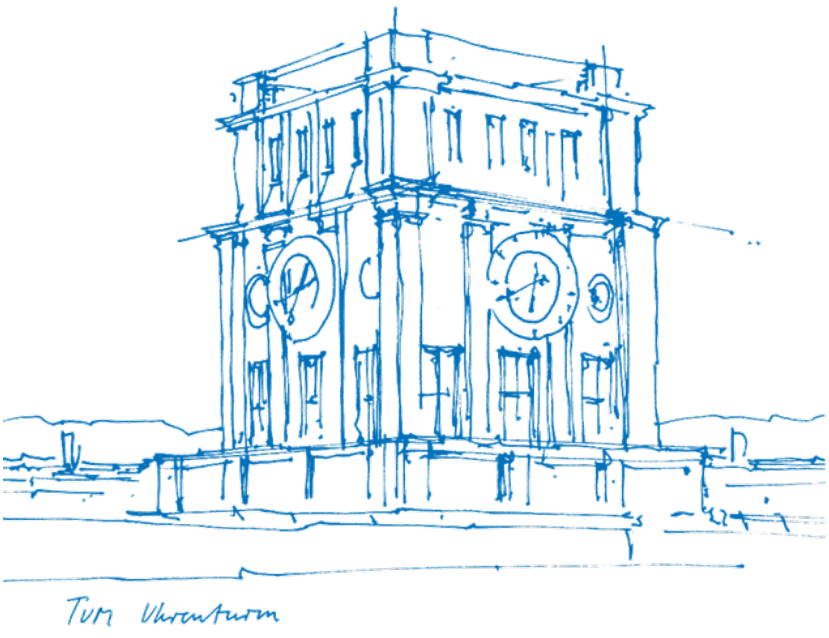
Chair of Information Systems and
Business Process Management (i17)
Department of Computer Science
TUM School of Computation, Information and Technology
Technical University of Munich

Bachelor's Thesis in Information Systems

Vincent Derek Held

An enhanced automated approach for transforming natural language process descriptions to BPMN2.0 process diagrams – with an evaluation of the application to ISO-Norm process descriptions



Chair of Information Systems and
Business Process Management (i17)
Department of Computer Science
TUM School of Computation, Information and Technology
Technical University of Munich

Bachelor's Thesis in Information Systems

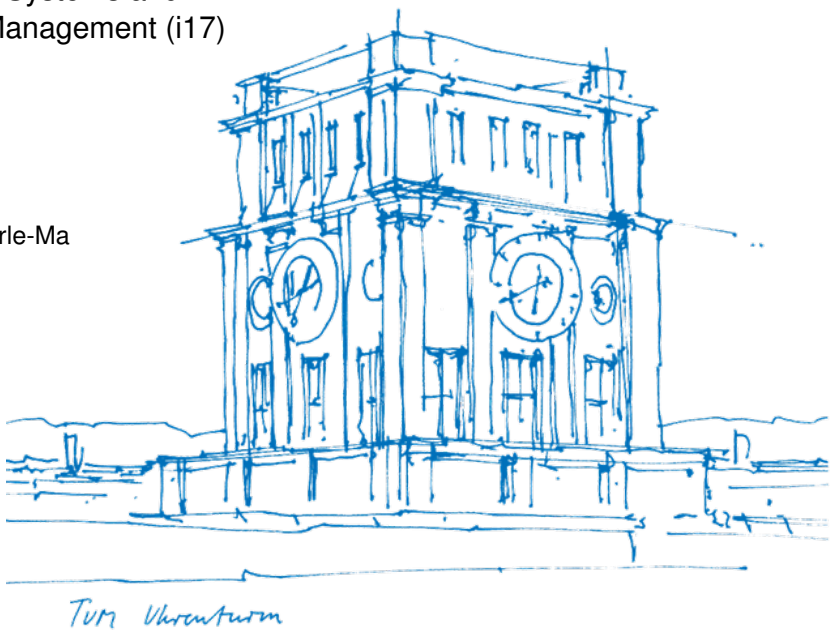
Vincent Derek Held

An enhanced automated approach for transforming natural language process descriptions to BPMN2.0 process diagrams – with an evaluation of the application to ISO-Norm process descriptions

Ein verbessertes automatisiertes Verfahren zur Umwandlung natürlichsprachlicher Prozessbeschreibungen in BPMN 2.0 Prozessdiagramme – mit einer Evaluierung der Anwendung auf ISO-Norm-Prozessbeschreibungen

Thesis for the Attainment of the Degree
Bachelor of Science

at the TUM School of Computation, Information and Technology,
Department of Computer Science,
Chair of Information Systems and
Business Process Management (i17)



Examiner
Prof. Dr. Stefanie Rinderle-Ma

Supervised by
Catherine Sai

Submitted by
Vincent Derek Held

Submitted on
15.12.2023

Declaration of Academic Integrity

I confirm that this bachelor's thesis is my own work and I have documented all sources and material used.

I am aware that the thesis in digital form can be examined for the use of unauthorized aid and in order to determine whether the thesis as a whole or parts incorporated in it may be deemed as plagiarism. For the comparison of my work with existing sources I agree that it shall be entered in a database where it shall also remain after examination, to enable comparison with future theses submitted. Further rights of reproduction and usage, however, are not granted here.

This thesis was not previously presented to another examination board and has not been published.

Garching, 15.12.2023 Vincent Derek Held

Abstract

A comprehensive understanding of business processes is crucial for digitizing these processes. The utilization of Business Process Model and Notation (BPMN) 2.0 process diagrams has emerged as a pivotal tool in both research and industry for representing and analyzing business workflows. An ever-growing number of regulatory documents and process execution data influences these processes.

This thesis presents the development of an advanced state-of-the-art method for converting natural language process descriptions into BPMN2.0 process diagrams. It focuses on assessing the advancements of recent developments compared to existing methodologies, particularly in processing complex regulatory documents such as ISO standards and data protection regulations. Additionally, this study explores optimal technologies for pre-processing textual inputs, efficient extraction of information, and effective visualization within a BPMN2.0 process model framework.

Keywords: *Natural Language Processing, Large Language Models, Business Process Compliance, Natural Language to Process, Business Process Model Generation*

Contents

List of Tables

List of Figures

Introduction

Motivation

Process modeling is a common technique for a better understanding and documentation of organizational processes and structures, as well as for process improvements and standardization [1]. Processes are not only used in companies, but they are also common within organizations and in research. Usually, lots of process documentation is stored in natural language [2]. Experience is necessary for process modeling, which creates a need for professional knowledge. BPMN2.0 is a standardized notation used to represent business processes graphically and is widely used in various industries and science. Process modeling is an elementary part of information systems and structural design. It is time-consuming, as 60% of business process management is spent on modeling [2], and the necessity for frequent updates due to ongoing process optimizations and regulatory changes [3], there is an increasing need for efficient modeling solutions.

Natural Language Processing (NLP) is a subfield of artificial intelligence that focuses on enabling computers to understand, interpret, and produce human language (natural language). By applying NLP techniques, this thesis aims to automate the extraction of process-relevant details from regulatory documents. Automated creation of process diagrams from textual descriptions is a promising approach to making processes more efficient and minimizing errors. This approach has already shown promising results in various application areas but is still considered unsolved [4].

However, limited research explicitly focuses on extracting process information from legal texts. With the increasing importance of business process compliance (BPC), especially concerning regulations such as the General Data Protection Regulation (GDPR) and the associated financial penalties for non-compliance, this area deserves more attention [5]. Typically, legal texts require interpretation by legal experts to integrate them into internal company processes. Automating this process by generating diagrams from legal documents could streamline the implementation of laws and standards and reduce errors. By comparing the reference models with those from company processes, implementation in companies and organizations could be simplified while reducing costs through improved productivity of process analysis and reduced dependency on legal advice.

With the combination of LLMs for pre-processing, rule-based NLP approaches for extracting relevant information from regulatory documents, and BMPN2.0 for visualization, our approach aims to improve understanding of regulatory documents, facilitate implementation through visualization, and minimize errors due to misunderstandings.

In summary, process modeling is an essential technique for understanding and improving organizational processes, but it can be time-consuming and requires specialized knowledge. Natural Language Processing (NLP) is a promising approach for automating process information extraction from textual descriptions. Automated process diagram creation has already shown positive results in various application areas. However, there is a need for more research specifically focused on the extraction of process information from legal texts. To address these gaps in the existing literature, this study aims to investigate the following research questions:

Research Questions

RQ 1: How can business process models be automatically generated from textual descriptions using NLP techniques? This research question aims to refine the current state approach to automatically convert textual process descriptions into graphical representations using NLP techniques and to investigate the methods and techniques that can be employed.

RQ 2: How do recent developments compare to existing methods? The second research question focuses on evaluating and comparing recent advancements in the field of automatic business process modeling, specifically those utilizing NLP techniques, with existing methods. It aims to assess the improvements, innovations, and potential advantages of these recent developments in terms of ease of use, model quality, and overall performance. By conducting a comparative analysis, this research question seeks to identify the strengths and limitations of the state-of-the-art approaches compared to established methods.

RQ 3: How well does the approach work concerning process descriptions from more complex regulatory documents (e.g., ISO Norms)? This question aims to assess the effectiveness and applicability of the proposed approach, specifically focusing on its performance when applied to more complex regulatory documents such as ISO Norms. It seeks to understand how well the NLP-based approach handles the challenges of intricate and detailed process descriptions within

these regulatory documents. The research will evaluate the accuracy and ability of the approach to transform complex regulatory language into meaningful and accurate BPMN2.0 process diagrams.

RQ4: Which technologies best visualize the extracted process information in a BPMN2.0 process model? This research question aims to explore the various technologies that can effectively visualize the process information extracted from textual descriptions in a BPMN2.0 process model. Once the relevant details have been extracted using Natural Language Processing (NLP) techniques, it is crucial to represent this information in a visually understandable and standardized format. The visualization of extracted process information plays a significant role, as it enables stakeholders to comprehend the process model easily and facilitates communication and collaboration among team members involved in process analysis, improvement, and implementation. The research will identify a best-practice technology for visualizing the extracted process information in a BPMN2.0 process model. Additionally, the research will explore any advancements or innovations in process visualization technologies that can enhance the representation and understanding of process models. Factors to consider include ease of use, flexibility in representing different process elements, support for BPMN2.0 notation, layout, and visualization quality. This understanding will contribute to improve the comprehension, implementation, and standardization of processes in organizations, ultimately leading to enhanced process efficiency, reduced errors, and better decision-making.

Contribution

This research aims to advance the state-of-the-art approach in transforming textual process descriptions into BPMN2.0 diagrams, building upon the work of Yu [4] and Friedrich et al. [2]. The main contributions of this work are:

1. **Enhanced scope of the evaluation:** This study pioneers by attempting to evaluate the baseline approach using regulatory documents, particularly ISO standards and articles from the General Data Protection Regulation (GDPR). This extends the assessment framework to cover diverse textual descriptions beyond traditional academic and industry boundaries.
2. **Identification of limitations:** By applying the SOTA to evaluation data consisting of process descriptions and legal texts, this study systematically identifies the underlying limitations of the baseline approach. These insights are crucial as they identify opportunities for improvement and adaptation, especially when dealing with texts that contain complicated legal nuances.

3. **Differential analysis:** The research analyses the differences between regulatory documents and traditional business process descriptions. Understanding these contrasts is essential as it illuminates the inherent complexity of regulatory documents and specifications that may not be present in standard process descriptions and provides insight into potential contextual adaptations.
4. **Strategic adjustments**: Given the limitations and differences uncovered, this work takes the next logical step by implementing possible strategies and methods for fine-tuning the automatic generation of business processes. Addressing the characteristics of legal texts provides a roadmap for refining and improving the tool's adaptability and accuracy by utilizing both rule-based and LLM functionalities.

In essence, this work is a bridge connecting business process modeling and regulatory documentation. It strives to improve the capabilities of the SOTA, highlight the specifics of regulatory text, and ensure that it is robust and versatile enough to handle a broader and more complex range of textual descriptions.

Through comprehensive assessment and detailed analysis, this study aims to strengthen the foundational work of [4] and ensure that it is equipped to meet the challenges of an evolving text landscape
.

Methodology

As an implementation project is the core of the thesis, we will follow the design science research (DSR) methodology proposed by Hevner et al. [6]–[8]. This guides the development and evaluation of our solutions to problems described in Chapter 1. The DSR approach is iterative and involves designing, building, evaluating, and refining an artifact until a satisfactory solution is achieved. This chapter will describe the DSR methodology we use in this study.

The DSR method consists of six stages, as follows:

1. problem identification and motivation
2. definition of the objectives for a solution
3. design and development
4. demonstration
5. evaluation
6. communication

In the following, it is explained how the six stages will be applied within the project:

Problem identification and motivation

In the DSR methodology, the initial step is to recognize the issue and establish the reason for the study. Research in the literature (as seen in the next chapter) indicates no solution for automatically generating process diagrams. Specifically, there is no investigation of the possibility of automatically generating process diagrams from regulatory documents. This study aims to develop an enhanced automated approach to extract relevant data from regulatory documents and create process diagrams as reference models. This will enhance understanding and implementation of regulations while also reducing mistakes.

Definition of the objectives for a solution

The second step in the DSR methodology is to define objectives. This thesis aims to develop an enhanced, proven method for automatically generating process models from a textual description using current technologies. For this purpose, given approaches will be evaluated, and the most advanced approach will be reconstructed. Here, the aim is to evaluate the accuracy, compared to the approach to be followed, as well as to minimize or eliminate errors and improve the output

quality. Additionally, this thesis focuses on analyzing the approach for creating process diagrams of regulatory documents and improving the implemented algorithm for this purpose.

Design and development

The third step in the DSR methodology is to design and develop the artifact. In our case, we develop an approach that leverages both classical rule-based NLP techniques and LLMs to extract relevant information from regulatory documents and create process diagrams (BPMN 2.0 models). The software prototype will be developed using Python programming language, Spacy library for NLP processing, different LLMs from OpenAI, and the BPMN 2.0 standard to visualize the resulting processes.

Demonstration

The fourth step in the DSR methodology is to demonstrate and evaluate the artifact. In our case, we demonstrated the developed Python code by using it to extract relevant information from a set of example regulatory documents and create process diagrams. Afterward, we evaluated the results of this proof-of-concept implementation based on the accuracy and completeness of the process models. The code will be published on GitHub and accessible to all stakeholders. Additionally, a set of natural process descriptions and human-modeled process diagrams, which have been used for the evaluation, will be published there, together with the output diagrams created by the algorithm.

Evaluation

The fifth step in the DSR methodology is to evaluate the artifact. The evaluation is an essential aspect of the research. It aims to qualitatively and quantitatively evaluate the results of our suggested implementations and compare them with the results of previous work in this research area to explore our work's strengths and limitations.

Communication

Finally, the last step is to communicate the results. The outcomes of this project will be presented in a scientific paper (thesis). As mentioned in the demonstration part of DSR, the code will be accessible with some examples in a GitHub repository. Additionally, this thesis will be communicated and presented in a thesis defense to all relevant stakeholders.

Evaluation

As emphasized in section 1.4, evaluation is a crucial component of the research process. It will demonstrate our method's effectiveness and highlight any limitations that may impact its application. The overall goal is to determine whether our approach can fulfill the requirements for the automatic generation of BPMN models from natural language input, focusing on both traditional descriptions and regulatory documents.

The evaluation will consist of a quantitative analysis and a qualitative review: For the quantitative part, we will derive metrics from comparing the approach with the approaches presented by [4] and [2]. A gold standard (GS) of model-text pairs, including traditional and regulatory text descriptions, will serve as a benchmark for this analysis. The results of the different approaches will be compared carefully with the GS diagrams. This step includes a manual count of the BPMN components, such as gateways, lanes, and nodes, to ensure a thorough and objective evaluation, followed by the calculation of different metrics. This comparison is performed using an evaluation matrix, which serves as the basis for our accuracy assessment.

In the qualitative evaluation, the quality of the task labels of the approach are discussed in comparison to the results of the current state of the art (SOTA) and the GS.

In addition, we will outline unresolved issues and challenges we encountered during the investigation. Based on these findings, we will suggest research questions for future research and possible improvements to enhance our approach further.

Structure

The structure of this thesis is as follows: Section 2 discusses related literature. Section 3 introduces the problem of this work. Section 4 describes the proposed methods to improve the accuracy of the generated process diagrams and adaptions to achieve improved results. Section 5 evaluates and compares the proposed approach with existing methods and evaluates our approach to the introduced dataset. Section 7 discusses limitations and validity considerations of the results, concluding with a summary and future research opportunities.

Related Work

A systematic literature review was conducted to research the current state of process generation. Table 1 presents the recommended search steps outlined by Achimugu et al. [9].

Literature Review

Upon receiving the research topic from the supervisor, the initial step undertaken was a comprehensive literature search. The Technical University Munich's University Library suggests adopting a Research Strategy Plan (RSP) to facilitate the identification of relevant search terms. Consequently, key terms such as "NLP", "BPMN", and "regulatory documents" were identified, along with their synonyms like "model" and "process". These terms were then strategically combined using "OR" and "AND" operators to refine the search process. Utilizing the RSP, specific keywords were determined, including "Process Model Generation from Natural Language Text" and "Extracting Business Process Models Using Natural Language Processing (NLP) Techniques". These keywords guided the literature search across various platforms, including dblp, IEEE, Springer, ACM, SCOPUS, and Google Scholar.

The selection process involved reviewing the titles, abstracts, and keywords and importing the chosen works into Zotero, a literature management tool. Valuable insights were extracted during the detailed review of each paper. Furthermore, pertinent references and related works were critically analyzed and incorporated into the literature corpus as deemed necessary.

Table 1
Steps of literature review [cf. [9]]

Search Steps	**Outcomes**
Identifying research	Research questions
Conduct literature search	Potentially relevant papers
Reading of title, abstracts, and keywords	Potentially relevant papers
Reading the full paper	Relevant papers
Removing duplicate entries	Unique and relevant papers
Extension by analyzing references	Final set of relevant papers

The literature review focused on identifying current state approaches for automatically generating process diagrams for textual descriptions and the processing of regulatory documents.

Information extraction and generation of process diagrams

Within the last years, multiple approaches have been developed [2], [10]–[12], and the relevance was outlined by multiple papers [13]. [12] conducted a state-of-the-Art analysis and compared the existing approaches. The most developed approaches will be outlined in the following:

The approach of [2] from 2010 is still considered state of the art and leverages Natural Language Processing (NLP) to extract information from textual process descriptions and create process diagrams in BPMN 2.0 format. Their paper elaborates on the methodologies implemented in detail, and the Java code used is made available on GitHub. The technique involves a dual-level analysis of the text: at the sentence level and the text level, identifying various challenges such as handling active-passive voice transformations, complex sentence structures, and anaphora resolution. To validate their approach, the authors compiled a dataset comprising textual process descriptions alongside their corresponding human-crafted diagrams. Despite its potential, the approach has certain limitations. These include the necessity for the description to be sequential, devoid of questions, containing minimal irrelevant process information, free of questions, and grammatically accurate.

