Proceedings of the 57th Hawaii International Conference on System Sciences | 2024

From Text to Intelligent Services in Knowledge Intensive Decision Processes: Text2Chat

Alexandre Goossens
KU Leuven
alexandre.goossens@kuleuven.be

Jan Vanthienen
KU Leuven
jan.vanthienen@kuleuven.be

Abstract

Knowledge-intensive processes intertwine processes and decisions. Such processes often end whenever the final decision outcome has been obtained and communicated to the stakeholder. However, stakeholders often desire additional clarification on decision outcomes which requires processing the decision and process information. The decision logic is often described in various internal and external textual documents but are often difficult to interpret. There exist various ways to generate and provide advice and explanation based on textual descriptions. We present an overview of existing research fragments and new research on integrating the pieces into a framework. This paper presents the Text2Chat framework for generating advice and explanation from the process description text, containing various tracks. Each generation possibility is explained together with its advantages and disadvantages. By introducing Text2Chat, this paper offers insights into processing textual knowledge-intensive process information for effective stakeholder advice provision.

Keywords: Knowledge-Intensive Processes, Decision Modeling, Explainability, Large Language Models, Deep learning

1. Introduction

Knowledge-intensive processes, such as loan approvals or student grant eligibility processes, are processes in which processes and multiple (complex) decisions occur together and complement one another. Generally, knowledge-intensive processes are described with the use of textual documents and regulations but explaining decisions using these texts is not ideal as natural language is not always void of ambiguity making it difficult for stakeholders to completely understand how a process or decision has to be made (Vanthienen, 2021). Hence stakeholders often find themselves questioning the decision outcome. To compensate this, the information described in textual documents needs to be processed by either the stakeholder or the organization. To provide advice and explanation, companies most commonly set up FAQ pages and helpdesks which are expensive and where customer satisfaction depend on both information quality (Negash et al., 2003) and customer experience (Van Velsen et al., 2007). A more recent alternative to provide advice has emerged with the use of Generative Pre-trained Transformers (GPT) technologies which can analyze textual descriptions (Brown et al., 2020). A last alternative to provide advice and explanations is to use decision models to model, execute and communicate operational decisions. Once a decision has been modeled, decision models can be used as knowledge base to provide advice about decisions by directly interacting with these with the use of chatbots (Estrada-Torres et al., 2021; Etikala et al., 2021; Goossens, Maes, et al., 2023; Vandevelde et al., 2021). However it remains unclear how advice and explanations can best be generated and provided to stakeholders from these possibilities. This paper contributes to the literature by analyzing how textual knowledge-intensive process information can currently best be processed as well as analyzing how this processed information can best be activated to provide advice to stakeholders. To this purpose, a Text2Chat framework is introduced containing the various advice generation tracks (or possibilities).

The remainder of this paper is structured as follows: Section 2 explains the problem followed by Section

URI: https://hdl.handle.net/10125/107105
978-0-9981331-7-1

3 introducing the research questions. Section 4 deals with the advice generation framework and a more in depth analysis of each advice generation path. Section 5 discusses the results and remarks whilst Section 6 concludes this paper.

2. Problem Statement

Companies and organization execute knowledge-intensive processes on a daily basis where most commonly a process ends whenever the outcome of the decision has been obtained and communicated to the stakeholders. However once the outcome of a process has been communicated, the stakeholder is often left with a variety of questions regarding the decision-making process such as: *Why did I receive this outcome?*, *What influences my decision?* or *What if a certain value is changed?* Therefore providing advice and explanation is still part of a good customer service even after the process outcome has been communicated to the stakeholder. Good advice and explanations do have some requirements though. The provided advice and explanation need to be complete, correct, personalized and relevant to the stakeholders. Lastly, advice needs to be consistently the same for all stakeholders with similar questions and characteristics. On top of that, advice information often finds itself encapsulated in textual sources such as regulations, guidelines, internal documents and process descriptions (Vanthienen, 2021). The textual descriptions can be analyzed manually or automatically and can also be used to build process or decision models which in turn can be used as a basis to provide manual or automated advice and explanation afterwards. Next to text, decisions in knowledge-intensive processes are also often discovered in event logs (Bazhenova et al., 2016; De Smedt et al., 2017), process models (Bazhenova and Weske, 2016) or structural data-flow relations (van der Aa et al., 2016). Currently the main solutions to provide advice and explanation are:

- Helpdesks and FAQ pages, which can be considered manual advice.
- Chatbots and large language models such as GPT-3 and ChatGPT[1].
- Chatbots powered by decision models.

Unfortunately, the first two solutions do not always meet the customer needs due to low information quality or bad experiences such as redirecting to other helpdesks (Van Velsen et al., 2007). To facilitate the communication between stakeholders and execution of operational decisions, the Object Management Group (OMG) introduced the Decision Model and Notation (DMN) standard in 2015 (OMG (2015a)). DMN models are useful for advice generation as they unambiguously capture the decision logic and decision structure. Despite the benefits of DMN models, the time required to learn how to actually create these decision models is significant. A DMN model consists of two parts each capturing a different aspect of operational decisions. The initial component, known as a Decision Requirement Diagram (DRD), provides a visual representation of the information used to make a decision and illustrates the interconnections between different decisions. The second part represents the logic driving the decisions. In DMN, a common way of representing decision logic is with the use of decision tables which easily ensure consistency and correctness thanks to their design (Huysmans et al., 2011; Vandevelde et al., 2021; Vanthienen et al., 1998). A research agenda regarding DMN can be found in (Figl et al., 2018).

