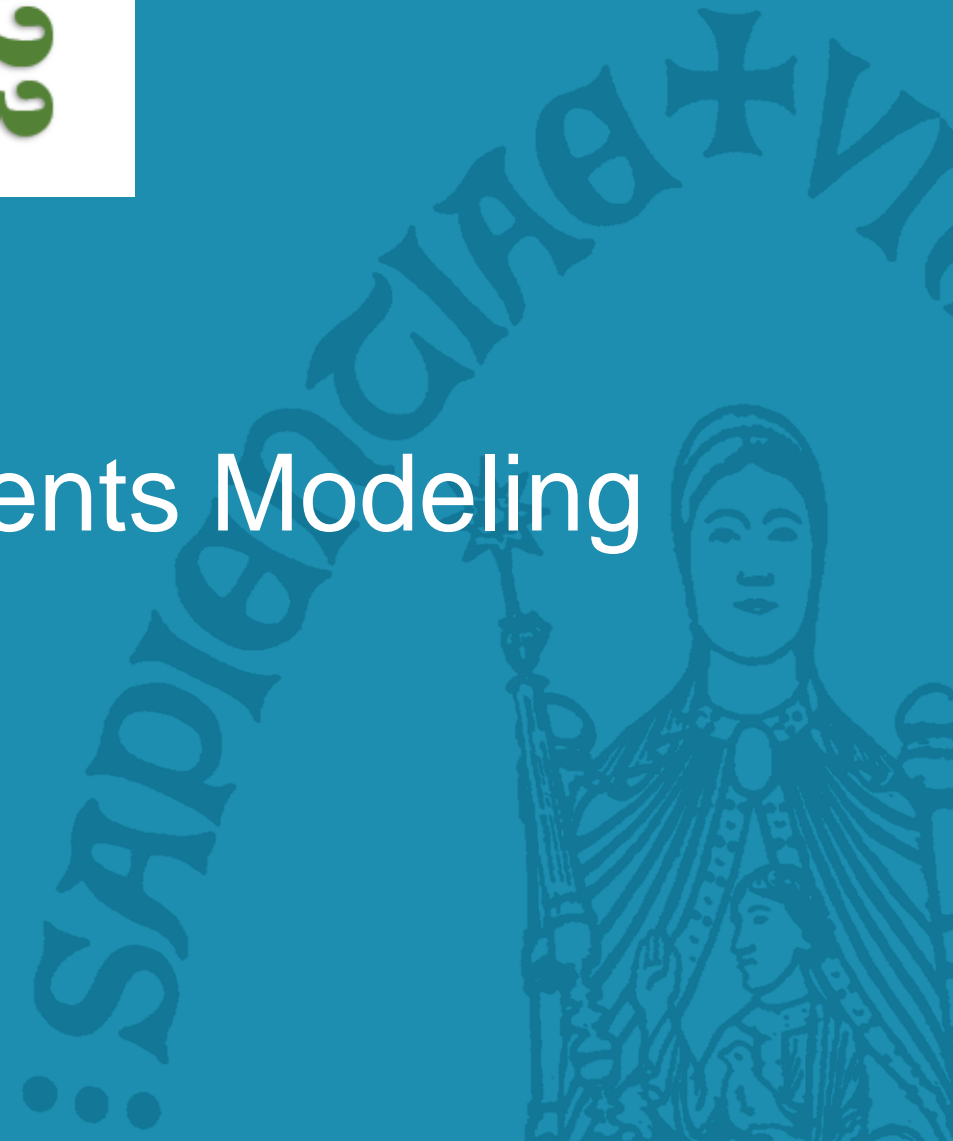




GPT-3 for Decision Requirements Modeling and Advice

Jan Vanthienen & Alexandre Goossens



Some **HOT** questions

Can LLMS (like GPT) replace decision modeling ->

1. Can GPT build Decision Models from text
 - a) Can GPT build Decision Requirements Diagrams (DRDs)?
 - b) Can GPT build Decision Logic Tables (DTs)?

2. Can GPT make decisions and provide explainability to decisions?

Can GPT be useful in decision model contexts?

1a) Deriving DRDs from text

Related Work

- **Pattern-based approach** to extract DRD and DMN models from unstructured text ((Etikala, V. et al., 2020), (Quishpi, L. et al., 2021))
- **Deep learning** pipeline to extract DRDs and logical statements from unstructured text (Goossens, A. et al., 2023)
- GPT-3 has already been used to extract from text:
 - **UML and BPMN** models (Fill, H.G. et al., 2023)

> Comparing GPT with deep learning for DRD extraction

Research questions

1. Can GPT-3 generate DRDs given a textual description?
2. What is the quality of these generated DRDs?
3. How does GPT-3 perform compared to the BERT-based approach (deep learning) that automatically extracts DRDs from texts

