



*La place des usagères et usagers dans les outils de fouille et d'exploration de données (PAUL)
Atelier de la Conférence EGC 2024*



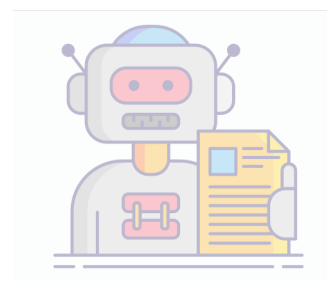
Towards Collaborative Business Intelligence with Conversational Agents: A User-Centric Approach

Olga CHEREDNICHENKO and Muhammad FAHAD

BI4People



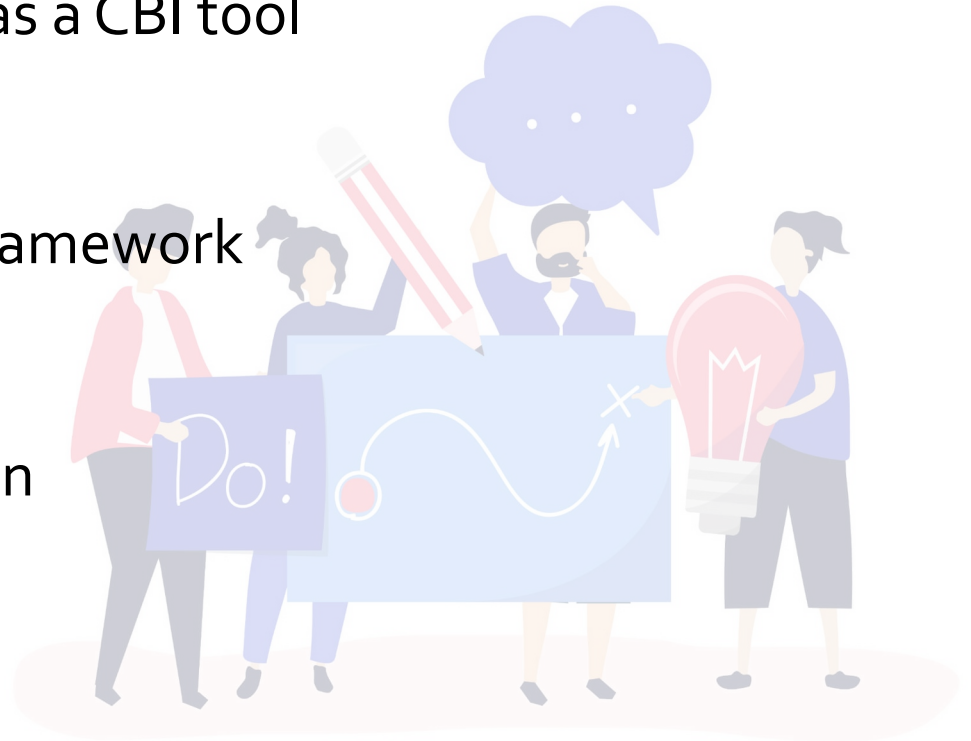
project ANR-19-CE23-0005 BI4people



UR ERIC, Univ-Lyon2
Dijon, 23/01/2024

Agenda

- Collaborative BI concept
- The ways of the CBI classification
- Conversational Agent as a CBI tool
- The main workflow
- CBI Virtual Assistant Framework
- Research questions
- Discussion & Conclusion

