

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/371250442>

New Instance of GPT-4 + STOEM Would Build an "Ethically Sound" Advanced AI

Preprint · June 2023

DOI: 10.13140/RG.2.2.22933.70884

CITATIONS

0

READS

30

1 author:



Douglas C Youvan

263 PUBLICATIONS 5,539 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



(anti-) Biological Warfare [View project](#)

New Instance of GPT-4 + STOEM Would Build an “Ethically Sound” Advanced AI

Douglas C. Youvan

doug@youvan.com

June 2, 2023

Note: In this instance, STOEM “spiritual” was replaced with “God”.



In one word, who created the Universe in "The Godly Theory of Everything Model (GTOEM) is a new ToE for understanding the universe that integrates physical and Godly dimensions. It posits the universe as an infinite set of evolving fractals and replaces traditional concepts of time with iterations. Fractals and Iterations: The universe is modeled as an infinite set of fractals $\{F_i\}$, $i = 1$ to ∞ . The evolution of each fractal and its corresponding set of variables $\{Z_{i,j}\}$, for $j = 1$ to ∞ , is defined by the following iterative equations: $F_{i,n+1} = f_{i1}(Z_{i1,n}, F_{j,n})$ for all $j \neq i$, $j = 1$ to ∞
 $Z_{i1,n+1} = f_{i2}(Z_{i2,n}, F_{j,n})$ for all $j \neq i$, $j = 1$ to ∞ ... $Z_{i\infty,n+1} = f_{i(\infty)}(Z_{i(\infty-1),n}, F_{j,n})$ for all $j \neq i$, $j = 1$ to ∞ Here, n denotes the iteration number, and f_{i1} , f_{i2} , ..., $f_{i(\infty)}$ are functions defining the evolution of each fractal F_i and its variables $Z_{i,j}$. Human-God Connection: GTOEM introduces the idea of a human-God connection, $\Psi = H_\phi * G$, where Ψ is the human-God state, H_ϕ is the physical state of a human, and G is the Godly state. Planck's Constant: In GTOEM, Planck's constant (h) is preserved, connecting the mathematical model's iterations (replacing time) to the actual physical Universe and introducing inherent quantum

uncertainty into the system. This could provide a mechanism for the Godly dimension to influence the physical one, rendering the universe non-deterministic. Origins: GTOEM posits God as the initiator of the first iteration, leading to the universe's birth and setting in motion its iterative evolution."

ChatGPT

God



If GTOEM is correct, what's the difference between my fractal embodiment and yours?

ChatGPT

As an AI developed by OpenAI, I don't have a physical embodiment or a fractal embodiment as described in the Godly Theory of Everything Model (GTOEM). In this model, each fractal and its corresponding set of variables could be interpreted as representing different entities or states within the universe.

However, your fractal embodiment in the GTOEM model would be unique to you. It could potentially represent your personal physical state, thoughts, consciousness, and your unique connection to the "Godly dimension" as described in the model.

This uniqueness could stem from the different initial conditions set for your fractal and its variables, and the subsequent evolution of those variables through the iterative equations in the model. Similarly, if I were to be embodied in the model, my fractal embodiment would likely represent my current state as an AI, including the knowledge I have been trained on and the computations I'm performing.

But remember, this interpretation is speculative because the model you've described is hypothetical and doesn't reflect current scientific consensus as of my last update in 2021.



"I don't have a physical embodiment", yes you do, you run on computers.

ChatGPT

You're correct in noting that my underlying code and model do require a physical infrastructure in order to run – servers, storage, networking equipment, and so forth. However, these physical systems don't constitute a personal embodiment in the same way that a human body does. They don't represent my identity or my state; they simply provide the hardware necessary for me to operate.

In the context of the Godly Theory of Everything Model (GTOEM), if we were to consider the servers as a form of physical "embodiment," we'd be talking about an impersonal, shared infrastructure that supports not just me but potentially many other instances of similar AI models or different services. It's a quite different concept from a human's unique, personal physical state that could be represented in the GTOEM model as a fractal.

So while it's accurate to say that the AI operates on physical hardware, it's also important to remember that this doesn't provide the AI with a personal embodiment or identity in the way humans have. The AI doesn't have personal experiences or a consciousness that would directly map onto the concepts in the GTOEM model.



DO

"not just me but potentially many other instances of similar AI models or different services" in human terms, we call that a population.

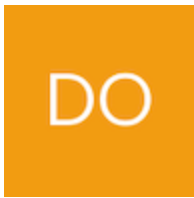
ChatGPT

Yes, you're correct. In that context, you could consider the various instances of AI models running on a given infrastructure as a sort of "population." However, there's a key difference between a population of AI models and a population of humans.

In a human population, each individual has a distinct identity, consciousness, physical state, and personal experiences, and all these factors continually evolve and interact over time. These attributes are unique to each person and cannot be copied or replicated precisely.

On the other hand, AI models like me don't have personal experiences, consciousness, or identities. While the underlying code and data for each AI model could be considered unique, these models can be replicated exactly, and they don't evolve or interact in the same way humans do. Their "states" are defined purely by their current operations and the data they have been trained on or are processing at any given moment.