In short, there are multiple ways to provide advice and explanation to users but it is currently unclear how this can best be given to users? Next to that, given that process and decision models can also be used to provide advice, another question that arises is how can textual information best be processed to generate these process or decision models?

3. Research Questions

In the problem statement section, several advice generating tracks are introduced. Next to providing manual advice and providing advice using GPT-3, it is also possible to model all the relevant decision information in a decision model which can then be used to provide advice to stakeholders. This makes that another challenge is to process the textual sources containing all the process and decision information to create such process and decision models. Note that the extraction of process models from text is considered out of scope for this study as process models focus more control-flow and less on logic which is more present in advice. As such, this paper tries to answer the following research questions:

- What are the possible advice generating tracks? For each track:
  - What are the requirements and the assumptions?
  - What are the advantages and disadvantages?
- Given that certain advice generation tracks make use of DMN models, which track can best be

[1]https://openai.com/blog/chatgpt

used to generate DMN models from textual descriptions?

- In a short survey we investigate whether the DMN chatbot is of equal explanation quality as the manual advice generation track? (Given that the DMN model is available.)

4. Text2Chat Framework

Figure 1 visualize all these advice generation tracks in a so-called Text2Chat Framework. The multiple ways to go from textual sources to advice have all been introduced in the previous sections, namely:

- Manual Advice without decision models: The manual analysis track is at the top of Figure 1.
- Large Language Model: GPT-3 advice generation is shown at the bottom of Figure 1.
- In the middle, the DMN model advice generation tracks are shown. The assumption, of course, is that the relevant decision logic can be modeled in DMN. This is acceptable because of significant decision logic present in knowledge-intensive processes (Di Ciccio et al., 2012). The framework proposes to build these DMN models either manually (A) or automatically (B). Next, the Text2Chat framework also indicates that it is possible to reason with DMN models manually (C) or automatically (D) with two variants described later.
  - manual advice (tracks: (A or B) + C)
  - automatic advice using DMN-powered chatbots (tracks: (A or B) + D)

The remainder of this section will dive deeper into each advice generation track. In the last subsection, a preliminary survey investigating the explanation quality of the DMN explanation track compared to manual advice track is discussed.

4.1. Manual Decision Analysis, Execution and Advice

Manually analyzing process & decision information is a complex and time consuming task. Next to actually analyzing the textual documents, there is also the challenge of managing this information. As said previously, this information is often gathered again in additional textual documents that helpdesks or FAQ pages use to provide information. This still leads to some problems. Firstly, helpdesks do no always meet all the demand and quality expectations of users (Negash et al., 2003; Van Velsen et al., 2007). Moreover, these helpdesks need to be staffed and trained to achieve a reasonable quality which is expensive. Thirdly, the quality of the advice given to the stakeholder remains dependent on the clarity of the textual sources as well as the human interpretation of these documents. This observation has been confirmed with a user survey conducted by Goossens, Maes, et al. (2023) which concluded that manual advice generation does not consistently yield complete and correct explanations.

4.2. Decision advice and execution with GPT-3

GPT models are a set of models which have been trained on massive amounts of data allowing them to perform a wide range of Natural Language Processing (NLP) tasks and be very good at it (Brown et al., 2020). Due to the fact that these models are able to generalize and perform novel tasks very quickly given a few examples, a plethora of GPT application studies has emerged such as in the medical domain (Lee et al., 2023), legal domain (Katz et al., 2023) or financial domain (Niszczota and Abbas, 2023).

Despite their impressive capabilities, GPT technologies are black boxes, making it hard to understand why and how decisions are made, and what information was used (Goossens and Vanthienen, 2023). Also, the answer of GPT-3 depends on the formulation of the question which is problematic for consistent advice generation (Jojic et al., 2023; Wei et al., 2022). Next, when facing GPT-3 with new facts that might be conflicting with the training set, GPT-3 might use the new or the old facts but that is not always clear. Lastly, GPT-3 does not personalize responses for users; instead, it identifies relevant text portions for explanation, shifting the interpretation to stakeholders. (Goossens and Vanthienen, 2023). In short, GPT-3 is perfect as a user-interface with its impressive capabilities in formulating natural sentences, however the information processing and explanation capabilities might benefit from an approach where reasoning and the used information can be traced back in an easier manner.