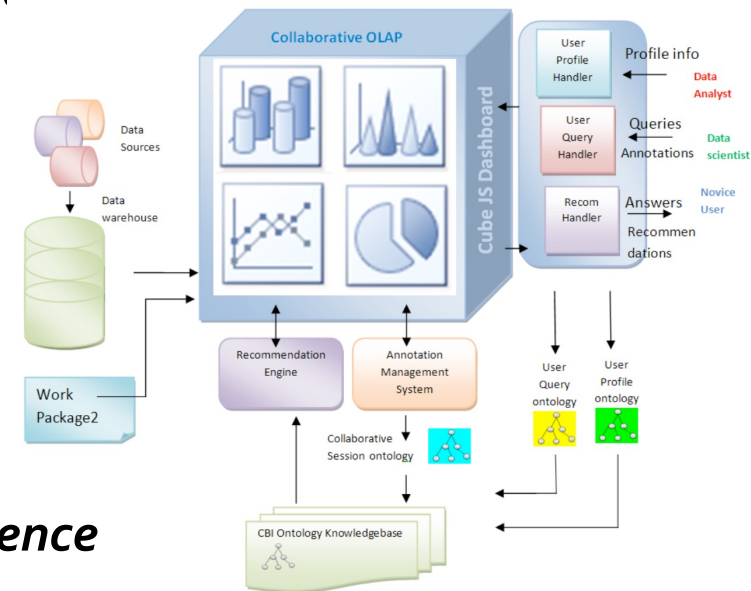
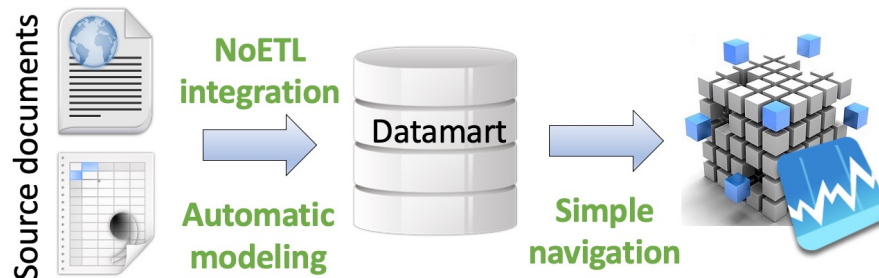


Business Intelligence for people



BI₄People
project ANR-19-CE23-0005

- **The aim** of BI₄people is to bring the power of OLAP interactive analysis **to the largest possible audience**, by implementing the data warehousing process in software-as-a-service mode, from multisource, heterogeneous data integration to very simple OLAP-like analysis and data visualization



Collaborative space & reusing experience

- The main idea is to **collect people in one** virtual or real **space** and **encourage them leave their comment or opinions** for general purpose. Moreover, **reusing another collaborators' results** or comments makes general BI - CBI.

The main points of the concept of CBI

- **Collaboration:** CBI is centered around the idea of collaboration and knowledge sharing among team members. By facilitating communication and collaboration, organizations can create a culture of data-driven decision-making.
- **Data sharing:** CBI involves sharing data across teams and departments. This means breaking down data silos and ensuring that everyone has access to the same data, regardless of where it is stored.
- **Visualization:** CBI leverages data visualization tools to make data more accessible and easier to understand. This allows team members to quickly identify trends and patterns and make informed decisions.
- **Agility:** CBI is agile, meaning that it enables teams to quickly adapt to changing business needs and make decisions based on the latest data.
- **Continuous Improvement:** CBI is a continuous process of improvement, with team members working together to identify opportunities for improvement and make changes to their processes accordingly.

Related Works

Approaches to implementing BI:

- Social BI,
- Mobile BI,
- Cloud-based BI,
- Self-service BI
- Embedded BI

Tools for facilitation CBI

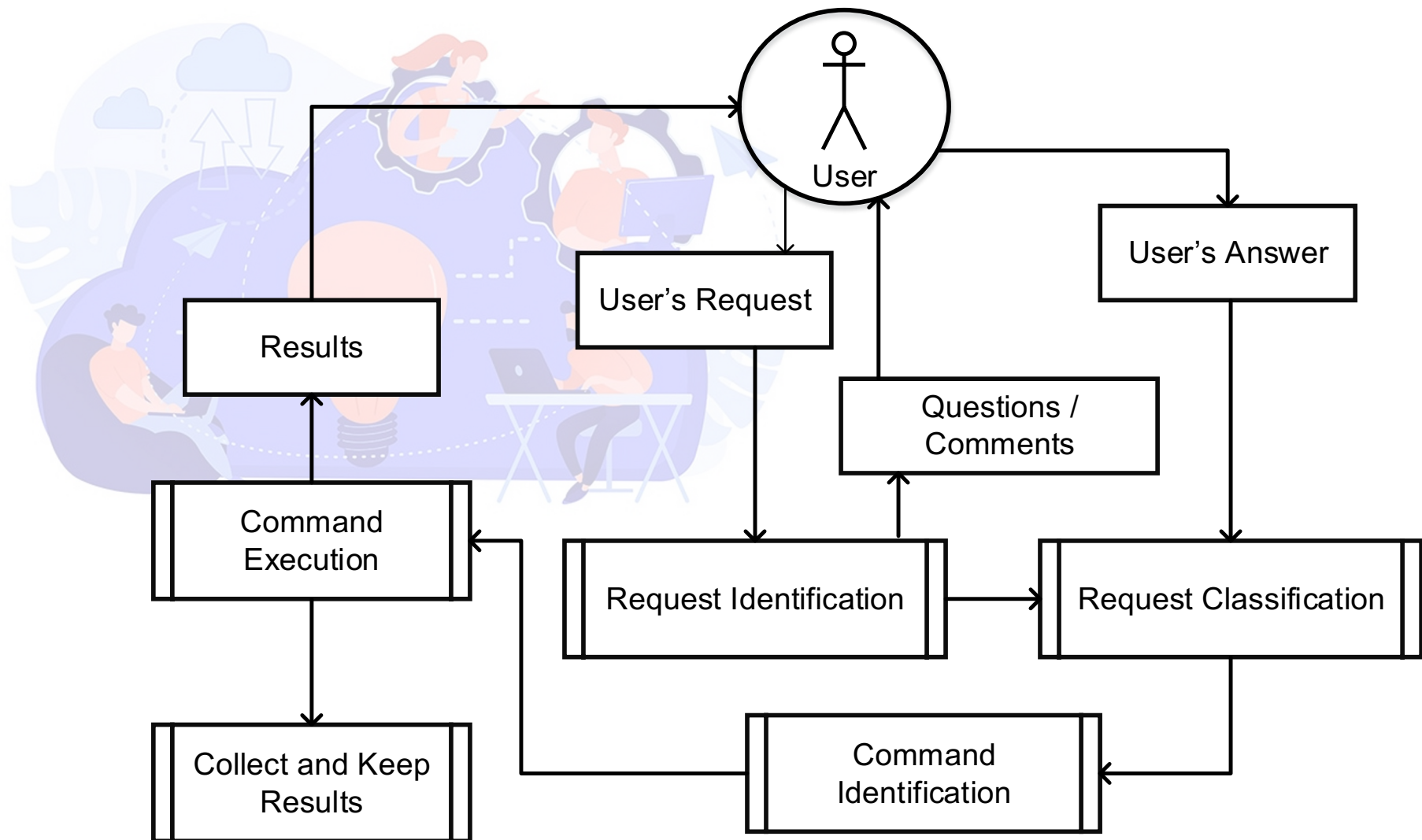
- Chatbots
- Conversational Agents
- Dialogue systems

CBI types:

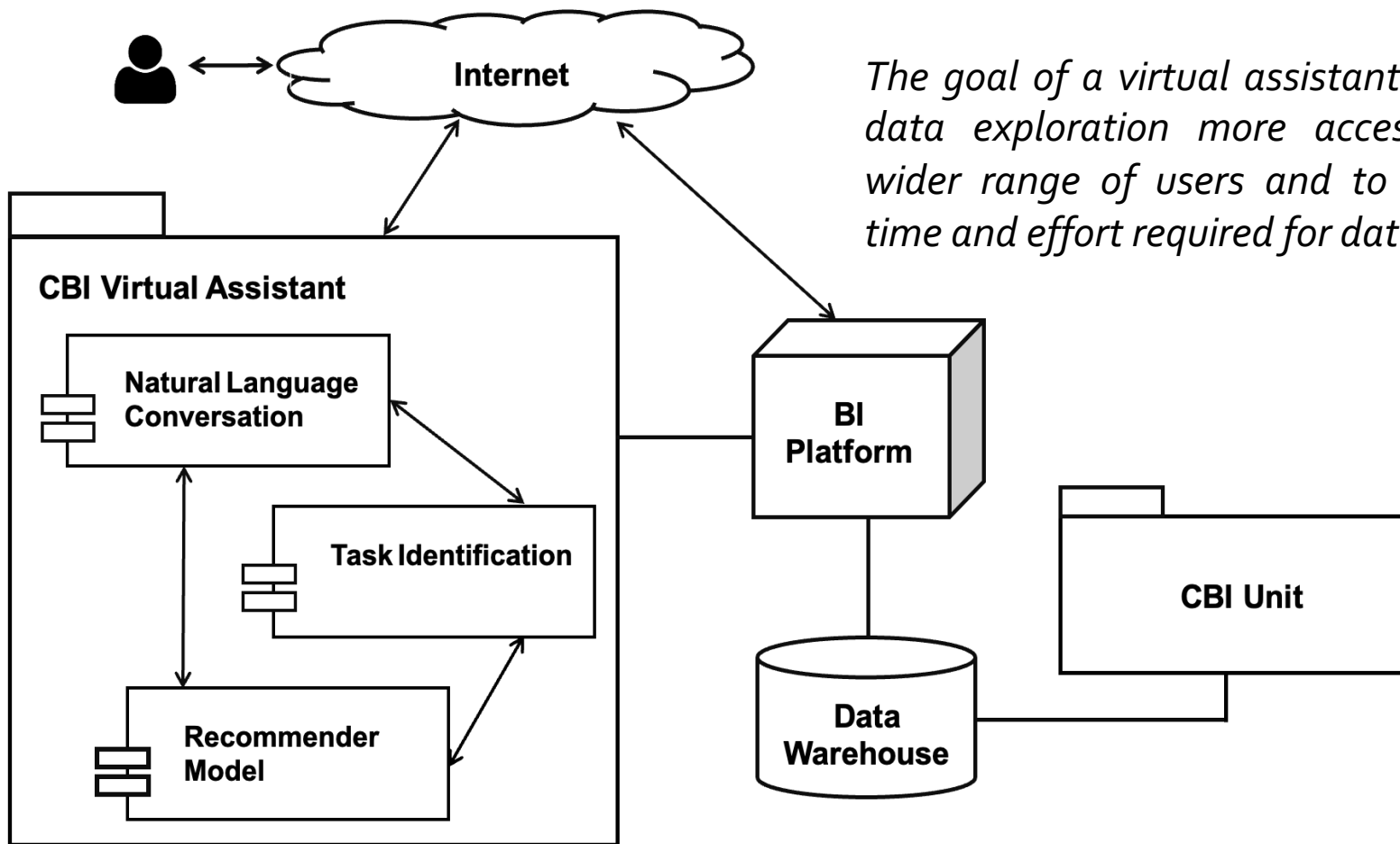
- Internal-collaborative BI
- Partnership in data
- Partnership in analysis