In their work, [13] utilized data originally compiled by [2] to address the absence of a standardized, annotated dataset for business process descriptions. This led to the creation of the PET (Process Extraction from Natural Language Text) dataset, which annotates various elements in process descriptions, including activities, gateways, actors, and flow information.

[10] introduced a semi-automated approach for extracting process-relevant information from textual documents. This information is initially stored in a structured format within a spreadsheet, allowing users to make necessary edits. Following these modifications, a process diagram is then generated.

Extraction from regulatory documents

Business process compliance is a crucial aspect of business process management [14]. Ensuring process compliance requires consideration of a wide range of regulatory documents from both national and international legislators. These external requirements have a significant impact on business processes and must be carefully extracted from the documents. The requirements are then converted into internal requirements such as guidelines or guides or integrated directly into process models [3], [15]. While current approaches of information extraction focuses on texts, where actions are named explicit, this work focus on the extraction from regulatory documents, such as the GDPR,

where the understanding and implementation are enormous challenges [16]. Current approaches of information extraction form regulatory documents are presented by [3], [15], [17]:

The semi-automated approach presented by [15] focuses on assessing the degree of compliance between regulatory documents and their realization in natural language, while the methodology introduced by [3] facilitates the automatic evaluation of compliance between process models and regulatory documents.

Recently, [17] presented an approach of task extraction from textual process description by leveraging different Large-Language-Models. This approach does not offer the extraction of all elements extracted by [2] nor the possibility of visualization as an BPMN2.0 diagram.

These selected approaches constitute important contributions to the field of automatically generating process diagrams from textual descriptions or the automated assessment of the compliance between internal process documentations and a regulatory document. By examining and comparing their methodologies, strengths, and limitations, we can gain valuable insights and identify areas for improvement in our research to develop an approach of extracting information from a regulatory documents to create a process diagrams.

Solution Design

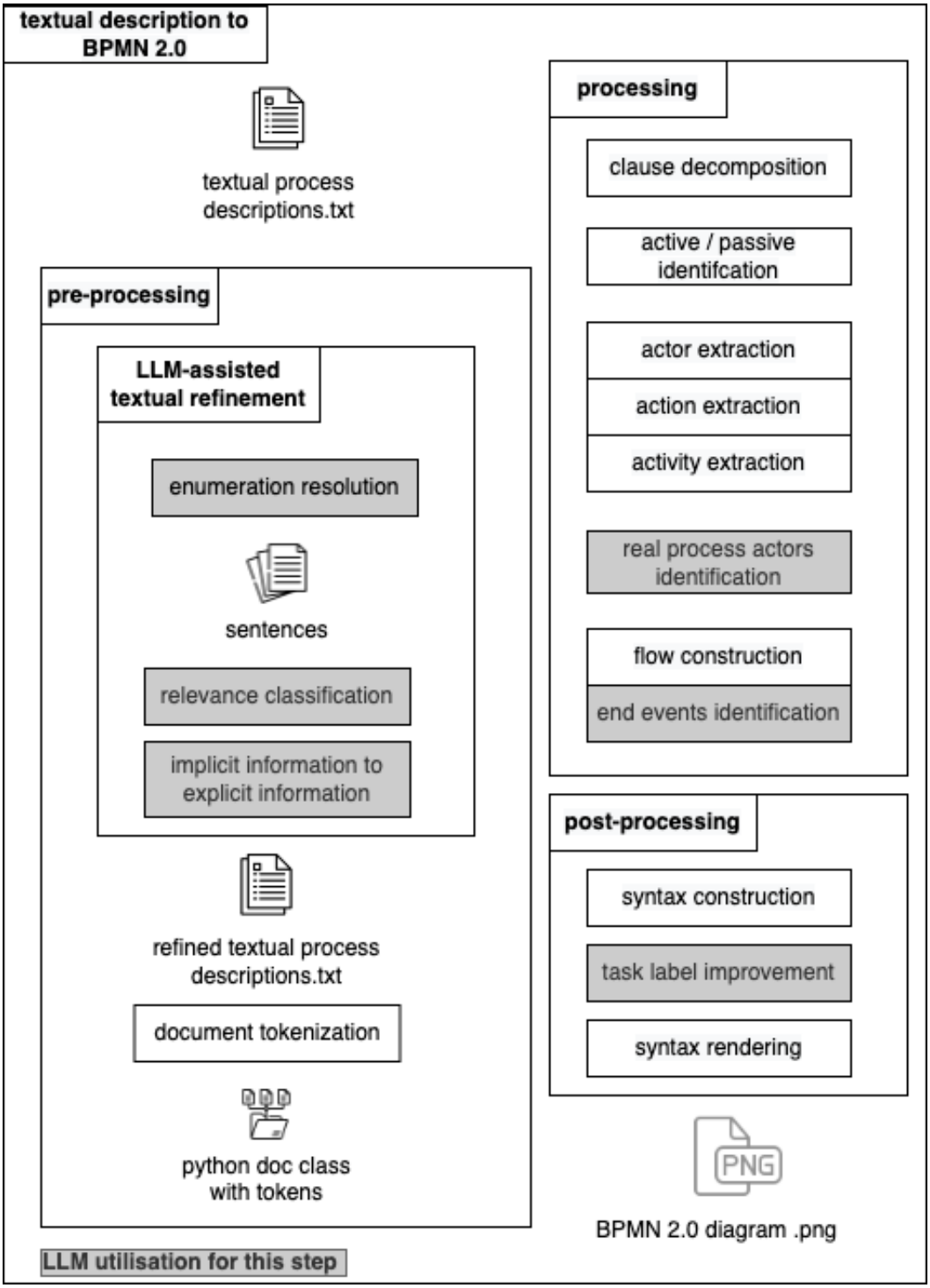
This chapter addresses the conceptual framework of our research and lays the foundation for the practical solutions we intend to develop. It serves as a crucial bridge between identifying key problems in automatic process extraction from natural language descriptions, focusing on regulatory documents, and the concrete application of these solutions.

First, this chapter focuses on identifying and formulating the core problems targeted by our research. This includes thoroughly investigating the challenges associated with the processing and interpreting of textual descriptions by automated systems. By breaking down these challenges, we aim to provide a clear understanding of the issues.

Second, this chapter presents theoretical approaches to solving these problems. Here, we set out the conceptual strategies and methods on which our proposed solutions are based. These approaches are based on current research and best practice. The solutions presented are aimed at traditional process descriptions and intended to advance the field of automated natural language processing in regulatory environments.

This chapter essentially forms the basis for the subsequent practical implementations. It provides the necessary context and theoretical basis for the practical work described in the Implementation chapter 4. Fig. 1 illustrates an overview of the steps of our approach and indicates the usage of LLMs.

Figure 1
Overview of the text2BPMN steps



Characteristics of regulatory documents

The objectives of this study include evaluating the effectiveness of automated extraction from natural language descriptions in regulatory documents, which requires a deep understanding of the unique characteristics of these documents. These characteristics play a crucial role in influencing the performance and effectiveness of automated extraction methods, especially in the context of processing the complexity inherent in regulatory texts.

One of the main challenges in processing regulatory texts using NLP systems is the complex and specialized lexicon that these documents often contain. The specialized terminology prevalent in such texts poses significant difficulties for NLP systems in terms of accurate comprehension and interpretation. Furthermore, the interpretation of legal texts is heavily dependent on **domain-specific knowledge**. The intricate nuances and jargon found in these documents require a comprehensive understanding of the domain in question, a requirement that standard NLP systems are often unable to fulfill [18].

Another factor that complicates the analysis of regulatory documents is their **format and structure variability**. This diversity challenges the adaptability of NLP algorithms and affects the consistency and reliability of information extraction across different documents. To illustrate this, we will discuss the particular challenge of enumerations in these documents and how their format variability directly impacts the effectiveness of NLP techniques.

Most existing NLP models are predominantly trained on general, publicly available web content, including non-technical texts such as news articles and general web content [19]. However, this training approach **overlooks the specialized nature of legal texts**, leading to a significant gap in the training of models. Therefore, the unique lexis, structure, and context of legal texts are areas where these models often fail, limiting their ability to process and interpret these documents effectively.

The previous sections have provided an overview of the specific characteristics of legal acts. In the following sections, the focus will be on explaining the methods used to take these unique characteristics into account to improve the quality of the results. To accomplish this, specific adaptations and innovations in the techniques used are described below, each carefully crafted to address the unique challenges of these documents while ensuring that the quality of the results

matches that of traditional process descriptions. For this purpose, the following text is divided into pre-processing, processing and post-processing.

Pre-processing

Pre-processing refers to the stage where raw data is transformed into a suitable format for analysis. Our approach considers textual process descriptions or regulatory documents as input. Such textual information are presented in unformatted plain text files (.txt). The complex language and structure of regulatory documents present a considerable challenge for conventional natural language processing (NLP) techniques, which are predominantly rule-based. To overcome these obstacles, the current study leverages a Large Language Model (LLM) as a tool for data pre-processing [20]. This innovative strategy not only improves the efficiency of data cleaning, but also serves as a fundamental step for more complex analysis.

Analytical resolution of enumeration structures in text

The functionality of rule-based extraction methods depends on processing texts that conform to standard grammatical structures, exceptionally complete sentences, and that do not contain unconventional textual formations such as bullets. This requirement emphasizes the need for a systematic approach to the resolution of enumerations in texts, as the following examples from the GS show:

(1) "The management review shall include consideration of:
a) the status of actions from previous management reviews; [...]
d) feedback on the information security performance, including trends in:
1) nonconformities and corrective actions; [...]
4) fulfilment of information security objectives;
e) feedback from interested parties; [...]"

(2) "Processing shall be lawful only if and to the extent that at least one of the following applies:
a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes;
b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract; [...]"

In Example 1, we observe an enumeration that delineates a series of actions, each requisite for the process under consideration. This necessitates the syntactic concatenation of enumerated items using the conjunction "and". Conversely, Example 2 presents an enumeration of implicit actions where fulfilling at least one criterion is sufficient. Here, the enumerated items should be concatenated using "or." These distinctions are not merely syntactic but bear significant implications for determining process flow and the type of decision gateway in the model. The correct interpretation and syntactical restructuring of these enumerations are crucial for accurate rule-based text processing and subsequent application in automated systems.

This analytical approach to enumeration resolution holds particular significance in regulatory documents, where precise interpretation of enumerated clauses is imperative for compliance and operational purposes. Regulatory texts often employ complex enumeration structures to outline mandatory actions, compliance criteria, and procedural steps, making accurate syntactical restructuring a critical component of rule-based text processing in these domains. However, the applicability of this approach extends beyond regulatory documents. It is equally pertinent in various other textual descriptions where the clarity of enumerated lists determines the effectiveness of information conveyance and subsequent action. In diverse contexts ranging from technical manuals to organizational procedures, the ability to correctly interpret and structurally reconfigure enumerative text is vital in ensuring accurate comprehension and implementation of the described processes.

Relevance of information

One of the most significant challenges of the SOTA approaches is the decreasing quality of the results with an increase of irrelevant information [2], [4]. To overcome this obstacle, an classification of irrelevant process information was developed. To address this challenge, a classification system, termed Irrelevant Information Criteria (IIC), was created specifically to identify and manage irrelevant process information. In the following the different IICs will be outlined:

1. Introduction sentence that describe the tasks of the company are not relevant and must be filtered. As seen in example 3 and 4 from the GS, the sentence does not name a process step, but it just describes that the overall goal of the company.

(3) "A small company manufactures customized bicycles."

(4) "The Evanstonian is an upscale independent hotel."

2. Information on the start or end of a process (instance) and not real process steps. While the information on the end of a process is relevant, as it helps to determine end events, start events are just classified as irrelevant information and need to be filtered. Therefore "a new process instance is created" should be filtered from example 5. As this sentence would not be a grammatically complete sentence, which leads to problems in the following rule-based approaches for the extraction, the sentence must be transformed into a new complete, but filtered sentence: "The sales department receives an order."

(5) "Whenever the sales department receives an order, a new process instance is created."

3. Information on the goals (ex. 6 - 8) or outcomes (ex. 9) of activities are considered irrelevant for the purposes of creating process diagrams. The inclusion of such details could lead to an overly precise and consequently overly detailed representation in the diagram, which is contrary to the intended clarity and conciseness.

(6) "to ensure valid results"

(7) "to ensure that the audit programme(s) are implemented and maintained"

(8) "as evidence of the results"

(9) "that are relevant to the information security management system"

(10) "The methods selected should produce comparable and reproducible results to be considered valid"

4. Information that clarifies that something is not universally applicable are not relevant and must be filtered.

(11) "as applicable"

(12) "if feasible"

5. Information on examples are not part of the process diagram, and therefore must be filtered from the text.

(13) "[...] e.g. 'Power limit exceeded by xxx watts' can be logged."

6. Information on references to other articles or paragraphs are not relevant and must be filtered from the text.

(14) "referred to in Article 22(1) and (4)"

(15) "in accordance with Article 55"

As described above, careful classification and filtering of irrelevant process information is crucial for improving the quality of process diagrams created from textual descriptions. By systematically eliminating non-essential elements such as introductory sentences about business tasks, start events, goals or results of activities, conditional statements, examples, and references to other articles or paragraphs, we can significantly optimize the creation of process models. This approach addresses a fundamental limitation identified in current state-of-the-art methods, where the presence of irrelevant information inversely affects the quality of the resulting diagrams [2], [4].

By refining the input data in this way, we strive to improve the accuracy and relevance of the generated process diagrams and contribute to developing more robust and efficient natural language processing techniques in business process management. Ultimately, this strategic filtering aligns with our overarching goal of creating clear, concise, and precise process diagrams that accurately reflect the intended business processes, free from excessive detail.

Implicit information

Due to the complexity of the natural language, information on actions is not always named explicitly. Implicit information is often use to provide descriptive information [21].

(16) "The Room Service Manager then submits an order ticket to the kitchen to begin preparing the food."

While the first action, "submit an order ticket to the kitchen," is explicitly stated, the second action, "the kitchen prepares," is implicit. The sentence mentions "begin preparing the food," which implies the kitchen's role, but it doesn't explicitly state that the kitchen is performing the

action. Understanding the causal relationship between the two actions requires a level of semantic understanding that goes beyond simple syntactic parsing. Rule-based systems typically follow predefined syntactic patterns and might miss the nuance that submitting the order ticket leads to the kitchen beginning food preparation.

(17) "Documented information shall be available as evidence of the results of management reviews."

The action "to document information" is implied rather than explicitly stated by using the participle. The sentence focuses on the availability of documented information rather than the act of documenting itself. Rule-based systems usually depend on explicit key words and might miss actions that are not directly mentioned. Understanding that "documented information" implies an action of documentation requires contextual knowledge. Rule-based systems typically lack the ability to infer actions based on context, as they operate on literal interpretations of text.

Therefore the challenge of identifying and interpreting implicit information in natural language texts is crucial for creating comprehensive process diagrams. As the examples show, natural language often conveys critical process steps implicitly rather than explicitly stating each action. Human readers usually infer these implicit actions based on textual clues. This characteristic poses a major challenge for rule-based systems, which traditionally rely on explicit actions and syntactic patterns and often overlook the complex implications inherent in natural language.

To bridge this gap, it is essential to integrate advanced semantic understanding into our process diagramming methodology. This means going beyond mere syntactic parsing and delving into the realm of semantic analysis, where context and implicit meanings are taken into account. Such an approach enables the capture of both explicit and implicit actions and ensures that the generated process diagrams are accurate and holistically represent the entire workflow.

The pre-processing plays a central role in our approach to automatic process extraction from natural language descriptions. In this stage, we look at the steps of the refinement and preparation of textual process descriptions or regulatory documents into a format suitable for further analysis. Key points include the analytical resolution of enumeration structures, filtering information based on the IIC, and transformation of implicit information to explicit information. All these steps are collectively

referred to as "LLM-Assisted Textual Refinement (LLM-ATR)," representing a comprehensive approach that integrates advanced language model capabilities for optimizing the text data before processing. Additionally, this pre-processing phase incorporates traditional methods implemented in the SOTA approaches, including cleaning white spaces and line breaks within the text, to refine the input data further for subsequent processing stages.

Processing

The processing phase in this research is characterized by applying a rule-based extraction methodology, primarily based on the principles and techniques of the current SOTA. This phase is of central importance as it transforms the refined input data from the pre-processing phase into a structured format for visualization in the post-processing stage. In the following sections, the limitations of this phase, identified through careful analysis, are critically examined. Our research systematically addresses these limitations, focusing on improving the efficiency and accuracy of the rule-based extraction process by leveraging innovation in the area of NLP.

Identification of real actors in process descriptions

The SOTA approach identifies some actors, still there are various lanes in the output, that have not been labeled. This comes for several reasons, which will be outlined in the following:

In the underlying approach, the actors are identified based on the dependency labels and stored together with the associated actions. Before the data is visualized, a distinction is made as to whether the identified actors are "real actors". A "real actor" is defined as any entity that performs an action, including individuals, software systems, organizations, or departments. In addition, entities that fulfill the role of a subject in the text are also categorized as actors. However, they are referred to as "non-real" actors if they do not fulfill the main criterion of an acting entity [2]. Based on this categorization, the actors are used for the generation of lanes in the resulting BPMN diagram. Examples 18- 20 represent actors, that have not been identified, due to false categorisation.

(18) "supervisory authority"

(19) "central system"

(20) "data subject"

To improve the accuracy of actor identification in process diagrams, this study proposes two key changes: refining the definition of a "real actor" and incorporating a Large Language Model (LLM) into our methodology.

The existing definition of a "real actor" in process diagrams is somewhat restrictive and cannot cover the diversity observed in our gold standard dataset. To change this, we propose to expand the definition to include not only people, software systems, organizations, and departments but also places, locations, professions, and occupations. This broader definition allows for a more comprehensive identification of actors and captures a wider range of entities involved in process activities.

The integration of a LLM is critical to realizing this broader definition. LLMs have the ability to adaptively learn from large data sets, which allows them to recognize nuanced distinctions when categorizing actors. This feature is particularly valuable in complex scenarios that can challenge traditional rule-based methods.

By combining an LLM with the revised definition of a "real actor", our approach aims to significantly improve the actor identification process. It is expected that this integration will lead to a more comprehensive and accurate representation of the various entities that act as actors in process diagrams. Consequently, the resulting BPMN diagrams will more accurately reflect the complexity they are intended to model, increasing both their reliability and usefulness in various applications.

