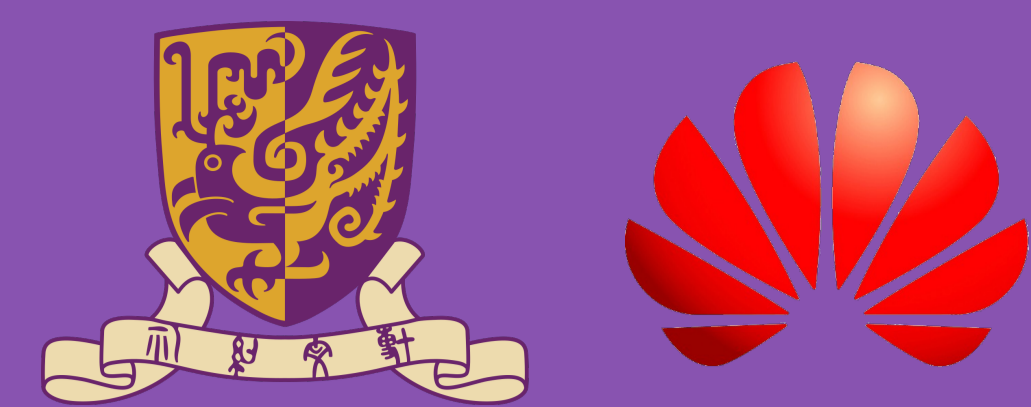


Large Language Models as Source Planner for Personalized Knowledge-Grounded Dialogue

Hongru Wang, Minda Hu, Yang Deng, Rui Wang, Fei Mi, Weichao Wang
Yasheng Wang, Wai-chung Kwan, Irwin King, Kam-Fai Wong
{hrwang, kfwong}@se.cuhk.edu.hk ydeng@nus.edu.sg

