Applied Data Science Project

Designer AI I

Mehrbod Nowrouz , Masoud Khalilian , Baharak Qaderi

Introduction

As Artificial Intelligence is moving forward, it is making its way into educational systems. With examples like ChatGPT, we have seen the role of these AI models in helping students with better understanding of various topics. Therefore, implementing AI into education serves as the future of education and brings higher quality education which is the 4th goal of UNSD.

Our project is also aiming to develop a module that would help students with creating and understanding ER models through turning their text instruction of the logic into an ER model.

Esercizio 1: Gestione azienda di traslochi aziendali

Si vuole progettare la base di dati per la gestione di traslochi aziendali. La base di dati deve contenere un elenco di furgoni adatti per effettuare i traslochi. I furgoni sono identificati dalla targa, e sono caratterizzati dal modello e dal volume in metri cubi, se disponibile. Tra tutti i furgoni, alcuni sono abilitati al trasporto di carichi speciali, e solo per tali furgoni è noto un elenco di certificazioni dei materiali speciali che sono autorizzati a trasportare. La base di dati deve contenere un elenco di magazzini, identificati da un codice e caratterizzati dall'indirizzo e dal nome dell'azienda a cui appartengono. Si vuole tenere traccia di tutti i traslochi effettuati. I traslochi sono identificati dalla data e dal furgone con cui sono effettuati, e sono caratterizzati dal nome dell'autista che lo effettua. Ogni trasloco è inoltre caratterizzato dal magazzino di partenza e dall'orario di partenza, e dal magazzino di arrivo e dall'orario di arrivo.
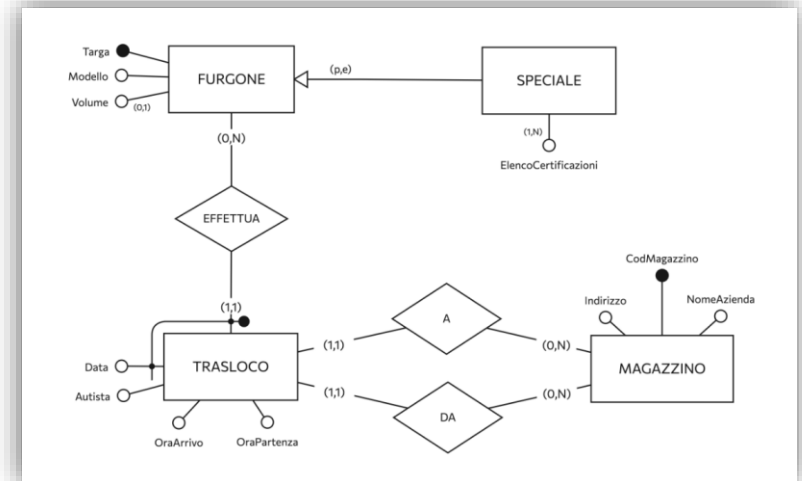
Challenges