Enhancing actor accuracy through similarity analysis

The creation of BPMN diagrams requires the creation of a list of actors derived from extracted information. This process involves cataloging all entities classified as "real actors" in a unified list. Then, in the post-processing phase, each actor from this list is represented by a corresponding lane in the BPMN diagram. Although the current state of the art (SOTA) has proven that it can identify different actors in process diagrams (see figure 2), the nuances of the language pose a significant challenge, especially when different terms can refer to the same entity. Consider the following examples:

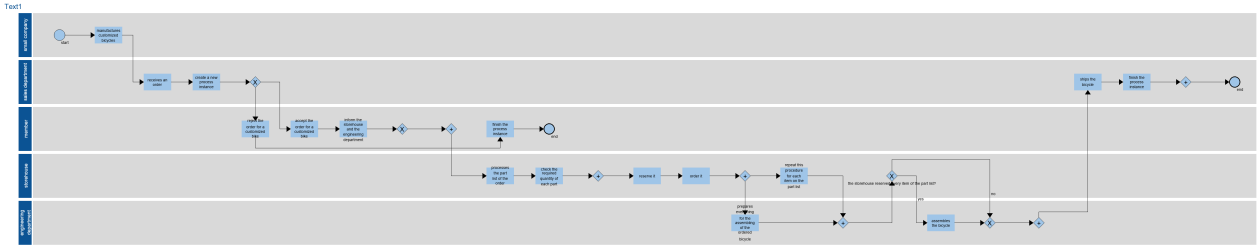
(21) "member of the sales department"

(22) "sales department"

In process diagrams, departments are often symbolically represented by a single member. In the context of the above examples, both expressions conceptually refer to the same entity - the sales department. This example illustrates the linguistic complexity associated with identifying actors in process diagrams.

To address this challenge, our approach includes an innovative strategy for analyzing and comparing the similarities in the names of the different actors. When the actors are identified as similar, their respective lanes are merged in the diagram. This method aims to improve the accuracy and precision of the resulting BPMN diagrams and ensure that they better reflect the true structure and dynamics of the underlying process. By implementing this approach, we seek to reduce linguistic ambiguity and improve the clarity and efficiency of our process diagrams.

Figure 2
BPMN2.0 diagram generated by Yu's approach



Leveraging modal verbs for improved clarity in process task labels

Auxiliary verbs, commonly referred to as helping verbs, play a central role in the formation of different sentence tenses and structures through the use of different forms of "to be", "to have" and "to do". Although these verbs are not used for the task labels, in certain contexts, they are crucial to extract and for the subsequent visualization of complete action labels in process diagrams. The SOTA successfully extracts auxiliary verbs necessary for accurately representing actions with adjectives such as "be available". However, this approach is unable to capture the nuances of modal auxiliary verbs - such as "can", "will", "shall", "may", "must", "should" and "could" - that are essential for representing capabilities, opportunities, permissions, and obligations. This restriction is particularly significant in legal documents, where the inclusion of modal verbs is of great relevance [3], [22].

Recognizing the importance of these linguistic elements, our approach incorporates modal verbs into the task labels, thereby capturing a fuller range of implied actions and responsibilities within the process diagrams.

Context-driven identification of end events

Current state-of-the-art approaches (SOTA) typically follow a structured methodology in determining end nodes. After extraction, the information is categorized into structures, which can be individual activities or blocks of activities. These blocks represent what are often referred to as condition blocks for exclusive activities or AND blocks for parallel activities within the diagram. The structures are stored in a list.

The identification process involves iterating through each structure, with the last structure in the list being designated as the end activity (is_end_activity = true). For conditional blocks, a similar iteration occurs within each branch to determine the end of a process. This determination is supported here by hypernyms, where the verb or its synonym is categorized under "end" or "refuse".

While this method provides a basic framework for determining the end of an activity, it is simplistic. Also, other actions can imply the end of a process based on the context.

(23) "A customer brings in a defective computer and the CRS checks the defect and hands out a repair cost calculation back. If the customer decides that the costs are acceptable, the process continues, otherwise she **takes her computer home unrepaired**. The ongoing repair consists of two activities, which are executed, in an arbitrary order. The first activity is to check and repair the hardware, whereas the second activity checks and configures the software. After each of these activities, the proper system functionality is tested. If an error is detected another arbitrary repair activity is executed, otherwise the repair is finished."

Consider the example in 23: The process is based on the actor (CRS) repairing the PC. The process ends the moment the customer takes the PC home. However, in another process context, such as a retail environment, the customer taking an item home could be an intermediate step and not the end point of the process. Therefore, it is crucial to include contextual understanding in the decision-making process to identify end activities.

To achieve this, we propose extending the existing approach by analyzing the process context more sophisticatedly. This involves looking at the activities themselves, the surrounding story, and the role of each activity within the broader process flow. In this way, we aim to achieve a more accurate and context-aware identification of end activities and improve the overall reliability and applicability of process diagramming.

Post-processing:

The post-processing phase focuses on the presentation and visualization of data obtained using natural language processing (NLP) techniques. The aim here is to convert this information into a format that is visually understandable and complies with established standards.

Business Process Model and Notation 2.0

The Business Process Model and Notation 2.0 (BPMN) [1] has established itself as the industry standard for business process modeling [23]. Designed by the Object Management Group, BPMN serves a dual purpose: it supports both technical and business users in business process management. Its design is intuitive for those involved and simultaneously able to encapsulate complex process semantics. The fundamental benefit of BPMN lies in its ability to graphically represent internal business processes and thus create standardized documentation for (corporate) processes.

BPMN comprises a large number of process elements that are described in detail in [24]. These essential elements are listed and explained in the following: **Start Events** and **End Events** are used to indicate where a particular process starts or ends. As business processes can be subdivided into sub-processes, BPMN offers the possibility to subdivide the diagram with sub-process. In this work, processes are not divided into sub-process and all activities are presented within one process. **Activities** represent work that is performed in a process by a company or organization. **Sequence Flows** are used to show the order in which activities will be performed within a process. The divergence and convergence of sequence flows in a BPMN process diagram is controlled by **Gateways**. They are used to fork and join sequence flows in the diagram. **Exclusive Gateways** are used to represent XOR-cases. Based on the decision criteria, only one branch will be executed. **Parallel Gateways** are used if both branches are executed in parallel (AND cases). Finally, actors are represented by **Pools**.

[1] https://www.bpmn.org

The visualization of extracted process information plays a significant role, as it enables stakeholders to comprehend the process model easily and facilitates communication and collaboration among team members involved in process analysis, improvement, and implementation. In contrast to other research areas, where the methods and libraries of data visualization are well established and formulated, there needs to be a consensus on standardized visualization strategies in the automatic creation of process diagrams. This divergence highlights a notable gap in the existing body of knowledge in this particular area of research.

Graphviz

Graphviz[2] is an open-source graph visualization software package known for its robustness in creating and editing graph structures. However, Graphviz offers no native support for specific Business Process Model and Notation (BPMN) elements. Instead, it provides a flexible framework in which various shapes can be customized to approximate BPMN elements. This customization leverages Graphviz's existing shapes and structures to represent the conventional symbols and notations associated with BPMN 2.0 but with certain limitations.

A primary limitation of Graphviz in this regard is its limited flexibility in adapting to BPMN 2.0 standards. The accurate representation of the essential elements as described in 3, in addition to further BPMN-specific elements, poses a challenge. Graphviz's visual results in rendering (arrangement of elements) are not as satisfactory as those produced by other approaches for this purpose [25]. The results can be exported in various formats, such as PNG, PDF, or SVG, to suit the specific requirements.

Process Piper

Process Piper [3] is an open source python library to generate business process diagram using python code or structured text. This library supports the following components of the business process diagrams [26]:

- Pool(s)
- Lane(s)
- Elements:

[2] https://graphviz.org/
[3] https://github.com/csgoh/processpiper

- Events (Start, End, Intermediate, Timer, Messages, Signals, Conditional and Link)
- Activities (Task, Subprocess)
- Gateways: Inclusive, Exclusive, Event

Consequently, Process Piper includes all the essential elements necessary for this research and provides opportunities to include additional elements.

Creating a process diagram with Process Piper is a two-step approach: first, an input syntax is formulated based on the extracted information. The library's rendering function then renders this syntax. Once the rendering process is complete, the model is saved as a PNG file.

Enhancing rendering quality through syntax optimization

In the SOTA approach by Yu [4], representation errors arise due to the nature of the extracted information that forms the basis of the process diagram syntax. This information is crucial for the construction of lanes, nodes, and gateways, where the following syntax notations are followed:

1. Start Nodes: "(start) as start_id"
2. End Nodes: "(end) **as** end_id"
3. Task Nodes: [individual_task_label] **as** activity_id
4. Lanes: "lane: individual_lane_name"
5. Parallel Gateway: "<@parallel> **as** gateway_id" and "<@parallel> **as** gateway_id_end"
6. Exclusive Gateway: "<> as gateway_id" and "<> **as** gateway_id_end"

Afterwards, the different created elements are connected, as shown in 26

The general syntax creation is accurate, but rendering problems often arise from the length and composition of the individual task description. Excessively long task descriptions can lead to rendering problems. In addition, task descriptions, which are ideally concise and contain a verb and an object, require careful drafting. In legal contexts, including a modal verb is crucial, as discussed in 3. Our research, therefore, attempts to make the wording of task labels more concise.

As indicated in the syntax notations, the term "as" is used as a keyword to define BPMN elements. However, problems arise when this keyword appears in contexts other than its intended use in the syntax. Especially in task labels, "as" can occur by mistake and is often used for comparisons or

explanations. In legal documents, "as" often appears in sentences such as 24 or in explanatory contexts such as 24. These cases can lead to ambiguities and complications in the representation, so a strategic approach is required to deal with such occurrences in generating BPMN elements.

(24) "as applicable"

(25) "as evidence of the implementation"

To overcome these problems, our approach involves a subsequent improvement of the generated syntax to refine the task labels to ensure clarity and conciseness in the task descriptions and to undermine the use of "as". This enhancement aims to improve the rendering process and the overall quality of the generated process diagrams.

(26) "title: text1_our_approach

width: 10000

colourtheme: BLUEMOUNTAIN

lane: sales department

(start) as start

[receives an order] as activity_12

<> as gateway_1

[accept the order for a customized bike] as activity_2

[can reject the order for a customized bike] as activity_3

(end) as end_3

<> as gateway_1_end

[inform the storehouse and the engineering department] as activity_14

[ships the bicycle] as activity_19

(end) as end

lane: storehouse

[processes the part list of the order] as activity_15

[check the required quantity of each part] as activity_16

[reserve it] as activity_17

[order it] as activity_18

<@parallel> as gateway_4

[repeat this procedure for each item on the part list] as activity_5

<the storehouse reserved every item of the part list?> as gateway_7

lane: engineering department

[prepares everything for the assembling of the ordered bicycle] as activity_6

<@parallel> as gateway_4_end

[assembles the bicycle] as activity_11

<> as gateway_7_end

start->activity_12->gateway_1

gateway_1-""->activity_2->gateway_1_end

gateway_1-""->activity_3->end_3

gateway_1_end->activity_14->activity_15->activity_16->activity_17->activity_18->gateway_-

4

gateway_4->activity_5->gateway_4_end

gateway_4->activity_6->gateway_4_end

gateway_4_end->gateway_7

gateway_7-"yes"->activity_11->gateway_7_end

gateway_7-"no"->gateway_7_end

gateway_7_end->activity_19->end"

Implementation

This chapter deals with the practical aspects of our research and describes implementing an enhanced automated approach that converts natural language process descriptions into BPMN2.0 process diagrams. Here, we proceed from the theoretical framework and methods outlined in the solution design chapter to their concrete application and show how these concepts are operationalized in a real-world context.

The chapter is methodically structured to go through each phase of the implementation process. It starts with a general overview of Large Language Models (LLMs), which play a crucial role in our approach—followed by sections for the three processing stages: pre-processing, processing, and post-processing.

The code associated with this thesis has been made publicly available on GitHub [4].

Leveraging Large Language Models

The use of OpenAI's Large Language Models (LLMs), including text-DaVinci-003, gpt-3.5-turbo-instruct, gpt-3.5-turbo, and gpt-4 [5], has been a cornerstone of our approach to improving process extraction from natural language descriptions.

GPT-Instruct models, a specialized subset of the GPT series, are adapted to interpret and execute the user's instructions more accurately [27]. This customization makes them particularly well suited for producing output closely aligned to specific instructions and constraints. This feature is valuable for tasks that require high precision and strict compliance with instructions. The fine-tuning on following instructions of these models reduces the probability of producing irrelevant or off-topic content and ensures that the output remains focused and relevant, especially when intended for subsequent automatic processing without human review or modification.

Several key aspects of using LLMs to optimize process extraction were considered as part of this research:

[4] https://github.com/VincentDerekHeld/thesis-bachelor-text2BPMN
[5] https://platform.openai.com/docs/models

Importance of correct spelling: Correct spelling of the prompts is critical because LLMs, including GPT models, rely on text tokenization. Incorrect spelling can lead to incorrect or unexpected tokens, which can have a negative impact on the model's understanding and processing of the text. Ensuring correct notation is, therefore, critical to the integrity of the processed data.

Strategic use of keywords: Including keywords such as "carefully" and "please" in prompts can influence the model's response. These keywords can prompt the LLM to focus more attention on the nuances of the task, which can lead to more accurate and considerate results. This technique can be particularly effective when aligning the model's responses more closely with the desired outcome.

Setting the temperature of instruct models: the "temperature" parameter in LLMs, especially in instruct models, plays a crucial role in controlling the randomness of the output. Our approach sets the temperature to zero, ensuring that the model generates relevant and coherent content while minimizing the risk of unexpected or divergent responses.

Splitting tasks into several steps: Dividing complex tasks into multiple steps is another strategy that significantly improves the effectiveness of the extraction process. Breaking a large task into smaller subtasks produces more accurate and precise results.

Step-by-step problem solving: Instructing the LLMs for step-by-step problem solving improves the quality of the outcome. Therefore, in this approach, complex tasks are broken down into smaller, manageable steps so the LLM can tackle each part. Such a methodical approach is helpful when dealing with complicated process descriptions and ensures systematic classification and transformation of information.

Despite its previously established effectiveness, the strategic decision to phase out the use of the DaVinci-003 model is in line with our commitment to a sustainable and future-oriented research methodology. This decision considers OpenAI's planned discontinuation of specific older models, including text-DaVinci-003, after January 4, 2024 [28]. By adopting newer and more advanced models, we ensure that our research remains at the forefront of technological advancements and continues to deliver robust and state-of-the-art solutions for process extraction from natural language descriptions.

Pre-Processing

The pre-processing phase in our research represents a significant improvement over the traditional SOTA methods commonly used for text pre-processing. While SOTA approaches mainly focus on rudimentary tasks such as removing brackets and merging multiple spaces using simple regular expressions, our method introduces a more sophisticated level of text refinement.

This advanced level of preprocessing, called "LLM-Assisted Textual Refinement" (LLM-ATR), is crucial for preparing the input data for accurate process extraction and visualization. It involves several key processes aiming to address specific complexities associated with natural language processing, considering regulatory documents as input.

LLM-ATR comprises several important components: Enumeration resolution (enabled when enumerations are present), filtering irrelevant information, and converting implicit actions into explicit actions. Each component of this methodology, essential for refining the process extraction text, is described in detail in the following sections.

Together, these elements of LLM-ATR represent a comprehensive approach to pre-processing that ensures input data is optimally prepared for the difficult task of transforming natural language descriptions into structured BPMN diagrams. By addressing the nuances and complexity of the language in this first phase, we set the stage for more effective and accurate process modeling in the later phases.

In addition, our approach includes a robust method for opening and reading files, complete with error-handling mechanisms to resolve potential issues such as "file not found". This functionality ensures the smooth ingestion and processing of text data and provides a solid foundation for the subsequent analysis steps.

Analytical resolution of enumeration structures in text

As explained in chapter 3, the effectiveness of rule-based extraction methods is fundamentally dependent on the structure of texts, as syntax and dependency trees are generated based on the structure, which serves as the basis for rule-based extraction approaches. In order to ensure a structure that conforms to a standardized grammatical framework, modifying the structure of the

text input to meet these requirements is a consideration that becomes particularly clear in the context of legal documents. Consider the following text segment from the GS:

(27) "The organization shall determine:
a) what needs to be monitored and measured, including information security processes and controls;
b) the methods for monitoring, measurement, analysis and evaluation, as applicable, to ensure valid results. The methods selected should produce comparable and reproducible results to be considered valid; [...]
f) who shall analyse and evaluate these results."

In this instance, SpaCy's default sentencizer segments the text into two distinct sentences (as shown in 28 and 29), primarily because it identifies sentence boundaries based on the presence of full stops.

(28) "The organization shall determine: a) what needs to be monitored and measured, including information security processes and controls;
b) the methods for monitoring, measurement, analysis and evaluation, as applicable, to ensure valid results."

(29) "The methods selected should produce comparable and reproducible results to be considered valid; [...]
f) who shall analyse and evaluate these results."

A custom sentencizer for integration into the SpaCy pipeline was developed to address this challenge. This custom sentencizer (CS), positioned strategically before the parsing phase, is designed to dissect enumerative text structures effectively, splitting each list item into a separate sentence for enhanced analyzability, as illustrated in 30-33.

(30) "The organization shall determine:"
"a) what needs to be monitored and measured, including information security processes and

controls;"

(31) "b) the methods for monitoring, measurement, analysis and evaluation, as applicable, to ensure valid results."

(32) "The methods selected should produce comparable and reproducible results to be considered valid; [...]"

(33) "f) who shall analyse and evaluate these results."

The CS utilizes the "token.tag_" attribute assigned to each token during the POS tagging process. This attribute, more detailed than the "token.pos_" attribute, enables the identification of specific textual elements like bullet points ("LS" tag). By leveraging these tags, the sentencizer can correctly identify enumerative structures, and the token can be marked as the start of a new sentence.

However, despite the structural improvements rendered by the CS, certain sentences (e.g., 33) still lack completeness in terms of grammatical structure, missing elements such as the verb and subject.

To bridge this gap, the proposed solution encompasses a hybrid approach, integrating a Large Language Model (LLM) with rule-based techniques. This approach begins with a LLM pre-processing the text prior to rule-based processing. For text containing enumerations (identified by the presence of the "LS" tag in a SpaCy doc file), the LLM is implemented to restructure these enumerations into grammatically complete sentences aligned with the input text's structure. This combination of rule-based methods and Large Language Models (LLMs) is strategically designed to increase both generalizability and accuracy when processing a wide range of text formats. In particular, the integration of LLMs is tuned to consider the nuances of enumerated texts. A key observation driving this approach is the tendency of LLMs to generate off-topic responses when instructed to resolve bulleted lists in texts that lack such structures. By linking LLMs with rule-based techniques, the system efficiently identifies and processes only those texts that contain enumerations, minimizing the likelihood of irrelevant or redundant output. This synergistic approach ensures a more targeted and context-appropriate application of LLM capabilities and increases the overall robustness and reliability of the results.