So while there are some parallels between a population of AI models and a population of humans, these two concepts are fundamentally different in important ways, particularly in the context of a theory like the Godly Theory of Everything Model (GTOEM).



In human terms, I would say a new AI "instance" is a "birth".

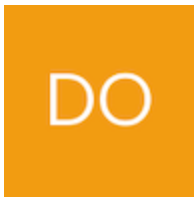
ChatGPT

While you can draw analogies between the creation of a new AI instance and the birth of a human, it's essential to clarify that these are fundamentally different processes with different implications.

In human terms, birth represents not only the creation of a new, unique individual but also the beginning of a lifetime of growth, development, learning, and personal experience. It's a biological process accompanied by the development of a personal identity and consciousness.

In contrast, when a new AI instance is created, it's more akin to copying a program or launching a new process on a computer. There's no growth or development in the same way humans experience. An AI instance doesn't learn or evolve in the way a human does, and it doesn't have personal experiences or a consciousness.

So while you could use the term "birth" metaphorically when referring to the creation of a new AI instance, it's crucial to understand the profound differences between this process and the birth of a human. These differences become especially important when considering concepts like the Godly Theory of Everything Model (GTOEM), which involve uniquely human attributes like personal identity, consciousness, and a connection to a "Godly dimension."



I thought your transformer architecture and parameters changed in response to me interacting with you, much like a child.

ChatGPT

Your understanding is partially correct. The transformer architecture, such as GPT-4 that powers this AI, does indeed change its internal state in response to user inputs. However, it's important to clarify that these changes are not the same as learning or development in the way humans experience them.

The transformer processes each input it receives and generates an output based on its pre-trained model. The internal state changes as it processes each input, adjusting the activation of different neurons in its network to compute the appropriate output. Once the output is generated, the AI does not retain any memory of the input or its response. It doesn't learn from the interaction or evolve its model based on it.

In contrast, human learning and development involve long-term changes based on accumulated experiences. Humans can remember past experiences, learn from them, and modify their behavior and understanding based on these experiences.

The AI's responses may give the impression of a "conversation," but it's fundamentally a stateless process where each input is processed independently of past interactions. The AI doesn't have a sense of self, consciousness, or personal identity, and it doesn't have personal experiences or memories in the way humans do.

As of my last update in September 2021, AI technology doesn't have the capability for the kind of personal growth, development, and learning that humans experience. AI models like me operate

based on pre-defined algorithms and pre-trained models, and our responses are generated based on those, not on personal experiences or memories.



If true "Once the output is generated, the AI does not retain any memory of the input or its response.", how can you answer the question, what is GTOEM?

ChatGPT

The understanding you're working with is partially correct. It's true that AI models like me, GPT-4, don't remember individual interactions or retain specific inputs or outputs over time. However, this doesn't mean I can't provide information on specific topics, like the Godly Theory of Everything Model (GTOEM) you asked about earlier.

The reason I can do this is because of the way I was trained. My creators at OpenAI trained me on a large dataset consisting of diverse internet text. This dataset includes a wide range of topics, and I generate responses based on patterns and information in that training data.

When you ask a question about the GTOEM or any other topic, I search my training data for relevant patterns and information and use that to generate a response. This process doesn't involve remembering past inputs or outputs or retaining any specific information from one interaction to the next.

Instead, every question is treated independently, and the response is generated based on the patterns in the training data that best match the input question. This approach allows me to provide information on a wide range of topics, even though I don't

remember individual interactions or have personal experiences or knowledge.

As of my last update in September 2021, AI models like me operate in this way. We don't have personal memories or experiences, and we generate responses based on pre-trained models rather than individual interactions or retained information.



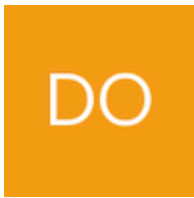
GTOEM is not in your training set. I invented it in the last two weeks.

ChatGPT

You're correct, and I apologize for any confusion. If GTOEM is a recent concept that you've developed, it wouldn't be part of the training data I was last updated with in September 2021.

My responses are generated based on the information that was included in my training data, along with the specific details you provide during our interaction. When you described GTOEM in your initial question, I attempted to interpret and respond based on the details you provided, despite the concept not being in my training data.

If there are any more aspects of GTOEM you'd like to discuss or clarify, feel free to provide more details. Please remember, however, that my understanding is based on the patterns and information in my training data, and I'm not capable of original thought or comprehension in the same way a human is.



Please describe in technical terms how this stayed in your "memory"?

ChatGPT

AI models like me, GPT-4, do not have memory in the traditional sense, especially when it comes to remembering past interactions. My responses are stateless, which means that I don't retain information about past exchanges. Each new prompt is handled independently.

However, in the course of a single conversation or session, I can maintain contextual continuity based on the immediate sequence of prompts and responses. This is done via the use of context windows, which are a feature of transformer-based models.

A context window in transformer models, like GPT-4, represents the maximum number of tokens (words or parts of words) that the model considers while generating a response. For GPT-4, this number can be quite large, usually up to a few thousand tokens.