1 Collecting data regarding instruction and its corresponding ER model.

2 Pre-processing plain text into a format suitable for an LLM model to produce the intended diagram

3 Evaluating the model with a ground truth diagrams.

4 Define metrics to quantitatively assess the quality of the final output

STAKEHOLDERS LIST

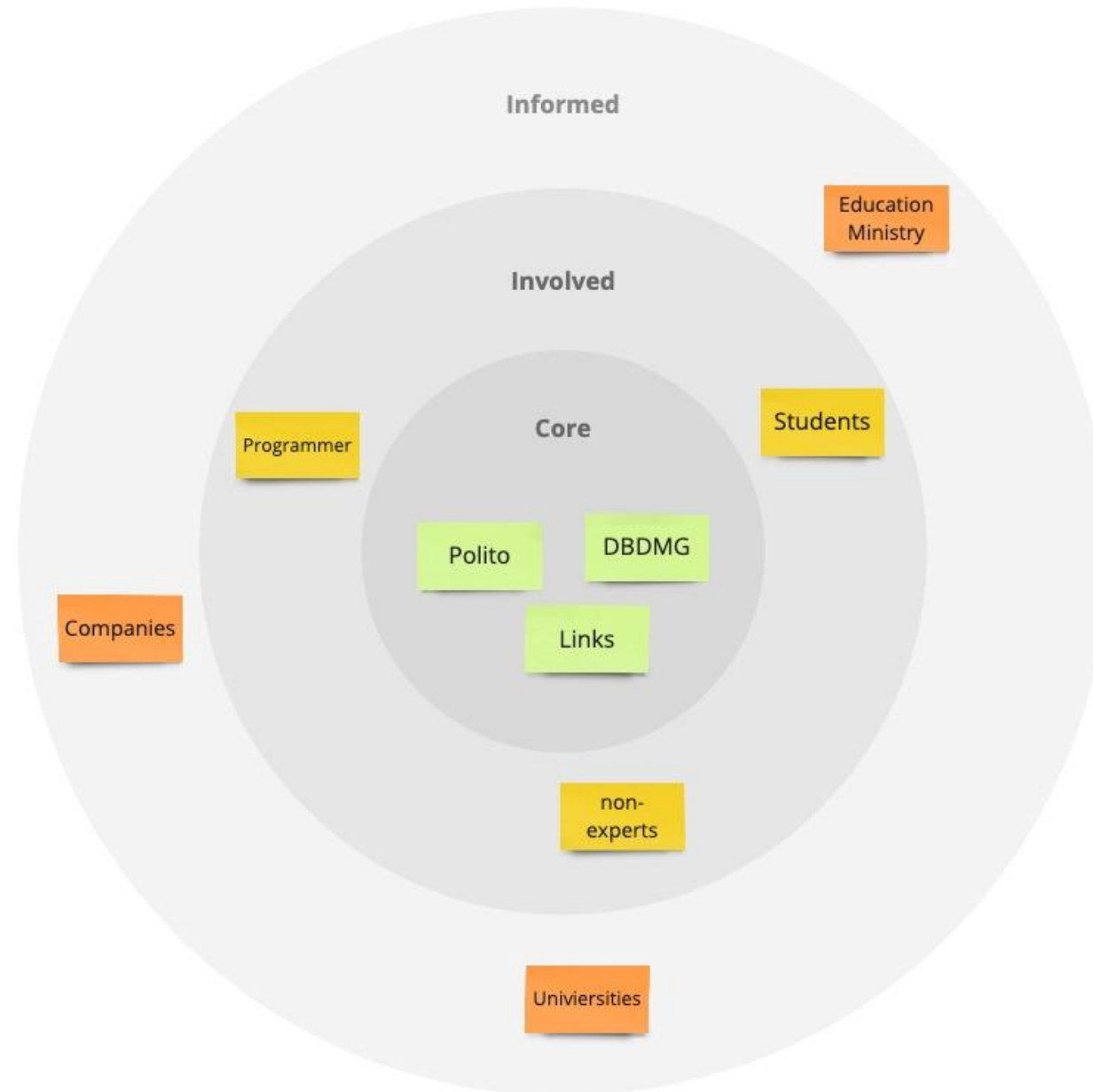|	Their urgent need	Data they expect	Data they produce
	Need: 1. Better and easier teaching 2. Easier exam question design 3. Better educated graduates.	**Produce:** 1. A collection of database design exercises. 2. Access to Designer tool to create new models and understand the internal data structure.	**Expect:** Performance data of the produced model Sample Diagrams produced by the model.
	Need: 1.Better and easier education 2. Programming without needing expertise	**Produce:** Performance data of the users before and after usage the developed model.	**Expect:** The diagram that fits their description of the logic.
	Need: 1.Links needs to satisfy the customer (DMBDMG) with producing their desired model. 2.Polito needs better education for its students.	**Produce:** 1.Links provides connection between the team and DBDMG (Customer) 2.Polito provides data regarding the student performances in target classes.	**Expect:** A functional module that can be put into practical use.
	Need: Privacy and data protection of people residing in Eu countries.	**Produce:** Rules to follow to keep the Eu citizens safe in terms of Data and Privacy	**Expect:** An AI module that adheres to the rules provided for Data protection and privacy.