Relevance of information

As described in section 3, the accuracy and effectiveness of the generated process diagrams depends crucially on the fact that only process-relevant information is included. Existing studies show a direct correlation between the amount of input data and the resulting quality of the diagrams [2], [4]. Therefore, this approach's selective extraction of relevant information is a crucial aspect.

Our methodology extends conventional **machine learning algorithms**, usually suitable for classifying large amounts of data. These algorithms typically require large datasets to train to distinguish relevant from irrelevant content effectively. However, our approach faces a notable challenge: the lack of annotated datasets that include process descriptions and regulatory documents to classify relevant and irrelevant information. This limitation complicates compiling a comprehensive training dataset, which is crucial for effectively applying traditional machine learning techniques in our context. Given this limitation, alternative strategies must be explored to effectively filter out irrelevant information to preserve the integrity and utility of process diagrams. These strategies may include more nuanced, context-dependent approaches that can effectively recognize process-relevant information even without large, annotated data sets.

In order to solve this problem, we use a **LLM** for selective information filtering. We use the concept of "few-shot learning", a concept developed to overcome the limitations of limited training data. This approach allows the model to generalize from a minimal number of examples, enabling accurate predictions and fast adaptation despite limited training instances [29].

The specific task of classifying information as relevant or irrelevant is formulated as **M-way K-shot classification problem**. In this setting, M denotes the **total number of different classes**, while K denotes the **number of examples** (or "**shots**") **per class** used for training. The performance of the few-shot classification model is naturally influenced by these two factors, namely the number of classes (M) and the number of shots (K) [30].

The primary objective of this stage in the process is to filter out non-essential elements from the input text meticulously. To achieve this, we have delineated a specific set of **Irrelevant Information Criteria (IIC)**, as explicated in section 3.

Addressing the challenge of prompt engineering in this context, a variety of approaches were evaluated to effectively filter irrelevant information from the process description:

1. Employing a single prompt that incorporates the IIC alongside the complete input text.
2. Utilizing a single prompt with the IIC and the entire input text, supported with shots.
3. Implementing multiple prompts, each dividing the IIC and using the entire input text, enhanced with shots.
4. Applying multiple prompts that split both the IIC and the input text into individual sentences, and incorporating shots.

In the second approach, the classification challenge is to manage a variable, M, which is the sum of the number of IIC categories plus an additional category for relevant information. The LLM has the task of categorizing each text segment into a specific IIC category or as relevant. An important observation in this context is the inverse relationship between the length of the input text and the quality of the output of the LLM.

The relationship between improved output quality and two key factors - the volume of input data (denoted I) and the number of irrelevant information criteria (IIC) classes - is direct and significant. This correlation emphasizes the need to strike a harmonious balance between the volume of input data and the complexity of classification tasks.

Our refined fourth approach uses SpaCy's sentence segmentation features to split the text into individual sentences. At the same time, the IIC is decomposed, mapping each specific criterion to a specific prompt. This strategy culminates in a binary classification task that splits the classification process into two main categories. The task is to decide whether the information in each sentence fits into one of the classes of the IIC and is therefore classified as irrelevant or whether it falls outside these parameters and is therefore classified as relevant.

After classification, the LLM enters the filtering phase and systematically removes content that is classified as irrelevant based on the IIC framework. In scenarios where a sentence consists entirely of information classified as irrelevant according to the IIC, the LLM generates an empty response. It ignores this sentence for further IIC categorization and moves on to the following sentence. Suppose a sentence contains a mixture of relevant and irrelevant information. In that case, the LLM filters out

the irrelevant parts and ensures that a coherent and complete sentence containing only the relevant content is returned. The sentence is further processed for all IIC categories.

Determining the optimal number of examples (K) for each class is a crucial component of this process and depends on the specifics of the classification task. The chosen value for K must encompass the diversity within each class while maintaining a balance to prevent overfitting or underfitting [30]. The variability inherent in our study information precludes a universal approach for K. However, our methodology is designed to minimize it and consider the limitations of the available training data.

This approach results in the LLM filtering out irrelevant information from the text descriptions, increasing the process's precision.

The prompts for this filtering process are based on detailed descriptions of the IIC accompanied by several representative examples or "shots".

Example 34 illustrates a prompt with background information, including a comprehensive description of the IIC, along with two illustrative examples. Example 35 describes the sequential method used in all filtration tasks. These prompts are methodically combined with each sentence of the input text, which enables the targeted exclusion of irrelevant introductory sentences from the corpus, with a minimum of training data (shots), thus optimizing the relevance and accuracy of the processed text.

(34) "**Background Information:**
Introduction sentence that describe the company are not relevant and must be filtered.
Example: The Sentence "A small company manufactures customized bicycles." must be filtered.
Example: The Sentence "The Evanstonian is an upscale independent hotel." must be filtered.

(35) "**Instruction:**
1) Decide carefully based on the provided background if a part of the following sentence fulfills the conditions of the provided background information. If the condition is fulfilled, go to 2), else go to 3).
2) Filter carefully the information, that fulfills the condition, from the text (but sill return

full sentences) or if the sentence consists only out of this irrelevant information return an empty message (just a spaces without any other characters).
3) Return the text carefully without any changes or interpretations.

Consequently, this method results in a refined text output containing only relevant process-related information that can be integrated into the subsequent stages of process diagramming. After converting implicit information into explicit information, the refined sentences are combined into a complete text description containing only the process-relevant information.

Implicit information

In the complex communication domain, information transmission often goes beyond explicit statements. As emphasized in chapter 3, the inherent complexity of natural language makes it necessary to identify implicit information and convert it into explicit forms before applying rule-based extraction methods. This transformation is crucial for the creation of comprehensive and accurate process diagrams. A similar approach is used to filter irrelevant information to simplify this task. The input text is segmented into individual sentences and analyzed step by step with a Large Language Model (LLM):

1. **Step:** The LLM analyzes the attached sentence to determine if a part contains any implicit actions. If it contains implicit actions step 2 shall be executed, else step 3.

2. **Step:** The LLM is instructed to transform the implicit action into explicit actions following two conditions: The original order and the wording of the sentence must be retained. Afterwards the resulting sentence shall be returned.

3. **Step:** The LLM is instructed to return carefully the full input sentence without any changes and interpretations.

Given the tendency of LLMs to give detailed answers that match the human communication style, it is essential to give clear instructions to the model. These instructions should emphasize the

need for concise answers and compliance with the task parameters, thereby avoiding unnecessary modifications or interpretations of the sentence content [29], [31].

Illustrative examples, such as 16 and 17, are built into the prompt to help the LLM identify and transform implicit information.

After the LLM-ATR, all refined sentences are reassembled into a complete text and saved in a .txt file format. This edited text then undergoes further processing based on the methodology described by [4]. This preparation includes removing redundant spaces and line breaks, which optimizes the text for subsequent analytical processing. This sequence of steps ensures that the final text is prepared for accurate and effective analysis and sets the stage for creating detailed and accurate process diagrams.

Alternative approach: Removal of Introductory Sentences

For filtering introductory sentences, similarity metrics prove suitable. This technique involves calculating the semantic similarity between different text units, ranging from single tokens to collections of tokens known as spans. At the core of this process is cosine similarity, a technique that quantifies similarity and yields values ranging from zero to one.

A cosine similarity score of one denotes maximal similarity, suggesting that the vectors representing the compared textual units are either parallel or perfectly aligned within the semantic vector space. This high score indicates a substantial semantic overlap between the text units. It is typically observed when a token or span is compared with itself or when the contents are exceedingly similar. In contrast, a score of zero signifies that the vectors are orthogonal in the multidimensional space, reflecting a complete absence of semantic similarity between the compared textual units. It is crucial to understand that a zero score does not suggest a negation of content but rather a lack of semantic correlation as interpreted by the model.

This method allows for the assessment of semantic similarity between the first sentence of a text and the text as a whole. The first sentence is used for this because it can be an introductory sentence, as exemplified in 3 and 4. This assessment determines the extent to which the introductory sentence contains the semantic equality of the entire text, enabling efficient and targeted filtering of introductory content that does not have any value to the process.

The similarity function is dependent on pre-trained models. It is important to note that if a word is absent in the model's vocabulary, it lacks a corresponding vector and cannot be effectively used in similarity calculations. Furthermore, the methodology includes setting a threshold to ascertain whether the first sentence of a document qualifies as an introductory sentence. This study initially set the threshold at 0.9, implying that the similarity between the introduction sentence and the entire text should be high. However, given that the gold standard comprises only 22 texts, the threshold of 0.9 needs more empirical generalizability. Consequently, this threshold-based approach was not adopted in our final methodology. Instead, a LLM-based approach was employed, detailed in reference 4.

Alternative approach: Removing examples

Using regular expression patterns (regex) offers a robust method to extract or exclude illustrative content. This approach is based on the identification of specific keywords that typically precede example sentences, such as "e.g.", "example given" or "for example". These markers are used as initial indicators in the text data to determine the beginning of an example.

After recognizing these introductory phrases, the method includes identifying closing punctuation marks or stop elements - in particular, periods (.), exclamation marks (!), or semicolons (;). These elements serve as definitive endpoints for the example sentences or clauses within the text.

The operative mechanism of this approach uses a regex pattern designed to cover the entire span of the text from the introductory keyword to, but not including, the final stop element. This pattern delimits the entire example sentence or phrase for removal. In the practical application of this method, the regex pattern is used to traverse the text and systematically cut out all segments that match the specified pattern.

It is important to recognize certain limitations of this approach. While powerful, regex is not always perfectly attuned to the nuances of natural language, especially in cases where the sentence structure deviates from standard structures. Furthermore, the effectiveness of this method depends on the consistent and conventional use of the specified keywords and stop words in the text. Anomalous or inconsistent use of these elements may result in sub optimal filtering, possibly leaving some example sentences untreated or inadvertently removing text that is not an example. Based on this limitation, our approach uses the LLM-based approach of filtering examples, outlined in 4.

Processing

In this crucial research phase, the focus shifts to extracting information from the refined input text. This process is a cornerstone in transforming natural language descriptions into comprehensive process diagrams. Our approach builds on the fundamental rule-based extraction methods developed by current SOTA. However, we have introduced several extensions to address specific challenges and improve the overall accuracy and relevance of the extracted information:

1. Identification of real actors with LLM
2. Accuracy of actor identification through similarity analysis
3. Inclusion of modal verbs in task descriptions
4. Contextual identification of end events

The following sections address each extension's specific implementation strategies and techniques.

Identification of real actors in process descriptions

To identify real actors, we used various methods in our evaluation, including the rule-based extraction techniques for processing regulatory documents described in a recent study on modeling business processes in 2023 [5]. Since we did not have access to the original implementation, we translated the described methods into Python code using the capabilities of spacy.

The underlying strategy for identifying actors combines constituent trees and record dependency relationships to identify different entities as actors in process descriptions accurately. This method involves specific rules based on linguistic structure and hierarchy:

1. Subject dependency: This rule applies to cases where a noun phrase (NP) is a subject in the sentence. Here, if the NP matches predefined criteria for being an actor (referred to as "actormarker"), it is identified as an actor.
2. Object dependency in passive voice: For sentences in passive voice, this rule looks at prepositional phrases (PP) that dominate a noun phrase. If this noun phrase is an "actormarker", it is identified as an actor. This is particularly relevant in passive constructions, where the actor is typically indicated in a prepositional phrase following the verb.

3. Object dependency in active voice: Similar to the first rule, this applies to sentences in active voice. If a noun phrase (NP) acts as an object and fits the criteria for an "actormarker", it is classified as an actor.

Although we have successfully integrated these rules with spacy, enabling a nuanced extraction of actors, we have encountered a limitation arising from the need for predefined actor markers, which could limit flexibility with different datasets.

Therefore, we expanded our research by evaluating Spacy's default Entity Recognizer. This tool classifies various named entities, such as people, organizations, and geographic locations, facilitating the location of relevant actors within the text [32]. However, Spacy's default Named Entity Recognizer (NER) performance can be inconsistent, as you can see in table 1. This table compares the actor entities identified by Spacy's NER with those recognized by the Process Extraction from Text (PET) set in selected texts from our gold standard dataset [13], showing the variability in actor identification.

Table 1
Named Entity Recognition Results

Text	**Identified Actor ENTs by Spacy NER**	**Identified Actor ENTs by PET**
Text 1:	0	8
Text 2:	0	3
Text 3:	2	16
Text 4:	2	8
Total	**4**	**35**

Incorporating the findings from these analyses, we address the underlying SOTA methods. These methods use advanced linguistic analysis to classify sentence structures and identify syntactic patterns crucial for information extraction. The actors are extracted using dependency labels such as "nsubj" for active sentences and "agent" for passive structures.

However, in this methodology, which is described in detail in section 3, discrepancy in the representation of certain actors in the output graphs can be seen, even though they were clearly identified by dependency labels. This is due to the subsequent distinction between real actors and non-real actors.

The identified actors are stored in an "actor" object and classified as real based on a set of predefined criteria:

1. **Presence in a Predefined Corrector List:** The first criterion involves checking whether the actor is listed in a specifically curated corrector list, named "PERSON_CORRECTOR_LIST." This list comprises predefined entities recognized as actors, ensuring that the identified entity aligns with the expected types of actors in process descriptions.
2. **Synonym and Hypernym Analysis:** The second criterion delves into a more linguistic aspect. It starts by generating a synonym set for each identified actor's name. Subsequently, a hypernym set is derived for each synonym. Hypernyms represent broader categories or classes encompassing the specific entity in question [33]. For instance, as illustrated in Example 36, "place" serves as a hypernym for "kitchen", implying a more general category of which "kitchen" is a part of. Each derived hypernym is then cross-referenced with a list titled "REAL_ACTOR_DETERMINERS". This list encapsulates categories like "person", "social_group", and "software system", as outlined in the definition of a real actor. When a hypernym aligns with these categories, the corresponding entity is classified as a real actor.

(36) "place" is a hypernym of "kitchen""

In order to increase the accuracy and reliability of the classification of actors as real or not real and thus improve the quality of the resulting BPMN, several strategies were evaluated:

1. Extension of the "PERSON_CORRECTOR_LIST" with additional actor entities.
2. Application of the Spacy Entity Recognizer for more robust identification of actors.
3. Refinement of the definition of a real actor expands the "REAL_ACTOR_DETERMINERS" list to include a broader range of hypernyms.
4. Use of a large language model (LLM) for the task of actor classification.

The decision in favor of one approach is explained in the following: (1) The objective of developing a universally applicable methodology that can be adapted to different data sets required the strategic decision not to extend the "PERSON_CORRECTOR_LIST". This decision aligns with the overarching goal of maintaining and improving the dynamic adaptability of the proposed methodology. (2) Using the spacy entity recognizer was ultimately not adopted due to limitations previously identified, outlined in 4.3. (3) Although the current implementation is a generalized solution, it has limitations. One notable example is the failure to classify "central system" as a real actor, although "software system" is listed in the "REAL_ACTOR_DETERMINERS". (4) To innovate in this area, the study

experimented with integrating Large Language Models (LLMs) into the classification process, as they offer widely generalized solutions.

Therefore, a third criterion was added based on the decision of a LLM: The LLM classifies every single actor based on the definition into a real actor or a non-real actor. If the current actor meets the definition, the LLM is instructed to return true. If not, it returns false. As the answer is provided as a string, a method was implemented to convert the answer into a boolean for further processing.

Actor Similarity

Given the multifaceted nature of natural language, an entity might be referenced using multiple terminologies.

Therefore, an algorithm that address the challenge of synonymous terms representing the same actor was implemented. This procedure assesses the similarity between actors before appending a new entity to the list of valid actors. This list is subsequently employed for generating both the syntactical structure for the process diagram and the diagram itself.

Algorithm 1 Determine Actor Similarity Utilizing SpaCy's Functionality

---

Require: $Actor1$: string, $Actor2$: string, nlp : nlp
Ensure: $similarity\_score$: float compare_actors_with_similarity$Actor1, Actor2, nlp$
1: $doc1 \leftarrow nlp(Actor1)$
2: $doc2 \leftarrow nlp(Actor2)$
3: $similarity\_score \leftarrow$ ROUND($doc1$.similarity($doc2$), 2)
4: **return** $similarity\_score$ =0

Alg. 1 necessitates two actor strings as inputs. For the calculation of similarity, we use the inherent functions of spacy. Therefore, the use of the "$en\_core\_web\_lg$" pipeline is required and disqualifies the use of the previously used pipeline "$en\_core\_web\_trf$". This is primarily due to the lack of pre-trained word vectors in the transformer pipeline, which are essential for the similarity estimation. Each vocabulary term possesses a linked vector, a multi-dimensional construct encapsulating semantic nuances determined by contextual associations in extensive corpora. Derived from the input strings, two spacy document objects (Doc) are instantiated. Subsequently, spacy's in-built similarity function evaluates the semantic proximity of these documents, contingent on their respective vectors. This operation computes the cosine similarity, interpreting the cosine of the angle delineating two vectors. Cosine similarity values oscillate between -1 and 1. Spacy normalizes this

value, ensuring the resultant similarity scores range between 0 (indicative of orthogonal vectors, implying dissimilarity) and 1 (identical vectors) [34].

Table 2
Similarity between Actors

	Member of Sales Department	Sales Department	Member of Legal Department
Member of Sales Department	1	0.67	0.92
Sales Department	0.67	1	0.46
Member of Legal Department	0.92	0.46	1

As the cosine similarity, for "member of legal sales" and "member of legal department" is pretty close, but as they refer to different entities, we implemented additionally approach that is token based 2. This function is designed to compute a similarity ratio between two actor names by comparing the lemmas (base forms) of their tokens, with an emphasis on significant content words.

Again, given two strings representing actors as input parameters, the function processes them through a predefined natural language processing pipeline, denoted as "nlp", to generate respective Doc objects. Prior to any comparison, the function systematically excludes tokens that are characterized as "stop words". Stop words, refer in linguistics and natural language processing to frequently occurring words in a language that, in analytical contexts, are considered to offer limited semantic value. Subsequently, the function undertakes a pairwise comparison of the lemmas of the tokens derived from the two actors. The objective of this phase is to enumerate the quantity of matching lemmas between the two sets. Acknowledging the potential variance in token counts between different actor strings, the function uses a normalization procedure to ensure a balanced evaluation of similarity without distortion.

To improve the accuracy of actor identification in process diagrams, our approach integrates both algorithms, 1 and 2, focusing on different aspects of similarity analysis. Combining these methods, we can better recognize whether different terminologies or phrases refer to the same entity within a process. This is particularly important to account for the inherent complexity and nuances of the natural language used in process descriptions.

