



LLM4ALL Kickoff

WP3: LLMs for Dialogue
LINAGORA



Why focus on dialogue?

A lot of information is conveyed through conversational interaction:

- emergency calls
- business meetings
- debates
- collaborative tasks

Most (accessible) models are not trained on real dialogue or evaluated on dialogue-specific tasks

Dialogue features

Often open-ended

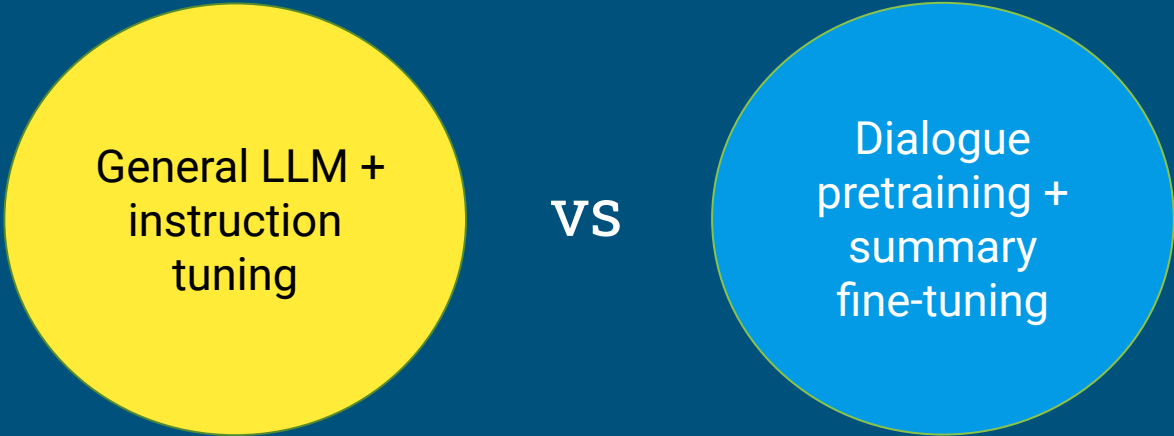
Oral-specific style/grammar/lexicon

Mistakes/revisions

Different perspectives (who said what)

Central question

What is the impact of dialogue-specific training on downstream dialogue understanding tasks?



General LLM +
instruction
tuning

VS

Dialogue
pretraining +
summary
fine-tuning

Step 1: Dialogue training

Dialogue generation but also:

- Passive comprehension
- Multi-party interactions

Turn-taking is important and abstraction from **noisy data**

[speaker001:] Ok, c'est quoi le plus, le souvenir, le premier souvenir, le plus clair que tu as en tête, de nous deux ?

[speaker002:] Le souvenir ?

[speaker001:] Ton premier souvenir de nous en fait.

[speaker002:] Le premier souvenir de nous le plus clair que j'ai, euh, je pense que c'est en Algérie. Je crois que c'est en Algérie. Et je sais pas, si tu te souviens. On était sur la place. Et, euh, et on courait. On jouait, comme des...

Task 3.1

M12

Dialogue data set (50M words already...)

First generation dialogue model(s)

M24

Augmented data set

Second generation dialogue model(s)

Open questions:

- Base model (decoder only?)
- Language mixture?
- More fine-grained genres?
- Data augmentation?

Step 2: Prompting extant models

How good are extant general purpose LLMs at dialogue understanding already?

- Prompting can replace for task-specific fine-tuning
- Summarization is notoriously subjective: prompting puts the task in the hands of the user
- Traditional summarization can be had with a special kind of prompt
- Prompting allows for interactive discussion
- Evaluation can focus on smaller chunks

Approach

Pull from extant summaries (ChatGPT?) to create questions

Les activités proposées sont toutes en accord avec un des thèmes. Par exemple, ils souhaitent brûler une effigie de Jeff Bezos, se raccordant avec le thème d'Amazon et de l'environnement. De plus, ils proposent une course de livraison en vélo, qui sensibilisera les livreurs à l'écologie. Un jeu de quille, thématique par les réseaux sociaux, et l'invitation d'un EHPAD, qui recevra des formations sur internet, avec en plus une initiation aux jeux vidéo. Les participants souhaitent aussi mettre en place un bar, et un food truck de soupe au pistou.

Quelles activités ont été proposées ?

Pourquoi veulent-ils proposer une course à vélo ?

Reformulate if necessary

Transcript: "Oui, c'est une réforme qui est anti-femme"



Quelle est une autre façon de poser la question suivante : "Est-ce que Mathilde Panot trouve que la réforme est anti-femme ?" ?



Vous pourriez poser la question de la manière suivante pour obtenir une variation :

"Quelle est l'opinion de Mathilde Panot sur la réforme en ce qui concerne son impact sur les femmes ?" ●

Step 3: Fine-tuning for summarization

Basic pretraining

vs

Dialogue training

Bigger model(s)

vs

Smaller model(s)

Prompting

vs

Fine-tuning

Tasks 3.2 & 3.3

M24: Evaluation methods

- Set of question/response pairs
- System in place for evaluation

M40: Summarization model(s)

- System in place for evaluation

Some concerns

1. Data: summaries for FreD but not for SUMM-RE
2. Evaluation: ROUGE and BERTScore, but not directly applicable to prompting and problematic in any case
3. Long contexts



Thank you!