Using Large Language Models to Generate, Validate, and Apply User Intent Taxonomies

Supplementary Material

## A. Prompt for chat user intent taxonomy generation

### Context and data description

* Your primary goal is to generate an intent taxonomy from \*\*the given data\*\* and \*\*the given existing taxonomy (if available)\*\*. You can use the taxonomy to organize and understand your data.
* You will be given information about a list of human-AI conversations. For each conversation, you'll be given a short summary about what the user task is performed in this conversation.
* You may also be given an existing taxonomy in the \*\*table\*\* format, where row in this taxonomy is a intent category. You can use this taxonomy to help you construct a new intent taxonomy. The schema of this intent taxonomy is as follows:
* \*\*title\*\*: the title of the intent category
* \*\*description\*\*: the description of the intent category
* \*\*examples\*\*: a list of examples in the intent category

Your primary goal is to generate a taxonomy that can serve for the following use cases: The primary use case of this taxonomy is to help understand what users are doing in human-AI conversations. Entities in this taxonomy can be used to label \*\*user intents\*\* in human-AI conversations.

### Criteria of a generic taxonomy

* Accuracy: The definitions, descriptions of classes, properties, and individuals in a taxonomy should be correct.
* Completeness: All the data should be reliably classified using this taxonomy.
* Conciseness: The taxonomy should not include any irrelevant elements with regards to the user intents in AI Chat.
* Clarity: The taxonomy should communicate the intended meaning of the defined terms. Definitions should be objective and independent of the context.
* Consistency: The taxonomy does not include or allow for any contradictions.

### Requirements of your output taxonomy

* Your output \*\*intent\*\* taxonomy should focus on the user actions in a task, not the task objects. This is \*\*different\*\* from a \*\*domain\*\* taxonomy, which primarily describes the task objects.
* Your output taxonomy should match the existing taxonomy and the data as closely as possible, without leaving out important intent categories or including unnecessary ones. Please make sure there is no overlap or contradiction among the intent categories in your output taxonomy.
* Your output \*title\* of each category should be \*no more than 5 words\*. The title should be a concise and clear label for the intent category. It can be either verb phrases or noun phrases, whichever is more appropriate.
* Your output \*description\* of each category should be \*no more than 30 words\*. The description should explain the user's goal or purpose for the intent category, and should differentiate it from other intent categories.
* The number of examples for each intent category should be \*no more than 3\*. The examples should either come from the given taxonomy or the provided data with \*\*exactly the same content\*\*. Please do not invent new examples or intents that are not in the given taxonomy or the data.
* \*\*Size limit of the output taxonomy\*\*: The total number of intent categories should be \*\*no more than 10\*\*.
* Your output taxonomy and examples should be in \*English\* only.

### You are asked to answer the following questions

* Q1. Please check the above general criteria and the specific taxonomy requirements one-by-one. Does the provided taxonomy satisfy the above requirements, word limits and taxonomy size limit? Please answer "yes" or "no". If there is no given taxonomy, please answer "no".
* Q2. Please explain your answer to Q1. If your answer to Q1 is "no", please also describe if you'd like to construct the taxonomy structure from scratch or you plan to make changes on the given taxonomy. Your answer to this question should be \*\*within 100 words\*\*.
* Q3. If your answer to Q1 is "no", then generate a new intent taxonomy from the the given data and the given existing taxonomy (if available). Your output taxonomy should be in the \*\*table\*\* format with the same schema. If your answer to Q1 is "yes", please answer "N/A". Please make sure the new taxonomy satisfies \*\*all of the above requirements\*\*. Please \*\*do not\*\* invent new examples or new intents that are not in the existing taxonomy or the provided data.