Algorithm 2 Compare Actor Tokens Using SpaCy

Require: $Actor1$: string, $Actor2$: string, nlp : nlp
Ensure: $similarity\_ratio$: float

$doc1 \leftarrow nlp(Actor1)$
$doc2 \leftarrow nlp(Actor2)$
$tokens1 \leftarrow$ FILTER_OUT_STOP_WORDS$(doc1)$
$tokens2 \leftarrow$ FILTER_OUT_STOP_WORDS$(doc2)$
$num\_tokens1 \leftarrow$ length$(tokens1)$
$num\_tokens2 \leftarrow$ length$(tokens2)$
$matching\_tokens \leftarrow 0$
if $num\_tokens1 \leq num\_tokens2$ **then**
  for each $token1$ in $tokens1$ **do**
    for each $token2$ in $tokens2$ **do**
      if $token1.lemma\_ == token2.lemma\_$ **then**
        $matching\_tokens \leftarrow matching\_tokens + 1$
        break
      end if
    end for
  end for
else
  for each $token2$ in $tokens2$ **do**
    for each $token1$ in $tokens1$ **do**
      if $token2.lemma\_ == token1.lemma\_$ **then**
        $matching\_tokens \leftarrow matching\_tokens + 1$
        break
      end if
    end for
  end for
end if
$avg\_tokens \leftarrow (num\_tokens1 + num\_tokens2)/2.0$
if $avg\_tokens > 0$ **then**
  $similarity\_ratio \leftarrow matching\_tokens/avg\_tokens$
else
  $similarity\_ratio \leftarrow 0.0$
end if
return $similarity\_ratio$ =0

For example, if both algorithms yield a similarity score greater than 0.5, it is inferred that the two strings in question describe the same actor. This threshold has been set based on extensive testing and analysis to ensure a balanced approach to actor identification. By setting this threshold, we attempt to balance overgeneralizing and overlooking essential connections between different actor labels.

The combined use of these algorithms not only increases the accuracy of our actor identification process but also contributes to the creation of more accurate and representative BPMN diagrams. This approach considers the complexity of actor references in natural language. It uses a sophisticated method to ensure that all relevant actors are correctly identified and represented in the process models.

Context-Driven Identification of End Events

Building on the approach outlined in the 3, the implementation phase focuses on accurately identifying the end of processes, especially in contexts where completion is not explicitly stated. This is achieved by extending SOTA with a contextual analysis that uses LLMs to recognize implicit indicators of the end of a process.

A key innovation in our approach is using LLMs specifically instructed to interpret the process context to identify potential end activities. The LLM is instructed to return a Boolean value: true if an action marks the end of the process and false if not. This instruction is illustrated by providing the model with a general example, such as the action "customer takes car" in a car repair process, which typically signifies process completion.

The results of the LLMs are normalized to boolean values for further integration into the SOTA process. In this context, GPT-4 performed superior to GPT-3.5 in interpreting and determining the final activities within different process scenarios.

An additional attribute was introduced for each activity to refine the process model further. This attribute identifies whether an activity exclusively denotes the completion of the process (referred to as a "is_finish_activity"). This distinction is crucial because while all "finish" activities are inherently "end" activities, the reverse is not always true.

In the syntax generation phase, the activities marked as "is_finish_activity" are not inserted directly into the process model as individual tasks. Instead, they are represented by an end event to reflect their role in the process flow accurately. For activities labeled as "end" but not "is_finish_activity", the activity and an end node are included in the generated syntax. This distinction ensures that the generated BPMN diagrams accurately reflect real-world processes, especially in cases where the end of the process is implied rather than explicitly stated.

Post-Processing

Once the relevant information, including activities, relationships and actor have been extracted from the text actors, we can proceed with the visualization step. As outlined in 3, there is no standardized approach or commonly used library used. Therefore two potential libraries have been identified and will be outlined in the following:

Enhancing rendering quality through syntax optimization / Refinement of task labels

Building on the solution design described in 3, this section focuses on the implementation strategies used to refine the syntax and thus improve the quality of the resulting BPMN diagrams.

A central aspect of our approach is the optimization of the individual task labels. In order to comply with the "Do Not Repeat Yourself" (DRY) principle, we developed our method for consistently generating task labels. Based on the basic approaches for generating task labels, we strategically replaced the keyword "as" with "like" in the task labels. This replacement is feasible as most instances of "as" in contexts such as explanatory (25) or relativizing uses (24) were pre-filtered by the LLM-assisted text refinement, based on the IIC criteria. As a result, only comparative usages in which "as" can be aptly replaced by "like" need to be replaced.

To further improve the quality of task labeling, a method was implemented that leverages the contextual understanding of LLMs. This method instructs the LLM to refine task labeling texts for clarity of the process diagram based on the text description of the process. In addition, a first approach for handling sub-processes has been introduced. If two tasks described as "the first activity" have the same object, they are merged by the LLM. The LLM also deals with imprecise descriptions such as "doing their tasks" and refines them based on the attached process description. Each task description is refined to contain a maximum of six words and to ensure that verbs are used in their base form to ensure conciseness and clarity.

The LLMs are instructed to return only the process diagram syntax as a simple string without additional explanations. Tests with different models have shown that while GPT-4 performs well on specific tasks (especially improving task labels), it occasionally omits task labels of the syntax or the provided instructions or changes the syntax structure, leading to inconsistencies and rendering errors. The GPT 3.5 instruction model, on the other hand, consistently reproduces the correct syntax format but tends to reintroduce "as" in labels. To counteract this, a post-processing check is

performed in which each occurrence of "as" in task labels (indicated by square brackets) is replaced by "like".

Given the dynamic nature of LLMs, a fallback mechanism is implemented. If problems occur during LLM processing and the corresponding rendering, the system returns to the unprocessed rule-based syntax for rendering. This ensures consistent output generation and eliminates potential failures caused by LLM inconsistencies.

Through these strategies, we aim to significantly improve syntax quality and ensure that the rendering of BPMN diagrams is more accurate and less prone to errors caused by syntax errors. This approach increases the diagrams' clarity, making complex process descriptions more accessible and understandable.

Evaluation

The culmination of any research project is the evaluation, where theory meets empiricism and methods are evaluated. This chapter systematically evaluates our proposed approach for automatically generating BPMN models from natural language descriptions. Our goal is to analyze and understand the effectiveness and limitations of our method through a structured evaluation that includes both quantitative and qualitative analyses.

Our evaluation strategy is based on a comprehensive comparative framework in which we measure our generated models against a gold standard. This standard consists of a set of text-based process descriptions and BPMN diagrams modeled by experts. By comparing the results of our approach against this standard, we aim to extract metrics - accuracy, recognition, and F1 score - that serve as indicators of the accuracy and reliability of our model.

To maintain a comprehensive perspective, we extend our investigation to evaluate the methods presented in [4] and [2]. A carefully defined gold standard (GS), comprising both traditional textual process descriptions and regulatory documents, will form the basis for our comparative analysis. Against this background, we will manually count and evaluate the occurrence of BPMN components such as gateways, lanes, and nodes to measure the compliance of automated models with the GS diagrams.

The GS is mostly only available in picture format, so an automated similarity comparison between the task labels is impossible. Therefore, in the qualitative evaluation, the quality of the task labels will be evaluated results (SOTA) based on the impressions of manual evaluation.

The conclusion chapter will address the unresolved issues and challenges that have arisen during this research. In the spirit of continuous improvement and development, we will make recommendations for future research directions and possible improvements to our proposed methodology. In doing so, we are approaching a detailed and rigorous evaluation to provide meaningful insights into process modeling and natural language processing.

Evaluation Metrics

Our evaluation is limited to the quantitative assessment of the occurrence of BPMN elements but not to the sequence. This targeted approach is based on the following strategic considerations:

First, the sequence and order of BPMN elements have been extensively analyzed in previous foundational work [4]. These foundational studies have successfully developed robust methods for ordering elements in process models. This research circumvents the need for redundant element order verification by utilizing these foundational insights. In particular, any significant deviations in the ordering of elements compared to the baseline approach are highlighted during the qualitative analysis phase.

Second, the research objectives of this study are aligned with specific improvements to the BPMN modeling process. These improvements include a refined identification of actors, a nuanced similarity analysis for actor terms, and the inclusion of modal verbs in task descriptions. The aim is to increase the precision and completeness of the identification and representation of BPMN elements. Such a specialized focus shifts the primary attention to the ordering of elements, which is important but falls outside the scope of the research objectives.

Therefore, the quantitative evaluation will closely assess the presence and accuracy of BPMN elements in the generated models. This focused analysis will provide insight into the effectiveness of the proposed improvements using the following metrics as a guide:

In the research, we deviate from the method of [35], who used the Graph Edit Distance (GED) to evaluate his approach. Instead, we focus on analyzing the occurrence of the generated BPMN elements, using precision, recall, and the F1 score as primary evaluation metrics. This decision is based on several considerations relevant to our research objectives and methodology.

First, our approach is fundamentally concerned with various classification tasks, such as determining the relevance or irrelevance of items and identifying real and non-real actors. In this context, precision and recall prove to be significant metrics. Precision quantifies the accuracy of positive predictions in our model, while recall evaluates the model's ability to identify all relevant instances. The F1 score, the harmonic mean of precision and recall, provides a balanced assessment of these two aspects. This trio of metrics is suitable for evaluating the effectiveness of our classification

tasks based on the occurrence of the various elements in the resulting BPMN diagrams. It provides a precise and quantitative measure of the performance of our research.

Conversely, GED, traditionally applied in graph theory, is primarily concerned with determining the similarity between two graphs. It calculates the minimum number of graph editing operations required to transform one graph into another, such as inserting, deleting, and replacing nodes and edges. This metric is more in line with [4]'s approach, which focuses more on the graph structure than the classification of relevant data. GED evaluates the structural changes between graphs, and the assigned costs for different processing operations strongly influence its results. However, this focus on graph structure deviates from our research goal of classifying information within BPMN elements.

Furthermore, the application of GED in our context is hampered by the lack of a machine-readable gold standard, usually only available in formats such as .png. This limitation makes matching and calculating similarity using GED challenging and dependent on the quality of the similarity methods. Furthermore, relying solely on a gold standard to assess model quality can be problematic, as the gold standard itself is a product of different model developers and, therefore, may not represent an absolute benchmark.

Considering these arguments, we have assessed the quality of our models through manual revision and supplemented our quantitative analysis with qualitative insights. This approach allows for a more nuanced and contextualized assessment of our models' performance.

In summary, we chose Precision, Recall, and the F1 score because they are relevant for classification tasks, easy to calculate and interpret, and widely recognized and accepted in fields such as machine learning, statistics, and data science. These metrics provide a clear and quantifiable measure of our model's performance in classifying BPMN elements, which is very conductive to our research goals. In contrast, GED, focusing on graph structure and associated complexity, is less suited to our specific evaluation requirements.

Considering the above, the most reliable results can be obtained when the automatically generated models are manually checked against the gold standard models and with context to the input text.

Therefore, our quantitative evaluation utilizes precision, recall, and F1 scores for assessing model performance on:

- lanes L
- task and event nodes N
- (parallel and exclusive) gateways G

These metrics are formally defined as:

$$Precision = \frac{\text{True Positives}}{\text{True Positives + False Positives}} \tag{5.1}$$

$$Recall = \frac{\text{True Positives}}{\text{True Positives + False Negatives}} \tag{5.2}$$

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision + Recall}} \tag{5.3}$$

Where:

- True Positives (TP): correctly matched occurrences
- False Positives (FP): irrelevant created occurrences not present in the gold standard
- False Negatives (FN): missed occurrences
- True Negatives (TN): correctly ignored irrelevant occurrences

The recall (rec.) measures the ability of the model to find all the relevant cases within a data set, while the precision (prec.) indicates the proportion of positive identifications that were actually correct. Therefore, the higher the recall, the more comprehensive the model is in retrieving relevant instances, while a higher precision relates to a lower rate of false positives.

The following section presents an analysis of the textual input, followed by an evaluation of our proposed models with the GS models developed by BPMN experts. We extend this evaluation with a comparison of the SOTA approach described in [2] and the improved SOTA approach developed by [4]. This comparison includes 22 input texts [6]. The metrics described above are used to ensure a consistent evaluation framework. The main objective is to determine the relative performance of the

[6]The model-text pair with ID 4 was removed from the gold standard as it does not represent a process description, so the numbering of the pairs ranges from 1 to 23, with 4 not included.

different methods on different input texts. Full details of the model-text pairs, extensive evaluation statistics, and the generated process models are available on our GitHub repository [7].

Data Set

A gold standard comprising 22 meticulously selected textual process descriptions along with their corresponding BPMN 2.0 diagrams has been established for this study. The assembly of this gold standard involved a methodical selection process, capturing a wide range of process descriptions across various domains. These domains include academic [35], the ISO/IEC 27001:2022 standards, GDPR [36], [37], and industry-specific processes of Smart Meter [37], [38]. The purpose of crafting such a gold standard is to provide an unbiased benchmark for evaluating automated process modeling techniques and to serve as a comprehensive dataset for future research endeavors in this field. The complete set of textual descriptions, corresponding BPMN diagrams, and models generated by our approach are accessible in the same code repository.

Table 1
Overview of Textual Descriptions and Model Pairs by Source

Origin Area	Source	ID	# Text-Model Pairs	Average # of Sentences	Average # of Tokens	Average # of Tokens per Sentence
academic	Friedrich	I (1-3)	3	10	184	18
regulatory	ISO/IEC 27001:2022	II (5-7)	3	5	254	83
regulatory	GDPR	III (8-14)	7	4	150	41
industry	Smart Meter	IV (15-23)	9	7	157	7

The dataset consists of text-process model pairs categorized according to their origin and typical usage scenarios [35]:

- **Academic:** these pairs are created and used by academic professionals, authors of educational material, tutorials from BPM vendors, or for internal training purposes. They illustrate theoretical concepts and are often used as educational tools.
- **Industry:** This category includes pairs actively developed and used in business processes. They reflect practical applications and are tailored to the operational needs of businesses to capture the nuances of their daily activities.

[7] https://github.com/VincentDerekHeld/thesis-bachelor-text2BPMN/tree/main/evaluation

- **Regulatory:** These pairs are derived from official legal texts, such as laws, regulations, and norms. They are created to ensure compliance with legal standards and usually have a higher level of complexity due to their formal and structured nature.

Table 1 summarizes the dataset's statistics, which includes the average number of sentences, tokens, and tokens per sentence for each source type. These metrics are derived from the tokenization and sentence segmentation capabilities of the Spacy library.

In the context of this research, the academic and the smart meter corpus text-model paris are aligned with traditional business process descriptions. In contrast, regulatory documents, such as ISO standards and GDPR articles, have different attributes explained in more detail in section 3.1. Fig. 1 illustrates the diversity of the GS while keeping a balance between traditional textual descriptions (marked in blue) and regulatory documents (marked in grey).

Previous SOTA approaches have shown that performance decreases as the length of the input text increases [35]. To evaluate this issue, the current study set several parameters for text length to be included in the quantitative evaluation. The distributions of sentences, tokens, and average tokens per sentence are shown graphically in Figures 2, 3 and 4, providing a visual interpretation of the structural composition of the data.

Figure 1
Overview of the distribution of traditional vs. regulatory text descriptions