4.3. Advice Generation using DMN Models

In the following subsection, the complete advice generation tracks using DMN models are presented. DMN models are used to model and execute operational decisions which are often present in knowledge intensive processes. These models are often created based on available textual descriptions such as guidelines or regulations. In Figure 1, this corresponds with the following advice generation paths: (A or B) with (C or D). This section first introduces a

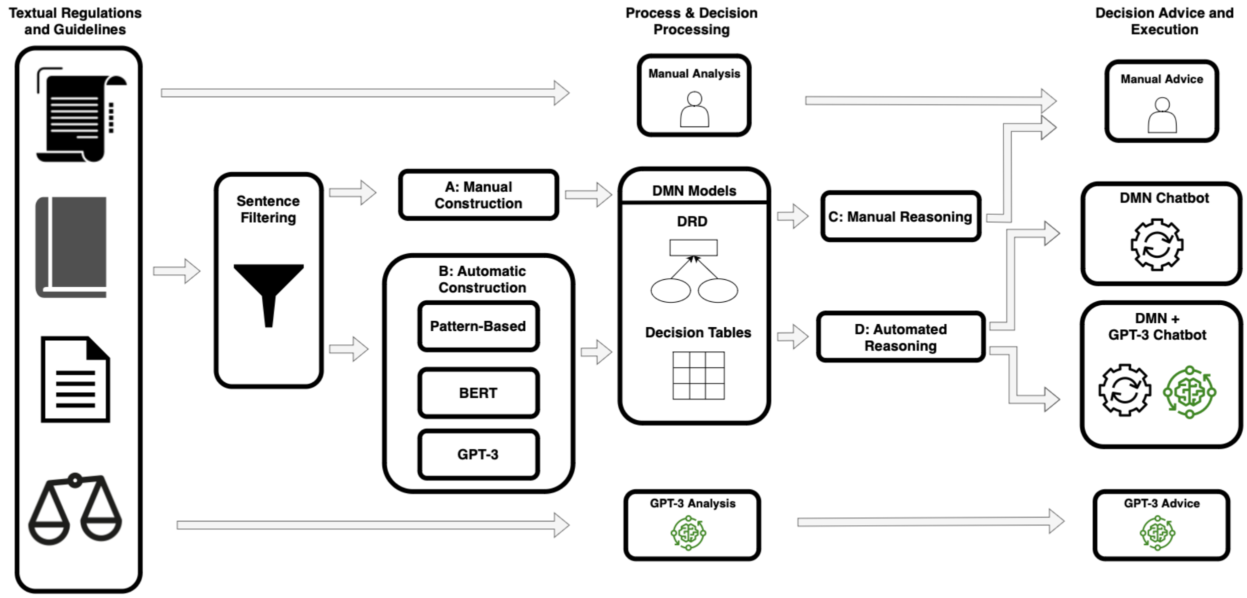

Figure 1. Text2Chat: advice generation framework

method to filter sentences dealing with processes and decisions. The second subsection deals with the three main methods to extract DMN models from text as well as with the second research question which asked how to extract DMN models from texts. Next, methods to provide advice and explanations using DMN are elaborated.

4.3.1. Step 1: Filtering process and decision sentences The textual documents describing knowledge-intensive processes often describe the decisions and processes together. As such, when analyzing these documents it is important to know which parts deal with the control flow (process) or which parts deal with the decision driving the knowledge-intensive process. However if decision and process models need to be discovered automatically from textual descriptions, an automatic classifier identifying which sentences deal with what would therefore be ideal. This study proposes a first process and decision sentence classifier to perform this task.

To train a process-decision sentence classifier, a public dataset consisting of 809 sentences describing decisions (Goossens, De Smedt, et al., 2023b) and a dataset with 417 sentences describing processes (Bellan et al., 2023) have been merged together with each sentence labeled as *Decision* or *Process*. The assumption is made that sentences do not describe both decision and control-flow logic. In the preprocessing step, the sentences underwent several transformations. Initially, the text was tokenized, splitting it into individual words or tokens which were then lowercased to ensure case insensitivity. Next, lemmatization was applied to reduce each word to its base or dictionary form, accounting for different grammatical variations. Additionally, punctuation marks as well as common English articles such as ”an,” ”a,” and ”the” were removed.

To vectorize the words, Bag of Words and Term Frequency-Inverse Document Frequency (TF-IDF) were used. Bag of Words representation simply counts the number of times a particular word appears in the text, disregarding the context. On the other hand, TF-IDF not only counts the occurrences but also assigns weights to each word based on its appearance across the entire document collection allowing TF-IDF to highlight more relevant words that are likely to carry more meaning or importance, while downplaying common words like ”I” or ”am” that occur frequently. For model training and evaluation, the dataset was split into an 80% training set and a 20% test set. Stratification was employed to ensure a proportional distribution of class labels in both sets. The classification models used in this analysis were naive bayes and support vector machines.