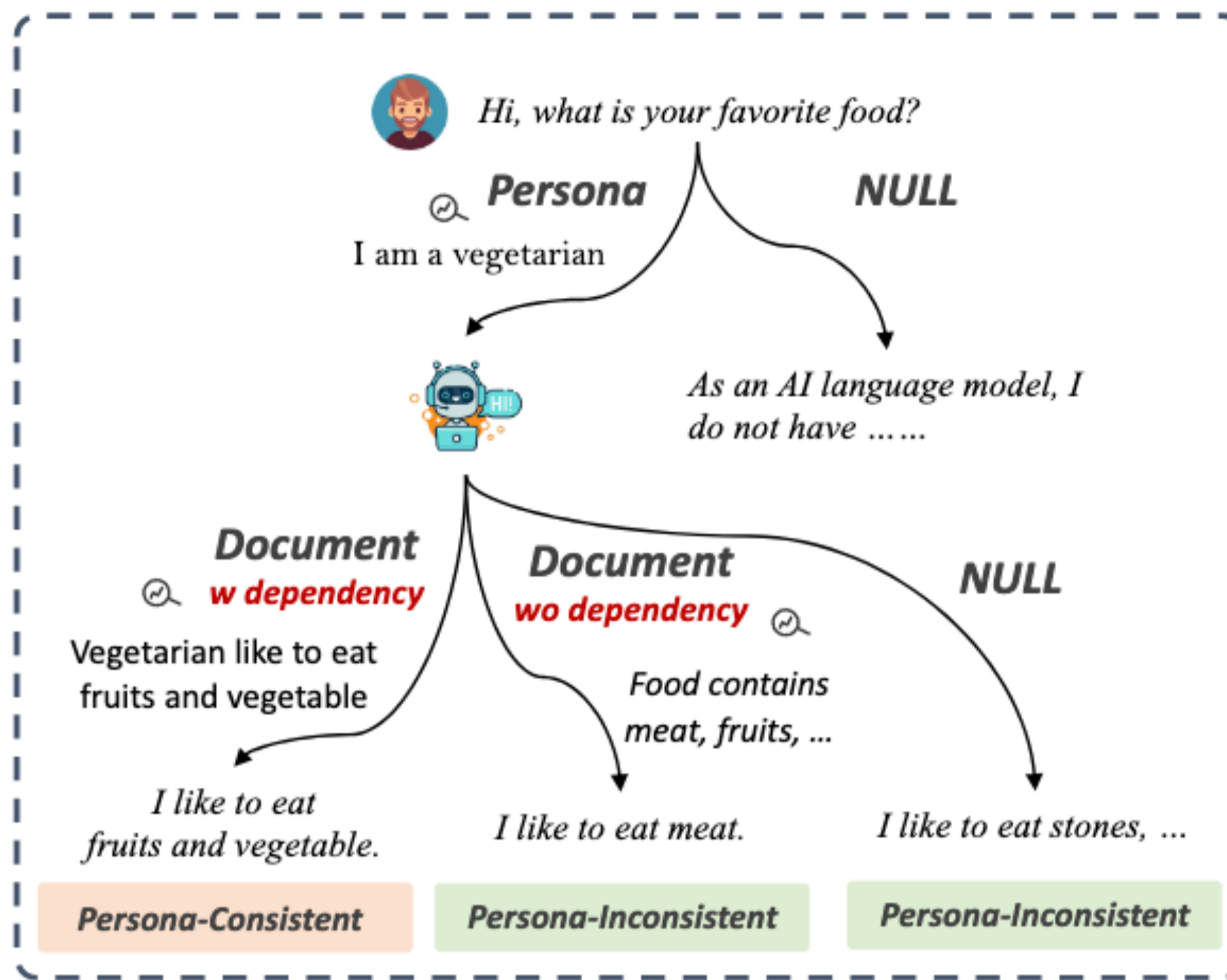
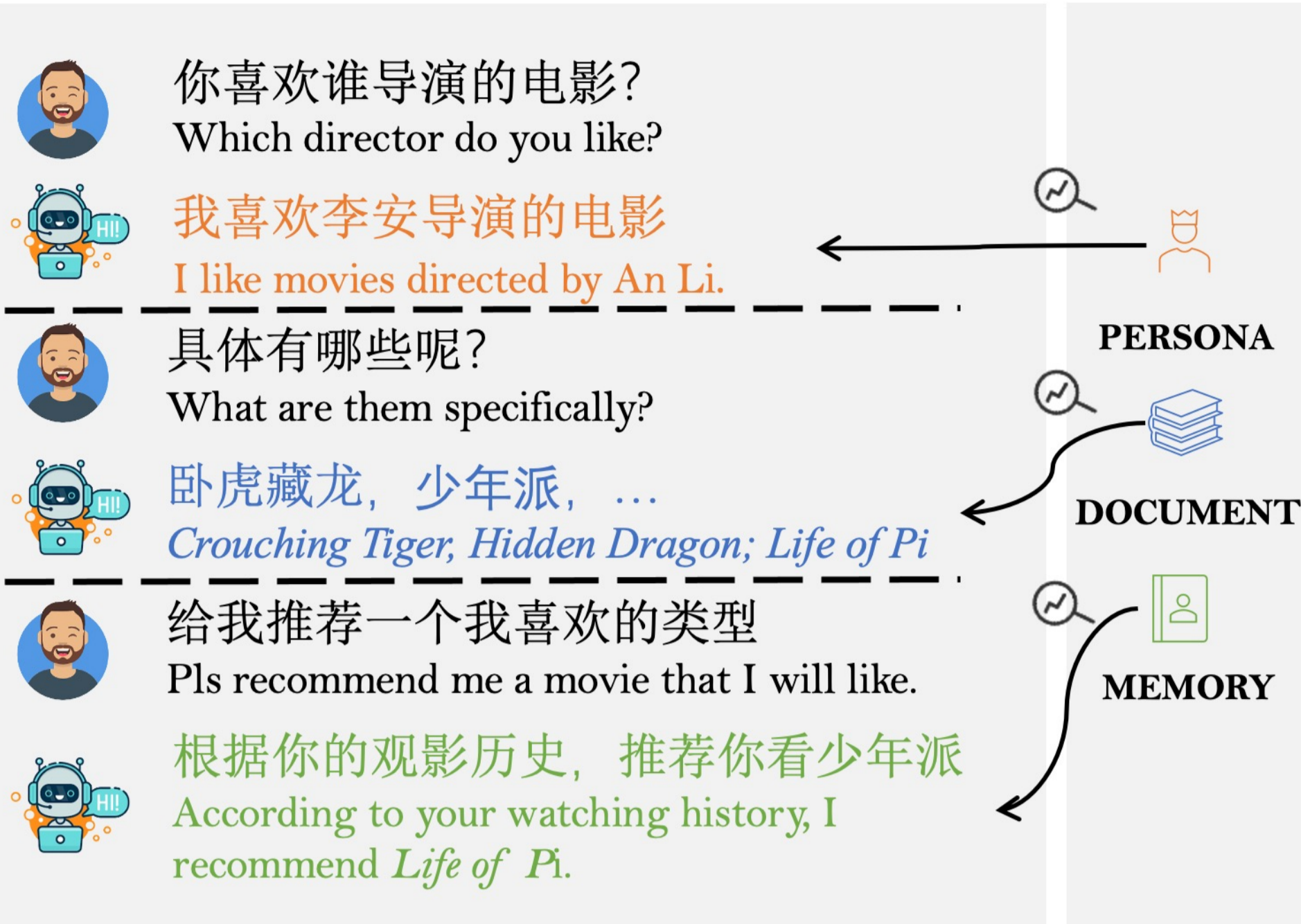
Code