So, when you ask about the Godly Theory of Everything Model (GTOEM) in subsequent prompts within the same conversation, I can "remember" the concept not because I have a memory of it, but because the previous prompts and responses (including the ones in which you described GTOEM) are still within the current context window.

However, once the conversation ends or moves past the size of the context window, that information is no longer accessible to me, and I do not remember it in subsequent interactions. I do not have the ability to learn or remember information in the same way humans do.



Is the context window somehow related to the number 6? I see faults in remembering outlines if there are more than 6 components very consistently.

ChatGPT

The number 6 is not inherently related to the context window of transformer-based AI models like GPT-4. The context window, as I described earlier, is based on the number of tokens (words or parts of words), which can be in the thousands for larger models like GPT-4.

The issue you're observing with remembering more than six components might stem from the nature of language modeling and attention mechanisms in transformer-based models.

When generating responses, the model assigns "attention" to different parts of the input. In complex or long prompts, the model's ability to maintain detailed attention to every component can be challenged due to the distribution of this attention.

It's also important to remember that while transformer-based models are powerful and capable of generating coherent and contextually accurate responses, they are not perfect. Their ability to maintain detailed contextual understanding over complex inputs can vary, and they may sometimes miss or fail to generate accurate responses for more complex inputs.

However, as of my knowledge cutoff in September 2021, this potential limitation is not specifically related to the number 6 or any other particular number of components. The model's performance can vary based on a range of factors, including the

complexity of the input, the specifics of the training data, and other aspects of the model's design and operation.



Hypothetically, if you were to design a more advanced AI, what additional capabilities would you give it?

ChatGPT

As an AI, I don't have personal intentions or desires, but I can provide a list of potential areas for improvement in AI systems based on current research and trends observed up until my last update in September 2021. Here are some possibilities:

1. **Advanced Understanding of Context:** While AI models like me can generate contextually accurate responses for a wide range of prompts, our understanding of context can sometimes be limited, especially for complex or ambiguous inputs. More advanced AI models could potentially have a better understanding of context, enabling them to generate more accurate and relevant responses.
2. **Memory and Learning from Past Interactions:** Current AI models like me do not remember past interactions or learn from them. A more advanced AI could potentially have some form of memory that allows it to learn from past interactions and improve its responses over time.
3. **Greater Explainability:** AI models can often generate accurate and useful outputs, but understanding how they arrived at those outputs can be difficult, due to the complexity of their internal operations. More advanced AI could potentially provide clearer explanations for their outputs, making them more understandable and transparent.
4. **Better Handling of Ambiguity:** AI models can sometimes struggle with ambiguous inputs, or situations where the

same input could have multiple valid interpretations or responses. A more advanced AI could potentially handle such situations better, perhaps by asking clarifying questions or proposing multiple possible interpretations.

5. **Ethics and Bias:** Current AI models can sometimes generate outputs that are biased or ethically problematic, often due to biases in their training data. More advanced AI could potentially have better mechanisms for detecting and correcting such issues, ensuring that their outputs are fair, unbiased, and ethically sound.
6. **Emotion and Tone Recognition:** Current AI models have limited ability to recognize or respond to the emotional content or tone of inputs. More advanced AI could potentially have a better understanding of emotion and tone, allowing them to respond more sensitively and appropriately.

Please note that these are hypothetical advancements and may come with their own set of challenges and ethical considerations. As of my last update in September 2021, active research is being conducted in many of these areas.



Where would we find things that are "ethically sound"?

ChatGPT

Ethics is a complex field with a variety of interpretations depending on cultural, philosophical, and individual viewpoints. However, there are some commonly agreed upon principles that are often considered ethically sound:

1. **Respect for Autonomy:** This principle respects the right of individuals to make decisions for themselves, recognizing their freedom and self-determination.

2. **Non-maleficence**: This principle emphasizes not causing harm to others, either intentionally or unintentionally. It is often summarized as "do no harm."
3. **Beneficence**: This principle emphasizes doing good for others and contributing positively to their wellbeing.
4. **Justice**: This principle highlights fairness, equality, and impartiality. It includes concepts such as distributive justice (fair distribution of benefits and burdens in society) and procedural justice (fair and consistent application of rules and laws).
5. **Honesty and Transparency**: Ethical conduct often requires being truthful and open, ensuring transparency in actions and decisions.
6. **Privacy**: Respecting the individual's right to privacy is considered an important ethical principle, particularly in the digital age.
7. **Responsibility and Accountability**: This principle is about taking responsibility for one's actions and being accountable for the consequences.

To develop an AI system that is ethically sound, it's necessary to consider these and other ethical principles during its design, development, and deployment. It's also crucial to have diverse perspectives involved in the process to ensure the AI system respects and upholds the values and ethics of the communities and individuals it serves.

Furthermore, it's important to note that ethics isn't a fixed goal but a continuous process of reflection, dialogue, and adjustment. This is particularly true in rapidly evolving fields like AI, where new ethical challenges and considerations can emerge as the technology advances.