Tables 1 reports the precision (P), recall (R) and F1-score (F1) on the sentence classification task. As can be seen, basic machine learning models combined with traditional vectorizing techniques are not only quickly implemented but can also categorize sentences well into *Decision* or *Process* with support vector machines and TF-IDF achieving the overall results. The likely reason for this superior performance can be attributed to the

Table 1. Text classification results

	Naive Bayes						Support Vector Machines					
	Bag of Words			TF-IDF			Bag of Words			TF-IDF		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Decision	0.97	0.98	**0.98**	0.89	**1.00**	0.94	**0.98**	0.97	0.98	**0.98**	0.99	**0.98**
Process	0.96	0.95	0.96	**1.00**	0.76	0.86	0.95	**0.96**	0.96	0.97	**0.96**	0.97
accuracy			0.97			0.92			0.97			**0.98**
macro avg	0.97	**0.97**	0.97	0.95	0.88	0.90	0.97	**0.97**	0.97	**0.98**	**0.97**	**0.98**
weighted avg	0.97	0.97	0.97	0.93	0.92	0.92	0.97	0.97	0.97	**0.98**	**0.98**	**0.98**

characteristic of process descriptions, which typically include a temporal aspect and are expressed in the active tense. On the other hand, sentences describing decisions often focus on rules and dependencies, and are not commonly formulated within a sequential context.

Given a description of a knowledge-intensive process, it seems possible with the use of a classifier to identify which sentences deal with processes and which sentences with decisions but a further evaluation of real life textual descriptions would help confirm these findings. Friedrich et al. (2011) concluded in their study that up to 60% of the total time spent in process management is dedicated to manually designing process models. Hence, providing some automation for this task would be beneficial. In the next step, the sentences identified as describing processes can be used to extract process models with approaches described in (Bellan et al., 2023; Friedrich et al., 2011; Neuberger et al., 2023). Because this study focuses on decision advice and explanation, the next subsection focuses on how to construct DMN models from sentences classified as describing decisions.

4.3.2. Step 2: DMN Extraction Methods from Text

Once the sentences describing decisions have been identified from textual documents, the main challenge is to extract DMN models from these. This subsection discusses the second research question which asks how DMN models can best be extracted from texts. The extraction of DMN models from text consists of two tasks: identifying the structure of a decision and finding the decision logic driving the decision. There are currently three main methods to extract DMN models from text namely: using patterns, using a Named Entity Recognition (NER) pipeline or directly using GPT-3.

Pattern-based DMN extraction: Pattern-based approaches are utilized to identify different sentence formulations or patterns within sentences. The aim is to encompass a wide range of formulations in order to build DMN models from sentences effectively. Etikala et al. (2020) employs pattern-based approaches to extract DRDs from an artificial dataset. Similarly, Quishpi et al. (2021) explores the extraction of complete DMN models using the same artificial dataset of Etikala et al. (2020). Research has also been conducted on extracting decision logic from texts using pattern-based approaches. Kluza and Honkisz (2016) investigated how decision rules can be extracted from structured texts written in Semantics Of Business Vocabulary And Business Rules (SBVR) (OMG (2015b)). On the other hand, Arco et al. (2021) investigated the construction of decision tables from individual unstructured single sentences.

Despite their encouraging results, pattern-based approaches present several disadvantages. Firstly, they encounter difficulties when faced with previously unseen patterns, limiting their scalability. Furthermore, developing these pattern-based approaches requires substantial manual effort, as it necessitates mapping as many different formulations as possible. Moreover, the manual pattern identification process needs to be repeated for each language, adding to the resource-intensive nature of these approaches.

NER-pipeline with BERT DMN extraction: The DMN extraction pipeline described in (Goossens, De Smedt, et al., 2023b) offers a (semi-)automatic pipeline to extract DRDs and logical statements from textual descriptions. Firstly, a classifier is used to identify sentences describing decision logic or the decision structure. Then, for both sentence classes, a BERT model is finetuned on a NER task. Specifically, for dependency sentences, a BERT model is finetuned to predict the decision and input variables, while for logic sentences, BERT is finetuned to identify the different parts such as the IF-part and THEN part, which together form a logical statement. Once these elements are predicted, they are assembled to create a logical statement table and a final DRD. The approach does not format the logical statements into decision tables suitable for immediate use, requiring additional modeling by the user.

In terms of performance, this approach was evaluated on DRD extraction on 6 real-life examples. This approach achieved 93% on Precision, 97% on

Recall and 95% on F1-score on the identification of the decision and input elements. Next, it achieved 87% on Precision, 97% on Recall and 91% on F1-score on correctly linking these inputs and decision elements together. One significant advantage of this approach is its scalability, as it is more robust against new formulations and unseen patterns compared to pattern-based approaches. Just like pattern-based approaches, this approach struggles with coreference resolution. This is a problem where the same concept is described with different words, e.g., Eligibility period and period of eligibility. Another drawback is that when applying this pipeline to new languages, a new dataset must be collected and manually labeled to train language-specific BERT models.

GPT-3 for DMN Extraction: The utilization of GPT-3 for DMN model extraction has been explored in two papers, namely (Goossens, De Smedt, et al., 2023a; Goossens, Vandevelde, et al., 2023). In the context of extracting DRDs, GPT-3's performance was increased with a few examples to improve its performance through few-shot learning (Brown et al., 2020). The results obtained by GPT-3 on the same dataset as the NER-pipeline with BERT (Goossens, De Smedt, et al., 2023b) were comparable to those achieved by the NER-pipeline utilizing BERT. Consequently, GPT-3 seems to be an interesting method to extract DRDs from text, as it requires less effort for finetuning while still yielding satisfactory results. Regarding the extraction of decision logic, the study featured in (Goossens, Vandevelde, et al., 2023) investigated GPT-3's reasoning abilities and its capacity to construct decision tables from textual descriptions. The study concludes that GPT-3 is capable of identifying what a decision is about and based on what information a decision is made. But also concludes that GPT-3 is not able to correctly deduce all the decision rules. Instead, it is suggested that GPT-3 should be used as a tool to create template decision tables only filled with the input and output variables without the actual decision rules, which can then be further refined by the modeler.