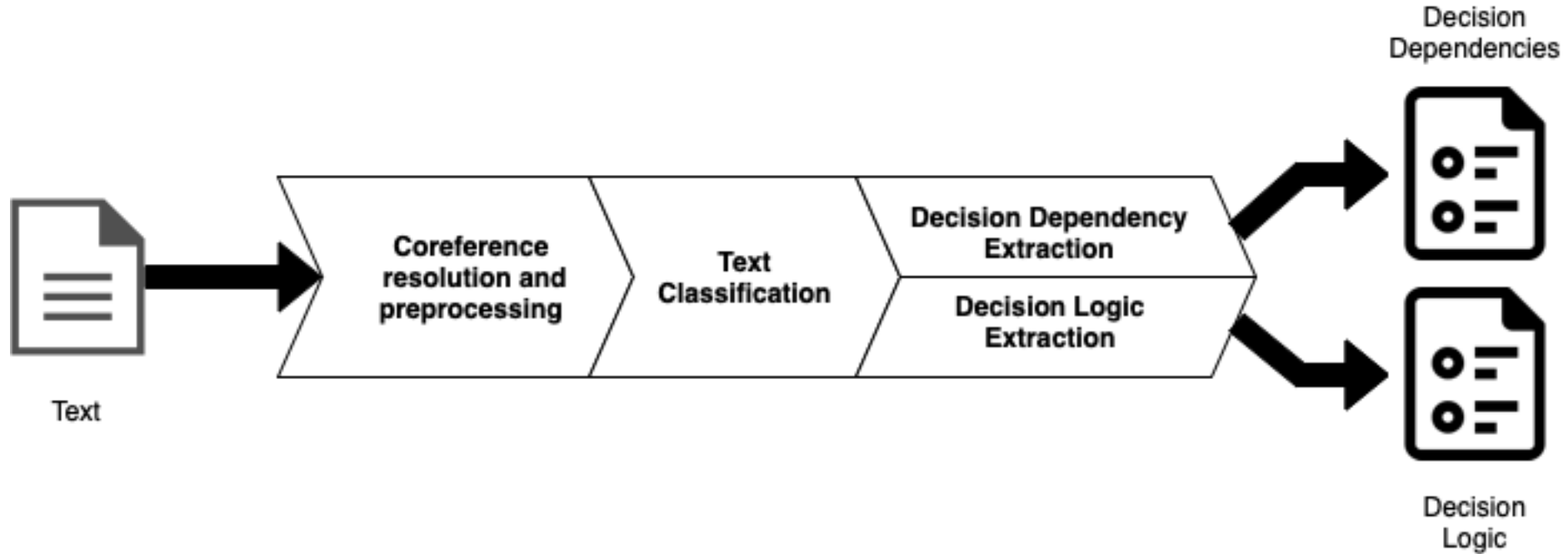
Goossens, A., De Smedt, J., Vanthienen, J.: Extracting decision model and notation models from text using deep learning techniques. Expert Systems with Applications 211, 118667 (2023)

Goossens, A., De Smedt, J., Vanthienen, J.: Comparing the performance of GPT-3 with BERT for decision requirements modeling. To be presented at CoopIS (2023)

Experiments on some real texts

- Six real textual descriptions collected from the Internet dealing with:
 - Housing loan eligibility, personal loan eligibility, student support eligibility, fraud rating score problem and ObamaCare support eligibility
 - No changes except replacement of bullet points with “and”.
- **Example:** *International students are generally self-supporting or supported by a scholarship. Our financial support is generally limited to Belgian nationals, although there are some exceptions. In order to be eligible for most types of support, you need to meet our general conditions. Our general conditions consist of nationality and academic requirements. We advise all international students to take a close look at the nationality requirements before applying. There may also be additional specific conditions per type of support.*
- Comparing GPT-3 and BERT for DRD extraction from text

BERT for DRD extraction from text



1. Coreference Resolution and Preprocessing
2. Text Classification
3. Decision Dependency Extraction and some Decision Logic Extraction

More details can be found in (Goossens, A. et al., Expert Systems with Applications, 2023)

GPT-3 for DRD extraction from text

Table 1. Hyperparameters of GPT-3

Parameter	Value
Model	text-davinci-003
Temperatures	[0, 0.2, 0.4, 0.6, 0.8, 1]
Max_tokens	3015

GPT-3 prompt:

1. *Do you know Decision Model and Notation and can you create a DRD?*
2. *Given a text description you are expected to provide a DRD in Graphviz code. Examples (4 examples):*
Input: sample text
Expected Graphviz Output: sample Graphviz output code
3. *Only provide the expected Decision Requirements Diagram in formatted graphviz code with inputs as ovals and decision as rectangles and arrows going bottom-up from inputs to decisions as shown in the examples. Do not write anything else. Analyze the following text: insert textual description.*

