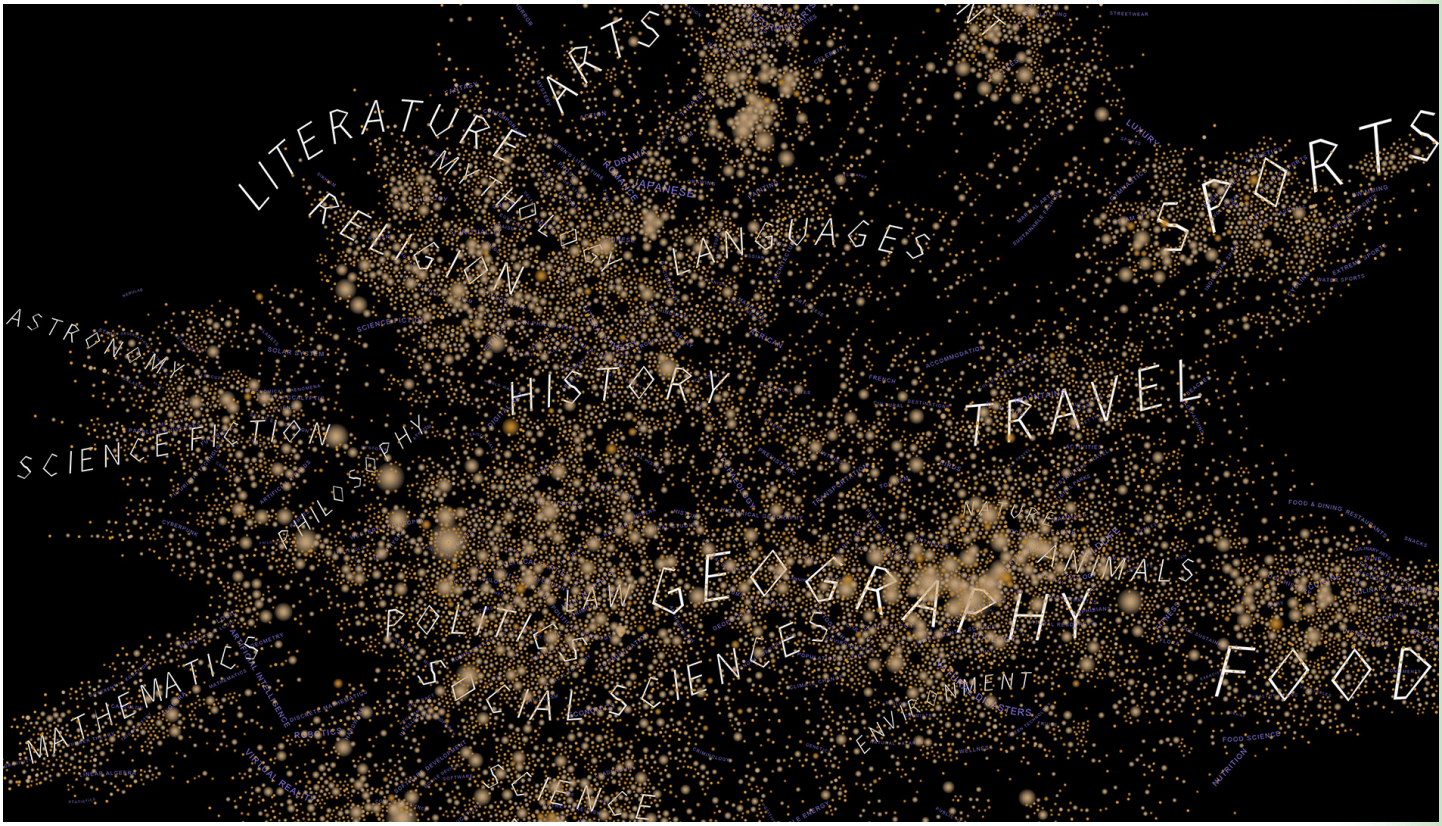
Artificial Worldviews



Kim Frederic Albrecht
k.albrecht@filmuniversitaet.de
Konrad Wolf Film University of Babelsberg, Potsdam, Germany,
metaLAB (at) Harvard & Berlin

How will “prompting” alter our perception? What types of aesthetics will large language models bring to the world? In what ways will technologies like ChatGPT affect notions, principles, and styles for the coming decade? *Artificial Worldviews* is a series of inquiries into the system underlying ChatGPT about its descriptions of the world. Utilizing prompting, data gathering, and mapping, this project investigates the data frames of “artificial intelligence” systems.

