



Large Language Models in Chatbot Applications

Ondřej Dušek

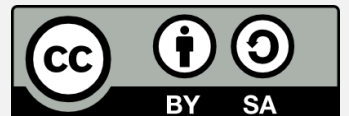
4EU+ AI Days

13.06.2024

Thanks: Vojtěch Hudeček




Charles University
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics




unless otherwise stated

LLMs are good at chat, right?

- Training scheme ~ be **convincing** but **not necessarily true**
- No access to external world by default
- Not reliable for QA: only uses information it memorized, “hallucinates”

 Who is Ondřej Dušek?

16, depending how you count

 Ondřej Dušek is a researcher and academic in the field of natural language processing (NLP). He is currently an Assistant Professor at the Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic.

Dušek's research focuses on developing computational models and algorithms for natural language generation, dialogue systems, and machine learning approaches for NLP tasks. He has published several research papers in top-tier NLP conferences and journals, and he has also served as a reviewer and program committee member for these conferences.

In addition to his research, Dušek is also actively involved in teaching and mentoring students in the field of NLP. He has supervised several Master's and PhD theses, and he teaches courses on topics such as computational linguistics, NLP, and machine learning.

no other tasks than generation & dialogue

no PhD graduate at the time of asking

I only teach dialogue systems

- **Assistant:** fulfill user requests (book a hotel/restaurant/taxi etc.)

- user:** I am looking for a train from Cambridge to London Kings Cross.
state: *{train {departure = cambridge, destination = london kings cross}}*
DB: *{train (70) {...}}*

	[count]	[departure]	[destination]
system:	There are 70	trains departing from Cambridge to London Kings Cross.	What day would you like to travel?
- user:** I would like to leave on Saturday after 18:45.
state: *{train {day = saturday, departure = cambridge, destination = london kings cross, leave at = 18:45}}*
DB: *{train (3) {arrive by = 19:51,21:51,23:51; id = TR0427,TR0925,TR4898; leave at = 19:00,21:00,23:00; ... }}*

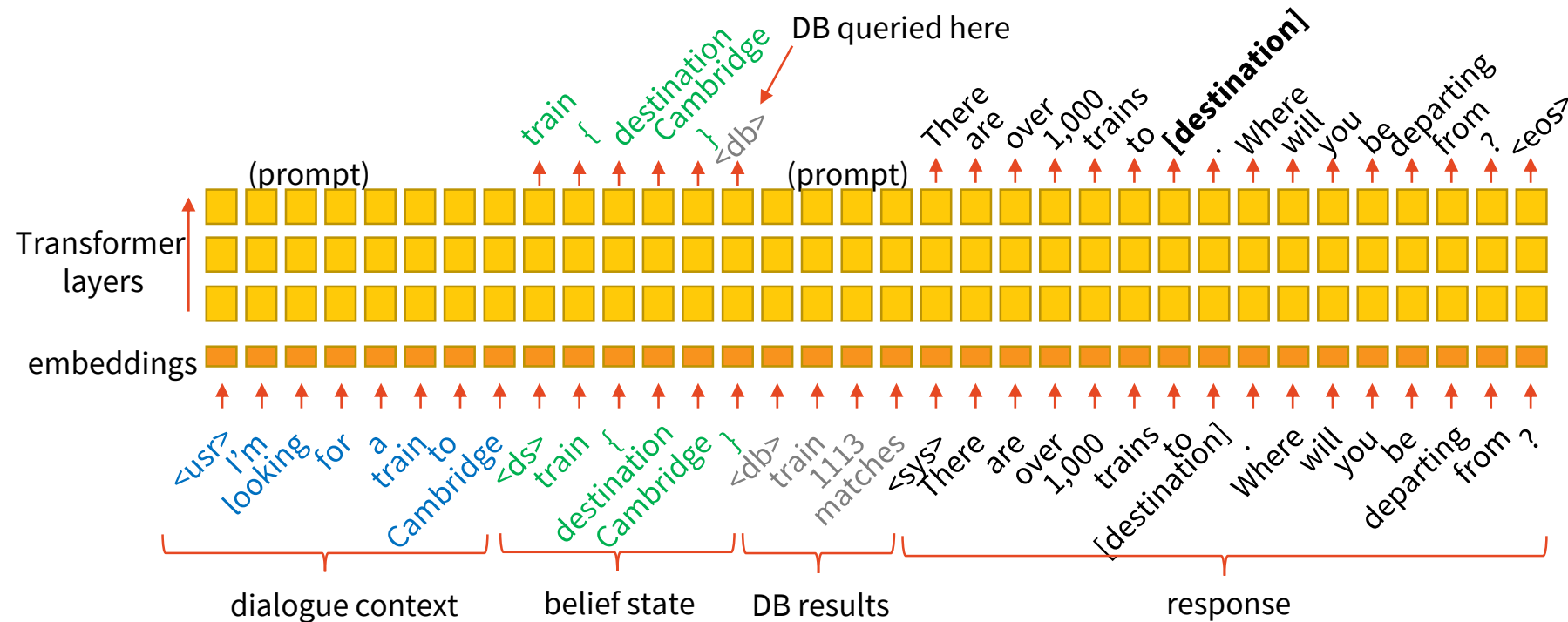
	[id]	[leave_at]	[arrive_by]
system:	TR0427	leaves at 19:00	and arrives by 19:51 . Would you like to book this train?
- user:** Yes, I would like to book it for eight people.