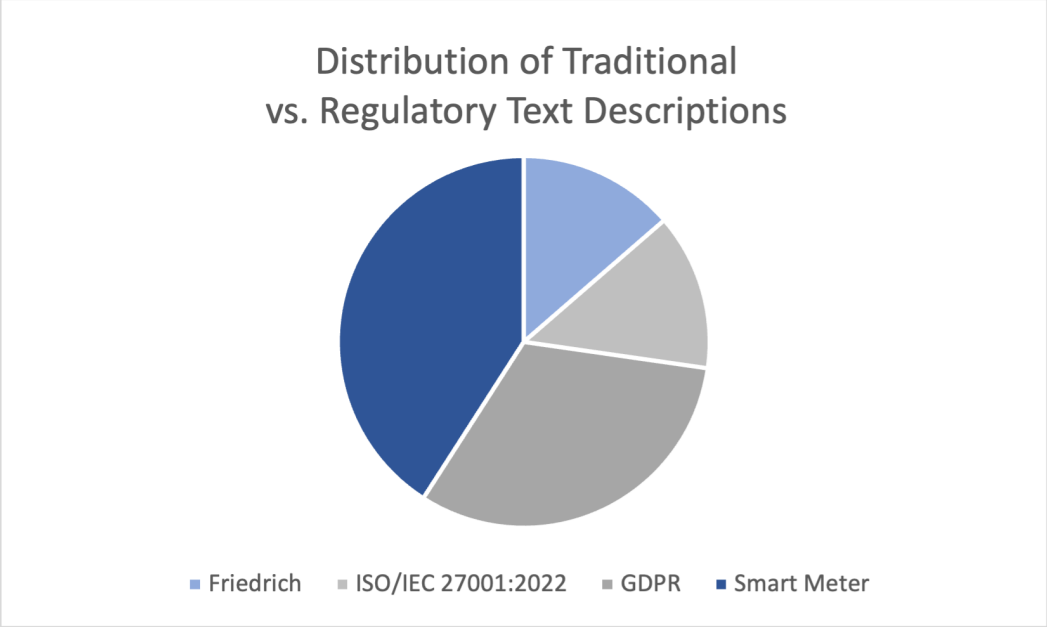


Figure 2
Overview of the number of sentences

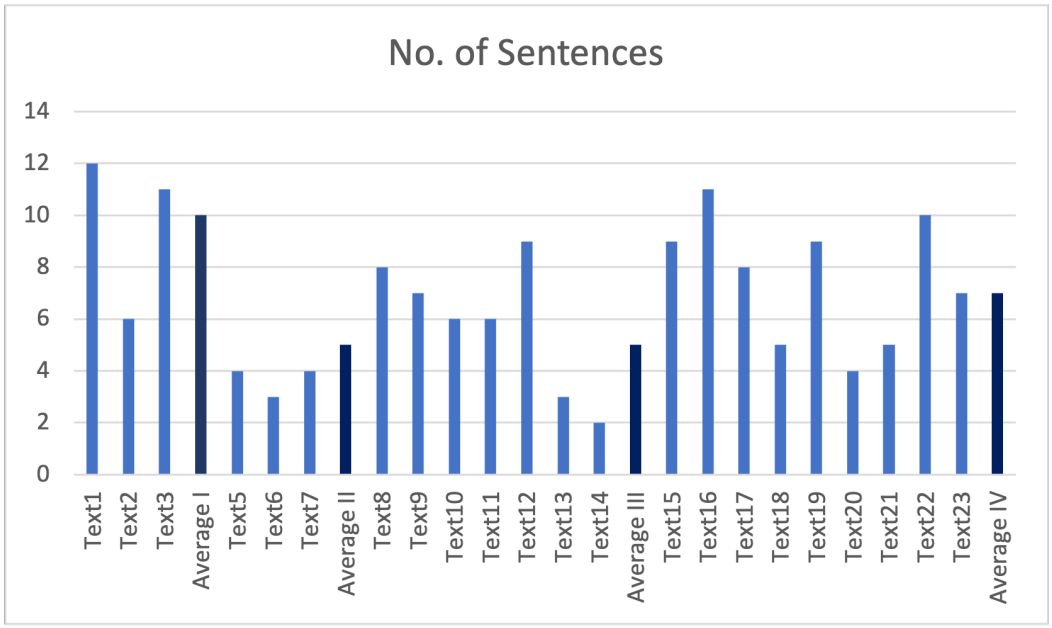


Figure 3
Overview of the number of tokens

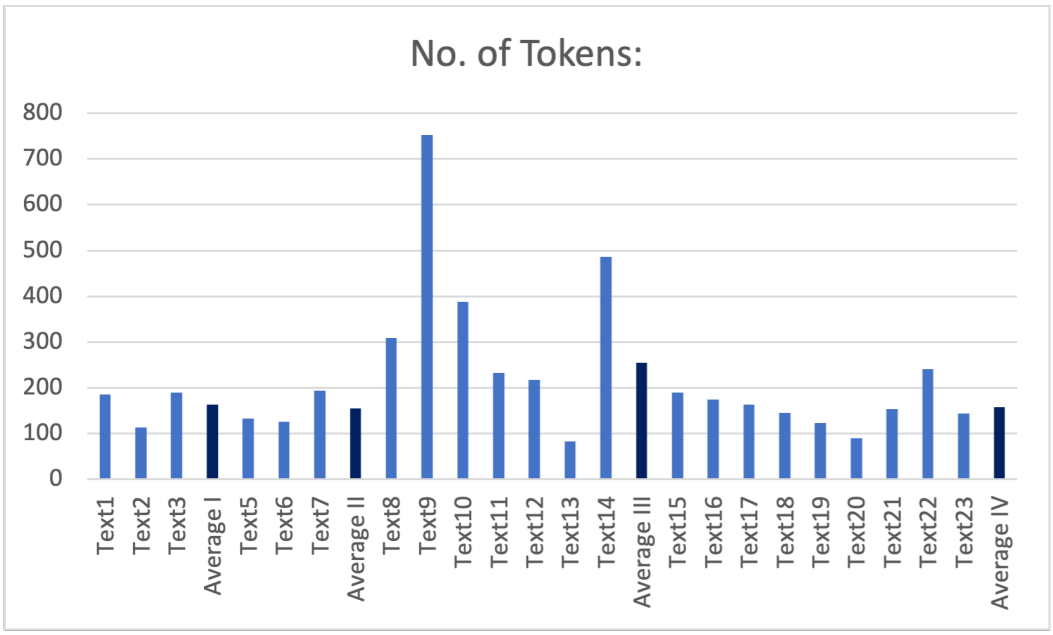
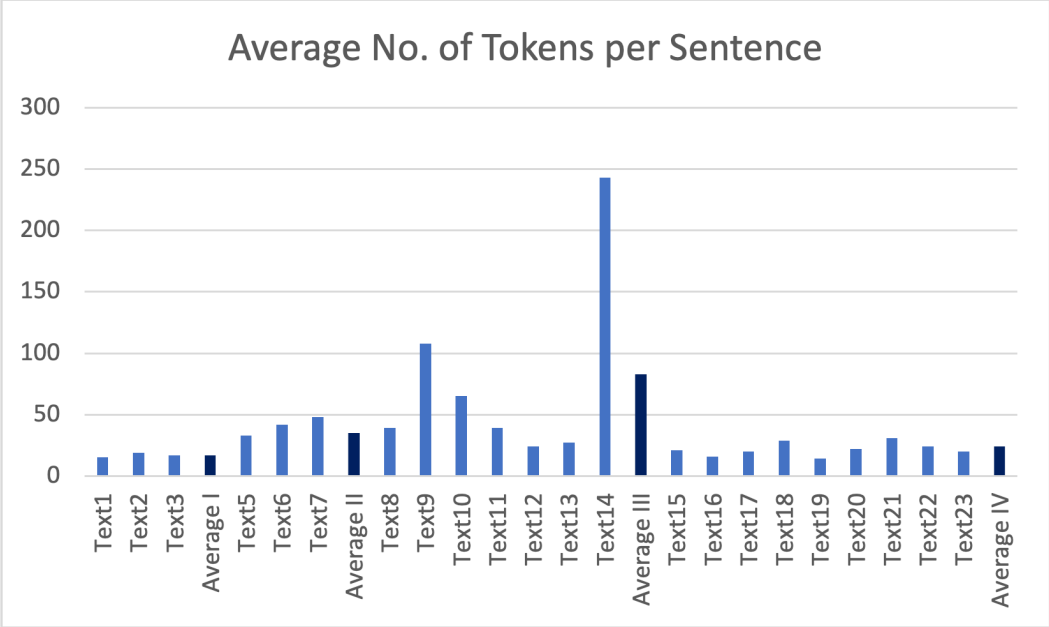


Figure 4
Overview of the average number of tokens per sentence



Qualitative Evaluation

While the quantitative evaluation will focus on the number of identified elements, this section will outline aspects considering the quality of the created diagrams, particularly the quality of the labels in the created diagrams. As already explained, an automated comparison of the task labels by determining the similarity between the different task labels is impossible, as most GS are only available as image files.

This section shifts the focus from the quantitative aspects, such as the number of identified elements, to the qualitative attributes, particularly the quality of the labels in the created diagrams. It is important to note that an automated comparison of the task labels was impossible, mainly because most gold standard references were only available as image files, preventing direct text analysis.

When the results were checked manually and the values subsequently determined, an improvement in the quality of the task labels was observed. The shortness and precision of the labels primarily characterize this improvement. Although the labels have become longer in some cases due to the inclusion of modal verbs (in the case of regulatory documents), this extension has contributed significantly to the precision of the labels. Including modal verbs effectively conveys the degree

of obligation or necessity associated with a task, improving the degree of descriptive quality of the labels.

A helpful example of this improvement is comparing the diagrams in the figures 5 and 6. The kitchen activity is only described as "doing their tasks" in the diagram created with the improved SOTA approach. In contrast, in this approach, the specific activity of the kitchen is described with a more precise and descriptive label: "prepares food". This example highlights the improved clarity and relevance of task labels in our approach and reflects a significant advance in the quality of BPMN diagramming.

Such qualitative advances in the clarity of the labels not only improve the readability of the diagrams but also contribute to a more accurate and meaningful representation of the underlying processes, thereby increasing the overall usefulness of the BPMN models for various stakeholders.

Figure 5
Automatic generated output diagram for Text 2 of this approach

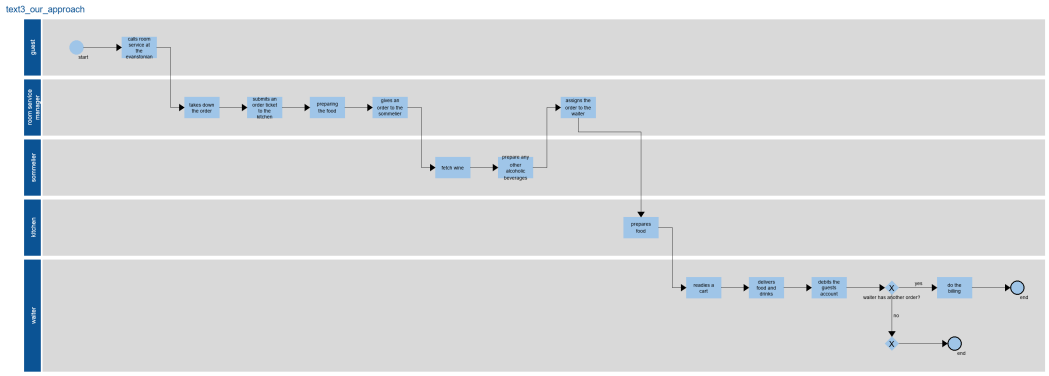
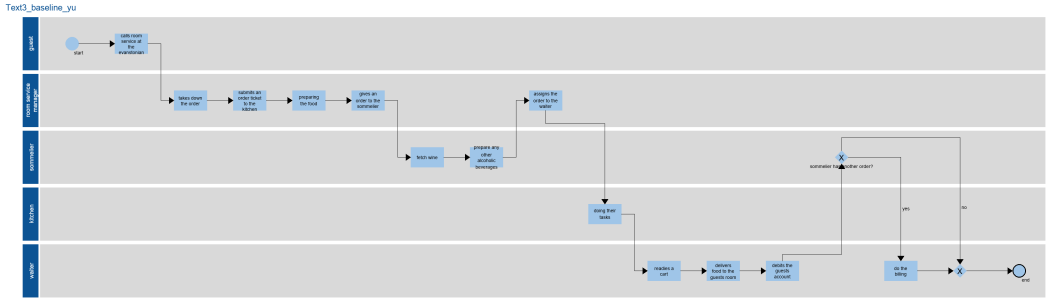


Figure 6
Automatic generated output diagram for Text 2 of the improved SOTA rule-based



Quantitative Evaluation

In the following the in 2,3 and 4 presented metrics will be outlined.

Evaluation of gateway identification (G)

Identifying gateways, denoted by **G**, presents a notable challenge across all examined approaches, as evidenced by universally low recall rates. This trend suggests a systemic difficulty in accurately detecting gateways within process models. Our methodology, together with the improved SOTA model of [4], shows a higher precision than the method proposed by Friedrich et al. [2], although the former has a slightly higher precision.

Upon dissecting the results across various text categories, Category I, which encompasses academic texts, consistently yields the highest precision and recall. This indicates that gateways within academic texts are the most effectively-identified by all three investigated approaches.

Category III, which comprises GDPR documents – often criticized for their complexity – exhibits more accurately identified gateways in both the refined SOTA method and our approach than texts from the industrial domain (Category IV). Nevertheless, it is essential to acknowledge that the SOTA approach outperforms this approach in gateway detection within Category III, as measured by precision. However, this comparison is somewhat biased as the improved SOTA approach did not produce three diagrams from this category, which may bias the comparison results.

For Categories II and IV, the improved baseline and this approach demonstrate an F1 score of zero, indicating no correct identification of gateways, significantly impacting the overall results across all categories. The underlying causes for this outcome will be examined and detailed in the qualitative analysis section.

Evaluation of nodes identification (N)

When evaluating node identification, labeled as **N**, all methods showed significantly improved performance compared to gateway identification. Nevertheless, the SOTA showed significantly inferior performance compared to both the improved SOTA method and this approach, with the latter showing a slight advantage.

The SOTA had difficulties recognizing relevant information and the precision of such identifications. In contrast, improved SOTA extracted more information from the input texts. However, this led to the unintended consequence that no BPMN diagrams could be created for certain texts, as explained in section 3.

For all models, the best node identification results were observed in Category I (academic texts), followed by Category IV (industry texts), suggesting that node detection is more effective in traditional process descriptions than legal texts.

While the improved SOTA model showed better results in identifying activities, it should be noted that the lack of diagrams for three input texts biases these results. In terms of precision, this implementation outperformed the others in all categories, highlighting the effectiveness of the approach.

Evaluation of actors identification (L)

In business process modeling, identifying actors represented as lanes (L) is crucial to creating accurate BPMN diagrams. Our research aimed to improve this facet of model creation, and based on the comparative metrics, our approach has outperformed baseline methods.

The results show that while the underlying approaches achieved high level of identification and accuracy in academic contexts (Category I), they declined in the further categories. This approach not only maintained a high recall across all categories, ensuring the identification of all relevant actors but also significantly improved precision. This suggests that this implementation performs better at accurately classifying actors as accurate.

Table 2
SOTA rule-based *occurrence results compared to the gold standard [cf. [2]]*

origin	group	**L**		**N**		**G**	
area	ID	rec.	prec.	rec.	prec.	rec.	prec.
academic	I	0,889	0,727	0,78	0,593	0,158	0,5
regulatory	II	0	0	0,196	0,262	0	0
regulatory	III	0,118	0,5	0,315	0,149	0	0
industry	IV	0,120	0,5	0,491	0,452	0	0
ALL	TOTAL (I-IV)	0,241	0,619	0,43	0,326	0,033	0,150
ALL	F1 Score	**0,347**		**0,371**		**0,054**	

Table 3
improved SOTA rule-based *occurrence results compared to the gold standard [cf. [4]]*

origin	group	**L**		**N**		**G**	
area	ID	rec.	prec.	rec.	prec.	rec.	prec.
academic	I	1	0,75	0,78	0,653	0,263	0,417
regulatory	II	0	0	0,304	0,548	0	0
regulatory	III	0,111	0,500	0,484	0,259	0,154	0,5
industry	IV	0,087	0,5	0,548	0,613	0	0
ALL	TOTAL (I-IV)	0,273	0,667	0,522	0,524	0,1	0,389
ALL	F1 Score	**0,387**		**0,523**		**0,159**	

Table 4
Our approach *occurrence results compared to the gold standard*

origin	group	**L**		**N**		**G**	
area	ID	rec.	prec.	rec.	prec.	rec.	prec.
academic	I	1	0,9	0,805	0,750	0,263	0,455
regulatory	II	1	1	0,393	0,786	0	0
regulatory	III	0,647	0,786	0,384	0,275	0,125	0,25
industry	IV	0,32	0,727	0,605	0,676	0	0
ALL	TOTAL (I-IV)	0,574	0,816	0,535	0,551	0,099	0,257
ALL	F1 Score	**0,674**		**0,543**		**0,143**	

Comparison of the identification of BPMN elements in traditional and regulatory texts

This study focuses on the differences in identifying BPMN elements between traditional process descriptions and regulatory texts, a topic that has been addressed repeatedly in the previous sections. This section consolidates these findings by combining the two data types across all BPMN elements.

The tables 5, 6 and 7 provide a summary overview comparing the performance of the different approaches against traditional and regulatory inputs. This methodology consistently shows superior performance on all BPMN elements compared to the established SOTA methods.

A differentiated view shows that the accurate identification of actors remains stable regardless of the text type. While identifying activities accurately is higher for traditional texts, the difference with regulatory texts must be clear. However, the sharp contrast can be seen in recognizing gateways - especially in legal texts, where gateways are rarely recognized correctly. Despite this challenge, traditional texts show sufficient accuracy in gateway recognition, even though this area has the most significant potential for improvement. The results highlight the need for further refinement of gateway identification within legal frameworks and confirm the effectiveness of the proposed improvements in the identification of actors and activities for processing traditional texts.

Table 5
SOTA rule-based *occurrence results compared between traditional process and regulatory descriptions [cf. [2]]*

Type of Input Text	Group ID	**L**		**N**		**G**	
		Rec.	Prec.	Rec.	Prec.	Rec.	Prec.
Traditional	I + IV	0.504	0.614	0.636	0.522	0.079	0.250
F1 Score for I + IV		**0.554**		**0.573**		**0.120**	
Regulatory	II + III	0.059	0.25	0.256	0.206	0	0
F1 Score for II + III		**0.095**		**0.228**		**0**	

Table 6
improved SOTA rule-based *occurrence results compared between traditional process and regulatory descriptions [cf. [4]]*

Type of Input Text	Group ID	**L**		**N**		**G**	
		Rec.	Prec.	Rec.	Prec.	Rec.	Prec.
Traditional	I + IV	0.543	0.625	0.664	0.633	0.132	0.208
F1 Score for I + IV		**0.581**		**0.648**		**0.161**	
Regulatory	II + III	0.042	0.167	0.394	0.404	0.077	0.25
F1 Score for II + III		**0.067**		**0.399**		**0.118**	

Table 7
Our approach *occurrence results compared between traditional process and regulatory descriptions*

Type of Input Text	Group ID	**L**		**N**		**G**	
		Rec.	Prec.	Rec.	Prec.	Rec.	Prec.
Traditional	I + IV	0.660	0.814	0.705	0.713	0.132	0.227
F1 Score for I + IV		**0.729**		**0.709**		**0.729**	
Regulatory	II + III	0.824	0.893	0.388	0.530	0.063	0.125
F1 Score for II + III		**0.857**		**0.448**		**0.083**	

As the results shows, [2] has moderate performance. In all metrics (F1-Score, precision and recall for each category of occurrences (Lanes, Nodes and Gateways)) the improved approach by [4] and this approach performance better. While [4] show in general better results, this approach outperforms both of the underlying approaches in most metrics, but in particular in the occurrence of the Lanes with a highest recall, precision and therefore highest overall F1 score.

Discussion

This research represents a significant enhancement in automating the creation of BPMN diagrams. This approach, demonstrated through rigorous evaluation, notably outperforms existing methods by transitioning from semi-automated to fully automated processes. This advancement is primarily attributed to the improved identification of actors, which previously required manual listing or contextual hypernym addition. The introduction of LLM-assisted text refinement (LLM-ATR) has also played a critical role, involving the breakdown of enumerations, filtration of irrelevant content, and transformation of implicit information into explicit form.

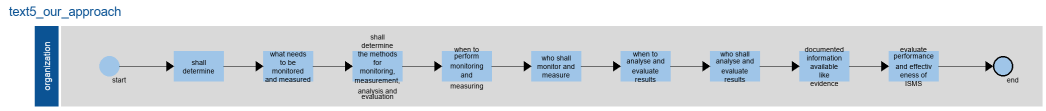
The challenge of transforming implicit information into explicit information must be considered unsolved. Different models have been evaluated for this task, and the text-da-vinci-003 model achieved the best results in identifying the implicit information and transforming it from implicit to explicit. We decided to adopt newer models based on OpenAI's decision to discontinue this model starting in 2024. This decision aligns with our goal to develop future-proof innovations and responsible research. However, as the results of our research represent, our implementation does not achieve the same level of results in this task as before. While the GPT3.5-instruct model cannot identify implicit actions sustainably, it returns precise answers and strictly follows the instructions. The GPT4.0 model, on the other hand, achieves good results in the identification of implicit information but tends to answer with an excessive length and add verbatim to the result.

Challenges still exist, especially in the extraction of gateways and the granularity of sentence splitting, which sometimes leads to excessive splitting and multiple activities for a single piece of information. Fig. 1 represents this problem, as the activity "shall determine" should not be an activity, but the label description should be a part of the following task label(s). A limitation of gateways is not only the limited identification, but also the inability to remove end gateways if a branch leads to the end of the process.

While identifying actors in legal texts shows promising results and the extraction of activities is satisfactory, identifying gateways remains a challenge. This problem exists not only with regulatory documents but also with traditional textual process descriptions. While from traditional textual process descriptions a structured process containing condition blocks and AND blocks can be

created, for regulatory documents the output results mostly in a linear process without either condition- or AND-bocks (Fig. 1).

Figure 1
Automatic generated output diagram for Text 5 of this approach



The introduction of ISO 24495-1:2023 and the ongoing development of ISO/AWI 24495-2 represent an emerging need for standardization in regulatory documents [39], [40]. These developments will likely profoundly impact future research directions and the practical application of our findings, which should be considered for further research.

