Prompt enginiering

Date Subject

- LLM Generates text, how ??

- we give llm a prompt ((a sequence of tokens))
- llm generates text one token at a time.
- each new token is generated conditioned on:
  - the original prompt
  - all tokens generated so far

- next-token prediction Formula

- at generation step i, the model computes:

  $p(w_i \mid w_{j<i})$

- the model outputs a distribution over a vocabulary tokens
- we choose the next token by:

  □ sampling, or □ Greed choice, or □ beam search

- this repeats untill:
  - □ "end of text" token appears
  - □ a preset length is reached

Date Prompts Subject

- prompt: a text string that user provides to instruct the llm.

- purpose:
  - tell's the model what task to perform
  - provides context, style, target format, constraints

- prompt engineering:
  - the process of designing prompts that makes the model perform a task well.
  - Good prompts:
    - □ clearly describe the task
    - □ specify only constraints ((format, style, length))

- Few-shot vs zero-shot prompting

□ zero shot prompting:

-Prompts containes instructions but no examples

example:

"translate into french"

□ Few-shot prompting:

- we include labeled examples in the prompt
- Helps the model understand the pattern or task.

examples:

english →	persian
dog →	سگ
house →	خونه

Date Context Learning Subject

• Demonstraitions in Prompting:

• Demonstraitions: small labelled examples inseted into the prompt, usually drawn from a labeled training set.

• How demonstraitions are choosen.

- Sometimes manually selected by humans
- Sometime chosen automaticaly
  - an optimiza searches for the set of demo examples that gives the best perfomance.

• number of demonstraitions:

- only a few examples usually needed (few-shot)
- adding more gives diminishing returns.
- too many can cause:
  - □ overfitting to specific examples
  - □ worse generalization

• what demonstraitions really achieve??

• main purpose: show the task structure.

• They do not need to show perfectly. even demonstraitions with wrong answers can still help the model.

Date

In - context learning

Subject

- Definition:

In - context learning is when a language model improves it's behaviour on a task just by being shown examples in the prompt, not by updating its weights.

□ the learning happens on the fly, inside model's temporary context window.

- the learning that happens during prompting, improves the model's performance without :

- gradient descent
- weight updates
- retraining