*LLM-driven process redesign* *Multi-disciplinary integration for prescriptive process monitoring*
Augmented Process Execution	[15]; [47]; [30]; [10]; [7]; [53]; [61]; [32]; [35]; [59]; [17]; [18]; [3]; [36]; [74]	RC9 RC10 RC11	*Trustworthy conversational corrections* *Conversational RPA* *Cognitive automation*

new research perspectives in the area of BPM. In particular, the paper [39] introduces the notion of Large Process Model (LPM) that combines the correlation power of LLMs with the analytical precision of knowledge-based systems and automated reasoning approaches. The authors envision the integration of LLM applications at different stages of BPM and posit the feasibility of implementing an LPM, while also underscoring inherent limitations and research challenges that must be addressed for its realization.

In [68], the authors focus on addressing the opportunities LLMs present in BPM by identifying six research directions in this respect. It is also worth mentioning the work in [41], where the authors perform a SLR and evaluate existing chatbots to assist conversational BP modeling leveraging a real-world test set. Based on this study, usage recommendations and further development in the identified area are consequently derived.

Our study distinguishes itself from previous literature in the same domain by focusing on challenges directly associated with the application of conversational techniques in BPM, offering a comprehensive and structured overview to foster future research to realize conversationally actionable ABPMSs.

9 Threats to Validity and Concluding Remarks

This survey has explored conversational techniques specifically designed for BPM, aiming to provide a comprehensive overview of this cutting-edge research domain. We opted for a survey instead of a SLR due to the relative novelty of the topic. Indeed, a survey approach seemed more appropriate to capture the current state of the field, as it may not have reached a level of maturity that meets the more stringent criteria associated with a complete SLR. To ensure the validity of our search protocol, we focused on addressing issues, such as incompleteness, by utilizing multidisciplinary search engines. Furthermore, we aimed to minimize selection bias by applying an accurate review of the studies and providing clear motivations for exclusions when necessary. It is also crucial to acknowledge the potential subjectivity inherent in the planning and implementation of the search protocol. To mitigate this threat, we clearly defined the survey's objectives, scope, and inclusion criteria, meticulously documenting each phase of the study to enhance transparency and reproducibility. Additionally, we adopted a validated methodology [40] for study selection and data extraction to mitigate interpretation biases. Our rigorous analysis pursued two primary objectives: (i) elucidating the principal conversational techniques applied across various BPM domains, and (ii) examining the extent to which these techniques can enable actionable conversations for ABPMSs. An overview of the identified RCs is outlined in Table 2.

Acknowledgments. This paper has been supported by the Sapienza projects DISPIPE and FOND-AIBPM, the PRIN 2022 project MOTOWN, and the PNRR MUR project PE0000013-FAIR. The work of Angelo Casciani has been carried out in the range of the Italian National Doctorate on AI run by Sapienza.

References

1. van der Aa, H., Balder, K.J., Maggi, F.M., Nolte, A.: Say it in your own words: Defining declarative process models using speech recognition. In: BPM Forum (2020)
2. Van der Aa, H., Carmona Vargas, J., et al.: Challenges and opportunities of applying natural language processing in business process management. In: COLING (2018)
3. van der Aa, H., Leopold, H.: Automatically identifying process automation candidates using natural language processing. Springer (2022)
4. Ackermann, L., Schönig, S., Zeising, M., Jablonski, S.: Natural language generation for declarative process models. In: CAiSE Workshops (2015)
5. de Almeida Bordignon, A.C., Thom, L.H., Silva, T.S., et al.: Natural language processing in business process identification and modeling: a systematic literature review. In: Brazilian Symposium on Information Systems (2018)
6. Badini, S., Regondi, S., et al.: Assessing the capabilities of ChatGPT to improve additive manufacturing troubleshooting. Adv. Ind. and Eng. Polymer Research (2023)
7. Bandlamudi, J., Mukherjee, K., Agarwal, P., et al.: Towards Hybrid Automation by Bootstrapping Conversational Interfaces for IT Operation Tasks. In: AAAI (2023)
8. Barbieri, L., Madeira, E., Stroeh, K., van der Aalst, W.: A natural language querying interface for process mining. Journal of Intelligent Inf. Sys. **61**(1) (2023)
9. Barón-Espitia, D., Dumas, M., González-Rojas, O.: Coral: Conversational What-If Process Analysis. In: ICPM (2022)
10. Beheshti, A., Yang, J., Sheng, Q.Z., et al.: ProcessGPT: Transforming Business Process Management with Generative Artificial Intelligence. In: IEEE ICWS (2023)
11. Bellan, P., Dragoni, M., Ghidini, C.: A Qualitative Analysis of the State of the Art in Process Extraction from Text. DP@ AI* IA (2020)
12. Brennig, K., Benkert, K., Löhr, B., Müller, O.: Text-Aware Predictive Process Monitoring of Knowledge-Intensive Processes: Does Control Flow Matter? In: BPM (2023)
13. Cabrera, L., Weinzierl, S., Zilker, S., Matzner, M.: Text-Aware Predictive Process Monitoring with Contextualized Word Embeddings. In: BPM Workshops (2022)
14. Caporale, T.: A tool for natural language oriented business process modeling. In: 8th Central-European Workshop on Services and their Composition (2016)
15. Chakraborti, T., Agarwal, S., Khazaeni, Y., et al.: D3BA: A Tool for Optimizing Business Processes Using Non-deterministic Planning. In: BPM Workshops (2020)
16. Chapela-Campa, D., Dumas, M.: From process mining to augmented process execution. Software and Systems Modeling (2023)
17. Dan, G., Claudiu, D., Alexandra, F., et al.: Multi-Channel Chatbot and Robotic Process Automation. In: IEEE Int. Conf. on Automation, Quality and Testing, Robotics (2022)
18. Do, S., Jeong, J.: Design and Implementation of RPA Based ChatMES System Architecture for Smart Manufacturing. WSEAS Trans. on Computer Research **10** (2022)
19. Dumas, M., Fournier, F., Limonad, L., Marrella, A., et al.: AI-augmented business process management systems: a research manifesto. ACM Trans. on Man. Inf. Sys. **14**(1) (2023)

20. Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A., et al.: Fundamentals of Business Process Management, vol. 1. Springer (2013)
21. Elmanaseer, S., Alkhatib, A.A., Albustanji, R.N.: A Proposed Technique for Business Process Modeling Diagram Using Natural Language Processing. In: ICIT (2023)
22. Fahland, D., Fournier, F., Limonad, L., et al.: Why are my Pizzas late? In: IJCAI (2023)
23. Ferreira, R.C.B., Thom, L.H., Fantinato, M.: A semi-automatic approach to identify business process elements in natural language texts. In: ICEIS (2017)
24. Fontenla-Seco, Y., Lama, M., Bugarín, A.: Process-To-Text: A Framework for the Quantitative Description of Processes in Natural Language. In: TAILOR (2020)
25. Fontenla-Seco, Y., Winkler, S., Gianola, A., et al.: The Droid You're Looking For: C-4PM, a Conversational Agent for Declarative Process Mining. In: BPM Forum (2023)
26. Freytag, T., Kanzler, B., Leger, N., Semling, D.: NLP as a service: An API to convert between process models and natural language text. In: BPM (PhD/Demos) (2021)
27. Friedrich, F., Mendling, J., Puhlmann, F.: Process model generation from natural language text. In: CAiSE (2011)
28. Fu, T., Gao, S., et al.: Learning towards conversational AI: A survey. AI Open **3** (2022)
29. Gao, J., Galley, M., Li, L.: Neural approaches to conversational AI. In: SIGIR (2018)
30. Grohs, M., Abb, L., Elsayed, N., Rehse, J.R.: Large Language Models Can Accomplish Business Process Management Tasks. In: BPM Workshops (2024)
31. Han, X., Hu, L., Mei, L., et al.: A-BPS: Automatic Business Process Discovery Service using Ordered Neurons LSTM. In: IEEE ICWS (2020)
32. Han, X., Hu, L., Sen, J., et al.: Bootstrapping Natural Language Querying on Process Automation Data. In: IEEE SCC (2020)
33. Haque, M.U., Dharmadasa, I., et al.: "I think this is the most disruptive technology": Exploring Sentiments of ChatGPT Early Adopters using Twitter Data. arXiv (2022)
34. Hnin, T., Oo, K.K.: Attention based LSTM with multi tasks learning for predictive process monitoring. In: Int. Workshop on Computer Science and Engineering (WCSE) (2019)
35. Hung, P.D., Trang, D.T., Khai, T.: Integrating Chatbot and RPA into Enterprise Applications Based on Open, Flexible and Extensible Platforms. In: CDVE (2021)
36. Huo, S., Mukherjee, K., Bandlamudi, J., et al.: Accelerating the Support of Conversational Interfaces for RPAs Through APIs. In: BPM Forum (2023)
37. Ivanchikj, A., Serbout, S., Pautasso, C.: Live process modeling with the BPMN Sketch Miner. Software and Systems Modeling **21**(5) (2022)
38. Jalayer, A., Kahani, M., Beheshti, A., et al.: Attention Mechanism in Predictive Business Process Monitoring. In: IEEE 24th EDOC (2020)
39. Kampik, T., Warmuth, C., Rebmann, A., et al.: Large Process Models: Business Process Management in the Age of Generative AI. arXiv (2023)
40. Kitchenham, B.: Procedures for performing systematic reviews. Keele, UK, Keele University **33**(2004) (2004)
41. Klievtsova, N., Benzin, J.V., Kampik, T., et al.: Conversational Process Modelling: State of the Art, Applications, and Implications in Practice. In: BPM Forum (2023)
42. Kobeissi, M., Assy, N., Gaaloul, W., et al.: Natural language querying of process execution data. Information Systems **116** (2023)
43. Lee, J.D., See, K.A.: Trust in automation: Designing for appropriate reliance. Human factors **46**(1) (2004)
44. Leopold, H., Mendling, J., Polyvyanyy, A.: Generating natural language texts from business process models. In: CAiSE (2012)
45. Lewis, P., Perez, E., Piktus, A., et al.: Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems **33** (2020)
46. Li, M., Wang, R., Zhou, X., et al.: ChatTwin: Toward Automated Digital Twin Generation for Data Center via Large Language Models. In: ACM BuildSys (2023)
47. Lins, L.F., Melo, G., Oliveira, T., et al.: PACAs: Process-Aware Conversational Agents. In: BPM Workshops (2021)
48. López, H.A., Debois, S., Hildebrandt, T.T., Marquard, M.: The process highlighter: From texts to declarative processes and back. In: BPM (2018)
49. Malik, S., Bajwa, I.S.: Back to origin: Transformation of business process models to business rules. In: BPM Workshops (2012)