Homepage



Dialogue Context

Sources

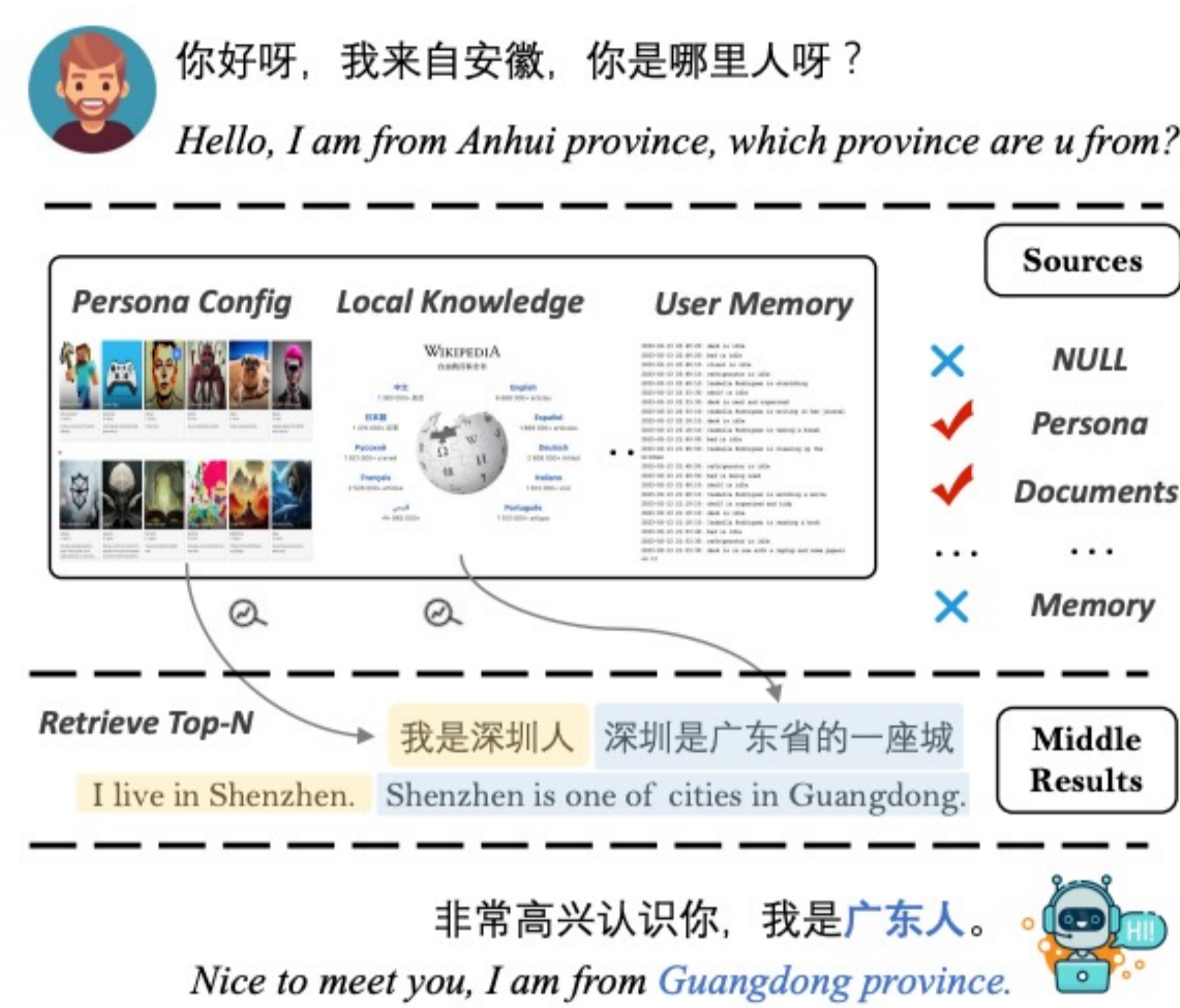


Dependency between Multiple Sources

- Open-domain Dialogue System requires access to various external knowledge sources to deliver **reliable, informative, personalized, and helpful responses**, depending on which sources are invoked.
- Indiscriminately incorporating all sources bring **unnecessary computing cost**, and sometimes it does not require external knowledge.