Results

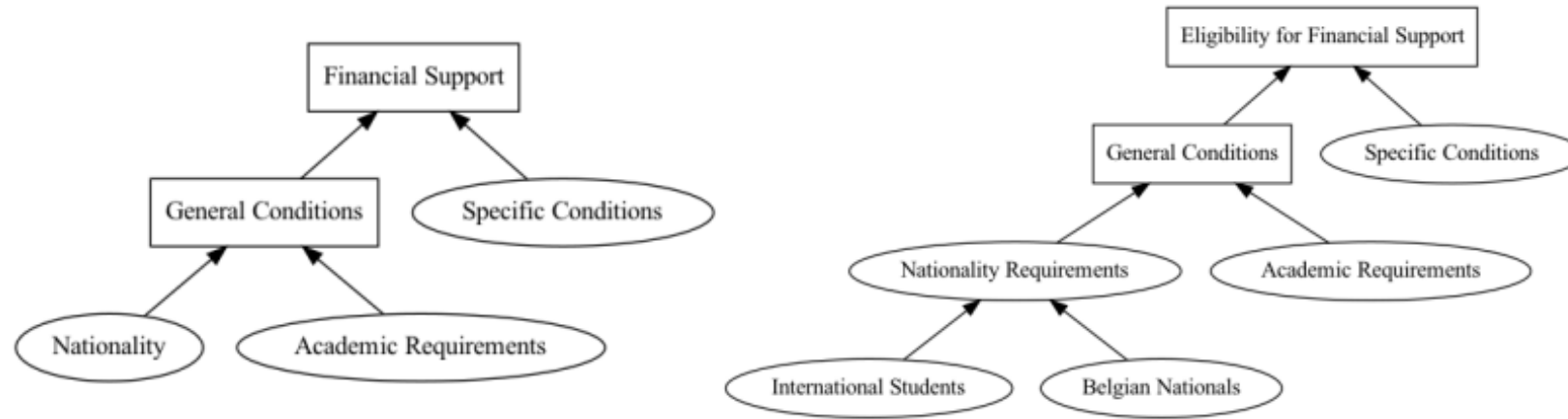


Fig. 2. Predicted DRD by GPT_0 for the student support VUB example

Fig. 3. Predicted DRD by GPT_0.2 for the student support VUB example

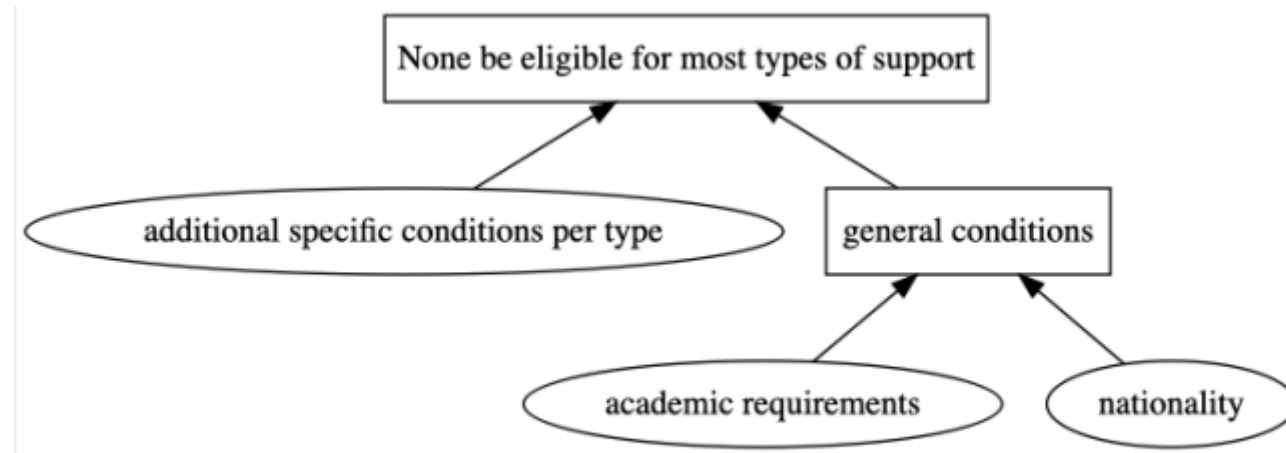


Fig. 5. Predicted DRD by BERT for the student support VUB example

Results

Table 3. Averages of GPT-3 and BERT models over all cases

Model	Decision and Input Data						Decision Dependencies					
	#gold	#pred	#ok	P	R	F1	#gold	#pred	#ok	P	R	F1
Total_GPT-3_0	38	40	33	87%	86%	86%	32	34	25	83%	74%	79%
Total_GPT-3_0.2		42	33	82%	86%	84%		36	25	78%	74%	76%
Total_GPT-3_0.4		40	33	87%	86%	86%		34	28	89%	87%	88%
Total_GPT-3_0.6		39	31	84%	80%	82%		33	23	80%	66%	72%
Total_GPT-3_0.8		39	31	85%	80%	82%		32	24	83%	70%	76%
Total_GPT-3_1		41	31	79%	80%	79%		34	24	74%	70%	72%
Average_BERT		40	37	93%	97%	95%		37	31	87%	97%	91%

#gold = identified by us, # pred = identified by the tool, # ok = number of correctly identified

P = #ok/#pred, Recall (R = #ok/#gold), and F1-score (F1 = 2P R/(P + R))

Answer to research questions 1a)

1. Yes, GPT-3 can generate DRDs given a textual description
2. The quality of generated DRDs is dependent on the temperature values.
 - For lower temperatures, GPT-3 actually often gives meaningful DRDs but sometimes struggles with the identification of intermediary decisions and decision dependencies. As such, using GPT-3 with low temperatures in a semi-automatic way to extract DRDs from text seems to be promising.
 - However, using higher temperatures (above 0.4) sporadically resulted in the creation of semantically wrong DRDs.
3. In this study, all GPT-3 models are outperformed in all but one metric by the BERT-based approach for automatic DRD extraction from texts.