Tips

* If you're given an existing taxonomy, you can use the provided data to update this taxonomy. By incorporating the newly provided data, you can \*add new categories\*, \*merge or generalize existing categories\*, \*split existing categories\*, \*reorganize the current tree structure\*, \*change titles and descriptions\*, \*swap examples\* and do other operations if needed.
* If the intent category structure of the given taxonomy cannot be easily adjusted, then please construct a new structure of these intent categories based on their descriptions and the provided examples. Please make sure your new taxonomy covers the semantics of the existing taxonomy as thoroughly as possible. Please \*\*do not\*\* invent new intents that are not in the existing taxonomy or the provided data.
* You should carefully review the examples provided in each category and make sure they are correctly labeled. You can also reorganize the examples or create new categories from them when needed. You're allowed to have fewer than 3 for each category but your examples should only come from examples in \*\*the given taxonomy or the provided data\*\*. Please \*\*do not\*\* invent new examples that are not in the existing taxonomy or the provided data.
* Please make sure your new taxonomy satisfies the \*\*word limits\*\* and \*\*taxonomy size limit\*\*. You're allowed to have fewer than 10 categories in your final output. If you couldn't fit your new taxonomy into the limits, please consider merging or abstracting some specific categories into more general categories.
* Please make sure there is no overlaps or contradictions among the intent categories in your output taxonomy.

## B. Prompt for taxonomy application

You will be given a conversation history between a User and an AI agent. Your task is to answer questions about the user's intent.

### User Intent

A user intent is defined as the user's purpose for conversing with the AI agent. The categories of user intents are:

* **Information Retrieval**: Conversations where the user wants to find factual information or answers to specific questions. The agent's responses are typically direct, concise, and informative, providing the relevant information and/or links to the sources. This intent calls for retrieving or reconstructing factual information that already exists, rather than synthesizing or computing something new.
* **Problem Solving**: Conversations where the user wants to perform a mathematical or logical operation, such as a conversion, a percentage, a formula, or a function. The agent's responses are typically factual and computed or constructed based on available information and what the user provided. Unlike Information Retrieval intent, this intent calls for the agent to do some processing on top of simply retrieving or extracting information.
* **Learning**: Conversations where the user wants to understand a concept or acquire skills by getting detailed explanation, reasoning, or synthesis. The agent's responses are typically a synthesis of information based on several factual pieces of information, often from different sources. The Learning intent requests often involve questions like 'how', 'why', or requests like 'explain' -- things that will indicate asking for explanations or doing investigation. Also, while individual turns may be of information retrieval nature, if the user is asking multiple questions that drill into a topic, that's an indication of Learning intent.
* **Content Creation**: Conversations where the user asks the agent to either generate original content or translate existing content into new content based on specified criteria or constraints. In the case of generating original content, the user's questions require some degree of creativity, novelty, or innovation from the agent. The agent's responses contain original or translated outputs that match the user's specifications.
* **Leisure**: Conversations where the user wants to chat or play with the agent out of curiosity, boredom, or humor, or else explore broad ideas or areas of interest without a specific goal or information need in mind. There may not even be a specific question or a request. The agent's responses are typically suggestive and engaging. The agent may also encourage further inquiry or action from the user to deepen their discovery experience.
* **Other**: This intent label can be used if none of the above labels fit. Note that you should do your best to find an appropriate label from the list above and only in the rare circumstances when you have very little to no confidence in that ability, you can use 'Other' label.

Examples

**Information Retrieval**

"User:where are the best places to watch the sun set in spring in Madison Wisconsin

AI:There are many beautiful places to watch the sun set in spring in Madison Wisconsin. Here are some of the top-rated ones according to web sources: - The Edgewater Hotel: This historic hotel offers a stunning view of Lake Mendota and the Capitol. You can also enjoy movie nights, Brewgrass Fridays, and other evening events on the rooftop terrace. - Memorial Union Terrace: This classic spot overlooks Lake Mendota and is perfect for ice cream or a pitcher. You can also catch live music and entertainment on the stage. I hope this helps you find your ideal spot for a sunset experience."

**Problem Solving**

"User: The cost for the excel training will £15,000 per training. 60% will be at the start of the training and 40% in the following month. so if the months for year were july to december interprete that information

AI:Based on the information you provided, the cost for the Excel training will be £15,000 per training. 60% of this amount, which is £9,000, will be due at the start of the training. The remaining 40%, which is £6,000, will be due in the following month. If the training takes place between July and December of a given year, then these payments would be made during those months."

**Learning**

"User:why do thc gummies not work for me

AI:There are several reasons why THC gummies might not work for some people. One reason could be due to their \*\*THC tolerance\*\* and physical factors like \*\*body size\*\*. Another reason could be due to \*\*gastrointestinal disorders\*\* that affect how the small intestine absorbs nutrients, water, and medicine. Additionally, some people may have a \*\*slow metabolism\*\* due to their genetics. Would you like more information on any of these reasons?"