The main workflow



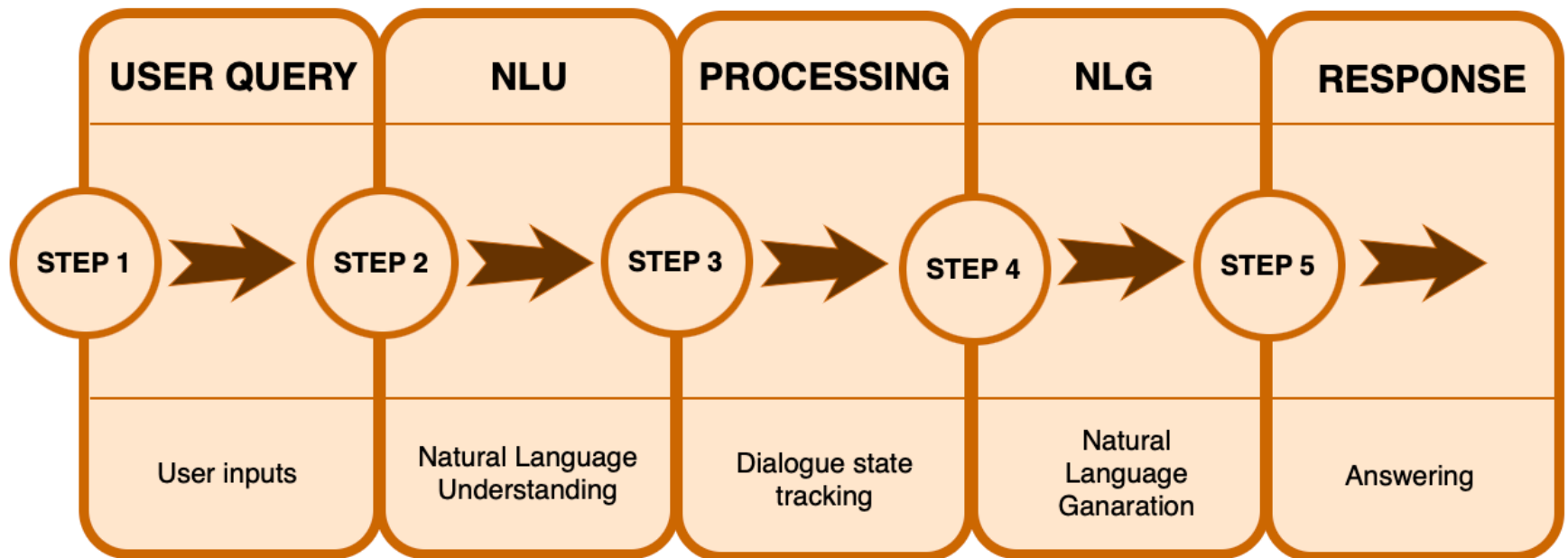
CBI Virtual Assistant Framework



The goal of a virtual assistant is to make data exploration more accessible to a wider range of users and to reduce the time and effort required for data analysis