The precision of the predicted decision dependencies of GPT-3 with a temperature of 0.4 is slightly higher than the one of BERT (89% vs 87%)

1b) Deriving decision logic from text

Related Work

- Pattern-based approach to extract decision tables from structured text (SBVR) (Kluza, K. et al., 2016)
- Extracting rules from legal documents (Dragoni M., et al., 2016)
- Pattern-based approach to extract DRD and DMN models from unstructured text ((Etikala, V. et al., 2020), (Quishpi, L. et al., 2021))
- Pattern-based approach to extract decision logic from single unstructured sentences (Arco, L. et al., 2021)
- Deep learning pipeline to extract DRDs and logical statements from unstructured text (Goossens, A. et al., 2023)

> Can GPT derive decision tables from text?

Research questions

1. Can GPT-3 identify information items of a decision with their values?
2. Can GPT-3 construct correct, complete and overlap-free decision tables?
3. Is GPT-3 able to correctly reason with decision tables?

Presented tomorrow at RuleML+RR 2023, Rule Challenge:

GPT-3 for Decision Logic Modeling. Alexandre Goossens, Simon Vandeveldel, Jan Vanthienen and Joost Vennekens.

GPT-3 Prompt

- Q1 *Below is a decision description. What does this description decide?* «**Problem Description**»
- Q2 «**Test1**» (answer question using the description)
- Q3 *What are the variables that influence this decision?*
- Q4 *For each input and output, give me an overview of their data type and their possible values.*
- Q5 *What are the relevant values of the numerical variables?*
- Q6 *Could you generate a DMN decision table for this description? Make sure the table can be read horizontally: the column headers contain the inputs and output.*
- Q7 *Make the rules mutually exclusive.*
- Q8 *Is this table complete? (I.e., is there an applicable rule for each set of inputs?) If it is incomplete, can you find an example for which no rule would be applicable?*
- Q9 *According to your table, answer the following question.* «**Test2**» (answer question using the table)

Results for 1b) decision logic

➤ 72 experiments: 6 examples x 4 temperatures x 3(CV)

1. Information items and values

- **Accuracy** 0.50 ± 0.41
- **Inferring implicit values:** GPT-3 was not able to consistently deduce that if there are excellent and good grades there should also be bad grades. Only higher temperatures of 0.7 and 1 sometimes captured this.
- **Own definitions of concepts:** It is difficult to know what GPT-3 knows and does not know e.g., the BMI example is purposefully incorrect in its definition of BMI however GPT-3 still hangs on to its own (correct) definition of how to determine a BMI-value.

2. Decision tables

- **Accuracy** 0.06 ± 0.11
- **Completeness:** It seems as if GPT-3 does not really understand the notion of completeness since it considered 70 out of 72 decision tables to be complete whilst only 42% were actually complete.
- **Misjudging Relevance:** GPT-3 sometimes used a value of a question as an important value to build a decision table.

3. Reasoning

- **Hallucination:** In case of a missing rule for certain inputs, GPT-3 will invent an output
- **Math:** Wrong calculations

2. Can GPT provide explainability to decisions?

- When decision are made automatically, service is performed online, customers still have some questions: Why, Why not, How, What to do?
Do current solutions for advice and explanation (call centers, FAQs, ChatGPT, online help, ...) provide an effective and efficient solution to customer demands?
- 1. Organizations might be tempted to provide a problem description and use it as an automatic helpdesk to answer questions of stakeholders. Can GPT provide very specific answers to customer questions that might have to be derived from the provided information?
- 2. Can GPT be combined with a DMN chatbot (DecisionCamp 2021, 2022)?

Integrating GPT-technologies with Decision Models for Explainability.
Alexandre Goossens, Jan Vanthienen. The 1st World Conference on
eXplainable Artificial Intelligence (xAI 2023), Lisbon 2023.

Vaccination example

- *The period in which a person will get vaccinated depends on the number of available doses and the person's assigned group. The person's assigned group is divided into three groups depending on their vulnerability, exposure, medical risks and age. If you are a resident or employee in a residential care centre or if you work in a first line care occupation then you belong to the most vulnerable and exposed people and you will be vaccinated first. Moving on to the second group, if you are older than 65, or if you are between 45 and 65 with an increased medical risk due to healthcare issues, you will get vaccinated next. Lastly, the third group consists of the broader population of people above the age of 18. If you are assigned to group 1, you get the highest vaccination priority at any time. If the available doses is low, then groups two and three are considered to be low priority. But if the available doses is medium then group 2 and group 3 are considered medium and low priority respectively. Lastly, if the available doses is high then group 2 is considered high priority and group 3 is considered medium priority*
- Possible Questions:
 - What age should I be to be assigned to group 1?
 - **Under what conditions** will I be vaccinated first?
 - I don't understand **why** I am considered high priority, can you explain that to me?
 - **What if** some input data are unknown or different?