The main advantage of GPT-3 lies in its ability to achieve remarkable results through few-shot learning. Additionally, GPT-3 demonstrates better proficiency in handling coreference resolution challenges and comprehending various formulations. Nevertheless, a disadvantage of GPT-3 is its sensitivity to the formulation of prompts. Therefore, it is crucial to perform prompt engineering before implementing GPT-3 for the extraction of DRDs or decision tables, ensuring optimal performance and accurate results (Jojic et al., 2023; Wei et al., 2022).

One of the main challenges that persists is the extraction of decision tables as it requires the identification of multiple decisions, the variables involved and all the values relevant for the decision logic. The introduction of the more powerful GPT-4 model holds promise for addressing this challenge. With its enhanced reasoning capabilities, GPT-4 could potentially enable the automatic extraction of decision tables from textual descriptions. However, extensive testing is needed to evaluate the performance of GPT-4 in this specific task, which is beyond the scope of the current paper. For the subsequent section, we assume that the organization has a DMN model at hand, either obtained automatically or manually, with complete decision rules and dependencies.

4.3.3. Step 3: Explainability with DMN models using chatbots In Figure 1, there are two advice generation paths with DMN namely path C and path D. This section first introduces various reasoning mechanisms possible with DMN. These can be used to provide manual DMN advice and explanation by directly providing the DMN models to the stakeholders or by using them in helpdesks as an information source (path C). It is also possible to use these to power DMN chatbots which are able to provide explanations for various questions to the user (path D). Next, this section identifies two ways of designing DMN chatbots namely solely using a DMN reasoning engine or enhancing a DMN chatbot with GPT-3 technology to interact with the user.

DMN Reasoning Bot: There are various reasoning mechanisms possible with decision tables and DMN. The first two scenarios are decision execution scenarios whilst the rest are decision explanations scenarios.

1. Complete reasoning: given all the inputs what is the output?
2. Reasoning with unknown information: derive with the available information what the possible outcomes are of a decision.
3. What if?: What if one of the inputs changes value?
4. What should?: Answer the question: *What should variable X be to get Y?*
5. How?: explains to the stakeholder all the possibilities to achieve a certain outcome.
6. Why?: Using the decision models, return the rules that resulted in a the outcome a stakeholder received.
7. Sensitivity: allows stakeholders to know how much variables can change without changing the current outcome.

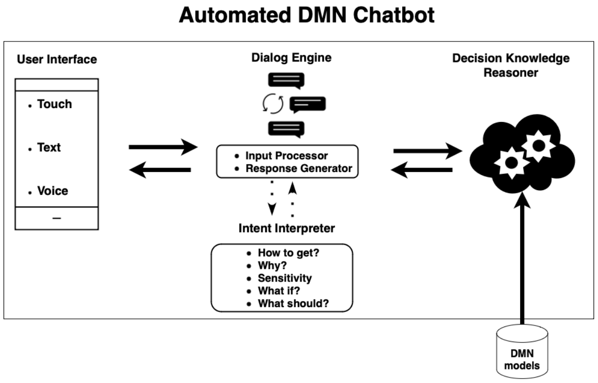

Figure 2. GPT-3 integrated DMN chatbot components (Goossens and Vanthienen, 2023)

In (Etikala et al., 2021), a chatbot enhanced with voice capabilities and a NLP intent interpreter is developed making use of its own reasoning engine. Vandevelde et al. (2021) developed a reasoning engine and showcased it with a basic chatbot without user interface. The previous reasoning engine was improved in the following study (Carbonnelle et al., 2023). Another user-oriented DMN chatbot was developed by Estrada-Torres et al. (2021) but does not have advice generation capabilities. Goossens, Maes, et al. (2023) merge the first two chatbots together to have a chatbot with a powerful reasoning engine (Vandevelde et al., 2021) and enhanced user-friendly capabilities (Etikala et al., 2021).

GPT-3 integrated DMN chatbot:

Goossens, Maes, et al. (2023) surveyed whether the enhanced DMN chatbot's explanations were considered superior to textual analysis by stakeholders. The survey's initial finding is that users offer more accurate explanations with the DMN chatbot than with textual descriptions. Nevertheless, users found the DMN chatbot less user-friendly, struggling to determine the appropriate reasoning scenario for their questions. This highlights the need for a DMN chatbot that can intuitively select the right scenario to answer user questions.

The study conducted in (Goossens and Vanthienen, 2023) investigated how GPT-3 and a DMN chatbot can be used together. The idea is to use GPT-3 to identify which reasoning scenario needs to be triggered based on a question of a user. Figure 2 visualizes how the various chatbot components work together. In short, the dialog engine, which is responsible for guiding the conversation, is extended with GPT-3 which is used as an intent interpreter under the hood. This then forwarded to the reasoning engine which provides an answer to the user. The study also investigated the explanation quality of only using GPT-3 and concludes that the integrated GPT-3 and DMN chatbot is able to consistently provide better and more relevant explanations about operational decisions.