The basic pipeline for conversational agent

Conversational agents (CAs) are computer programs designed to engage in natural conversations with human users. They can be categorized as either chatbots for informal chatting or task-oriented agents for providing users with specific information related to a task



Motivation



- ▣ *The text generation models are pre-trained on a vast amount of text.*
- ▣ *To use the models effectively some instructions and several examples are included in a prompt. Using demonstrations to show how to perform a task is often called "few-shot learning."*
- ▣ *Fine-tuning improves on few-shot learning by training on many more examples than can fit in the prompt, letting you achieve better results.*
- ▣ *Fine-tuning can make the text generation models better, but it requires a careful investment of time and effort.*
- ▣ *The way to get good results is it with prompt engineering, prompt chaining.*
- ▣ *Typically, the best results are obtained when combining prompt chaining and tool use with fine-tuning.*

Few-shot learning vs Fine-tuning => Crafting prompts

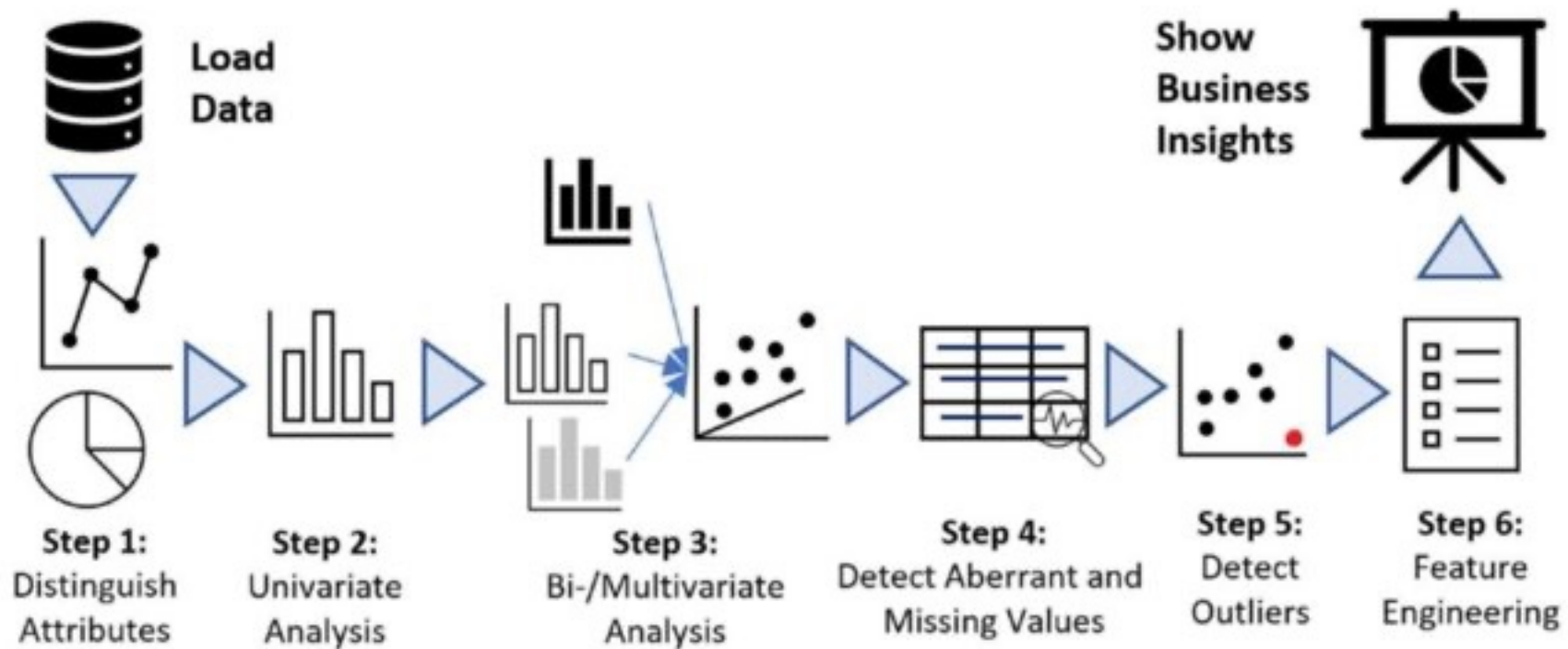
Research questions

RQ1 : How can pretrained generative language models be effectively utilized for data query construction to support non-expert users ?