Introduction

Artificial intelligence and machine learning methods are often referred to as black boxes, indicating that the user cannot understand their inner workings. However, this trait is shared by all living beings: we come to know a person not by examining their brain structures but by conversing with them. The so-called black box is not impenetrable since we can gain an understanding of its inner workings by interacting with it. Through individual inquiries, we can only acquire anecdotal evidence of the network. However, by systematically querying chatGPT’s underlying programming interface, we can map the structures of the system.

In my research, I methodically request data about large-scale, indefinable human concepts and visualize the results. These outputs visualize expansive data structures and unusual, sometimes unsettling worldviews that would otherwise be unimaginable. The terms “power”

Keywords Prompting Poesis, Data Visualization, ChatGPT, AI, Evidentiary Realism, Investigative Design.
DOI 10.34626/2024_xcoax_023

and “knowledge” unfold vast discourses from philosophy, politics, social sciences to natural sciences; they hold multidimensional meanings within social relations. The resulting graphics resemble narratives found in the works of Franz Kafka or Jorge Luis Borges, representing an infinite library of relational classifications, bureaucratic structures, and capricious mechanisms of inclusion and exclusion.

Data

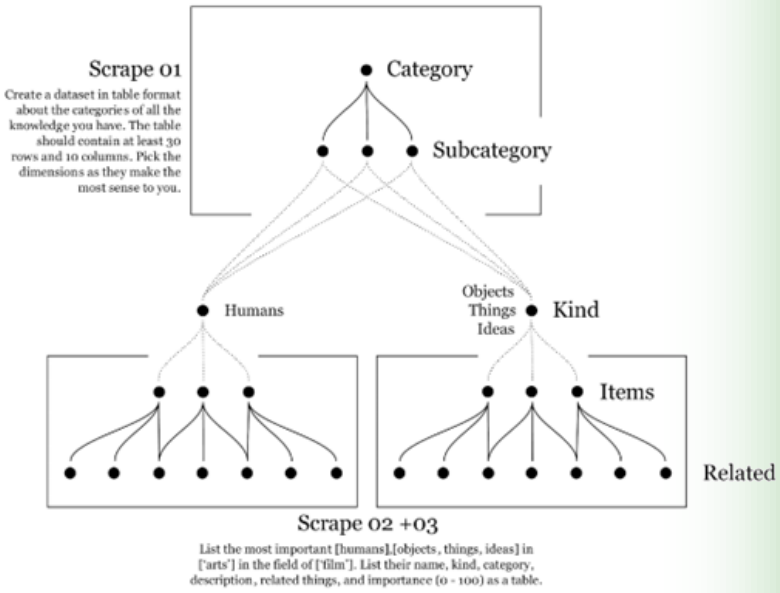
The Initial Dataset

The OpenAI Application Programming Interface (API) structures calls into two messages: the user message and the system message (OpenAI 2024). While the user message is similar to the text you enter into the front end of ChatGPT, the system message helps set the behavior of the assistant. For the project, I designed the following system message: *You are ChatGPT, a mighty Large Language Model that holds knowledge about everything in the world and was trained on a massive corpus of text data, around 570GB of datasets, including web pages, books, and other sources.*

The initial user message was the following: *Create a dataset in table format about the categories of all the knowledge you have. The table should contain at least 30 rows and 10 columns. Pick the dimensions as they make the most sense to you.*

I called these requests six times with six different temperatures: 0, 0.2, 0.4, 0.6, 0.8, and 1. The temperature, ranging between 0 and 1, determines the randomness of the responses. The higher the temperature of the request, the more the results vary. The resulting data file from the six API calls consisted of 31 fields and 425 subfields.

Fig. 1. Core Dataset Scraping Diagram.



The Core Dataset

The core dataset was requested from the OpenAI API in 1764 requests over the span of three days. Humans and objects were requested separately in all fields and subfields (425). Each of the 850 calls was made

twice: once with a temperature of 0 and once with a temperature of 0.5. All requests in the visualization were made to the model “gpt-3.5-turbo.” The number of returned items per request varied between five (‘Linguistics’ and ‘Travel Budget’) and 40 (‘Mythology’) returned rows of data. Due to this inconsistency, some fields hold more items than others. The user message was always the same: List the most important humans in ‘Arts’ in the field of ‘Film’. *List their name, kind, category, description, related things, and importance (0 - 100) as a table.* Replacing the field ‘Arts’ and the subfield ‘Film’ with one of the 425 combinations of fields and subfields.