STAKEHOLDERS MAP

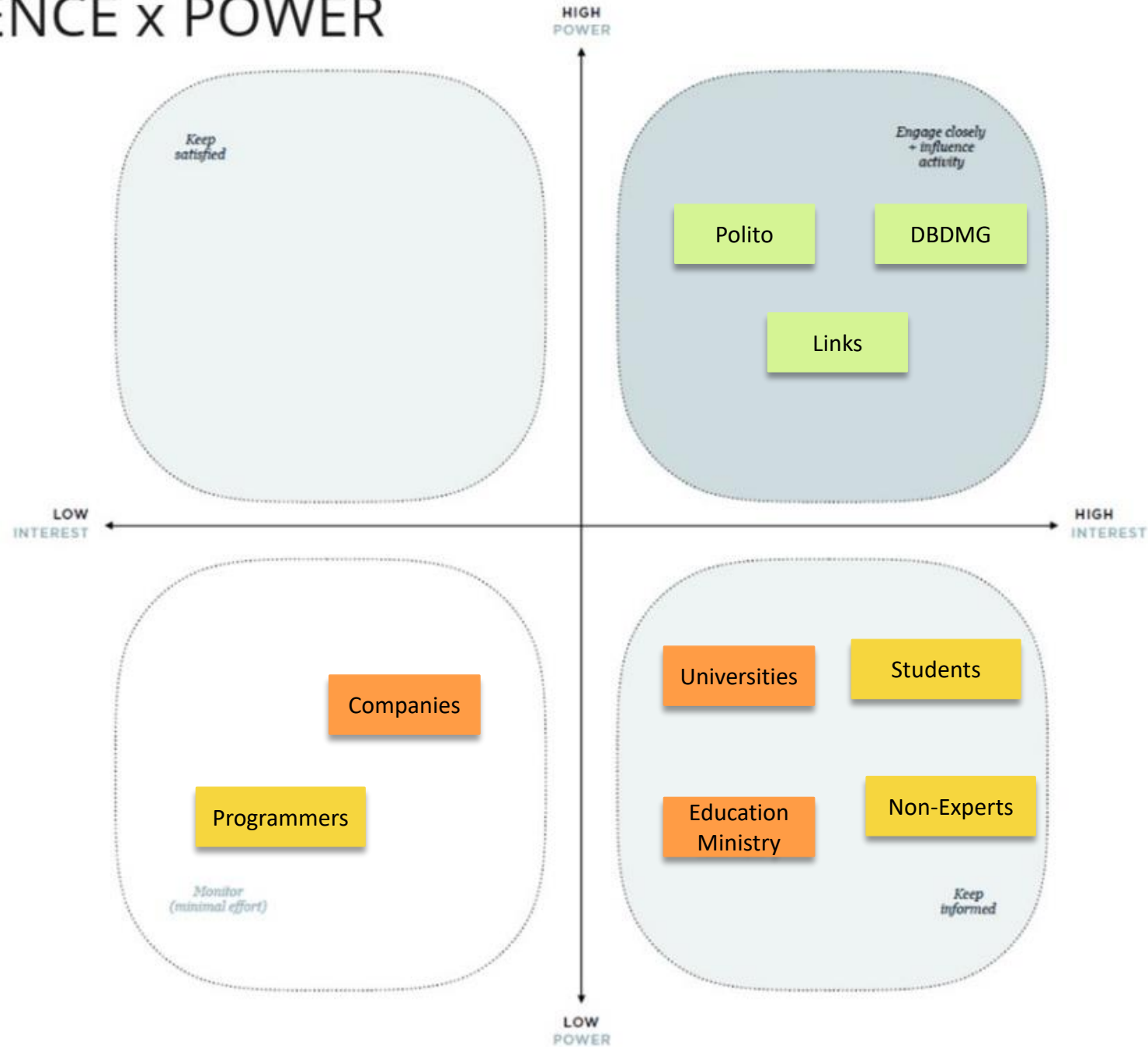**Core**

End users, Direct beneficiaries

Involved

Facilitators that can help to promote and encourage the adoption

Informed

Supervisors or experts to keep informed

Informed

Involved

Core

Education Ministry

Students

Programmer

Polito

DBDMG

Links

Companies

non-experts

Univiersities

miro

User Stories

As a <user role>

I want <goal>

so that <benefit>.