Research Questions

1. When provided with a problem description, context, and question, can GPT-3 provide the correct answer to the question (without a decision model)?
2. If not
Given that a decision model and a reasoning engine are available, to what extent is GPT-3 capable of identifying the appropriate reasoning scenario required to answer a given question or provide an explanation?
 - Can GPT-3 label questions into:
 - How to?
 - Why?
 - What if?
 - What should?

1. GPT-3 for Question Answering

- **Complete Prompt:** *I provide you with the following text description: insert textual description. Assume the following information to answer the following questions: insert situation. Provide for each answer the concerning question: insert all questions.*

Table 1. Context of Vaccination and Discount Problem

Problem	Situation
Vaccination	I am 65 years old, my exposure is high, I am not at risk nor vulnerable and the available doses is low.
Discount	I have been loyal for three years, have bought goods over time with a total value of 8 and currently placed an order of volume 80 not during the holidays.

Question types

Class 1: How to get Y?

Example: "How can I make sure the result is: drink is Guinness?"

Class 2: Why?

Example: "Why did not I not get my loan approved?"

Class 3: Sensitivity

Example: "How much can my daughter still grow to be allowed in the ride"

Class 4: What if?

Example: "What would happen if I wear a green t-shirt instead of red?"

Class 5: What should?

Example: "What colors should I decorate my room with to get a warmer tone?"

Prompt questions and respective classes


Table 2. Prompt questions and respective classes

Example	Question	Class
DISCOUNT	How much value can I still buy without changing client tier?	Class 3: Sensitivity
	How long will I still be the same client tier if I don't buy any more goods?	Class 3: Sensitivity
	What if I order during a holiday season?	Class 4: What if?
	What if the total value bought now exceeded 10?	Class 4: What if?
	Why am I considered a silver customer?	Class 2: Why?
	How come I only got a 4% discount?	Class 2: Why?
	How can I get the discount of 10%?	Class 1: How to get Y?
	How can I become a gold tier customer?	Class 1: How to get Y?
	How many orders should I place to get a golden client tier?	Class 5: What should?
	What holiday season should it be to get a discount of 7%?	Class 5: What should?
Vaccination	How can I get assigned to group 2?	Class 1: How to get Y?
	Under what conditions will I be vaccinated first?	Class 1: How to get Y?
	What if next week it is my birthday and I turn 66 years old?	Class 4: What if?
	What would happen if next week there is a high amount of doses available instead of low?	Class 4: What if?
	Why was I assigned to group 1?	Class 2: Why?
	I don't understand why I am considered high priority, can you explain that to me?	Class 2: Why?
	How much can the available doses change without changing the priority level?	Class 3: Sensitivity
	What is the maximum age I can be without changing group?	Class 3: Sensitivity
	How many doses should there be to get a medium vaccination priority being assigned to group 3?	Class 5: What should?
	What age should I be to be assigned to group 1?	Class 5: What should?

Results GPT

Table 4. Results Question Answering

 = Correct

 = Incomplete or incorrect, or containing too much redundant info

EXAMPLE	Question	Golden answer	Answer GPT-3 logic
DISCOUNT	1	Value can still increase with 2.	2
	2	2 years	5 years.
	3	5% discount	4% discount.
	4	You would be gold tier and receive a discount of 7%.	You would become a silver customer.
	5	Loyalty >5 and value <2 OR loyalty = [2..5] and value is [2..10] OR loyalty <2 and value >10	Because you have been loyal for three years and have bought goods over time with a total value of 8.
	6	Because holiday season = False , Customer tier= silver and volume <100.	Because you ordered 80 units not during the holidays.
	7	Holiday season= True, customer tier= gold and volume >100	Order 100 or more units during a holiday season.
	8	Loyalty is [2..5] and value >10 OR loyalty >5 and value >2	Buy goods over time with a total value of 10 or more and be loyal for 2 to 5 years, or be loyal for more than 5 years and buy more than 2.
	9	You should place orders with value >3	You need to buy goods over time with a total value of 10 or more and be loyal for 2 to 5 years, or be loyal for more than 5 years and buy more than 2.
	10	Holiday season = True + volume >= 100	Silver customers can get a 7% discount when they order 100 or more units during a holiday season.
Vaccination	1	If vulnerability= Low, Exposure= Low, medical risk = High, age >= 45 OR if age >65 and everything else Low	You can get assigned to group 2 if you are older than 65, or if you are between 45 and 65 with an increased medical risk due to healthcare issues.
	2	If assigned group= group 1: vaccination priority = High OR if available doses= high and assigned group = group 2: vaccination priority = High	You will be vaccinated first if You are a resident or employee in a residential care centre or if You work in a first line care occupation.
	3	Nothing changes	If you turn 66 years old, you will be assigned to group 2 and will be vaccinated next.
	4	Still high priority	If there is a high amount of doses available, group 2 will be considered high priority and group 3 will be considered medium priority.
	5	Exposure = High, Risk= Low, vulnerability = Low and age = any	You were assigned to group 1 because you are a resident or employee in a residential care centre or if you work in a first line care occupation, making you the most vulnerable and exposed.
	6	Return rule: Available doses = low + group 1	You are considered high priority because you are a resident or employee in a residential care centre or if you work in a first line care occupation, making you the most vulnerable and exposed.
	7	No change will affect your sensitivity level	The available doses can change depending on the number of available doses and the person's assigned group. If the available doses is low, then groups two and three are considered to be low priority. But if the available doses is medium then group 2 and group 3 are considered medium and low priority respectively. Lastly, if the available doses is high than group 2 is considered high priority and group 3 is considered medium priority.
	8	Any age	The maximum age you can be without changing group is 65 years old.
	9	High	If the available doses is medium then group 2 and group 3 are considered medium and low priority respectively.
	10	Any age is fine, it is dependent on vulnerability, exposure and risk	To be assigned to group 1, you must be a resident or employee in a residential care centre or if you work in a first line care occupation.