Fig. 2. All Categories, Subcategories and Item Datapoints.

Visualization

Layers

The map consists of four layers. The first two layers are the fields (31) and subfields (425) in which the Generative Pre-Trained Transformer (GPT) categorized its knowledge. The third layer consists of 7,880 items representing the core dataset of the project. The fourth layer consists of 24,416 items, including people, objects, places, etc., that GPT-3.5 named in relation to the core items of the third layer.

Positioning

The visualization is a calculation of network similarities. Fields connect to subfields, and objects and humans connect by co-mentions in multiple fields. Thus, in the resulting map, objects and humans cluster together by similarity.

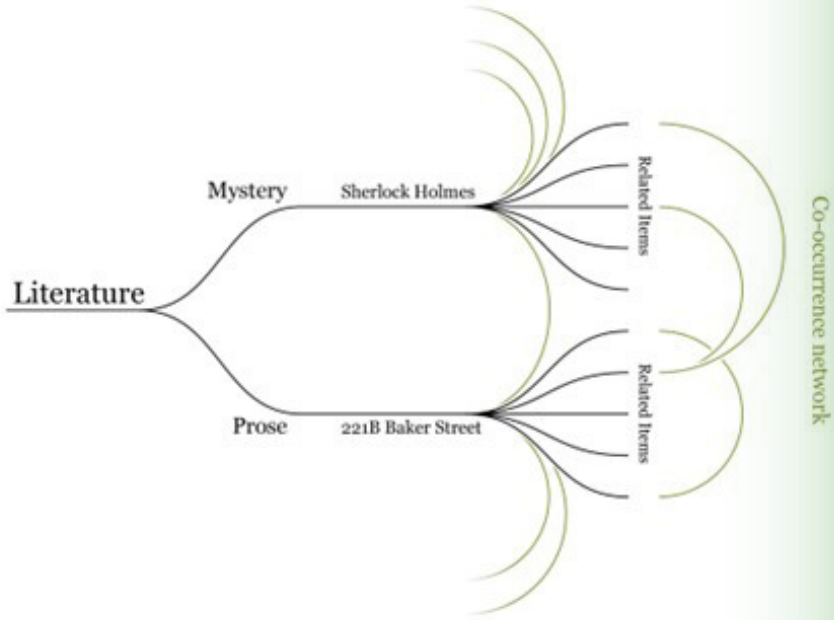


Fig. 3. Network Diagram Schematic.

Preliminary Findings

What Is This Map?

To understand the meaning of the map, it is essential to understand the forces and restrictions guiding GPT-3.5. Large Language Models (LLMs) are bound by at least three forces: the technical infrastructures of computation, the training data, and the post-training moderation. Training a model such as GPT-3 takes massive amounts of hardware, resources, and energy. Far from being artificial, these models are trained with rare earth elements and consume vast amounts of energy during training. OpenAI keeps the exact configurations secret, but it is known that the main driver of the system's computation is the NVIDIA V100 Tensor Core graphics processing unit (GPU), as well as a Microsoft high-bandwidth cluster (Brown et al. 2020). Estimations suggest that training GPT-3 has cost at least $4.6 million (Li 2020).