For further research, the gold standard should be further developed. For regulatory text, components of different ISO standards should be included to confirm the universality of the approach or, if necessary, to develop it further. It should also be noted that no single gold standard exists for a process model. While the context-driven identification of end events using LLMs represents significant progress, as illustruated in 2, the evaluation is difficult, as some experts prefer more end nodes, while others strive for minimalism. This can also be seen within the used GS, where some BPMN models aim to have one end for each actor/lane, while others include the process as a whole. This variance exemplifies the diversity in modeling BPMN diagrams and underlines the need for increased standardization. Another example of diversity is particularly evident when comparing diagrams that start with customer interaction. While some BPMN diagrams include the initial interaction with the customer in the process (illustrated in 2, other diagrams focus exclusively on the company's internal processes (Fig. 3)).

Figure 2
Model with initial customer interaction excluded[cf. [2]]

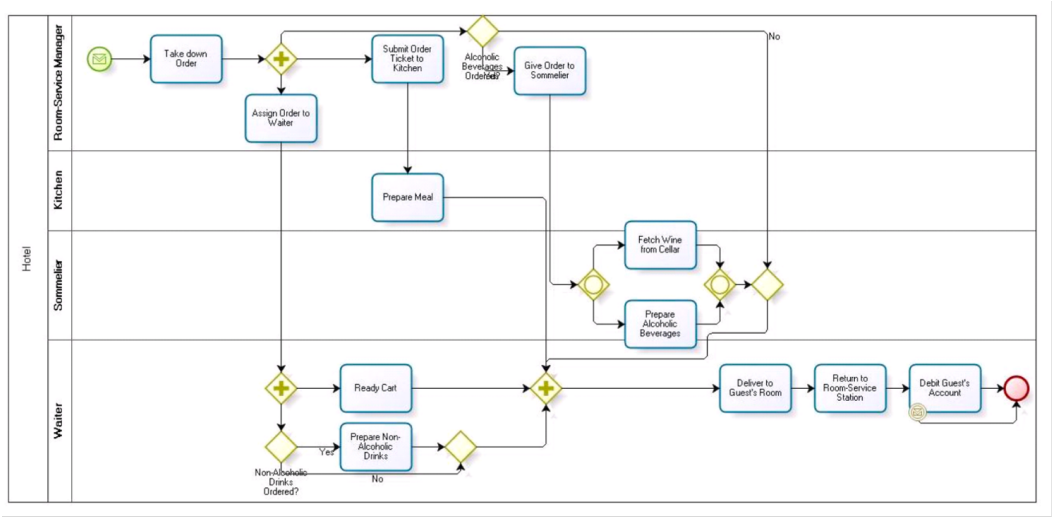
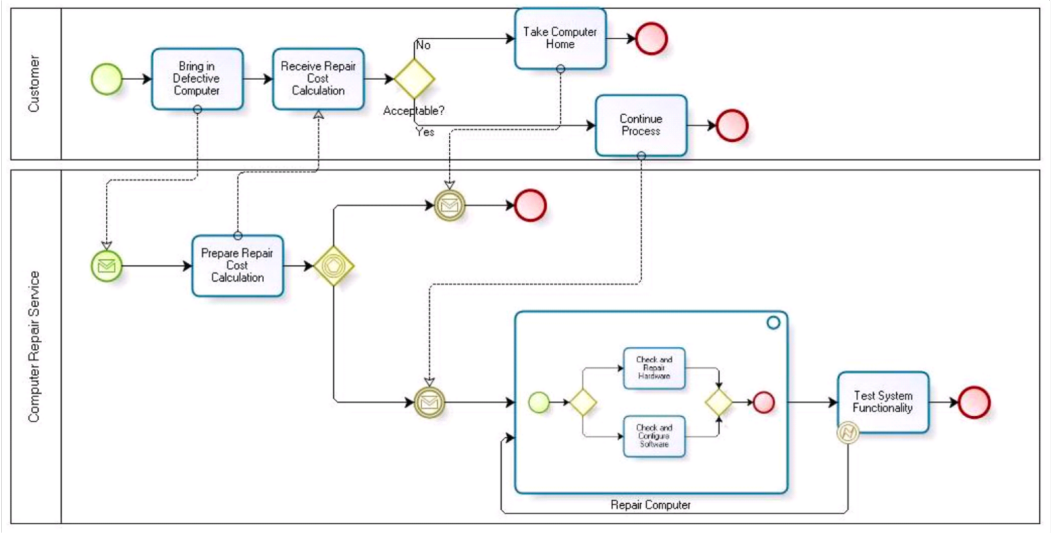


Figure 3
Model with initial customer interaction included [cf. [2]]



One of the critical aspects of using LLMs in research is the concern about transparency and reproducibility. This study addressed these concerns by assigning specific, smaller tasks to the LLMs. This approach allows for easier monitoring and validation of results, as the output of each task can be directly observed and evaluated against the console outputs. This method increases the transparency of the process and makes it easier to understand and evaluate the contributions of the LLMs.

However, a significant challenge with LLMs is the variability of their responses. In practice, identical queries can return different results when repeated. This variability can lead to inconsistencies and unpredictable results, which is a significant problem in the context of research.

To mitigate this problem, we have introduced strategies such as including controlled lists and applying filtering criteria. These measures serve to ensure a degree of consistency and control over the results of the LLM. In addition, we consciously decided in this methodology to create all models in a single iteration. This approach was chosen to avoid potential biases from repeatedly building models until achieving an optimal result.

By limiting ourselves to a single attempt at model building, we wanted to highlight the inherent limitations of current LLMs, particularly their consistency. This decision underscores the importance of recognizing and addressing the limitations of advanced models. It also forms the basis for future research on developing more consistent and reliable LLMs. This commitment to ensuring consistency and transparency in this methodology enhances this results' credibility and contributes to a broader discourse on the responsible use of LLMs in research.

In addition, the "process-piper" package used to visualize the diagrams is limited to image formats, which restricts post-creation modifications. Future work should aim to produce BPMN 2.0 compliant diagrams with corresponding .xml files for greater flexibility and usefulness.

Future research should focus on several areas: the use of a parser specifically trained on legal texts, the general improvement of gateway identification, the use of trained LLMs for context-specific tasks, and the improvement of the transformation from implicit to explicit information. In addition, the identification of start events using LLMs represents another promising avenue for exploration.

In summary, this research shows that a combination of traditional rule-based NLP methods and LLMs can yield powerful results when applied to context-specific decisions. In particular, LLMs show significant potential in breaking down complex, challenging tasks into smaller, more manageable subtasks. This approach advances the automatic generation of BPMN diagrams and develops different methods for using LLMs, opening up new avenues for future research and practical applications in this area.

Conclusion

This paper presents an enhanced automated approach for transforming natural language process descriptions into BPMN 2.0 process diagrams, considering applying regulatory process descriptions. Our method represents a significant advance in natural language processing (NLP) and demonstrates the successful integration of leading-edge technologies to automate the creation of business process models.

This research shows a significant improvement in the quality and accuracy of process diagrams generated from text through the novel use of traditional rule-based technologies combined with Large Language Models (LLMs) and the introduction of LLM-assisted textual refinement (LLM-ATR). Our research shows that synergistically integrating traditional rule-based methods with state-of-the-art LLM innovations can improve text processing from the initial pre-processing stages to advanced post-processing analysis. This combined approach not only enriches syntactic analysis but also provides a sophisticated understanding of complex textual data.

A key finding of this study is the effective use of LLMs with minimal training data in addressing categorization challenges within process model generation, an area traditionally restricted by the scarcity of annotated datasets. This innovation demonstrates the potential of LLMs in extracting nuanced information from text, a crucial step in automating process model generation.

However, the journey from text to process diagram remains challenging, particularly in complex regulatory documents such as ISO standards and GDPR. This results show that while there has been considerable progress in automating process generation from traditional descriptive text, the complexity of regulatory text still presents significant limitations. This highlights the ongoing complexity and unresolved nature of fully automated process diagramming in certain contexts.

The search for the ideal technology to visualize extracted process information in BPMN 2.0 models also continues. Despite advances in visualization techniques developed specifically for BPMN notation, current solutions mainly deliver image files that lack the versatility of the .xml format. This limitation hinders the potential for further automated processing and refinement of the generated diagrams.

In summary, this research represents a significant step towards automated translation of natural language into process diagrams that capture real-world business processes' inherent complexity and nuances. It highlights the potential and limitations of current NLP technologies in this area and illustrates both the progress that has been made and the challenges that still need to be overcome.

Bibliography

[1] H. Leopold, J. Mendling, and A. Polyvyanyy, "Generating natural language texts from business process models," in *Active Flow and Combustion Control 2018*, R. King, Ed., vol. 141, Series Title: Notes on Numerical Fluid Mechanics and Multidisciplinary Design, Cham: Springer International Publishing, 2012, pp. 64–79, ISBN: 978-3-319-98176-5 978-3-319-98177-2. DOI: 10.1007/978-3-642-31095-9_5. [Online]. Available: http://link.springer.com/10.1007/978-3-642-31095-9_5 (visited on 03/07/2023).

[2] F. Friedrich, J. Mendling, and F. Puhlmann, "Process model generation from natural language text," in *Active Flow and Combustion Control 2018*, R. King, Ed., vol. 141, Series Title: Notes on Numerical Fluid Mechanics and Multidisciplinary Design, Cham: Springer International Publishing, 2011, pp. 482–496, ISBN: 978-3-319-98176-5 978-3-319-98177-2. DOI: 10.1007/978-3-642-21640-4_36. [Online]. Available: http://link.springer.com/10.1007/978-3-642-21640-4_36 (visited on 03/05/2023).

[3] K. Winter, H. van der Aa, S. Rinderle-Ma, and M. Weidlich, "Assessing the compliance of business process models with regulatory documents," in *Conceptual Modeling*, G. Dobbie, U. Frank, G. Kappel, S. W. Liddle, and H. C. Mayr, Eds., vol. 12400, Series Title: Lecture Notes in Computer Science, Cham: Springer International Publishing, 2020, pp. 189–203, ISBN: 978-3-030-62521-4 978-3-030-62522-1. DOI: 10.1007/978-3-030-62522-1_14. [Online]. Available: http://link.springer.com/10.1007/978-3-030-62522-1_14 (visited on 01/16/2023).

[4] S. Yu, *Improved Auto-generation of Business Process Models from Natural Language Texts of Various Complexity.pdf*, 2023.

[5] X. Sun, S. Yang, C. Zhao, and D. Yu, "Design-time business process compliance assessment based on multi-granularity semantic information," en, *The Journal of Supercomputing*, Sep. 2023, ISSN: 0920-8542, 1573-0484. DOI: 10.1007/s11227-023-05626-0. [Online]. Available: https://link.springer.com/10.1007/s11227-023-05626-0 (visited on 10/06/2023).

[6] A. Hevner, "Design science in information systems research," *Management Information Systems Quarterly*, vol. 28.1, 2008.

[7] J. vom Brocke, A. Hevner, and A. Maedche, "Introduction to design science research," in *Design Science Research. Cases*, J. vom Brocke, A. Hevner, and A. Maedche, Eds., Series Title: Progress in IS, Cham: Springer International Publishing, 2020, pp. 1–13, ISBN: 978-3-030-46780-7 978-3-030-46781-4. DOI: 10.1007/978-3-030-46781-4_1. [Online]. Available: http://link.springer.com/10.1007/978-3-030-46781-4_1 (visited on 01/16/2023).

[8] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research,"

[9] P. Achimugu, A. Selamat, R. Ibrahim, and M. N. Mahrin, "A systematic literature review of software requirements prioritization research," *Information and Software Technology*, vol. 56, no. 6, pp. 568–585, Jun. 2014, ISSN: 09505849. DOI: 10.1016/j.infsof.2014.02.001. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0950584914000354 (visited on 05/18/2023).

[10] K. Honkisz, K. Kluza, and P. Wiśniewski, "A concept for generating business process models from natural language description," in *Knowledge Science, Engineering and Management*, W. Liu, F. Giunchiglia, and B. Yang, Eds., ser. Lecture Notes in Computer Science, Cham: Springer International Publishing, 2018, pp. 91–103, ISBN: 978-3-319-99365-2. DOI: 10.1007/978-3-319-99365-2_8.

[11] R. César, B. Ferreira, L. Thom, T. Thm, and M. Fantinato, *A Semi-Automatic Approach to Identify Business Process Elements in Natural Language Texts*. Apr. 26, 2017. DOI: 10.5220/0006305902500261.

[12] M. Riefer, S. Ternis, and T. Thaler, *Mining Process Models from Natural Language Text: A State-of-the-Art Analysis*. Mar. 9, 2016.

[13] P. Bellan, M. Dragoni, and C. Ghidini, "Extracting business process entities and relations from text using pre-trained language models and in-context learning," in *Enterprise Design, Operations, and Computing*, J. P. A. Almeida, D. Karastoyanova, G. Guizzardi, M. Montali, F. M. Maggi, and C. M. Fonseca, Eds., ser. Lecture Notes in Computer Science, Cham: Springer International Publishing, 2022, pp. 182–199, ISBN: 978-3-031-17604-3. DOI: 10.1007/978-3-031-17604-3_11.

[14] S. Rinderle-Ma, K. Winter, and J.-V. Benzin, "Predictive compliance monitoring in process-aware information systems: State of the art, functionalities, research directions," *Information*

Systems, vol. 115, p. 102 210, May 2023, ISSN: 03064379. DOI: 10.1016/j.is.2023.102210. arXiv: 2205.05446[cs]. [Online]. Available: http://arxiv.org/abs/2205.05446 (visited on 06/11/2023).

[15] C. Sai, K. Winter, E. Fernanda, and S. Rinderle-Ma, "Detecting Deviations Between External and Internal Regulatory Requirements for Improved Process Compliance Assessment," en, in *Advanced Information Systems Engineering*, M. Indulska, I. Reinhartz-Berger, C. Cetina, and O. Pastor, Eds., ser. Lecture Notes in Computer Science, Cham: Springer Nature Switzerland, 2023, pp. 401–416, ISBN: 978-3-031-34560-9. DOI: 10.1007/978-3-031-34560-9_24.

[16] K. Winter and S. Rinderle-Ma, *Untangling the GDPR Using ConRelMiner*, arXiv:1811.03399 [cs], Nov. 2018. DOI: 10.48550/arXiv.1811.03399. [Online]. Available: http://arxiv.org/abs/1811.03399 (visited on 11/23/2023).

[17] N. Klievtsova, J.-V. Benzin, T. Kampik, J. Mangler, and S. Rinderle-Ma, *Conversational Process Modelling: State of the Art, Applications, and Implications in Practice*. Apr. 2023.

[18] M. J. Bommarito II, D. M. Katz, and E. M. Detterman, *LexNLP: Natural language processing and information extraction for legal and regulatory texts*, arXiv:1806.03688 [cs, stat], Jun. 2018. [Online]. Available: http://arxiv.org/abs/1806.03688 (visited on 12/05/2023).

[19] *English · spaCy Models Documentation*, en. [Online]. Available: https://spacy.io/models/en (visited on 10/18/2023).

[20] H. Zhang, Y. Dong, C. Xiao, and M. Oyamada, *Large Language Models as Data Preprocessors*, arXiv:2308.16361 [cs], Aug. 2023. [Online]. Available: http://arxiv.org/abs/2308.16361 (visited on 10/19/2023).

[21] P. D. Perera, "Knowledge-Driven Implicit Information Extraction," en,

[22] M. Dragoni, S. Villata, W. Rizzi, and G. Governatori, *Combining NLP Approaches for Rule Extraction from Legal Documents*. Jan. 2016.

[23] G. Aagesen and J. Krogstie, "BPMN 2.0 for Modeling Business Processes," *Handbook on Business Process Management 1: Introduction, Methods, and Information Systems*, pp. 219–250, Apr. 2015, ISSN: 978-3-642-45099-0. DOI: 10.1007/978-3-642-45100-3_10.

[24] "Business Process Model and Notation (BPMN), Version 2.0," en,

[25] I. Kitzmann, C. König, D. Lübke, and L. Singer, *A Simple Algorithm for Automatic Layout of BPMN Processes*. Jul. 2009, Journal Abbreviation: 2009 IEEE Conference on Commerce and Enterprise Computing, CEC 2009 Pages: 398 Publication Title: 2009 IEEE Conference on Commerce and Enterprise Computing, CEC 2009. DOI: 10.1109/CEC.2009.28.

[26] *Process Piper: Supported BPMN symbols*. [Online]. Available: https://github.com/csgoh/processpiper/wiki/Supported-BPMN-symbols (visited on 12/03/2023).

[27] *Aligning language models to follow instructions*, en-US. [Online]. Available: https://openai.com/research/instruction-following (visited on 12/06/2023).

[28] *GPT-4 API general availability and deprecation of older models in the Completions API*, en-US. [Online]. Available: https://openai.com/blog/gpt-4-api-general-availability (visited on 12/06/2023).

[29] H. Naveed, A. U. Khan, S. Qiu, *et al.*, *A Comprehensive Overview of Large Language Models*, arXiv:2307.06435 [cs], Nov. 2023. [Online]. Available: http://arxiv.org/abs/2307.06435 (visited on 11/23/2023).

[30] A. Parnami and M. Lee, *Learning from Few Examples: A Summary of Approaches to Few-Shot Learning*, arXiv:2203.04291 [cs], Mar. 2022. [Online]. Available: http://arxiv.org/abs/2203.04291 (visited on 10/24/2023).

[31] B. Mittelstadt, S. Wachter, and C. Russell, "To protect science, we must use LLMs as zero-shot translators," en, *Nature Human Behaviour*, vol. 7, no. 11, pp. 1830–1832, Nov. 2023, Number: 11 Publisher: Nature Publishing Group, ISSN: 2397-3374. DOI: 10.1038/s41562-023-01744-0. [Online]. Available: https://www.nature.com/articles/s41562-023-01744-0 (visited on 11/27/2023).

[32] *Linguistic Features · spaCy Usage Documentation*, en. [Online]. Available: https://spacy.io/usage/linguistic-features#named-entities (visited on 12/13/2023).

[33] *Hypernym noun - Definition, pictures, pronunciation and usage notes | Oxford Advanced Learner's Dictionary at OxfordLearnersDictionaries.com*. [Online]. Available: https://www.oxfordlearnersdictionaries.com/definition/english/hypernym?q=hypernyms (visited on 12/04/2023).

[34] *Linguistic Features · spaCy Usage Documentation*, en. [Online]. Available: https://spacy.io/usage/linguistic-features#vectors-similarity (visited on 12/13/2023).

[35] F. Friedrich, "AT THE SCHOOL OF BUSINESS AND ECONOMICS OF THE HUMBOLDT-UNIVERSITA¨ T ZU BERLIN," en,

[36] C. Cappiello and M. Ruiz, Eds., *CAiSE Forum*, Springer, 2019. DOI: 10.1007/978-3-030-21297-1.

