



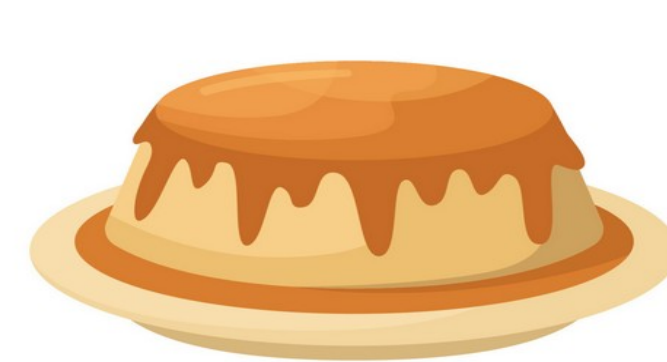
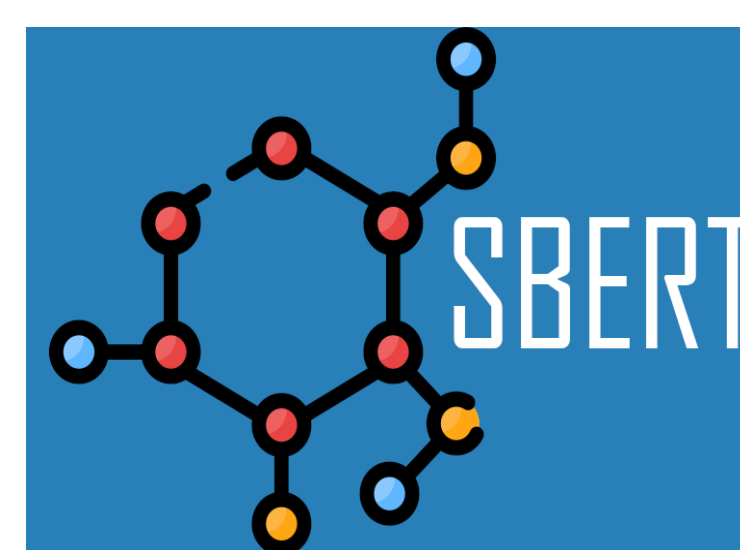
# Prompting, Retrieval, Training: An exploration of different approaches for task-oriented dialogue generation

Gonçalo Raposo, Luísa Coheur, and Bruno Martins

## “What’s the best approach for task-oriented dialogue generation?”

RETRIEVAL VS GENERATION

Models:



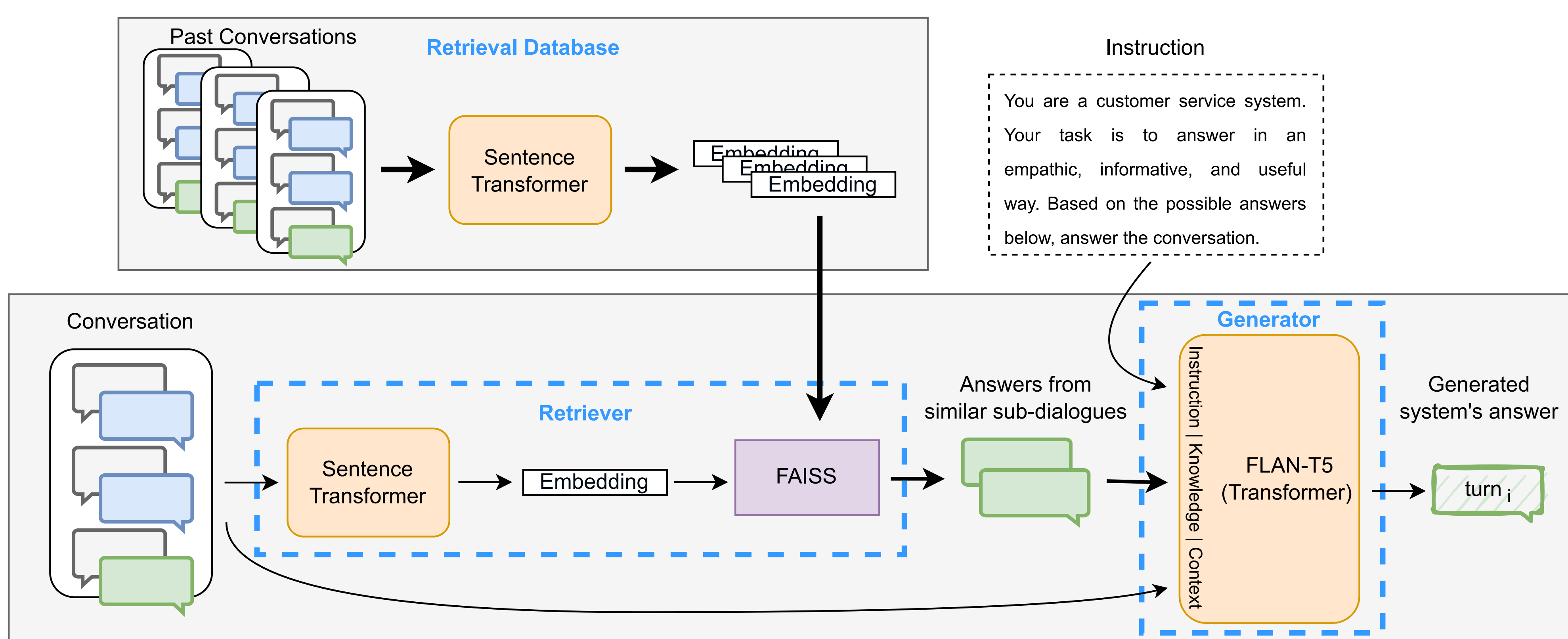
FLAN-T5

GPT-4

Datasets:

- MultiWOZ 2.2
- Taskmaster-2

- Zero-shot
- Few-shot
- Fine-tuning



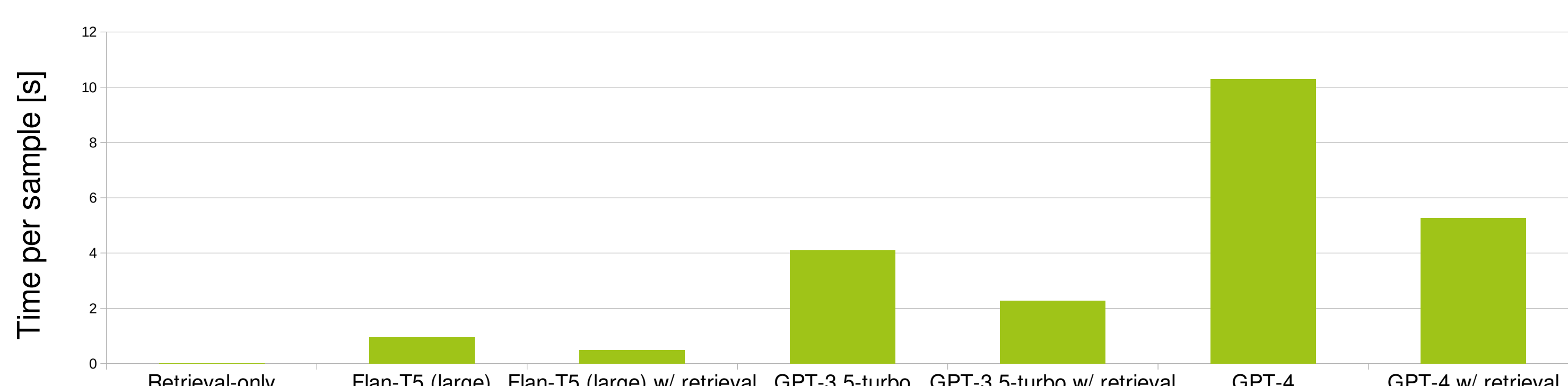
User: I'm needing a train leaving on wednesday and arriving by 11:00.

Reference: Sure, where are you leaving from and traveling to?

Prediction: Certainly. Where will you be departing from?

Table: Evaluation of different models and variations in the MultiWOZ dataset.

Model	Fine-tuned	ROUGEL-F1	BLEURT
Retrieval-only	-	0.1767	0.4022
Retrieval-only	retrieval	0.2622	0.4762
Flan-T5	-	0.1456	0.3840
Flan-T5 few-shot	retrieval	0.1804	0.4036
Flan-T5	generation	0.2795	0.4925
Flan-T5 few-shot	both	<b>0.2976</b>	<b>0.5033</b>
GPT-3.5-turbo	-	0.1761	0.4638
GPT-3.5-turbo few-shot	retrieval	0.2503	0.5009
GPT-4	-	0.1537	0.4581
GPT-4 few-shot	retrieval	0.2532	0.4868



## Conclusions:

- ✓ Few-shot prompting using retrieved possible answers works the best for task-oriented dialogues.
- ✓ If training data is available and it does not differ much from the test data, a fine-tuned generation model is better than GPT-3.5 and GPT-4.
- ✓ If the testing data differs from the training data, using a more general large language model and only fine-tuning the retrieval component is a better procedure.

### Acknowledgments:

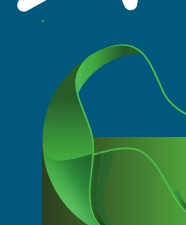
This research was supported by the Portuguese Recovery and Resilience Plan through the project C645008882-00000055 (Center for Responsible AI), and through Fundação para a Ciência e a Tecnologia (FCT), specifically through the P2020 program LISBOA-01-0247-FEDER-045909 (MAIA), and through the INESC-ID multi-annual funding with reference UIDB/50021/2020.



Center for Responsible AI



Fundação para a Ciência e a Tecnologia



PRR



REPUBLICA PORTUGUESA



Financiado pela União Europeia NextGenerationEU