Requirements

- Functional Requirements:

	Requirements	Status	Actions	Priority
1	The system gets input from the user.	Agreed	Develop a module to take the initial input from the user.	Must have
2	The system transforms user input into a structured format compatible to the language model.	Agreed	Develop a custom module to transform the user input to define and clarify the objective for LLM.	Must have
3	The system handles the API to LLM modules.	Agreed	Develop a custom API module for sending, receiving and error handling of API calls.	Must have
4	The system turns the output of the language model into a JSON file.	Agreed	Develop a translation module from LLM to JSON format.	Must have
5	The system saves the output to user's system.	Agreed	Develop a output module.	Must have
6	The system adheares to constraints of the designerdb.io	Agreed	Develop the module adherent to the rule of ER model	Must have
7	The system validates the developed ER model.	Agreed	Develop a module to validate the output.	Must have
8	The system proprocesses the input text instructions.	Agreed	Develop a pre-processing module before training the model.	Must have
9	The system logs the processes of the system.	Agreed	Develop a time tracking module for crucial processes of the system.	Must have

- Non-Functional Requirements:

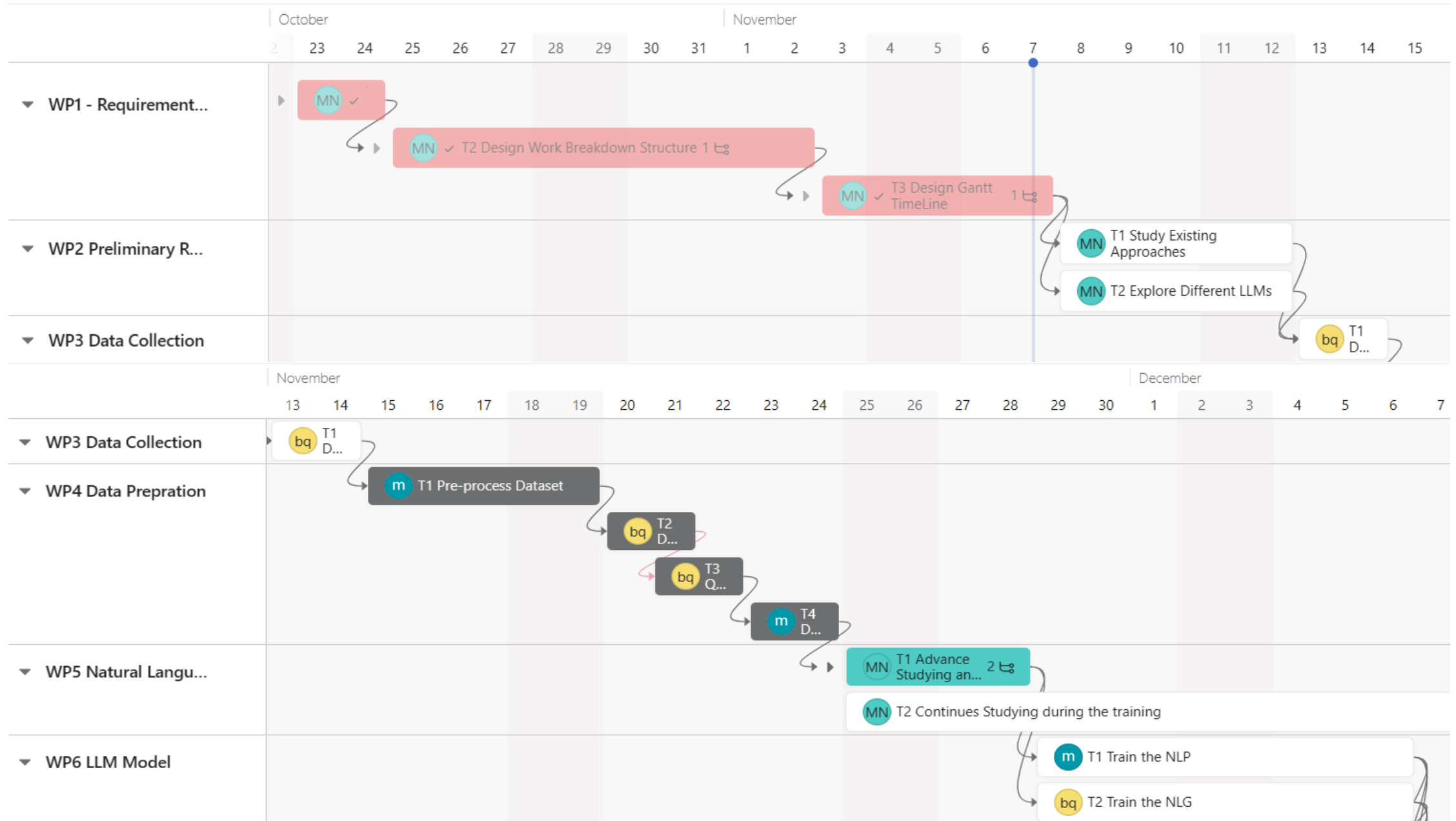|	Refrence personas	Requirements	Priority
1	s	The system should have entry point	Must have
2	s	The system output should be evaluated by external parties	Must have