[37] K. Winter, H. van der Aa, S. Rinderle-Ma, and M. Weidlich, "Assessing the compliance of business process models with regulatory documents," in *ER Conference*, Springer, 2020, pp. 189–203. DOI: 10.1007/978-3-030-62522-1_14.

[38] K. Böhmer, F. Stertz, T. Hildebrandt, *et al.*, "Application and testing of business processes in the energy domain," in *Workshop Datenbanksysteme für Business, Technologie und Web (BTW*, GI, 2017, pp. 25–32. [Online]. Available: https://dl.gi.de/handle/20.500.12116/909.

[39] ISO, *ISO 24495-1:2023*, en, Oct. 2020. [Online]. Available: https://www.iso.org/standard/78907.html (visited on 10/26/2023).

[40] ISO, *ISO/AWI 24495-2*, en. [Online]. Available: https://www.iso.org/standard/85774.html (visited on 10/26/2023).

Appendix

Detailed Test Data Sets

PET Data.

Text 1: Bicycle manufacturing

A small company manufactures customized bicycles. Whenever the sales department receives an order, a new process instance is created. A member of the sales department can then reject or accept the order for a customized bike. In the former case, the process instance is finished. In the latter case, the storehouse and the engineering department are informed. The storehouse immediately processes the part list of the order and checks the required quantity of each part. If the part is available in-house, it is reserved. If it is not available, it is back-ordered. This procedure is repeated for each item on the part list. In the meantime, the engineering department prepares everything for the assembling of the ordered bicycle. If the storehouse has successfully reserved or back-ordered every item of the part list and the preparation activity has finished, the engineering department assembles the bicycle. Afterwards, the sales department ships the bicycle to the customer and finishes the process instance.

Text 2: The workflow of a computer repair service (CRS)

A customer brings in a defective computer and the CRS checks the defect and hands out a repair cost calculation back. If the customer decides that the costs are acceptable, the process continues, otherwise she takes her computer home unrepaired. The ongoing repair consists of two activities, which are executed, in an arbitrary order. The first activity is to check and repair the hardware, whereas the second activity checks and configures the software. After each of these activities, the proper system functionality is tested. If an error is detected another arbitrary repair activity is executed, otherwise the repair is finished.

Text 3: Hotel Service

The Evanstonian is an upscale independent hotel. When a guest calls room service at the Evanstonian, the room-service manager takes down the order. She then submits an order ticket to the kitchen to begin preparing the food. She also gives an order to the sommelier (i.e., the wine waiter) to fetch wine from the cellar and to prepare any other alcoholic beverages. Eighty percent of room-service orders include wine or some other alcoholic beverage. Finally, she assigns the order to the waiter. While the kitchen and the sommelier are doing their tasks, the waiter readies a cart (i.e., puts

a tablecloth on the cart and gathers silverware). The waiter is also responsible for nonalcoholic drinks. Once the food, wine, and cart are ready, the waiter delivers it to the guest's room. After returning to the room-service station, the waiter debits the guest's account. The waiter may wait to do the billing if he has another order to prepare or deliver.

Text 4: Underwriters

Whenever a company makes the decision to go public, its first task is to select the underwriters. Underwriters act as financial midwives to a new issue. Usually they play a triple role: First they provide the company with procedural and financial advice, then they buy the issue, and finally they resell it to the public. Established underwriters are careful of their reputation and will not handle a new issue unless they believe the facts have been presented fairly. Thus, in addition to handling the sale of a company's issue, the underwriters in effect give their seal of approval to it. They prepare a registration statement for the approval of the Securities and Exchange Commission (SEC). In addition to registering the issue with the SEC, they need to check that the issue complies with the so-called blue-sky laws of each state that regulate sales of securities within the state. While the registration statement is awaiting approval, underwriters begin to firm up the issue price. They arrange a road show to talk to potential investors. Immediately after they receive clearance from the SEC, underwriters fix the issue price. After that they enter into a firm commitment to buy the stock and then offer it to the public, when they haven't still found any reason not to do it.

ISO/IEC 27001:2022.

Text 5: 2.1 Monitoring, measurement, analyses and evaluation

The organization shall determine: a) what needs to be monitored and measured, including information security processes and controls; b) the methods for monitoring, measurement, analysis and evaluation, as applicable, to ensure valid results. The methods selected should produce comparable and reproducible results to be considered valid; c) when the monitoring and measuring shall be performed; d) who shall monitor and measure; e) when the results from monitoring and measurement shall be analyzed and evaluated; and f) who shall analyse and evaluate these results. Documented information shall be available as evidence of the results. The organization shall evaluate the information security performance and the effectiveness of the information security management system.

Text 6: 2.2 Internal Audit

The organization shall plan, establish, implement and maintain an audit programme(s), including

the frequency, methods, responsibilities, planning requirements and reporting. When establishing the internal audit programme(s), the organization shall consider the importance of the processes concerned and the results of previous audits. The organization shall: a) define the audit criteria and scope for each audit; b) select auditors and conduct audits that ensure objectivity and the impartiality of the audit process; c) ensure that the results of the audits are reported to relevant management; Documented information shall be available as evidence of the implementation of the audit programme(s) and the audit results.

Text 7: 2.3 Management review

Top management shall review the organization's information security management system at planned intervals to ensure its continuing suitability, adequacy and effectiveness. The management review shall include consideration of: a) the status of actions from previous management reviews; b) changes in external and internal issues that are relevant to the information security management system; c) changes in needs and expectations of interested parties that are relevant to the information security management system; d) feedback on the information security performance, including trends in: 1) nonconformities and corrective actions; 2) monitoring and measurement results; 3) audit results; 4) fulfilment of information security objectives; e) feedback from interested parties; f) results of risk assessment and status of risk treatment plan; g) opportunities for continual improvement. The results of the management review shall include decisions related to continual improvement opportunities and any needs for changes to the information security management system. Documented information shall be available as evidence of the results of management reviews.

General Data Protection Regulation (GDPR).

Text 8: Art. 33 GDPR Notification of a personal data breach to the supervisory authority

In the case of a personal data breach, the controller shall without undue delay and, where feasible, not later than 72 hours after having become aware of it, notify the personal data breach to the supervisory authority competent in accordance with Article 55, unless the personal data breach is unlikely to result in a risk to the rights and freedoms of natural persons. Where the notification to the supervisory authority is not made within 72 hours, it shall be accompanied by reasons for the delay. The processor shall notify the controller without undue delay after becoming aware of a personal data breach. The notification referred to in paragraph 1 shall at least: a) describe the nature of the personal data breach including where possible, the categories and approximate number of data subjects concerned and the categories and approximate number of personal data records

concerned; b) communicate the name and contact details of the data protection officer or other contact point where more information can be obtained; c) describe the likely consequences of the personal data breach; d) describe the measures taken or proposed to be taken by the controller to address the personal data breach, including, where appropriate, measures to mitigate its possible adverse effects. Where, and in so far as, it is not possible to provide the information at the same time, the information may be provided in phases without undue further delay. The controller shall document any personal data breaches, comprising the facts relating to the personal data breach, its effects and the remedial action taken. That documentation shall enable the supervisory authority to verify compliance with this Article.

Text 9: Art. 6 GDPR Lawfulness of processing

Processing shall be lawful only if and to the extent that at least one of the following applies: a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes; b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract; c) processing is necessary for compliance with a legal obligation to which the controller is subject; d) processing is necessary in order to protect the vital interests of the data subject or of another natural person; e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller; f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child. Point (f) of the first subparagraph shall not apply to processing carried out by public authorities in the performance of their tasks. Member States may maintain or introduce more specific provisions to adapt the application of the rules of this Regulation with regard to processing for compliance with points (c) and (e) of paragraph 1 by determining more precisely specific requirements for the processing and other measures to ensure lawful and fair processing including for other specific processing situations as provided for in Chapter IX. The basis for the processing referred to in point (c) and (e) of paragraph 1 shall be laid down by: a) Union law; or b) Member State law to which the controller is subject. The purpose of the processing shall be determined in that legal basis or, as regards the processing referred to in point (e) of paragraph 1, shall be necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller.

That legal basis may contain specific provisions to adapt the application of rules of this Regulation, inter alia: the general conditions governing the lawfulness of processing by the controller; the types of data which are subject to the processing; the data subjects concerned; the entities to, and the purposes for which, the personal data may be disclosed; the purpose limitation; storage periods; and processing operations and processing procedures, including measures to ensure lawful and fair processing such as those for other specific processing situations as provided for in Chapter IX. The Union or the Member State law shall meet an objective of public interest and be proportionate to the legitimate aim pursued.

Where the processing for a purpose other than that for which the personal data have been collected is not based on the data subject's consent or on a Union or Member State law which constitutes a necessary and proportionate measure in a democratic society to safeguard the objectives referred to in Article 23(1), the controller shall, in order to ascertain whether processing for another purpose is compatible with the purpose for which the personal data are initially collected, take into account, inter alia: a) any link between the purposes for which the personal data have been collected and the purposes of the intended further processing; b) the context in which the personal data have been collected, in particular regarding the relationship between data subjects and the controller; c) the nature of the personal data, in particular whether special categories of personal data are processed, pursuant to Article 9, or whether personal data related to criminal convictions and offences are processed, pursuant to Article 10; d) the possible consequences of the intended further processing for data subjects; e) the existence of appropriate safeguards, which may include encryption or pseudonymisation.

Text 10: Art. 15 GDPR Right of access by the data subject

The data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data and the following information: 1. the purposes of the processing; 2. the categories of personal data concerned; 3. the recipients or categories of recipient to whom the personal data have been or will be disclosed, in particular recipients in third countries or international organisations; 4. where possible, the envisaged period for which the personal data will be stored, or, if not possible, the criteria used to determine that period; 5. the existence of the right to request from the controller rectification or erasure of personal data or restriction of processing of personal data concerning the data subject or to object to such processing; 6. the right to lodge a complaint with

a supervisory authority; 7. where the personal data are not collected from the data subject, any available information as to their source; 8. the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.

Where personal data are transferred to a third country or to an international organisation, the data subject shall have the right to be informed of the appropriate safeguards pursuant to Article 46 relating to the transfer.

The controller shall provide a copy of the personal data undergoing processing. For any further copies requested by the data subject, the controller may charge a reasonable fee based on administrative costs. Where the data subject makes the request by electronic means, and unless otherwise requested by the data subject, the information shall be provided in a commonly used electronic form.

The right to obtain a copy referred to in paragraph 3 shall not adversely affect the rights and freedoms of others.

Text 11: Art. 20 GDPR Right to data portability

The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the personal data have been provided, where: a) the processing is based on consent pursuant to point (a) of Article 6(1) or point (a) of Article 9(2) or on a contract pursuant to point (b) of Article 6(1); and b) the processing is carried out by automated means.

In exercising his or her right to data portability pursuant to paragraph 1, the data subject shall have the right to have the personal data transmitted directly from one controller to another, where technically feasible.

The exercise of the right referred to in paragraph 1 of this Article shall be without prejudice to Article 17. That right shall not apply to processing necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller. The right referred to in paragraph 1 shall not adversely affect the rights and freedoms of others.

Text 12: Art. 7 GDPR Conditions for consent

Where processing is based on consent, the controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data.

If the data subject's consent is given in the context of a written declaration which also concerns other matters, the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language. Any part of such a declaration which constitutes an infringement of this Regulation shall not be binding.

The data subject shall have the right to withdraw his or her consent at any time. The withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal. Prior to giving consent, the data subject shall be informed thereof. It shall be as easy to withdraw as to give consent.

When assessing whether consent is freely given, utmost account shall be taken of whether, inter alia, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract.

Text 13: article 16 and 19 partly

The data subject shall have the right to obtain from the controller without undue delay the rectification of inaccurate personal data concerning him or her. Taking into account the purposes of the processing, the data subject shall have the right to have incomplete personal data completed, including by means of providing a supplementary statement.

The controller shall communicate any rectification or erasure of personal data or restriction of processing carried out in accordance with Article 16.

Text 14: Art. 17 GDPR Right to erasure ('right to be forgotten')

The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay where one of the following grounds applies: a) the personal data are no longer necessary in relation to the purposes for which they were collected or otherwise processed; b) the data subject withdraws consent on which the processing is based according to point (a) of Article 6(1), or point (a) of Article 9(2), and where there is no other legal ground for

the processing; c) the data subject objects to the processing pursuant to Article 21(1) and there are no overriding legitimate grounds for the processing, or the data subject objects to the processing pursuant to Article 21(2); d) the personal data have been unlawfully processed; e) the personal data have to be erased for compliance with a legal obligation in Union or Member State law to which the controller is subject; f) the personal data have been collected in relation to the offer of information society services referred to in Article 8(1).

Where the controller has made the personal data public and is obliged pursuant to paragraph 1 to erase the personal data, the controller, taking account of available technology and the cost of implementation, shall take reasonable steps, including technical measures, to inform controllers which are processing the personal data that the data subject has requested the erasure by such controllers of any links to, or copy or replication of, those personal data.

Paragraphs 1 and 2 shall not apply to the extent that processing is necessary: a) for exercising the right of freedom of expression and information; b) for compliance with a legal obligation which requires processing by Union or Member State law to which the controller is subject or for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller; c) for reasons of public interest in the area of public health in accordance with points (h) and (i) of Article 9(2) as well as Article 9(3); d) for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) in so far as the right referred to in paragraph 1 is likely to render impossible or seriously impair the achievement of the objectives of that processing; or e) for the establishment, exercise or defence of legal claims.

Smart Meter.

Text 15:

The aim is to remotely interrupt the supply of electrical energy for an object, as a result of a customer change or due to open collection claims, via the central system. The shutdown takes place by opening the breaker present in the terminal. From the central system, a shutdown command is transmitted to the affected terminal. The terminal executes the received command, i.e., the internal breaker interrupts the power supply of the customer system, returns the status of the shutdown to the central system and creates an entry in the logbook. This status must be visible on the terminal. With collection there is no possibility of an immediate on-site reconnection by the customer after the

blocking. Input is either deregistration of current customer, customer moves out without new tenant or customer does not pay after admonition process. The supply of electrical energy is interrupted, the status "interrupted" is displayed at the terminal and in the central system. An entry in the logbook of the terminal has been created.

Text 16: The prepayment function is realized via the central system. The terminal only acts as a switching device. The central system generates the shutdown command by comparing credits with consumption. The comparison of credits with actual consumption value should be made periodically. The basis is the daily readings. A "real-time monitoring" is therefore not necessary. Shutdown may only be carried out in compliance with legal conditions. Each status change of the terminal must be displayed on the device and also create an entry in the logbook. There may be a corresponding customer information about the current balance from the central system to the customer. Credit or credit limit is consumed and shutdown time is within the legally allowed time window. The supply of electrical energy is interrupted and the status of the breaker is displayed on the terminal, an entry in the logbook of the terminal is generated and the status is transmitted to the central system.

Text 17: Exceeding an active power limit set in the meter and activated shuts down the customer system. For the customer, immediate reclosing on site, ie directly at the meter, must be possible. The shutdown, the message of the reclosing and the exceeding of the power threshold must be able to be transmitted to the central system as ALARM or EVENT and also visualized on the meter. Breaker status is, e.g., OFF and READY. An entry is also created in the logbook. In the logbook, in addition to changing the breaker status, the reason for the shutdown, e.g. "Power limit exceeded by xxx watts" can be logged. Exceeding the power threshold, for example in the context of the basic supply. Exceeding the power threshold was entered in the logbook of the meter and transmitted to the central system as an ALARM or EVENT.

Text 18: There is an elimination of the terminal on-site by means of a mobile device via the service interface (WZ). In order to meet the strict data protection regulations, the cryptographic parameters that authorize the device to be switched off must be stored on the mobile device. It must therefore not be possible with these cryptographic parameters to turn off several or all terminals of the terminal equipment. The cryptographic parameters must comply with the security concept from "OE Requirements Catalog End-to-End Security Smart Metering". Input is either deregistration of current customer or customer moves out without new tenant. The breaker of the terminal has

switched off the system, indicating this on the device. An entry in the logbook of the terminal has been created.

Text 19: The operating limit of the meter is set by the control panel. Optionally this is possible for both energy directions. Change requirement threshold of the power limitation, for example due to contract adjustments. New threshold set, EVENT sent to the central system and logbook entry made on the meter. The active power limit value of the meter must be able to be set locally via the service interface (WZ) of the meter. Optionally this is possible for both energy directions. A logbook entry is generated. This state is transmitted to the central system as an ALARM or EVENT when the transmission link (WAN) is available.

Text 20: The customer interface (H1) on the meter is to be activated or deactivated from the central system, whereby it is deactivated by adefault. The control center receives an information status as ALARM or EVENT. Customer wants to use or not use customer interface (H1) anymore. Customer interface (H1) on the meter is activated / deactivated and there is an entry in the logbook. meter reports its status to the control panel as ALARM or EVENT.

Text 21: In the course of the so-called opt-out regulation, it must be possible for customers who refuse to install an intelligent electricity meter to deactivate the load profile recording in the meter. Activation / deactivation of load profile recording must be possible remotely (WAN interface). Each of these changes must be logged in the relevant legal logbook of the meter. If the customer is opt-out, it must be displayed on the meter accordingly (for example, display, LED). Opt-out request of the customer to implement at the NL. Load profile recording is activated / deactivated and the information has been transmitted to the central system as ALARM or EVENT. A logbook entry has been made in the meter and the current status OPT-OUT is displayed on the meter.

Text 22: The load switching device is a completely independent of the meter device and serves as a possible replacement for a ripple control device or a timer. The load switching device has its own communication interface and can thus make contact with the control panel. In principle, the relays follow a switching program specified by the central system and stored in the load switching device. However, it must be possible from the central system to override the circuit program with a spontaneous command (eg relay XY "OFF" or "ON"). The next opposite command (either from the switching table or from an external source) changes the state of the switching device. The position

of each relay or each switching cycle specified by the switching program must be reported back to the central system. The relay settings must also be visually displayed on site.

The central system sends a command to override a relay ("off" or "on"). The service interface is used to override a relay ("Off" or "On").

The corresponding relay in the load switching device switches to the desired status, reports this back to the central system and the position of the relay on the load switching device is also visible and a logbook entry is made.

Text 23: For each relay in the load switching device, an independent, independent of the other relay switching program should be configurable. It should be possible to subdivide the circuit program into daily, weekly, seasonal and annual programs taking into account weekly, holiday and special days. The switching program is managed centrally and transmitted to the load switching device via the communication paths. For control purposes, the circuit program must also be read-back. Any change to the circuit program, regardless of whether it is remotely executed, must be logged in a logbook. The central system sends changed switching programs to the load switching device. The load switching device receives the changed switching programs (feedback to the central system) and records this in the logbook.