RQ2 : What are the essential components of a prompt created for a generative language model to construct data queries effectively?

The generation of prompts, integral to user interaction, necessitates an understanding of the user's context, preferences, and analytical objectives. The research question centers on the efficacy of a generative language model in facilitating data exploration for non-expert users. The experimental pursuit seeks to delineate a linguistic framework for prompt construction and the development of a user query reference model.

Exploratory Data Analysis (EDA)



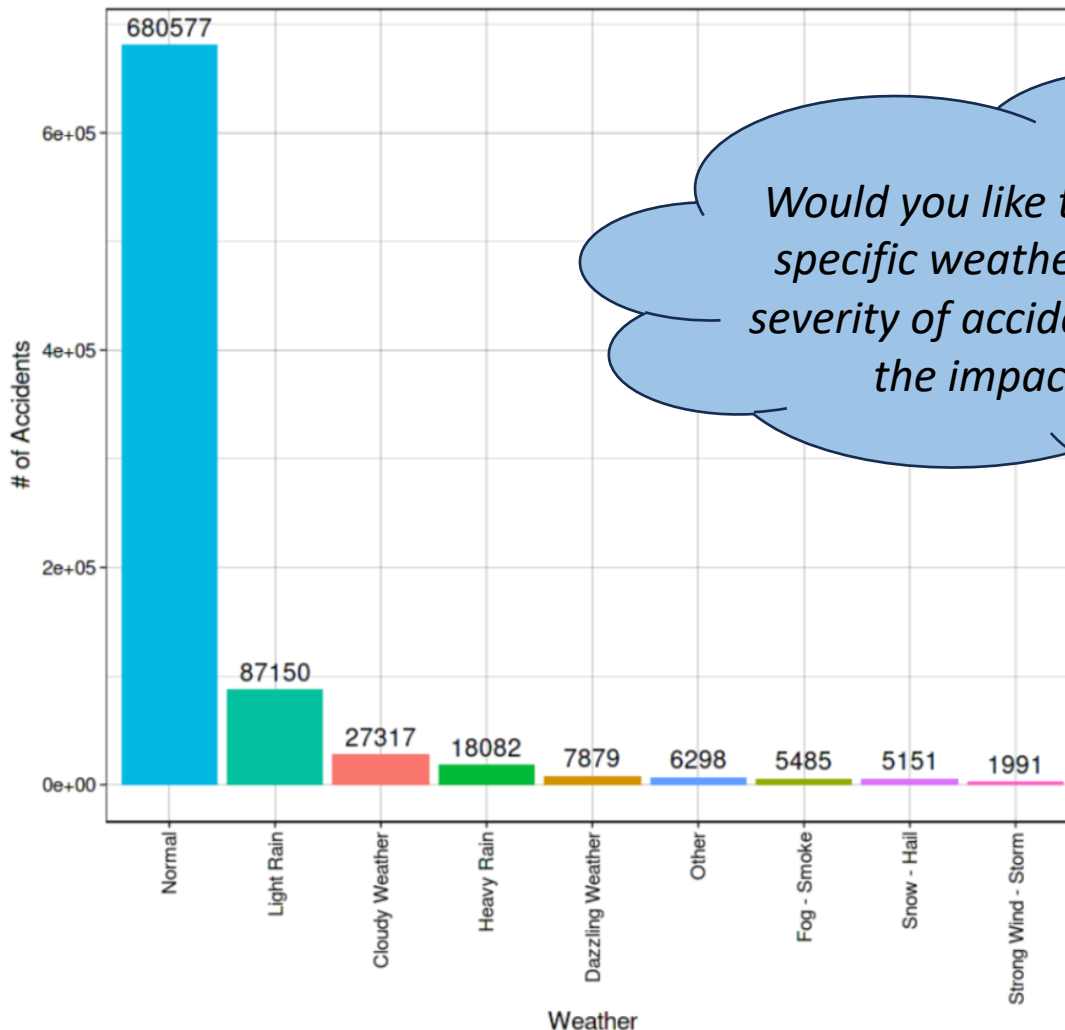
For our experimental endeavors, a deliberate emphasis is placed on univariate and bivariate analyses, predicated on the presumption that the non-expert user is the focal investigator. The experiment involves manipulating collocations from EDA steps, generating prompts, user queries, and employing them to interact with the dataset. The generated grades are analyzed to build a prompt framework facilitating non-expert users in initiating EDA.