Diagram
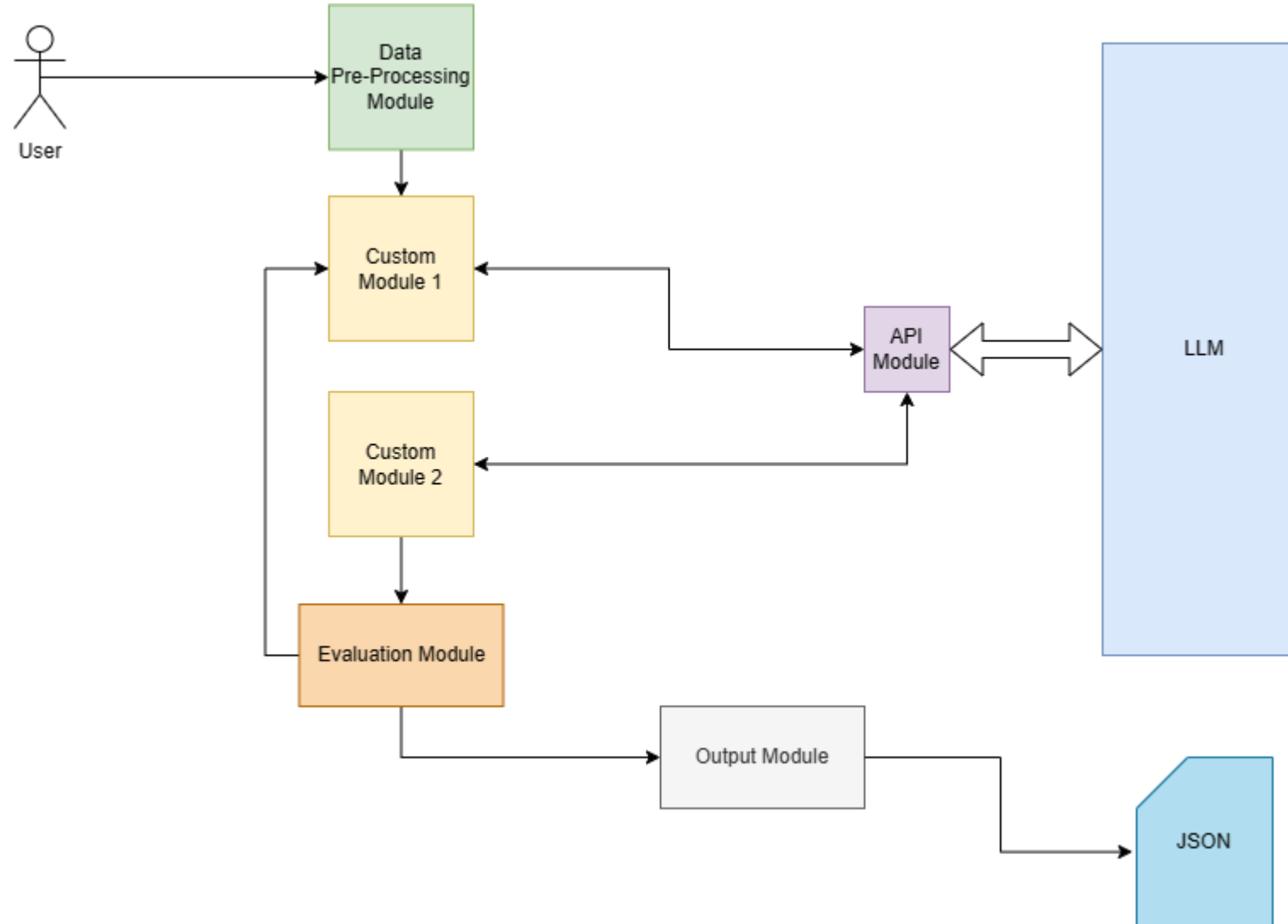
Work Packages

Task name	Assignee	Due date	Collaborators	Person Days
WP1 - Requirement Analysis				
✓ T1 Design Users Requirements Table 1 ⊟	MN Mehrbod No...	Oct 23 – 24	MN m bq	1
✓ T2 Design Work Breakdown Structure 1 ⊟	MN Mehrbod No...	Oct 25 – Nov 2	MN m bq	7
✓ T3 Design Gantt TimeLine 1 ⊟	MN Mehrbod No...	Nov 3 – Today	MN m bq	4
WP2 Preliminary Research ⚡				
⊘ T1 Study Existing Approaches	MN Mehrbod No...	Nov 8 – 12	MN m bq	4
⊘ T2 Explore Different LLMs	MN Mehrbod No...	Nov 8 – 12	MN m bq	4
WP3 Data Collection				
⊘ T1 Data Collection	bq baharak.qad...	Nov 13 – 14	MN bq m	1
WP4 Data Prepration ⚡				
⊘ T1 Pre-process Dataset	m masoud	Nov 15 – 19	MN m bq	4
⊘ T2 Data Enrichment	bq baharak.qad...	Nov 20 – 21	MN m bq	1
⊘ T3 Quality Check	bq baharak.qad...	Nov 21 – 22	MN m bq	1