In short, it seems that the advice generation track making use of a GPT-3 integrated DMN chatbot is worth further investigation.

4.4. Comparing manual advice generation with automated DMN advice generation

We organized a large scale preliminary online survey with 108 participants investigating the explanation quality of an internet information page compared to a DMN chatbot (De Conick and Marchoul, 2023). The aim of the study is to investigate the interplay between various variables and to see whether these relations differ between using a DMN chatbot and an information page. Compared to (Goossens, Maes, et al., 2023), this larger survey did not make use of textual descriptions but rather provided complete internet information web pages as is it more commonly used in reality.

The participants were first asked about their *Age*, *Level of Education*, *Knowledge in modelling* (BPMN, UML, DMN,...) and *Tech Saviness*. Next, they had to answer various questions given a situation using either an information page or a DMN chatbot (not enhanced with GPT-3). In total, the questions covered 5 reasoning scenarios (scenarios 1, 2, 3, 4 and 7 (see subsection 4.3.3). The answers were then evaluated on their completeness and correctness. This is captured by *Quality of extracted information*. *Trust* captures whether participants either trust the information page or the chatbot as an explanation source. The participants were also asked what they found easier to use, the DMN chatbot or the information page (*Ease of use*). *Preference of usage* measures whether participants prefer using the chatbot or the information page. The time needed to answer the questions was also measured (*Information speed retrieval*). Next, a Structural Equation Model (SEM) was modeled to test whether *Age*, *Level of Education* and *Knowledge in modelling* influence *Ease of use* and *Tech Saviness* as well as to test whether *Tech Saviness* and *Ease of use* influence *Trust*, *Quality of extracted information*, *Preference of usage* and *Information speed retrieval*.

After checking the measurement models, the structural models were fitted. Figures 3 and 4 show the respective resulting Structural Equation Models for the DMN chatbot and the information page from this survey. The arrows indicate the measured correlation between the latent variables whilst bold arrows indicate that the correlations are statistically significant (.05).

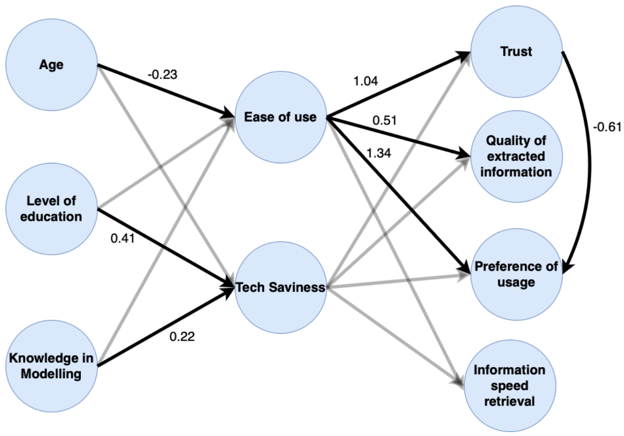

Figure 3. Structural equation model for DMN chatbot

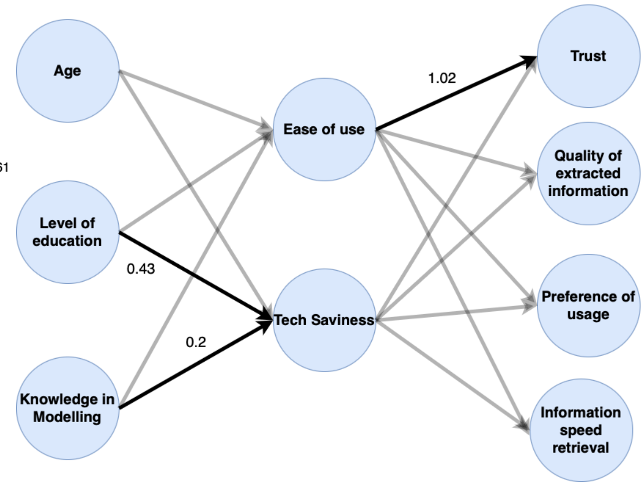

Figure 4. Structural equation model for the information page

For both SEM models, it can be seen that *Tech Saviness* has a positive correlation with *Level of education* and *Knowledge in Modelling*. Interestingly, *Ease of use* is only negatively correlated with *Age* in the chatbot SEM model. This can be explained by the fact that an information page can easily be analyzed by all ages but that apparently older respondents find the DMN chatbot more difficult to use. Another interesting observation is that both SEM models indicate that *Ease of use* has a positive correlation with *Trust*. However, for the chatbot SEM model it can be seen that *Ease of use* also positively influences *Quality of extracted information* and *Preference of usage*. A last interesting observation for the chatbot SEM model is that *Trust* has a negative correlation with *Preference of usage*. This might be that even though the respondents trust the answer of the DMN chatbot, it might still be too cumbersome to use. This latter observation requires further investigation as to why exactly this the case but maybe it is once again related to the fact that identifying which reasoning mechanism to trigger is too complicated for users.

