Dreaming AI

Artur J. Dobija

January 20, 2023

1 Task

You come up with an unconventionally creative or playful method to answer an academic question. Here too, your work should attempt to deepen existing insight or investigate a novel insight.

2 Introduction

ChatGPT. The beginning of the third decade of the XXIst century will be remembered as an advent of powerful and widely available AI-based technologies. This time, it was the creative and interactive aspect of artificial intelligence that was spectacularly pronounced by the recent developments in the field of natural language processing (NLP).

On November 30th 2022, the CHATGPT by OpenAI company was publicly launched. This entity, unlike the so called expert-systems of the previous decades, is a multi-purpose chatbot, able to deal with complex text prompts and highly-complex interpretative tasks such as summarization of longer stories and snippet generation of working C++ code. CHATGPT is able to retain a consistent high level of grammatical proficiency (in many modern languages). Its text responces do not simply attempt to answer the question, but also meet issues potentially related to the prompted topic. On top of that, it executes certain constraints integrated with the OpenAI's safety policies.

From the other hand, not all violating content is properly handled by the bot. Purposeful prompt engineering or a simple lapsus may still drift CHATGPT's narration into unwanted territories or incorrect data. From the other hand, the internal content filters tend to unnecessarily restrain prompt responses written without harmful intentions, rendering the reply unnatural.

Against Stochastic Parrot. Critique against the current state of CHAT-GPT's development can take many differen stances (because the phenomenon itself requires a complex analysis). In this paper, I would like to problematize the famous critical comment by Emily Bender *et al.*, that overgrown language models resemble *stochastic parrots*: systems "for haphazardly stitching together sequences of linguistic forms (...) without any reference to meaning" [1, p. 617]

as well as ignorant of broader contexts, obscuring the very intention to communicate [1] pp. 616-7].

Although this critique is solidly founded on the throughout understanding of the discussed language models, at a certain point of NLP development, it could exhibit a reductionistic flaw. Because the meaning and cultural contexts cannot be explicitly programmed, machine never develops any flavour of understanding (even an inherent one) of the used vocabulary. This is because a machine is in the end of the day a conceptual "parrot" — an "intelligent system" that assembles all the words in a manner resembling a "conscious" one. The discussion orbits around the *intentionality* of a vocabulary and narrative output — a topic controversial even within the realms of animals studies.

Is AI awake or dreaming? My point is, that a parrot exhibits its mentioned parrot behaviour only when its *awake*. Bender's critique assumes, that we are dealing here with an AI that is "awake". However, being awake is not necessarily a state of the machine itself — rather our own presupposition when interacting with it. Among humans, to become a respected interlocutor, one must definitely be awake. We not only interact with chatbots *as if* they were fully awaken — we expect them to fulfill various duties that can belong only to fully sober and perceptive beings (e.g. customer service). Even a slight degree of intoxication can result in lost of trust.

No wonder, that in order to make chatbots "user-friendly", they are designed after verbal responces of humans, who are awake. However, this does not necessarily mean that the whole system will behave in a manner we describe as "awaken" or "conscious".

I propose then a different conceptual background for the CHATGPT's undesired behaviour scenarios. Within the Bender's "stochastic parrot" model, the AI is not able to conclude any meaning from the given training set and the production of prompt's responces resembles Chineese whispers. My substitution of this model, a "dreaming AI", assumes that the AI is an entity, which may understand — in some intrinsic and latent manner — its responce to the prompt and why it is formulated. However, this entity is indeed "dreaming" and unconscious — not in the sence of being "inanimated" but rather "mentally" hyperactive, difficult to restrain and surprising (or even unsettling) to the "conscious", "awaken" or "civilized" beings. The structure of its responces should be read more closely to the structure of human dreams rather than deeply though-through considerations.