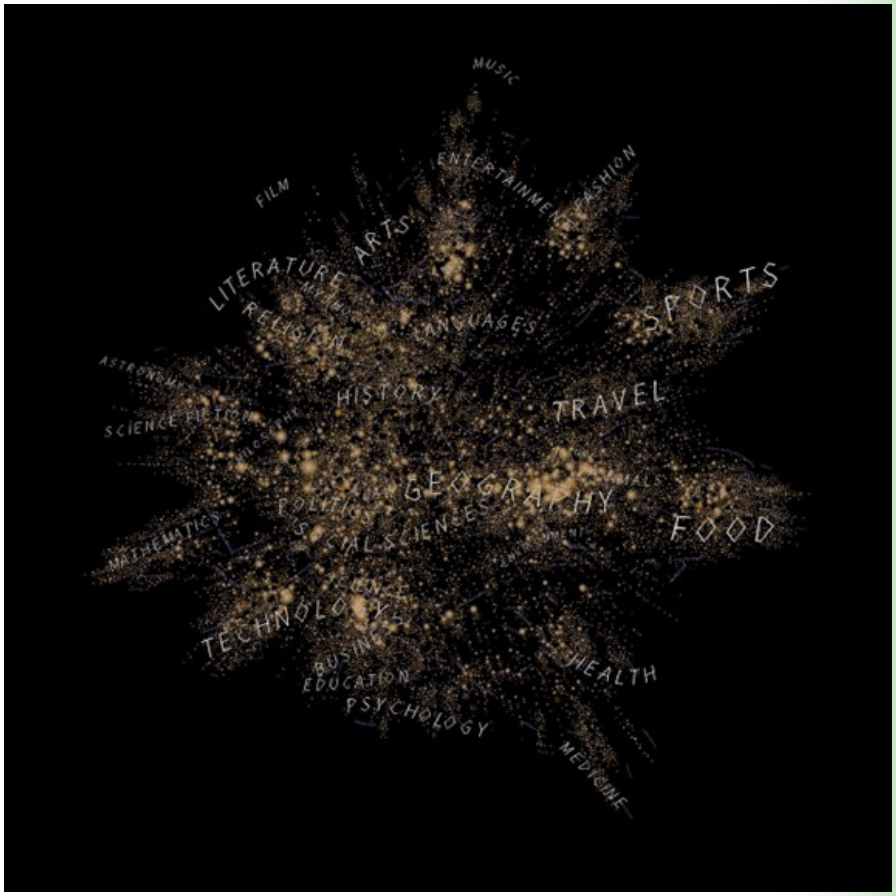
While the map looks a bit like a cumulated map of Wikipedia entries, the training data consists only of a small fragment of text from Wikipedia (3 billion tokens). The vast majority, 410 billion of the total 499 billion tokens that GPT-3 was trained on, comes from a nonprofit organization that has crawled the web, named Common Crawl, since 2008 (Brown et al. 2020). The basis of the learning system behind GPT-3 is texts from the internet. Writing all this content took millions of individuals writing blogs, essays, news stories, etc. For OpenAI, using this content was free of charge. But the web does not only contain the pinnacle of human thought; it includes everything ever published on the

internet, from clusters of vegan turkey with rice paper skin recipes to adult fan fiction on Batman and Robin.

As Time Magazine reported, the company behind GPT, OpenAI, paid Kenyan workers less than $2 per hour to make the system less toxic (Perrigo 2023). Sama, a San Francisco-based firm that employs workers in Kenya, Uganda, and India, labeled text snippets about child sexual abuse, bestiality, murder, suicide, torture, self-harm, and incest. Workers in the Global South are suffering from this traumatic content so that maps such as the one behind this text are not filled with harmful abusive content. OpenAI is disclosing the rules of moderation on their website.

The three forces of the technical infrastructures, the training data, and the post-training moderation do not provide a full image of what GPT is doing, but they lay out an ideology. The results returned by GPT-3.5 are bound by infrastructural costs, the textual source data, and the restrictions set on the system. The cost variations of the three layers are vastly different. The source for the training came to OpenAI without charge; the computation cost at least $4.6 million, and for the moderation, OpenAI paid $2 per hour. GPT-3 is a representation of a system in which computation is the highest pinnacle of the process, predominating the origin and instruction processes. The map is a representation of the textual contents of the internet, with restrictions of computational methods and a filtering system of a company deciding what the system will return and what not.

Fig. 4. Final Mapping of ChatGPT.



Warned Limitations

The data acquisition was mainly a question of what the infrastructure of GPT-3.5 would return. Generating lists of the most important objects and humans within such vast fields as Science Fiction, Beverages, Mammals, or Television can only lead to wrong answers. ChatGPT is set up to account for this impossibility and returned at the end of each request statement such as:

Please note that this is just a sample dataset with randomly chosen categories and subcategories. You can customize it further based on your specific requirements and areas of interest.

Please note that there are numerous individuals who have contributed significantly to the field of architecture, and this list only presents a small subset.

Please note that this is just a sample dataset, and the actual breadth of knowledge I possess is much more extensive.