Data set

Dataset of road accidents in France (**Bases de données annuelles des accidents corporels de la circulation routière - Années de 2005 à 2021**)

https://www.data.gouv.fr/en/datasets/bases-de-donnees-annuelles-des-accidents-corporels-de-la-circulation-routiere-annees-de-2005-a-2021/#_

Accidents by Weather



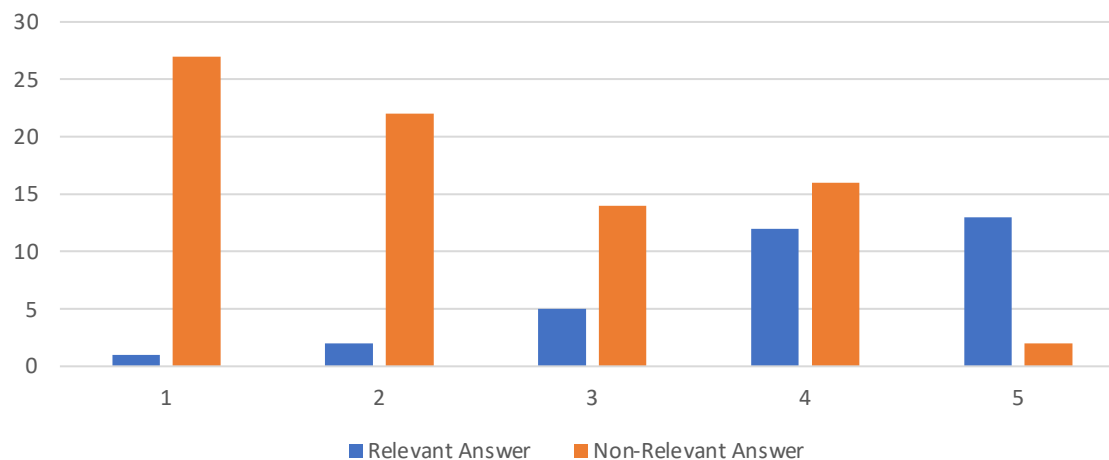
Would you like to explore the correlation between specific weather conditions and the frequency or severity of accidents in France to better understand the impact of weather on road safety?

	weather	weather_n
1	Normal	680577
2	Light Rain	87150
3	Heavy Rain	18082
4	Snow - Hail	5151
5	Fog - Smoke	5485
6	Strong Wind - Storm	1991
7	Dazzling Weather	7879
8	Cloudy Weather	27317
9	Other	6298

Experimenting

You are a data analysis assistant. The user is not an expert. You explore data about accidents in France. You have dataset and its description. Suggest a question to the data. the question should be in the form "would you like to explore ...?"

- ✓ *Would you like to explore the distribution of reasons for travel in the accidents dataset to gain insights into the primary factors contributing to accidents in France?*
- ✓ *Would you like to explore a breakdown of the different reasons for travel in the dataset?*



- ✗ *Can you provide a breakdown of the different reasons for travel in the dataset?*
- ✗ *What are the different reasons for travel (traject) recorded in the dataset, and how frequently do they occur?*

Conclusion and Discussion

- The pretrained models, such as GPT-3.5, can be harnessed to facilitate data exploration.
- By framing the interaction within the context of a data analysis assistant catering to a non-expert user, we achieved relevant prompts that contributed to a more user-friendly data exploration experience.
- Our experiments revealed that the most pivotal elements include a detailed description of the query and explicit identification of the feature to be analyzed.
- The specific command plays a less decisive role in comparison to the dataset information and feature name.
- By understanding the nuanced requirements of constructing prompts, researchers and practitioners can harness the potential of such models to enhance the accessibility and usability of data analysis tools for a broader audience.

THANK YOU FOR YOUR ATTENTION!