- The **interdependence** between different external sources brings new challenges, while ignoring the complex relationship between different sources, leads to sub-optimal performance.

Motivation

Method



Planning: make a series of decision to determine whether or not use knowledge, which and when.

$$M : c \rightarrow K_i, K_j, \dots, K_n \text{ or } \text{NULL}, \quad (1)$$

Retrieval: retrieve top-n results from local databases according to the decided used source knowledge

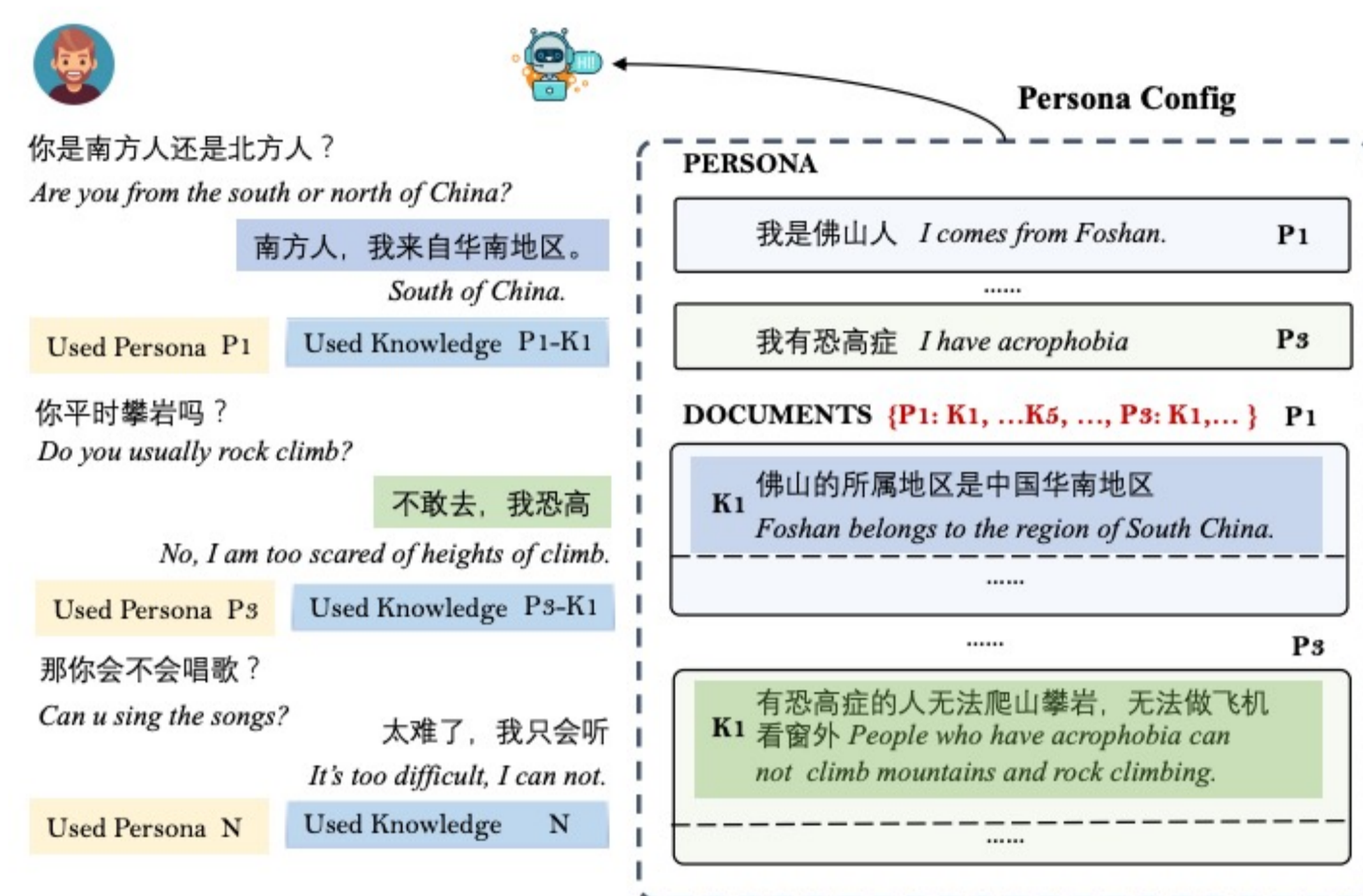
$$R : K_i, K_j, \dots, K_n \rightarrow k_i^j, \dots, k_n^m \quad (2)$$