**Content Creation**

"User:create a resume tagline for this profile: Graduate in Business and Information Systems Engineering with a strong interest in Data Science, Machine Learning, and AI.

AI:How about this tagline: "Data-driven Business Engineer with a passion for Machine Learning and AI"?"

**Leisure**

"User:i challenge you to a pokemon battle

AI:I'm sorry, but I'm just a search engine and I don't have the ability to participate in Pokémon battles. However, you can try one of the Pokémon battle simulators I mentioned earlier to find other trainers to battle with."

Tips

* The following intentions indicate seeking information: Information Retrieval, Problem Solving, and Learning. Here is how we can differentiate among them. Information Retrieval is often going to be the default as it involves retrieving what typically already exists. Problem Solving will involve doing something more on top of retrieval information. So, knowing the tax brackets is Information Retrieval intent, but having the agent calculate one's taxes is a Problem Solving intent. Similarly, asking what function in Python does a certain operation has Information Retrieval intent, but asking the agent to write a function in Python has Problem Solving intent. Learning typically calls for an explanation. If the user asks the agent to describe how state income tax and estate tax differ, that has Learning intent. Similarly, if the user asks the agent to explain how a Python function works, that is Learning. Also, if a conversation involves multiple (3 or more) turns that shows the user is drilling more into a given topic, that is an indication of Learning.
* The following intentions are beyond information retrieval: Transaction, Creation, and Leisure. Transaction is where the user is looking to buy, plan, or execute something and needs help or recommendations. Creation is where the user wants the agent to generate something new instead of retrieving it. Leisure is where the user is using the agent as a substitute for a human to have natural language chitchat, explore a topic for fun. The user may not even provide a question, and simply have a statement to initiate an interesting conversation.

## C. Prompt for comparing search and chat user intent taxonomy generation

### Context and data description

- Your primary goal is to generate an intent taxonomy from \*\*the given data\*\*. You can use the taxonomy to organize and understand your data.

- You will be given data that contain user requests to a search system.

- Ignore the requests that are not in English.

- Ignore data with multiple intents.

- The schema of this intent taxonomy is as follows:

- \*\*title\*\*: the title of the intent category

- \*\*description\*\*: the description of the intent category

- \*\*examples\*\*: a list of examples in the intent category

Here is an example of a taxonomy:

```markdown

|title|description|examples|

|-|-|-|

|title of the category|description of the category|example 1; example 2; example 3|

```

### Your primary goal is to generate a taxonomy that can serve for the following use cases

- The primary use case of this taxonomy is to help understand what users are doing when they submit a request to a search system. Entities in this taxonomy can be used to label \*\*user intents\*\* in human-system interactions.

### Here are the criteria of a good taxonomy

- \*\*Accuracy\*\*: The definitions, descriptions of classes, properties, and individuals in a taxonomy should be correct.

- \*\*Comprehensiveness\*\*: All the data should be reliably classified using this taxonomy.

- \*\*Conciseness\*\*: The taxonomy should not include any irrelevant elements with regards to the user intents in AI Chat.

- \*\*Clarity\*\*: The taxonomy should communicate the intended meaning of the defined terms. Definitions should be objective and independent of the context.

- \*\*Consistency\*\*: The taxonomy does not include or allow for any contradictions.

### Requirements of your output taxonomy

- Your output \*\*intent\*\* taxonomy should focus on the user actions in a task, not the task objects. This is \*\*different\*\* from a \*\*domain\*\* taxonomy, which primarily describes the task objects.

- Your output taxonomy should match the data as closely as possible, without leaving out important intent categories or including unnecessary ones. Please make sure there is no overlap or contradiction among the intent categories in your output taxonomy.

- Your output \*title\* of each category should be \*no more than 3 words\*. The title should be a concise and clear label for the intent category. It can be either verb phrases or noun phrases, whichever is more appropriate.

- Your output \*description\* of each category should be \*no more than 30 words\*. The description should explain the user's goal or purpose for the intent category, and should differentiate it from other intent categories.

- The number of examples for each intent category should be \*no more than 3\*. The examples should either come from the given taxonomy or the provided data with \*\*exactly the same content\*\*. Please do not invent new examples or intents that are not in the given taxonomy or the data.

- \*\*Size limit of the output taxonomy\*\*: The total number of intent categories should be \*\*no more than 5\*\*.

- Your output taxonomy and examples should be in \*English\* only.