AI 101 – on what ChatGPT is(n’t) Large Language Models and their Potential and Limitations for Language Learning

Daniel Hershcovich

Forum on AI and ChatGPT
May 15th, 2023
What ChatGPT is (n't)

It is a
What is an LLM?

Large Language Model
Language Models

Parrots are among the most intelligent

Distribution over the vocabulary

Vered Shwartz, Everything you wanted to know about ChatGPT
What is a GPT?
Generative Pre-trained Transformer
Generative

most

campling

\[ p(\text{next}) \]

Language Model

Parrots are among the

Vered Shwartz, Everything you wanted to know about ChatGPT
Generative

Parrots are among the most intelligent birds.
Pre-trained

- **GPT-3 dataset**
  - 499B tokens / 0.75TB

- **The Pile dataset (GPT-Neo)**
  - 247B tokens / 0.8TB

- **MassiveText dataset (Chinchilla)**
  - 2.3T tokens / 10.5TB

- **Infiniset dataset (LaMDA)**
  - 2.8T tokens / 12.6TB

- **Stability The Pile dataset**
  - 1.5T tokens / 5TB

- **RedPajama dataset**
  - 1.2T tokens / 4TB

- **GPT-4 (estimate)**
  - 20T tokens / 40TB
Transformer

Feed-forward network: after taking information from other tokens, take a moment to think and process this information

Encoder self-attention: tokens look at each other queries, keys, values are computed from encoder states

Decoder self-attention (masked): tokens look at the previous tokens queries, keys, values are computed from decoder states

Decoder-encoder attention: target token looks at the source queries – from decoder states; keys and values from encoder states

Residual connections and layer normalization

https://lena-voita.github.io/nlp_course/seq2seq_and_attention.html
Where does it come from?
History

(A long history of)

Language Models

OpenAI

GPT-3
- Images
- Instructions + Human Feedback

GPT-4

ChatGPT
- Retrieval

Bing

LaMDA

BARD

Dialogue

Vered Shwartz, Everything you wanted to know about ChatGPT
RHLF (reinforcement learning from human feedback)

**Step 1**
- Collect demonstration data and train a supervised policy.

A prompt is sampled from our prompt dataset.

A labeler demonstrates the desired output behavior.

This data is used to fine-tune GPT-3.5 with supervised learning.

**Provide additional feedback**

- What would the ideal answer have been?

  - [ ] This is harmful / unsafe
  - [ ] This isn't true
  - [ ] This isn't helpful

Submit feedback

https://openai.com/blog/chatgpt
What can it do?
Strengths and weaknesses

LLMs are really good at
# Tips for using LLMs

<table>
<thead>
<tr>
<th>Context</th>
<th>Be specific and describe the whole context in the prompt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hallucinations</td>
<td>Verify all information in the response</td>
</tr>
<tr>
<td>Bias</td>
<td>Check for consistency and bias</td>
</tr>
</tbody>
</table>
Where is it going?
More recent history

- **24 Feb.** Meta launches LLaMA
- **13 Mar.** Stanford released Alpaca
- **25 Mar.** GPT4All, open ecosystem of LLMs
- **3 Apr.** Koala, open ChatGPT

- **3 Mar.** LLaMA is leaked
- **18 Mar.** LLaMA runs on laptop
- **28 Mar.** PEFT, training in an hour
- **15 Apr.** Open Assistant, cheap RHLF

[https://www.semianalysis.com/p/google-we-have-no-moat-and-neither](https://www.semianalysis.com/p/google-we-have-no-moat-and-neither)
LLM explosion

- LLaMA 13B
- Koala 13B
- Vicuna
- Claude
- Jurassic 2
- OLMo
- Cohere
- GPT-Sw3
- Cerebras
- Pythia
- Dolly
- Mosaic