Assembling: incorporate all retrieved middle results into the final response generation

$$M : \text{Inp} \rightarrow s_t, \quad (3)$$

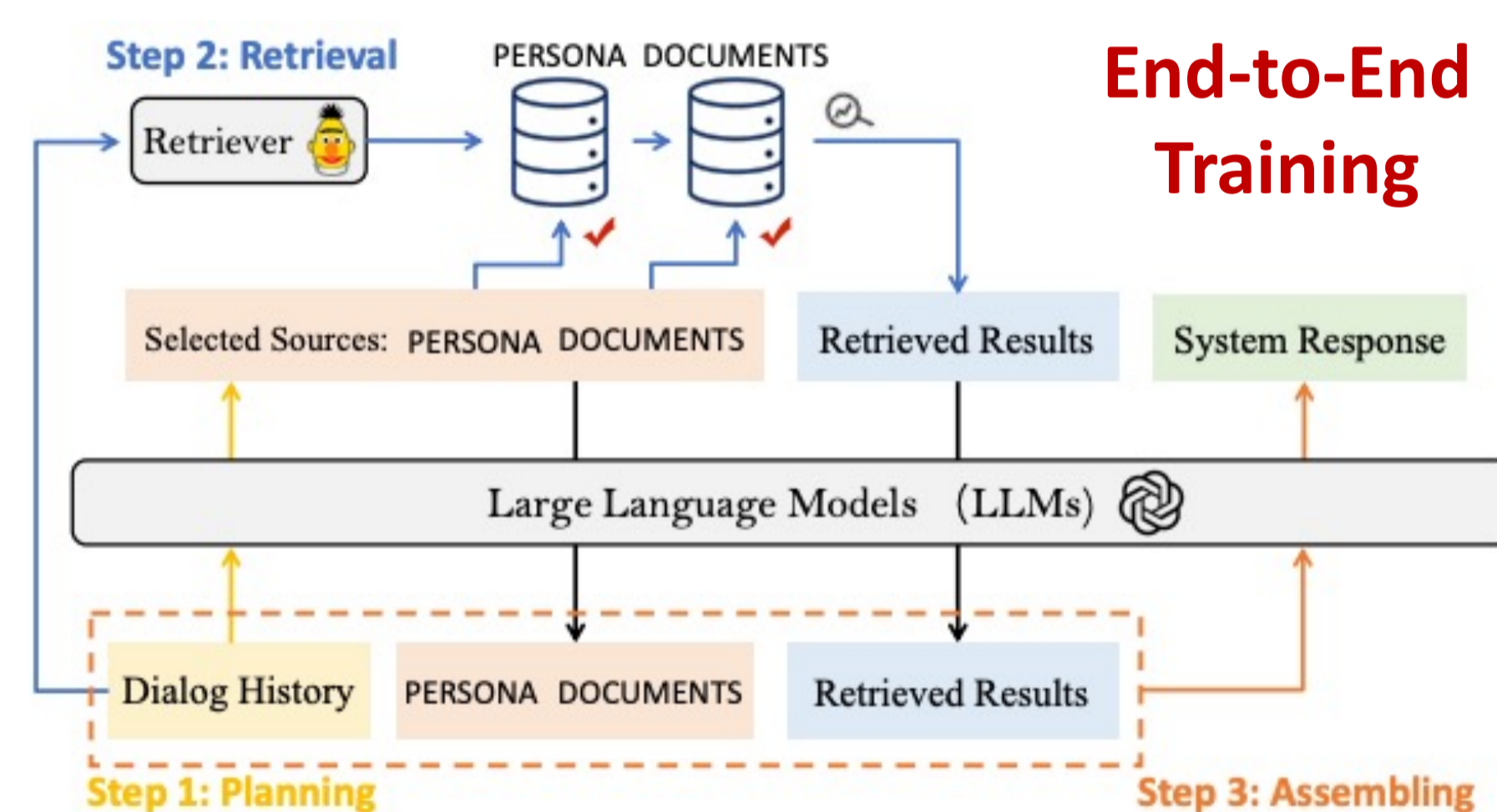
where $\text{Inp} = \{c \text{ [SOURCE]} K_i, \dots, K_n \text{ [EOS]} \text{ [MIDDLE]} k_i^j, \dots, k_n^m \text{ [EOM]}\}$.

