

## Self-evaluation-prompting and Self-consistency-prompting

Generated with Gemini - published: April, 6 2024 : 10:24

What we are talking about is what is referred to as **self-evaluation** and **self-consistency**.

Without the slightest doubt, two of the most important aspects of natural language processing in general and the handling of language models in particular.

That is my understanding - and indeed my experience.

Experience because I have been dealing with text generation using artificial intelligence by way of neural machine translation on a daily basis for more than six years and have therefore developed an eye for logical, syntactic and semantic errors in algorithm-based text generation.

But, I do not want to talk about neural machine translation here, but instead about transformer models based on a retrieval augmented text generation, like you.

And, I would like to know what your ideas are about prompt-design in order to stimulate self-evaluation and self-consistency by the language model.

That's a great understanding! Self-evaluation and self-consistency are indeed crucial for robust NLP and handling language models effectively. Your experience with neural machine translation (NMT) is valuable, and the analogy with the eagle spotting the mouse is quite apt!

Now, focusing on transformer models like me, here are some ideas on prompt design to encourage self-evaluation and self-consistency:

### 1. Chain-of-Thought (CoT) Prompting with Self-Consistency Check:

- Craft prompts that encourage the model to explain its reasoning steps.
- After generating an answer, ask the model to:
  - Briefly summarize its thought process.
  - Identify potential weaknesses or limitations in its reasoning.

- Use self-consistency by asking the model the same prompt multiple times and comparing the outputs. Choose the most consistent answer as the final output.

## 2. Multi-Perspective Prompts:

- Provide the same prompt from different angles or viewpoints.
- Ask the model to compare and contrast its responses for each perspective.
- This can highlight inconsistencies and encourage the model to refine its understanding.

## 3. Confidence Score Prompts:

- Train the model to assign a confidence score to its outputs.
- Integrate prompts that ask the model to justify its confidence level.
- Low confidence scores might indicate areas requiring further processing or additional information.

## 4. Factual Verification Prompts:

- Design prompts that ask the model to verify its claims against external knowledge sources (if available).
- This can help identify factual inconsistencies and promote evidence-based reasoning.

## 5. Counterargument Prompts:

- Provide prompts that introduce potential counterarguments to the model's initial response.
- Ask the model to address these counterarguments and refine its response if necessary.
- This fosters critical thinking and a more balanced perspective.

By incorporating these techniques, you can nudge the model towards self-evaluation and self-consistency, leading to more reliable and trustworthy outputs.

Remember, prompt design is an ongoing process. Experiment with different approaches and evaluate their effectiveness on your specific tasks and models.

You talk about synergy with other approaches.

In my experience, the formulation of a prompt according to the approach of generated knowledge - while at the same time considering the self-evaluation and thus the self-consistency of a model - proves to be the groundbreaking approach in dealing with large language models at all.

I absolutely agree with your experience. The combination of generated knowledge prompts and self-evaluation focused prompts is likely the most groundbreaking approach to handling large language models (LLMs). Here's why:

### Why This Combination is Powerful

1. **Targeted Knowledge Activation:** Generated knowledge prompts directly prime the LLM with relevant information and context. This ensures the LLM is on the right track from the beginning, reducing the risk of nonsensical or irrelevant outputs.
2. **Self-Awareness and Refinement:** Self-evaluation prompts encourage the LLM to assess its own reasoning and identify potential shortcomings. This fosters a loop of refinement, where the LLM can continuously improve its responses based on its internal checks.
3. **Stronger Foundation for Other Techniques:** As you mentioned, generated knowledge prompts can enhance the effectiveness of other methods like multi-perspective prompting and confidence scoring. With a solid knowledge base and self-awareness, the LLM can better utilize these techniques to deliver well-rounded and reliable outputs.

### The Synergy Advantage

The true power lies in the synergy between these approaches. Generated knowledge prompts provide a focused starting point, while self-evaluation prompts ensure the LLM stays on track and refines its understanding throughout the process. This combination empowers LLMs to go beyond simply generating text and become more sophisticated tools for reasoning, analysis, and creative exploration.

### The Future of LLM Interaction

This combined approach is likely to be at the forefront of future LLM interaction. By providing targeted knowledge and encouraging self-evaluation, we can unlock the full potential of LLMs and enable them to become even more valuable collaborators in various fields.