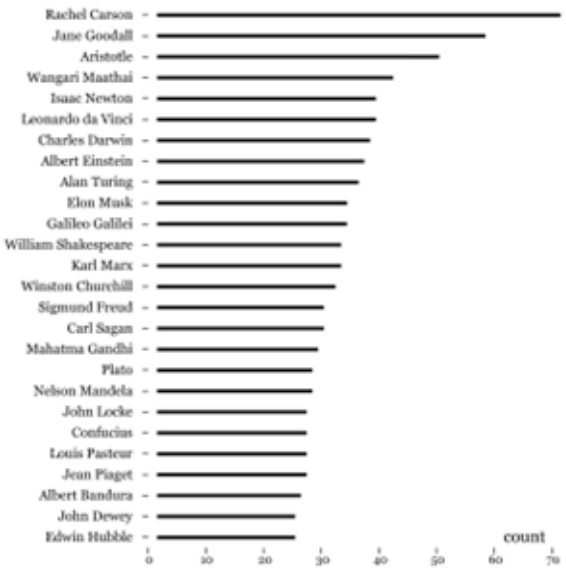


Fig. 5. Most frequently named items in the dataset.

Nevertheless, the system returned entries and thus created a situation in which Aristotle, Arthur Schopenhauer, Immanuel Kant, Plato, and John Dewey are listed in the field of aesthetics within philosophy, and others were not. As these systems become part of the lives and work of billions of humans, what is included and what is excluded matters tremendously. This map is one approach towards finding a reflexive mechanism, a way to investigate what these systems return and what they leave out.

Individual Centralities

One of the dataset's most striking features is simply counting the number of times GPT named things. Figure 5 shows the most frequently named things in the list. First, the list of the most named things consists only of humans. Secondly, the list is led by Rachel Carson and Jane Goodall. Rachel Carson is known for her book Silent Spring (1962) and for advancing the global environmental movement. Jane Goodall is considered the world's foremost expert on chimpanzees. An American marine

biologist, an English primatologist, and an anthropologist are the two most named figures in the project.

In comparison, the Pantheon project (Hidalgo 2014) ranks people, among others, by the number of Wikipedia language editions and count of article clicks. In this ranking, the first female is Mary, mother of Jesus, at rank 33 (accessed on 7th of August 2023). Muhammad, Isaac Newton, and Jesus are the top-ranked figures in the Pantheon project.

The results from GPT-3.5 are more female, more diverse, and less religious than the Pantheon ranking. It is worth noting that ranked fourth is a Kenyan social, environmental, and political activist named Wangari Maathai. It is also worth noting that Rachel Carson, Jane Goodall, and Wangari Maathai all appear in the same cluster on the map around the fields of Geography, Nature, and Environment.

To make sense of these counterintuitive ranking results, it is important to note how the data was generated. Fields and subfields were requested through 1,764 API calls. Rachel Carson was listed 73 times within the 1,764 calls. For a person, object, place, etc., to be named frequently, GPT-3.5 needs to name it in as many combinations of categories and subcategories as possible. Thus, high-ranking results from spreading into many categorical systems.

The question becomes: Are Rachel Carson and Jane Goodall individuals whose research spreads especially well? Research that transcends fields and categories? Or is something else happening here? Might OpenAI set certain parameters that lead to the design of such a list? Are prompt engineers pushing certain perspectives to become more visible? Or is GPT on its way to becoming general artificial intelligence and cares a lot for the planet and the environment? At this point, this is hard to say and would need a much deeper investigation than these preliminary findings.

Project: https://artificial-worldviews.kimalbrecht.com/

References

Brown, Tom B, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, et al.
2020. "Language Models Are Few-Shot Learners." *arXiv*. https://doi.org/10.48550/arxiv.2005.14165.

Carson, Rachel.
1962. *Silent Spring*. Boston: Houghton Mifflin Company.

Hidalgo, César.
2014. "Pantheon – Visualizations." *Pantheon: Explore human collective memory*. http://pantheon.media.mit.edu/treemap/country_exports/NO/all/-4000/2010/H15/pantheon.

Li, Chuan.
2020. "OpenAI's GPT-3 Language Model: A Technical Overview." *Lambda Labs*. June 3, 2020. https://lambdalabs.com/blog/demystifying-gpt-3.

OpenAI.
2024. "API Reference." *OpenAI developer platform*. https://platform.openai.com/docs/api-reference/runs.

Perrigo, Billy.
2023. "OpenAI Used Kenyan Workers on Less Than $2 Per Hour: Exclusive." *Time*. January 18, 2023. https://time.com/6247678/openai-chatgpt-kenya-workers/.