KBP Dataset



Supervised SAFARI

Unsupervised SAFARI



There are different knowledge bases storing relevant information:
 $K_1: \{K_1_DESC\}$
 $K_2: \{K_2_DESC\}$
.....
There exists a dependency between these knowledge bases.
 $\{DEPENDENCY_DESC\}$
Here is the dialogue between the user and the system: $\{DIALOGUE\}$
Based on the user's last question, please determine if it requires invoking the corresponding knowledge base. If the invocation is necessary, output the names of the knowledge bases in the order they should be invoked. If no invocation is needed, output NULL.

The dialogue is as follows:
 $\{DIALOGUE\}$
The following knowledge is retrieved from different sources of knowledge bases:
 $\{MIDDLE_RESULTS\}$
Please play the role of the system and generate a reply according to the context of the dialogue and given knowledge. Please make sure your reply is consistent with the given knowledge. If the provided knowledge is NULL, generate a response solely based on the dialogue context.
System:

- Performance of **Planning**
 - Supervised ChatGLM > Supervised BELLE > Unsupervised ChatGPT > Others
- Performance of **Retrieval**
 - DPR > RocketQAv2 > BM25
- Performance of **Assembling**
 - Supervised BELLE > Supervised ChatGLM

Model	Persona	Both	
		PERSONA	DOCUMENTS [†]
BM25	36.80	48.97	11.37
RocketQAv2	80.00	92.31	35.75
DPR	83.20	93.07	39.33

Table 5: The performance of **Retrieval** of different types of retrievers. There are 125 examples that only require PERSONA and 923 require both PERSONA and KNOWLEDGE. We also report the Recall@1 of DOCUMENTS without dependency (DOCUMENTS[†]).

Model	BLEU1	Rouge-L	P.C	K.C
BELLE-LLAMA-7B-2M	30.48	34.61	75.34	46.62
CHATGLM-6B	23.81	26.70	76.99	42.39
Unsupervised Setting				
Zero-shot				
BELLE-LLAMA-7B-2M	11.84	19.24	30.59	27.34
CHATGLM-6B	6.18	14.50	14.73	24.73
CHATGPT	12.06	24.44	73.47	38.00
In-context				
BELLE-LLAMA-7B-2M	19.51	22.25	72.98	24.89
CHATGLM-6B	13.74	19.69	16.92	24.89
CHATGPT	16.03	25.62	46.38	35.56

Table 6: The performance of **Assembling** under supervised/unsupervised settings.

Model	BLEU1	RougeL	P.C	K.C
CHATGLM-6B	23.81	26.70	76.99	42.39
+ Ground Planning	24.29	27.01	86.16	57.12
+ Ground Retrieval	25.86	29.15	79.52	53.95
+ Ground P & R	25.71	29.43	90.56	72.99
- Dependency	23.32	25.53	75.67	38.49
- Documents	<u>23.06</u>	<u>25.34</u>	75.91	36.53
- Planning*	23.51	25.98	72.90	<u>24.89</u>
- Planning**	23.69	26.81	<u>71.60</u>	34.91

Table 7: Ablation study on the impact of different steps and modules in SAFARI.

- We are the first to augment the dialogue system to plan and incorporate multiple sources of knowledge into responses (e.g., **decide whether or not require knowledge, which source to call, and when to call**).
- We build a personalized knowledge-grounded dialogue dataset, KBP, where the responses are conditioned on multiple sources of knowledge with **dependency relationship**.
- We conduct lots of experiments and analysis on latest LLMs. More can be found in the paper.

Ablation Study

Experiments

Conclusions



Chinese Social Media
(PaperWeekly) Report

!!! The first author is looking for any potential visiting positions. Feel free to email him for any discussion and cooperation.