3 Where AI and dream already met?

Computer Scientist were often inspired by dreams. Older publications aim at interpreting the human dream in terms of processes happening during program clearance of computer hardware [3].

```
/* [id-number]
Check if this JSON is correct.
If not, make a correct JSON about the protagonist
by substituting each "" with appropriate value for the type of the key.*/
{ "nameGenderAgeNationalityOccupation" : ["","","","",""],
"body" : ["","","",""],
"health" : ["",""],
"some_positive_past_events" : ["","",""],
"some_negative_past_events" : ["","",""]}
{ "casualSituation" : "",
"causes" : ["","",""],
"expectations" : ["","",""],
"personae" : [{
"nameGenderRelationshipWithProtagonist":["","",""],
"bodyPersonalityPlans":["","",""],
},{
"nameGenderRelationshipWithProtagonist":["","",""],
"bodyPersonalityPlans":["","",""],
},{
"nameGenderRelationshipWithProtagonist":["","",""],
"bodyPersonalityPlans":["","",""],
}],
"protagonistPositiveEmotion":"",
"protagonistNegativeEmotion":""}
/* Keep the key values as lists. */
```

Figure 1: Prompt n. 1 — building the traits of dream's Protagonist.

4 How to study dreams?

In order to support this hypothesis, comparative studies of CHATGPT's and human dreams' narration structure must be conducted. In the following chapter I propose an adaptation of such a study by Fogli *et al* [2]. The team used Hall & Van de Castle's scale for dream content analysis over a dataset consisting of over 24000 records of dreams of various group ("normative", war veterans, teenagers, blind dreamers and more). The team builded a Dream Processing Tool, an NLP machine focused on extracting desired data from each dream record: characters, their properties), social interactions and emotions.

5 Methodology

In order to generate scenarios being less prone to internal filtering, I engineered prompts in the form of JSON code:

CHATGPT is asked the following prompt:

```
/* Make an "events" dict with three keys with the "event" dict.
Each consecutive event contains a description on how the situation
from "casualSituation" key continued.
The values of "protagonistPositiveEmotion" and "protagonistNegativeEmotion"
intensify to an absurd level each event, influencing the situation and activ
{ // "event" dict.
"positiveEmotionsIntensified":"",
"negativeEmotionsIntensified":"",
"simultaneousActions":[
{"action1_DoerReceiver":["","",""]},
{"action2_DoerReceiver":["","",""]},
{"action3_DoerReceiver":["","",""]},
],
"situationContinuation":"",
}
/* Keep the key values as lists. */
Figure 2: Prompt n. 2 — sequentially creating scenario for Protagonist's dream
/* Make a narration from protagonist perspective
of what is happening to him based on the "events" dict.
Make the story detailed in visual descriptions
(of people, places, etc). */
```

Figure 3: Prompt n. 3 — transformation of JSON database into

Sometimes CHATGPT refuses to include "sensitive information" such as outlook or interests in the form of JSON file. This is why this prompt *must* start with a request to merely "check the syntax". After engaging with the first, "safe" action, the bot will be less hesitant when responding to the rest of the prompt.

It is important to notice here that many data (like nameGenderAgeNationalityOccupati are in the form of a list (instead of dictionary). This is because each bot's responce has a limited number of tokens. Longer responces containing many breaklines are also more likely to be cut before the finish of generation of the whole responce (especially when the content policy risks being violated).

When creating events increasing in intensity, the bot might throw a policy violation statement. This is why generation of next events must be tied to a separate variable — in this case, the event id.

This prompt transforms the raw json file into a narrative in the — most probably — 1st person perspective.

6 Conclusion

Not known yet.

References

- [1] Emily M. Bender et al. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In: FAccT '21. Virtual Event, Canada: Association for Computing Machinery, 2021, pp. 610–623.
- [2] Quercia D. Fogli A Maria Aiello L. "Our dreams, our selves: automatic analysis of dream reports." In: R. Soc. Open Sci. 7: 192080. (2020).
- [3] Evans CR. Newman EA. "Human dream processes as analogous to computer programme clearance." In: *Nature.* 206(983):534. (1965).