50. Mateev, M.: Predictive Analytics Based on Digital Twins, Generative AI, and ChatGPT. In: World Multi-Conference on Systemics, Cybernetics and Informatics, WMSCI (2023)
51. McTear, M.: Conversational ai: Dialogue systems, conversational agents, and chatbots. Springer Nature (2022)
52. Mustansir, A., Shahzad, K., Malik, M.K.: Towards automatic business process redesign: an NLP based approach to extract redesign suggestions. Automated Software Engineering **29**(1) (2022)
53. Mustroph, H., Barrientos, M., Winter, K., Rinderle-Ma, S.: Verifying Resource Compliance Requirements from Natural Language Text over Event Logs. In: BPM (2023)
54. Nasiri, S., Adadi, A., Lahmer, M.: Automatic generation of business process models from user stories. Int. Journal of Elect. and Comp. Eng. **13**(1) (2023)
55. Neuberger, J., Ackermann, L., Jablonski, S.: Beyond Rule-Based Named Entity Recognition and Relation Extraction for Process Model Generation from Natural Language Text. In: CoopIS (2023)
56. Qian, C., Wen, L., Kumar, A.: BEPT: A behavior-based process translator for interpreting and understanding process models. In: CIKM (2019)
57. Rebmann, A., van der Aa, H.: Extracting Semantic Process Information from the Natural Language in Event Logs. In: CAiSE (2021)
58. Resinas, M., del Río-Ortega, A., van der Aa, H.: From Text to Performance Measurement: Automatically Computing Process Performance Using Textual Descriptions and Event Logs. In: BPM (2023)
59. Rizk, Y., Isahagian, V., Boag, S., et al.: A Conversational Digital Assistant for Intelligent Process Automation. In: BPM Forum (2020)
60. Rooein, D., Bianchini, D., Leotta, F., et al.: aCHAT-WF: Generating conversational agents for teaching business process models. Software and Systems Modeling **21**(3) (2022)
61. Sai, B., Thanigaivelu, S., Shivaani, N., et al.: Integration of Chatbots in the Procurement Stage of a Supply Chain. In: CSITSS (2022)
62. Sherstinsky, A.: Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena **404** (2020)
63. Sintoris, K., Vergidis, K.: Extracting business process models using natural language processing (NLP) techniques. In: IEEE 19th Conference on Business Informatics (2017)
64. Siyaev, A., Valiev, D., Jo, G.S.: Interaction with Industrial Digital Twin Using Neuro-Symbolic Reasoning. Sensors **23**(3) (2023)
65. Sokolov, K., Timofeev, D., Samochadin, A.: Process extraction from texts using semantic unification. In: IC3K (2015)
66. Sonbol, R., Rebdawi, G., Ghneim, N.: A Machine Translation Like Approach to Generate Business Process Model from Textual Description. SN CS **4**(3) (2023)
67. Vaswani, A., Shazeer, N., Parmar, N., et al.: Attention is all you need. Advances in Neural Information Processing Systems **30** (2017)
68. Vidgof, M., Bachhofner, S., Mendling, J.: Large Language Models for Business Process Management: Opportunities and Challenges. In: BPM Forum (2023)
69. Warmuth, C., Leopold, H.: On the Potential of Textual Data for Explainable Predictive Process Monitoring. In: ICPM Workshops (2022)
70. Weizenbaum, J.: ELIZA—a computer program for the study of natural language communication between man and machine. Communications of the ACM **9**(1) (1966)
71. Yan, R.: "Chitty-Chitty-Chat Bot": Deep Learning for Conversational AI. In: IJCAI (2018)
72. Yeo, H., Khorasani, E., Sheinin, V., et al.: Natural Language Interface for Process Mining Queries in Healthcare. In: IEEE Big Data (2022)
73. Zeltyn, S., Shlomov, S., Yaeli, A., Oved, A.: Prescriptive Process Monitoring in Intelligent Process Automation with Chatbot Orchestration. In: PMAI (2022)
74. Zeng, Z., Watson, W., Cho, N., et al.: FlowMind: Automatic Workflow Generation with LLMs. In: ACM International Conference on AI in Finance (2023)
75. Zhao, W.X., Zhou, K., Li, J., et al.: A Survey of Large Language Models. arXiv (2023)