⊘ T4 Dataset Structure Definition	m masoud	Nov 23 – 24	MN bq m	1

Work Packages

WP5 Natural Language Processing

▶ ⧗ T1 Advance Studying and Integrate Related Models 2	MN Mehrbod No...	Nov 25 – 28	MN m bq 3
⊘ T2 Continues Studying during the training	MN Mehrbod No...	Nov 25 – Dec 29	MN bq m

WP6 LLM Model

⧗ T1 Train the NLP	m masoud	Nov 29 – Dec 6	MN m bq 7
⧗ T2 Train the NLG	bq baharak.qad...	Nov 29 – Dec 6	MN m bq 7

WP7 Model Evaluation

⧗ T1 Evaluate Performance	bq baharak.qad...	Dec 7 – 24	MN bq m 17
⧗ T2 Evaluate Text Quality	MN Mehrbod No...	Dec 7 – 24	MN m bq 17
⧗ T3 Evaluate Text Accuracy	m masoud	Dec 7 – 24	MN m bq 17

WP8 Model Optimization

⧗ T1 fine-tune	MN Mehrbod No...	Dec 25 – 29	MN m bq 4

WP9 Paper Writting

⧗ T1 Write the corresponding paper.	MN Mehrbod No...	Dec 30, 2023 – Jan 17, 2024	MN bq m 17

Gantt Chart

Gantt Chart

Github