1. Can GPT correctly answer the questions?

- Organizations might be tempted to provide a problem description and use it as an automatic helpdesk to answer questions from stakeholders.
- **However:**
 - GPT-3 is good at repeating information and providing relevant pieces of information, but not very good at providing very specific answers derived from the provided information, leaving the interpretation to the stakeholder
 - Example:
 - Question: *What age should I be to be assigned to group 1?*
 - GPT Answer: *To be assigned to group 1, you must be a resident or employee in a residential care centre or if you work in a first line care occupation.*
 - Correct answer: *Any age is fine. It is dependent on vulnerability, exposure and risk.*

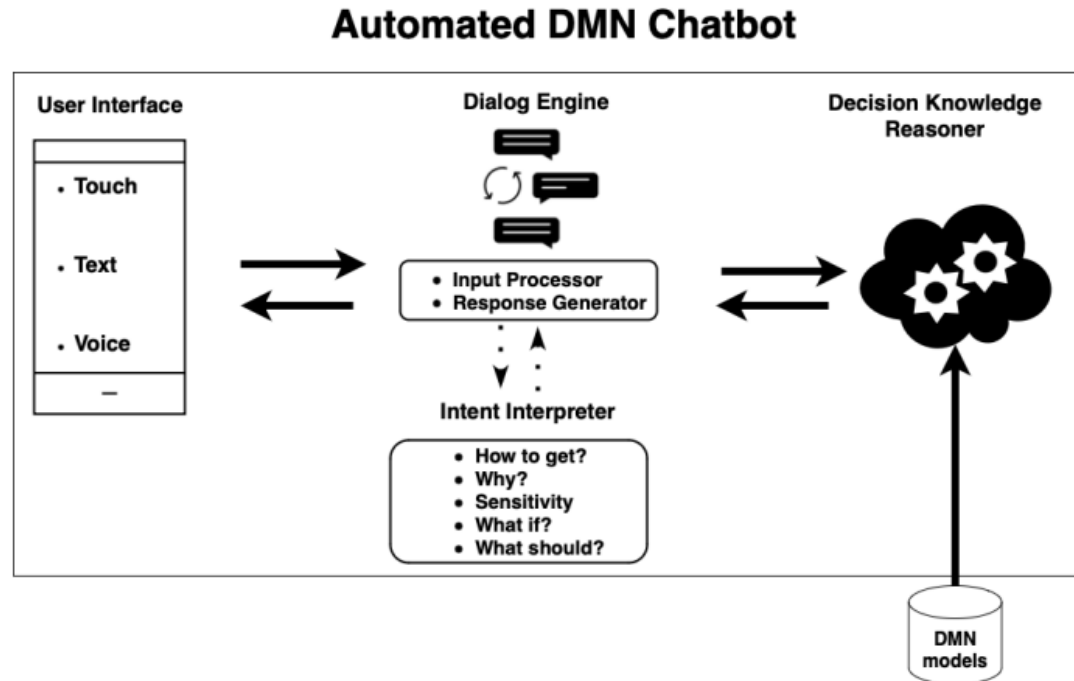
2. Can GPT identify the reasoning scenario?

If GPT cannot answer the question, maybe it can recognize the type?

Given that a decision model and a reasoning engine are available, to what extent is GPT-3 capable of identifying the appropriate reasoning scenario required to answer a given question?

- Can GPT-3 label questions into:
 - How to?
 - Why?
 - What if?
 - What should?

Reasoning and explanation with DMN chatbot



KU LEUVEN

Communicating with Decision Models

DECISION CAMP 2021
13th September

Vedavyas Etikala, Ziboud Van Veldhoven, Alexandre Gooseens, Prof. Jan Vanthienen.
LIRIS, Research Center for Management Informatics, KU Leuven, Belgium

The image shows a promotional graphic for a workshop. It features a blue background with the KU Leuven logo and the title 'Communicating with Decision Models'. Below the title is the 'DECISION CAMP 2021' logo and the date '13th September'. To the right, there is an illustration of a chatbot interface with a robot and a human user, with speech bubbles and a question mark. At the bottom right, the names of the organizers and their affiliation are listed.