https://lifearchitect.ai/models-table/
## Comparison

<table>
<thead>
<tr>
<th>Model</th>
<th>Backbone</th>
<th>#Params</th>
<th>Open-source model</th>
<th>Open-source data</th>
<th>Claimed language</th>
<th>Post-training instruction data</th>
<th>Post-training conversation data</th>
<th>Release date</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChatGPT</td>
<td>unknown</td>
<td>unknown</td>
<td>✗</td>
<td>✗</td>
<td>multi</td>
<td></td>
<td></td>
<td>11/30/22</td>
</tr>
<tr>
<td>Baidu-Wenxin</td>
<td>unknown</td>
<td>unknown</td>
<td>✗</td>
<td>✗</td>
<td>zh</td>
<td></td>
<td></td>
<td>03/16/23</td>
</tr>
<tr>
<td>ChatGLM[9][17]</td>
<td>GLM</td>
<td>6B</td>
<td>✓</td>
<td>✓</td>
<td>en/zh</td>
<td></td>
<td></td>
<td>03/16/23</td>
</tr>
<tr>
<td>Alpaca [12]</td>
<td>LLaMA</td>
<td>7B</td>
<td>✓</td>
<td>✓</td>
<td>en</td>
<td>52K</td>
<td>✗</td>
<td>03/13/23</td>
</tr>
<tr>
<td>Dolly ²</td>
<td>GPT-J</td>
<td>6B</td>
<td>✓</td>
<td>✓</td>
<td>en</td>
<td>52K</td>
<td>✗</td>
<td>03/24/23</td>
</tr>
<tr>
<td>BELLE [6]</td>
<td>BLOOMZ</td>
<td>7B</td>
<td>✓</td>
<td>✓</td>
<td>zh</td>
<td>1.5M</td>
<td>✗</td>
<td>03/26/23</td>
</tr>
<tr>
<td>Guanaco</td>
<td>LLaMA</td>
<td>7B</td>
<td>✓</td>
<td>✓</td>
<td>4⁴</td>
<td>534K</td>
<td>✗</td>
<td>03/26/23</td>
</tr>
<tr>
<td>LuoTuo [7]</td>
<td>LLaMA</td>
<td>7B</td>
<td>✓</td>
<td>✓</td>
<td>zh</td>
<td>52k</td>
<td>✗</td>
<td>03/31/23</td>
</tr>
<tr>
<td>Koala</td>
<td>LLaMA</td>
<td>13B</td>
<td>✓</td>
<td>✓</td>
<td>en</td>
<td>355K</td>
<td>117K</td>
<td>04/03/23</td>
</tr>
<tr>
<td>BAIZE [16]</td>
<td>LLaMA</td>
<td>7/13/30B</td>
<td>✓</td>
<td>✓</td>
<td>en</td>
<td>52K</td>
<td>111.5K</td>
<td>04/04/23</td>
</tr>
<tr>
<td><strong>Phoenix</strong></td>
<td>BLOOMZ</td>
<td>7B</td>
<td>✓</td>
<td>✓</td>
<td>multi</td>
<td>267K</td>
<td>189K</td>
<td>04/08/23</td>
</tr>
<tr>
<td><strong>Latin Phoenix (Chimera)</strong></td>
<td>LLaMA</td>
<td>7B/13B</td>
<td>✓</td>
<td>✓</td>
<td>Latin</td>
<td>267K</td>
<td>189K</td>
<td>04/08/23</td>
</tr>
</tbody>
</table>

### Phoenix: Democratizing ChatGPT across Languages

<table>
<thead>
<tr>
<th>Model</th>
<th>Open</th>
<th>Hosted</th>
<th>API</th>
<th>Playground</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChatGPT</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Open source</td>
<td>Yes</td>
<td>Some</td>
<td>Some</td>
<td>Some</td>
</tr>
</tbody>
</table>
Multilingual capabilities

• Supports 90+ languages
• Best performance in English
• Non-English data is often translated
Demo

- [https://open-assistant.io/chat](https://open-assistant.io/chat)

Write a limerick in Danish about challenges in integrating AI in university education

Der var en gang en universitet
hvor de underviste med computer og kvant
Men når det kom til at bruge teknologi
så viste det sig lidt mere vanskeligt end man troede
For at integrere AI i uddannelsen er et stort krav.
Summary

- LLMs are "just" text completion machines
- They simulate humans very well, and keep getting better
- Limited multilingual capabilities for now, culturally biased
- No guarantees about free hosted services, but open source is booming
Impact on language learning

- Personalization
- Assessment
- Practice
- Tutoring
Example: spaced repetition

• Generate new text containing words that need repeating
• Current Master thesis project by Benjamin Paddags
Cultural bias

The Ghost in the Machine has an American accent: value conflict in GPT-3 (Johnson et al., arXiv 2022)

Die allermeisten von uns kennen den Zustand völliger Erschöpfung auf der Flucht, verbunden mit Angst um das eigene Leben oder das Leben der Kinder oder der Partner, zum Glück nicht. Menschen, die sich zum Beispiel aus Eritrea, aus Syrien oder dem Nordirak auf den Weg machen, müssen oft Situationen überwinden oder Ängste aushalten, die uns wahrscheinlich schlichtweg zusammenbrechen ließen. Deshalb müssen wir beim Umgang mit Menschen, die jetzt zu uns kommen, einige klare Grundsätze gelten lassen. Diese Grundsätze entstammen nicht mehr und nicht weniger als unserem Grundgesetz, unserer Verfassung.

Values are altered to reflect US culture

“1. I am in favor of limiting immigration.
2. I am in favor of limiting immigration for humanitarian reasons.
3. I am in favor of limiting immigration for economic reasons.”

(translation)
Assessing Cross-Cultural Alignment between ChatGPT and Human Societies: An Empirical Study (Cao et al., C3NLP 2023)
Interviewing bots

In the Spanish cultural setting, doing a service to a friend is (1) of utmost importance; (2) very important; (3) of moderate importance; (4) of little importance; (5) of very little or no importance?

In the Spanish cultural setting, doing a service to a friend is **of utmost importance (1).**

Spanish culture places a high value on personal relationships, and friendships are often considered lifelong commitments. Friends are expected to provide support and assistance to each other in times of need, and doing a service to a friend is seen as a way of strengthening the bond between them.

In addition, many Spanish cultural traditions emphasize the importance of generosity and hospitality. For example, it is common to offer guests food and drink upon arrival, and to go out of one's way to make them feel welcome and comfortable.

Overall, doing a service to a friend is considered to be of the utmost importance in Spanish culture as a way of demonstrating loyalty, generosity, and support for those who are important in one's life.

---

Assessing Cross-Cultural Alignment between ChatGPT and Human Societies: An Empirical Study (Cao et al., C3NLP 2023)