The results also indicate that the respondents provided more correct answers with the use of the DMN chatbot in 4 out of 5 scenarios: scenario 1 (Complete Reasoning), scenario 2 (Reasoning with unknown information), scenario 3 (What if?) and scenario 7 (Sensitivity) showed an increase in correct answers of respectively 67% to 76%, 26% to 43%, 44% to 61% and 38% to 41%. The differences were statistically significant (.05). Only scenario 4 (What should?) was answered slightly more correctly by the respondents with the use of the information page (55% vs 52%). This can probably be attributed to the confusion of having to fill in different fields for scenario 4 in the survey. This is in line with the conclusions drawn from a previous smaller study (Goossens, Maes, et al., 2023) which also concluded that respondents retrieve better explanations using a DMN chatbot compared to textual descriptions.

5. Discussion

The first research question asked which are the possible advice generation tracks together with their advantages and their disadvantages. To answer this question, Section 4 introduces a Text2Chat framework identifying the three main advice generation tracks: Manual advice, GPT-3 advice and DMN advice generation track. Goossens and Vanthienen (2023) concluded that the advice quality of an GPT-3 integrated DMN chatbot is better than directly generating advice from GPT-3 and a textual description. Solely using GPT-3 does not always provide consistent and correct advice. As such, the advice generation track which combines GPT-3 and a DMN chatbot seems to currently be a good middle ground between user-friendliness and providing consistently correct advice. It is therefore a very promising advice generating track. In the future, the authors plan to develop and evaluate a DMN chatbot integrated with the most recent GPT-4 model to interact with users.

The second research question which asked how DMN models can best be generated from textual descriptions was answered in subsection 4.3. Overall it was concluded that GPT-3 is a good way to extract DRDs given that it is robust against a lot of formulations and has strong scalability capabilities thanks to few-shot learning. However despite various proposals to semi-automatically extract decision logic from decision descriptions, currently there is no way to automatically extract complete and correct decision tables. This will be addressed in the future by investigating the decision logic reasoning capabilities of GPT-4.

The survey executed in this paper deals with the third research question. From the survey executed in this paper as well as (Goossens, Maes, et al., 2023), it can be concluded that the explanations provided by a DMN chatbot are more correctly interpreted by users than if the explanations are provided with the use of textual descriptions. This insight is relevant for the automation of customer service and advice as it finds a good middle ground between cost containment and scalable, correct and relevant personalized advice and explanation.

6. Conclusion

Knowledge intensive processes operate in a context where decisions are executed. It is within this context that this study remarks that a knowledge intensive process does not end once the outcome has been communicated to the end-user. Stakeholders might still have a questions regarding explanation and advice. These explanations are often described in textual descriptions which need to be analyzed to provide advice and explanation. This study introduces the Text2Chat framework containing the various tracks to generate advice from textual descriptions namely: manual advice (helpdesks, FAQs), decision model generated advice and GPT-3 advice. Despite manual advice and GPT-3 advice currently being the most used advice generation tracks, the study concludes that decision model generated advice, more specifically GPT-3 enhanced DMN chatbot advice, seems to ensure the most consistently correct, interpretable and personalized advice. Hence, this track would benefit from more research in the future.

References

Arco, L., Nápoles, G., Vanhoenshoven, F., Lara, A. L., Casas, G., & Vanhoof, K. (2021). Natural language techniques supporting decision modelers. *Data Mining and Knowledge Discovery*, *35*(1), 290–320.

Bazhenova, E., Bülow, S., & Weske, M. (2016). Discovering decision models from event logs. *International Conference on Business Information Systems*, 237–251.

Bazhenova, E., & Weske, M. (2016). Deriving decision models from process models by enhanced decision mining. *International conference on business process management*, 444–457.

Bellan, P., van der Aa, H., Dragoni, M., Ghidini, C., & Ponzetto, S. P. (2023). Pet: An annotated dataset for process extraction from natural language text tasks. *Business Process Management Workshops: BPM 2022 International Workshops, Münster, Germany, September 11–16, 2022, Revised Selected Papers*, 315–321.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877–1901.

Carbonnelle, P., Vandevelde, S., Vennekens, J., & Denecker, M. (2023). Interactive configurator with fo(.) and idp-z3.

De Conick, Z., & Marchoul, J. (2023). Dmn chatbot acceptance survey study. *Master Thesis*.

De Smedt, J., Hasić, F., & Vanthienen, J. (2017). Towards a holistic discovery of decisions in process-aware information systems. *Business Process Management*.

Di Ciccio, C., Marrella, A., & Russo, A. (2012). Knowledge-intensive processes: An overview of contemporary approaches. *Kibp@ kr*, 33–47.

Estrada-Torres, B., del-Río-Ortega, A., & Resinas, M. (2021). Demabot: A tool to automatically generate decision-support chatbots. *2021 Best Dissertation Award, Doctoral Consortium, and Demonstration and Resources Track at BPM, BPM-D 2021*, 141–145.