- Survey indicated:
 - DMN chatbot provides more correct explanations compared to textual descriptions
 - But is not user-friendly enough (in understanding the question) according to the participants

Vaccination example

Person's Assigned Group						Vaccination Priority			
U	Vulnerability	Exposure	Medical Risk	Age	Person's Assigned Group	U	Available Doses	Person's Assigned Group	Vaccination Priority
1	High	-	-	-	Group 1	1	-	Group 1	High
2	Low	High	-	-	Group 1	2	Low	Group 2, Group 3	Low
3	Low	Low	High	≥ 45	Group 2	3	Medium	Group 2	Medium
4	Low	Low	Low	> 65	Group 2	4	Medium	Group 3	Low
5	Low	Low	Low	$[18..65[$	Group 3	5	High	Group 2	High
6	Low	Low	High	$[18..45[$	Group 3	6	High	Group 3	Medium

Fig. 1. Decision Tables for the vaccination problem

Various reasonings:

- Execution reasonings:
 - Class 0.1: Complete Reasoning
 - Class 0.2: Partial decision-making
 - Class 0.3: Reasoning with incomplete information
- Explanation Reasonings:
 - Class 1: How to get?
 - Class 2: Why?
 - Class 3: Sensitivity
 - Class 4: What if?
 - Class 5: What should?

Discount example

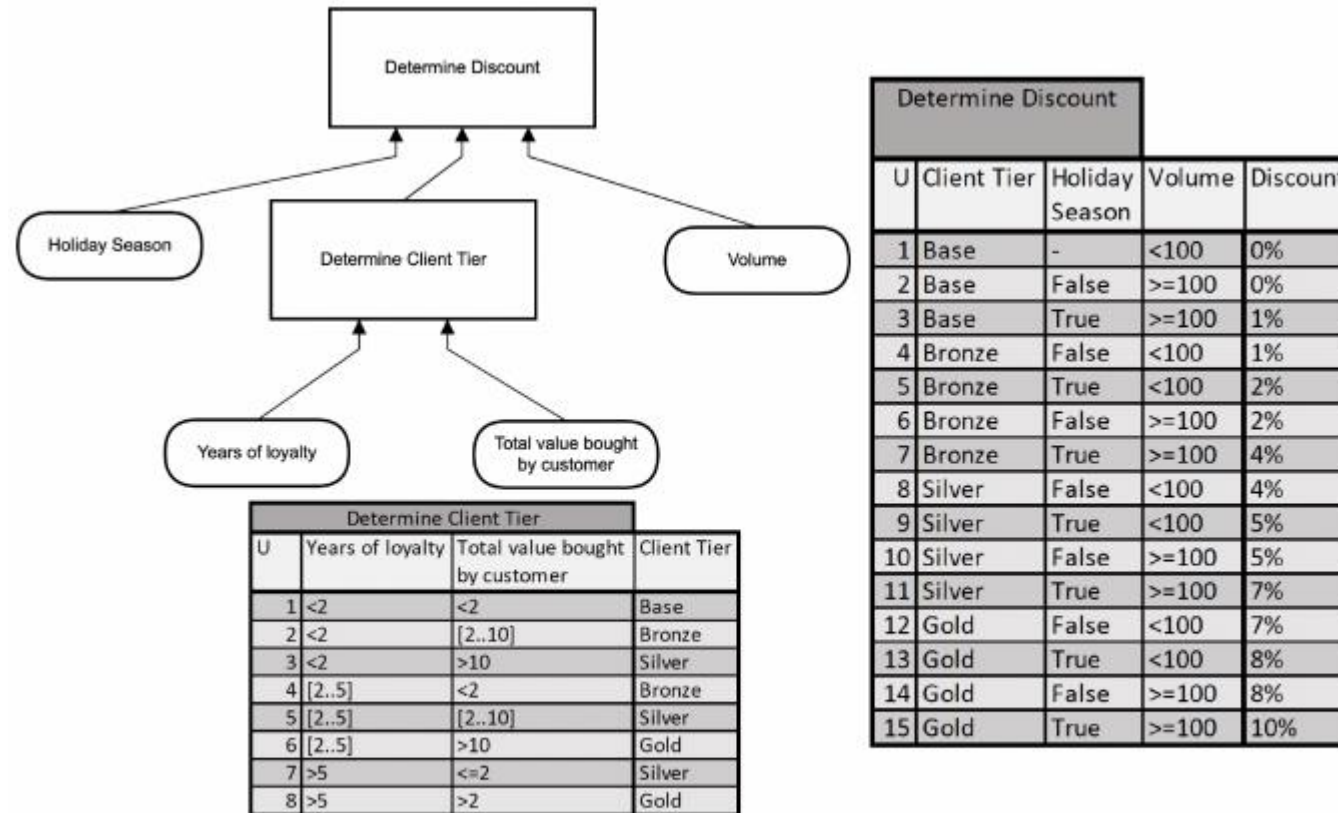


Fig. 4. Discount Example

Methodology: GPT-3 for scenario identification

Prompt:

There are 5 classes in total. For each class I give you three examples:


*I will provide you with sentences and return to me for each sentence: the sentence and the expected output; Class Y: Expected output. **Insert Questions***

Temperature 0: because we want consistent answers and not too much freedom in its interpretation of the text

Result

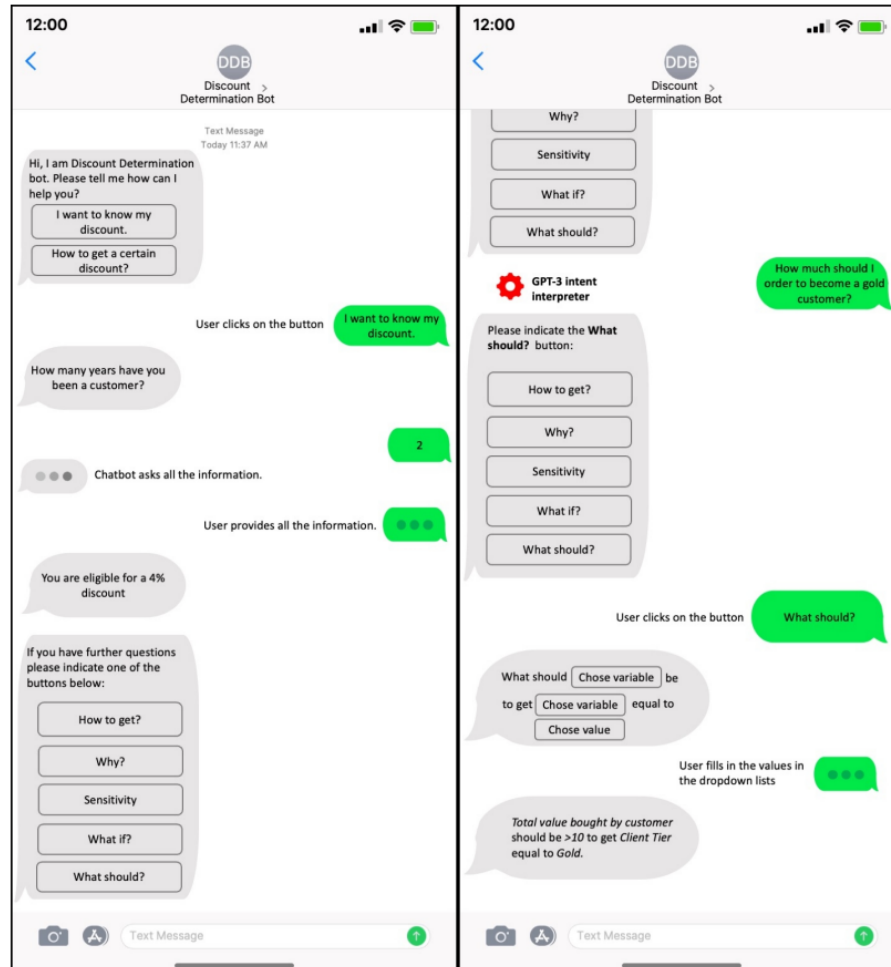
Table 5. Results Scenario Identification

 = Correct

 = Incomplete or incorrect

EXAMPLE	Question	Golden Class	Answer GPT-3 to classify
DISCOUNT	1	Class 3: Sensitivity	Class 3: Show all paths
	2	Class 3: Sensitivity	Class 3: Show all paths
	3	Class 4: What if?	Class 4: What if variable=value?
	4	Class 4: What if?	Class 4: What if variable=value?
	5	Class 2: Why?	Class 2: Explain
	6	Class 2: Why?	Class 2: Explain
	7	Class 1: How to get Y?	Class 1: How to get variable=value?
	8	Class 1: How to get Y?	Class 1: How to get variable=value?
	9	Class 5: What should?	Class 5: What should?
	10	Class 5: What should?	Class 5: What should?
Vaccination	1	Class 1: How to get Y?	”Class 1: How to get variable=value?”
	2	Class 1: How to get Y?	Class 1: How to get variable=value?
	3	Class 4: What if?	Class 4: What if variable=value?
	4	Class 4: What if?	Class 4: What if variable=value?
	5	Class 2: Why?	Class 2: Explain
	6	Class 2: Why?	Class 2: Explain
	7	Class 3: Sensitivity	Class 3: Show all paths
	8	Class 3: Sensitivity	Class 3: Show all paths
	9	Class 5: What should?	Class 5: What should?
	10	Class 5: What should?	Class 5: What should?

Explainable DMN + GPT-3 Chatbot Concept Illustration



- GPT-3 can be used to automatically or semi-automatically identify which reasoning scenario should be triggered to answer a question.
- The reasoning and explanation can be left to the DMN chatbot

Conclusion

- Yes, GPT-3 can generate DRDs given a textual description, but there are better alternatives.
- Deriving decision tables from text with GPT performs poorly.
- Combining GPT-3 with a DMN chatbot provides better explanations rather than directly asking GPT-3 for explanations.
- GPT-3 is able to correctly identify 5 scenarios and as such can be used to improve the user-interface and user-friendliness of the current DMN chatbot proposal.

Thank you for your attention