	[reference]
system:	I have booked it for you. Your reference number is 00000057. Is there anything else I can help you with?
- user:** I am also looking for an expensive restaurant in the centre.
belief: *{restaurant {area = centre, price range = expensive} train {...}}*
DB: *{restaurant (33) {area = centre (33); name=Curry Garden, ...; ...}, ...}*

	[count]	[price_range]	[area]
system:	There are 33	expensive restaurants in the centre.	Is there a particular type of food you would like?

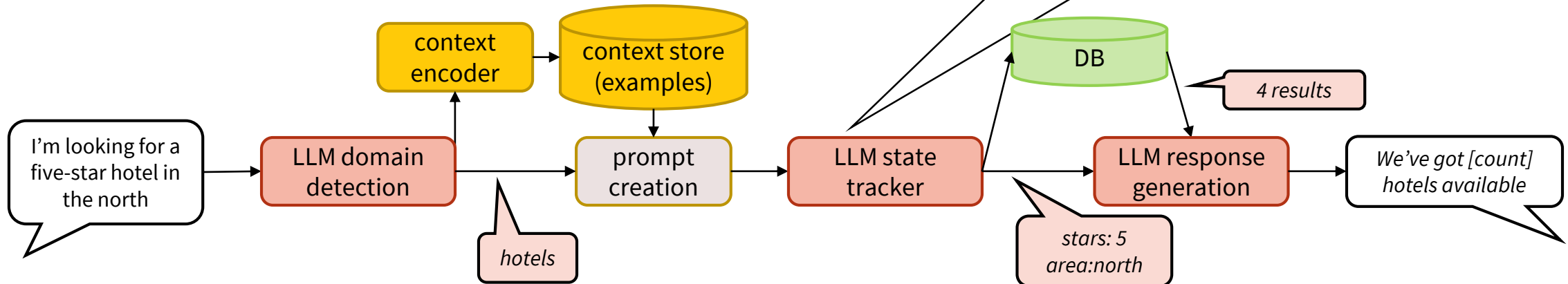
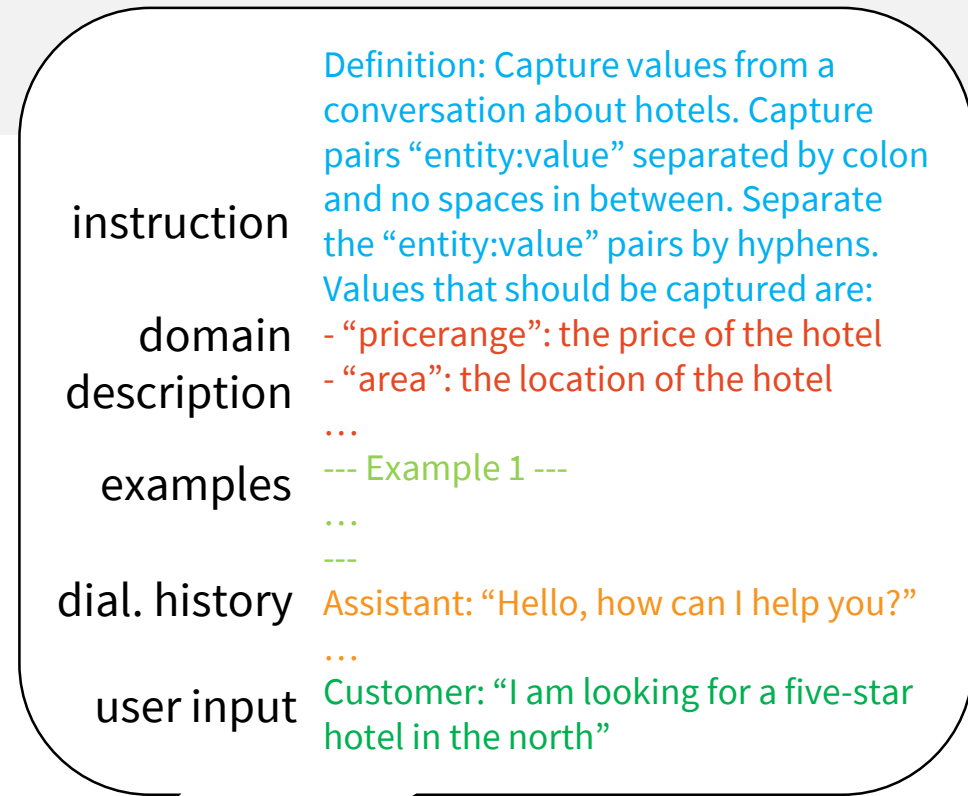
- Works with a database/external API (pre-set **fixed slots**)
- Needs **rules** or 1000s of **training dialogues** (previous neural models)

- SotA neural models: **belief state tracking** → DB query → **response**
 1. input prefix → generate belief state (diff)
 2. input DB results → generate delexicalized response



LLMs for Task-Oriented Dialogue

- Can we use the same with LLMs & prompting?
 - still: input → **state** → **DB** → **response**
 - additional 1st step: **domain detection**
 - tracking & response prompts domain specific
- “Zero” or few-shot setting
 - few-shot: FAISS context store, 10 ex./domain
 - little or no data needed: wide potential
- ChatGPT, Tk-Instruct, Alpaca, GPT-NeoX, OPT



Results

- Domain detection ~ 70%+
 - Alpaca & TklInstruct OK, ChatGPT almost perfect
 - good enough to get relevant examples & prompts
- Belief state – not great
 - much worse than SotA
 - examples help (ChatGPT, TklInstruct: ~50-60% F1, Alpaca 8%), 10 ex./domain enough

- Responses:
OKish

Dialogue Success		ChatGPT	TklInstruct
1-step (corpus)	predicted state	44%	19%
	gold state	68%	46%
expert eval (end-to-end, with recoveries)		76%	64%

- More potential with better prompt engineering

Thanks

Contacts:

Ondřej Dušek

odusek@ufal.mff.cuni.cz

<https://tuetschek.github.io>

@tuetschek

Links

These slides: <https://bit.ly/4euaidays-od>

Paper: <https://aclanthology.org/2023.sigdial-1.21>



Thanks:



Vojtěch Hudeček