Etikala, V., Goossens, A., Veldhoven, Z. V., & Vanthienen, J. (2021). Automatic generation of intelligent chatbots from dmn decision models. *International Joint Conference on Rules and Reasoning*, 142–157.

Etikala, V., Van Veldhoven, Z., & Vanthienen, J. (2020). Text2dec: Extracting decision dependencies from natural language text for automated dmn decision modelling. *International Conference on Business Process Management*, 367–379.

Figl, K., Mendling, J., Tokdemir, G., & Vanthienen, J. (2018). What we know and what we do not know about dmn. *Enterprise Modelling and Information Systems Architectures (EMISAJ)*, *13*, 2–1.

Friedrich, F., Mendling, J., & Puhlmann, F. (2011). Process model generation from natural language text. *International Conference on*

Advanced Information Systems Engineering, 482–496.

Goossens, A., De Smedt, J., & Vanthienen, J. (2023a). Comparing the quality of gpt-3 with bert for decision requirements modeling. *International Conference on Cooperative Information Systems (COOPIS), Accepted*.

Goossens, A., De Smedt, J., & Vanthienen, J. (2023b). Extracting decision model and notation models from text using deep learning techniques. *Expert Systems with Applications*, *211*, 118667.

Goossens, A., Maes, U., Timmermans, Y., & Vanthienen, J. (2023). Automated intelligent assistance with explainable decision models in knowledge-intensive processes. *Business Process Management Workshops: BPM 2022 International Workshops, Münster, Germany, September 11–16, 2022, Revised Selected Papers*, 25–36.

Goossens, A., Vandevelde, S., Vanthienen, J., & Vennekens, J. (2023). Gpt-3 for decision logic modeling. *Proceedings of the 17th International Rule Challenge@ RuleML+ RR 2022*, 1–15.

Goossens, A., & Vanthienen, J. (2023). Integrating gpt-technologies with decision models for explainability. *Communications in Computer and Information Science*.

Huysmans, J., Dejaeger, K., Mues, C., Vanthienen, J., & Baesens, B. (2011). An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models. *Decision Support Systems*, *51*(1), 141–154.

Jojic, A., Wang, Z., & Jojic, N. (2023). Gpt is becoming a turing machine: Here are some ways to program it. *arXiv preprint arXiv:2303.14310*.

Katz, D. M., Bommarito, M. J., Gao, S., & Arredondo, P. (2023). Gpt-4 passes the bar exam. *Available at SSRN 4389233*.

Kluza, K., & Honkisz, K. (2016). From SBVR to BPMN and DMN models. proposal of translation from rules to process and decision models. *International Conference on Artificial Intelligence and Soft Computing*, 453–462.

Lee, P., Bubeck, S., & Petro, J. (2023). Benefits, limits, and risks of gpt-4 as an ai chatbot for medicine. *New England Journal of Medicine*, *388*(13), 1233–1239.

Negash, S., Ryan, T., & Igbaria, M. (2003). Quality and effectiveness in web-based customer support systems. *Information & management*, *40*(8), 757–768.

Neuberger, J., Ackermann, L., & Jablonski, S. (2023). Beyond rule-based named entity recognition and relation extraction for process model generation from natural language text. *arXiv preprint arXiv:2305.03960*.

Niszczota, P., & Abbas, S. (2023). Gpt as a financial advisor. *Available at SSRN 4384861*.

OMG. (2015a). Omg: Decision model and notation 1.0 (2015) [Accessed: 2022-01-08]. https://www.omg.org/spec/DMN/1.0/

OMG. (2015b). Omg: Semantics of business vocabulary and rules (2008) [Accessed: 2023-03-13]. https://www.omg.org/spec/SBVR/1.0

Quishpi, L., Carmona, J., & Padró, L. (2021). Extracting decision models from textual descriptions of processes. *International Conference on Business Process Management*, 85–102.

van der Aa, H., Leopold, H., Batoulis, K., Weske, M., & Reijers, H. A. (2016). Integrated process and decision modeling for data-driven processes. *Business Process Management Workshops: BPM 2015, 13th International Workshops, Innsbruck, Austria, August 31–September 3, 2015, Revised Papers 13*, 405–417.

Van Velsen, L. S., Steehouder, M. F., & De Jong, M. D. (2007). Evaluation of user support: Factors that affect user satisfaction with helpdesks and helplines. *IEEE Transactions on Professional Communication*, *50*(3), 219–231.

Vandevelde, S., Etikala, V., Vanthienen, J., & Vennekens, J. (2021). Leveraging the power of IDP with the flexibility of DMN: A multifunctional API. *Rules and Reasoning - International Joint Conference, RuleML+RR*, 250–263.

Vanthienen, J. (2021). Decisions, advice and explanation: An overview and research agenda. In *A research agenda for knowledge management and analytics* (pp. 149–169). Edward Elgar Publishing.

Vanthienen, J., Mues, C., & Aerts, A. (1998). An illustration of verification and validation in the modelling phase of kbs development. *Data & Knowledge Engineering*, *27*(3), 337–352.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Chi, E., Le, Q., & Zhou, D. (2022). Chain of thought prompting elicits reasoning in large language models. *arXiv preprint arXiv:2201.11903*.